

INTELIGENCIA ARTIFICIAL GENERATIVA COMO APOYO EXPLORATORIO EN LA REVISIÓN DE TESIS DE POSGRADO

Generative Artificial Intelligence as Exploratory Support Resource in the Process of Reviewing Thesis Drafts within the Graduate Context

Corini Guarachi, Alan Rodrigo

Universidad Mayor de San Andrés

acorinig@fcpn.edu.bo

ORCID: <https://orcid.org/0000-0003-0629-6474>

La Paz – Bolivia

López Montalvo, Rosalía

Universidad Mayor de San Andrés

rlopezm@fcpn.edu.bo

ORCID: <https://orcid.org/0009-0007-3505-8623>

La Paz – Bolivia

Resumen

La revisión académica de tesis de posgrado constituye un proceso complejo que requiere rigurosidad metodológica, coherencia estructural y validación científica. En este contexto, la inteligencia artificial generativa puede constituirse en una herramienta de apoyo para la identificación preliminar de inconsistencias metodológicas, problemas estructurales, errores de citación y debilidades argumentativas. La presente investigación analiza, desde un enfoque exploratorio y descriptivo basado en un estudio de caso, el comportamiento de distintos modelos de inteligencia artificial generativa en tareas de revisión académica de una tesis de posgrado se utilizó un prompt estructurado de revisión académica aplicado a modelos como ChatGPT, Claude, Gemini, Qwen y Kimi. Las observaciones generadas fueron clasificadas en categorías metodológicas, estructurales y bibliográficas, permitiendo comparar de manera descriptiva el tipo, cantidad y profundidad de los hallazgos producidos por cada modelo. Los resultados evidenciaron

diferencias observables entre los modelos evaluados respecto a la profundidad analítica, trazabilidad de las observaciones, detección de inconsistencias académicas y revisión de aspectos bibliográficos, se observó que la estructura del prompt influye directamente en la calidad de las respuestas generadas, aunque los resultados no deben interpretarse como generalizables debido al carácter exploratorio del estudio y al uso de un único caso de análisis. La inteligencia artificial generativa puede contribuir a los procesos preliminares de revisión académica, siempre que sea utilizada como herramienta de apoyo y bajo supervisión humana especializada, debido a limitaciones relacionadas con alucinaciones, validación científica, sesgos y razonamiento contextual.

Palabras clave: inteligencia artificial generativa, revisión académica, tesis de posgrado, estudio exploratorio, ingeniería de prompts

Abstract

The academic review of graduate theses constitutes a complex process that requires methodological rigor, structural coherence, and scientific validation. Within this context, generative artificial intelligence can serve as a support tool for the preliminary identification of methodological inconsistencies, structural problems, citation errors, and argumentative weaknesses. Implementing an exploratory and descriptive case-study approach, this research analyzes different generative artificial intelligence models performance in academic review tasks for a graduate thesis. A structured academic review prompt was applied to models such as ChatGPT, Claude, Gemini, Qwen, and Kimi. The generated observations were classified into methodological, structural, and bibliographic categories, allowing a descriptive comparison of the type, quantity, and depth of the findings produced by each model. The results revealed observable differences among the evaluated models regarding analytical depth, traceability of observations, academic inconsistencies detection, and bibliographic review. It was observed that the structure of the prompt directly influences the quality of the generated responses, though

these findings should not be interpreted as generalizable due to the study's exploratory nature and the use of a single case of analysis. Generative artificial intelligence can contribute to preliminary academic review processes, provided it is used as a support tool and under specialized human supervision, due to limitations related to hallucinations, scientific validation, biases, and contextual reasoning.

Keywords: generative artificial intelligence, academic review, graduate thesis, exploratory study, prompt engineering

Introducción

La transformación digital en la educación superior ha impulsado el desarrollo de nuevas herramientas tecnológicas orientadas al fortalecimiento de procesos académicos y científicos (Selwyn, 2019).

Entre estas tecnologías, la inteligencia artificial generativa ha adquirido especial relevancia debido a su capacidad para producir contenido textual, analizar información compleja y apoyar procesos de evaluación documental (Kasneci et al., 2023).

En el ámbito de posgrado, la revisión de tesis representa una actividad de alta complejidad metodológica y científica, debido a que implica verificar coherencia estructural, consistencia argumentativa, calidad metodológica y cumplimiento de estándares académicos de diferentes trabajos en diferentes áreas académicas (Creswell y Creswell, 2018), sin embargo, los procesos tradicionales de revisión demandan largos periodos de evaluación y dependen significativamente de la experiencia del tutor o tribunal académico. La evaluación de investigaciones especializadas exige que los tribunales académicos revisen literatura científica para emitir criterios fundamentados.

Recientemente, los modelos de lenguaje de gran escala han demostrado capacidades relevantes en tareas de análisis textual, corrección académica y generación automatizada de observaciones (Bommasani et al., 2022). Estas capacidades han generado interés en su aplicación dentro de contextos educativos y científicos (Lo, 2023).

El uso de inteligencia artificial generativa en procesos de revisión académica también presenta riesgos asociados con la generación de contenidos sin fundamento teórico o validación científica “alucinaciones”, generación de referencias inexistentes, sesgos y problemas de confiabilidad científica (OpenAI et al., 2024).

La presente investigación plantea un estudio exploratorio piloto orientado al análisis del comportamiento de distintos modelos de inteligencia artificial generativa en tareas de revisión académica automatizada. Para ello, se utilizó una tesis de posgrado como caso de estudio experimental, aplicando un mismo prompt estructurado de revisión académica sobre diferentes modelos de lenguaje. El propósito del estudio no fue reemplazar la evaluación científica humana, sino explorar el potencial de la inteligencia artificial generativa como herramienta de apoyo preliminar en procesos de revisión académica.

Materiales y métodos

Enfoque de investigación

La investigación se desarrolló bajo un enfoque exploratorio descriptivo mediante un estudio piloto de caso aplicado al análisis del comportamiento de modelos de inteligencia artificial generativa en procesos de revisión académica de tesis de posgrado.

El estudio se centró en evaluar la capacidad de distintos modelos de lenguaje para identificar inconsistencias metodológicas, problemas estructurales, errores de citación y debilidades argumentativas a partir de un prompt académico estructurado, debido al carácter exploratorio de la investigación, los resultados obtenidos fueron interpretados desde una perspectiva descriptiva y comparativa no inferencial.

El propósito principal del estudio fue analizar el potencial de la inteligencia artificial generativa como herramienta de apoyo preliminar en procesos de revisión académica automatizada, sin reemplazar el criterio científico humano.

Diseño metodológico

La metodología fue estructurada en cuatro fases principales:

1. diseño del prompt académico de revisión,
2. selección del caso de estudio,
3. generación automatizada de observaciones mediante modelos de inteligencia artificial generativa,
4. sistematización y análisis descriptivo de resultados.

Como caso experimental se utilizó una tesis de posgrado procesada mediante distintos modelos de inteligencia artificial generativa utilizando el mismo prompt estructurado de revisión académica, las observaciones generadas por cada modelo fueron posteriormente organizadas y clasificadas en categorías metodológicas, estructurales y bibliográficas para su análisis comparativo descriptivo.

El estudio buscó explorar diferencias observables en el comportamiento de los modelos respecto a profundidad analítica, trazabilidad de observaciones y capacidad de detección de inconsistencias académicas.

La selección de una única tesis permitió controlar la variabilidad documental, evaluando todos los modelos bajo las mismas condiciones (documento y prompt), de este modo, las diferencias observadas corresponden al comportamiento de cada modelo, en el marco de un estudio piloto exploratorio.

Diseño del prompt

El prompt utilizado en la investigación fue construido utilizando técnicas de prompt engineering y estructuración jerárquica de instrucciones orientadas a procesos de revisión académica preliminar.

El diseño del prompt incorporó criterios metodológicos, estructurales, bibliográficos y argumentativos comúnmente utilizados en procesos universitarios de evaluación de tesis.

El prompt experimental utilizado en el estudio se encuentra disponible públicamente en un repositorio GitHub (<https://github.com/darkpro/prompt>) con fines de trazabilidad y reproducibilidad académica.

El uso de un único prompt estructurado estandarizó la evaluación, asegurando que las diferencias observadas correspondieran únicamente al comportamiento de cada modelo.

Variables de análisis

Los indicadores utilizados en la investigación corresponden a categorías descriptivas de análisis orientadas a sistematizar las observaciones generadas por los modelos de inteligencia artificial generativa. Debido al carácter exploratorio del estudio, las variables no fueron utilizadas como métricas inferenciales ni como instrumentos estandarizados de evaluación científica.

Tabla 1. Variables utilizadas en la evaluación experimental

Variable	Descripción
Precisión de observaciones	Capacidad del modelo para identificar posibles inconsistencias dentro de la tesis.
Profundidad metodológica	Nivel técnico y científico de las observaciones generadas.
Consistencia estructural	Detección de problemas organizativos y estructurales.
Calidad bibliográfica	Identificación de errores APA y referencias inconsistentes.
Capacidad argumentativa	Detección de contradicciones y inconsistencias académicas.
Reproducibilidad	Consistencia de respuestas frente a criterios similares de evaluación.
Trazabilidad académica	Capacidad de vincular observaciones con páginas y fragmentos específicos.

Análisis descriptivo

La investigación incorporó un análisis descriptivo orientado a sistematizar la frecuencia y tipo de observaciones generadas por los modelos de inteligencia artificial generativa.

Las observaciones fueron agrupadas en categorías previamente definidas dentro del prompt académico, considerando aspectos metodológicos, estructurales, bibliográficos y argumentativos.

Los resultados obtenidos fueron utilizados únicamente con fines comparativos descriptivos y exploratorios.

Debido al carácter exploratorio del estudio y al uso de un único caso experimental, no se aplicaron procedimientos estadísticos inferenciales, el análisis se limitó a una comparación descriptiva de las observaciones generadas por cada modelo, sin pretender establecer diferencias estadísticamente significativas entre ellos.

Análisis cualitativo

El análisis cualitativo se centró en analizar descriptivamente la calidad interpretativa de las observaciones generadas por el modelo, considerando:

- Claridad técnica
- Precisión conceptual
- Profundidad metodológica
- Utilidad académica
- Pertinencia científica
- Capacidad crítica

Asimismo, se evaluó la capacidad del modelo para identificar relaciones entre objetivos, metodología, resultados y conclusiones dentro de las tesis analizadas (Lo, 2023).

Consideraciones éticas

La investigación fue desarrollada bajo criterios de uso responsable de inteligencia artificial generativa en contextos académicos, los documentos utilizados fueron procesados únicamente con fines investigativos y de análisis metodológico.

Se consideraron riesgos relacionados con generación de información incorrecta, alucinaciones, referencias inexistentes y sesgos algorítmicos, aspectos ampliamente documentados en investigaciones recientes sobre modelos de lenguaje (Ji et al., 2023)

En consecuencia, la inteligencia artificial generativa fue considerada exclusivamente como herramienta de apoyo académico y no como mecanismo sustitutivo del juicio científico humano.

Asimismo, los resultados obtenidos en la investigación no fueron considerados mecanismos automáticos de evaluación científica, sino herramientas de apoyo preliminar sujetas a validación y revisión humana especializada.

Resultados

El análisis exploratorio describió el comportamiento de distintos modelos de inteligencia artificial generativa en tareas de revisión académica automatizada. Los modelos evaluados fueron ChatGPT, Qwen, Claude, Gemini y Kimi.

Los resultados obtenidos evidenciaron diferencias observables en la capacidad descriptiva para identificar posibles inconsistencias académicas, profundidad metodológica de las observaciones y precisión en la identificación de problemas estructurales y bibliográficos entre los modelos evaluados.

Debido al carácter exploratorio del estudio y al uso de una única tesis como caso de análisis, los resultados fueron interpretados desde una perspectiva descriptiva y comparativa no inferencial, los indicadores presentados en las tablas corresponden a métricas

estimativas orientadas a sistematizar las observaciones generadas por cada modelo.

Los índices descriptivos presentados en las Tablas 3, 4 y 5 corresponden a estimaciones heurísticas generadas por cada modelo de inteligencia artificial a partir del mismo conjunto de criterios definidos en el prompt estructurado, estos valores fueron utilizados exclusivamente como indicadores descriptivos para facilitar la comparación exploratoria entre modelos y no constituyen métricas estandarizadas ni resultados de validación experimental.

Detección de inconsistencias

La Tabla 2 presenta los principales indicadores cuantitativos obtenidos durante el análisis experimental.

Tabla 2. *Evaluación académica automatizada*

Indicador	ChatGPT	Qwen	Claude	Gemini	Kimi
Total, de observaciones detectadas	196	24	87	26	68
Total, de páginas con errores	58	12	35	35	45
Total, de inconsistencias críticas	24	4	8	5	8

Los resultados evidenciaron diferencias entre los modelos evaluados. ChatGPT registró la mayor cantidad de observaciones (196) e inconsistencias críticas (24), destacándose por la amplitud y profundidad de la revisión. Claude presentó un desempeño intermedio con 87 observaciones y una alta capacidad de análisis estructural. Gemini generó menos observaciones (26), pero mostró una alta profundidad metodológica. Kimi obtuvo 68 observaciones, con resultados moderados en los indicadores analizados. Por su parte, Qwen presentó la menor cantidad de observaciones (24), aunque mantuvo consistencia en la evaluación de aspectos estructurales y de coherencia académica.

Evaluación de coherencia estructural

La coherencia estructural fue evaluada considerando aspectos relacionados con organización documental, numeración de capítulos, referencias cruzadas, secuencia metodológica y consistencia interna del documento.

La Tabla 3 presenta indicadores descriptivos relacionados con coherencia estructural observada durante el proceso de revisión automatizada.

Tabla 3. *Indicadores descriptivos de coherencia estructural*

Modelo	Índice	Interpretación
ChatGPT	72/100	Coherencia estructural media
Qwen	88/100	Alta coherencia estructural
Claude	52/100	Problemas estructurales significativos
Gemini	82/100	Alta coherencia estructural
Kimi	35/100	Baja coherencia estructural

Los modelos lograron identificar problemas relacionados con capítulos incompletos, numeración inconsistente, subsecciones duplicadas y referencias cruzadas incorrectas, se detectaron inconsistencias entre índice y contenido real del documento, especialmente en tesis con estructuras metodológicas extensas.

Detección preliminar de inconsistencias metodológicas

En relación con validación metodológica, los modelos fueron capaces de detectar inconsistencias entre objetivos, metodología, resultados y conclusiones.

Entre los principales problemas identificados destacaron:

- ausencia de validación experimental,
- métricas mencionadas, pero no implementadas,
- debilidad en definición de variables,

- inconsistencias entre enfoque metodológico y resultados,
- problemas de reproducibilidad,
- interpretación insuficiente de resultados experimentales.

La Tabla 4 resume los índices estimados de solidez metodológica obtenidos durante la evaluación.

Tabla 4. *Indicadores descriptivos de revisión metodológica*

Modelo	Índice	Interpretación
ChatGPT	48/100	Solidez metodológica baja
Qwen	72/100	Solidez metodológica media-alta
Claude	61/100	Implementación metodológica parcialmente consistente
Gemini	85/100	Alta solidez metodológica
Kimi	40/100	Deficiencias metodológicas significativas

Los resultados descriptivos sugieren que los prompts estructurados pueden contribuir a mejorar la detección preliminar de vacíos metodológicos y debilidades experimentales.

Revisión bibliográfica automatizada

La mayor cantidad de observaciones correspondió a validación bibliográfica y cumplimiento de normas APA.

Entre las principales observaciones bibliográficas detectadas por los modelos se encontraron:

- Referencias no citadas,
- Citas sin referencia,
- Duplicados bibliográficos,
- Fuentes no académicas,
- Referencias incompletas,
- Errores de formato APA,

- Posibles referencias inexistentes.

La Tabla 5 resume los índices de calidad bibliográfica detectados.

Tabla 5. *Evaluación de calidad bibliográfica*

Modelo	Índice	Interpretación
ChatGPT	67/100	Calidad bibliográfica media
Qwen	85/100	Alta calidad bibliográfica
Claude	55/100	Problemas bibliográficos moderados
Gemini	80/100	Calidad bibliográfica media-alta
Kimi	45/100	Problemas bibliográficos significativos

Los resultados muestran que los modelos con mayor capacidad contextual lograron identificar patrones complejos de inconsistencias bibliográficas, especialmente relacionados con duplicación de referencias y errores de citación académica.

Los resultados obtenidos deben interpretarse considerando el carácter exploratorio del estudio y la utilización de una única tesis como caso experimental, por lo que las observaciones descritas no representan conclusiones generalizables sobre el comportamiento global de los modelos evaluados, por lo tanto las diferencias observadas deben interpretarse exclusivamente dentro del contexto experimental analizado y no como evidencia concluyente del desempeño general de los modelos de inteligencia artificial.

Discusión

Los resultados obtenidos evidencian que la inteligencia artificial generativa posee un alto potencial como herramienta de apoyo en procesos de revisión académica de tesis de posgrado, los modelos evaluados mostraron capacidades descriptivas preliminares para identificar inconsistencias metodológicas, problemas estructurales, errores de citación e inconsistencias académicas, aspectos que se consideran dentro de procesos universitarios de evaluación científica.

Los resultados evidenciaron una influencia directa del diseño del prompt sobre la calidad de las observaciones generadas, los resultados muestran que prompts estructurados permiten mejorar significativamente la profundidad metodológica, la precisión de las observaciones y la trazabilidad académica de los análisis realizados por los modelos de inteligencia artificial generativa, este comportamiento coincide con investigaciones recientes relacionadas con prompt engineering y modelos de lenguaje de gran escala, donde la calidad de las instrucciones influye directamente en el desempeño del modelo.

Se observó que los modelos evaluados presentan diferencias importantes en términos de capacidad analítica, profundidad crítica y consistencia académica, aunque ChatGPT generó la mayor cantidad de observaciones y mostró un alto nivel de detalle técnico, particularmente en detección de inconsistencias metodológicas y errores estructurales, también presentó mayor tendencia a generar observaciones redundantes y análisis excesivamente extensos.

Sin embargo, el estudio evidencia que la inteligencia artificial generativa no debe considerarse un sustituto del juicio científico humano, aunque los modelos demostraron capacidad para detectar patrones complejos de inconsistencias académicas, aún presentan limitaciones relacionadas con el razonamiento científico y validación metodológica, pero también permitió identificar observaciones preliminares que podrían pasar desapercibidas durante revisiones iniciales.

El estudio presenta limitaciones inherentes a su carácter exploratorio y al uso de una única tesis como caso experimental, esta decisión permitió controlar la variabilidad documental y evaluar todos los modelos bajo un mismo documento y un mismo prompt estructurado; sin embargo, restringe la posibilidad de generalizar los resultados, por lo tanto los hallazgos deben interpretarse únicamente dentro del contexto experimental analizado, asimismo, los resultados pueden verse influenciados por el diseño del prompt, así como por las versiones específicas y las capacidades contextuales de los modelos de inteligencia artificial utilizados.

El estudio no incorporó validación cruzada mediante evaluadores humanos ni análisis estadístico inferencial, por lo que los resultados deben interpretarse desde una perspectiva descriptiva orientada a exploración preliminar y no como una evaluación definitiva de las diferencias entre modelos.

Los hallazgos obtenidos permiten considerar a la inteligencia artificial generativa como una herramienta complementaria dentro de procesos de apoyo académico, especialmente en etapas preliminares de revisión documental. Sin embargo, su utilización requiere supervisión especializada y validación científica humana para evitar interpretaciones incorrectas, sesgos o generación de observaciones no fundamentadas.

Conclusiones

La presente investigación permitió explorar el uso de la inteligencia artificial generativa como herramienta de apoyo preliminar en procesos de revisión académica de tesis de posgrado mediante prompts estructurados orientados a detección de inconsistencias metodológicas, estructurales y bibliográficas.

Los resultados obtenidos sugieren que los modelos de inteligencia artificial generativa poseen capacidades descriptivas preliminares para apoyar procesos de revisión académica, particularmente en identificación de inconsistencias estructurales, observaciones metodológicas, errores de citación y revisión bibliográfica. Asimismo, se observó que la incorporación de prompts estructurados puede influir en la profundidad y organización de las observaciones generadas por los modelos evaluados.

No obstante, la inteligencia artificial generativa no debe considerarse un mecanismo sustitutivo, su uso dentro de procesos académicos requiere supervisión especializada debido a limitaciones relacionadas con alucinaciones, sesgos, interpretación contextual y generación de observaciones potencialmente incorrectas o no verificables científicamente.

Como trabajo futuro se propone ampliar el estudio mediante la evaluación de múltiples tesis pertenecientes a distintas áreas del conocimiento, incorporar validación cruzada con especialistas en investigación y aplicar procedimientos estadísticos que permitan contrastar cuantitativamente los resultados obtenidos.

Referencias

- Bommasani, R., Hudson, D., Adeli, E., Altman, R., y Arora, S. (12 de Julio de 2022). *On the Opportunities and Risks of Foundation Models*. Retrieved 20 de Mayo de 2026, from <https://arxiv.org/abs/2108.07258>
- Creswell, J., y Creswell, D. (2 de Enero de 2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. 5th ed. SAGE Publications*. Retrieved 18 de Mayo de 2026, from <https://www.amazon.com/Research-Design-Qualitative-Quantitative-Approaches/dp/1506386709>
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., y Xu, Y. (03 de Marzo de 2023). *Survey of Hallucination in Natural Language Generation*. Retrieved 19 de Mayo de 2026, from <https://dl.acm.org/doi/10.1145/3571730>
- Kasneci, E., Sessler, K., Küchemann, S., y Bannert, M. (Abril de 2023). *Should Robots Replace Teachers? AI and the Future of Education*. Retrieved 15 de Mayo de 2026, from <https://www.sciencedirect.com/science/article/abs/pii/S1041608023000195?via%3Dihub>
- Lo, C. (13 de Abril de 2023). *What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature*. Retrieved 25 de Mayo de 2026, from <https://www.mdpi.com/2227-7102/13/4/410>
- OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., y Akkaya, I. (4 de Marzo de 2024). *GPT-4 Technical Report*. Retrieved 27 de Mayo de 2026, from <https://arxiv.org/abs/2303.08774>
- Selwyn, N. (Noviembre de 2019). *Should robots replace teachers? AI and the Future of Education*. Retrieved 20 de Mayo de 2026, from <https://www.wiley.com/en-us/shop/cultural-studies-general/should-robots-replace-teachers-ai-and-the-future-of-education-p-9781509528967>

Fecha de recepción: 15 de mayo de 2026

Fecha de aceptación: 15 de junio de 2026