

A figyelem határai: pozícióérzékenység vizsgálata RAG-rendszerekben

Antal Margit

Sapientia EMTE, Marosvásárhely, Románia

manyi@ms.sapientia.ro,

A visszakereséssel kiegészített generálás (RAG) hatékonysága nemcsak azon múlik, hogy milyen releváns dokumentumokat sikerül visszakeresni, hanem azon is, hogy ezek hol helyezkednek el a nyelvi modell kontextusablakában. Tanulmányunk azt vizsgálja, hogy az úgynevezett "középen elveszett" jelenség – amikor a modell rosszabbul hasznosítja a kontextus közepén elhelyezkedő információt – mennyire általánosan érvényesül valós RAG-rendszerekben: különböző adathalmazok és nyelvi modellek esetén, milyen dokumentumjellemzőkkel függ össze, és hogy egy kereszt-enkóder alapú újrarangsorolás enyhítheti-e a hatást.

Kísérleteinkben a helyes választ tartalmazó dokumentumot szisztematikusan az öt visszakeresett dokumentum elejére, közepére, illetve végére helyeztük, majd egy külön nyelvi modellel értékeltettük a generált válaszok minőségét. Az eredmények következetes és szignifikáns teljesítményromlást mutatnak, amikor a releváns dokumentum a kontextus közepén található – ez a hatás azonban jelentősen eltérő mértékű: rövid, gondosan szerkesztett dokumentumok esetén szerény, míg hosszabb, heterogénebb dokumentumok esetén sokkal kifejezettebb. Ez arra utal, hogy a "középen elveszett" jelenség erőssége nagyban függ a dokumentumok jellegétől is.

Kulcsszavak: visszakereséssel kiegészített generálás (RAG), "középen elveszett" hatás, kontextuspozíció-torzítás

References

- [1] Liu N.F. et al. (2024) Lost in the Middle: How Language Models Use Long Contexts *Transactions of the Association for Computational Linguistics*, vol. 12, pp. 157–173 10.1162/tac1_a_00638
- [2] Gabín, J., Pérez, A., Parapar, J. (2026) Lost in the Evidence? Reproducing Document Position and Context Size Effects in RAG, *Proceedings of the 49th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '26*, pp. 1–10 [url-https://doi.org/10.1145/3805712.3808569](https://doi.org/10.1145/3805712.3808569)
- [3] Antal, M., Buza, K. (2025) Evaluating Open-Source LLMs in RAG Systems: A Benchmark on Diploma Theses Abstracts Using Ragas. *Acta Univ. Sapientiae Inform.* 17, 5 (2025). <https://doi.org/10.1007/s44427-025-00006-3>