

UNEMI

UNIVERSIDAD ESTATAL DE MILAGRO

REPÚBLICA DEL ECUADOR

UNIVERSIDAD ESTATAL DE MILAGRO
FACULTAD DE POSGRADO

VICERRECTORADO DE INVESTIGACIÓN Y POSGRADO

ARTÍCULOS PROFESIONALES DE ALTO NIVEL
PREVIO A LA OBTENCIÓN DEL TÍTULO DE:

MAGÍSTER EN INTELIGENCIA ARTIFICIAL PARA LA EDUCACION

TEMA:

GEMINI COMO HERRAMIENTA DE APRENDIZAJE MULTIMODAL:
APLICACIONES Y ESTRATEGIAS PARA LA EDUCACION SUPERIOR

Autor:

VALLE MEDINA FERNANDO DAVID

Director:

DAVILA SANTILLAN LISBETH NARCISA

Milagro, 2026

**GEMINI COMO HERRAMIENTA DE APRENDIZAJE MULTIMODAL:
APLICACIONES Y ESTRATEGIAS PARA LA EDUCACIÓN SUPERIOR**

RESUMEN

El desarrollo de los Modelos de Lenguaje Multimodal de Gran Tamaño (MLLMs), con su capacidad para procesar simultáneamente texto, imagen, audio y video, ha transformado el panorama de la Inteligencia Artificial en la educación superior, abriendo una nueva era de entornos de aprendizaje enriquecidos e interactivos. Sin embargo, la literatura científica carece de modelos de integración pedagógica estructurados que orienten la incorporación didáctica de estos modelos en contextos universitarios. Este artículo **propone** un marco tecno-pedagógico para la integración de Google Gemini en la educación superior a partir de una revisión cualitativa sistematizada de literatura y un análisis temático. El enfoque propuesto articula una taxonomía de modalidades de interacción multimodal, un conjunto de estrategias instruccionales operacionalizadas con roles docentes y estudiantiles diferenciados, y lineamientos éticos para la evaluación de aprendizajes mediados por inteligencia artificial. Las aplicaciones identificadas abarcan desde la tutoría visual interactiva y el andamiaje para la escritura académica hasta el análisis contextual en tiempo real, modalidad que extiende el aprendizaje hacia entornos físicos situados. A pesar de las oportunidades identificadas, el estudio subraya que ninguna estrategia de integración adquiere sentido pedagógico pleno sin un marco ético simultáneo que garantice la autoría intelectual del estudiante, prevenga la dependencia cognitiva y proteja su privacidad informacional. Este trabajo aporta un lenguaje conceptual compartido para docentes, diseñadores instruccionales e instituciones de educación superior, y abre líneas de investigación empírica sobre el impacto diferencial de las modalidades de interacción en el aprendizaje universitario.

Palabras clave: aprendizaje multimodal; inteligencia artificial generativa; educación superior; diseño instruccional; Google Gemini

ABSTRACT

The development of Multimodal Large Language Models (MLLMs), with their capacity to simultaneously process text, image, audio, and video, has transformed the landscape of Artificial Intelligence in higher education, opening a new era of enriched and interactive learning environments. However, the scientific literature lacks structured techno-pedagogical frameworks to guide the didactic incorporation of these models in university contexts. This article constructs a pedagogical integration framework for Google Gemini in higher education based on a systematized literature review combined with in-depth thematic analysis. The proposed framework articulates a taxonomy of multimodal interaction modalities, a set of operationalized instructional strategies with differentiated teacher and student roles, and ethical guidelines for the assessment of AI-mediated learning. The identified applications range from interactive visual tutoring and academic writing scaffolding to real-time contextual analysis, a modality that extends learning toward physically situated environments. Despite the opportunities identified, this study underscores that no integration strategy acquires full pedagogical meaning without a simultaneous ethical framework ensuring student intellectual authorship, preventing cognitive dependency, and protecting informational privacy. This work contributes to establishing a shared conceptual language for educators, instructional designers, and higher education institutions, and opens empirical research lines on the differential impact of interaction modalities on university learning outcomes.

Keywords: multimodal learning; generative artificial intelligence; higher education; instructional design; Google Gemini

1. Introducción

La integración de la Inteligencia Artificial Generativa (IAG) en la educación superior ha generado una de las disrupciones pedagógicas más significativas del siglo XXI, transformando radicalmente las dinámicas de enseñanza y aprendizaje (Bidry et al., 2025). Esta transición tecnológica no solo representa una evolución en las herramientas digitales disponibles, sino que plantea un desafío urgente frente a la obsolescencia de los modelos instruccionales tradicionales, los cuales resultan inadecuados para ecosistemas educativos hiperconectados (Sozon et al., 2025). A medida que las instituciones académicas se enfrentan a tasas crecientes de adopción de Inteligencia Artificial no regulada por parte de los estudiantes, surge la necesidad imperativa de replantear las políticas restrictivas vigentes. En su lugar, se requieren estrategias de integración curricular fundamentadas en evidencia empírica (Cordero et al., 2024). En este escenario, la capacidad de la IAG para personalizar el aprendizaje exige un análisis exhaustivo sobre sus efectos en el desarrollo cognitivo y profesional del estudiantado, con implicaciones significativas para la formación académica y profesional a escala global (Zhai et al., 2024).

Históricamente, la adopción de la inteligencia artificial en el ecosistema educativo se limitó a sistemas de tutoría basados en reglas y algoritmos predictivos con capacidades de interacción altamente restringidas (Ouyang & Jiao, 2021). La aparición de los Modelos de Lenguaje Grande (LLMs) democratizó el acceso a interfaces conversacionales, permitiendo la generación de texto coherente, la optimización de procesos de redacción académica y el ofrecimiento de retroalimentación formativa inmediata (Chu et al., 2026). No obstante, la naturaleza estrictamente textual de esta primera ola de IAG evidenció limitaciones al contrastarse con las teorías del procesamiento cognitivo, las cuales postulan que el aprendizaje humano es inherentemente multisensorial (Czerkawski, 2025). Esta disonancia impulsó la necesidad de herramientas que superaran el paradigma unimodal para alinearse de manera efectiva con los mecanismos perceptuales de asimilación de conocimiento propios del cerebro humano.

Para superar estas barreras, la investigación educativa contemporánea ha transitado hacia el aprendizaje multimodal, un enfoque que reconoce la complejidad de la cognición integrando simultáneamente estímulos visuales, auditivos y textuales. La literatura sobre analíticas de aprendizaje multimodal (MMLA) demuestra que la triangulación de diversos tipos de datos facilita una representación holística de los procesos cognitivos y conductuales, optimizando la predicción del rendimiento estudiantil (Varga-Atkins et al., 2025). Al proporcionar representaciones interconectadas de un mismo concepto teórico, se reduce drásticamente la carga cognitiva extrínseca y se fomenta una asimilación más robusta, especialmente en disciplinas científicas complejas. En consecuencia, el diseño instruccional universitario requiere imperiosamente tecnologías capaces de procesar información a través de múltiples canales para atender la variabilidad individual de habilidades, preferencias y estrategias de aprendizaje.

En este punto de inflexión tecnológica, Google Gemini emerge como un referente crítico al consolidarse como uno de los modelos fundacionales pioneros con una arquitectura multimodal nativa desde su concepción (Google Deepmind et al., 2023). A diferencia de los sistemas previos que acoplaban módulos de visión de forma periférica a un núcleo textual, Gemini fue entrenado mediante secuencias entrelazadas de texto, imágenes, audio y código de programación. Esta integración estructural subyacente le permite exhibir capacidades avanzadas de razonamiento intermodal, tales como la interpretación analítica de diagramas, la resolución visual de problemas matemáticos y la síntesis de respuestas audiovisuales coherentes. Por consiguiente, Gemini trasciende la función de procesador de lenguaje para constituirse en una plataforma de mediación cognitiva interactiva, redefiniendo las fronteras operativas del diseño tecno-pedagógico en las instituciones de educación superior.

A pesar de este significativo avance, la literatura científica actual exhibe un sesgo temporal, concentrando la inmensa mayoría de sus esfuerzos empíricos en la evaluación de LLMs unimodales de primera generación. Diversos estudios han documentado exhaustivamente las limitaciones críticas de estos sistemas tempranos, enfatizando riesgos sistémicos como la propagación de alucinaciones algorítmicas, el sesgo en los datos y la vulneración de la integridad académica (Lin et al., 2024; Shi et al., 2026; Xu et al., 2025). De igual manera, las investigaciones desde la psicología educativa advierten que el uso no guiado de generadores de texto puede derivar en la erosión de la memoria a largo plazo y en un aprendizaje eminentemente superficial (Zhai et al., 2024). Estas preocupaciones válidas han monopolizado el discurso académico, eclipsando la evaluación de herramientas multimodales que poseen mecánicas de interacción cualitativamente distintas.

En consecuencia, se evidencia un vacío sustancial en el estado del arte respecto a la implementación pedagógica de modelos multimodales nativos como Gemini en el ámbito de la educación superior (Imran & Almusharraf, 2024). Si bien la teoría del aprendizaje multimodal se encuentra sólidamente fundamentada, existe una disponibilidad limitada de marcos metodológicos contemporáneos que instruyan a los docentes sobre cómo orquestar las capacidades intermodales de este modelo específico. Cuando Gemini es abordado en la literatura, tiende a ser evaluado bajo los mismos parámetros restrictivos de los generadores de texto, ignorando su capacidad intrínseca para la tutoría audiovisual y el análisis de datos complejos. Este déficit de investigación aplicada impide que el sector académico capitalice la inversión tecnológica, dejando a los educadores desprovistos de directrices empíricas para una integración curricular innovadora y éticamente responsable.

El problema central de esta investigación radica en la carencia de estrategias tecno-pedagógicas estructuradas que permitan aprovechar las capacidades intermodales de Google Gemini para favorecer el aprendizaje universitario (Allazo, 2024). Sin un enfoque de aplicación empíricamente validado, el uso de esta herramienta por parte de la comunidad académica tiende a ser intuitivo, fragmentado y propenso a reproducir los patrones de dependencia cognitiva observados en la IA de primera generación (AIBlooshi, 2026). En consecuencia, se desconoce sistemáticamente cuáles son las adaptaciones didácticas más efectivas para integrar la utilidad multimodal de Gemini en la consecución de resultados de aprendizaje y en la mitigación de las limitaciones inherentes a los LLMs tradicionales.

Para abordar esta problemática, el presente estudio plantea la siguiente pregunta de investigación: ¿Cuáles son las principales aplicaciones didácticas y estrategias metodológicas reportadas en la literatura para integrar el modelo multimodal Gemini en la educación superior? De esta interrogante se desprenden tres objetivos específicos: 1) identificar casos de uso en los que la arquitectura multimodal de Gemini presenta potenciales ventajas pedagógicas según la literatura analizada; 2) diseñar un marco estructurado de estrategias instruccionales; y 3) formular lineamientos para la evaluación ética de estos aprendizajes. Desde una perspectiva teórica, se plantea que la integración regulada de las capacidades multimodales de Gemini, articulada mediante un diseño instruccional centrado en el estudiante, puede contribuir a una mayor profundidad del procesamiento analítico en comparación con enfoques unimodales.

2. Desarrollo

2.1. Evolución hacia la multimodalidad

La trayectoria de la inteligencia artificial generativa ha experimentado un salto cualitativo al transitar de Modelos de Lenguaje Grande (LLMs) estrictamente textuales a Modelos de Lenguaje Multimodal de Gran Tamaño (MLLMs). Esta evolución responde a la limitación inherente de los sistemas unimodales, los cuales, al procesar únicamente secuencias de texto, carecían de una conexión directa con mundo físico (Yin et al., 2024). En contraste, los MLLMs logran alinear diversas modalidades en un espacio de representación vectorial compartido, lo que permite realizar inferencias cruzadas entre datos de distinta naturaleza. Esta capacidad de interpretar y sintetizar información audiovisual y textual de manera simultánea representa un avance significativo, al ampliar las capacidades de los sistemas de IA más allá del procesamiento secuencial de texto hacia enfoques multimodales de análisis e integración de información.

2.2. Arquitectura y capacidades de Gemini

Dentro de este nuevo ecosistema tecnológico, Google Gemini se distingue estructuralmente por poseer una arquitectura multimodal nativa desde su fase de preentrenamiento fundacional. A diferencia de modelos contemporáneos que ensamblan redes neuronales independientes para procesar visión y texto de manera secuencial, Gemini fue diseñado para ingerir y procesar secuencias entrelazadas de texto, imagen, audio y código computacional de

forma simultánea (Google Deepmind et al., 2023). En el ámbito de la educación superior, estos avances permiten plantear que sistemas como Gemini podrían desempeñar funciones de apoyo tutorial, integrando el análisis de representaciones gráficas, la comprensión de consultas verbales y la generación de retroalimentación multimodal en tiempo real, lo cual abre nuevas posibilidades para entornos de aprendizaje interactivos y adaptativos.

2.3. Teoría del aprendizaje multimodal y cognición dual

La pertinencia pedagógica de herramientas como Gemini encuentra respaldo teórico en la Teoría Cognitiva del Aprendizaje Multimedia, la cual postula que el cerebro humano procesa la información visual y verbal a través de canales neurológicos independientes (Mayer, 2001). Desde este marco, la interacción con un modelo multimodal podría favorecer la activación coordinada de canales visuales y verbales, siempre que el diseño de la tarea evite sobrecarga y distribuya de forma pertinente la información.

Adicionalmente, el diseño instruccional apoyado en una IA multimodal nativa se alinea con los postulados centrales de la Teoría de la Carga Cognitiva (Sweller et al., 2019). La presentación simultánea y contextualizada de explicaciones textuales integradas con diagramas analizados por IA minimiza el efecto de atención dividida, reduciendo la carga cognitiva extrínseca impuesta sobre la memoria de trabajo. De este modo, al externalizar hacia la arquitectura de Gemini el procesamiento estructural que no constituye el objetivo central de aprendizaje, los estudiantes pueden redirigir sus recursos cognitivos hacia la asimilación del material y la resolución de problemas, optimizando la adquisición de competencias académicas de nivel superior.

3. Metodología

3.1 Enfoque de Investigación

El presente estudio adopta un enfoque cualitativo de orientación interpretativa, el cual resulta epistemológicamente coherente con la naturaleza del objeto de estudio: comprender, describir y analizar cómo las capacidades multimodales de Google Gemini pueden ser pedagógicamente integradas en contextos de educación superior. Bajo esta premisa, la investigación privilegia la densidad conceptual frente a la generalización estadística, facilitando una comprensión profunda mediante el análisis sistemático de la literatura especializada (Creswell & Poth, 2016). Esta orientación metodológica permite sintetizar

perspectivas de la psicología cognitiva, el diseño instruccional y la inteligencia artificial educativa en un modelo integral y pragmático. Por consiguiente, este enfoque no restringe el alcance analítico; por el contrario, lo potencia al supeditar la cuantificación de variables aisladas a la construcción de significados contextualizados.

3.2. Tipo de Estudio

La investigación se clasifica como descriptivo-interpretativa. En su dimensión descriptiva, el estudio caracteriza sistemáticamente las aplicaciones pedagógicas que la arquitectura multimodal nativa de Gemini habilita en entornos universitarios, identificando sus funcionalidades diferenciales respecto a los modelos de lenguaje unimodales de generación previa. En su dimensión interpretativa, el estudio trasciende la mera enumeración funcional para analizar los mecanismos mediante los cuales dichas aplicaciones inciden en los procesos cognitivos del estudiantado, en la reducción de la carga cognitiva extrínseca y en la profundización del aprendizaje.

Esta doble dimensión es indispensable para alcanzar los tres objetivos específicos formulados. La articulación de ambos niveles analíticos garantiza que los hallazgos superen el plano descriptivo y contribuyan con interpretaciones teóricamente fundamentadas sobre las condiciones que favorecen la efectividad tecno-pedagógica del modelo.

3.3. Diseño de Investigación

El diseño metodológico corresponde a una revisión cualitativa sistematizada de literatura combinada con análisis temático inductivo-deductivo. Este diseño es reconocido en la literatura especializada como adecuado para la construcción de marcos conceptuales y pedagógicos en áreas donde la evidencia empírica directa es limitada o heterogénea (Snyder, 2019).

Construcción del corpus documental. La identificación de fuentes se realizó mediante búsquedas sistemáticas en las bases de datos Scopus, Web of Science, ERIC y Google Scholar. Los descriptores de búsqueda empleados fueron: *multimodal learning*, *Google Gemini education*, *multimodal large language models*, *generative AI higher education* e *instructional design artificial intelligence*, en sus variantes en español e inglés, combinados mediante los operadores booleanos AND y OR. El período de búsqueda abarcó los años 2022-2026, priorizando publicaciones posteriores a 2022 dada la naturaleza emergente del objeto de estudio.

Criterios de inclusión. Se consideraron fuentes tanto en español como en inglés sobre: a) artículos científicos arbitrados, capítulos de libro académico o reportes técnicos institucionales; b) publicaciones que abordaran explícitamente la integración de IA generativa o modelos multimodales en contextos educativos formales; c) trabajos que discutieran implicaciones pedagógicas, cognitivas o éticas del uso de LLMs o MLLMs en educación superior; y d) disponibilidad del texto completo en acceso abierto o mediante suscripción institucional.

Criterios de exclusión. Se descartaron: a) materiales de divulgación sin proceso de revisión por pares; b) documentos cuyo objeto de estudio fuera exclusivamente la IA en contextos no educativos; c) publicaciones anteriores a 2019, salvo obras teóricas fundacionales en cognición y diseño instruccional; y d) textos duplicados o versiones previas de trabajos ya incluidos.

Proceso de selección y corpus final. La búsqueda inicial arrojó 312 registros. Tras eliminar duplicados se obtuvieron 287 registros únicos. La revisión de títulos y resúmenes redujo el conjunto a 61 documentos potencialmente relevantes. Finalmente, tras la lectura completa y aplicación estricta de los criterios señalados, el corpus definitivo quedó conformado por 28 fuentes primarias y secundarias.

Este diseño permitió no solo identificar patrones, categorías y tendencias recurrentes en la literatura revisada, sino también desarrollar una elaboración conceptual orientada a la construcción de un marco pedagógico para la integración de Gemini en la educación superior.

3.4. Construcción del marco pedagógico

A partir de la codificación temática del corpus, se consolidaron categorías relativas a modalidades de interacción, funciones pedagógicas, demandas cognitivas e implicaciones éticas. Sobre esa base, el marco fue construido en tres niveles: clasificación de modalidades, traducción de hallazgos en secuencias instruccionales y sistematización de orientaciones éticas operativas. Es importante precisar que esta sección describe el procedimiento seguido para construir dicho marco; los productos resultantes se presentan en la sección de Resultados. En el primer nivel, las categorías emergentes del análisis temático fueron organizadas y clasificadas según tres criterios operativos: la modalidad de interacción predominante identificada en la literatura, el nivel en la Taxonomía revisada de Bloom que

cada modalidad tendía a activar según los estudios analizados, y el tipo de retroalimentación formativa que cada interacción habilitaba.

En el segundo nivel, las categorías clasificadas fueron traducidas en secuencias metodológicas explícitas mediante un proceso de especificación operacional. Para cada estrategia instruccional se definieron cuatro componentes a partir de la literatura analizada: el rol esperado del estudiante, el rol del docente facilitador, el tipo de interacción multimodal requerida y los indicadores observables de profundización cognitiva. En el tercer nivel, las implicaciones éticas identificadas de forma recurrente en el corpus fueron sistematizadas y organizadas en torno a tres dimensiones críticas: la verificación de la autoría intelectual, la transparencia en el uso de la IA, la prevención de la dependencia cognitiva y la equidad en el acceso tecnológico. Para cada dimensión se formularon orientaciones operativas dirigidas a comités curriculares y docentes universitarios.

4. Resultados

Esta sección presenta los resultados del estudio entendidos como una síntesis analítica de la literatura revisada y, al mismo tiempo, como una propuesta de marco pedagógico elaborada a partir de dicha revisión. Por ello, los apartados que siguen combinan hallazgos identificados en el corpus con una elaboración conceptual propia, orientada a organizar taxonomías, estrategias instruccionales y lineamientos éticos para el uso de Gemini en educación superior.

4.1. Taxonomía de aplicaciones didácticas de Gemini

La primera categoría emergente agrupa los casos de uso en los que la arquitectura multimodal nativa de Gemini presenta potenciales ventajas pedagógicas respecto a los modelos de lenguaje unimodales en determinadas tareas. A partir del análisis temático del corpus, se construyó una taxonomía de cuatro dimensiones articuladas: modalidad de interacción, nivel cognitivo activado según la Taxonomía Revisada de Bloom (Anderson & Krathwohl, 2001), tipo de retroalimentación formativa habilitada y disciplinas con mayor potencial de aplicación. Esta estructura permite a los docentes seleccionar el tipo de interacción más pertinente para cada objetivo de aprendizaje declarado en sus programas curriculares.

Tabla 1

Taxonomía de aplicaciones didácticas de Gemini en educación superior

Modalidad de interacción	Nivel Bloom (2001)	Retroalimentación habilitada	Disciplinas con mayor potencial
Texto a Imagen (análisis visual)	Comprender / Analizar	Corrección conceptual inmediata sobre representaciones gráficas.	Anatomía, ingeniería, arquitectura, geografía, biología.
Imagen a Texto (descripción e interpretación)	Comprender / Evaluar	Orientación interpretativa sobre gráficos, tablas y diagramas complejos.	Medicina, química, física, economía, ciencias sociales, informática.
Voz ↔ Voz (síntesis auditiva)	Comprender / Analizar	Explicación multicanal de conceptos abstractos con refuerzo auditivo.	Idiomas extranjeros, comunicación, pedagogía, psicología educativa.
Video en tiempo real → Análisis contextual (interpretación espacial)	Aplicar / Evaluar	Retroalimentación inmediata sobre procedimientos, objetos y entornos físicos capturados por cámara.	Laboratorios de ciencias, clínicas de salud, talleres de ingeniería, práctica docente, trabajo de campo.
Texto ↔ Texto (diálogo conceptual y argumentativo)	Comprender / Analizar / Evaluar	Reformulación explicativa, diálogo socrático simulado y retroalimentación argumentativa sobre producción escrita en desarrollo.	Todas las disciplinas con componente de escritura académica y argumentación.
Texto ↔ Código (generación y depuración)	Aplicar / Crear	Retroalimentación iterativa sobre errores lógicos y sintácticos.	Programación, estadística computacional, ciencias de datos, matemáticas.

Nota: Las disciplinas indicadas son orientativas; la aplicabilidad puede extenderse a otros campos según el objetivo instruccional.

El análisis del corpus evidenció que la interacción texto a texto y la conversión de texto a imagen constituyen los casos de uso con mayor respaldo en el período estudiado, particularmente en disciplinas STEM, donde la interpretación analítica de diagramas, modelos y representaciones matemáticas gráficas representa una demanda cognitiva recurrente para el estudiante universitario (Bewersdorff et al., 2025). Se espera que la capacidad de Gemini para observar una representación visual y proporcionar explicación

textual y auditiva simultánea active ambos canales de procesamiento cognitivo descritos en la Teoría Cognitiva del Aprendizaje Multimedia, reduciendo el efecto de atención dividida y facilitando la construcción de esquemas mentales más integrados.

Por su parte, la modalidad texto a código presenta, por su naturaleza procedimental y constructiva, condiciones favorables para la activación de niveles cognitivos superiores, particularmente los niveles de aplicación y creación. La retroalimentación iterativa sobre errores lógicos, la explicación de por qué un fragmento de código no funciona y la generación de variantes alternativas configuran un ciclo de aprendizaje activo alineado con los Primeros Principios de Instrucción, propuestos por Merrill (2002) donde el conocimiento es demostrado, aplicado e integrado en el mundo real del estudiante. La modalidad de video en tiempo real mediante Gemini Live representa posiblemente la expansión de mayor potencial transformador del paradigma multimodal hacia el aprendizaje situado. A diferencia de las interacciones precedentes, que operan sobre representaciones digitales de la realidad —imágenes, documentos, código—, esta funcionalidad permite que Gemini observe directamente el entorno del estudiante a través de la cámara del dispositivo, analizando objetos, procedimientos, espacios y situaciones en tiempo real para proporcionar retroalimentación inmediata y contextualmente anclada (Imran & Almusharraf, 2024).

4.2. Estrategias para el uso de Gemini

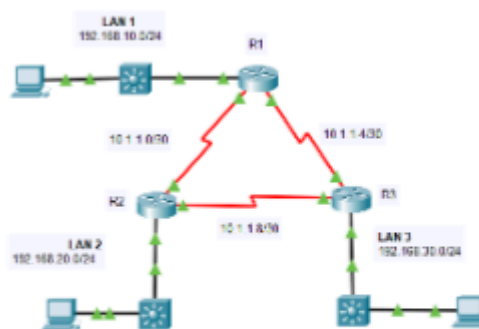
La segunda categoría de resultados corresponde al marco de aplicaciones y estrategias instruccionales construido a partir de la codificación temática de la documentación. Cada estrategia fue definida a partir de cuatro componentes operacionales: rol del estudiante, rol del docente facilitador, tipo de interacción multimodal predominante e indicadores observables de profundización cognitiva. Las estrategias se organizan en función del nivel cognitivo que priorizan y son aplicables a distintas modalidades de enseñanza —presencial, híbrida y en línea—.

4.2.1. Estrategia de tutoría visual interactiva

Esta estrategia se apoya en el principio de modalidad de la Teoría Cognitiva del Aprendizaje Multimedia (Mayer, 2001), según el cual la presentación simultánea de información en canales visual y verbal optimiza la capacidad de procesamiento en la memoria de trabajo. Bajo este enfoque, el uso de Gemini se presenta como una herramienta de mediación, donde el MLLM actúa como puente cognitivo capaz de interpretar y analizar imágenes científicas complejas en tiempo real. La multimodalidad permite realizar un análisis espacial de la

imagen, identificando los principales elementos, contribuyendo con un contexto profundo, añadiendo información técnica y reduciendo la abstracción de la imagen. Gemini transforma una representación estática en una experiencia de aprendizaje dinámica, clara y adaptada al nivel de comprensión de cada usuario.

El estudiante carga en Gemini una imagen de un fenómeno complejo —diagrama científico, plano arquitectónico, gráfico estadístico, imagen histológica, topología de red— y formula una pregunta analítica. Gemini describe la imagen, identifica sus componentes estructurales, establece relaciones causales y proporciona retroalimentación correctiva si el estudiante ofrece una interpretación inicial incorrecta. El docente diseña previamente la consigna de interacción y opera como guía metacognitiva durante la sesión.



Actúa como experto en redes de comunicaciones, explicame de manera resumida la topología con relación al protocolo de enrutamiento OSPF.



Esta topología en malla (full-mesh) es ideal para **OSPF**, ya que utiliza enlaces punto a punto (/30) que eliminan la necesidad de elegir DR/BDR, optimizando la convergencia.

Todo se configura en el **Área 0**, donde el protocolo calculará automáticamente la ruta más eficiente hacia las LAN basándose en el costo del ancho de banda, aprovechando la redundancia para garantizar alta disponibilidad.

¿Quieres que te prepare los comandos de configuración para uno de los routers?

Figura 1. Ejemplo de interacción visual imagen-texto: el estudiante sube un esquema gráfico de red y el modelo Gemini facilita información explicativa. Nota. Captura de pantalla de interacción realizada por el autor con Google Gemini.

Actúa como experto en redes, certificaciones Cisco y simulador packet tracer. Genera la topología de una red WAN con tres routers centrales mediante enlaces seriales punto a punto. Cada router actúa como nodo de acceso para una red local (LAN) independiente, las cuales están integradas mediante switches.

Ver razonamiento

¡Claro que sí! Aquí tienes una topología de red WAN profesional, diseñada como si estuvieras trabajando directamente en el simulador Cisco Packet Tracer.

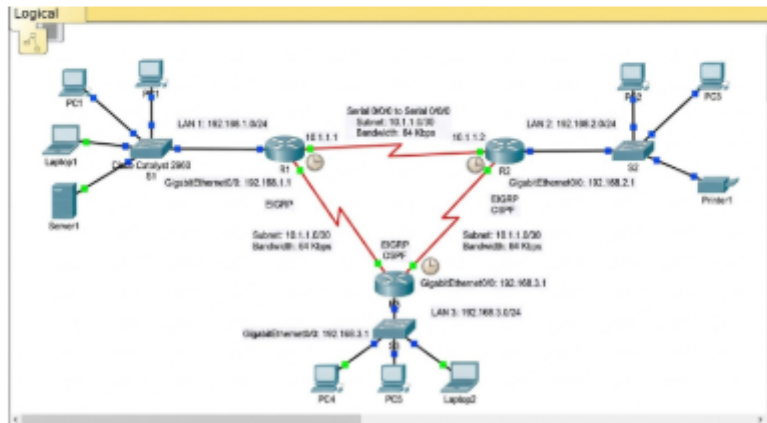


Figura 2. Ejemplo de interacción visual texto-imagen: el estudiante comparte información y el modelo Gemini genera una imagen ilustrativa de la topología de red.

Tabla 2. Componentes operacionales de la estrategia de tutoría visual interactiva

Rol del estudiante	Rol del docente	Interacción multimodal	Indicadores de logro
Activo: interroga, interpreta, reformula hipótesis con la guía de Gemini.	Facilitador: diseña la consigna visual, monitorea el proceso y orienta la reflexión metacognitiva.	Carga de imagen + interrogación textual + respuesta audiovisual de Gemini.	Precisión en la descripción de componentes; calidad de la inferencia causal; corrección conceptual progresiva.

4.2.2. Estrategia de tutoría auditiva interactiva

Esta estrategia aprovecha el modo de voz nativo —Gemini Live— para establecer una interacción continua mediada por lenguaje hablado entre el estudiante y el modelo. Su fundamento teórico se sustenta en evidencia empírica sobre el impacto de los sistemas conversacionales basados en IA en el desarrollo de competencias comunicativas: estudios

recientes señalan que estos sistemas abordan desafíos pedagógicos persistentes —como la limitada práctica, la ansiedad ante la expresión pública y la falta de apoyo individualizado— al ofrecer entornos personalizados y de baja presión (Ding & Muhyiddin B Yusof, 2025) . En este sentido, esta modalidad supera a los chatbots textuales convencionales al incorporar prosodia, entonación y mayor naturalidad conversacional, elementos que aproximan la experiencia al diálogo humano y favorecen la disposición del estudiante a comunicarse en contextos académicos exigentes.

A través de este tipo de interacción, el estudiante activa principalmente procesos de comprensión, análisis y evaluación, en función del diseño instruccional propuesto por el docente. Una de sus aplicaciones con mayor potencial pedagógico es el diálogo socrático, en el que el estudiante sostiene una conversación argumentativa sobre un concepto o problema disciplinar. En este escenario, el modelo no ofrece respuestas directas, sino que formula preguntas progresivas que examinan la coherencia del razonamiento, lo que exige al estudiante defender sus posturas, identificar contradicciones y reformular sus argumentos en tiempo real. Esta dinámica implica un procesamiento verbal activo y sostenido que la modalidad textual no reproduce con igual inmediatez ni con la misma distribución de la carga cognitiva en el canal auditivo

Otra aplicación relevante es la preparación de exposiciones y debates académicos. El estudiante puede emplear esta herramienta como audiencia simulada para ensayar presentaciones, recibir retroalimentación sobre claridad expositiva, precisión conceptual, estructura argumentativa y fluidez, y repetir el ejercicio antes de una evaluación formal. La evidencia disponible indica que estos entornos de práctica mediados por IA mejoran el desempeño comunicativo y el compromiso, al tiempo que reducen la ansiedad y favorecen una participación sostenida (Fathi et al., 2024). sta combinación de retroalimentación formativa y ausencia de juicio evaluativo convierte a esta modalidad en un andamio eficaz para estudiantes con altos niveles de ansiedad comunicativa, quienes encuentran en ella un espacio de práctica seguro y repetible.

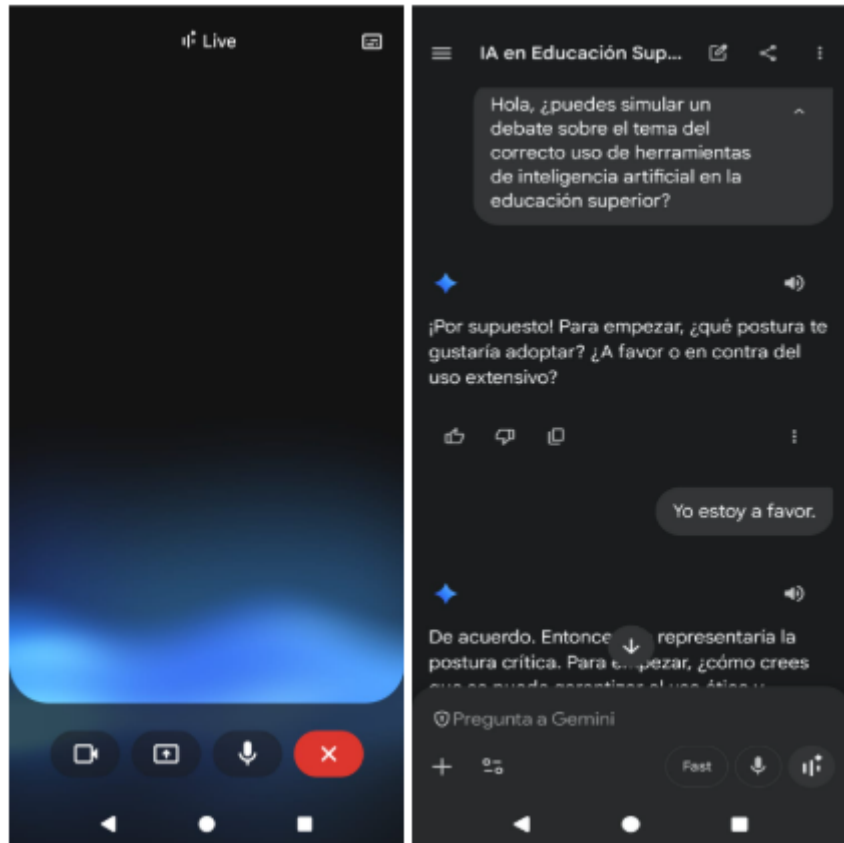


Figura 3. Ejemplo de interacción auditiva: el estudiante realiza una petición de manera verbal y el modelo utiliza el modo voz para dar respuesta. Nota. Captura de pantalla de interacción realizada por el autor con Google Gemini Live.

La tercera aplicación corresponde a la práctica de idiomas extranjeros. En este uso, el estudiante interactúa con Gemini Live en la lengua objetivo, solicitando al modelo que corrija su pronunciación, identifique errores de estructura gramatical o amplíe el vocabulario en contexto. La naturaleza bidireccional y adaptativa del modo de voz permite replicar situaciones comunicativas reales —entrevistas profesionales, presentaciones académicas, negociaciones— con un nivel de exigencia lingüística calibrable según el perfil del estudiante. El docente cumple aquí un rol de diseñador del escenario comunicativo: define la situación, el registro lingüístico exigido y los criterios de evaluación que permitirán distinguir el progreso genuino del estudiante de la simple fluidez conversacional con la IA.

Tabla 3: Componentes operacionales de la Estrategia de Tutoría Auditiva Interactiva

Rol del estudiante	Rol del docente	Interacción multimodal	Indicadores de logro
Activo-oral: argumenta, expone, debate y reformula posiciones en tiempo real mediante voz.	Diseñador de escenario: define el contexto comunicativo, el registro exigido y los criterios evaluativos; supervisa la calidad del diálogo.	Interacción oral bidireccional con Gemini en modo voz; retroalimentación auditiva inmediata sobre contenido, fluidez y estructura argumentativa.	Coherencia argumentativa oral; precisión léxica y gramatical; calidad de la exposición en evaluación.

4.2.3. Estrategia de tutoría contextual en tiempo real

A diferencia de las modalidades previas, que operan sobre representaciones digitales de la realidad, Gemini Live permite que el modelo observe directamente el entorno físico del estudiante a través de la cámara del dispositivo, analizando objetos, instrumentos y procedimientos en tiempo real para proporcionar retroalimentación contextual inmediata. Su aplicación más relevante se sitúa en entornos de laboratorio universitario, donde el estudiante puede apuntar la cámara hacia un montaje experimental, un instrumento de medición o una reacción química en curso, y recibir de Gemini identificación de errores procedimentales, orientación sobre la secuencia correcta de pasos y explicación de los principios teóricos subyacentes, sin interrumpir la práctica. Los MLLMs poseen la capacidad de evaluar contenido textual y visual —como gráficos y diagramas— de manera simultánea, lo que permite una retroalimentación personalizada y comprensiva que facilita una comprensión más profunda de los procesos cognitivos del estudiante y de su habilidad para comunicar ciencia compleja en modos escritos y visuales (Bewersdorff et al., 2025).

La relevancia pedagógica de esta modalidad se sustenta en el ofrecimiento de retroalimentación adaptativa en tiempo real constituye el rol de mayor impacto de la IA en entornos de aprendizaje situado, con concentración predominante en entrenamiento procedimental y educación STEM (Liu et al., 2025). Esta interacción activa los niveles de aplicación y evaluación, con potencial de alcanzar el nivel de creación cuando el análisis contextual se utiliza como insumo para diseñar soluciones originales a problemas identificados en el entorno observado. Dado que su aplicación pedagógica sistemática se encuentra en fase emergente de documentación empírica, su incorporación curricular debe tratarse como hipótesis de diseño instruccional sujeta a validación futura.

Tabla 4: Componentes operacionales de la estrategia de tutoría contextual en tiempo real

Rol del estudiante	Rol del docente	Interacción multimodal	Indicadores de logro
Activo-procedimental: ejecuta, muestra y consulta en tiempo real mediante la cámara del dispositivo.	Diseñador de práctica: define el protocolo de uso, los criterios procedimentales y los indicadores evaluativos del desempeño observado.	Transmisión de video en tiempo real → análisis contextual de Gemini → retroalimentación oral e inmediata sobre el entorno físico capturado.	Corrección procedimental progresiva; precisión en la identificación de instrumentos; retroalimentación sin interrupción el flujo de trabajo.

4.2.4. Estrategia de andamiaje iterativo para la escritura académica

Basada en la interacción texto a texto y en el principio de retroalimentación formativa progresiva, esta estrategia utiliza a Gemini como andamio cognitivo externo que apoya al estudiante en el desarrollo de competencias de escritura sin sustituir su agencia intelectual. Activa los niveles de análisis y evaluación de la taxonomía de Bloom. Su pertinencia es especialmente útil en contextos de escritura científica, donde el dominio del lenguaje técnico representa uno de los principales obstáculos para el estudiante universitario: la literatura documenta que quienes se inician en la escritura académica suelen percibir el lenguaje científico como altamente restrictivo y difícil de apropiar (Aldayel et al., 2026). En este escenario, Gemini actúa como mediador lingüístico capaz de ampliar de forma adaptativa la información textual, explicando terminología técnica en contexto, reformulando definiciones con distinto nivel de abstracción o clarificando conceptos encontrados durante la lectura de fuentes primarias, complementando así el material unimodal previo con una segunda capa de información textual accesible y personalizada.

El estudiante carga un borrador de texto académico —ensayo, informe o tesis parcial— y solicita retroalimentación específica sobre dimensiones declaradas: coherencia argumentativa, pertinencia de evidencias, precisión conceptual o estructura lógica. Gemini analiza el documento y ofrece comentarios diferenciados por sección, sin reescribir el texto por el estudiante. La interacción es cíclica: el estudiante revisa, vuelve a cargar el documento mejorado y recibe una nueva ronda de retroalimentación. El docente debe supervisar la evaluación de los contenidos de manera fundamental para prevenir la dependencia cognitiva y asegurar la integridad académica (Eigner & Händler, 2024).

Tabla 5: Componentes operacionales de la estrategia de andamiaje iterativo para la escritura académica

Rol del estudiante	Rol del docente	Interacción multimodal	Indicadores de logro
Activo-reflexivo: carga borradores, solicita retroalimentación sobre dimensiones específicas, revisa y reformula su propio texto de forma cíclica.	Evaluador dual: supervisa los registros de interacción con Gemini y valida que el producto final refleje el razonamiento propio del estudiante.	Carga de documento textual → análisis seccional de Gemini → retroalimentación diferenciada por dimensión → revisión del estudiante → nueva carga iterativa.	Mejora progresiva de la coherencia argumentativa; precisión en el uso del lenguaje técnico disciplinar; evidencia de autoría intelectual en la versión final.

4.3. Lineamientos éticos para la evaluación de aprendizajes mediados por Gemini

La tercera categoría de resultados articula los lineamientos éticos derivados del análisis temático, organizados en torno a tres dimensiones críticas identificadas en la literatura: la verificación de la autoría intelectual, la prevención de la dependencia cognitiva y la privacidad de datos y protección del estudiante. Estas dimensiones responden directamente a las preocupaciones sistémicas documentadas respecto a los riesgos éticos de la inteligencia artificial en contextos educativos y constituyen orientaciones operativas dirigidas a comités curriculares y docentes universitarios.

4.3.1. Verificación de la autoría intelectual

La integración de Gemini en actividades evaluativas exige que las instituciones adopten protocolos explícitos para distinguir la contribución cognitiva del estudiante de los contenidos generados por la IA. El presente documento propone tres mecanismos complementarios de verificación: (a) el portafolio de proceso, que documenta la trayectoria de interacciones entre el estudiante y Gemini, evidenciando el desarrollo progresivo del pensamiento propio; (b) la defensa oral del producto, en la que el estudiante explica y argumenta sus decisiones frente al docente sin apoyo de la herramienta; y (c) la declaración de uso de IA, un componente reflexivo donde el estudiante describe explícitamente cómo utilizó Gemini y qué aportó por sí mismo al trabajo final.

El portafolio constituye el mecanismo de verificación de mayor riqueza evidencial dentro del marco propuesto, dado que su función no se limita a demostrar el producto final sino a

reconstruir la trayectoria cognitiva completa a lo largo de la actividad. La evidencia disponible señala que la evaluación en entornos con IA debe transformarse para cultivar habilidades de aprendizaje autorregulado, aprendizaje responsable e integridad, dimensiones que el portafolio de proceso materializa al exigir al estudiante una gestión consciente y documentada de su interacción con la herramienta (Xia et al., 2024).

En términos operativos, el docente puede, mediante la rúbrica de evaluación, requerir la incorporación de recursos multimedia —principalmente video e imagen— que documenten y evidencien los pasos seguidos por el estudiante durante el desarrollo de la actividad. Esta exigencia cumple una doble función pedagógica: por una parte, incrementa la autenticidad de la evidencia al anclarla a registros que la IA no puede generar en nombre del estudiante; por otra, activa competencias de documentación y comunicación visual que son en sí mismas objetivos de aprendizaje en múltiples disciplinas.

La defensa oral constituye actualmente el mecanismo de verificación de mayor practicidad para distinguir la autoría intelectual genuina frente al contenido generado por IA, precisamente porque opera en el único registro que la tecnología actual no puede suplantar: el razonamiento en tiempo real, sin andamio y bajo interpelación directa. La investigación empírica evidencia que las actividades escritas resultan altamente vulnerables a la manipulación por IA generativa, y que los docentes no logran distinguir con fiabilidad los trabajos asistidos por IA de los producidos íntegramente por el estudiante (Hadra et al., 2026), lo que lleva a recomendar el desplazamiento progresivo hacia evaluaciones sincrónicas e interpersonales —exámenes orales, presentaciones en vivo y entrevistas reflexivas— estrategias que dificultan la delegación de la tarea a un MLLM (Alsharefeen & Al Sayari, 2025).

La declaración de uso de IA es un componente reflexivo y metacognitivo que complementa los mecanismos anteriores al exigir al estudiante una narración explícita sobre cómo utilizó Gemini, con qué propósito, en qué momentos del proceso y qué aportó por sí mismo al producto final. Su valor pedagógico no radica únicamente en la dimensión de integridad sino en que obliga al estudiante a tomar conciencia del límite entre el andamio cognitivo y la delegación del pensamiento, fortaleciendo la autorregulación del aprendizaje. Sin embargo, la evidencia empírica advierte sobre sus limitaciones como mecanismo aislado: el miedo a las repercusiones académicas, las directrices ambiguas, la aplicación inconsistente por parte de los docentes y la influencia del grupo de pares son barreras determinantes para el

cumplimiento honesto de las declaraciones de uso de LLMs (Gonsalves, 2025). Por ello, la declaración de uso de IA adquiere pleno valor pedagógico únicamente cuando se enmarca en una cultura institucional de transparencia consolidada, no como requisito burocrático aislado, sino como práctica reflexiva integrada en el diseño instruccional desde el inicio del curso. Estos mecanismos no tienen como propósito sancionar el uso de la herramienta, sino asegurar que la IA opere como soporte del aprendizaje y no como sustituto de él.

4.3.2. Prevención de la dependencia cognitiva

El riesgo de dependencia cognitiva representa una dimensión ética urgente en la integración de IA generativa en la educación superior. Esta práctica se caracteriza por la tendencia del estudiante a sustituir el pensamiento independiente en actividades cognitivas de alto nivel —como la comprensión profunda, el razonamiento, la conceptualización y el juicio— lo que a largo plazo puede consolidarse en inercia cognitiva. Es pertinente establecer la siguiente distinción: no todo uso frecuente de Gemini constituye dependencia, sino únicamente aquel en que el estudiante externaliza los procesos cognitivos que debería desarrollar internamente. La sobredependencia en sistemas de IA ocurre cuando los usuarios aceptan las recomendaciones generadas sin cuestionarlas, lo que conduce a una reducción progresiva de la capacidad de toma de decisiones autónomas.

Para mitigar este riesgo en el contexto multimodal de Gemini, el análisis temático permitió identificar tres principios operativos que el docente debe incorporar de forma explícita en el diseño instruccional. El primero es el principio de interrogación progresiva, que establece que las consignas deben diseñarse de modo que el estudiante formule preguntas analíticas a Gemini antes de recibir respuestas, invirtiendo la lógica de consumo pasivo e impulsando un procesamiento activo previo a la interacción. El segundo es el principio de validación independiente, que exige al estudiante contrastar sistemáticamente las respuestas de Gemini con fuentes primarias académicas, ejerciendo un juicio evaluativo propio sobre la calidad y pertinencia de la información recibida. El tercero es el principio de transferencia sin andamio, que incluye en la evaluación final tareas de demostración competencial en las que el estudiante resuelve un problema equivalente sin acceso a la herramienta, verificando que el aprendizaje fue genuinamente internalizado y no simplemente delegado.

4.3.3. Privacidad de datos y protección del estudiante

La integración de Gemini en actividades evaluativas genera, de forma inherente, un flujo continuo de datos sobre las interacciones y procesos de aprendizaje del estudiante: las preguntas formuladas, los errores cometidos, los patrones de interacción, las dudas recurrentes y las estrategias de resolución de problemas son procesados por empresas externas a la institución, sobre las cuales el estudiante tiene un control limitado o nulo. Esta dimensión convierte la privacidad de datos en una condición ética previa a cualquier implementación pedagógica de la herramienta, no en una consideración secundaria. En el caso específico de Gemini, la ausencia de políticas institucionales explícitas sobre el tratamiento de los datos generados durante las interacciones académicas sitúa al estudiante en una posición de vulnerabilidad informacional que las instituciones de educación superior tienen la responsabilidad de mitigar.

El lineamiento propuesto establece tres obligaciones institucionales operativas. La primera es la obligación de información previa: antes de exigir el uso de Gemini en actividades evaluativas, la institución debe informar al estudiante de forma explícita y comprensible qué datos genera su interacción con la herramienta, cómo son tratados por Google, bajo qué jurisdicción legal se procesan y durante qué período se conservan. La segunda es la obligación de política curricular de privacidad: cada asignatura que integre Gemini debe incluir en su sílabo una sección específica sobre las condiciones de privacidad aplicables al uso de la herramienta, diferenciando los datos que permanecen en el dispositivo del estudiante de los que se transmiten a servidores externos. Los marcos éticos vigentes identifican la privacidad, el sesgo y la transparencia como las tres preocupaciones centrales del uso de IA en educación, señalando que los estudiantes tienen un control limitado sobre sus datos y entornos digitales, lo que exige políticas institucionales explícitas de protección y gobernanza (Ofosu-Asare, 2025). La tercera es la obligación de alternativas equivalentes: ningún estudiante debe verse penalizado académicamente por negarse a utilizar una plataforma comercial que procese sus datos personales; en consecuencia, el docente debe diseñar actividades alternativas equivalentes para quienes no consientan el procesamiento de sus datos por parte de Gemini, garantizando que la protección de la privacidad no se convierta en una desventaja académica.

Tabla 6: Síntesis de los lineamientos éticos para la evaluación de aprendizajes mediados por Gemini.

Dimensión ética	Riesgo que mitiga	Mecanismo operativo propuesto
Verificación de la autoría intelectual.	Suplantación del trabajo cognitivo del estudiante por contenido generado por IA	Portafolio de proceso con evidencia multimedia + defensa oral sin andamio + declaración reflexiva de uso de IA
Prevención de la dependencia cognitiva.	Erosión del pensamiento crítico, inercia cognitiva y reducción de la toma de decisiones autónoma	Interrogación progresiva + validación independiente con fuentes primarias + transferencia sin andamio en evaluación final
Privacidad de datos y protección del estudiante	Vulnerabilidad informacional por procesamiento de datos de aprendizaje en plataformas comerciales externas	Obligación de información previa + política curricular de privacidad en el syllabus + diseño de actividades alternativas equivalentes

5. Discusión

Este estudio propone una arquitectura conceptual para pensar su integración didáctica en educación superior desde una perspectiva multimodal, instruccional y ética. Las capacidades multimodales nativas de Google Gemini pueden ser pedagógicamente integradas en la educación superior y responden directamente al vacío identificado en la literatura: la ausencia de propuestas didácticas que trasciendan la evaluación de los modelos de lenguaje de primera generación y capitalicen las funcionalidades diferenciales de los modelos fundacionales multimodales.

El presente estudio presenta una taxonomía de seis modalidades de interacción que potencian el accionar de Gemini en dimensiones concretas. Este resultado permite sostener que la integración regulada de las capacidades multimodales, bajo un diseño instruccional centrado en el estudiante, ofrece condiciones más favorables para la profundización del procesamiento analítico que el uso aislado de herramientas unimodales. La taxonomía construida evidencia que la ventaja pedagógica de Gemini no reside únicamente en su capacidad para generar texto coherente —atributo que comparte con los LLMs de generación previa— sino en su arquitectura de procesamiento simultáneo de múltiples modalidades, la cual activa de forma coordinada los canales visual y verbal de procesamiento cognitivo descritos en la Teoría Cognitiva del Aprendizaje Multimedia. Este alineamiento entre arquitectura tecnológica y mecanismos cognitivos es, precisamente, lo que distingue a Gemini como objeto de estudio pedagógico diferenciado.

El marco de estrategias instruccionales planteadas traduce esta ventaja arquitectónica en

secuencias metodológicas aplicables. Resulta significativo que las cuatro estrategias desarrolladas no sean intercambiables ni jerárquicas, sino complementarias: cada una activa un perfil cognitivo distinto dentro de la taxonomía de Bloom y responde a necesidades instruccionales específicas según la disciplina y la modalidad de enseñanza. Esta diferenciación contrasta con los enfoques documentados en la literatura sobre uso de IA en educación superior, los cuales tienden a proponer soluciones genéricas aplicables indistintamente a cualquier contexto, sin considerar que la eficacia pedagógica de una herramienta multimodal depende críticamente del diseño de la tarea con la que se articula. La estrategia de tutoría contextual en tiempo real mediante Gemini Live constituye un caso especialmente relevante en este punto: representa la extensión más significativa del paradigma multimodal hacia el aprendizaje situado, al eliminar la mediación de representaciones digitales y operar directamente sobre el entorno físico del estudiante.

Importante

Resulta primordial que las estrategias mencionadas sean coordinadas en relación con los lineamientos éticos propuestos en el presente estudio. La articulación de tres dimensiones éticas operativas no constituye un apéndice normativo al margen del enfoque pedagógico, sino una condición de posibilidad del mismo: ninguna estrategia instruccional adquiere pleno sentido pedagógico si no está acompañada de mecanismos que garanticen la autoría intelectual del estudiante, prevengan la dependencia cognitiva y protejan su privacidad informacional. Este principio de integralidad representa una contribución que diferencia al presente estudio de los enfoques predominantes en la literatura, los cuales abordan las dimensiones éticas y las dimensiones instruccionales de la IA en educación como objetos de investigación separados, raramente articulados en un modelo unificado de implementación.

Desde la perspectiva de las limitaciones, el estudio presenta tres restricciones inherentes a su diseño metodológico que deben ser reconocidas con transparencia. La primera es que el marco propuesto es de naturaleza teórica: las categorías construidas a través del análisis temático de la documentación ofrecen coherencia conceptual y fundamentación empírica indirecta, pero no han sido sometidas a validación experimental en contextos universitarios reales. La segunda limitación refiere a la velocidad de cambio del objeto de estudio: Gemini es un modelo en desarrollo activo, y algunas de las funcionalidades descritas —especialmente Gemini Live con cámara en tiempo real— pueden evolucionar de forma significativa entre la publicación de este artículo y su aplicación práctica, lo que exige que los lectores traten este

enfoque como una arquitectura adaptable y no como un protocolo fijo. La tercera limitación concierne a la restricción geográfica y lingüística del corpus analizado: dado que las fuentes consultadas son predominantemente en inglés, el marco resultante puede no capturar investigaciones relevantes publicadas en español, portugués u otros idiomas, lo que limita su representatividad en contextos educativos latinoamericanos y restringe la incorporación de perspectivas pedagógicas propias de la región.

Estas limitaciones condicionan el alcance del estudio y deben considerarse al interpretar la propuesta presentada. En un campo donde la producción empírica sobre modelos multimodales nativos es todavía incipiente, la construcción de propuestas conceptuales rigurosas que orienten la investigación futura y provean a los docentes de categorías interpretativas constituye un aporte metodológico de primer orden. El presente estudio establece una taxonomía inicial, un conjunto de estrategias instruccionales diferenciadas y un sistema de lineamientos éticos integrados que pueden servir como punto de partida para investigaciones de mayor alcance empírico, en particular aquellas que busquen medir el impacto diferencial de las modalidades de interacción sobre la profundidad del procesamiento cognitivo, la retención a largo plazo y el desarrollo de competencias académicas de nivel superior en contextos universitarios diversos.

6. Conclusiones

Los hallazgos obtenidos permiten afirmar que las capacidades multimodales nativas del modelo Gemini no constituyen simplemente una mejora incremental sobre los sistemas de lenguaje previos, sino una reconfiguración cualitativa de las posibilidades pedagógicas disponibles para el diseño instruccional universitario. Esta distinción es el punto de partida conceptual del que se derivan las demás contribuciones del estudio.

La taxonomía de aplicaciones didácticas evidencia que la ventaja pedagógica no es genérica ni uniforme: emerge de manera diferenciada según la modalidad de interacción activada y el diseño de la tarea con la que se articula. Las seis modalidades identificadas configuran un espectro de posibilidades cualitativamente distintas, cada una con un perfil cognitivo propio definido por el nivel taxonómico que activa, el tipo de retroalimentación formativa que habilita y las disciplinas con mayor potencial de aplicación. De entre ellas, la interacción en tiempo real mediante cámara representa la expansión de mayor potencial transformador del paradigma multimodal hacia el aprendizaje situado, aunque su validación empírica

sistemática permanece como agenda pendiente para la investigación futura.

El marco de estrategias instruccionales desarrollado convierte ese potencial en secuencias metodológicas directamente aplicables. Su contribución principal no reside únicamente en describir qué puede hacer Gemini, sino en orientar cómo debe diseñarse la interacción para que la herramienta funcione como amplificador del pensamiento del estudiante y no como sustituto de él. En este sentido, el docente deja de ser un usuario ocasional de la herramienta para convertirse en arquitecto de experiencias de aprendizaje mediadas por inteligencia artificial, con criterios explícitos para seleccionar, implementar y evaluar cada tipo de interacción multimodal en función de los objetivos de aprendizaje declarados.

Los lineamientos éticos articulados en torno a la verificación de la autoría intelectual, la prevención de la dependencia cognitiva y la protección de la privacidad de datos completan el marco al establecer que ninguna estrategia instruccional adquiere pleno sentido pedagógico si no está acompañada de condiciones que garanticen la agencia intelectual del estudiante y la responsabilidad institucional en el uso de plataformas comerciales externas. La multimodalidad no es una característica técnica adicional de Gemini: es la condición que lo convierte en una herramienta pedagógicamente diferente. Comprender esa diferencia con rigor conceptual, traducirla en estrategias instruccionales fundamentadas y enmarcarla en principios éticos operativos es el aporte central de esta investigación al campo de la tecnología educativa.

Finalmente, dotar a docentes, diseñadores instruccionales y tomadores de decisiones institucionales de un lenguaje compartido para evaluar y planificar la integración de modelos fundacionales multimodales en los entornos universitarios del siglo XXI constituye, en última instancia, la contribución que este estudio aspira a consolidar. Para ello, se abren al menos tres líneas prioritarias de investigación futura: en primer lugar, estudios experimentales que validen empíricamente la eficacia diferencial de cada modalidad de interacción sobre el procesamiento cognitivo y la retención de conocimiento en disciplinas específicas; en segundo lugar, investigaciones que examinen el impacto de estas estrategias en contextos educativos latinoamericanos, donde las condiciones de acceso tecnológico, formación docente y cultura institucional pueden modular significativamente los resultados; y en tercer lugar, estudios longitudinales que evalúen el efecto a largo plazo de la integración regulada de MLLMs sobre el desarrollo de la autonomía cognitiva y la integridad académica del estudiantado universitario.

Referencias Bibliográficas

- AlBlooshi, S. (2026). Artificial intelligence in higher education, opportunities, and challenges: A review. *Frontiers in Education, 10*. <https://doi.org/10.3389/feduc.2025.1683968>
- Aldayel, H. S., Pravitasari, H., & Fadillah, N. (2026). Factors affecting academic writing difficulties among English as a Foreign Language (EFL) university students. *Acta Psychologica, 263*, 106250. <https://doi.org/10.1016/j.actpsy.2026.106250>
- Allazo, E. A. V. (2024). Aplicación de la inteligencia artificial en la enseñanza multimodal: Revisión sistemática de la literatura. *Revista Científica Disciplinares, 3*(4), 21-21. <https://doi.org/10.71727/disciplinares.v3i4.292>
- Alsharefeen, R., & Al Sayari, N. (2025). Examining academic integrity policy and practice in the era of AI: A case study of faculty perspectives. *Frontiers in Education, 10*. <https://doi.org/10.3389/feduc.2025.1621743>
- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives : complete edition*. Addison Wesley Longman, Inc. <https://eduq.info/xmlui/handle/11515/18824>
- Bewersdorff, A., Hartmann, C., Hornberger, M., Seßler, K., Bannert, M., Kasneci, E., Kasneci, G., Zhai, X., & Nerdel, C. (2025). Taking the next step with generative artificial intelligence: The transformative role of multimodal large language models in science education. *Learning and Individual Differences, 118*, 102601. <https://doi.org/10.1016/j.lindif.2024.102601>
- Bidry, M., Hanine, M., Ouaguid, A., & Obidallah, W. J. (2025). Transforming Education With Generative AI: A Comprehensive Review of Advancements, Challenges, and Future Opportunities. *IEEE Access, 13*, 202938-202955. <https://doi.org/10.1109/ACCESS.2025.3636891>
- Chu, Z., Wang, S., Xie, J., Zhu, T., Yan, Y., Ye, J., Zhong, A., Hu, X., Liang, J., Yu, P. S., & Wen, Q. (2026). *LLM Agents for Education: Advances and Applications* (arXiv:2503.11733). arXiv. <https://doi.org/10.48550/arXiv.2503.11733>
- Cordero, J., Torres-Zambrano, J., & Cordero-Castillo, A. (2024). Integration of Generative Artificial

- Intelligence in Higher Education: Best Practices. *Education Sciences*, 15(1).
<https://doi.org/10.3390/educsci15010032>
- Creswell, J. W., & Poth, C. N. (2016). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. SAGE Publications.
- Czerkawski, B. (2025). AI and the Learning Experience Design: From Divergent Creativity to Convergent Precision. *TechTrends*, 69(2), 260-270.
<https://doi.org/10.1007/s11528-024-01032-2>
- Ding, D., & Muhyiddin B Yusof, A. (2025). Investigating the role of AI-powered conversation bots in enhancing L2 speaking skills and reducing speaking anxiety: A mixed methods study. *Humanities and Social Sciences Communications*, 12(1), 1223.
<https://doi.org/10.1057/s41599-025-05550-z>
- Eigner, E., & Händler, T. (2024). *Determinants of LLM-assisted Decision-Making* (arXiv:2402.17385). arXiv. <https://doi.org/10.48550/arXiv.2402.17385>
- Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System*, 121, 103254. <https://doi.org/10.1016/j.system.2024.103254>
- Gonsalves, C. (2025). Addressing student non-compliance in AI use declarations: Implications for academic integrity and assessment in higher education. *Assessment & Evaluation in Higher Education*, 50(4), 592-606. <https://doi.org/10.1080/02602938.2024.2415654>
- Google Deepmind, Anil, R., Borgeaud, S., Alayrac, J.-B., Yu, J., Soricut, R., Schalkwyk, J., Dai, A. M., Hauth, A., Millican, K., Silver, D., Johnson, M., Antonoglou, I., Schrittwieser, J., Glaese, A., Chen, J., Pitler, E., Lillicrap, T., Lazaridou, A., ... Vinyals, O. (2023). *Gemini: A Family of Highly Capable Multimodal Models* (arXiv:2312.11805). arXiv. <https://doi.org/10.48550/arXiv.2312.11805>
- Hadra, M., Cambridge, K., & Mesbah, M. (2026). Evaluating the accuracy and reliability of AI content detectors in academic contexts. *International Journal for Educational Integrity*, 22(1), 4. <https://doi.org/10.1007/s40979-026-00213-1>
- Imran, M., & Almusharraf, N. (2024). Google Gemini as a next generation AI educational tool: A

- review of emerging educational technology. *Smart Learning Environments*, 11(1), 22.
<https://doi.org/10.1186/s40561-024-00310-z>
- Lin, Z., Guan, S., Zhang, W., Zhang, H., Li, Y., & Zhang, H. (2024). Towards trustworthy LLMs: A review on debiasing and dehallucinating in large language models. *Artificial Intelligence Review*, 57(9), 243. <https://doi.org/10.1007/s10462-024-10896-y>
- Liu, Z., Cheon, S., Stanbury, A., Jiao, X., Xing, W., & Kang, H. (2025). Towards contextual-based AI: A scoping review of artificial intelligence in X reality for personalized learning. *Computers and Education: Artificial Intelligence*, 10, 100523.
<https://doi.org/10.1016/j.caeai.2025.100523>
- Mayer, R. E. (Ed.). (2001). A Cognitive Theory of Multimedia Learning. En *Multimedia Learning* (pp. 41-62). Cambridge University Press. <https://doi.org/10.1017/CBO9781139164603.004>
- Merrill, M. D. (2002). First principles of instruction. *Educational Technology Research and Development*, 50(3), 43-59. <https://doi.org/10.1007/BF02505024>
- Ofosu-Asare, Y. (2025). Guiding the future: Developing an ethical framework for generative AI use in education. *International Journal of Information and Learning Technology*, 43(1), 26-44.
<https://doi.org/10.1108/IJILT-06-2024-0113>
- Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence*, 2, 100020. <https://doi.org/10.1016/j.caeai.2021.100020>
- Peláez-Sánchez, I. C., Velarde-Camaqui, D., & Glasserman-Morales, L. D. (2024). The impact of large language models on higher education: Exploring the connection between AI and Education 4.0. *Frontiers in Education*, 9. <https://doi.org/10.3389/feduc.2024.1392091>
- Shi, Y., Yu, K., Dong, Y., & Chen, F. (2026). Large language models in education: A systematic review of empirical applications, benefits, and challenges. *Computers and Education: Artificial Intelligence*, 10, 100529. <https://doi.org/10.1016/j.caeai.2025.100529>
- Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333-339. <https://doi.org/10.1016/j.jbusres.2019.07.039>
- Sozon, M., Parnter, C., Wei Lun, W., & Chowdhury, M. A. (2025). Generative AI in higher education: Navigating benefits and challenges in the technological era. *Journal of Applied*

Research in Higher Education, 1-16. <https://doi.org/10.1108/JARHE-02-2025-0103>

- Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive Architecture and Instructional Design: 20 Years Later. *Educational Psychology Review*, 31(2), 261-292. <https://doi.org/10.1007/s10648-019-09465-5>
- Varga-Atkins, T., Saunders, S., Beckingham, S., Keshishi, N., Lacković, N., Li, N., Jiatong, X., Lindsay, R., Wen, R., Winder, I., Ndiaye, H., Jaway, A., Li, S., Jiaotong, X., & Yuan, Y. (2025). *An Educators' Guide to Multimodal Learning and Generative AI*. <https://doi.org/10.17638/03194252>
- Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. F. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education*, 21(1), 40. <https://doi.org/10.1186/s41239-024-00468-z>
- Xu, Z., Jain, S., & Kankanhalli, M. (2025). *Hallucination is Inevitable: An Innate Limitation of Large Language Models* (arXiv:2401.11817). arXiv. <https://doi.org/10.48550/arXiv.2401.11817>
- Yin, S., Fu, C., Zhao, S., Li, K., Sun, X., Xu, T., & Chen, E. (2024). *A survey on multimodal large language models*. <https://dx.doi.org/10.1093/nsr/nwae403>
- Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, 11(1), 28. <https://doi.org/10.1186/s40561-024-00316-7>

CARTA DE ACEPTACIÓN

Por medio de la presente, se hace constar que los autores:

Fernando David Valle Medina, Gustavo Israel Valle Medina y Lisbeth Narcisa Dávila Santillán

presentaron ante el Comité Editorial de la "Revista de Investigación Científica y Social - Reincisol", ISSN 2953-6421, editada en Bahía de Caráquez, Ecuador, el artículo titulado:

Gemini como herramienta de aprendizaje multimodal: aplicaciones y estrategias para la educación superior

Luego de un riguroso proceso de revisión por pares, fue aceptado para su publicación en la revista de frecuencia continua.

El cual podrá visualizarse en el siguiente enlace:

<https://www.reincisol.com/ojs/index.php/reincisol/index>

Recibido: 16/02/2026

Aceptado: 08/03/2026

Bahía de Caráquez, Ecuador, 01/04/2026

Código de verificación: DOC-20260401160521-23961F57DE14

Para verificar la autenticidad de este documento, visite:

<https://reincisol.com/verificar.php?code=DOC-20260401160521-23961F57DE14>


Prof. Gustavo Santacruz Sanabria, PhD.
EDITOR



UNEMI

UNIVERSIDAD ESTATAL DE MILAGRO

¡Evolución académica!

@UNEMIEcuador

