DERLEME
DOI:10.4274/nts.galenos.2026.64936
Nucl Med Semin 2026;12:234-238

Büyük Dil Modellerinin Klinik Kullanım Alanları

Clinical Applications of Large Language Models

Aysu Dursun[1], Nazım Coşkun[2]

[1]Sağlık Bilimleri Üniversitesi, Ankara Etlik Şehir Hastanesi, Nükleer Tıp Kliniği, Ankara, Türkiye
[2]Sağlık Bilimleri Üniversitesi, Ankara Bilkent Şehir Hastanesi, Nükleer Tıp Kliniği, Ankara, Türkiye

Öz

Büyük dil modelleri (LLM), doğal dil işleme alanındaki hızlı gelişmelerle birlikte sağlık hizmetlerinde giderek daha fazla kullanılmaktadır. Nükleer tıpta LLM'ler klinik karar desteği, raporlama, akademik yazım, literatür taraması, eğitim ve idari süreçlerde önemli bir potansiyel taşımaktadır. Multimodal modeller ise pozitron emisyon tomografisi ve tek foton emisyonlu bilgisayarlı tomografi görüntülerinin klinik ve metinsel verilerle birlikte değerlendirilmesine olanak sağlayabilir. Bununla birlikte halüsinasyon, önyargı, açıklanabilirlik eksikliği, sayısal verilerin değerlendirilmesindeki sınırlılıklar, veri güvenliği ve hukuki sorumluluk önemli sorunlar olmaya devam etmektedir. Mevcut kanıtlar, LLM'lerin özellikle yapılandırılmış metin ve idari görevlerde yararlı olabileceğini, ancak klinik karar verme ve hasta güvenliğini etkileyen uygulamalarda hekim denetiminin vazgeçilmez olduğunu göstermektedir.
Anahtar Kelimeler: Büyük dil modelleri, yapay zeka, nükleer tıp, PET/BT, klinik karar desteği

Abstract

Large language models (LLMs) are increasingly being used in healthcare as natural language processing technologies rapidly evolve. In nuclear medicine, LLMs have considerable potential in clinical decision support, reporting, academic writing, literature review, education, and administrative workflows. Multimodal models may further enable the combined interpretation of positron emission tomography and single-photon emission computed tomography images with clinical and textual information. However, hallucinations, bias, limited explainability, weaknesses in handling quantitative data, data privacy, and legal responsibility remain important challenges. Current evidence suggests that LLMs may already provide meaningful benefits in structured text-based and administrative tasks. In contrast, applications involving clinical reasoning and patient safety still require careful validation and physician oversight. Responsible implementation should therefore maintain the physician as the final decision-maker.
Keywords: Large language models, artificial intelligence, nuclear medicine, PET/CT, clinical decision support

Giriş

Güçlü hesaplama kaynaklarının ve eğitim için büyük veri kümelerinin geliştirilmesi ilerledikçe, sağlık hizmetlerinde derin öğrenme teknikleri daha önemli rol oynamaya başlamıştır. Nükleer tıp kliniklerinde raporlama, standardizasyon, dökümantasyon hata ve tutarsızlık tespiti, klinik karar verme sürecinde ayırıcı tanı önerileri, hasta anamnezlerinin incelenmesi, radyofarmasötik geliştirme, teranostik hasta seçimi ve doz planlama gibi çok fazla alanda ve konuda büyük dil modellerinin (LLM) kullanımı ivmelenerek yaygınlaşmaktadır (1,2).

Bu derlemede, nükleer tıp alanında LLM'nin güçlü ve sınırlı yönlerinin ele alınması ve klinik kullanımlarına ilişkin genel bir çerçeve sunulması amaçlanmıştır.

LLM Nedir?

LLM'ler çok büyük metin veri kümeleri üzerinde (makaleler, kitaplar ve diğer internet tabanlı içerikler)

Sorumlu Yazar/Corresponding Author
Dr. Aysu Dursun, Sağlık Bilimleri Üniversitesi, Ankara Etlik Şehir Hastanesi, Nükleer Tıp Kliniği, Ankara, Türkiye
E-posta: usyasund@gmail.com **ORCID ID:** orcid.org/0009-0000-1082-3973
Geliş Tarihi/Received: 04.10.2026 **Kabul Tarihi/Accepted:** 04.11.2026 **Yayınlanma Tarihi/Publication Date:** 26.11.2026
Cite this article as/Atıf: Dursun A, Coşkun N. Clinical applications of large language models. Nucl Med Semin. 2026;12:234-238

Copyright© 2026 Yazar(lar). Türkiye Nükleer Tıp Derneği adına Galenos Yayınevi tarafından yayımlanmıştır.
Creative Commons Atıf-GayriTicari 4.0 (CC BY-NC 4.0) Uluslararası Lisansı ile lisanslanmış, açık erişimli bir makaledir.
Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of Turkish Society of Nuclear Medicine.
This is an open access article under the Creative Commons Attribution Non-Commercial 4.0 International (CC BY-NC 4.0) License.

eğitilmiş yapay zeka sistemleridir. Kelimeler arasındaki karmaşık ilişkisel bağlantıları temsil ederek insan benzeri metin üretmek üzere derin öğrenmeyi kullanan sinir ağı mimarileri olarak tanımlanabilir. LLM'ler kelimelerin dilde nasıl ilişkisel biçimde kullanıldığını öğrenir ve bu öğrenilen kalıpları doğal dil işleme görevlerini tamamlamak için uygular (3).

Doğal dil işleme (*natural language processing*), dilin insan yeteneğini taklit edecek biçimde otomatik olarak analiz edilmesini ve üretilmesini amaçlayan bir hesaplamalı bilim alanıdır. LLM, bu alanın en gelişmiş ürünüdür. Transformer mimarisi sayesinde çok büyük veri hacimlerini paralel olarak işleyebilmekte ve dili anlama ile üretme kapasitesinde önemli bir sıçrama sağlamaktadır (3).

Transformer Mimarisi

Modern LLM'lerin temelinde 2017 yılında tanıtılan transformer mimarisi yatar. *Transformer*'ın ayırt edici bileşeni, bir metindeki tüm kelime ve kelime parçacıkları (token'lar) arasındaki ilişkileri aynı anda değerlendiren "dikkat" (*attention*) mekanizmasıdır. Bu sayede model, uzun metinlerdeki bağlamı koruyabilir ve büyük veri hacimlerini paralel işleyerek klasik modellerin ötesinde bir dil anlama ve üretme yeteneği kazanır. Bu yaklaşımın klinik açıdan dikkat edilmesi gereken sonucu, modelin tutarlı ve akıcı metin üretme kapasitesinin, üretilen bilginin doğruluğunu garanti etmemesidir (3).

Tokenizasyon ve Dil Modellemesi Nedir?

LLM'ler metinsel verileri doğrudan kelime kelime değil, "token" adı verilen daha küçük parçalar halinde işler. "Token" bir kelime, kelimenin bir bölümü veya bir noktalama işareti olabilir. Bir yapbozun parçaları gibi düşünüldüğünde model parçalar arasındaki örüntüye dayanarak sıradaki yapboz parçalarını tahmin eder (Şekil 1). Bu durum, modelin bilgiyi bir veri tabanından çekerek gerçekten "bilmediğini"; bunun yerine eğitildiği verilerdeki dilsel örüntülerden yola çıkarak metinler ürettiğini gösterir. Bu ayrım da aslında LLM'lerin neden bazen kendinden emin biçimde ama yanlış çıktılar (diğer bir ismi ile halüsinasyonlar) üretebildiğini açıklayan temel noktadır. Token'lar arasında yapılan seçimler, aslında konunun bilgisel yüküne olan hakimiyetten daha çok bir sonraki token'in tahmini ile ilerlemektedir (3).

Multimodal Görüntü-Dil Modelleri

İlk kuşak LLM'ler yalnızca metinler ile çalışırken, multimodal (büyük çok-kipli) modeller metnin yanında görüntüyü de girdi olarak alabilir. Nükleer tıp kliniği açısından bu, pozitron emisyon tomografisi (PET) veya tek foton emisyon bilgisayarlı tomografisi (SPECT) görüntülerinin metinsel raporlarla birlikte yorumlanabilmesi potansiyeli demektir (3). PET/bilgisayarlı tomografi (BT) incelemelerinin görüntü değerlendirmesinden raporlamaya kadar tüm süreçlerini gerçekleştirmeye yönelik çok kipli sistemler geliştirilmektedir. Ancak bu sistemler henüz erken araştırma ve geliştirme aşamasındadır. Klinik uygulamada güvenilir ve etkin biçimde kullanılabileceklerini gösteren kapsamlı doğrulama çalışmaları henüz tamamlanmamıştır (4).

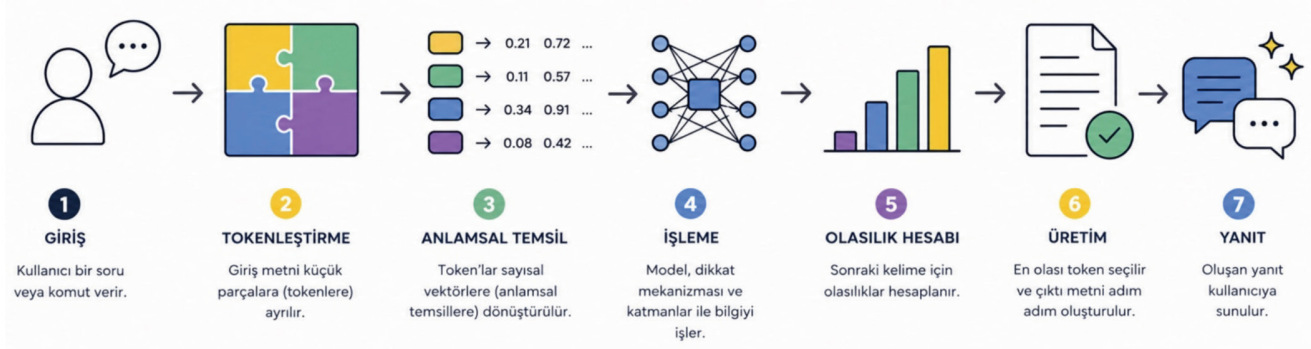
Yaygın LLM'ler

Günümüzde klinik ve akademik ortamlarda karşılaşılan başlıca genel amaçlı modeller arasında OpenAI tarafından geliştirilen ChatGPT, Anthropic tarafından geliştirilen Claude, Google tarafından geliştirilen Gemini ve xAI tarafından geliştirilen Grok sayılabilir. Bunların yanında, tıbbi metinler üzerinde özelleştirilmiş alan-özgü modeller (örneğin, Radiology-GPT, Radiology-Llama2) yer almaktadır. Modeller ve sürümleri hızla değişmekte olup performans değerlendirmelerinin kısa sürede güncelliğini yitirebileceği akılda tutulmalıdır.

Nükleer Tıpta LLM Kullanım Alanları

Klinik Karar Desteği

LLM'ler, farklı kaynaklardan gelen bilgiyi sentezleyerek klinisyene bağlama duyarlı öneriler sunabilir; uygun görüntüleme yönteminin seçilmesi ve karmaşık olguların yorumlanmasında yardımcı olabilir (5). Nükleer tıp kliniğine özgü örneklerde, onkolojik veri kümeleri üzerinde eğitilen modeller PET/BT'de tutulum örüntülerini özetleyip olası ayırıcı tanılar önerebilmekte; lenfoma değerlendirmesinde Deauville skorlaması için ince ayarlı modeller yüksek doğruluk gösterebilmektedir (5). Erişim-artırımlı üretim (*retrieval-augmented generation*, RAG) yaklaşımı, modeli kurumun kendi rapor arşivine bağlayarak doğruluğu belirgin biçimde artırmaktadır; 211.000'den fazla PET raporundan oluşan bir veri tabanıyla çalışan bir RAG sisteminde benzer olguların erişiminde %84 başarı bildirilmiştir (5,6).



Şekil 1. Büyük dil modellerinin çalışma prensibi

Akademik Yazım

LLM'ler; makale taslağı hazırlama, dil düzeltme ve metin özetleme gibi görevlerde araştırmacılara yardımcı olabilmektedir (2). Ancak bu kullanım, yazarlık, intihal ve akademik dürüstlük konularında ciddi sorular doğurmaktadır. Erken dönemde ChatGPT'nin bazı yayınlarda ortak yazar olarak listelenmesi tartışma yaratmış, birçok dergi yapay zeka destekli yazıma ilişkin editoryal politikalar geliştirmek zorunda kalmıştır (1,2).

Modellerin ikna edici ancak uydurma kaynaklar içeren özetler üretebildiği ve bunların zaman zaman hakemleri yanıltabildiği gösterilmiştir. Dolayısıyla LLM'ler akademik yazımda bir araç olarak kullanılabilirse de üretilen her ifadenin ve kaynağın yazar tarafından doğrulanması zorunludur ve bu sorumluluk tamamen yazara aittir (1,2).

Raporlama

Otomatik rapor üretimi, LLM'lerin en dönüştürücü uygulamalarından biridir. Modeller ilk izlenim taslaklarını hazırlamada, bulguları özetlemede ve rapor tutarlılığını sağlamada yardımcı olabilir (2,3). PACS veya RIS sistemleriyle entegre edildiğinde dikte süresini kısaltarak hekimin daha karmaşık olgulara odaklanmasına olanak tanır. LLM'ler ayrıca geçmiş raporları tarayarak çelişkileri, eksik bulguları ve tutarsız takip önerilerini saptayabilir (2,5). Bununla birlikte, LLM tarafından üretilen izlenimler tamlık ve doğrulukta güçlü olsa da özlülük ve gerçekçilik (*verisimilitude*) açısından hekim raporlarının gerisinde kalabilmektedir. Bu nedenle rapor yazma prompt'u ile elde edilen her çıktı için hekim denetimi gereklidir.

Literatür Tarama

LLM'ler, geniş literatürü tarayıp özetleyerek ve eğitim gören hekimlere kişiselleştirilmiş rehberlik sağlayarak günümüzde fazlasıyla kullanılmaktadır (2). RAG yaklaşımıyla güvenilir kaynaklara bağlandığında, örneğin; teranostik, özellikle peptit reseptör radyonüklid tedavisi gibi alt alanlarda yüksek doğrulukta yanıtlar üretilebildiği bildirilmiştir (5). Yine de modellerin var olmayan atıflar üretebilme eğilimi nedeniyle, literatür taramasında elde edilen her kaynağın birincil metinden teyit edilmesi gerekir.

İdari Süreçler ve Eğitim

Randevu planlaması, hasta hazırlık talimatları, DRG kodlaması ve sigorta yazışmaları gibi idari görevler, LLM'lerin en güvenli ve en yüksek getirili kullanım alanı olarak öne çıkmaktadır (5). Bu görevlerde hasta zararı riski düşük, iş yükü kazancı ise yüksektir. Prospektif olarak toplanan gerçek hasta sorularının kullanıldığı bir çalışmada, idari sorulara verilen LLM yanıtları tüm ölçütlerde uzman yanıtlarına eşdeğer ya da üstün bulunmuş; uzman olmayan değerlendiriciler idari soruların %86'sında LLM yanıtını insan yanıtına tercih etmiştir (7).

Kısıtlılıklar, Etik Sorunlar ve Veri Güvenliği

Doğruluk ve Halüsinasyon

Halüsinasyon LLM'lerin en temel kısıtlılığı olarak ele alınmakta olup modelin akla yatkın ancak yanlış içeriği yüksek bir özgüven ile üretmesi şeklinde tanımlanmaktadır (2,3). Erken bir değerlendirmede ChatGPT, Avrupa Nükleer Tıp Board Sınavı örnek sorularının yalnızca %34'ünü doğru yanıtlamış (rastgele şans yaklaşık %24) ve düzeltildikten sonra bile seçeneklerde/şıklarda yer almayan yanlış yanıtlar üretmiştir (1). Gerçek dünya koşullarında hasta sorularını değerlendiren daha önce değinilen prospektif bir çalışmada ise LLM yanıtları idari sorularda uzman performansını yakalamış veya aşmış; buna karşılık tıbbi

soruların %33'ünde uydurma/tahrif edilmiş bilgi ve %7'sinde potansiyel zarar riski saptanmıştır (7). Modelin kendi bilgi sınırının farkında olması (*self-awareness*) boyutu ise zayıf bulunmuştur (7).

Açıklanabilirlik, Önyargı ve Sayısal Veri Sınırı

LLM çıktılarının nasıl üretildiğinin şeffaf olmaması, klinisyen güvenini zedeleyen önemli bir engeldir. Zincirleme düşünme (*chain-of-thought*) ve belirsizlik ölçümü (*uncertainty quantification*) gibi tekniklerle şeffaflığın artırılması önerilmektedir. Modellerin eğitiminde kullanılan verilerdeki yanlılık önemli sorunlara sebep olabilmektedir. Bu da klinik kararlarda eşitsizliğe yol açabilir (2,5). Ayrıca dil temelli modeller ham sayısal veriyi çözümlemede ve istatistiksel hesaplamada görece zayıf olarak raporlanmış olup bu nedenle nükleer tıptaki asıl katkıları niceliksel görüntü değerlendirmesinden çok metin analizi ve rapor üretimiyle sınırlı kalmaktadır.

Bugün için LLM'lerden görüntünün kendisini yorumlamaları beklenmekten çok, en güvenilir katkıları sayısal ve görsel analiz sonuçlarının yazıya dökülmesi ve raporlanması olacaktır.

Veri Güvenliği ve Mevzuat

LLM'lerin klinik iş akışlarına entegrasyonu, hasta verisinin gizliliği ve Kişisel Verilerin Korunması Kanunu, *Health Insurance Portability and Accountability Act* gibi düzenlemelere uyum açısından önemli sorunlar doğurmaktadır. Elektronik sağlık kayıtlarının işlenmesinde veri güvenliği, yerinde (*on premises*) veya federatif öğrenme gibi gizliliği koruyan yaklaşımların benimsenmesini gerektirmektedir. Mevzuata uygunluk, sürekli doğrulama ve denetim, sorumlu bir benimseme için vazgeçilmezdir (2,5).

Etik ve Sorumluluk

Sorumluluk sorunu bu noktada karşımıza çıkabilecek en önemli ve henüz net biçimde çözülmemiş hukuki ve etik sonuçlar barındıran problemdir. Tıbbi düzenleyiciler hekimlerin kendi bilgi ve yetkinlik sınırları içinde uygulama yapmasını beklerken, LLM'ler bu sınırı bilmeksizin yüzeysel ve yanıltıcı yanıtlar verebilmektedir. Bu konuda yapılmış hemen hemen her çalışmada yer alan ortak vurgu sistemlerinin etik zorluklarının sistematik biçimde ve acilen ele alınması gerektiğidir (2,5).

Gelecek Perspektifi

Nükleer tıpta LLM araştırmalarının önümüzdeki dönemde birkaç öncelikli eksende yoğunlaşması beklenmektedir. Birincisi, görüntüleme, klinik notlar ve laboratuvar verisini birleştiren multimodal modeller, özellikle karmaşık olgularda tanısal kesinliği artırma potansiyeli taşımaktadır. İkincisi, elektronik sağlık kaydına entegre modellerle hastaya özgü, bağlama duyarlı öneriler sunan kişiselleştirilmiş karar desteği geliştirilebilir (2,5). Üçüncüsü, kaynak kısıtlı ortamlar için hafif (*lightweight*) modeller ve gizliliği koruyan federatif öğrenme yoluyla ölçeklenebilir dağıtım hedeflenmelidir. Son olarak, teranostik uygulamaları nükleer tıp ve onkoloji kliniklerinin merkezine yerleştikçe, üretken yapay zekanın ilaç tasarımı, hedef tahmini ve bireyselleştirilmiş tedavi protokollerine katkısı önem kazanacaktır. Tüm bu ilerlemelerin ortak koşulu, prospektif ve çok merkezli çalışmalarla, klinik sonuç odaklı ve standardize değerlendirme çerçeveleri altında sağlam kanıt üretilmesidir (2,5).

Sonuç

LLM, nükleer tıp kliniğinde iş yükünü azaltma, raporlamayı standardize etme, klinik karar desteğini güçlendirme ve eğitime katkı sunma bakımından belirgin bir potansiyel taşımaktadır. Bununla birlikte mevcut kanıtlar, faydanın göreve göre büyük ölçüde değiştiğini ortaya koymaktadır. İdari işlerde ve yapılandırılmış metin görevlerinde LLM'ler halihazırda insan performansını yakalayabilirken, klinik muhakeme, sertifikasyon düzeyinde bilgi ve hasta güvenliği gerektiren alanlarda henüz denetimsiz kullanıma hazır değildir.

Bu nedenle günümüzdeki esas soru "insan mı, yapay zeka mı" değil, "hangi görevin hangi düzeyde denetimle kime bırakılacağıdır". Halüsinasyon, önyargı, açıklanabilirlik eksikliği, veri güvenliği ve sorumluluk gibi sorunlar çözülmeden LLM'ler tıbbi kararda tek başına güvenilir kabul edilmemelidir. Bu modellerin klinikte kullanımı, insan denetimini merkeze alan, sorumluluğun hekimde olduğu ve son kararı yine hekimin verdiği bir benimseme yaklaşımını zorunlu kılmaktadır.

Dipnotlar

Çıkar Çatışması: Yazarlar tarafından çıkar çatışması bildirilmemiştir.

Finansal Destek: Yazarlar tarafından finansal destek almadıkları bildirilmiştir.

Yapay Zeka (AI) Kullanım Beyanı: Dil düzeltme ve metin düzenleme desteği alınmıştır.

Kaynaklar

1. Alberts IL, Mercolli L, Pyka T, et al. Large language models (LLM) and ChatGPT: what will the impact on nuclear medicine be? Eur J Nucl Med Mol Imaging. 2023;50:1549-1552.
2. Hirata K, Matsui Y, Yamada A, et al. Generative AI and large language models in nuclear medicine: current status and future prospects. Ann Nucl Med. 2024;38:853-864. Erratum in: Ann Nucl Med. 2025;39:404-405.
3. Bradshaw TJ, Tie X, Warner J, Hu J, Li Q, Li X. language models and large multimodal models in medical imaging: a primer for physicians. J Nucl Med. 2025;66:173-182.
4. Choi H, Bae S, Na KJ. End-to-end PET/CT interpretation and quantification with an LLM-orchestrated AI agent: a real-world pilot study. J Nucl Med. 2026;67:1668-1674.
5. Papageorgiou PS, Christodoulou RC, Pitsillos R, et al. The role of large language models in improving diagnostic-related groups assignment and clinical decision support in healthcare systems: an example from radiology and nuclear medicine. Appl Sci. 2025;15:9005.
6. Choi H, Lee D, Kang YK, Suh M. Empowering PET imaging reporting with retrieval-augmented large language models and reading reports database: a pilot single center study. Eur J Nucl Med Mol Imaging. 2025;52:2452-2462.
7. Latifoltojar A, Orunmuyi A, Wang M, et al. Real-world evaluation of large language model for patients medical and administrative queries in nuclear medicine. NPJ Digit Med. 2026;9:754.