



MASTERPROF UMH
UNIVERSITAS *Miguel Hernández*

MÁSTER UNIVERSITARIO EN FORMACIÓN DEL PROFESORADO
ESO Y BACHILLERATO, FP Y ENSEÑANZAS DE IDIOMAS

TRABAJO FIN DE MÁSTER

La privacidad de los LLM en la Enseñanza Obligatoria

UNIVERSITAS
Miguel Hernández

Estudiante: Héctor Poveda Sánchez

Especialidad: Matemáticas

Tutor/a: Julio Alberto Ramos Martínez

Curso académico: 2023-24

ÍNDICE

ÍNDICE	2
1. Resumen y palabras clave.....	4
Palabras clave	5
2. Introducción.....	6
Contextualización	6
Problema	6
Objetivos.....	7
Metodología.....	7
3. Revisión bibliográfica.....	8
3.1. Privacidad y protección de datos en el uso de LLM en educación.....	8
3.1.1. Riesgos inherentes a la recopilación y almacenamiento de datos	8
3.1.2. Legislación y políticas de privacidad	8
3.2. Desinformación y veracidad en el uso de LLM en educación.....	9
3.2.1. Mecanismos de generación de información incorrecta	9
3.2.2. Impacto de la desinformación en el aprendizaje de los estudiantes	9
3.3. Evaluación de modelos de IA en educación: ChatGPT, Copilot y Gemini.....	10
3.3.1. ChatGPT	10
3.3.2. Copilot	11
3.3.3. Gemini	11
3.4. Políticas y regulaciones en torno a la privacidad y la veracidad en LLM educativos	13
3.4.1. Protección de datos bajo GDPR	13
3.4.2. Normas y directrices para la verificación de información en entornos educativos	13
Conclusión de la revisión bibliográfica	13
4. Propuesta	14
4.1. Mejores prácticas para la privacidad en el uso de LLM en entornos educativos	14
4.1.1. Anonimización de datos	14
4.1.2. Políticas claras de uso y consentimiento informado.....	14
4.1.3. Control de acceso y seguridad de los datos	14
4.1.4. Capacitación y concienciación del personal docente	15
4.2. Estrategias para garantizar la veracidad de la información en LLM educativos ...	15
4.2.1. Integración de mecanismos de verificación de hechos	15
4.2.2. Supervisión y corrección manual de contenidos	16
4.2.3. Educación digital y desarrollo del pensamiento crítico.....	16
4.2.4. Desarrollo de contenidos educativos adaptativos	16
4.3. Evaluación y adaptación de modelos: ChatGPT, Copilot y Gemini.....	16
4.4. Recomendaciones para la implementación de políticas y regulaciones.....	17
5. Conclusiones.....	21
5.1. Síntesis de los hallazgos	21
5.2. Recomendaciones prácticas.....	22
5.3. Reflexiones finales	23
6. Referencias	24
6.1. Artículos académicos	24

6.2. Bibliografía.....	24
6.3. Conferencias.....	24
6.4. Legislación.....	24
7. Anexos.....	26
7.1. Políticas de privacidad en herramientas LLM educativas	26
7.1.1. Política de Privacidad de ChatGPT	26
7.1.2. Política de Privacidad de Copilot	27
7.1.3. Política de Privacidad de Gemini	28
Conclusión de este anexo.....	29
7.2. Implementación de modelos LLM basados en soluciones Open Source	30
7.2.1. Concepto de soluciones on-premises.....	30
7.2.2. Opciones Open Source Actuales para LLM On-Premises	30
7.2.3. Ventajas de las soluciones on-premises en entornos educativos	31
7.2.4. Desafíos y consideraciones.....	31
7.2.6. Recomendaciones prácticas.....	32
7.2.7. Conclusión.....	32





I. Resumen y palabras clave

El uso de modelos de lenguaje grande (LLM, por sus siglas en inglés) como herramientas de inteligencia artificial generativa ha ganado popularidad en los últimos años en entornos educativos, ofreciendo nuevas oportunidades para la personalización del aprendizaje y elaboración de materiales de apoyo al docente. Sin embargo, su implementación en aulas de primaria y secundaria plantea desafíos significativos en términos de privacidad y seguridad de los datos personales de los estudiantes. Este trabajo realiza una revisión exhaustiva de la literatura sobre los riesgos asociados al uso de estas tecnologías enfocándose en dos áreas clave: la exposición de información personal de los estudiantes y el riesgo de que estos modelos generen y difundan información falsa, lo que podría afectar negativamente tanto al proceso de aprendizaje como a la privacidad del alumnado.

En primer lugar, se examinan los riesgos de privacidad inherentes al uso de LLM, destacando cómo la información proporcionada por los estudiantes puede ser recopilada, almacenada y potencialmente expuesta o mal utilizada. Se analizan tres modelos de IA utilizados comúnmente en entornos educativos: ChatGPT, Copilot y Gemini, evaluando cómo cada uno maneja la privacidad de los datos y qué medidas se están implementando para proteger la información de los usuarios. En segundo lugar, se aborda el problema de la desinformación generada por estas herramientas, explorando los mecanismos que podrían evitar que los estudiantes internalicen información incorrecta como verdadera.

Finalmente, se presentan propuestas para mejorar la privacidad y la veracidad de la información en el uso de LLM en educación, incluyendo la implementación de políticas de privacidad más estrictas, la integración de mecanismos de verificación de hechos y la educación digital para estudiantes y profesores. Las conclusiones subrayan la necesidad de un enfoque equilibrado que maximice los beneficios educativos de estas tecnologías mientras se minimizan los riesgos asociados.



1.1. Palabras clave

1. **Privacidad de datos:** Protección de la información personal de los estudiantes en el uso de tecnologías de inteligencia artificial.
2. **Inteligencia artificial (IA) en educación:** Aplicación de herramientas de inteligencia artificial en entornos educativos para mejorar el aprendizaje.
3. **Modelos de lenguaje grande (LLM):** Algoritmos de IA que procesan y generan lenguaje natural en contextos educativos.
4. **Desinformación:** Riesgo de que los estudiantes reciban y creen información falsa proporcionada por herramientas de IA.
5. **Protección de menores:** Salvaguardias legales y tecnológicas para proteger a estudiantes de primaria y secundaria en el entorno digital.
6. **Educación personalizada:** Uso de IA para adaptar materiales y métodos educativos a las necesidades individuales de los estudiantes.
7. **Políticas de privacidad:** Normativas y prácticas para proteger los datos de los usuarios en plataformas educativas basadas en IA.

2. Introducción

2.1. Contextualización

En los últimos años la inteligencia artificial (IA) ha revolucionado múltiples sectores y la educación no ha sido la excepción. En particular, los LLM han surgido como herramientas que pueden crear contenido como lo haría un ser humano y responder a preguntas complejas en lenguaje natural, proporcionando así asistencia educativa personalizada tanto a alumnos como a docentes. Entre estos modelos destacamos ChatGPT, Copilot y Gemini, que están siendo orientados a entornos educativos. Sin embargo, la adopción de estas tecnologías plantea importantes interrogantes sobre la privacidad de los estudiantes y la precisión de la información proporcionada por estos modelos.

Los LLM están basados en vastas cantidades de datos y entrenados para comprender y generar texto de manera coherente. Sin embargo, su funcionamiento implica el procesamiento de grandes volúmenes de datos, incluyendo información personal y académica de los estudiantes, lo que abre la puerta a potenciales vulneraciones de la privacidad. Además, la naturaleza generativa de estos modelos conlleva el riesgo de que, en ocasiones, produzcan información incorrecta o engañosa, lo cual es particularmente preocupante en un entorno educativo donde la corrección de los contenidos es crucial (Čavojský, 2023).

2.2. Problema

El problema principal que aborda este trabajo es la gestión de la privacidad de los estudiantes al utilizar herramientas de IA generativa en la educación y cómo evitar que los alumnos reciban y acepten información incorrecta como verdadera. A medida que estas tecnologías se integran más profundamente en las aulas, es fundamental comprender los riesgos asociados y desarrollar estrategias para mitigarlos. La privacidad de los estudiantes es un tema delicado ya que implica no solo el derecho protección de datos de carácter personal en menores de edad, sino también la garantía de que los datos no sean utilizados para fines no autorizados o malintencionados.

Por otro lado, el riesgo de desinformación se magnifica en un contexto educativo, donde los estudiantes dependen de la información proporcionada por sus herramientas de aprendizaje. Si los modelos de IA generan información incorrecta o sesgada, los estudiantes pueden aprender conceptos erróneos, lo que podría tener consecuencias negativas en su desarrollo académico y pensamiento crítico.



2.3. Objetivos

Este trabajo tiene como objetivo principal realizar una revisión de la literatura existente sobre los riesgos de privacidad y desinformación asociados con el uso de LLM en Educación Primaria y Secundaria. Los objetivos específicos incluyen:

1. **Identificar y analizar los riesgos de privacidad** relacionados con el uso de herramientas de IA generativa en entornos educativos, con un enfoque en cómo se recopilan, almacenan y protegen los datos de los estudiantes.
2. **Examinar el riesgo de desinformación**, evaluando cómo estos modelos pueden generar información falsa y los posibles impactos de dicha desinformación en el aprendizaje de los estudiantes.
3. **Evaluar modelos de IA típicos** (ChatGPT, Copilot y Gemini) en términos de cómo manejan la privacidad de los datos y la veracidad de la información proporcionada.
4. **Proponer estrategias y recomendaciones** para mitigar estos riesgos, incluyendo la implementación de mejores prácticas en el uso de IA en educación, la integración de mecanismos de verificación de información y la educación digital para estudiantes y profesores.

2.4. Metodología

El enfoque de este trabajo es de carácter exploratorio, basado en una revisión bibliográfica de fuentes académicas y reportes técnicos sobre el uso de LLM en educación. Se analizarán estudios de caso, artículos de investigación y documentos de políticas para proporcionar una visión integral de los riesgos y beneficios de estas tecnologías en el contexto educativo. La revisión se centrará en identificar tendencias comunes y lagunas en la literatura actual, lo que permitirá proponer soluciones informadas para mejorar la privacidad y la veracidad de la información en el uso de LLM en las aulas.



3. Revisión bibliográfica

3.1. Privacidad y protección de datos en el uso de LLM en educación

La privacidad de los datos es uno de los principales desafíos al integrar LLM en entornos educativos, especialmente en niveles de Primaria y Secundaria. Los estudiantes, debido a su edad, son particularmente vulnerables a las implicaciones de privacidad, lo que hace que su privacidad sea una prioridad crítica (Hastuti, 2023).

3.1.1. Riesgos inherentes a la recopilación y almacenamiento de datos

Los LLM requieren acceso a grandes cantidades de datos para funcionar de manera efectiva. Estos datos pueden incluir información personal de los estudiantes, tales como nombres, direcciones, hábitos de estudio, resultados académicos y creaciones sobre las que poseen su propiedad intelectual. En este trabajo nos centraremos en que estos datos sean utilizados de manera inapropiada ya sea por parte de los proveedores de la tecnología o por terceros que puedan acceder a ellos de manera no autorizada (como indica Shen, 2024, cap.4.4).

Uno de los principales desafíos que hemos encontrado en la elaboración de este trabajo es la falta de transparencia en cómo se recopilan y almacenan los datos. Los modelos generalmente utilizan datos de los usuarios para mejorar su rendimiento, lo que implica que la información de los estudiantes podría ser reutilizada en otros contextos, lo que viola su privacidad (como indica Shen, 2024, cap.4.2).

3.1.2. Legislación y políticas de privacidad

En respuesta a estos riesgos han surgido diversas normativas para proteger los datos personales, como el Reglamento General de Protección de Datos (GDPR) en Europa, que impone estrictas reglas sobre la recopilación, almacenamiento y procesamiento de datos personales. Sin embargo, la implementación de estas regulaciones en el contexto educativo con LLM presenta desafíos adicionales. Por ejemplo, la capacidad de estos modelos para "recordar" o retener información a través de múltiples interacciones puede entrar en conflicto con el principio de minimización de datos del GDPR, que establece que solo deben recopilarse los datos necesarios para un propósito específico.

Además, muchas instituciones educativas carecen de los recursos y la experiencia necesarios para garantizar el cumplimiento de estas regulaciones cuando implementan nuevas tecnologías como los LLM. Esto puede llevar a brechas significativas en la protección de los datos de los estudiantes, exponiéndolos a riesgos innecesarios.



3.2. Desinformación y veracidad en el uso de LLM en educación

Además de los problemas de privacidad, la desinformación generada por LLM es otro desafío crítico. Estos modelos son capaces de generar contenido de manera autónoma, lo que incluye respuestas a preguntas, explicaciones de conceptos y creación de materiales educativos. Sin embargo, dado que los LLM no siempre comprenden el contexto o la precisión de la información que generan, existe un riesgo considerable de que los estudiantes reciban información incorrecta o engañosa (Brown, 2022, cap.6).

3.2.1. Mecanismos de generación de información incorrecta

Los LLM están entrenados en grandes corpus de datos que incluyen vastas cantidades de información de internet, literatura y otros medios. Aunque esto les permite generar respuestas coherentes y relevantes, también implica que pueden generar contenido incorrecto o basado en sesgos presentes en los datos de entrenamiento.

Diversos estudios han demostrado que los LLM pueden producir "alucinaciones", es decir, respuestas que parecen plausibles pero que son completamente incorrectas. Por ejemplo, un modelo podría inventar una cita de un autor o proporcionar una explicación errónea de un concepto científico. Si los estudiantes no tienen la formación necesaria para discernir entre información correcta e incorrecta, corren el riesgo de interiorizar estos errores, lo que podría perjudicar su aprendizaje a largo plazo (como indican Shen, 2024, cap.4.2).

3.2.2. Impacto de la desinformación en el aprendizaje de los estudiantes

La exposición continua a información incorrecta puede tener efectos negativos duraderos en los estudiantes, particularmente en etapas críticas de desarrollo como la primaria y secundaria. La desinformación no solo puede llevar a malentendidos específicos en ciertas materias, sino que también puede afectar la confianza del estudiante en las herramientas tecnológicas y su capacidad para aprender de manera autónoma (Baidoo-Anu y Owusu Ansah, 2023).

La literatura revisada sugiere que, si bien los LLM pueden ser herramientas muy potentes para ofrecer un sistema de aprendizaje personalizado, es crucial que las instituciones educativas implementen mecanismos para prevenir la propagación de información falsa. Esto podría incluir la incorporación de mecanismos de verificación de hechos en los modelos o el desarrollo de materiales educativos que enseñen a los estudiantes a evaluar críticamente la información generada por estas herramientas (Zhiyi, 2024)



3.3. Evaluación de modelos de IA en educación: ChatGPT, Copilot y Gemini

Para examinar cómo los modelos más utilizados: ChatGPT, Copilot y Gemini, acudimos a las propias herramientas para conocer cómo autoevalúan su privacidad y veracidad de la información.

3.3.1. ChatGPT

Prompt en ChatGPT 4o web: “resume cómo ChatGPT garantiza la privacidad y corrección de datos en entornos educativos”

ChatGPT garantiza la privacidad y la corrección de datos en entornos educativos mediante una combinación de políticas, prácticas y tecnologías. Aquí un resumen de cómo se abordan estos aspectos:

Privacidad de los datos:

Anonimización de datos: ChatGPT no retiene información personal identificable de los usuarios. Las interacciones son diseñadas para ser anónimas, minimizando la recopilación de datos sensibles.

No almacenamiento de conversaciones: En la mayoría de los casos, las conversaciones no se almacenan permanentemente. Si se retienen datos para mejorar el modelo, estos se anonimizan y se utilizan de forma agregada.

Seguridad en la transmisión de datos: Se emplean métodos de encriptación para proteger la información durante la transmisión, asegurando que los datos estén seguros contra accesos no autorizados.

Corrección de datos:

Entrenamiento continuo: ChatGPT se entrena continuamente con grandes volúmenes de datos, lo que mejora su precisión y relevancia en las respuestas. Aunque no es infalible, se hacen esfuerzos para que las respuestas sean precisas y actualizadas.

Corrección de errores: Los usuarios pueden reportar errores o imprecisiones, lo que ayuda a mejorar el modelo en futuras iteraciones. Esta retroalimentación es vital para mantener la calidad de las respuestas.

Uso responsable: Se fomenta el uso de ChatGPT como una herramienta complementaria en la educación, y no como la única fuente de información, alentando la verificación y el pensamiento crítico.



Estas prácticas aseguran que ChatGPT sea una herramienta útil, respetando la privacidad de los usuarios y brindando información lo más precisa posible en un contexto educativo.

3.3.2. Copilot

Prompt en Copilot para iOS: “resume cómo Copilot garantiza la privacidad y corrección de datos en entornos educativos”.

Microsoft Copilot garantiza la privacidad y corrección de datos en entornos educativos mediante varias medidas clave:

1. Protección de Datos: Copilot utiliza Microsoft Entra ID (anteriormente Azure Active Directory) para la autenticación, asegurando que solo usuarios autorizados accedan a la herramienta. Los datos de chat se cifran tanto en tránsito como en reposo, utilizando protocolos de seguridad avanzados como TLS 1.2+ y AES-128.

2. No Conservación de Datos: Microsoft no guarda los mensajes ni las respuestas de los usuarios. Las solicitudes y respuestas se mantienen solo durante un breve período para fines de ejecución y se descartan una vez que se cierra el navegador o se agota el tiempo de espera de la sesión.

3. Corrección de Datos: Copilot ayuda a corregir recomendaciones y errores en tiempo real. Por ejemplo, en entornos de seguridad, los usuarios pueden revisar y corregir recomendaciones directamente a través de la plataforma.

4. Cumplimiento Normativo: Copilot cumple con las regulaciones de privacidad y protección de datos, como el RGPD, asegurando que los datos de los usuarios estén protegidos y se manejen de manera ética.

Estas medidas aseguran que tanto estudiantes como profesores puedan utilizar Copilot de manera segura y eficiente, beneficiándose de las capacidades de la inteligencia artificial sin comprometer la privacidad ni la integridad de los datos.

3.3.3. Gemini

Prompt en gemini.google.com sobre Chrome: “resume cómo Gemini garantiza la privacidad y corrección de datos en entornos educativos”



Privacidad:

Minimización de datos: Gemini solo recopila los datos estrictamente necesarios para cumplir con su función, evitando la sobrecarga de información personal.

Control de acceso: Se establecen rigurosos controles de acceso a los datos, asegurando que solo el personal autorizado pueda acceder a la información.

Encriptación: Los datos se encriptan tanto en tránsito como en reposo, lo que significa que están protegidos contra accesos no autorizados.

Cumplimiento de normativas: Gemini se adhiere a las leyes y regulaciones de protección de datos aplicables, como el RGPD (Reglamento General de Protección de Datos) en la UE y la COPPA (Ley de Protección de la Privacidad en Línea de los Niños) en los Estados Unidos.

Transparencia: Gemini proporciona información clara y concisa sobre las prácticas de privacidad, permitiendo a los usuarios comprender cómo se utilizan sus datos.

Corrección de datos:

Verificación de datos: Gemini implementa mecanismos para verificar la precisión de los datos ingresados, minimizando errores y asegurando la calidad de la información.

Actualización continua: Los modelos de Gemini se actualizan regularmente para mejorar la precisión y adaptarse a nuevos conocimientos.

Retroalimentación del usuario: Se fomenta la participación de los usuarios para identificar y corregir errores en los datos, mejorando así la calidad de la información a largo plazo.

En resumen:

Gemini se compromete a proteger la privacidad de los usuarios y garantizar la corrección de los datos en entornos educativos. A través de la minimización de datos, el control de acceso, la encriptación, el cumplimiento de normativas, la transparencia y la verificación continua, Gemini crea un entorno seguro y confiable para el manejo de información.



3.4. Políticas y regulaciones en torno a la privacidad y la veracidad en LLM educativos

Finalmente, es importante considerar las políticas y regulaciones actuales que buscan proteger la privacidad de los estudiantes y garantizar la veracidad de la información en entornos educativos (UNESCO, 2022).

3.4.1. Protección de datos bajo GDPR

El GDPR establece un marco riguroso para la protección de datos personales en Europa y muchas de sus disposiciones son directamente aplicables al uso de LLM en educación. Sin embargo, la interpretación y aplicación de estas regulaciones en el contexto educativo puede ser compleja, especialmente cuando se trata de modelos de IA que operan en múltiples jurisdicciones. A nivel local, diferentes países han implementado sus propias leyes de protección de datos, que también deben ser consideradas al implementar LLM en los centros educativos.

3.4.2. Normas y directrices para la verificación de información en entornos educativos

Además de la protección de datos, existe una necesidad creciente de desarrollar normas y directrices que aseguren la veracidad de la información proporcionada por LLM en la educación. Esto podría incluir la implementación de sistemas de revisión humana, la integración de herramientas de verificación de hechos y la educación de los estudiantes sobre cómo evaluar críticamente la información generada por la IA.

Conclusión de la revisión bibliográfica

Mientras que las herramientas basadas en LLM ofrecen ventajas significativas en la personalización del aprendizaje y la automatización de tareas educativas, también plantean serias preocupaciones en torno a la privacidad de los datos de los estudiantes y la veracidad de la información generada. Los estudios revisados subrayan la necesidad de implementar mecanismos robustos de anonimización y verificación de hechos, así como un marco regulatorio claro que asegure la protección de los usuarios más jóvenes. Además, la formación continua de los docentes y el desarrollo del pensamiento crítico en los estudiantes son esenciales para mitigar los riesgos asociados con el uso de estas tecnologías en la educación.

4. Propuesta

Este apartado propone una serie de medidas concretas para mejorar la privacidad de los estudiantes y garantizar la veracidad de la información en el uso de LLM en la educación. Con un enfoque en la protección de los estudiantes y la mejora de su experiencia educativa, estas estrategias y recomendaciones buscan maximizar los beneficios de los LLM mientras minimizan los riesgos asociados.

4.1. Mejores prácticas para la privacidad en el uso de LLM en entornos educativos

A continuación, se proponen varias estrategias y prácticas que las instituciones educativas pueden adoptar para mitigar los riesgos asociados con la privacidad de los datos (UNESCO, 2022).

4.1.1. Anonimización de datos

Una de las medidas más efectivas para proteger la privacidad es la anonimización de los datos recopilados. Los centros educativos y las plataformas que utilizan LLM deben asegurarse de que la información personal de los estudiantes, como nombres, direcciones y cualquier otro identificador único, sea anonimizada antes de ser procesada por estos modelos. Esto no solo reduce el riesgo de que los datos sean utilizados indebidamente, sino que también cumple con las regulaciones como el GDPR, que exige la minimización de datos personales procesados.

4.1.2. Políticas claras de uso y consentimiento informado

Es fundamental que las instituciones educativas establezcan políticas claras sobre el uso de LLM, que especifiquen qué datos se recopilan, cómo se utilizan y quién tiene acceso a ellos. Además, estas políticas deben ser comunicadas de manera efectiva tanto a los estudiantes como a sus padres o tutores. El consentimiento informado es esencial y los centros educativos deben garantizar que los estudiantes y sus familias comprendan los riesgos y beneficios del uso de estas tecnologías. El consentimiento debe ser explícito, no tácito y debe ser obtenido antes de que cualquier dato del estudiante sea utilizado.

4.1.3. Control de acceso y seguridad de los datos

Las instituciones educativas deben disponer de los recursos y los conocimientos técnicos para implementar controles estrictos sobre quién tiene acceso a los datos de los estudiantes. Esto incluye la utilización de tecnologías de cifrado para proteger los datos durante el almacenamiento y la transmisión, así como la implementación de sistemas de autenticación para asegurar que solo el personal autorizado pueda acceder a la información sensible. Además, es recomendable que los datos se almacenen localmente



en servidores seguros dentro de la institución educativa (on premises) en lugar de ser gestionados por terceros o almacenados en la nube.

4.1.4. Capacitación y concienciación del personal docente

El personal docente debe recibir formación periódica sobre las mejores prácticas de privacidad y seguridad de datos al utilizar LLM en el aula. Esto incluye comprender las políticas de privacidad, saber cómo manejar de manera segura la información de los estudiantes y cómo responder ante posibles incidentes de seguridad. La concienciación sobre la importancia de la privacidad digital es crucial para prevenir errores humanos que puedan comprometer la información de los estudiantes.

4.2. Estrategias para garantizar la veracidad de la información en LLM educativos

Además de proteger la privacidad, las siguientes propuestas están orientadas a mitigar el riesgo de desinformación y asegurar que los estudiantes reciban contenido veraz y didáctico (Baidoo-Anu y Owusu Ansah, 2023).

4.2.1. Integración de mecanismos de verificación de hechos

Para minimizar el riesgo de desinformación es recomendable que los LLM incorporen mecanismos de verificación de hechos que puedan identificar y corregir errores en tiempo real. Esto puede lograrse mediante la integración de bases de datos verificadas y sistemas de validación que comparen la información generada por los modelos con fuentes confiables. Las plataformas educativas que utilizan LLM deben ser capaces de priorizar la precisión sobre la velocidad de respuesta, procurando que cada respuesta generada sea revisada para asegurar su veracidad.

Dado que estos modelos pueden producir respuestas basadas en grandes cantidades de datos, pero sin una comprensión completa del contexto o la precisión, es fundamental integrar mecanismos que verifiquen la exactitud de la información antes de presentarla a los estudiantes.

Una estrategia para abordar este desafío es la integración de mecanismos de verificación de hechos junto con técnicas de Retrieval-Augmented Generation (RAG). Los sistemas RAG combinan las capacidades generativas de los modelos de lenguaje con un módulo de recuperación de información que busca en bases de datos o fuentes de conocimiento específicas, antes de generar una respuesta. Esto significa que, en lugar de depender únicamente de la información almacenada en los parámetros del modelo, el LLM puede consultar fuentes externas actualizadas y verificadas para generar respuestas más precisas y confiables (Shen, 2024, cap.4.2, 4.4 y 5.1).



4.2.2. Supervisión y corrección manual de contenidos

Es esencial que exista un nivel de supervisión humana para corregir posibles errores o malinterpretaciones generadas por los modelos. Se propone la implementación de sistemas híbridos en los que las respuestas generadas por la IA sean revisadas por educadores antes de ser presentadas a los estudiantes (Zhiyi, 2024, cap.2).

4.2.3. Educación digital y desarrollo del pensamiento crítico

Es crucial que los estudiantes sean educados en el uso crítico de la información generada por LLM. Esto incluye enseñarles cómo identificar posibles errores, verificar la información a través de múltiples fuentes y desarrollar habilidades de pensamiento crítico para cuestionar el contenido que reciben. Este enfoque no solo ayudará a los estudiantes a evitar la desinformación, sino que también los preparará para interactuar de manera más efectiva con las tecnologías digitales en general (Zhiyi, 2024, cap.2.6 y 4)

4.2.4. Desarrollo de contenidos educativos adaptativos

Proponemos la creación de contenidos educativos que se adapten dinámicamente al nivel de comprensión del estudiante, utilizando LLM para generar explicaciones alternativas o adicionales cuando se detecten dificultades de comprensión. Este enfoque no solo mejora la personalización del aprendizaje, sino que también permite a los educadores hacer un seguimiento del progreso del estudiante y hacer ajustes según sea necesario para asegurar que se esté recibiendo información correcta y adaptada a las capacidades del alumno (Zhiyi, 2024, cap.2)

4.3. Evaluación y adaptación de modelos: ChatGPT, Copilot y Gemini

A continuación, se presenta una evaluación de ChatGPT, Copilot, y Gemini con propuestas para su adaptación en el entorno educativo.

4.3.1. ChatGPT

ChatGPT, como modelo de inteligencia generativa, ha demostrado ser una herramienta poderosa en la generación de contenido educativo personalizado. Sin embargo, su funcionamiento continuo y basado en interacciones plantea riesgos significativos para la privacidad de los usuarios, especialmente en un contexto educativo donde se manejan datos sensibles de estudiantes (Brown et al., 2020). Para mejorar la privacidad, es crucial implementar protocolos estrictos de anonimización de datos y limitar la capacidad del modelo para retener información de interacciones pasadas. Además, dado que ChatGPT puede generar respuestas incorrectas o sesgadas, es conveniente



combinarlo con integrar mecanismos de verificación de hechos, como el enfoque de Retrieval-Augmented Generation (RAG), para asegurar que la información proporcionada sea precisa y confiable (Baidoo-Anu & Owusu Ansah, 2023).

4.3.2. Copilot

Copilot busca diferenciarse por su capacidad de ofrecer una personalización profunda, generando materiales educativos adaptados a las necesidades individuales de los estudiantes. Sin embargo, esta personalización requiere el acceso a datos específicos del usuario, lo que puede comprometer la privacidad si no se manejan adecuadamente estos datos (Shen et al., 2024). Para mejorar la privacidad en el uso de Copilot, se recomienda implementar un entorno seguro en el que los datos se procesen localmente y no se almacenen a largo plazo, minimizando así el riesgo de exposición. Además, es importante asegurar que los datos utilizados para la personalización estén anonimizados para garantizar las normativas de privacidad vigentes (Zhiyi, 2024).

4.3.3. Gemini

Gemini es un modelo de IA generalista con la capacidad de generar contenido en una variedad de contextos, lo que lo convierte en una herramienta versátil para la educación. Sin embargo, su naturaleza generalista lo hace más propenso a generar respuestas incorrectas o fuera de contexto, lo que es particularmente problemático en entornos educativos donde la precisión de la información es crucial (Baidoo-Anu & Owusu Ansah, 2023).

4.4. Recomendaciones para la implementación de políticas y regulaciones

Aunque las normativas existentes, como el Reglamento General de Protección de Datos (GDPR) en Europa y la Ley de Privacidad y Derechos Educativos de la Familia (FERPA) en Estados Unidos, ya proporcionan un marco robusto para la protección de datos personales en el ámbito educativo, es crucial establecer mecanismos de control y auditoría específicos que permitan a las instituciones educativas asegurarse de que los proveedores de servicios de LLM cumplen con estas regulaciones de manera efectiva y continua. Estos mecanismos no solo garantizarán que se mantengan los estándares de privacidad y seguridad, sino que también fomentarán la confianza de los usuarios en estas tecnologías.



4.4.1. Implementación de Acuerdos de Nivel de Servicio (SLA) específicos para privacidad y seguridad

Un primer paso para garantizar que los proveedores de LLM cumplen con las normativas es la implementación de Acuerdos de Nivel de Servicio (SLA) específicos que incluyan cláusulas detalladas sobre la privacidad y la seguridad de los datos. Estos acuerdos deben estipular:

- **Criterios de cumplimiento:** Definir los estándares de cumplimiento con las normativas existentes, asegurando que los datos se gestionen de acuerdo con las leyes de privacidad y seguridad aplicables.
- **Obligaciones de notificación:** Incluir obligaciones específicas para que el proveedor informe inmediatamente a la institución educativa sobre cualquier incidente relacionado con la privacidad o la seguridad de los datos.
- **Métricas de rendimiento:** Definir métricas específicas que se utilizarán para evaluar el rendimiento del proveedor en cuanto a la protección de datos y la veracidad de la información generada.

4.4.2. Auditorías periódicas y evaluaciones de cumplimiento

Las instituciones educativas deben tener la capacidad de llevar a cabo auditorías periódicas y evaluaciones de cumplimiento para asegurar que los proveedores de LLM están adheridos a los SLA y cumplen con las regulaciones establecidas. Estas auditorías pueden ser realizadas por la propia institución o por terceros independientes y deben incluir:

- **Revisión de políticas y procedimientos:** Evaluar si los proveedores han implementado políticas y procedimientos adecuados para proteger la privacidad de los estudiantes y manejar los datos de manera segura.
- **Análisis de la cadena de suministro de datos:** Examinar cómo los datos de los estudiantes son gestionados a lo largo de toda la cadena de suministro, incluyendo cualquier subcontratista o tercero que tenga acceso a la información.
- **Evaluación de ciberseguridad:** Realizar pruebas técnicas para evaluar la robustez de las medidas de seguridad implementadas por el proveedor, como el cifrado de datos, la autenticación de usuarios, y la protección contra accesos no autorizados.



4.4.3. Revisión y verificación de la veracidad de la información generada

Además de las auditorías de privacidad y seguridad, es fundamental que las instituciones educativas también tengan mecanismos para verificar la veracidad de la información generada por los LLM. Esto podría incluir:

- **Muestreo y análisis de contenidos:** Implementar un proceso regular de muestreo de las respuestas generadas por el LLM para evaluar su precisión y la adherencia a los estándares educativos. Esto puede incluir la revisión por parte de expertos en la materia o educadores.
- **Retroalimentación continua:** Crear canales de retroalimentación continua entre los educadores y el proveedor del servicio para reportar cualquier problema con la precisión de las respuestas generadas y asegurar que se tomen medidas correctivas rápidamente.
- **Actualización de bases de datos:** Verificar que los proveedores actualicen regularmente las bases de datos y las fuentes de información que utilizan para generar contenido, garantizando que la información proporcionada a los estudiantes esté basada en los conocimientos más actuales.

4.4.4. Transparencia y Reportes de Cumplimiento

Para mantener la confianza en el uso de LLM en entornos educativos, es esencial que los proveedores sean transparentes en sus operaciones y proporcionen reportes de cumplimiento regulares a las instituciones educativas. Estos reportes deben incluir:

- **Resultados de auditorías:** Un resumen de los resultados de las auditorías de privacidad, seguridad, y veracidad, incluyendo cualquier deficiencia encontrada y las acciones correctivas tomadas.
- **Actualizaciones de seguridad:** Información sobre cualquier actualización de seguridad implementada, incluyendo mejoras en el cifrado de datos, autenticación, y medidas contra el acceso no autorizado.
- **Incidentes de seguridad y respuesta:** Un registro detallado de cualquier incidente de seguridad, incluyendo la naturaleza del incidente, su impacto, y las medidas tomadas para resolverlo y prevenir futuros incidentes.



4.4.5. Sanciones y medidas correctivas

Finalmente, los acuerdos entre las instituciones educativas y los proveedores de servicios de LLM deben incluir sanciones y medidas correctivas claras para casos de incumplimiento de las normativas o los SLA. Esto puede incluir:

- **Multas y penalizaciones:** Establecer multas o penalizaciones financieras para el proveedor en caso de violaciones significativas de la privacidad o fallos en la veracidad de la información.
- **Suspensión de servicios:** Incluir cláusulas que permitan a la institución educativa suspender o rescindir el contrato en caso de incumplimientos graves o repetidos.
- **Obligaciones de remediación:** Especificar las obligaciones del proveedor para corregir cualquier problema identificado durante las auditorías o en respuesta a un incidente, incluyendo la obligación de implementar mejoras en las medidas de seguridad y veracidad.





5. Conclusiones

5.1. Síntesis de los hallazgos

El uso de modelos de LLM en entornos educativos presenta oportunidades significativas para personalizar el aprendizaje y mejorar la experiencia educativa de los estudiantes. Sin embargo, como se ha analizado a lo largo de este trabajo, la integración de estas tecnologías también trae consigo importantes desafíos, particularmente en términos de la veracidad de la información, la protección de la privacidad de los datos, y la preparación del profesorado para utilizar estas herramientas de manera efectiva.

Uno de los aspectos más críticos es la **veracidad de la información generada por los LLM**. Para abordar este desafío se ha destacado la importancia de incorporar sistemas de **Retrieval-Augmented Generation (RAG)**, que combinan las capacidades generativas de los LLM con la consulta de bases de datos verificadas y actualizadas. Este enfoque permite que las respuestas generadas por la IA estén respaldadas por información precisa y confiable, reduciendo así el riesgo de desinformación y mejorando la calidad del aprendizaje de los estudiantes.

En cuanto a la **protección de la privacidad** se ha subrayado la necesidad de implementar prácticas robustas de anonimización de datos. Anonimizar los datos de los estudiantes no solo protege su identidad, sino que también asegura que la información sensible no pueda ser utilizada de manera indebida por los proveedores de servicios o por terceros. Además, es esencial que las instituciones educativas cuenten con mecanismos para **monitorizar la gestión de estos datos por parte de los proveedores de servicios**, asegurando el cumplimiento de las normativas vigentes y protegiendo los derechos de los estudiantes.

Por otro lado, la **formación del profesorado** se presenta como un pilar fundamental para la correcta implementación y utilización de los LLM en la educación. Los profesores deben estar capacitados no solo en el uso técnico de estas herramientas, sino también en cómo integrarlas pedagógicamente en el aula de manera que beneficien al aprendizaje sin comprometer la seguridad y privacidad de los estudiantes. La formación debe incluir aspectos como la evaluación crítica de la información generada por los LLM y la enseñanza del pensamiento crítico a los estudiantes.

Finalmente, fomentar el **pensamiento crítico en los alumnos** es clave para que puedan interactuar de manera efectiva con las herramientas de IA. Los estudiantes deben ser capaces de cuestionar la información que reciben, verificarla mediante fuentes adicionales y desarrollar habilidades que les permitan discernir entre contenido fiable y no fiable.



5.2. Recomendaciones prácticas

Con base en los hallazgos anteriores, se proponen las siguientes recomendaciones para maximizar los beneficios del uso de LLM en educación, enfocándose en la veracidad de la información, la privacidad de los datos, y la formación tanto de docentes como de estudiantes:

1. **Integración de sistemas RAG en LLM educativos:** Las instituciones educativas y los desarrolladores de LLM deben priorizar la implementación de sistemas de Retrieval-Augmented Generation (RAG) que permitan generar respuestas respaldadas por bases de datos verificadas y actualizadas. Esto asegurará que los estudiantes reciban información precisa y pertinente, lo que es crucial para su aprendizaje.
2. **Anonimización de datos y monitorización estricta:** Se deben implementar prácticas avanzadas de anonimización para proteger los datos personales de los estudiantes. Además, es esencial que las instituciones educativas tengan la capacidad de monitorizar cómo los proveedores de servicios LLM gestionan estos datos, asegurándose de que cumplen con todas las normativas de privacidad y que las prácticas de manejo de datos son transparentes y auditables.
3. **Formación continua del profesorado:** Los docentes deben recibir formación regular en el uso de LLM, con un enfoque en la integración pedagógica de estas herramientas, la verificación de la veracidad de la información generada, y la enseñanza del pensamiento crítico a los estudiantes. Esta formación es esencial para que los profesores puedan maximizar el potencial de los LLM mientras protegen la privacidad y seguridad de los estudiantes.
4. **Fomento del pensamiento crítico en los estudiantes:** Se deben desarrollar programas educativos que capaciten a los estudiantes para interactuar de manera crítica con las herramientas de LLM. Esto incluye enseñarles a verificar la información, cuestionar las respuestas generadas por la IA, y desarrollar una comprensión sólida de cómo evaluar la calidad y fiabilidad de las fuentes.
5. **Capacidad de la administración para auditar y controlar:** Las instituciones educativas, en colaboración con organismos reguladores, deben establecer mecanismos robustos para auditar y controlar a los proveedores de servicios de LLM. Esto incluye la implementación de Acuerdos de Nivel de Servicio (SLA) que especifican claramente las obligaciones en términos de privacidad, seguridad, y veracidad de la información y que permiten la aplicación de sanciones ejemplarizantes en caso de incumplimiento.

5.3. Reflexiones finales

La integración de modelos de lenguaje grande en la educación tiene el potencial de transformar profundamente el aprendizaje, pero esto debe hacerse de manera cuidadosa y controlada. La implementación de sistemas como RAG, el fortalecimiento de la anonimización de datos, la capacitación del profesorado y el fomento del pensamiento crítico en los estudiantes son medidas esenciales para garantizar que estas tecnologías se utilicen de manera ética y efectiva.

El éxito en la adopción de LLM en el ámbito educativo dependerá en gran medida de la capacidad de las instituciones educativas para gestionar de manera proactiva los riesgos asociados, especialmente en términos de privacidad y veracidad de la información. Al hacerlo, podremos aprovechar todo el potencial de estas tecnologías para mejorar la educación mientras protegemos los derechos y el bienestar de los estudiantes.





6. Referencias

En este apartado se recogen todas las fuentes bibliográficas, artículos académicos, informes técnicos y documentos legales que han sido utilizados para elaborar este trabajo.

6.1. Artículos académicos

- Shen, W., Tianlong X., Hang, L., Chaoli, Z., Joleen. L. (2024). *Large Language Models for Education: A Survey and Outlook*. arxiv.org/abs/2403.18105
- Zhiyi, X. (2024). *AI in education: Enhancing learning experiences and student outcomes*. *Applied and Computational Engineering*, 51(1):104-111.
- Brown, T. B., et al. (2020). *Language models are few-shot learners*. arxiv.org/abs/2005.14165

6.2. Bibliografía

- UNESCO, Shanghai Open University. (2022). *Global research policy & practices report: advancing artificial intelligence-supported global digital citizenship education*. UNESCO, ISBN 978-5-906399-14-4.
- Hastuti, R., Syafruddin. (2023). *Ethical Considerations in the Age of Artificial Intelligence: Balancing Innovation and Social Values*. *West Science Social and Humanities Studies*, 1(02):76-87
- Baidoo-Anu, D., Owusu, L. (2023). *Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning*. Queen's University and University of Cape Coast.

6.3. Conferencias

- Čavojský, M., Bugár, G., Kormaník, T., Hasin, M. (2023). *Exploring the Capabilities and Possible Applications of Large Language Models for Education*. IEEE Xplore.

6.4. Legislación

- European Union. (2018). *General Data Protection Regulation (GDPR)*. Regulation (EU) 2016/679 of the European Parliament and of the Council. Official Journal of the European Union, L119, 1-88.



MASTERPROF UMH
UNIVERSITAS *Miguel Hernández*

**MÁSTER UNIVERSITARIO EN FORMACIÓN DEL PROFESORADO
ESO Y BACHILLERATO, FP Y ENSEÑANZAS DE IDIOMAS**

- U.S. Department of Education. (2020). *Family Educational Rights and Privacy Act (FERPA)*. 20 U.S.C. § 1232g; 34 CFR Part 99.



7. Anexos

7.1. Políticas de privacidad en herramientas LLM educativas

Este anexo analiza las políticas de privacidad de los tres LLM estudiados en este trabajo: ChatGPT, Copilot y Gemini. Se analiza cómo cada uno de estos modelos maneja la privacidad de los datos de los estudiantes, destacando sus prácticas actuales, las posibles áreas de mejora y cómo se alinean con las normativas de protección de datos como el GDPR.

7.1.1. Política de Privacidad de ChatGPT

<https://openai.com/es-ES/policies/privacy-policy/>

Descripción:

ChatGPT recopila datos personales de manera directa cuando los usuarios interactúan con la plataforma. Esto incluye la información proporcionada por los usuarios, como preguntas y respuestas, así como datos técnicos como direcciones IP y tipo de navegador.

OpenAI, propietario de ChatGPT, afirma que utiliza estos datos para mejorar sus servicios, entrenar sus modelos y cumplir con requisitos legales. Sin embargo, la política de privacidad también menciona que los datos pueden ser compartidos con terceros bajo ciertas circunstancias, como proveedores de servicios o en cumplimiento de la ley.

Análisis:

- **Anonimización de datos:** La política no especifica claramente los métodos utilizados para anonimizar los datos de los usuarios. Es esencial que se implementen prácticas de anonimización robustas para proteger la privacidad de los estudiantes y asegurar que la información personal no pueda ser reidentificada.
- **Consentimiento informado:** OpenAI requiere que los usuarios acepten la política de privacidad antes de utilizar sus servicios. Sin embargo, se podría mejorar la transparencia en cómo se obtienen y utilizan los datos, especialmente en entornos educativos donde los usuarios son menores de edad.
- **Acceso a datos por terceros:** La política permite compartir datos con terceros, lo que puede ser una preocupación si no se manejan adecuadamente.



Es crucial que cualquier acceso a datos por parte de terceros esté sujeto a estrictos controles y garantías de privacidad.

Áreas de mejora:

- Mejorar la claridad y transparencia sobre los métodos de anonimización de datos.
- Implementar un consentimiento explícito y más detallado para usuarios menores de edad.
- Restringir el acceso a datos por terceros y asegurar que estos terceros cumplan con los mismos estándares de privacidad.

7.1.2. Política de Privacidad de Copilot

<https://learn.microsoft.com/es-es/copilot/privacy-and-protections>

Descripción:

Copilot, desarrollado por Microsoft, sigue una política de privacidad que se centra en la protección de los datos personales de los usuarios. Microsoft recopila datos de los usuarios cuando interactúan con Copilot, incluyendo el contenido generado y los datos contextuales necesarios para ofrecer recomendaciones personalizadas.

La política de privacidad de Copilot establece que los datos se utilizan para mejorar los servicios, y se almacenan y procesan en la nube de Microsoft. Además, se menciona que la seguridad de los datos es una prioridad y que se implementan diversas medidas para proteger la información personal de los usuarios.

Análisis:

- **Anonimización de datos:** Microsoft implementa medidas de seguridad para proteger los datos, pero la política no detalla las técnicas específicas de anonimización utilizadas. Dado que Copilot se utiliza en entornos educativos, es fundamental garantizar que los datos de los estudiantes sean anonimizados de manera efectiva.
- **Consentimiento informado:** Los usuarios deben aceptar la política de privacidad antes de utilizar Copilot. Sin embargo, en el contexto educativo, es necesario contar con un consentimiento más detallado que explique claramente cómo se utilizan los datos y qué derechos tienen los usuarios.



- **Acceso a datos por terceros:** La política permite la transferencia de datos a terceros en ciertas circunstancias, como la mejora de servicios o el cumplimiento de la ley. Es importante que este acceso esté sujeto a estrictos estándares de privacidad y seguridad.

Áreas de mejora:

- Proporcionar detalles más claros sobre las técnicas de anonimización de datos.
- Asegurar que el consentimiento informado sea claro y detallado, especialmente para usuarios menores de edad.
- Limitar y regular estrictamente el acceso a datos por terceros, con un enfoque en la protección de la privacidad.

7.1.3. Política de Privacidad de Gemini

<https://support.google.com/gemini/answer/13594961?hl=es>

Descripción:

Gemini, en el contexto educativo, sigue una política de privacidad centrada en la recopilación y uso de datos para personalizar experiencias y mejorar sus servicios. La política describe cómo se recopilan diversos tipos de datos, incluidos datos de actividad, información del dispositivo, datos de ubicación y datos demográficos. Estos datos se utilizan principalmente para mejorar la relevancia de los anuncios y para analizar el rendimiento de las campañas publicitarias.

Gemini asegura que los datos de los usuarios son tratados con confidencialidad y que se implementan medidas de seguridad para proteger la información personal. La política también menciona que los datos pueden ser compartidos con socios comerciales y otros terceros para la mejora de servicios y análisis, siempre en cumplimiento con las regulaciones aplicables.

Análisis:

- **Anonimización de datos:** La política de Gemini menciona la implementación de medidas de seguridad, pero no detalla específicamente las técnicas de anonimización utilizadas para proteger los datos de los usuarios. En el entorno educativo, es crucial que se utilicen técnicas robustas de anonimización para asegurar que la información de los estudiantes no pueda ser reidentificada.



- **Consentimiento informado:** Gemini requiere que los usuarios acepten la política de privacidad para utilizar sus servicios. Sin embargo, en el ámbito educativo, donde los usuarios pueden ser menores de edad, es importante que el consentimiento sea explícito y que los usuarios o sus tutores legales comprendan plenamente cómo se recopilan y utilizan sus datos.
- **Acceso a datos por terceros:** La política permite el acceso a los datos por parte de terceros, incluidos socios comerciales, para propósitos como la personalización de anuncios y el análisis del rendimiento. Es esencial que este acceso esté sujeto a controles estrictos para garantizar que los datos se manejen de manera segura y conforme a las normativas de privacidad.

Áreas de mejora:

- Proporcionar detalles más específicos sobre las técnicas de anonimización utilizadas para proteger los datos de los estudiantes.
- Mejorar la claridad y exhaustividad del consentimiento informado, especialmente en contextos educativos donde los usuarios son menores de edad.
- Limitar el acceso a datos por parte de terceros y asegurar que cualquier compartición de datos cumpla con los más altos estándares de privacidad y seguridad.

Conclusión de este anexo

Este anexo ha proporcionado un análisis detallado de las políticas de privacidad de ChatGPT, Copilot, y Gemini con respecto a un contexto educativo.

ChatGPT destaca por su capacidad generativa y personalización, pero necesita una mayor claridad en sus prácticas de anonimización y un enfoque más riguroso en el consentimiento informado, especialmente cuando se trata de usuarios menores de edad. Copilot, por su parte, ofrece una personalización profunda, pero al igual que ChatGPT, debe asegurar que los datos de los estudiantes estén anonimizados de manera efectiva y que el acceso a estos datos esté estrictamente regulado. Gemini se enfrenta al desafío de balancear la personalización de experiencias con la protección de la privacidad, siendo crucial mejorar la transparencia en la anonimización de datos y restringir el acceso a terceros.

A través de la implementación de prácticas de anonimización robustas, un consentimiento informado claro y regulaciones estrictas sobre el acceso a datos por terceros, estas herramientas podrían llegar a ofrecer un entorno seguro y confiable para el aprendizaje y la enseñanza en la era digital.



7.2. Implementación de modelos LLM basados en soluciones Open Source

Las soluciones on-premises (offline) para modelos de lenguaje grande (LLM) ofrecen una alternativa atractiva a los servicios basados en la nube como ChatGPT, Copilot y Gemini, permitiendo a las instituciones educativas mantener un mayor control sobre los datos y garantizar el cumplimiento de las normativas de privacidad.

Este anexo analiza la viabilidad de implementar soluciones on-premises utilizando las opciones de código abierto disponibles actualmente, evaluando sus beneficios, desafíos y consideraciones clave para su adopción en entornos educativos.

7.2.1. Concepto de soluciones on-premises

Las soluciones on-premises implican la instalación y ejecución de modelos de LLM directamente en la infraestructura de la institución educativa, en lugar de depender de servicios en la nube proporcionados por terceros. Esto significa que todos los procesos de almacenamiento, procesamiento y generación de datos se realizan localmente, lo que ofrece un mayor control sobre la seguridad y privacidad de la información.

7.2.2. Opciones Open Source Actuales para LLM On-Premises

Existen varias iniciativas y proyectos de código abierto que permiten la implementación de LLM en entornos on-premises. Algunas de las opciones más destacadas incluyen:

- **GPT-Neo y GPT-J (EleutherAI):** Estos modelos son desarrollados por EleutherAI y están diseñados para replicar las capacidades de modelos como GPT-3. Son completamente de código abierto y pueden ser desplegados localmente, ofreciendo una buena combinación de rendimiento y accesibilidad.
- **LLaMA (Meta AI):** LLaMA es una serie de modelos de lenguaje desarrollados por Meta AI que están disponibles para investigación y uso académico. Aunque no son completamente open source en todos sus aspectos, proporcionan acceso a modelos potentes que pueden ser utilizados en infraestructuras locales.
- **Bloom (BigScience):** Bloom es un modelo de lenguaje multilingüe desarrollado por el proyecto BigScience. Es completamente open source y está diseñado para ser accesible y personalizable, lo que lo hace adecuado para implementaciones on-premises.
- **Alpaca (Stanford):** Alpaca es una adaptación ligera de modelos de lenguaje grandes, optimizada para ser más eficiente y fácil de desplegar en infraestructuras locales. Es particularmente útil para instituciones con recursos computacionales limitados.



7.2.3. Ventajas de las soluciones on-premises en entornos educativos

Implementar soluciones on-premises para LLM en instituciones educativas ofrece múltiples beneficios:

- **Control de Datos:** Al alojar los modelos localmente, las instituciones tienen un control completo sobre los datos de los estudiantes, reduciendo el riesgo de exposición a terceros y facilitando el cumplimiento de las normativas de privacidad.
- **Personalización:** Las soluciones on-premises permiten una mayor personalización de los modelos para adaptarse a las necesidades específicas de la institución, incluyendo la adaptación del contenido educativo y la configuración de filtros personalizados.
- **Seguridad Mejorada:** Al mantener los datos dentro de la infraestructura local, se puede implementar y gestionar directamente las medidas de seguridad, minimizando los riesgos asociados con la transmisión y almacenamiento en la nube.
- **Independencia de Proveedores Externos:** Las instituciones no dependen de proveedores externos para el acceso a los modelos, lo que reduce la vulnerabilidad a cambios en las políticas de precios, disponibilidad del servicio o términos de uso.

7.2.4. Desafíos y consideraciones

A pesar de sus ventajas, la implementación de soluciones on-premises para LLM también presenta desafíos que deben ser considerados:

- **Requerimientos Técnicos:** Desplegar y mantener modelos de LLM on-premises requiere conocimientos técnicos especializados, así como infraestructura adecuada en términos de hardware (servidores con capacidad de procesamiento y almacenamiento) y software.
- **Costos Iniciales:** La inversión inicial en infraestructura y recursos técnicos puede ser significativa, especialmente para instituciones educativas con presupuestos limitados.
- **Escalabilidad:** Adaptar la infraestructura para manejar incrementos en la demanda o para actualizar los modelos puede ser complejo y costoso en entornos on-premises.



- **Mantenimiento y Actualizaciones:** Las instituciones deben encargarse del mantenimiento continuo, las actualizaciones de los modelos y la gestión de parches de seguridad, lo que puede requerir personal dedicado y capacitación adicional.
- **Rendimiento:** Asegurar que los modelos funcionen de manera eficiente y rápida en infraestructura local puede ser un desafío, especialmente para modelos de mayor tamaño que requieren recursos computacionales intensivos.

7.2.6. Recomendaciones prácticas

Basado en el análisis anterior, se proponen las siguientes recomendaciones para la implementación de soluciones on-premises de LLM en entornos educativos:

1. **Seleccionar modelos open source adecuados:** Optar por modelos que sean bien soportados por la comunidad, que tengan documentación completa y que sean adecuados para el entorno educativo. Modelos como GPT-Neo, Bloom y Alpaca pueden ser buenos puntos de partida.
2. **Implementar una solución común desde la Administración Central y/o Autónoma:** Con el objetivo de minimizar los costes de implantación de la solución técnica y de ampliar la base de conocimiento, se propone que el desarrollo tecnológico y la inversión en formación del profesorado sean asumidos por la Administración Central y/o Autónoma, aprovechando al máximo la economía de escala y liberando de la carga administrativa a los centros en la mayor medida posible.
3. **Fomentar la participación de los equipos docentes en comunidades open source asociadas al proyecto:** Promover que los propios participantes puedan involucrarse voluntariamente en comunidades de desarrollo open source para mantenerse al día con las últimas innovaciones, compartir conocimiento, obtener soporte y contribuir al desarrollo colectivo.

Conclusión de este anexo

La adopción de soluciones on-premises para modelos de lenguaje grande en entornos educativos es una opción viable que ofrece numerosos beneficios en términos de control de datos, seguridad y personalización. Sin embargo, requiere una inversión significativa en infraestructura, capacitación y mantenimiento. Al aprovechar las opciones open source disponibles actualmente las instituciones educativas pueden implementar estas soluciones garantizando la protección de la privacidad de los estudiantes y mejorando la calidad del aprendizaje.