American Journal of Epidemiology, 2025, **194**, 322–326
https://doi.org/10.1093/aje/kwae215
Advance access publication date July 16, 2024
Invited Commentary

Invited commentary: deep learning—methods to amplify epidemiologic data collection and analyses

D. Alex Quistberg*[,1,2], Stephen J. Mooney[3], Tolga Tasdizen[4,5], Pablo Arbelaez[6,7], Quynh C. Nguyen[8]

[1]Urban Health Collaborative, Dornsife School of Public Health, Drexel University, Philadelphia, PA 19104, United States
[2]Department of Environmental and Occupational Health, Dornsife School of Public Health, Drexel University, Philadelphia, PA 19104, United States
[3]Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA 98195, United States
[4]Department of Electrical and Computer Engineering, College of Engineering, University of Utah, Salt Lake City, UT 84112, United States
[5]The Scientific Computing and Imaging Institute, University of Utah, Salt Lake City, UT 84112, United States
[6]Department of Biomedical Engineering, Universidad de los Andes, Bogota 111711, Colombia
[7]Centro de Investigacion y Formacion en Inteligencia Artificial (CinfonIA), Universidad de los Andes, Bogota 111711, Colombia
[8]Department of Epidemiology and Biostatistics, School of Public Health, University of Maryland, College Park, MD 20742, United States

*Corresponding author: D. Alex Quistberg, Urban Health Collaborative, 3600 Market Street, 7th Floor, Philadelphia, PA 19104 (daq26@drexel.edu)

Abstract

Deep learning is a subfield of artificial intelligence and machine learning, based mostly on neural networks and often combined with attention algorithms, that has been used to detect and identify objects in text, audio, images, and video. Serghiou and Rough (*Am J Epidemiol*. 2023;192(11):1904-1916) presented a primer for epidemiologists on deep learning models. These models provide substantial opportunities for epidemiologists to expand and amplify their research in both data collection and analyses by increasing the geographic reach of studies, including more research subjects, and working with large or high-dimensional data. The tools for implementing deep learning methods are not as straightforward or ubiquitous for epidemiologists as traditional regression methods found in standard statistical software, but there are exciting opportunities for interdisciplinary collaboration with deep learning experts, just as epidemiologists have with statisticians, health care providers, urban planners, and other professionals. Despite the novelty of these methods, epidemiologic principles of assessing bias, study design, interpretation, and others still apply when implementing deep learning methods or assessing the findings of studies that have used them.

Key words: artificial intelligence; deep learning; neural networks; epidemiologic methods; data collection; data analysis; computer vision.

***Editor's note:** The opinions expressed in this article are those of the authors and do not necessarily reflect the views of the* American Journal of Epidemiology.

Artificial intelligence via deep learning models underlies many of our daily technological tools and online social interactions, encompassing activities such as object identification in images and videos, speech recognition, text analysis and search, among many others, and is beginning to enter the biomedical and public health research arenas.[1-6] Deep learning has also been used for identifying health conditions from radiological images and genomic data,[7-12] extracting data from clinical charts and notes,[13-15] identifying racist content on social media,[16,17] building chatbots to deliver health information,[18,19] and classifying the built and social environments.[20-26] In this issue, Serghiou and Rough[27] provide a timely primer on the fundamentals of deep learning for epidemiologic researchers, focused on describing the mathematical and statistical basis for these methods. In this commentary, we focus on the role deep learning could take in epidemiology, how deep learning could be useful to epidemiologists, and how epidemiologists should approach these methods.

Deep learning presents a potentially powerful means to speed and expand some traditional data collection methods used in epidemiology by replacing some of the human labor required for extraction from qualitative data such as patient charts,[14] free responses from surveys and interviews,[28] social media,[17] neighborhood measures from in-person and virtual audits or GIS (geographic information system),[21,25] and audio and video recordings,[29-31] although it also, of course, presents other challenges and issues. Studies using these data-collection methods are frequently limited by the time and effort required for data extraction, as well as the geographic area they can cover.[20,25] For example, suppose a researcher wants to examine the relationship between some built environment feature, like sidewalk conditions, and some health outcome, like older adult falls. In that case, they will need spatially precise measures of sidewalk conditions.[32,33] Obtaining such measures often requires manual collection efforts that can be time-consuming and costly, including examining maps, images, or other sources to identify and quantify the sidewalk conditions. Similar efforts could be expended with deep learning by using the collected data as training data for neural networks, allowing researchers to automatically quantify the presence and condition of sidewalks more efficiently in a much larger geographic area using archived imagery such as Google Street View or Mapillary.[25]

Received: February 1, 2023. **Accepted:** July 12, 2024
© The Author(s) 2024. Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

Downloaded from https://academic.oup.com/aje/article/194/2/322/7714789 by guest on 02 June 2026

The potential to leverage such efficiency, as described by Serghiou and Rough,[27] relies on the training data from which neural networks can learn to identify or detect the object of interest (eg, sidewalk conditions). These training data are usually developed by trained human raters who identify the objects, features, or words of interest in the raw data. It follows that future-thinking epidemiologists might choose to plan and design present extraction protocols to prepare for the results to be used as training data, either in their current research or for future use. Even if no such steps can be taken, protocols previously relying on human labor could use such labor only to collect training data, allowing deep learning to replace much of the originally planned manual collection, potentially expanding the study to a larger geographic area or population. Just as with any statistical analysis, however, deep learning approaches also require training data, as well as rigorous development of artificial intelligence models through training, testing, and validation that may also require substantial time and costs.

Deep learning can also contribute to or supplement and augment current and traditional statistical modeling approaches and data analyses used by epidemiologists, particularly in cases with large or high-dimensional data.[34,35] Some recent evidence suggests that deep learning models may perform better than traditional models for creating propensity scores,[36] prediction for screening and diagnostic instruments,[37] and for causal inference in observational studies,[38-40] whereas for multiple imputation there are conflicting findings.[41-44] There are also efforts and possibilities to use deep learning for predicting individual event censoring time,[45] analyzing count data,[46,47] improving nutritional epidemiology models,[48] and estimating population prevalence or risk of diseases and mortality.[49,50] Such approaches, essentially considering deep learning as a tool that exploits computational power to wring every bit of information from existing data, are simultaneously exciting and warrant caution regarding the overfitting of those models to their training data.[51] Additionally, as these approaches are still novel and have not been applied extensively in epidemiologic and public health research, the extent of their limitations may not yet be fully understood.

Deep learning may also unlock new opportunities for cross-site data sharing.[52] For example, deep learning-driven natural language processing and image recognition algorithms could be used with potentially personally identifying electronic health record data, including clinical notes or diagnostic images, to generate nonidentifying data abstracts that could be shared with researchers not approved to see the original, identifying data. In principle, such approaches could dramatically lower the burden of the assembly of pooled datasets, significantly increasing the scope of data collaboratives.[53,54] However, novel approaches to data sharing also raise novel methodological concerns and may deepen or complicate existing concerns. As another example, suppose a natural language processing algorithm that enables cross-site sharing of clinical notes identifies clinical conditions better on sites that use a particular electronic health record system or point-of-care note-taking tool. Such a scenario might induce selection bias (eg, the algorithm is used to identify cases only and not controls) or other systematic measurement error (eg, the algorithm is used to code exposure or other variables of interest). Researchers must be aware of this risk before launching the study to gather enough training data on sites other than the site on which the algorithm was trained if they hope to estimate the error and put bounds on the bias.

As briefly mentioned by Serghiou and Rough,[27] there are novel approaches in the field that readers should be aware of, particularly in generative pretrained transformer models (also known as foundation models). Transformer models have been a significant recent advance in artificial intelligence. They have gained substantial public awareness, especially during the past year with publicly available generative text (eg, ChatGPT), image (eg, Midjourney or DALL-E), video (eg, Make-A-Video), programming code (eg, Github Copilot), and other models.[55,56] The transformer model was first described by researchers from Alphabet[57] for natural language–processing models and later extended to computer vision by researchers at Facebook Artificial Intelligence Research.[58] Unlike traditional neural networks, transformer models can learn valuable representations in unsupervised scenarios (ie, they do not require the research team to train the model to recognize a particular outcome) but typically require extremely large datasets (eg, billions of data elements or parameters as opposed to millions in the largest traditional neural networks) to outperform traditional neural networks. Training a transformer from scratch, however, is not always necessary (let alone feasible) for researchers interested in using these models; rather, existing trained models can be combined with new data of interest to create models more specific to the research question or task of interest. These models or traditional neural networks can also be enhanced with fine-tuning, such as using pretraining and self-supervised learning.[59,60]

How should epidemiologists evaluate evidence from studies using deep learning and what should those articles include? As with all epidemiologic studies, issues of validity, reliability, and bias remain paramount. To allow readers to assess these threats, studies incorporating deep learning should describe the training data, the models used, model parameters, how the model was trained, any fine-tuning steps to improve the models, model performance, and ethical aspects.[61] Describing the training data is extremely important, as much of the bias and limitations from models may originate from them. This should include how the training data were collected (eg, sampling methods used), the reliability and validity of those data (eg, interrater reliability or other quality control statistics), and making training data available, if possible, for others to examine and use. Sharing the models is a very common practice among researchers in deep learning, and if possible, sharing model weights is another consideration so the findings can be replicated and applied to other datasets. Several online communities, such as Github, HuggingFace.co, and Deepai.org, enable researchers to share code, models, and data. Shared, common data resources for the epidemiologic community could help improve the rigor and generalizability of these models as their use grows. Providing open access to the data and models from epidemiologic studies could enable replicability, as well as provide some of the resources needed to make effective use of artificial intelligence approaches. For example, existing benchmark datasets widely used already for deep learning are ImageNet, COCO (common objects in context), PaLM, LLaMA-2, Mapillary Vistas, but few of these may meet the needs of epidemiologic research, particularly for more specialized topics.[62] The quality of these existing datasets also varies substantially and may not meet the reliability needed for epidemiologic research, especially for complicated or specific tasks, such as detecting and identifying small objects of the built environment, food and nutrition, racism and discrimination, medical images, and other video, audio, or text data relevant to public health and epidemiology. Epidemiologic research could greatly benefit from a set of curated models and training datasets that have been validated, with high reliability.

Downloaded from https://academic.oup.com/aje/article/194/2/322/7714789 by guest on 02 June 2026

Just as epidemiologists collaborate with statisticians, health care professionals, and other research professionals, those seeking to use deep learning in their research and practice should seek out partners and collaborators who are experts in deep learning to ensure fidelity and appropriate implementation and interpretation in epidemiologic research. Just as with other disciplines, initial collaborations can be challenging as we seek common language and methods, but are well worth the effort. We can expect many exciting and innovative uses of these methods in epidemiology and public health research and implementation, as well as misuses and ethical challenges, as they become more common and accessible. As with any new methods or technological advances adopted by the field, the underlying epidemiologic principles of interpretation, study design, bias, causality, and others will still apply.

Funding

This work was supported by the Fogarty International Center of the National Institutes of Health (grant K01TW011782 to D.A.Q.); by the National Library of Medicine (grants R00LM012868 to SJM and R01LM012849 to Q.C.N.); and by the National Institute on Minority Health and Health Disparities (grant R01MD016037 to Q.C.N.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of interest

We have no conflicts to report.

Data availability

Data was not used for this work.

References

1. Carin L, Pencina MJ. On deep learning for medical image analysis. *JAMA*. 2018;320(11):1192-1193. https://doi.org/10.1001/jama.2018.13316
2. Hinton G. Deep learning—a technology with the potential to transform health care. *JAMA*. 2018;320(11):1101-1102. https://doi.org/10.1001/jama.2018.11100
3. Stead WW. Clinical implications and challenges of artificial intelligence and deep learning. *JAMA*. 2018;320(11):1107-1108. https://doi.org/10.1001/jama.2018.11029
4. Wang F, Casalino LP, Khullar D. Deep learning in medicine—promise, progress, and challenges. *JAMA Intern Med*. 2019;179(3):293-294. https://doi.org/10.1001/jamainternmed.2018.7117
5. Miotto R, Wang F, Wang S, et al. Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform*. 2018;19(6):1236-1246. https://doi.org/10.1093/bib/bbx044
6. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. *Nat Med*. 2019;25(1):24-29. https://doi.org/10.1038/s41591-018-0316-z
7. Winkler JK, Fink C, Toberer F, et al. Association between surgical skin markings in dermoscopic images and diagnostic performance of a deep learning convolutional neural network for melanoma recognition. *JAMA Dermatol*. 2019;155(10):1135-1141. https://doi.org/10.1001/jamadermatol.2019.1735
8. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA*. 2017;318(22):2199-2210. https://doi.org/10.1001/jama.2017.14585
9. AlDubayan SH, Conway JR, Camp SY, et al. Detection of pathogenic variants with germline genetic testing using deep learning vs standard methods in patients with prostate cancer and melanoma. *JAMA*. 2020;324(19):1957-1969. https://doi.org/10.1001/jama.2020.20457
10. Yoo H, Kim KH, Singh R, et al. Validation of a deep learning algorithm for the detection of malignant pulmonary nodules in chest radiographs. *JAMA Netw Open*. 2020;3(9):e2017135. https://doi.org/10.1001/jamanetworkopen.2020.17135
11. Pokaprakarn T, Prieto JC, Price JT, et al. AI estimation of gestational age from blind ultrasound sweeps in low-resource settings. *NEJM Evidence*. 2022;1(5):EVIDoa2100058. https://doi.org/10.1056/EVIDoa2100058
12. Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health*. 2019;1(6):e271-e297. https://doi.org/10.1016/S2589-7500(19)30123-2
13. Wu S, Roberts K, Datta S, et al. Deep learning in clinical natural language processing: a methodical review. *J Am Med Inform Assoc*. 2020;27(3):457-470. https://doi.org/10.1093/jamia/ocz200
14. Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records. *NPJ Digital Med*. 2018;1(1):18. https://doi.org/10.1038/s41746-018-0029-1
15. Obeid JS, Dahne J, Christensen S, et al. Identifying and predicting intentional self-harm in electronic health record clinical notes: deep learning approach. *JMIR Med Inform*. 2020;8(7):e17784. https://doi.org/10.2196/17784
16. Badjatiya P, Gupta S, Gupta M, et al. Deep learning for hate speech detection in tweets. *Proceedings of the 26th International Conference on World Wide Web Companion*. Perth, Australia: International World Wide Web Conferences Steering Committee; 2017:759-760.
17. Dadvar M, Eckert K. *Cyberbullying Detection in Social Networks Using Deep Learning Based Models*. Springer; 2020.
18. Kandpal P, Jasnani K, Raut R, et al. Contextual chatbot for healthcare purposes (using deep learning). *Presented at 2020 Fourth World Conference on smart trends in systems, security and sustainability (WorldS4), 27-28 July 2020*. IEEE; 2020:625-634.
19. Kurup G, Shetty SD. AI conversational chatbot for primary healthcare diagnosis using natural language processing and deep learning. In: Das AK, Nayak J, Naik B, et al, eds. *Computational Intelligence in Pattern Recognition Proceedings of CIPR 2021*. 2022:259-272. https://doi.org/10.1007/978-981-16-2543-5_22
20. Gebru T, Krause J, Wang Y, et al. Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. *Proc Natl Acad Sci*. 2017;114(50):13108-13113. https://doi.org/10.1073/pnas.1700035114
21. Keralis JM, Javanmardi M, Khanna S, et al. Health and the built environment in United States cities: measuring associations using Google Street View–derived indicators of the built environment. *BMC Public Health*. 2020;20(1):215. https://doi.org/10.1186/s12889-020-8300-1
22. Li X, Zhang C, Li W, et al. Assessing street-level urban greenery using Google Street View and a modified green view index. *Urban For Urban Green*. 2015;14(3):675-685. https://doi.org/10.1016/j.ufug.2015.06.006
23. Maharana A, Nsoesie E. Use of deep learning to examine the association of the built environment with prevalence of neighborhood adult obesity. *JAMA Netw Open*. 2018;1(4):e181535. https://doi.org/10.1001/jamanetworkopen.2018.1535

Downloaded from https://academic.oup.com/aje/article/194/2/322/7714789 by guest on 02 June 2026

24. Naik N, Kominers SD, Raskar R, et al. Computer vision uncovers predictors of physical urban change. *Proc Natl Acad Sci U S A*. 2017;114(29):7571-7576. https://doi.org/10.1073/pnas.1619003114
25. Nguyen QC, Sajjadi M, McCullough M, et al. Neighbourhood looking glass: 360 automated characterisation of the built environment for neighbourhood effects research. *J Epidemiol Community Health*. 2018;72(3):260-266. https://doi.org/10.1136/jech-2017-209456
26. Seltenrich N. Remote-sensing applications for environmental health research. *Environ Health Perspect*. 2014;122(10):A268-A275. https://doi.org/10.1289/ehp.122-A268
27. Serghiou S, Rough K. Deep learning for epidemiologists: an introduction to neural networks. *Am J Epidemiol*. 2023;192(11):1904-1916. https://doi.org/10.1093/aje/kwad107
28. Leeson W, Resnick A, Alexander D, et al. Natural language processing (NLP) in qualitative public health research: a proof of concept study. *Int J Qual Methods*. 2019;18:160940691988702. https://doi.org/10.1177/1609406919887021
29. Muzammel M, Salam H, Othmani A. End-to-end multimodal clinical depression recognition using deep neural networks: a comparative analysis. *Comput Methods Programs Biomed*. 2021;211:106433. https://doi.org/10.1016/j.cmpb.2021.106433
30. Roshanzamir A, Aghajan H, Soleymani BM. Transformer-based deep neural network language models for Alzheimer's disease risk assessment from targeted speech. *BMC Med Inform Decis Mak*. 2021;21(1):92. https://doi.org/10.1186/s12911-021-01456-3
31. Kidziński Ł, Yang B, Hicks JL, et al. Deep neural networks enable quantitative movement analysis using single-camera videos. *Nat Commun*. 2020;11(1):4054. https://doi.org/10.1038/s41467-020-17807-z
32. Plascak JJ, Rundle AG, Babel RA, et al. Drop-and-spin virtual neighborhood auditing: assessing built environment for linkage to health studies. *Am J Prev Med*. 2020;58(1):152-160. https://doi.org/10.1016/j.amepre.2019.08.032
33. Aghaabbasi M, Moeinaddini M, Zaly Shah M, et al. A new assessment model to evaluate the microscale sidewalk design factors at the neighbourhood level. *J Transp Health*. 2017;5:97-112. https://doi.org/10.1016/j.jth.2016.08.012
34. Jorm LR. Commentary: towards machine learning-enabled epidemiology. *Int J Epidemiol*. 2021;49(6):1770-1773. https://doi.org/10.1093/ije/dyaa242
35. Najafabadi MM, Villanustre F, Khoshgoftaar TM, et al. Deep learning applications and challenges in big data analytics. *J Big Data*. 2015;2(1):7. https://doi.org/10.1186/s40537-014-0007-7
36. Whata A, Chimedza C. Evaluating uses of deep learning methods for causal inference. *IEEE Access*. 2022;10:2813-2827. https://doi.org/10.1109/ACCESS.2021.3140189
37. Tomašev N, Harris N, Baur S, et al. Use of deep learning to develop continuous-risk models for adverse event prediction from electronic health records. *Nat Protoc*. 2021;16(6):2765-2787. https://doi.org/10.1038/s41596-021-00513-5
38. Luo Y, Peng J, Ma J. When causal inference meets deep learning. *Nat Mach Intell*. 2020;2(8):426-427. https://doi.org/10.1038/s42256-020-0218-x
39. Rao S, Mamouei M, Salimi-Khorshidi G, et al. Targeted-BEHRT: deep learning for observational causal inference on longitudinal electronic health records. *arXiv*. 2022.
40. Blakely T, Lynch J, Simons K, et al. Reflection on modern methods: when worlds collide—prediction, machine learning and causal inference. *Int J Epidemiol*. 2021;49(6):2058-2064. https://doi.org/10.1093/ije/dyz132
41. de Jong J, Emon MA, Wu P, et al. Deep learning for clustering of multivariate clinical patient trajectories with missing values. *GigaScience*. 2019;8(11):giz134. https://doi.org/10.1093/gigascience/giz134
42. Getz K, Hubbard RA, Linn KA. Performance of multiple imputation using modern machine learning methods in electronic health records data. *Epidemiology*. 2023;34(2):206-215. https://doi.org/10.1097/EDE.0000000000001578
43. Kim J-S, Gao X, Rzhetsky A. RIDDLE: Race and ethnicity Imputation from Disease history with Deep LEarning. *PLoS Comput Biol*. 2018;14(4):e1006106. https://doi.org/10.1371/journal.pcbi.1006106
44. Wang Z, Akande O, Poulos J, et al. Are deep learning models superior for missing data imputation in surveys? Evidence from an empirical comparison. *Surv Methodol*. 2022;48(2):375-399. https://www150.statcan.gc.ca/n1/pub/12-001-x/2022002/article/00009-eng.htm
45. Jeong J-H, Jia Y. CausalDeepCENT: deep learning for causal prediction of individual event times. *arXiv*. https://doi.org/10.48550/arXiv.2203.10207, March 19, 2022, preprint: not peer reviewed.
46. Montesinos-Lopez OA, Montesinos-Lopez JC, Salazar E, et al. Application of a Poisson deep neural network model for the prediction of count data in genome-based prediction. *Plant Genome*. 2021;14(3):e20118. https://doi.org/10.1002/tpg2.20118
47. Montesinos-López OA, Montesinos-López JC, Singh P, et al. A multivariate Poisson deep learning model for genomic prediction of count data. *G3*. 2020;10(11):4177-4190. https://doi.org/10.1534/g3.120.401631
48. Morgenstern JD, Rosella LC, Costa AP, et al. Perspective: big data and machine learning could help advance nutritional epidemiology. *Adv Nutri*. 2021;12(3):621-631. https://doi.org/10.1093/advances/nmaa183
49. Bej S, Sarkar J, Biswas S, et al. Identification and epidemiological characterization of type-2 diabetes sub-population using an unsupervised machine learning approach. *Nutr Diabetes*. 2022;12(1):27. https://doi.org/10.1038/s41387-022-00206-2
50. Weng SF, Vaz L, Qureshi N, et al. Prediction of premature all-cause mortality: a prospective general population cohort study comparing machine-learning and standard epidemiological approaches. *PloS One*. 2019;14(3):e0214365. https://doi.org/10.1371/journal.pone.0214365
51. Mooney SJ, Keil AP, Westreich DJ. Thirteen questions about using machine learning in causal research (you won't believe the answer to number 10!). *Am J Epidemiol*. 2021;190(8):1476-1482. https://doi.org/10.1093/aje/kwab047
52. Ha YJ, Lee G, Yoo M, et al. Feasibility study of multi-site split learning for privacy-preserving medical systems under data imbalance constraints in COVID-19, X-ray, and cholesterol dataset. *Sci Rep*. 2022;12(1):1534. https://doi.org/10.1038/s41598-022-05615-y
53. Festag S, Spreckelsen C. Privacy-preserving deep learning for the detection of protected health information in real-world data: comparative evaluation. *JMIR Form Res*. 2020;4(5):e14064. https://doi.org/10.2196/14064
54. Jin H, Luo Y, Li P, et al. A review of secure and privacy-preserving medical data sharing. *IEEE Access*. 2019;7:61656-61669. https://doi.org/10.1109/ACCESS.2019.2916503
55. Bommasani R, Hudson DA, Adeli E, et al. On the opportunities and risks of foundation models. *arXiv*. https://doi.org/10.48550/arXiv.2108.07258, July 12, 2022, preprint: not peer reviewed.
56. Sevilla J, Heim L, Ho A, et al. Compute trends across three eras of machine learning. *arXiv*. February 11, 2022, preprint:

Downloaded from https://academic.oup.com/aje/article/194/2/322/7714789 by guest on 02 June 2026

not peer reviewed. https://doi.org/10.1109/IJCNN55064.2022.9891914
57. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. *arXiv*. August 2, 2023, preprint: not peer reviewed. https://doi.org/10.1109/IJCNN55064.2022.9891914
58. Carion N, Massa F, Synnaeve G, et al. *End-to-End Object Detection With Transformers*. Springer; 2020.
59. Zoph B, Ghiasi G, Lin T-Y, et al. Rethinking pre-training and self-training. In: Larochelle H, Ranzato M, Hadsell R, et al, eds. *Proceedings of the 34th International Conference on Neural Information Processing Systems*. Curran Associates Inc; 2020. p. 3833-3845.
60. Reed CJ, Yue X, Nrusimha A, et al. Self-supervised pretraining improves self-supervised pretraining. *Presented at 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, January 3-8, 2022. IEEE; 2022.
61. Mitchell M, Wu S, Zaldivar A, et al. Model cards for model reporting. *FAT∗ '19: Proceedings of the Conference on Fairness, Accountability, and Transparency*. ACM Digital Library; 2019.
62. Blagec K, Kraiger J, Frühwirt W, et al. Benchmark datasets driving artificial intelligence development fail to capture the needs of medical professionals. *J Biomed Inform*. 2023;137:104274. https://doi.org/10.1016/j.jbi.2022.104274

Downloaded from https://academic.oup.com/aje/article/194/2/322/7714789 by guest on 02 June 2026

© The Author(s) 2024. Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.
American Journal of Epidemiology, 2025, **194**, 322–326
https://doi.org/10.1093/aje/kwae215
Invited Commentary