Junta de Andalucía
Consejería de Cultura,
Patrimonio Histórico y Deporte
Instituto Andaluz del Patrimonio Histórico

REVISTA PH
ISSN 2340-7565

Sección Debate (***revista PH*** 120, febrero 2027)

Debate 36: IA y patrimonio cultural. ¿Qué deberíamos tener en cuenta para el uso social, ético y sostenible de la IA en el ámbito del patrimonio?

Textos provisionales

La inteligencia artificial generativa (IAG) y el patrimonio incómodo: una pareja mal avenida

Tania González-Cantera | Universidad del País Vasco (EHU)

La incorporación de la inteligencia artificial generativa (IAG) al ámbito del patrimonio cultural suscita interrogantes sobre quién produce los relatos del pasado, así como qué intereses, valores y perspectivas condicionan su contenido. Estas cuestiones adquieren especial relevancia en el caso del patrimonio incómodo, cuya interpretación está atravesada por conflictos políticos y sociales a menudo perpetuados en el tiempo.

El problema no reside únicamente en que estos sistemas puedan generar información errónea o inventada, sino también en la reproducción de sesgos que privilegian narrativas dominantes y relegan perspectivas históricamente (González-Cantera y Bonacchi 2025). Mediante la selección de información, el lenguaje y la representación desigual de voces, estos sistemas pueden consolidar determinadas interpretaciones del pasado e influir en cómo la sociedad comprende los acontecimientos históricos, reconoce a sus víctimas y atribuye significado a los lugares de memoria.

Los modelos generativos producen respuestas a partir de patrones aprendidos y, en el caso de los grandes modelos lingüísticos (LLM), predicen secuencias de palabras. Por ello, se han comparado con loros (Bender et ál. 2021) al reproducir el lenguaje sin comprender realmente su significado.

Estos modelos no son repositorios neutrales de conocimiento ni garantizan la veracidad de sus resultados. Entre sus limitaciones destacan las alucinaciones, que generan información incorrecta o referencias inexistentes con apariencia de certeza (Alkaissi y McFarlane 2023), y el *common token bias*, que favorece la reproducción de expresiones frecuentes en los datos de entrenamiento (Asprovska y Hunter 2024). Sin embargo, en el patrimonio incómodo resulta especialmente relevante el sesgo discursivo (Giacalone 2026), que opera mediante omisiones, eufemismos, elecciones terminológicas o una visibilidad desigual de los actores históricos, sin necesidad de recurrir a falsedades explícitas. Así, una respuesta mayoritariamente correcta puede ofrecer una interpretación sesgada, especialmente en contextos donde el lenguaje, la contextualización de la violencia y el reconocimiento de las víctimas forman parte de las disputas por la memoria.

En un estudio publicado en 2025, la Dra. Bonacchi y yo analizamos las respuestas de ChatGPT-4o sobre espacios de memoria asociados con regímenes dictatoriales. El modelo GPT-4o, retirado de ChatGPT por OpenAI en febrero de 2026, tiene actualmente un interés principalmente histórico para el presente debate.

Aunque las versiones más recientes han mejorado la precisión de las respuestas y la capacidad para resolver tareas complejas, esto no significa que los sesgos detectados hayan desaparecido (Afridi et ál. 2026). Esto se debe a que la calidad de las respuestas no depende únicamente de su precisión factual. También importa qué perspectivas incorpora, cuáles excluye y cómo presenta las controversias. Un modelo puede mejorar su capacidad para localizar información y, al mismo tiempo, continuar reproduciendo desequilibrios presentes en las fuentes disponibles.

En el citado artículo analizábamos las respuestas de ChatGPT-4o sobre tres espacios de memoria vinculados a dictaduras de países hispanohablantes –el Valle de Cuelgamuros (España), el Museo y Sitio de Memoria ESMA (Argentina) y el monumento a Salvador Allende (Chile)– mediante el enfoque histórico-discursivo de Ruth Wodak.

En el caso español, las respuestas tendían a alinearse con la narrativa del Gobierno de Pedro Sánchez sobre la resignificación del monumento. Sin embargo, también incorporaban vocabulario y marcos interpretativos franquistas, como los conceptos de “caídos” o “cruzada”, y reproducían la idea del complejo como espacio de “reconciliación”. Asimismo, presentaban ciertas posiciones de la extrema derecha como demandas de neutralidad democrática, calificaban de “radical” la propuesta de Unidas Podemos de exhumar los restos del dictador frente a la postura “cauta” del Partido Popular por preservarlos, omitían actores relevantes (como la oposición de la comunidad benedictina del valle) y reproducían inexactitudes históricas.

Estos resultados muestran que los sesgos no siempre se expresan mediante una defensa ideológica explícita, sino también a través del lenguaje, la selección de acontecimientos y la legitimación desigual de las posiciones políticas. La coexistencia de narrativas institucionales con discursos franquistas y revisionistas evidencia cómo el modelo reproduce las narrativas más presentes en internet, incluidas aquellas que legitiman la simbología franquista o minimizan la represión de la dictadura.

La aparición de alternativas a ChatGPT, como Claude (Anthropic) y Gemini (Google), amplía las opciones disponibles, pero no garantiza una mayor pluralidad interpretativa. Aunque los estudios comparativos entre plataformas identifican diferencias en sus orientaciones políticas, estas se mantienen dentro de marcos ideológicos hegemónicos (Choudhary 2025; Yuksel, Catalbas y Oc 2025). Sin embargo, todavía no disponemos de estudios equivalentes que permitan determinar si estas diferencias se reproducen en la interpretación del patrimonio cultural y los pasados conflictivos.

Comparar modelos resulta, por tanto, necesario para analizar cómo construyen sus interpretaciones, qué perspectivas privilegian y cuáles excluyen. Más que aspirar a una IAG completamente neutral –un horizonte difícilmente alcanzable–, se trata de identificar los factores que determinan sus respuestas y evaluar sus implicaciones para la representación y transmisión de las memorias conflictivas.

La relación entre inteligencia artificial generativa (IAG) y memoria es bidireccional: los modelos reproducen patrones presentes en sus datos de entrenamiento y sus respuestas pueden difundirse por internet, donde pasan a formar parte del entorno informativo que consultan otros usuarios y sistemas. Esta retroalimentación puede reforzar la visibilidad de las narrativas dominantes y perpetuar la marginalización de otras voces.

En el ámbito del patrimonio incómodo, estas dinámicas pueden condicionar la interpretación de la violencia, la atribución de responsabilidades históricas y la representación de las memorias en conflicto. No se trata de otorgar la misma validez a todas las interpretaciones, sino de evitar que la búsqueda de pluralidad derive en una falsa equivalencia entre el conocimiento respaldado por evidencias y los discursos revisionistas. Por ello, el uso de la IAG en el ámbito patrimonial requiere analizar qué relatos reproduce y qué voces deja fuera. Que sus respuestas resulten convincentes no significa que ofrezcan una interpretación rigurosa del pasado: su valoración sigue dependiendo del criterio de quienes investigan y gestionan el patrimonio.

BIBLIOGRAFÍA

- Afridi, H., Ullah, H., Khan, S. D. y Ullah, M. (2026) From GPT-3 to GPT-5: Mapping their capabilities, scope, limitations, and consequences. *arXiv*, arXiv:2604.10332. Disponible en: https://doi.org/10.48550/arXiv.2604.10332 [Consulta: 09/11/2026]
- Alkaissi, H. y McFarlane, S.I. (2023) Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. *Cureus*, vol. 15, n.º 2, e35179. Disponible en: https://doi.org/10.7759/cureus.35179 [Consulta: 09/11/2026]
- Asprovska, M. y Hunter, N. (2024) The Tokenization Problem: Understanding Generative AI's Computational Language Bias. *Ubiquity Proceedings*. Disponible en: https://doi.org/10.5334/uproc.123 [Consulta: 09/11/2026]
- Bender, E.M., Gebru, T., McMillan-Major, A. y Shmitchell, S. (2021) On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. En: *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. New York: Association for Computing Machinery, pp.610-623. DOI: https://doi.org/10.1145/3442188.3445922 [Consulta: 09/11/2026]
- Choudhary, T. (2025) Political Bias in Large Language Models: A Comparative Analysis of ChatGPT-4, Perplexity, Google Gemini, and Claude. *IEEE Access*, vol. 13, pp. 11341-11379. Disponible en: https://doi.org/10.1109/ACCESS.2024.3523764
- Giacalone, M. (2026) Discursive behavior of generative language models in geopolitical and humanitarian contexts. *Discover Artificial Intelligence*, vol. 6, art. 230. DOI: https://doi.org/10.1007/s44163-026-00965-2
- González-Cantera, T. y Bonacchi, C. (2025) ChatGPT's interpretation of contested memoryscapes favours the voice of current governments, capitalism and the far-right. *International Journal of Heritage Studies*, vol. 31, n.º 11, pp. 1508-1530. DOI: https://doi.org/10.1080/13527258.2025.2543736
- Yuksel, D., Catalbas, M.C. y Oc, B. (2025) Language-Dependent Political Bias in AI: A Study of ChatGPT and Gemini. *arXiv*, arXiv:2504.06436. Disponible en: https://doi.org/10.48550/arXiv.2504.06436 [Consulta: 09/11/2026]