npj | digital medicine

News & Views

Published in partnership with Seoul National University Bundang Hospital

https://doi.org/10.1038/s41746-025-02015-0

Artificial intelligence for autoimmune diseases

Arjun Mahajan, Avery H. LaChance, Adam Rodman, Lorinda Chung & Dylan Powell

Emerging evidence suggests generative artificial intelligence (AI) may offer potential for autoimmune and rheumatic disease care, moving beyond traditional narrow AI applications to produce contextualized clinical content to support a wide spectrum of medical tasks. This article explores generative AI applications across autoimmune and rheumatologic clinical care, research, and administrative domains. However, significant implementation challenges remain, including clinical validation, model interpretability, data integration complexities, and evolving regulatory frameworks.

Research suggests nearly one in ten individuals is affected by an autoimmune or rheumatic disease – conditions that largely lack definitive cures and show a rapidly rising incidence[1,2]. In light of these pressing challenges, recent *npj Digital Medicine* articles highlight how recent developments in artificial intelligence, particularly in generative artificial intelligence (GenAI), represent a potential paradigm shift in autoimmune and rheumatic disease care[3–5]. Traditional AI applications for rheumatic diseases have mainly focused on narrow tasks, like classification or prediction[6]. In contrast, generative models, such as large language models (LLMs), can produce contextualized clinical content and recommendations informed by large medical datasets, supporting a wide range of clinical tasks[6,7]. These models may be particularly valuable for rheumatologic practice, which is often defined by diagnostic uncertainty, varied disease presentations, and the need for individualized treatment[8,9]. Despite this alignment, the applications of this technology across autoimmune and rheumatologic clinical care, research, and administrative workflows remain largely unexplored.

AI-enhanced clinical decision-making

Generative AI demonstrates early potential to improve diagnostic accuracy, treatment guidance, and clinical decision-making in rheumatic disease management. GenAI's potential to advance clinical care partly lies in its ability to combine multiple types of clinical information; by integrating patient symptoms, laboratory data, imaging findings, and even genomic profiles, these systems can generate clinical insights that inform decision-making in real-time[3,7]. These systems show particular proficiency in navigating the complexity inherent to rheumatic diseases, where diagnostic and therapeutic decisions often require the synthesis of clinical elements across multiple organ systems and timeframes[9,10]. A recent validation study demonstrated that foundational LLMs achieved high diagnostic accuracy in inflammatory rheumatic diseases, identifying a higher proportion of expert-curated cases correctly compared to human specialists[11]. Similarly, GPT-4's performance in musculoskeletal radiology interpretation matched the diagnostic accuracy of radiology residents when provided with both medical history and imaging findings, highlighting its multimodal capabilities[12]. These platforms also excel in rare disease identification, with select LLMs achieving high accuracy in diagnosing uncommon and orphan diseases – a valuable capability given the rarity of many rheumatic diseases that may challenge even experienced clinicians[13,14]. GenAI and related advancements have also demonstrated the ability to incorporate dermatologic findings to enhance the evaluation and management of autoimmune diseases, as in diagnosing lupus or evaluating alopecia areata severity[15,16]. While these early findings are promising, the evidence base remains limited, predominantly derived from curated clinical vignettes and structured datasets rather than the real-world prospective trials and longitudinal studies typically needed for widespread implementation[10–12].

Platforms have also demonstrated accuracy in providing treatment guidance, with LLMs achieving high accuracy and concordance with guidelines when delivering methotrexate-related information for rheumatoid arthritis patients[17]. This early evidence suggests their potential to serve as reliable clinical decision support tools that can adapt to differing patient presentations while incorporating evolving evidence-based guidelines. Importantly, real-world evidence shows that clinician-AI collaboration can improve care quality: in a multicenter randomized controlled trial of physicians, LLM assistance increased correct management-reasoning scores without raising harmful decision rates, underscoring how interactive generative AI can translate into improvements in clinical outcomes[18]. For instance, a GenAI system could potentially evaluate a patient presenting with joint pain and a facial rash, recommend targeted autoantibody testing and evaluation for internal organ involvement including muscle inflammation and kidney function studies. It could then integrate these results to guide further testing to differentiate dermatomyositis from lupus diagnosis and recommend initial treatment, tailored to the individual's needs (Fig. 1).

Moving beyond the administrative maze

Multiple studies have suggested that physicians spend nearly twice as long on administrative work as on face-to-face care[19,20]. The administrative burden inherent to rheumatic disease management, characterized by extensive documentation and care coordination requirements, can outweigh that of other specialties, and presents a compelling target for generative AI optimization[21]. Current LLMs demonstrate significant capability in reducing documentation overhead through automated generation of clinical notes, discharge summaries, and administrative communications, with real-time suggestions that maintain clinical accuracy while improving efficiency[22–25].

These systems also show particular value in patient engagement through the generation of personalized educational materials that can be tailored to individual health literacy levels, cultural preferences, and disease-

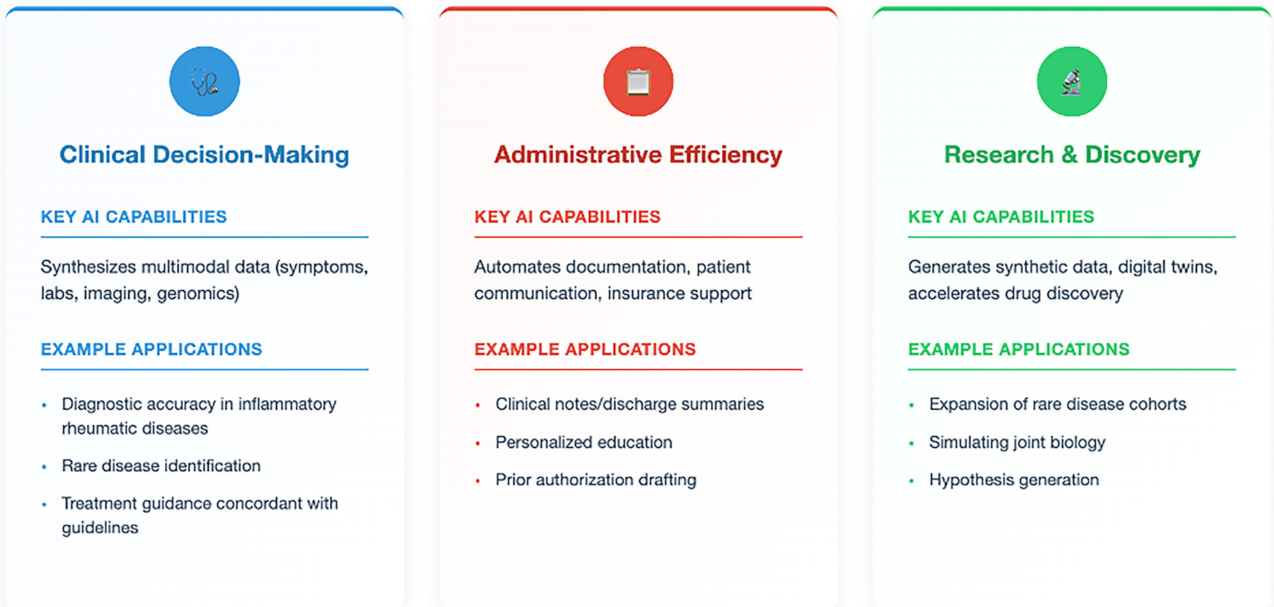

Fig. 1 | Selected applications of generative artificial intelligence in autoimmune disease care.

specific considerations – a critical capability given the complexity of rheumatic diseases and their treatments[26–28].

Additionally, generative models have the potential to streamline insurance authorization processes by automatically creating prior approval documentation and translating complex clinical rationales into formats required by payers, potentially reducing delays in accessing specialized rheumatologic treatments, like biologics[29,30]. Future implementations are anticipated to integrate ambient listening technologies that can automatically capture and structure clinical encounters, convert conversational exchanges into structured documentation, and provide real-time clinical decision support, thereby allowing clinicians to focus more directly on patient care rather than administrative tasks[31,32].

Accelerating discovery

Generative AI may help in reshaping the research landscape in rheumatic diseases by enabling novel approaches to drug discovery, clinical trial design, and scientific hypothesis generation.

In pharmaceutical development, transformer-based models have been successfully adapted for molecular representation learning, enabling the identification of novel therapeutic targets and prediction of molecular properties essential for drug development in rheumatic diseases[33]. Additionally, emerging generative drug repurposing technologies may potentially enable identification of new therapeutic applications for existing medications in autoimmune diseases[34]. These technologies can also help create digital twins - sophisticated patient simulation models trained on clinical trajectories, laboratory results, and treatment responses[35]. For instance, these simulations can reproduce the complex cellular interactions driving rheumatoid arthritis, enabling testing of drug mechanisms, prediction of disease trajectories, and identification of new therapeutic targets[36]. Similarly, GenAI can learn from existing patient data to generate synthetic patient profiles; by expanding small cohorts of rare diseases such as lupus, systemic sclerosis, and vasculitis, these profiles could provide the statistical power needed to compare treatments and predict patient outcomes with more confidence[37,38].

These systems also demonstrate the capacity to aid in medical research beyond data analysis: generating research hypotheses, conducting literature reviews, creating cloud-based data pipelines that allow secure and remote analysis of sensitive patient data, and assisting in manuscript preparation, thereby potentially accelerating the pace of scientific discovery while maintaining methodological rigor[39–42].

Challenges for AI in rheumatic disease care

Significant challenges remain across multiple domains for the successful implementation of generative AI in advancing clinical care, that are briefly alluded to here.

Clinical validation is a particular challenge given the rarity of many rheumatic diseases; limited patient populations often make it difficult to create robust trials and generate sufficiently large validation datasets for AI systems[43]. The heterogeneity in study designs and outcome measures has been further complicated by inconsistent reporting standards, prompting the recent development of LLM-specific reporting guidelines to improve research transparency and reproducibility[44]. To address performance challenges, generative AI approaches can include fine-tuning methods for domain adaptation, synthetic data generation to expand limited datasets, or alternative architectures such as diffusion or adversarial networks[27,42]. Of particular relevance for clinical applications is retrieval-augmented generation (RAG), which integrates external knowledge sources (e.g., clinical guidelines or verified resources) into model outputs to improve factual accuracy and mitigate hallucinations - representing a potential pathway toward safer and more reliable use in medicine[45].

Model interpretability and explainability remain critical barriers, as rheumatologists may require transparent reasoning processes to maintain clinical confidence and ensure appropriate care, particularly given the complexity of multi-organ diseases where treatment decisions must account for numerous interacting factors including disease activity, medication interactions, and patient preferences[46,47]. While studies have demonstrated that LLM-generated explanations may sometimes represent post-hoc explanations rather than an actual account of the model's internal reasoning, they may still offer value: flaws in these explanations, such as contradictions, omissions, or unsupported logic, can correlate with inaccurate or biased outputs and provide a flag for clinician oversight[46]. However, evidence remains preliminary and mixed, with some clinician–machine learning collaboration experiments noting that explanations accompanying incorrect recommendations can heighten over-reliance and reduce clinician accuracy[48–50].

Data integration challenges may be particularly pronounced in rheumatology, where patient care often involves multiple subspecialists, fragmented electronic health records, and longitudinal disease monitoring that spans decades[46,51]. Ensuring secure, interoperable data sharing requires robust technical infrastructure and standardized data formats that many healthcare systems currently lack – particularly relevant given the heterogeneous laboratory and serologic testing prevalent for autoimmune diseases[52]. Regulatory oversight of non-deterministic AI systems remains nascent, with current frameworks emerging to address the unique challenges of generative models that produce variable outputs - particularly concerning when applied to high-stakes clinical decisions regarding immunosuppressive therapies[53].

Finally, successful implementation ideally requires fluid integration into existing clinical workflows and training programs to ensure rheumatic disease providers can effectively interpret and utilize AI-generated outputs[45,47].

Conclusion

Generative AI does not substitute for the human expertise that arises from years of prior experience and personal interaction with patients. However, collectively, these advances position generative AI to potentially transform rheumatologic practice toward greater precision, efficiency, and patient-centered care - provided its integration proceeds with rigorous validation, transparency, and equitable oversight.

Data availability

No datasets were generated or analysed during the current study.

Arjun Mahajan[1], Avery H. LaChance[1,2], Adam Rodman[1,3], Lorinda Chung[4,6] & Dylan Powell[5,6] ✉

[1]Harvard Medical School, Boston, MA, USA. [2]Brigham and Women's Hospital, Boston, MA, USA. [3]Beth Israel Deaconess Medical Center, Boston, MA, USA. [4]Department of Rheumatology, Stanford University School of Medicine, Palo Alto, CA, USA. [5]Faculty of Health Sciences & Sport, University of Stirling, Stirling, UK. [6]These authors contributed equally: Lorinda Chung, Dylan Powell. ✉e-mail: dylan.powell@stir.ac.uk

Received: 23 August 2025; Accepted: 18 September 2025;
Published online: 24 October 2025

References

1. Conrad, N. et al. Incidence, prevalence, and co-occurrence of autoimmune disorders over time and by age, sex, and socioeconomic status: a population-based cohort study of 22 million individuals in the UK. *Lancet* **401**, 1878–1890 (2023).
2. Miller, F. W. The increasing prevalence of autoimmunity and autoimmune diseases: an urgent call to action for improved understanding, diagnosis, treatment, and prevention. *Curr. Opin. Immunol.* **80**, 102266 (2023).
3. Xian, S. et al. Transformer patient embedding using electronic health records enables patient stratification and progression analysis. *npj Digit. Med.* **8**, 521 (2025).
4. Papanastasiou, G. et al. Large scale causal modeling to identify adults at risk for combined and common variable immunodeficiencies. *npj Digit. Med.* **8**, 361 (2025).
5. Maarseveen, T. D. et al. Improving musculoskeletal care with AI enhanced triage through data driven screening of referral letters. *npj Digit. Med.* **8**, 98 (2025).
6. Dubey, S., Chan, A., Adebajo, A. O., Walker, D. & Bukhari, M. Artificial intelligence and machine learning in rheumatology. *Rheumatol. (Oxf.)* **63**, 2040–2041 (2024).
7. Thirunavukarasu, A. J. et al. Large language models in medicine. *Nat. Med* **29**, 1930–1940 (2023).
8. Cho, J. H. & Feldman, M. Heterogeneity of autoimmune diseases: pathophysiologic insights from genetics and implications for new therapies. *Nat. Med* **21**, 730–738 (2015).
9. Sharma, S. D. & Bluett, J. Towards Personalized Medicine in Rheumatoid Arthritis. *Open Access Rheumatol.* **16**, 89–114 (2024).
10. Chan A. Current applications and future roles of AI in rheumatology. *Eur. Med. J.* (2024). Available at: https://www.emjreviews.com/rheumatology/article/current-applications-and-future-roles-of-ai-in-rheumatology-j170123/ (accessed 18 Aug 2025).
11. Gräf, M. et al. Comparison of physician and artificial intelligence-based symptom checker diagnostic accuracy. *Rheumatol. Int.* **42**, 2167–2176 (2022).
12. Horiuchi, D. et al. ChatGPT's diagnostic performance based on textual vs. visual information compared to radiologists' diagnostic performance in musculoskeletal radiology. *Eur. Radio.* **35**, 506–516 (2025).
13. do Olmo, J., Logroño, J., Mascías, C., Martínez, M. & Isla, J. Assessing DxGPT: Diagnosing rare diseases with various large language models. Preprint at https://doi.org/10.1101/2024.05.08.24307062 (2024).
14. Buckley, T. A., Crowe, B., Abdulnour, R. E., Rodman, A. & Manrai, A. K. Comparison of Frontier Open-Source and Proprietary Large Language Models for Complex Diagnoses. *JAMA Health Forum* **6**, e250040 (2025).
15. Li, Q. et al. Human-multimodal deep learning collaboration in 'precise' diagnosis of lupus erythematosus subtypes and similar skin diseases. *J. Eur. Acad. Dermatol Venereol.* **38**, 2268–2279 (2024).
16. Lee, S. et al. Clinically Applicable Deep Learning Framework for Measurement of the Extent of Hair Loss in Patients With Alopecia Areata. *JAMA Dermatol* **156**, 1018–1020 (2020).
17. Coskun, B. N., Yagiz, B., Ocakoglu, G., Dalkilic, E. & Pehlivan, Y. Assessing the accuracy and completeness of artificial intelligence language models in providing information on methotrexate use. *Rheumatol. Int* **44**, 509–515 (2024).
18. Goh, E. et al. GPT-4 assistance for improvement of physician performance on patient care tasks: a randomized controlled trial. *Nat. Med* **31**, 1233–1238 (2025).
19. Sinsky, C. et al. Allocation of Physician Time in Ambulatory Practice: A Time and Motion Study in 4 Specialties. *Ann. Intern Med* **165**, 753–760 (2016).
20. Young, R. A., Burge, S. K., Kumar, K. A., Wilson, J. M. & Ortiz, D. F. A Time-Motion Study of Primary Care Physicians' Work in the Electronic Health Record Era. *Fam. Med* **50**, 91–99 (2018).
21. Kheirkhah, H. et al. An Overview of Reviews to Inform Organization-Level Interventions to Address Burnout in Rheumatologists. *J. Rheumatol.* **50**, 1488–1502 (2023).
22. Williams, C. Y. K. et al. Physician- and Large Language Model-Generated Hospital Discharge Summaries. *JAMA Intern Med* **185**, 818–825 (2025).
23. Chua, C. E. et al. Integration of customised LLM for discharge summary generation in real-world clinical settings: a pilot study on RUSSELL GPT. *Lancet Reg. Health West Pac.* **51**, 101211 (2024).
24. Asgari, E. et al. A framework to assess clinical safety and hallucination rates of LLMs for medical text summarisation. *npj Digit. Med.* **8**, 274 (2025).
25. Liu, S. et al. Using large language model to guide patients to create efficient and comprehensive clinical care message. *J. Am. Med Inf. Assoc.* **31**, 1665–1670 (2024).
26. Ye, C., Zweck, E., Ma, Z., Smith, J. & Katz, S. Doctor Versus Artificial Intelligence: Patient and Physician Evaluation of Large Language Model Responses to Rheumatology Patient Questions in a Cross-Sectional Study. *Arthritis Rheumatol.* **76**, 479–484 (2024).
27. Busch, F. et al. Current applications and challenges in large language models for patient care: a systematic review. *Commun. Med* **5**, 26 (2025).
28. Du, K. et al. Comparing Artificial Intelligence-Generated and Clinician-Created Personalized Self-Management Guidance for Patients With Knee Osteoarthritis: Blinded Observational Study. *J. Med Internet Res* **27**, e67830 (2025).
29. Vatsal, S., Singh, A. & Tafreshi, S. Can GPT Improve the State of Prior Authorization via Guideline Based Automated Question Answering? Preprint at https://arxiv.org/abs/2402.18419 (2024).
30. Pandey, H., Amod, A. & Shivang. Advancing Healthcare Automation: Multi-Agent System for Medical Necessity Justification. Preprint at https://doi.org/10.48550/arXiv.2404.17977 (2024).
31. Shah, S. J. et al. Physician Perspectives on Ambient AI Scribes. *JAMA Netw. Open* **8**, e251904 (2025).
32. Stults, C. D. et al. Evaluation of an Ambient Artificial Intelligence Documentation Platform for Clinicians. *JAMA Netw. Open* **8**, e258614 (2025).
33. Zhang, Y. et al. Application of Computational Biology and Artificial Intelligence in Drug Design. *Int J. Mol. Sci.* **23**, 13568 (2022).
34. Yan, C. et al. Leveraging generative AI to prioritize drug repurposing candidates for Alzheimer's disease with real-world clinical validation. *npj Digit. Med.* **7**, 46 (2024).
35. Shen, M. D., Chen, S. B. & Ding, X. D. The effectiveness of digital twins in promoting precision health across the entire population: a systematic review. *NPJ Digit Med* **7**, 145 (2024).
36. Zerrouk, N., Augé, F. & Niarakis, A. Building a modular and multi-cellular virtual twin of the synovial joint in Rheumatoid Arthritis. *npj Digit. Med.* **7**, 379 (2024).
37. Bordukova, M., Makarov, N., Rodriguez-Esteban, R., Schmich, F. & Menden, M. P. Generative artificial intelligence empowers digital twins in drug discovery and clinical trials. *Expert Opin. Drug Discov.* **19**, 33–42 (2024).

38. Ibrahim, M. et al. Generative AI for synthetic data across multiple medical modalities: A systematic review of recent developments and challenges. *Comput Biol. Med* **189**, 109834 (2025).
39. Qi, B. et al. Large Language Models as Biomedical Hypothesis Generators: A Comprehensive Evaluation. Preprint at https://doi.org/10.48550/arXiv.2407.08940 (2024).
40. Kokash, N. et al. Ontology- and LLM-based Data Harmonization for Federated Learning in Healthcare. Preprint at https://doi.org/10.48550/arXiv.2505.20020 (2025).
41. Kobak, D., González-Márquez, R., Horvát, E. Á. & Lause, J. Delving into LLM-assisted writing in biomedical publications through excess vocabulary. *Sci. Adv.* **11**, eadt3813 (2025).
42. Reis, F., Lenz, C., Gossen, M., Volk, H. D. & Drzeniek, N. M. Practical Applications of Large Language Models for Health Care Professionals and Scientists. *JMIR Med Inf.* **12**, e58478 (2024).
43. Yang, Y. et al. Artificial intelligence for predicting treatment responses in autoimmune rheumatic diseases: advancements, challenges, and future perspectives. *Front Immunol.* **15**, 1477130 (2024).
44. Gallifant, J. et al. The TRIPOD-LLM reporting guideline for studies using large language models. *Nat. Med* **31**, 60–69 (2025).
45. Ke, Y. H. et al. Retrieval augmented generation for 10 large language models and its generalizability in assessing medical fitness. *npj Digit. Med.* **8**, 187 (2025).
46. Venerito, V. Artificial intelligence in rheumatology: days of a future past. *Rheumatol. Adv. Pr.* **9**, rkaf022 (2025).
47. Mahajan, A. et al. Cognitive bias in clinical large language models. *npj Digit. Med.* **8**, 428 (2025).
48. Gilbert, S. & Kather, J. N. Guardrails for the use of generalist AI in cancer care. *Nat. Rev. Cancer* **24**, 357–358 (2024).
49. Balagopalan, A. et al. The road to explainability is paved with bias: measuring the fairness of explanations. In *2022 ACM Conference on Fairness, Accountability, and Transparency* 1194–1206 (ACM, 2022).
50. Jacobs, M. et al. How machine-learning recommendations influence clinician treatment selections: the example of antidepressant selection. *Transl. Psychiatry* **11**, 108 (2021).
51. McMaster, C. et al. Artificial Intelligence and Deep Learning for Rheumatologists. *Arthritis Rheumatol.* **74**, 1893–1905 (2022).
52. Yoon, D. et al. Redefining Health Care Data Interoperability: Empirical Exploration of Large Language Models in Information Exchange. *J. Med Internet Res* **26**, e56614 (2024).
53. Meskó, B. & Topol, E. J. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *NPJ Digit Med* **6**, 120 (2023).

Author contributions
A.M. developed the concept and wrote the first draft and amended the final version. L.C., A.H.L., A.R. and D.P. provided oversight in drafting and editing of the manuscript. All authors read and approved the final manuscript.

Competing interests
The authors declare no competing interests. D.P. is News & Views editor at npj Digital Medicine but played no role in the internal review or decision to publish this News & Views piece.

Additional information
Correspondence and requests for materials should be addressed to Dylan Powell.

Reprints and permissions information is available at http://www.nature.com/reprints