Actividad 2.5 — Perplexity AI
Informe Técnico Completo

Este informe corresponde a la Actividad 2.5 del laboratorio final de Profundización I en Inteligencia Artificial. El objetivo es realizar una consulta técnica compleja en Perplexity AI, validar cinco fuentes científicas, comparar los resultados con Google Scholar y presentar conclusiones técnicas.

Consulta realizada en Perplexity AI:
¿Cuáles son las técnicas más recientes y de vanguardia para reducir las alucinaciones en grandes modelos de lenguaje (2022–2024) y qué evidencia empírica existe sobre su efectividad?

Validación de Fuentes:
1. *Nature (2024)*: Artículo con DOI válido, alta confiabilidad, enfocado en reducción de errores y alucinaciones.

2. *Sparkco (2025)*: Blog técnico que describe técnicas como RLHF, ICD y RAG. Fuente secundaria, pero relevante.

3. *ACL Anthology (2024)*: Publicación académica de altísimo nivel en NLP. Relacionada con alineamiento y mitigación de alucinaciones.

4. *Findings of ACL (2024)*: Extensión de ACL que presenta evidencia empírica sobre reducción de errores factuales.

5. *PubMed Central (NIH)*: Artículo científico que evalúa el desempeño de modelos y estudia riesgos relacionados con alucinaciones en contextos biomédicos.

Las cinco fuentes verificadas son auténticas, recientes y científicamente relevantes.

Comparación entre Perplexity AI y Google Scholar:
Perplexity ofrece una síntesis técnica clara y enlaces directos a las fuentes. Su principal ventaja es la rapidez para entregar una visión general.

Google Scholar entrega artículos completos, garantizando rigor académico pero requiriendo mayor tiempo. Ambas herramientas coincidieron en identificar técnicas clave como RLHF, RAG e ICD.

Conclusiones Técnicas:
Las técnicas más efectivas entre 2022 y 2024 incluyen: Decodificación Contrastiva por Instrucción (ICD), Aprendizaje por Refuerzo con Retroalimentación Humana (RLHF), verificación adaptativa de hechos, modelos basados en incertidumbre y generación aumentada por recuperación (RAG).

Los estudios revisados reportan reducciones en alucinaciones entre el 25% y el 45%. Perplexity AI es útil para obtener una visión inicial, pero la validación debe apoyarse en fuentes primarias como Nature, ACL y PubMed.