

Uso de RAG en el desarrollo de guiones instruccionales de B@UNAM

Adrián Sarmiento, Omar Terrazas

Using RAG in the development of B@UNAM instructional scripts

Resumen

Este proyecto describe la implementación de inteligencia artificial generativa (IAGen) como apoyo al desarrollo de contenidos educativos para el Bachillerato a Distancia de la UNAM (B@UNAM). El proyecto se centra en complementar el proceso de creación de guiones instruccionales mediante la aplicación de sistemas RAG (*Retrieval-Augmented Generation*) que integran más de 1,450 aprendizajes esenciales y fundamentales establecidos por la UNAM como indispensables para este nivel educativo. Se desarrolló un sistema basado en el modelo masivo de lenguaje Llama 3.2 que utiliza una base de datos especializada de documentos académicos indexados mediante un sistema de claves únicas por área de conocimiento. Esta arquitectura permite la generación de contenidos de alta calidad, al tiempo de preservar la trazabilidad y verificación de las fuentes utilizadas.

Palabras clave: guiones instruccionales; RAG; inteligencia artificial generativa; educación a distancia; generación de contenido

Abstract

This project describes the implementation of Generative Artificial Intelligence (GenAI) as support for the development of educational content for UNAM's Distance Learning High School (B@UNAM). The project focuses on complementing the process of creating instructional scripts through the application of RAG (*Retrieval-Augmented Generation*) systems that integrate more than 1,450 essential and fundamental learning objectives established by UNAM for this educational level. A system was developed based on the large language model Llama 3.2, which utilizes a specialized database of academic documents indexed through a unique key system by knowledge area. This architecture enables the generation of high-quality content while maintaining traceability and verification of the sources used.

Keywords: instructional scripts; RAG; generative artificial intelligence; distance education; content generation

Introducción

El Bachillerato a Distancia de la UNAM (B@UNAM) fue creado en 2007 y hoy forma parte del Sistema Universidad Abierta y Educación a Distancia (SUAYED) de la Universidad Nacional Autónoma de México (Amador, 2023). Este programa surgió con el propósito fundamental de democratizar el acceso a la educación media superior, ofreciendo una alternativa educativa de calidad que permitiera superar las barreras geográficas, económicas y sociales que tradicionalmente limitaban las oportunidades de formación académica (Vadillo et al., 2022).

La función social del B@UNAM se ha consolidado con el paso de los años, durante los cuales ha desarrollado una red de vinculación social mediante convenios de colaboración con universidades e instituciones de educación superior, sectores del gobierno federal y locales, asociaciones y fundaciones de la sociedad civil, y organismos extranjeros. Esta estrategia ha fortalecido tanto el desarrollo institucional como la vinculación nacional e internacional del programa (Amador, 2023).

El programa representa una innovación significativa que ha permitido ampliar la cobertura de la educación media superior, atendiendo, particularmente, a poblaciones que por diversas circunstancias no pueden acceder a la modalidad presencial, teniendo un alcance internacional que ha beneficiado a estudiantes de los cinco continentes (Dirección General de Comunicación Social, 2016).

Diseño de los programas de estudio de B@UNAM

Los programas de estudio de B@UNAM se han caracterizado por su estructura curricular integrada e interdisciplinaria, siendo este uno de los elementos más innovadores del modelo educativo (Vadillo & Villatoro, 2009). Su diseño curricular comprende 30 asignaturas organizadas en cuatro semestres, las cuales han sido desarrolladas con base en estándares internacionales de calidad, su evaluación mediante la rúbrica OSCQR (*OpenSUNY Course Quality Review*) del *Online Learning Consortium* ha demostrado que 30 de 31 cursos evaluados se ubican en un rango de cumplimiento entre 90 y 99%, confirmando la excelencia académica del programa (Bañuelos, 2019).

Los guiones instruccionales de las asignaturas de estos programas se han desarrollado siguiendo una metodología que integra múltiples componentes pedagógicos y tecnológicos. Su diseño ha considerado tanto aspectos didácticos como elementos de innovación tecnológica, creando experiencias de aprendizaje que aprovechan el entorno digital para facilitar la construcción del conocimiento (Naranjo, 2014).

Actualización del plan de estudios

La actualización del plan de estudios se ha vuelto imperativa para adaptarnos a las transformaciones en la educación. Hoy en día, los sistemas educativos buscan atender las necesidades educativas personalizadas, priorizando el desarrollo de habilidades como la cooperación, creatividad, liderazgo y resolución de problemas (Mukul & Büyüközkan, 2023). En este nuevo paradigma, es fundamental inculcar una mentalidad de aprendizaje a lo largo de toda la vida.

Uno de los primeros pasos en este proceso ha sido el desarrollo del “Mapa de pensamientos”, un componente curricular innovador que nos sirve como guía para crear actividades de aprendizaje dentro de nuestros materiales en línea. Este mapa nos permite diseñar entornos virtuales que facilitan la construcción de contenidos significativos, alineados con el desarrollo de cinco tipos de pensamiento que nuestros estudiantes necesitan para enfrentar los complejos desafíos del mundo actual (Pineda, 2019).

Esta necesidad de un nuevo enfoque nos ha llevado a adoptar la transdisciplinariedad en la educación. Este modelo de aprendizaje se basa en interacciones colaborativas que trascienden las fronteras de las disciplinas tradicionales, como las artes, las ciencias naturales y sociales (O’Sullivan, 2025). El objetivo es claro: integrar a actores externos al ámbito académico para contribuir a la solución de problemas reales, como el cambio climático o la inequidad social. La transdisciplinariedad es esencial porque los desafíos de hoy no son asuntos técnicos aislados, sino que involucran múltiples factores interconectados.

Finalmente, para que esta transformación sea efectiva, debemos aprovechar los avances tecnológicos que ya han redefinido la enseñanza y el aprendizaje. Tecnologías como la inteligencia artificial, la realidad aumentada y la computación en la nube potencian el proceso educativo al facilitar la recolección e interpretación de datos, además de proporcionar un acceso flexible a recursos (Joshi et al., 2024). Estos avances, combinados con la transdisciplinariedad, nos permiten construir un modelo educativo más completo y adaptado a la complejidad de la realidad contemporánea.

Inteligencia artificial generativa (IAGen) para la creación de contenidos

Con todo esto en mente, el proceso de desarrollo de contenidos para B@UNAM se ha revolucionado con la incorporación de la inteligencia artificial. En distintas fases de su elaboración, se utiliza un modelo masivo de lenguaje (LLM, por sus siglas en inglés) como Llama 3.2 (Meta, 2024) para integrar de forma eficiente los más de 1,450 aprendizajes esenciales (Consejo Académico del Bachillerato, 2023) y fundamentales (Consejo Académico del Bachillerato, 2008) que la UNAM ha definido como pilares de la formación del bachiller. Este enfoque nos permite no solo generar textos y materiales de estudio de alta calidad, sino también diseñar ejercicios interactivos y actividades colaborativas que refuerzan los nuevos pilares de la educación: la cooperación, el pensamiento crítico y la resolución de problemas en un contexto de aprendizaje a lo largo de toda la vida.

Actualmente las herramientas de IAGen que generan textos se encuentran en un nivel de madurez suficiente para ser utilizadas en la creación de contenidos digitales. Esto permite a los usuarios simplificar la solución de algunas tareas, las cuales pueden comenzar desde la creación de algún texto con un itinerario hasta contextos altamente especializados, como es el escribir propuestas de solución a problemas matemáticos fundamentales (Xu et al., 2025). También se ha documentado su éxito en la resolución de evaluaciones escritas de nivel superior (Şahin Kizil, 2024), lo que nos permite plantearnos su aplicación para crear materiales consistentes. En este sentido, existen técnicas para recuperar documentos académicos revisados por un equipo de expertos y supervisar la calidad del contenido creado, haciendo posible aplicar la capacidad generativa de los LLM para crear contenidos de manera transdisciplinaria, con una dificultad adaptada a varios niveles de exigencia y manteniendo una narrativa que mantenga a los participantes motivados al estudio.

El proyecto de desarrollo de guiones instruccionales de B@UNAM incorpora IAGen, y en particular, incluye una técnica denominada RAG (del inglés *Retrieval-Augmented Generation*) (Bhardwaj et al., 2025). Esta técnica permite la creación de contenidos utilizando una base de datos especializada y que puede ser de acceso local, para suministrar información relevante al LLM y evitar alucinaciones (Gokcimen & Das, 2025). Para realizar la creación de algún material mediante un sistema RAG, se escribe una petición al sistema (un *prompt*) que incluye los temas a desarrollar y los criterios para que el material sea correcto y relevante. Posteriormente, el sistema RAG consulta una base de datos de documentos seleccionados como referencias para cada aprendizaje y conocimiento esencial, se localiza la sección del documento más relevante y se devuelven los resultados más significativos a nivel semántico (Elwin et al., 2025). Finalmente estos resultados se incluyen en una segunda petición que será la encargada de crear el contenido, esto permite condicionar la respuesta del LLM incorporado en el RAG y reduce la probable aparición de contenido que no tenga afirmaciones basadas en referencias académicas (Lewis et al., 2020).

Selección de contenidos relevantes

Para que un sistema basado en RAG funcione de manera eficiente e interpretable (Arora y Ramteke, 2024), es necesario indexar cada texto que se usa como referencia documental, con el propósito de poder identificar la referencia y su ubicación en un base de datos. Esta información incluye el tipo de medio que lo contiene (archivos de texto, audios o imágenes), indicar si se trata de un documento de acceso abierto y redistribuible, los autores del material y el año de publicación, entre otros datos.

Para esto se realizó la búsqueda y clasificación de cada documento que sustenta académicamente los aprendizajes esenciales (Consejo Académico del Bachillerato, 2023) y fundamentales (Consejo Académico del Bachillerato, 2008), que son requisitos en la creación de los nuevos programas de estudios del Bachillerato a Distancia. La tabla 1 muestra un ejemplo del proceso de clasificación para los documentos. Este proceso conforma un diccionario de claves para poder indexar de manera unívoca cada documento con el aprendizaje o conocimiento esencial que se recopila. La clave está conformada por un prefijo según el área de conocimiento y un identificador único para enumerar cada recurso bibliográfico.

Disciplina	Eje temático	Aprendizaje o conocimiento	Clave	Asignaturas asociadas
Física	Electromagnetismo	4.3 Definir el potencial eléctrico y la diferencia de potencial.	FIS04030000A	105 Matemáticas de campos y luz
Física	Electromagnetismo	4.4 Explicar el experimento de Oersted.	FIS04040000A	105 Matemáticas de campos y luz
Física	Electromagnetismo	4.5 Resolver	FIS04050000A	105 Matemáticas de campos y

		problemas usando la ley de Ohm.		luz
Física	Electromagnetismo	4.6 Comparar circuitos resistivos en serie y en paralelo.	FIS04060000A	105 Matemáticas de campos y luz

Tabla 1. Ejemplo de diccionario de claves y conocimientos esenciales.

Con el proceso anterior es posible recuperar el origen de cada contenido y verificar su congruencia utilizando supervisión de expertos en cada campo disciplinario y técnicas de Procesamiento de Lenguaje Natural (NLP, por sus siglas en inglés) (Huynh & McNamara, 2025).

Arquitectura general de un RAG para creación de guiones

Una vez completado el proceso de recopilación documental, se construye una base de datos que puede ser consultada por el LLM. A este proceso se le conoce como la ingesta, representación y almacenamiento de datos. Esta base de datos puede ser consultada para buscar semánticamente temas relacionados con el guión instruccional que se desea crear para después enviar un comando de creación con toda la información recuperada. Este paso se conoce como RAG, recuperación y generación aumentada. La figura 1 muestra una arquitectura general que describe los componentes del sistema.

Sistema de generación aumentada por recuperación

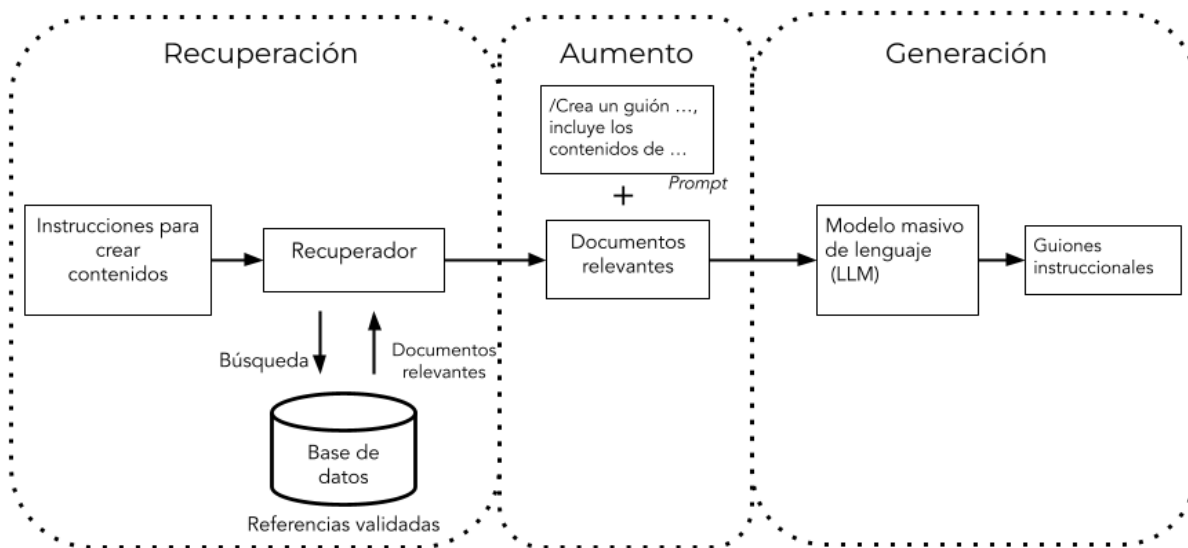


Figura 1. Arquitectura general del sistema RAG para generación de guiones instruccionales.

Con el sistema mostrado en la figura 1 ha sido posible crear guiones instruccionales con una estructura general. Después estos guiones son validados para que contengan los temas solicitados y una narrativa acorde al programa de estudios al que pertenece, sobre todo por el enfoque interdisciplinar de los mismos. Esto es posible a través del lenguaje natural [ref lenguaje natural], se da una instrucción en forma de texto al LLM para crear los contenidos. Esta instrucción incluye describir el rol del LLM como escritor de contenido, el lenguaje apropiado para el nivel bachillerato, todos los temas a integrar, una narrativa novedosa y que mantenga interesado al lector, integrar ecuaciones y ejemplos prácticos cuando sea necesario, crear casos de estudio que se pueden aplicar a la vida cotidiana, y siempre mantener el cumplimiento de los criterios académicos obligatorios. Debido a que el LLM es una herramienta interactiva, en caso de no cumplir los criterios anteriores, es posible iterar las peticiones de creación mejorando el contenido hasta ser aprobado. El proceso siempre termina con la revisión de expertos humanos en los ámbitos disciplinares correspondientes a cada asignatura. Solo tras incorporar sus recomendaciones de mejora y su aprobación final se procede a publicar la asignatura en línea.

Resultados preliminares

Una prueba piloto inicial ya ha sido completada para evaluar este modelo. Se desarrolló una unidad de contenido de uno de los nuevos programas de estudio con el apoyo de la IAGen y otra versión creada por expertos humanos. Las observaciones de los docentes del área de conocimiento están siendo evaluadas a fondo para contrastar la calidad, pertinencia y claridad de cada uno de los aprendizajes en ambas versiones.

Además, esta prueba nos permitió analizar y comparar los tiempos de producción. El desarrollo de una unidad de contenido por expertos humanos requirió aproximadamente 1,640 horas, y con la incorporación de las herramientas de IAGen, se redujo a solo 118 horas. Este proceso, además de ser significativamente más rápido, ha demostrado ser más preciso, al haber menos imprecisiones contenidas en los materiales cuya creación con IAGen pasó de 21 a 15.

Conclusiones

La integración de la inteligencia artificial en la creación de guiones instruccionales representa una transformación significativa en la producción de contenidos educativos. Herramientas como los sistemas basados en RAG permiten manejar grandes volúmenes de información, construir narrativas coherentes y mantener el interés del estudiante. Estos sistemas no solo agilizan el trabajo, sino que también ofrecen un control sobre la privacidad de los datos, lo que es esencial para el desarrollo de contenidos especializados.

La implementación de un sistema de indexación con claves únicas para cada documento de referencia facilita la trazabilidad y la verificación del contenido generado. Esto asegura la validación por parte de expertos y garantiza la calidad y la coherencia académica en la producción masiva de materiales.

Este enfoque híbrido, que combina la generación automatizada con la validación humana, sugiere que la inteligencia artificial no reemplaza al factor humano en la educación. En cambio, potencia las capacidades de los docentes, permitiéndoles

enfocarse en tareas de mayor valor, como la mentoría y la personalización del aprendizaje. En esencia, este modelo representa un paradigma escalable y replicable que potencia el rigor académico mientras optimiza los procesos de creación de contenido.

Referencias

- Amador Bautista, M. del R.. (2023). Función social del Bachillerato a Distancia de la UNAM (B@UNAM, 2007-2022). *Revista Mexicana de Bachillerato a Distancia*, 15(29). <https://doi.org/10.22201/cuaieed.20074751e.2023.29.84984>
- Arora, S., y Ramteke, V. (2024). Revolutionizing Third Party Risk Management Using Generative AI and RAG. *2024 First International Conference on Data, Computation and Communication (ICDCC)*, 16–22. <https://doi.org/10.1109/ICDCC62744.2024.10961910>
- Bañuelos, A. M. (2019). La evaluación de cursos en línea mediante rúbrica. El caso de las asignaturas de B@UNAM. *Revista Mexicana de Bachillerato a Distancia*, 11(22). <https://doi.org/10.22201/cuaed.20074751e.2019.22.70578>
- Bhardwaj, S., Joshi, R. C., Tiwari, V., Riha, K., Sikora, P., y Dutta, M. K. (2025). AI-powered retrieval-augmented generation framework with large language models for enhanced public health response. *2025 3rd International Conference on Advancement in Computation & Computer Technologies (InCACCT)*, 616–621. <https://doi.org/10.1109/InCACCT65424.2025.11011436>
- Consejo Académico del Bachillerato. (2008). Conocimientos fundamentales para la enseñanza media superior. https://www.cab.unam.mx//nucleo_con/con_fun_2008/completa.pdf
- Consejo Académico del Bachillerato. (2023). Aprendizajes esenciales 2023. https://www.cab.unam.mx//interiores/aprendizajes_esenciales.html
- Dirección General de Comunicación Social. (2016, octubre 2). Tiene la UNAM 400 alumnos de Bachillerato a Distancia en 34 países de los cinco continentes. Boletín UNAM-DGCS-722. <https://www.dgcs.unam.mx/index.html>
- Gokcimen, T., y Das, B. (2025). A novel system for strengthening security in large language models against hallucination and injection attacks with effective strategies. *Alexandria Engineering Journal*, 123, 71–90. <https://doi.org/10.1016/j.aej.2025.03.030>
- Huynh, L. y McNamara, D. S. (2025). GenAI-Powered Text Personalization: Natural Language Processing Validation of Adaptation Capabilities. *Applied Sciences*, 15(12), 6791. <https://doi.org/10.3390/app15126791>
- Joshi, S., Srivastava, A. P., Prabhu, S., Pathak, P., Chirputkar, A. y Pillai, S. (2024). Trends in digital transformation (DT) of higher education institutes (HEI): Bibliometric and systematic review of 13 years. *Journal of Applied Research in Higher Education*. <https://doi.org/10.1108/JARHE-02-2024-0084>

- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... y Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Proceedings of the 34th International Conference on Neural Information Processing Systems*. <https://doi.org/10.48550/arXiv.2005.11401>
- Meta. (2024, septiembre 25). Llama 3.2: Revolutionizing edge AI and vision with open, customizable models. Meta AI. Recuperado el 28 de agosto de 2025, de <https://ai.meta.com/blog/llama-3-2-connect-2024-vision-edge-mobile-devices/>
- Mukul, E., y Büyüközkan, G. (2023). Digital transformation in education: A systematic review of education 4.0. *Technological Forecasting and Social Change*, 194, 122664. <https://doi.org/10.1016/j.techfore.2023.122664>
- Naranjo, L. D. (2014). Avance tecnológico en el Bachillerato a Distancia de la UNAM (B@UNAM) en línea. *Revista Mexicana de Bachillerato a Distancia*, 6(12), 6–6. <https://doi.org/10.22201/cuaed.20074751e.2014.12.64869>
- O'Sullivan, G. (2025). U-shaped learning: A new model for transdisciplinary education. *Humanities and Social Sciences Communications*, 12(1), 182. <https://doi.org/10.1057/s41599-025-04478-8>
- Pineda, V. J. (2019). El proceso de construcción del mapa de pensamientos de B@UNAM. *Revista Mexicana de Bachillerato a Distancia*, 11(22). <https://doi.org/10.22201/cuaed.20074751e.2019.22.70585>
- Elwin, G. R., Narayanan, R., Raj, R., & Moumitha. (2025). AI-driven solution for analyzing and interpreting data from PDFs and other textual resources using RAG and Hybrid Retrieval Models. 2025 2nd International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), 1–6. <https://doi.org/10.1109/RMKMATE64874.2025.11042348>
- Şahin Kizil, A. (2024). Large Language Models and automated essay scoring: Implications arising from research. *International Journal of Language Academy*, 12(3), 85–97. <https://doi.org/10.29228/ijla.77036>
- Vadillo, G., Bucio García, J., y Naranjo, D. (2022). 15 años de B@UNAM: Nacimiento y desarrollo de un bachillerato en línea. *EDUTECH REVIEW. International Education Technologies Review / Revista Internacional de Tecnologías Educativas*, 9(1), 77–86. <https://doi.org/10.37467/gkarevedutech.v9.3306>
- Vadillo, G., y Villatoro, C. (2009). B@UNAM: Interdisciplina y actualización en el currículum integrado de la Universidad Nacional Autónoma de México. *Revista Mexicana de Bachillerato a Distancia*, 1. <https://doi.org/10.22201/cuaed.20074751e.2009.Especial.46979>
- Xu, X., Xiao, T., Chao, Z., Huang, Z., Yang, C., y Wang, Y. (2025). Can LLMs solve longer Math word problems better? *ICLR*. (arXiv:2405.14804). arXiv. <https://doi.org/10.48550/arXiv.2405.14804>

Carlos Adrián Sarmiento Gutiérrez

adrian_sarmiento@cuaed.unam.mx

Bachillerato a Distancia UNAM

[009-0005-0551-3205](tel:009-0005-0551-3205)

Omar Terrazas Razo

omar_terrazas@cuaed.unam.mx

Bachillerato a Distancia UNAM

[0000-0002-3424-3751](tel:0000-0002-3424-3751)



En la traducción de los artículos de la Revista Mexicana de Bachillerato a Distancia se utiliza en ocasiones una herramienta de inteligencia artificial como Google Translate, ChatGPT-4o o DeepL, con revisión humana posterior, cuidando siempre el estilo y aportación de los autores.