

Sovereign RAG Engine for Inclusive Education: A Co-Creation Case between FAMAF-UNC and Prócer Tecnologías

Abstract. The gap in access to educational materials for students with visual or learning disabilities represents a critical challenge in Argentina and Latin America. This work describes the collaboration methodology between the Faculty of Mathematics, Astronomy, Physics, and Computing (FAMAF-UNC) and the company Prócer Tecnologías for the development of an inclusion-oriented Retrieval-Augmented Generation (RAG) engine. Unlike a purely academic approach, the project focused on the technology transfer of open-source large language models (LLMs) into a pre-existing production environment created by Prócer Tecnologías. It details the collaboration process through the Secretariat of Innovation and Technology Transfer, the industrial requirements for data sovereignty, and the integration of multimodal capabilities for the interpretation of classroom contexts.

Keywords: University–Industry Collaboration · Artificial Intelligence · Inclusive Education · Technology Transfer · Data Sovereignty.

Motor RAG Soberano para Educación Inclusiva: Un Caso de Co-creación entre FAMAF-UNC y Prócer Tecnologías

Autores omitidos para revisión ciega

Institución omitida para revisión ciega

Resumen La brecha en el acceso a materiales educativos para estudiantes con discapacidades visuales o de aprendizaje representa un desafío crítico en Argentina y Latinoamérica [22,3]. Este trabajo describe la metodología de colaboración entre la Facultad de Matemática, Astronomía, Física y Computación (FAMAF-UNC) y la empresa Prócer Tecnologías para el desarrollo de un motor de Generación Aumentada por Recuperación (RAG) orientado a la inclusión. A diferencia de un enfoque puramente académico, el proyecto se centró en la transferencia tecnológica de modelos de lenguaje de código abierto (LLM) hacia un entorno productivo preexistente, creado por la empresa Prócer Tecnologías. Se detalla el proceso de vinculación a través de la Secretaría de Innovación y Vinculación Tecnológica, los requerimientos industriales de soberanía de datos y la integración de capacidades multimodales para la interpretación de contextos áulicos.

Keywords: Vinculación Universidad–Industria · Inteligencia Artificial · Educación Inclusiva · Transferencia Tecnológica · Soberanía de Datos.

1. Introducción

1.1. Institución y Empresa

FAMAF–UNC dependiente del rectorado de la Universidad Nacional de Córdoba, se caracteriza por su fuerte impronta en investigación básica, tiene una planta de aproximadamente 300 docentes, 2800 estudiantes de grado y 300 estudiantes de posgrado. Actualmente cuenta con una Secretaría de Innovación y Vinculación Tecnológica (SIVT), con la misión de establecer el vínculo con el medio productivo local y regional para aportar a la resolución de problemas concretos de base científico–tecnológica. A través de la SIVT se canalizan proyectos de I+D+i, consultorías técnicas, capacitaciones *ad hoc* para la industria e iniciativas de emprendedurismo para estudiantes e investigadora/es [2,6].

Prócer Tecnologías SAS es una empresa de base tecnológica enfocada en el desarrollo de soluciones basadas en inteligencia artificial aplicada a la asistencia en la educación, con operaciones en Argentina y otros países de América.

1.2. Historial y naturaleza de la vinculación

A partir de reuniones llevadas a cabo durante el año 2025 entre la empresa e investigadora/es de la FAMAF, se planteó la necesidad de trabajar conjuntamente en problemáticas de I+D y transferencia de saberes.

Desde agosto de 2025, se definió abordar el problema de la recuperación de información en entornos escolares, formalizando un proyecto de I+D de cuatro meses de duración para implementar un sistema RAG (Retrieval-Augmented Generation) [16,10]. El plan abarca desde el diseño de embeddings hasta la integración de un MVP funcional en el dispositivo de la empresa.

Para esto, Prócer, con el aval de la FAMAF, presentó el plan en las convocatorias Voucher de Innovación Colaborativa del Gobierno de la Provincia de Córdoba [37], resultando adjudicados en diciembre de 2025. El vínculo se formalizó mediante un servicio de asistencia técnica [34], realizando la reunión de kick-off el 18 de diciembre de 2025.

1.3. Motivación del proyecto

Millones de estudiantes en Argentina y LATAM enfrentan barreras para una educación equitativa. En Argentina, el 10,2 % de la población de 6+ años presenta alguna dificultad según el INDEC [13], incluyendo 70.000 niños con deficiencia visual. La dislexia afecta a cerca del 10 % [36], impactando la lectoescritura. A nivel regional, UNICEF reporta 19,1 millones de niños y adolescentes con discapacidad con potencial limitado [36]. Los jóvenes con discapacidad en LATAM tienen una probabilidad un 23 % menor de completar la educación secundaria (Banco Mundial [20]).

Estos estudiantes, particularmente aquellos con discapacidad visual, dislexia o discapacidad intelectual, enfrentan grandes barreras para acceder a contenidos escolares multiformato (textos, gráficos, tablas, mapas). Las herramientas actuales, como lectores de pantalla o apps, no comprenden bien el contenido visual complejo, no ofrecen trazabilidad en las respuestas que proporcionan, y los modelos de Inteligencia Artificial existentes suelen presentar errores o “alucinaciones” que afectan negativamente la calidad educativa [22,28]. Esta situación no sólo dificulta el aprendizaje individual y autónomo, sino que también perpetúa la brecha de desigualdad educativa.

La propuesta de Prócer es abordar este desafío con un dispositivo de asistencia auditiva [30], que permita el acceso autónomo, auditivo y accesible a la información completa. El sistema ingiere materiales escolares (principalmente textuales) cargados por usuarios y los indexa como texto para poder realizar consultas sobre los mismos. La solución busca transformar la inclusión educativa, permitir evaluaciones más equitativas, avanzar hacia la reducción de dependencia de terceros e integrar IA confiable y eficiente al aula [35,3].

Aunque los sistemas RAG existen y Prócer ya implementaba un sistema funcional sobre servicios de Microsoft Azure [23] en la nube para su componente textual, la propuesta de co-creación con FAMAF representa un salto estratégico fundamental hacia la soberanía tecnológica. El objetivo principal es desarrollar

una arquitectura de IA propia y soberana para el motor RAG textual y el modelo de lenguaje generador de respuestas. Este desarrollo permitirá tener un motor RAG textual propio, personalización avanzada para diversos casos de uso educativo y requerimientos de hardware y la propiedad intelectual de un activo tecnológico central en este dominio.

1.4. Naturaleza de la Colaboración

La vinculación se estructura como un proyecto de I+D y transferencia tecnológica con una duración de cuatro meses. El plan de trabajo, financiado parcialmente mediante el programa Voucher de Innovación Colaborativa de la Provincia de Córdoba, abarca desde el análisis de requerimientos y diseño de embeddings hasta la integración de un MVP (Producto Mínimo Viable) funcional en el dispositivo creado por la empresa.

2. Solución Técnica y Transferencia

Un aspecto central de esta colaboración fue la transición estratégica de Prócer desde servicios de terceros hacia una arquitectura propia y soberana. Esta decisión no fue sólo técnica, sino que responde a tres imperativos:

1. **Privacidad y Ética:** En el ámbito educativo, se procesan datos de menores y materiales sensibles. Mantener la lógica de inferencia en servidores propios (o en el dispositivo) garantiza que la información no sea utilizada para entrenar modelos externos ni quede sujeta a políticas de privacidad de empresas extranjeras.
2. **Sostenibilidad Económica:** El modelo de pago por “token” de las APIs comerciales representa un riesgo de escalabilidad para una PyME local. La migración a modelos *open source* permite a Prócer predecir costos y ofrecer soluciones asequibles a escuelas públicas de la región [19,8,21,18].
3. **Independencia Operativa y Offline-first:** La dependencia de la conectividad es una barrera crítica en el interior de Argentina y LATAM. La co-creación con FAMAF se centró en optimizar modelos para que, en un futuro cercano, el motor RAG pueda funcionar totalmente *on-device*, eliminando la dependencia de internet [39,12].

Esta sinergia permitió que FAMAF aporte el rigor científico en la validación de estos modelos locales (asegurando que su precisión fuera comparable a la de los modelos privativos), mientras que la empresa aportó la visión de mercado necesaria para convertir la teoría en un activo tecnológico propio.

La arquitectura resultante, denominada *Prócer RAG*, fue diseñada para ser agnóstica al modelo de lenguaje subyacente pero optimizada para el idioma español y contextos pedagógicos [38,18].

Tabla 1. Componentes del sistema transferido a la industria.

Capa	Tecnología	Criterio Industrial
API / Orquestación	FastAPI + LangChain [33,4]	Escalabilidad y modularidad.
Base de Datos	PostgreSQL + pgvector [15]	Estabilidad y soporte de búsqueda híbrida.
Modelos (LLM)	DeepSeek / Llama 3 [8,19]	Balance entre precisión factual y costo.
Infraestructura	Ollama / On-premise [39]	Soberanía tecnológica y offline-first.

2.1. Evaluación Experimental y Métricas

La validación se realizó con RAGAS [9] sobre datos reales del sistema en la nube de Prócer: 31 perfiles activos, 76 documentos educativos en español y 396 pares pregunta-respuesta generados con GPT-5.2 [26] y revisados manualmente. Se evaluó el sistema completo transferido: base vectorial, búsqueda híbrida (vectorial + *full-text search*) y reranking con BAAI/bge-reranker-v2-m3 [7,5]. Esta configuración combina recuperación semántica, coincidencias léxicas exactas y reordenamiento por pertinencia antes de entregar el contexto al LLM.

Se compararon 10 LLMs locales con Ollama y cuantización Q4_K_M, entre 4B y 14B parámetros [11,14,19,8,31,24,1]. Las métricas incluyeron Factual Correctness, Answer Accuracy, BLEU, ChrF y ROUGE-L [32,25,27,29,17].

Tabla 2. Resumen de resultados con el sistema completo.

Métrica	Rango observado	Mejor resultado
Factual Correctness	0.587–0.633	Gemma 3 12B (0.633)
Answer Accuracy	0.734–0.857	Qwen 3 8B (0.857)
BLEU	0.164–0.229	Gemma 3 12B (0.229)
ChrF	0.475–0.542	Gemma 3 12B (0.542)
ROUGE-L	0.309–0.447	Gemma 3 12B (0.447)

Los resultados muestran respuestas consistentes en todos los modelos evaluados. Qwen 3 8B alcanzó la mayor Answer Accuracy, mientras que Gemma 3 12B lideró en Factual Correctness y métricas textuales, sugiriendo que la calidad de la recuperación y el reranking pesan tanto como el tamaño del modelo.

2.2. Multimodalidad y Accesibilidad

Para abordar contenidos visuales (como fotos de pizarras), la metodología incluyó un puente hacia modelos de visión (VLM). La imagen es procesada externamente solo cuando es estrictamente necesario, convirtiendo el contenido visual en contexto textual que luego alimenta al RAG soberano [11,31]. Esto permite que el dispositivo de la empresa mantenga la mayor parte de su inteligencia de forma local y privada.

3. Impacto Social e Industrial

El éxito de esta vinculación se mide tanto en el beneficio social como en el crecimiento industrial. El **impacto educativo** es directo: estudiantes con dislexia o discapacidad intelectual acceden ahora a una consulta autónoma de textos, mitigando la brecha que hoy reduce en un 23 % sus posibilidades de completar la secundaria en la región [20]. Paralelamente, se produce un **fortalecimiento empresarial** estratégico, donde Prócer Tecnologías evoluciona su producto hacia un “tutor inteligente”. Este salto cualitativo permite la apertura de nuevas unidades de negocio basadas en el licenciamiento de modelos de IA soberanos para gobiernos y organismos educativos, consolidando la propiedad intelectual desarrollada conjuntamente con la Universidad.

4. Conclusiones

El vínculo entre FAMAF-UNC y Prócer Tecnologías demuestra que la investigación en IA no debe ser un proceso aislado de las realidades productivas y sociales [2,6]. La metodología de colaboración empleada permitió transformar papers académicos sobre RAG en una solución funcional que mejora la vida de miles de personas. El futuro de esta alianza apunta a la optimización de modelos para su ejecución totalmente *on-device*, eliminando las barreras de conectividad en escuelas rurales y reforzando el liderazgo de la tecnología cordobesa en el campo de la inclusión educativa.

Agradecimientos

Este trabajo utilizó recursos computacionales de UNC Supercómputo (CCAD) de la Universidad Nacional de Córdoba (<https://supercomputo.unc.edu.ar>), que forman parte del Sistema Nacional de Computación de Alto Desempeño (SNCAD) de la República Argentina.

Referencias

1. Abdin, M., et al.: Phi-4 technical report. arXiv preprint arXiv:2412.08905 (2024)
2. Britto, F.A., Lugones, G.E.: Bases y determinantes para una colaboración exitosa entre ciencia y producción. CIECTI (2020)
3. CAF – Banco de Desarrollo de América Latina: Inteligencia artificial en el sector de la educación. Tech. rep., CAF – Banco de Desarrollo de América Latina (2021), <https://scioteca.caf.com/handle/123456789/1928>
4. Chase, H., et al.: LangChain. <https://github.com/langchain-ai/langchain> (2022)
5. Chen, J., Xiao, S., Zhang, P., Luo, K., Lian, D., Liu, Z.: M3-Embedding: Multilinguality, multi-functionality, multi-granularity text embeddings through self-knowledge distillation. arXiv preprint arXiv:2402.03216 (2024)
6. Confraria, H., Vargas, F.: Scientific systems in latin america: performance, networks, and collaborations with industry. The Journal of Technology Transfer **44**, 874–915 (2019). <https://doi.org/10.1007/s10961-017-9631-7>

7. Cormack, G.V., Clarke, C.L., Buettcher, S.: Reciprocal rank fusion outperforms Condorcet and individual rank learning methods. In: Proceedings of the 32nd Annual ACM SIGIR Conference. pp. 758–759. ACM (2009)
8. DeepSeek-AI: DeepSeek-R1: Incentivizing reasoning capability in LLMs via reinforcement learning. arXiv preprint arXiv:2501.12948 (2025)
9. Es, S., James, J., Espinosa-Anke, L., Schockaert, S.: RAGAS: Automated evaluation of retrieval augmented generation. arXiv preprint arXiv:2309.15217 (2023)
10. Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M., Wang, H.: Retrieval-augmented generation for large language models: A survey. arXiv preprint arXiv:2312.10997 (2023)
11. Gemma Team, Google DeepMind: Gemma 3 technical report. arXiv preprint arXiv:2503.19786 (2025)
12. Gerganov, G., et al.: llama.cpp: LLM inference in C/C++. <https://github.com/ggerganov/llama.cpp> (2023)
13. Instituto Nacional de Estadística y Censos (INDEC): Estudio nacional sobre el perfil de las personas con discapacidad: Resultados definitivos 2018. Tech. rep., INDEC, Ciudad Autónoma de Buenos Aires, Argentina (2018), <https://www.indec.gob.ar/ftp/cuadros/poblacion/estudio.discapacidad.12.18.pdf>, iISBN 978-950-896-532-5
14. Jiang, A.Q., et al.: Mistral 7B. arXiv preprint arXiv:2310.06825 (2023)
15. Kane, A.: pgvector: Open-source vector similarity search for Postgres. <https://github.com/pgvector/pgvector> (2021)
16. Lewis, P., et al.: Retrieval-augmented generation for knowledge-intensive NLP tasks. In: Advances in Neural Information Processing Systems. vol. 33, pp. 9459–9474 (2020)
17. Lin, C.Y.: ROUGE: A package for automatic evaluation of summaries. In: Text Summarization Branches Out. pp. 74–81. ACL (2004)
18. Lin, M.P.C., Huang, J.Y., Chang, D.H., Tembrevilla, G., Bowen, G.M., Poitras, E., Janarthanan, V., Ryoo, J.: Open-source large language models in education: A narrative review of evidence, pedagogical roles, and learning outcomes. *AI Educ.* **2**(1), 4 (2026). <https://doi.org/10.3390/aieduc2010004>
19. Llama Team, Meta AI: The Llama 3 herd of models. arXiv preprint arXiv:2407.21783 (2024)
20. Male, C., Wodon, Q.T.: Disability gaps in educational attainment and literacy. Report, World Bank Group, Washington, D.C. (2017), <http://documents.worldbank.org/curated/en/396291511988894028>
21. Manchanda, J., Boettcher, L., Westphalen, M., Jasser, J.: The open source advantage in large language models (LLMs). arXiv preprint arXiv:2412.12004 (2024)
22. Miao, F., Holmes, W., Huang, R., Zhang, H.: AI and Education: Guidance for Policy-makers. UNESCO (2021), <https://unesdoc.unesco.org/ark:/48223/pf0000376709>
23. Microsoft: Microsoft Azure. <https://azure.microsoft.com> (2010)
24. Mistral AI: Mistral NeMo. <https://mistral.ai/news/mistral-nemo/> (2024)
25. NVIDIA Corporation: Answer Accuracy – RAGAS documentation. https://docs.ragas.io/en/latest/concepts/metrics/available_metrics/nvidia_metrics/#answer-accuracy (2025)
26. OpenAI: Introducing GPT-5.2. <https://openai.com/es-419/index/introducing-gpt-5-2/> (2025)
27. Papineni, K., Roukos, S., Ward, T., Zhu, W.J.: BLEU: a method for automatic evaluation of machine translation. In: Proceedings of the 40th Annual Meeting of the ACL. pp. 311–318 (2002)

28. Pombo, C., Cabrol, M., González Alarcón, N., Sánchez Ávalos, R.: fAIr LAC: Adopción ética y responsable de la inteligencia artificial en américa latina y el caribe. Tech. rep., Banco Interamericano de Desarrollo (2020). <https://doi.org/10.18235/0002169>, <https://publications.iadb.org/es/fair-lac-adopcion-etica-y-responsable-de-la-inteligencia-artificial-en-america-latina-y-el-caribe>
29. Popović, M.: chrF: character n-gram F-score for automatic MT evaluation. In: Proceedings of the Tenth Workshop on Statistical Machine Translation. pp. 392–395. ACL (2015)
30. Procer Tecnologías S.A.S.: Procer 4: Lector auditivo portátil con inteligencia artificial. <https://www.procertecnologias.com/products/procer4.es> (2024)
31. Qwen Team: Qwen3 technical report. arXiv preprint arXiv:2505.09388 (2025)
32. Factual Correctness – RAGAS documentation. https://docs.ragas.io/en/latest/concepts/metrics/available_metrics/factual_correctness/ (2025)
33. Ramírez, S.: FastAPI. <https://fastapi.tiangolo.com> (2018)
34. Resolución decanal 758/2025: Servicio de desarrollo de la Facultad de Matemática, Astronomía, Física y Computación (FAMAF) de la Universidad Nacional de Córdoba (UNC) para la empresa Procer Tecnologías SAS. Digesto Electrónico de la UNC <https://digesto.unc.edu.ar/handle/123456789/598620> (2025)
35. UNESCO Regional Office in Santiago: Observatory on artificial intelligence in education for latin america and the caribbean. <https://www.unesco.org/en/articles/observatory-artificial-intelligence-education-latin-america-and-caribbean> (2026)
36. UNICEF, Mir, C., Rubio, M., Fiszbein, A., Escaroz, G., Gomez, C.: Protección social inclusiva para la niñez con discapacidad en latinoamérica y el caribe: Marco regional. Tech. rep., Fondo de las Naciones Unidas para la Infancia (UNICEF), Panamá, República de Panamá (nov 2023), <https://www.unicef.org/lac/media/44326/file/Protecci%C3%B3n%20social%20inclusiva%20para%20la%20ni%C3%B1ez%20con%20discapacidad%20en%20Latinoam%C3%A9rica%20y%20el%20Caribe:%20Marco%20Regional.pdf>, iniciativa conjunta de UNICEF y Diálogo Interamericano
37. Voucher de innovación colaborativa (prensa agencia innovar y emprender córdoba). <https://prensa.cba.gov.ar/informacion-general/innovacion-abierta-y-financiamiento-lanzan-dos-programas-clave-para-empresas-y-startups/> (2025)
38. Xu, H., Gan, W., Qi, Z., Wu, J., Yu, P.S.: Large language models for education: A survey. arXiv preprint arXiv:2405.13001 (2024)
39. Ollama. <https://ollama.com> (2023)