3° Πανελλήνιο Συνέδριο
Ιατρικής Φυσικής

27° Διαπανεπιστημιακό Συνέδριο
Ακτινολογίας

Retrieval-augmented generation (RAG) για μεγάλα γλωσσικά μοντέλα (LLMs) θεμελιωμένα σε έγκυρη γνώση στην ακτινολογία: βιβλιογραφική ανασκόπηση

Retrieval-augmented generation (RAG) for knowledge-grounded large language models (LLMs) in radiology: a narrative review

Κωνσταντίνος Σπανός

Ακτινολογικό Εργαστήριο, Αξονικός & Μαγνητικός Τομογράφος, Γ.Ν.Θ. «Γ. Παπανικολάου», Θεσσαλονίκη

Επικοινωνία: **spakon@yahoo.gr**

30, 31 ΟΚΤ + 1 ΝΟΕ
2026

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΥΓΕΙΑΣ
- ΤΜΗΜΑ ΙΑΤΡΙΚΗΣ, ΛΑΡΙΣΑ

Οργάνωση

ΕΦΙΕ
Ένωση Φυσικών Ιατρικής Ελλάδος

ΕΛΛΗΝΙΚΟ ΚΟΛΛΕΓΙΟ ΑΚΤΙΝΟΛΟΓΙΑΣ

Εισαγωγή

Γιατί χρειάζεται η RAG

Παραισθήσεις (hallucinations)
Εύλογες αλλά εσφαλμένες απαντήσεις, χωρίς προειδοποίηση.

Παρωχημένη γνώση
Η γνώση δεν επικαιροποιείται μετά την εκπαίδευση του μοντέλου.

Αδιαφάνεια πηγών
Η απάντηση δεν τεκμηριώνεται και δεν ελέγχεται.

Τι κάνει η RAG

Συνδέει το μοντέλο με **εξωτερική, επαληθεύσιμη γνώση** κατά την παραγωγή της απάντησης. Τα σχετικά αποσπάσματα ανακτώνται τη στιγμή του ερωτήματος και η απάντηση συντάσσεται με βάση αυτά, χωρίς επανεκπαίδευση του μοντέλου και με αναφορά στις πηγές της.

Πώς λειτουργεί

Κλινικό ερώτημα → **Ανάκτηση σχετικών αποσπασμάτων** → **Σύνθεση απάντησης από το LLM** → **Απάντηση με παραπομπές**

Βάση γνώσης

- Βιβλιογραφία και κατευθυντήριες οδηγίες
- Πρωτόκολλα εξετάσεων
- Προηγούμενες γνωματεύσεις και περιστατικά

Είτε διαδικτυακά σε πραγματικό χρόνο είτε τοπικά, με προστασία του απορρήτου.

Σκοπός
Να συνοψιστούν οι χρήσεις της RAG στην ακτινολογία ως προς την κλινική εφαρμογή, τις πηγές γνώσης, τον τρόπο ανάκτησης, τη μέθοδο αξιολόγησης και τους περιορισμούς.

Μέθοδοι & Υλικά

Σχεδιασμός της ανασκόπησης

1 Αναζήτηση

PubMed και αποθετήρια προδημοσιεύσεων (preprints), περίοδος 2023 έως 2026.

2 Επιλογή

Μελέτες από περιοδικά, προδημοσιεύσεις και ανασκοπήσεις για την ιατρική απεικόνιση, με ανάγνωση πλήρους κειμένου.

3 Χαρτογράφηση

Κάθε μελέτη καταγράφηκε σε τέσσερις άξονες. Ο χάρτης τεκμηρίωσης οργανώνει τη σύνθεση των ευρημάτων.

16
μελέτες πλήρους κειμένου

4
άξονες χαρτογράφησης

2023 έως 2026
περίοδος αναζήτησης

Χάρτης τεκμηρίωσης σε τέσσερις άξονες

Α Κλινική εφαρμογή

Ποιο έργο υποστηρίζει το σύστημα: απάντηση ερωτημάτων, σύνταξη γνωματεύσεων, εξαγωγή δεδομένων, υποστήριξη αποφάσεων.

Β Πηγή γνώσης και σχεδιασμός ανάκτησης

Τι ανακτάται, από πού και πώς: βάσεις δεδομένων, πρωτόκολλα, αρχεία περιστατικών, τοπικά ή διαδικτυακά.

Γ Μεθοδολογία αξιολόγησης

Πώς μετρήθηκε το όφελος: κρίση ειδικών, μετρικές ακρίβειας, εξειδικευμένα εργαλεία ελέγχου.

Δ Περιορισμοί

Τι αναφέρεται ως αδυναμία: ποιότητα ανάκτησης, δεδομένα, υπολογιστικοί πόροι, έλεγχος σε πραγματικές συνθήκες.

Αποτελέσματα

Α · Κλινικές εφαρμογές

Απάντηση διαγνωστικών ερωτημάτων

Αύξηση της ακρίβειας και λιγότερες εσφαλμένες απαντήσεις, με όφελος που εξαρτάται από το μοντέλο [5].

Παράδειγμα: βάση 3.689 άρθρων RadioGraphics, άνοδος του GPT-4 από 75,5% σε 81,2% σε ακτινολογική εξέταση [5].

Σύνταξη και σύνοψη γνωματεύσεων

Η ανάκτηση ομοίων περιστατικών και προηγούμενων γνωματεύσεων βελτιώνει την πληρότητα και την ορθότητα [6-9].

Παράδειγμα: ανάκτηση ομοίων ακτινογραφιών θώρακα και φράσεων από προηγούμενες γνωματεύσεις [8, 9].

Εξαγωγή δομημένων δεδομένων

Κατηγοριοποίηση CAD-RADS, βαθμολόγηση BT-RADS και σταδιοποίηση από το κείμενο των γνωματεύσεων [10, 11].

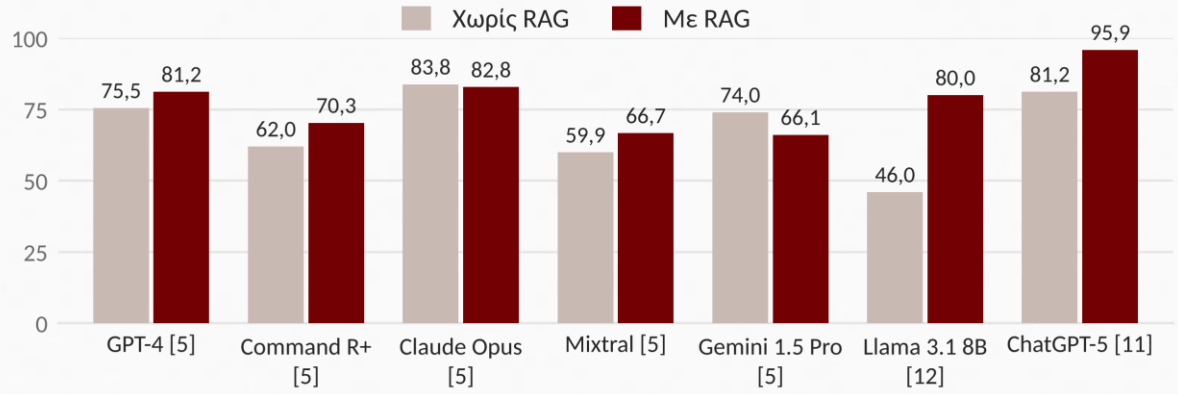
Παράδειγμα: ακρίβεια κατηγοριοποίησης CAD-RADS από 81,2% σε 95,9% με RAG, σε 320 γνωματεύσεις CCTA [11].

Υποστήριξη κλινικών αποφάσεων

Επιλογή της κατάλληλης εξέτασης βάσει οδηγιών (ACR), καθορισμός πρωτοκόλλου, σύσταση σκιαγραφικού [12, 13].

Παράδειγμα: τοπικό Llama με οδηγίες ACR, κατάλληλες συστάσεις σε 80% των σεναρίων έναντι 46% χωρίς RAG [12].

Ακρίβεια (%) χωρίς και με RAG, ίδιο μοντέλο ανά ζεύγος



Διαφορετικό έργο ανά μελέτη: ακτινολογική εξέταση [5], συστάσεις απεικόνισης κατά ACR [12], κατηγοριοποίηση CAD-RADS [11]. Σε τοπικό μοντέλο η RAG μηδένισε τις παραισθήσεις (8% σε 0%) [13].

Β · Πηγές γνώσης και ανάκτηση

Βιβλιογραφικές βάσεις δεδομένων, πρωτόκολλα εξετάσεων και αρχεία περιστατικών, είτε διαδικτυακά σε πραγματικό χρόνο είτε τοπικά.

Νεότερες κατευθύνσεις: graph RAG και agentic RAG [1, 4]. Οφέλη: συνέπεια, διαφάνεια, προσαρμοστικότητα χωρίς επανεκπαίδευση.

Γ · Αξιολόγηση

Δεν είναι τυποποιημένη: κρίση ειδικών, μετρικές ακρίβειας και εξειδικευμένα εργαλεία, όπως το RAG-Check [14].

Διαφάνεια στην αναφορά: λίστες ελέγχου όπως το MI-CLEAR-LLM καθορίζουν τι πρέπει να δημοσιεύεται για κάθε μελέτη LLM [15, 16].

Συμπεράσματα

Η RAG είναι ένας ρεαλιστικός τρόπος να γίνουν τα LLMs ασφαλέστερα και πιο χρήσιμα στην κλινική πράξη. Το όφελος όμως δεν είναι δεδομένο και πρέπει να μετριέται.

Βασικά μηνύματα

1. **Θεμελίωση σε επαληθεύσιμη γνώση**
   Λιγότερες παραισθήσεις, επικαιροποιημένη γνώση και απαντήσεις με πηγές, χωρίς επανεκπαίδευση του μοντέλου.

2. **Το όφελος εξαρτάται από το πλαίσιο**
   Διαφέρει ανάλογα με το μοντέλο, το κλινικό έργο και την ποιότητα των ανακτημένων πηγών. Σε ορισμένα μοντέλα δεν καταγράφεται όφελος.

3. **Τοπική λειτουργία με προστασία του απορρήτου**
   Η RAG σε τοπικά εγκατεστημένα μοντέλα προσφέρει αξιόπιστη υποστήριξη χωρίς τα δεδομένα των ασθενών να εξέρχονται από το νοσοκομείο.

Δ · Περιορισμοί της τρέχουσας τεκμηρίωσης

- Άντληση μη συναφών ή λανθασμένων πληροφοριών, που περνούν στην απάντηση.
- Εστίαση πολλών μελετών στην ακτινογραφία θώρακα.
- Χρήση προσομοιωμένων αντί για πραγματικά δεδομένα.
- Υψηλές απαιτήσεις σε υπολογιστικούς πόρους.
- Έλλειψη ελέγχου σε πραγματικές συνθήκες εργασίας.

Τι χρειάζεται

- **Τυποποιημένη αξιολόγηση και διαφανής αναφορά, κατά τη λίστα ελέγχου MI-CLEAR-LLM**
- **Ένταξη και έλεγχος σε πραγματικές συνθήκες, στην καθημερινή ροή εργασίας**

Βιβλιογραφία

1. Fink A, Rau A, Reisert M, et al. Retrieval-augmented generation with large language models in radiology: from theory to practice. *Radiol Artif Intell. 2025, 7(4): e240790.* doi:10.1148/ryai.240790

2. Soni N, Ora M, Agarwal A, et al. A review of the opportunities and challenges with large language models in radiology: the road ahead. *AJNR Am J Neuroradiol. 2025, 46(7): 1292-1299.* doi:10.3174/ajnr.A8589

3. Neha F, Bhati D, Shukla DK. Retrieval-augmented generation (RAG) in healthcare: a comprehensive review. *AI. 2025, 6(9): 226.* doi:10.3390/ai6090226

4. Salehi S, Singh Y, Horst KK, et al. Agentic AI and large language models in radiology: opportunities and hallucination challenges. *Bioengineering. 2025, 12(12): 1303.* doi:10.3390/bioengineering12121303

5. Weinert DA, Rauschecker AM. Enhancing large language models with retrieval-augmented generation: a radiology-specific approach. *Radiol Artif Intell. 2025, 7(3): e240313.* doi:10.1148/ryai.240313

6. Sun L, Zhao J, Han M, Xiong C. Fact-aware multimodal retrieval augmentation for accurate medical radiology report generation. *arXiv. 2024.* doi:10.48550/arXiv.2407.15268

7. Hou W, Cheng Y, Xu K, et al. RADAR: enhancing radiology report generation with supplementary knowledge injection. *arXiv. 2025.* doi:10.48550/arXiv.2505.14318

8. Park J, Yoon B, Kim S, Choi K. RA-RRG: multimodal retrieval-augmented radiology report generation with key phrase extraction. *arXiv. 2025.* doi:10.48550/arXiv.2504.07415

9. Samanta HS. Grounded multimodal retrieval-augmented drafting of radiology impressions using case-based similarity search. *arXiv. 2026.* doi:10.48550/arXiv.2603.17765

10. Jabal MS, Warman P, Zhang J, et al. Open-weight language models and retrieval-augmented generation for automated structured data extraction from diagnostic reports: assessment of approaches and parameters. *Radiol Artif Intell. 2025, 7(3): e240551.* doi:10.1148/ryai.240551

11. Kaba E, Çubukçu Y, Uzunibrahimoğlu B, et al. Retrieval-augmented generation-enhanced large language models for comprehensive CAD-RADS 2.0 categorization from structured coronary CTA reports. *Radiologie (Heidelb). 2026.* doi:10.1007/s00117-026-01580-z

12. Gupta A, Rangarajan K, Krishna Kumar RG, Anshal S. Locally deployed context-aware chatbot outperforms generic large language models for guideline-concordant pediatric imaging recommendations. *Pediatr Radiol. 2026, 56(9): 2048-2055.* doi:10.1007/s00247-025-06453-6

13. Wada A, Tanaka Y, Nishizawa M, et al. Retrieval-augmented generation elevates local LLM quality in radiology contrast media consultation. *NPJ Digit Med. 2025, 8(1): 395.* doi:10.1038/s41746-025-01802-z

14. Mortaheb M, Khojastepour MA, Chakradhar ST, Ulukus S. RAG-Check: evaluating multimodal retrieval augmented generation performance. *arXiv. 2025.* doi:10.48550/arXiv.2501.03995

15. Park SH, Suh CH, Lee JH, et al. Minimum reporting items for clear evaluation of accuracy reports of large language models in healthcare (MI-CLEAR-LLM). *Korean J Radiol. 2024, 25(10): 865-868.* doi:10.3348/kjr.2024.0843

16. Park SH, Suh CH, Lee JH, et al. Minimum reporting items for clear evaluation of accuracy reports of large language models in healthcare (MI-CLEAR-LLM): 2025 updates. *Korean J Radiol. 2025, 26(12): 1123-1132.* doi:10.3348/kjr.2025.1522