Houston Journal of Health Law & Policy

ERRORS IN THE EMR:

UNDER-RECOGNIZED HAZARD FOR AI IN HEALTHCARE

E. H. Morreim, JD, PhD*

* The author acknowledges with gratitude the very helpful comments provided on earlier versions of this paper by Douglas McNair, MD, PhD, Gordon Apple, JD, Jonathan Stewart, PhD, Jessica Mantel, JD, Leah Fowler, JD, Jessica Roberts, JD, Michael Ewer, MD.

I. OVERVIEW

Some of us are old enough to remember the hype and the hope as electronic medical records (EMRs) emerged. They will save time, reduce errors, and save lives! What a glorious future we anticipated, even while acknowledging the transition would have its bumps.

We're still waiting.[1]

Another, even more momentous development we likewise greet with hope and hype, but with greater recognition that its promise brings perils: artificial intelligence (AI).[2] In healthcare, AI's greatest progress to date has focused on improving interpretation of static, largely machine-generated images such as radiographs, electrocardiograms (EKGs) and pathology slides, to enhance diagnoses.[3] In contrast, many of AI's more recent ventures, geared more toward therapeutics and financial management than diagnostics, rely mainly on datasets heavily comprised of human observations recorded in EMRs and in derivative administrative databases such as insurance records.[4] This Article focuses mainly on the latter.

We already know that AI in healthcare faces significant challenges, such as to exacerbate racial and other biases and to perform poorly in unexpected ways. These issues are briefly summarized in Part II.

Less well recognized is another likely source of poor performance. We know that "[a]ny algorithm is only as good as the data on which it

---

1 Although EMRs have greatly reduced some of the problems inherent in hand-written records, e.g. drug errors due to indecipherable handwriting, physicians now spend significantly more time in documentation efforts, thereby exacerbating burnout. *See generally* Lisa Rosenbaum, *Transitional Chaos or Enduring Harm? The EHR and the Disruption of Medicine*, 17 NEW ENGL. J. MED. 373, 1585–88 (2015); Raj M. Ratwani et al., *Improving Electronic Health Record Usability and Safety Requires Transparency*, 24 JAMA 320, 2533–34 (2018); *see also* David C. Classen et al., *Inpatient EHR User Experience and Hospital EHR Safety Performance*, 6 JAMA NETWORK OPEN e2333152 (2023); Don Eugene Detmer & Andrew Gettinger, *Essential Electronic Health Record Reforms for This Decade*, 329 JAMA 1825, 1825-26 (2023).

2 NAT'L ACAD. MED., ARTIFICIAL INTELLIGENCE IN HEALTH CARE: THE HOPE, THE HYPE, THE PROMISE, THE PERIL 7–36 (Michael Matheny et al. eds., 2022), https://nam.edu/wp-content/uploads/2019/12/AI-in-Health-Care-PREPUB-FINAL.pdf.

3 *Id.*, at 145–179; Rajpurkar Pranav & Lungren Matthew, *The Current and Future State of AI Interpretation of Medical Images*, 388 N. ENGL. J. MED. 1981 (2023).

4 Rajpurkar & Lungren, *supra* note 3, at 1981-90.

is trained."[5] "In using black box clinical AI systems, physicians and hospitals 'place trust not only in the equation of the model, but also in the entire database used to train it and, in the handling (e.g., labelling) of that database by the designers.'"[6] Unfortunately, evidence suggests that the EMRs serving as the main dataset for therapeutics-focused AI are considerably more error-laden than we commonly realize.[7] Parts III and IV discus EMR errors. First, Part III describes *unintentional* errors, such as missing or mis-recording relevant history and physical information. As then explained in Part IV, *intentional* errors largely track payment methods, predominantly either fee-for-service (FFS) or capitation. In FFS, providers are paid for each service, hence must document their interventions in the EMR—with the obvious temptation to document worse symptoms than the patient actually has (e.g., to justify prior authorization of an otherwise non-covered intervention), or to document more than was actually done, thereby "gaming the system"[8] to garner larger payment.

Capitation provides a flat fee, usually per-patient-per-month, in exchange for all the care that patient needs.[9] Such payments are commonly risk-adjusted so that sicker patients (i.e., those with more diagnoses) bring higher compensation. Here, as will be seen in litigation

---

5 Glenn I. Cohen, *What Should ChatGPT Mean for Bioethics*?, 23 AMER. J. BIOETH. 8, 8–16 (2023).

6 Benny Chan, *Applying a Common Enterprise Theory of Liability to Clinical AI Systems*, 47 AM. J. LAW MED. 351, 364 (2021); Joan M. Teno, *Garbage in, Garbage out—Words of Caution on Big Data and Machine Learning in Medical Practice*, 4 JAMA HEALTH FORUM 1, 1 (2023); Prathit A. Kulkarni & Hardeep Singh, *Artificial Intelligence in Clinical Diagnosis: Opportunities, Challenges, and Hype*, 330(4) JAMA, 317–18 ("[T]he quality of diagnostic data in the EHR will directly determine the accuracy of the AI-generated diagnosis."); Stefan Harrer, *Attention Is Not All You Need: The Complicated Case of Ethically Using Large Language Models in Healthcare and Medicine*, 90 EBIOMEDICINE 1, 2 ("[I]f models have been trained on a vast corpus of internet data with limited filtering (as for example is the case for ChatGPT or stable diffusion), they have ingested facts as much as misinformation, biased content as much as fair content, harmful materials as much as harmless ones. Without a means to assess any of these criteria before answering a prompt, LLMs are at risk—and numerous examples have demonstrated they have fallen to it—of reproducing, amplifying, and disseminating problematic content and misinformation.").

7 *See* text accompanying footnotes Parts III and IV.

8 *See* Haavi Morreim, *Gaming the System: Dodging the Rules, Ruling the Dodgers*, 151 AM. J. BIOETHICS 3, 443–47 (2012).

9 Sri Lekha Tummalapalli et al., *Capitated Versus Fee-For-Service Reimbursement and Quality of Chronic Disease: A US Cross-Sectional Analysis*, 22 BMS HEALTH SERVICES RESEARCH 1, 1–3 (2022).

against several Medicare Advantage plans, the temptation is to insert unwarranted, unsubstantiated diagnoses into patients' EMRs to improve risk adjustments and thereby enhance payments to the organization.

Errors like those discussed in this Article can potentially have a variety of adverse effects, such as to steer the resulting AI clinical decision support algorithms or AI-generated chart notes off course in ways that, per Part V, may be difficult to detect and even more difficult to correct. Part VI explores some further implications of EMR errors, such as the use of AI for insurance coverage or pre-authorization decisions.

II. AI: BASICS AND KNOWN DRAWBACKS

A. Basics

Artificial intelligence is defined as "a computer or machine simulating or imitating one or more forms of human intelligence."[10] Machine learning (ML), in turn, "permits a machine to teach itself, based on data provided, to solve problems that humans innately know how to solve but that are not easily broken down into discrete steps for a computer—such as recognizing human faces."[11]

To date, AI's accomplishments include interpretation of radiographic images, pathology slides, and signal processing such as EKGs and EEGs,[12] sometimes outperforming physicians of the relevant specialties.[13] In healthcare, machine learning (ML, a subset of AI) is often "supervised" rather than "unsupervised," as it discerns patterns from example cases labeled by humans as correct.[14] Large language models

---

10 Erin M. Eislein & Anna-Liisa Mullis, *Corporate Practice of Medicine Statutes in the Age of Artificial Intelligence*, 22 HEALTH L. CONNECTIONS 1, 2 (2021); *see also* NAT'L ACAD. MED., *supra* note 2, at 13–16.

11 Eislein & Mullis, *supra* note 10, at 2.

12 *See* NAT'L ACAD. MED., *supra* note 2.

13 Charlotte J. Haug & Jeffrey M. Drazen, *Artificial Intelligence and Machine Learning in Clinical Medicine*, 388 N. ENGL. J. MED. 1201, 1201–08 (2023).

14 *See, e.g.*, Jonathan H. Chen et al., *Decoding Artificial Intelligence to Achieve Diagnostic Excellence: Learning From Experts, Examples, and Experience*, 328 JAMA 709, 709–10 (2022) ("by exposing machine learning algorithms to thousands of retinal images that include cases of diabetic retinopathy labeled by ophthalmologists, these systems can make the diagnosis in

(LLMs), in turn, "are a specific subset of machine learning designed to understand and generate human language by enabling computers to convert language and unstructured text into machine-readable, organized data."[15]

Emerging uses for ML and LLM in healthcare are numerous. Providers and payors alike have begun using AI to streamline coding, claims management, payment and prior authorization decisions.[16] In the clinical setting, LLMs using voice recognition can write clinical encounter notes, potentially saving time and reducing provider burnout.[17] AI-generated Clinical Decision Support (CDS) tools can

---

future images without being told what to look for."); *see also, e.g.*, Melissa D. McCradden et al., *A Research Ethics Framework for the Clinical Translation of Healthcare Machine Learning*, 22 AM. J. BIOETHICS 8, 12 (2022) (noting that in supervised algorithm development, one portion of the dataset is then trained to establish the initial model, which is then tested on other portions of the dataset or on external data, often multiple times).

15 Scott Gottlieb & Lauren Silvis, *How to Safely Integrate Large Language Models Into Health Care*, 4 JAMA HEALTH FORUM e233909 (2023) ("By incorporating data from countless individual decisions, these models can mimic a person's responses by calculating the probability of each potential response and then selecting the most appropriate one—either by choosing the response with the highest likelihood of being correct or by sampling from a distribution of likely outcomes. A key difference between machine learning and LLMs is functionality in that machine learning devices are trained to perform specific tasks, whereas LLMs can understand and generate free-form text, potentially making them effective tools for expanding interactions with patients. For example, LLMs may generate prompts to obtain additional information about a patient's symptoms or response to therapy."); *see also* Nikhil R. Sahni & Brandon Carrus, *Artificial Intelligence in U.S. Health Care Delivery*, 389(4) N. ENGL. J. MED. 348, 349 (2023) ("Two types of AI have generally been pursued in health care delivery: machine learning, which involves computational techniques that learn from examples instead of operating from predefined rules, and natural language processing, which is the ability of a computer to transform human language and unstructured text into machine-readable structured data that reliably reflect the intent of the language".).

16 Victoria Bailey, *How Do Healthcare Organizations Feel About Autonomous Coding?*, REVCYCLE INTELLIGENCE, (Oct. 27, 2023), https://revcycleintelligence.com/news/how-do-healthcare-organizations-feel-about-autonomous-coding; *see also* Jacqueline LaPointe, *Revenue Cycle Sets Sights on Autonomous Coding, Claims Automation*, REVCYCLE INTELLIGENCE, (Nov. 9, 2023), https://revcycleintelligence.com/news/revenue-cycle-set-sights-on-autonomous-coding-claims-automation; *see also* Daniel Payne, *AI's Big Test: Making Sense of $4 Trillion in Medical Expenses*, POLITICO (Dec. 31, 2023), https://www.politico.com/news/2023/12/31/ai-medical-expenses-001325577; *see also* Michael Anne Kyle & Zirui Song, *The Consequences and Future of Prior-Authorization Reform*, 389 NEW ENG. J. MED. 291, 291–93 (2023); Sahni & Carrus, *supra* note 15, at 348–58 (AI functions also include "[c]onsumer, continuity of care, network and market insights, clinical operations, clinical analysis, quality and safety, value-based care, reimbursement, corporate functions.).

17 *See* Jeffrey E. Harris, *An AI-Enhanced Electronic Health Record Could Boost Primary Care Productivity*, 330 JAMA 801, 801–02 (2023); *see also* Mindy Duffourc et al., *Generative*

potentially provide a richer array of diagnostic possibilities than the clinician might produce,[18] identify which patients may soon need a higher level of care,[19] help surgeons to plan and perform surgeries more precisely,[20] improve pain care,[21] create personalized risk profiles,[22] or provide psychotherapy through conversational agents.[23]

B. Known drawbacks

Nevertheless, an "AI chasm" has emerged—a substantial gap between AI and clinically useful, reliable applications.[24] AI algorithms

---

Artificial Intelligence in Health Care and Liability Risks for Physicians and Safety Concerns for Patients, 330 JAMA 313, 313–14 (2023); *see also* Peter Lee et al., *Benefits, Limits, and Risks of GPT-4 as an Artificial Intelligence Chatbot for Medicine*, 388 N. ENGL. J. MED., 1233, 1233–39 (2023); *see also* Ashwin Nayak et al., *Comparison of History of Present Illness Summaries Generated by a Chatbot and Senior Internal Medicine Residents*, 183 JAMA 1026, 1026–27 (2023).

18 *See* Katherine E. Goodman et al., *Preparing Physicians for the Clinical Algorithm Era*, 389 N. ENGL. J. MED. 483, 483–87 (2023).

19 *Id.*; *see also* NAT'L ACAD. MED., *supra* note 2, at 73; *see also* Heather B. Deixler et al., *Shooting for the Moon: The Evolution of Key AI/ML Regulations Governing Certain Health Care Products and Services*, 4 HEALTH L. CONNECTIONS 1, 1–8 (2023) (Per proposed federal rules, "CDS" may be replaced by "decision support interventions" (DSI) to be regulated as devices in various settings.).

20 Daniel Schiff & Jason Borenstein, *How Should Clinicians Communicate With Patients About the Roles of Artificially Intelligent Team Members?*, 21 AMA J. OF ETHICS E138, E138–45 (2019); *see also* Jayakumar Prakash et al., *Comparison of an Artificial Intelligence–Enabled Patient Decision Aid vs Educational Material on Decision Quality, Shared Decision-Making, Patient Experience, and Functional Outcomes in Adults With Knee Osteoarthritis: A Randomized Clinical Trial*, 4 JAMA NETWORK OPEN e2037107 (2021); *see also* Dimitris Bertisimas et al., *Artificial Intelligence to Find the Optimal Margin Width in Hepatectomy for Colorectal Cancer Liver Metastases*, 157 JAMA SURGERY e221819 (2022).

21 John D. Piette et al., *Patient-Centered Pain Care Using Artificial Intelligence and Mobile Health Tools: A Randomized Comparative Effectiveness Trial*, 182 JAMA INTERNAL MED. 975, 975–83 (2022).

22 Jennifer Blumenthal-Barby et al., *Research on the Clinical Translation of Health Care Machine Learning: Ethicists Experiences on Lessons Learned,* 22(5) AMER. J. BIOETHICS 1, 1–3 (2022) (One such use assists in evaluating patients for LVADs (left ventricular assist devices).

23 John D. McGreevey III et al., *Clinical, Legal, and Ethical Aspects of Artificial Intelligence–Assisted Conversational Agents in Health Care*, 324 JAMA 552, 552–53 (2020).

24 INST. MED. et al., CROSSING THE QUALITY CHASM: A NEW HEALTH SYSTEM FOR THE 21ST CENTURY (2006); McCradden et al., *supra* note 14, at 8; Nicole Martinez-Martin & Mildred K. Cho, *Bridging the AI Chasm: Can EBM Address Representation and Fairness in Clinical Machine Learning?,* 22 AMER. J. BIOETHICS 30, 30–32 (2022); Jonathan H. Lu et al., *Assessment of Adherence to Reporting Guidelines by Commonly Used Clinical Prediction Models From a*

have proved vulnerable in ways not entirely anticipated at the outset. For instance, when IBM's Watson, which had performed brilliantly in complex tasks such as chess and "Jeopardy!" was applied to oncology, it recommended incorrect and unsafe cancer treatments.[25] Similarly, changes in demographics of the population to which an algorithm is applied, or shifts in physicians' practice habits, or variations in the prevalence of the condition being addressed, or differences in software from one otherwise-identical scanner to the next, or differences among the people who operate the hardware, can all cause AI performance to deteriorate.[26] Thus, "dataset shift" "can result when there is a mismatch between the data set with which an AI system was developed and the data on which it is subsequently deployed."[27]

Single Vendor A Systematic Review, 5 JAMA NETWORK OPEN e2227779 (2022); Kevin Roose, *We Need to Talk About How Good Artificial Intelligence Is Getting*, N.Y. TIMES (Aug. 24, 2022), https://www.nytimes.com/2022/08/24/technology/ai-technology-progress.html.

[25] Chan, *supra* note 6, at 352, citing Casey Ross & Ike Swetlitz, *IBM pitched its Watson supercomputer as a revolution in cancer care. It's nowhere close.*, STAT (Sept. 5, 2017), https://www.statnews.com/2017/09/05/watson-ibm-cancer/ (Watson for Oncology "recommended bevacizumab (Avastin) to a patient with evidence of severe bleeding, despite a clear contraindication and a warning from the FDA.")

[26] *See* Yun Liu et al., *How to Read Articles That Use Machine Learning*, 322 JAMA 1806, 1814 (2019); *see* Chan, *supra* note 6, 351–385; Andrew L. Beam et al., *Challenges to the Reproducibility of Machine Learning Models in Health Care*, 323 JAMA 305, 305–06 (2020); In one example, researchers "trained a neural network to diagnose pneumonia from patient radiographs in one hospital system and evaluated its diagnostic ability on external radiographs from different hospital systems, with their results showing that performance on external datasets was significantly degraded." NAT'L ACAD. MED., *supra* note 2, at 166 (citing J.R. Zech et al., *Variable Generalization Performance of a Deep Learning Model to Detect Pneumonia in Chest Radiographs: A Cross-Sectional Study*, 15 PLOS MEDICINE e1002683 (2018)); *see also* Andrew Wong et al., *External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients*, 181 JAMA INTERN. MED. 1065, 1065–70 (2021) (cited in *Blueprint for an AI Bill of Rights: Making automated systems work for the American People*, (Oct. 2022), https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf.).

[27] Andrew L. Beam et al., *Artificial Intelligence in Medicine*, 388 NEW ENGL. J. MED. 1220, 1221 (2023) (citing Samuel G. Finlayson et al., *The Clinician and Dataset Shift in Artificial Intelligence*, 385 N. ENGL. J. MED. 283, 283–86 (2021)); *see also* Andrew S. Lea et al., *Mind the Gap – Machine Learning, Dataset Shift, and History in the Age of Clinical Algorithms*, 390 NEW ENGL. J. MED. 293, 293 (2024); NAT'L ACAD. MED., *supra* note 2, at 166 (If a hospital uses lactate measurements to confirm suspicion of sepsis, an AI attuned to that practice pattern may deteriorate significantly for a population that includes higher numbers of drug users, whose lactate may be elevated.); *see also* Ravi B. Parikh et al., *Addressing Bias in Artificial Intelligence in Health Care*, 322 JAMA 2377, 2377–78 (2019); Rajpurkar & Lungren, *supra* note 3, at 1986.

Analogously, questions have arisen regarding how much and what sort of data will be available in the first place. Litigation has already arisen, for instance, surrounding the sharing of otherwise private data from patients' EMR with creators of AI content.[28] Analogously, litigation has arisen where AI companies allegedly infringed copyright law by relying on proprietary content for its dataset.[29] To the extent that copyright and privacy laws restrict AI access to a rich and representative dataset, the result for healthcare is likely to be distorted and, potentially, dangerously inaccurate.

Moreover, even as supervised machine learning requires training via correct examples of whatever phenomena the AI model is designed to assess, such systems need "massive data streams"[30] of "manually labeled training data"[31] that, at the same time, too often lack a *bona fide* "gold standard" for correctness.[32]

Additionally it is well recognized that, where patients' EMRs and insurance records provide datasets, any racial, gender, age or other

---

[28] Mindy Dufforc, *Health Care AI and Patient Privacy - Dinerstein v Google*, 331 JAMA 909, 909–10 (2024) (noting that although plaintiff patients' lawsuit was rejected, the issue is not likely dead).

[29] In *Andersen v. Stability AI Ltd.*, for instance, artists allege that Stability AI wrongly used billions of copyrighted images in developing its AI software. Anderson v. Stability AI Ltd., No. 23-CV-00201-WHO, 2023 WL 7132064 (N.D. Cal. Oct. 30, 2023); The case in *Getty Images (US) Inc. v. Stability AI Ltd.* is similar, but concerns photographic images rather than artist-created visual art. Complaint at ¶ 1, *Getty Images v. Stability AI LTD*, No. 1:23-cv-00135-UNA (D. Del. Feb. 3, 2023); In *Concord Music Group Inc. v. Anthropic PBC*, major music publishers similarly alleges that the company trained its AI on plaintiffs musical works. Complaint at ¶ 6, *Concord Music Group Inc. v. Anthropic PBC*, No. 3:23-cv-01092 (M.D. Tenn 2024); The same basic case is alleged by book authors in *Kadrey v. Meta Platforms Inc.* and in *Authors Guild v. Open AI Inc.*. Kadrey v. Meta Platforms, Inc., No. 23-CV-03417-VC, 2023 WL 8039640 (N.D. Cal. Nov. 20, 2023); Complaint at ¶ 1, *Authors Guild v. Open AI Inc.*, No. 1:23-cv-8292 (S.D. NY 2023).

[30] NAT'L ACAD. MED., *supra* note 2; David James Hunter & Christopher Holmes, *Where Medical Statistics Meets Artificial Intelligence*, 389 N. ENGL. J. MED. 1211, 1212 (2023) ("The checking of AI-supported findings is particularly important in the emerging field of generative AI through self-supervised learning, such as large language models and medical science chatbots that may be used, among many applications, for medical note taking in electronic health records. Self-supervised learning by these foundation models involves vast quantities of undocumented training data and the use of broad objective functions to train the models with trillions of parameters (at the time of this writing).").

[31] Chen et al., *supra* note 14.

[32] Adewole S. Adamson & H. Gilbert Welch, *Machine Learning and the Cancer-Diagnosis Problem — No Gold Standard*, 381 N. ENGL. J. MED. 2285, 2285–87 (2019).

social biases embedded therein not only can be replicated, but exacerbated in the resulting AI algorithm.[33] Such bias can arise from diverse sources.[34] The challenge is sufficiently formidable that the Department of Health and Human Services has proposed that Section 1557 of the Affordable Care Act, which prohibits discrimination, should encompass clinical algorithms.[35]

Moreover, AI algorithms can enter the market without meeting any uniform, validated criteria for accuracy and reliability. A recent study showed, for instance, that a variety of AI products for breast cancer screening were cleared by the FDA under the 510k pathway, yet did not share any consistent basis for "reporting of data sources, data

---

33 Nicole Martinez-Martin & Mildred K. Cho, *Bridging the AI Chasm: Can EBM Address Representation and Fairness in Clinical Machine Learning?*, 22 AMER. J. BIOETHICS 30, 30-32 (2022); Ashraf Fazy et al. *Racial and Ethnic Discrepancy in Pulse Oximetry and Delayed Identification of Treatment Eligibility Among Patients With COVID-19*, 182 JAMA INTERN MED. 730, 730–38 (2022); Christina Jewett, *Doctors Warn F.D.A. of Health Risks Posed by Flawed Oxygen Devices: Especially for Black patients, inaccurate readings have imperiled care and may have contributed to deaths during the pandemic, experts told an advisory panel*, N.Y. TIMES (Nov. 2, 2022), https://www.nytimes.com/2022/11/02/health/pulse-oximeters-black-patients.html.

34 Bias can arise from multiple sources. *See* Parikh et al., *supra* note 27 (arising from minorities' more limited access to care); *see also* Elliott Crigger & Christopher Khoury, *Making Policy on Augmented Intelligence in Health Care*, 21 AMA J. ETHICS E188, E188–91 (2019) (arising from minorities' more limited access to care); *see also* Darshali A. Vyas et al., *Hidden in Plain Sight — Reconsidering the Use of Race Correction in Clinical Algorithm*, 383 NEW ENG. J. MED. 874 (2020) (arising from expressly racial elements in the clinical algorithms used to treat patients); *see* Parikh et al., *supra* note 27 (arising simply from the reality that communities and the healthcare to which they have access can differ significantly across geographic regions); *see also* Chexi Yuan, *Algorithmic Fairness of Machine Learning Models for Alzheimer Disease Progression*, 6 JAMA NETWORK OPEN e2342203 (2023); *see also* Jiageng Wu et al., *Clinical Text Datasets for Medical Artificial Intelligence and Large Language Models — A Systematic Review*, 1 NEW ENG. J. MED. AI 1 (2024) (An additional source of bias arises as the public availability of clinical datasets to be used for clinical AI generally come from more affluent settings – healthcare organizations that can afford to buy, implement and maintain expensive EMR systems. Thereby "[t]he limited accessibility of clinical text data impedes the development of clinical artificial intelligence systems and hampers research participation from resource-poor regions and medical institutions, thereby exacerbating health care disparities.").

35 Nondiscrimination in Health Programs and Activities, 87 Fed. Reg. 47824, 47824-478920 (Aug. 4, 2022) (codified at 42 C.F.R. pts. 438, 440, 457, 460, 80, 84, 86, 91, 92, 147, 155, 156); *see also* Carmel Shachar & Sara Gerke, *Prevention of Bias and Discrimination in Clinical Practice Algorithms,* 329 JAMA 283, 283–84 (2023); Katherine E. Goodman et al., *Clinical Algorithms, Antidiscrimination Laws, and Medical Device Regulation*, 329 JAMA 285, 285–86 (2023); Goodman, *supra* note 18.

set type, validation approach, and clinical utility assessment."[36] Reproducibility is elusive at best,[37] and replicating a study can be prohibitively costly or actually impossible where data are owned by corporations that choose not to share them.[38] And LLMs, for their part, are known sometimes to produce pure fiction, known as "hallucinations."[39]

The foregoing challenges are well known. However, a rather different, and seriously underrecognized, challenge for AI in clinical healthcare concerns the quality of data from which various kinds of AI tools may be generated. CDS algorithms, for instance, require that massive amounts of data be obtained and cleaned via pre-processing methods.[40] The data must be high quality, and yet at this moment the "[m]ethods to assess data validity and reproducibility are often ad hoc," and "the assessment of the quality of data that are available and the methodology to create a high-quality dataset are not standardized or often are nonexistent."[41] More to the point, a closer look reveals that EMRs actually are laden with documentation errors.

As noted above, "[a]rtificial intelligence and machine learning are limited by the quality of data on which they are trained."[42] Whereas imaging and signal functions (CT, MRI, EKG etc.) are largely produced by machines,[43] EMRs are comprised heavily of the observations

---

36 Kunal C. Potnis et al., *Artificial Intelligence in Breast Cancer Screening Evaluation of FDA Device Regulation and Future Recommendations*, 182 JAMA INTERNAL MED. 1306, 1306 (2022); *see also* Jessica T. Lee et al., *Analysis of Devices Authorized by the FDA for Clinical Decision Support in Critical Care,* 183 JAMA INTERNAL MED. 1399, 1399–1401 (2023) (discussing CDS tools in the setting of critical care).

37 Peter Lee et al., *Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine*, 388 NEW ENG. J. MED. 2399, 2399–2400 (2023).

38 Blumenthal-Barby, *supra* note 22 (noting FDA has begun to regulate Clinical decision support (CDS) tools as medical devices); *see also* Gottlieb, *supra* note 15.

39 Peter Lee et al., *Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine*, 388 NEW ENG. J. MED. 1233, 1233–39 (2023).

40 McCradden et al., *supra* note 14, at 12.

41 NAT'L ACAD. MED., *supra* note 2, at 133.

42 Parikh et al., *supra* note 27; *see also* NAT'L ACAD. MED., *supra* note 2, at 147; *see also* Cohen, *supra* note 5; *see also* Chan, *supra* note 6.

43 Julie M. Vose, *Electronic Medical Records and Physicians: A Love-Hate Relationship*, 38 ONCOLOGY 103, 103 (2024). Even radiographic and signal images, of course, require human involvement and can thereby be affected by human error. For instance, EKG leads must be placed correctly on the patient, patients must be positioned properly, and imaging equipment

human beings make and the descriptive notes they compose—with all the attendant vagaries and nuances, the varying structures for EMR data entry, and a host of other factors that can introduce errors at myriad points.[44] As we will now see, EMRs labor under a variety of errors, both unintentional and, unfortunately, also intentional.

III. DOCUMENTATION ERRORS IN THE MEDICAL RECORD: UNINTENTIONAL

We begin with unintentional errors and omissions, starting with missing data. Medical records commonly fail to include a variety of information that may be relevant for clinical and thereby AI purposes.[45] Some of this dearth, for instance, comes from the fact that much of EMR software, designed initially to maximize billing and then additionally to capture quality data,[46] consists of substantial amounts of "box-checking" that can leave little time for the substantive narrative notes that can reveal important nuance.[47] Additionally, EMRs do not always reflect the fact that the patient has died. One study, for instance, reviewed a cohort of seriously ill primary care patients. When comparing this list with the California Department of Public Health Public Use Death File, researchers found that nineteen percent who

must be periodically serviced and calibrated. Rohini Kopparam et al., *Incorrect Electrocardiogram Lead Placement in ST-Segment-Elevation Myocardial Infraction*, 183 JAMA INTERN. MED. 1156, 1156–57 (2023).

[44] *See id.*

[45] *See* Teno, *supra* note 6; Rachel Knevel & Katherine Liao, *From Real-World Electronic Health Record Data to Real-World Results Using Artificial Intelligence*, 82 ANNALS OF THE RHEUMATIC DISEASES 306, 306–11 (2022); *see also* Lukasz Mazur et al., *Association of the Usability of Electronic Health Records With Cognitive Workload and Performance Levels Among Physicians*, 2 JAMA NETWORK OPEN 1, 1–11 (2019); Kenneth Mandl & Isaac Kohane, *A 21st-Century Health IT System — Creating a Real-World Information Economy*, 376 NEW ENG. J. MED. 1905, 1905-07 (2017); Allan Goroll, *Emerging from EHR Purgatory — Moving from Process to Outcomes,* 376 NEW ENG. J. MED. 2004, 2004-06 (2017); Marissa Hendrickson et al., *The Review of Systems, the Electronic Health Record, and Billing*, 322 JAMA 115, 115–16 (2019).

[46] *See* Rebecca Voelker & Yulin Hswen, *Clinical AI Tools Must Be Fed the Right Data, Stanford Health Care's Chief Data Scientist Says*, 330 JAMA 2137, 2137 (2023); *see also* Katherine Goodman et al., *AI-Generated Clinical Summaries Require More Than Accuracy*, 331 JAMA 637 (2024).

[47] *See* Detmer & Gettinger, *supra* note 1.

were identified as alive in the health plan's EMR, were in fact dead.[48] Outreach from the practice had been made to 80% of those deceased patients.[49]

Outright errors have many origins. For example, "many diagnostic errors are related to core clinical skills, including history taking, physical examination, and other data-gathering activities."[50] As a result, "the quality of diagnostic data in the EHR will directly determine the accuracy of the AI-generated diagnosis."[51] Similarly, another study revealed that in the case of hospitalized patients who either died or were transferred to intensive care, 23% had experienced some sort of diagnostic error while 17% incurred harm or death because of such an error.[52]

Providers' documentation burdens, and various ways to address them, introduce additional opportunities for error. Copy-and-paste, intended to save time, can and does corrupt EMRs.[53] Even if a multiply-copied entry is initially correct—which should not always be presumed—over time it may become incorrect as the patient's condition evolves. A patient who initially had a fever may no longer be febrile, for instance, but a copy-and-paste note may state s/he still has a fever. Although an AI program can recognize such repetition, it can be

---

48 Neil Wenger et al., *Consequences of a Health System Not Knowing Which Patients Are Deceased,* 184 JAMA INTERNAL MED. 213, 213–14 (2024).

49 *Id.*

50 Kulkarni & Singh, *supra* note 6 (citing Hardeep Singh et al., *Types and Origins of Diagnostic Errors in Primary Care Settings*, 173 JAMA INTERNAL MED. 418, 418–25 (2013)).

51 Kulkarni & Singh, *supra* note 6, at 317 (citing Viralkumar Vaghani et al., *Validation of an electronic trigger to measure missed diagnosis of stroke in emergency departments*, 28 J. AM. MED. INFORMATICS ASS'N 2202, 2202–11 (2021)).

52 Andrew D. Auerbach et al., *Diagnostic Errors in Hospitalized Adults Who Died or Were Transferred to Intensive Care*, 184 JAMA INTERNAL MED. 164, 173 (2024); Grace Y. Zhang & Cary P. Gross, *Protecting Patients by Reducing Diagnostic Error*, 184 JAMA INTERNAL MED. 173, 173 (2024); *see also* Gary L. Freed et al., *Error Rates in Race and Ethnicity Designation Across Large Pediatric Health Systems*, 7 JAMA NETWORK OPEN 1, 8 (2024) (noting that in a somewhat similar problem, a recent study showed surprisingly large discrepancies regarding patient's attributed race and ethnicity among EMRs).

53 Michael D. Wang et al., *Characterizing the Source of Text in Electronic Health Record Progress Notes*, 177 JAMA INTERNAL MED. 1212, 1212 (2017) (noting that one recent study found, upon analyzing 23,603 notes written by 460 clinicians, that "[i]n a typical note, 18% of the text was manually entered; 46%, copied; and 36%, imported. Residents manually entered less … and copied more … than did medical students … or direct care hospitalists …").

difficult for the program to discern precisely at what point the copied information became incorrect.

As another instance, although scribes' main purpose is to record notes for physician-patient encounters,[54] errors can arise[55] given that scribes are typically college students who, though perhaps trained in medical terminology, may not always have mastered the linguistic nuances of human communication.

LLMs have now been introduced to function as scribes—"listening" to physician-patient encounters and then writing notes.[56] The EMR software company, Epic, now offers "Nuance" technology to "hear" and document clinical encounter conversations directly into EMRs.[57] Microsoft offers a similar technology,[58] while Amazon Web Services likewise has announced its comparable tool, HealthScribe.[59]

---

54 Sarah Florig et al., *Medical Record Closure Practices of Physicians Before and After the Use of Medical Scribes*, 328 JAMA 1350, 1350–51 (2022); *see also* George A. Gellert et al., *The Rise of the Medical Scribe Industry: Implications for the Advancement of Electronic Health Records*, 313 JAMA 1315, 1315–16 (2015); Paul R. Shafer et al., *Design and Implementation of a Cluster Randomized Trial Measuring Benefits of Medical Scribes in the VA*, 106 CONTEMP. CLINICAL TRIALS 455 (2021); Kristin Ullman et al., *The Effect of Medical Scribes in Cardiology, Orthopedic, and Emergency Departments: A Systematic Review*, 61 J. EMERGENCY MED. 19 (2021).

55 Atul Gawande, *Why Doctors Hate Their Computers*, NEW YORKER (Nov. 5, 2018), https://www.newyorker.com/magazine/2018/11/12/why-doctors-hate-their-computers.

56 Brock Turner, *Epic, Nuance bring more GPT-4 tools to EHRs*, MOD. HEALTHCARE (June 27, 2023, 4:00 AM), https://www.modernhealthcare.com/digital-health/epic-gpt-4-ehr-nuance (noting that Microsoft's Nuance Communications, for instance, has now been adopted by Epic, a major EMR software provider, and will soon serve as "scribe" for many physicians); *see also* Geoff Brumfiel, *Doctors Are Drowning in Paperwork. Some Companies Claim AI Can Help*, NAT'L PUB. RADIO (Apr. 5, 2023, 4:13 AM), https://www.wknofm.org/2023-04-05/doctors-are-drowning-in-paperwork-some-companies-claim-ai-can-help.

57 *See id.*

58 Leroy Leo, *Teladoc expands Microsoft tie-up to document patient visits with AI*, REUTERS (July 18, 2023, 1:08 PM), https://www.reuters.com/business/healthcare-pharmaceuticals/teladoc-health-expands-microsoft-tie-up-automate-documentation-with-ai-2023-07-18/.

59 Mary Roeloffs, *AI at The Doctor? Amazon Launches New Service as Google, Microsoft Aim At Merging Healthcare With Artificial Intelligence*, FORBES (July 26, 2023, 2:25 PM), https://www.forbes.com/sites/maryroeloffs/2023/07/26/ai-at-the-doctor-amazon-launches-new-service-as-google-microsoft-aim-at-merging-healthcare-with-artificial-intelligence/?sh=7fae3f6824f7; *see also* Katyanna Quach, *Healthcare Org With Over 100 Clinics Uses Openai's GPT-4 to Write Medical Records*, THE REGISTER (June 6, 2023), https://www.theregister.com/2023/06/06/carbon_health_deploys_gpt4powered_tools/; Meanwhile, Mayo Clinic, Kaiser and CVS are likewise investing in a generative AI company. Brock Turner, *Mayo,*

As with other uses of LLMs, these AI-scribe services can produce inaccuracies such as hallucinations. In one study, for instance, a chatbot generated summaries of histories and physical examinations. Although those AI-generated summaries became increasingly accurate as prompts improved, nevertheless the AI responses commonly fabricated patients' age and gender.[60] Additionally, Automated Speech Recognition can mis-"hear" the words of minority groups whose accents or speech patterns do not fit the speech patterns on which the software is trained. In an attempt to rectify the problem, Google and Howard University have launched "Project Elevate Black Voices".[61]

Unintentional errors are not new, but their frequency has been highlighted by "Open Notes"—patients' legally authorized access to view nearly all their own medical records, without charge and without delay.[62] With easier access, patients have identified errors. In one

---

Kaiser, CVS Invest in Generative AI Company, MOD. HEALTHCARE (Oct. 26, 2023, 10:07 AM), https://www.modernhealthcare.com/digital-health/abridge-ai-mayo-clinic-cvs-health-ventures-epic.

60 Ashwin Nayak et al., *Comparison of History of Present Illness Summaries Generated by a Chatbot and Senior Internal Medicine Residents*, 183 JAMA INTERNAL MED. 1026, 1026–27 (2023); *see also* Eric Ward & Cary Gross, *Evolving Methods to Assess Chatbot Performance in Health Sciences Research,* 183 JAMA INTERNAL MED. 1030, 1030–31 (2023); *see also* Davey Alba & Ike Swetlitz, *Google Taps AI to Revamp Costly Health-Care Push Marred by Flops*, BLOOMBERG (July 30, 2024), https://www.bloomberg.com/news/features/2024-07-30/google-sees-ai-as-the-key-to-a-health-care-revolution (noting AI tool intended to create nurse-to-nurse handoff notes failed to include such medically crucial data as patient allergies); *see also* Garance Burke & Hilke Schellmann, *Researchers Say an AI-Powered Transcription Tool Used in Hospitals Invents Things No One Ever Said*, AP NEWS (Oct. 26, 2024), https://apnews.com/article/ai-artificial-intelligence-health-business-90020cdf5fa16c79ca2e5b6c4c9bbb14 (noting, *inter alia*, a study showing widespread fabrications in trials of various AI-powered transcription tools used to summarize recordings of medical encounters).

61 Candace McDuffie, *Google and Howard University Are Changing The Future Of Voice Technology With Project Elevate Black Voices*, THE ROOT INST. (Sept. 20, 2023), https://www.theroot.com/google-and-howard-university-are-changing-the-future-of-1850855438; *see also* Zion Mengesha et al., *"I don't Think These Devices are Very Culturally Sensitive."— Impact of Automated Speech Recognition Errors on African Americans*, 4 FRONTIERS A.I. 1, 1–12 (2021).

62 21st Century Cures Act, H.R. 34, 114th Cong. (2016) (requiring that patients must have access to nearly everything in their EMR – including progress notes, consultation notes, discharge summaries, history and physical notes, imaging reports, pathology reports, and procedure notes (exception: psychotherapy notes)); *see* U.S. Federal Rule Mandates Open Notes, https://www.opennotes.org/onc-federal-rule/ (last visited Mar. 7, 2024) (beginning October 6, 2022, the definition of "electronic health information" expanded beyond the "core data" identified in the law's initial version, to encompass essentially all the data identified under

widely cited study, Bell and colleagues found that, of nearly 137,000 patients who responded to a survey, over 21% reported a perceived mistake and, of these, 42.3% indicated that the error was somewhat or very serious.[63] Other studies have found rates of 21%[64], 17%[65], 18%[66], 12.4%[67] and 70%.[68] Anecdotes from patients interviewed for these studies are illustrative.

> "I am a nonsmoker and do not abuse alcohol." "Note said my mother had Alzheimer's. Not correct. My mother died of congestive heart failure while recovering in a skilled nursing facility from spinal meningitis."[69]
>
> "I was not born in 1910."[70]

---

the HIPAA law as "protected health information"); Liz Salmi et al., *US policy requires immediate release of records to patients*, 372 BRIT. MED. J. 426 (2021); *see also* Betty Lengyel-Gomez, *21st Century Cures Act—A Summary*, HEALTHCARE INFO. & MGMT. SYS. SOC'Y, https://www.himss.org/resources/21st-century-cures-act-summary (Nov. 20, 2018); *see also* Mario Fucinari, *Empowering Patients with Their Health Record in a Modern Health IT Economy*, CHIROHEALTHUSA (May 1, 2021), https://www.chirohealthusa.com/consultants/empowering-patients-with-their-health-record-in-a-modern-health-it-economy/ (discussing when a delay in fulfilling requests is an interference under the information blocking regulation).

63 Sigall K. Bell et al., *Frequency and Types of Patient-Reported Errors in Electronic Health Record Ambulatory Care Notes*, 3 JAMA NETWORK OPEN 1, 1-16 (2020).

64 Sigall K. Bell et al., *A Patient Feedback Reporting Tool for OpenNotes: Implications for Patient-Clinician Safety and Quality Partnerships*, 26 BRIT. MED. J QUALITY & SAFETY 312, 316 (2017).

65 Barbara D. Lam et al., *Speaking up About Patient-Perceived Serious Visit Note Errors: Patient and Family Experiences and Recommendations,* 28 J. AM. MED. INFORMATICS ASS'N 685, 688 (2020).

66 Rachael Lear et al., *Patients' Willingness and Ability to Identify and Respond to Errors in Their Personal Health Records: Mixed Methods Analysis of Cross-Sectional Survey Data*, 24 J. MED. INTERNET RSCH. 1, 9 (2022); *see also* Bell, *supra* note 64.

67 Lisa Freise et al., *Assessment of Patients' Ability to Review Electronic Health Record Information to Identify Potential Errors: Cross-Sectional Web-Based Survey*, 5 JMIR FORMATIVE RSCH. 1, 1-10 (2021); *see also* Sari Kujala et al., *Patients' Experiences of Web-Based Access to Electronic Health Records in Finland: Cross-Sectional Survey*, 24 J. MED. INTERNET RSCH. 1, 8 (2022).

68 Cecilia Pyper et al., *Patients' Experiences When Accessing Their On-line Electronic Patient Records in Primary Care*, 54 BRIT. J. GEN. PRAC. 38, 38–43 (2004).

69 Bell, *supra* note 64.

70 Pyper, *supra* note 68.

Additionally, patients identify important information that should be present but is missing, such as medications, vaccinations, allergies, test results. Likewise, patients also identify missing events such adverse reactions to medication, operations, tuberculosis, premature childbirth, miscarriage, irritable bowel syndrome, severe migraine, glaucoma.[71]

> "In one surgical encounter the procedure was cancelled at the beginning, after I was under anesthesia, and they woke me up to cancel the procedure. Nevertheless, there was a note in my chart that described the procedure in detail and claimed it had gone smoothly."[72]

> "It mentioned my 22-year-old son living in a trailer behind my house who was providing me with some support. I don't have a 22-year-old son, nor do I have a trailer."[73]

IV. DOCUMENTATION ERRORS IN THE MEDICAL RECORD: INTENTIONAL

Intentional errors can occur for many reasons. Sometimes physicians may misrecord events in an effort to ward off potential tort liability. A somewhat similar error arises as a common annoyance in contemporary EMRs is seemingly endless box-checking, which can trigger a variety of responses.[74] For instance, billed amounts for outpatient encounters often rise, the more body areas and organ systems the physician examines.[75] Hence many EMR software packages require

---

71 *Id.*

72 Personal experience communicated by colleague; on file with the author.

73 Personal experience communicated by colleague; on file with the author.

74 *See* Detmer & Gettinger, *supra* note 1.

75 *See, e.g.*, *Evaluation and Management (E&M) Guidelines*, COLUM. U. IRVING MED. CTR. OFF. BILLING COMPLIANCE, https://www.compliance.cuimc.columbia.edu/compliance-standards/evaluation-and-management-e-m-guidelines (last visited Oct. 19, 2024) (Section II (Documentation of E/M Services) discusses seven component that are used in defining the levels of E/M services for payment, of which the most important are history, examination and medical decision making. The examination portion, usually focusing on physical examination, requires documentation that various body areas and/or organ systems were examined. Section II.B. (Documentation of Examination) notes that coding can discern whether the level of medical decision-making should best be characterized as straightforward, low complexity, moderate complexity, or high complexity. Section II.C. (Documentation of the complexity of medical decision-making) notes that services such as coordinating care can be properly

physicians to check a box or enter other notation for each body area/organ system they examine and for ascertaining various sorts of quality measures.[76] Anecdotally, where physicians believe it is unnecessary to examine a particular body part, the easier course may simply be to check the box and get on with the rest of the visit. The EMR may then indicate that actual examinations corresponded to each checked box, and that no problems were found. If the underlying reality is otherwise, then the EMR's reflection of "examined, no problem found" may be erroneous.

That said, the great bulk of intentional errors stems from health insurers' modes of payment. On one hand, traditional FFS payment compensates providers more, the more individual services they provide.[77] In contrast, "capitation" pays providers a flat fee, per-patient-per-month, in exchange for all the care that patient needs.[78] That payment, in turn, is typically risk-adjusted so that sicker patients bring higher compensation to offset their higher costs.[79]

A. Fee-for-service

FFS, as the name suggests, provides a specified fee for each separate service a provider delivers.[80] It carries the obvious temptations to provide unnecessary services, to document the provision of services that were not in fact delivered, to exaggerate the severity of the patient's condition in order to either secure authorization for a service not otherwise covered, or to "justify" an insurer's paying higher amounts for more serious conditions and procedures.

---

characterized as problem-focused, detailed, comprehensive, etc. Section II.D. (Documentation of an encounter dominated by counseling or coordination of care)).

76 *See* Detmer & Gettinger, *supra* note 1.

77 Sri Lekha Tummalapalli et al., *supra* note 9; *see also Capitation vs. Fee For Service*, DIFFEN, https://www.diffen.com/difference/Capitation_vs_Fee_For_Service (last visited Mar. 14, 2024).

78 Sri Lekha Tummalapalli et al., *supra* note 9; *see also Capitation vs. Fee For Service*, *supra* note 77.

79 Jeff Lagasse, *Medicare Advantage growth fueled by people switching from original Medicare*, HEALTHCARE FIN. (Sept. 7, 2023), https://www.healthcarefinancenews.com/news/medicare-advantage-growth-fueled-people-switching-original-medicare.

80 Sri Lekha Tummalapalli et al., *supra* note 9; *see also Capitation vs. Fee For Service*, *supra* note 77.

Beginning in the 1970s and 1980s, insurers responded to burgeoning healthcare costs with "managed care" processes requiring physicians to document that their care was truly necessary and appropriately provided.[81] Requirements to document medical necessity—usually, with specific kinds of wording—wrought endless cycles of gamesmanship as physicians learned what words to use to secure coverage, followed by insurers' changing those rules, to again ramp up bases for denying coverage.[82] Thus, in the days before mammography was routinely covered, for instance, physicians learned to write "rule out breast cancer" instead of "screen for breast cancer."[83] And then, when insurers caught on, they learned to write "possible breast mass" instead of "rule out."[84]

Medical literature has documented such gamesmanship. Novack and colleagues found in 1989, for instance, that 87% of physicians expressed willingness to commit deception at least occasionally "to circumvent 'ridiculous rules'"[85] e.g., as 53% of respondents would manipulate the record to secure coverage for (then-noncovered) routine mammograms.[86] A decade later Freeman et al. found that physicians were willing to deceive in the medical record to obtain coverage for coronary bypass surgery (57.7%) or for arterial revascularization (56.2%).[87] Although these and similar studies focused on hypothetical

---

81 E. HAAVI MORREIM, BALANCING ACT: THE NEW MEDICAL ETHICS OF MEDICINE'S NEW ECONOMICS 8–42 (H. Tris Engelhardt et al. eds., 1995).

82 E. Haavi Morreim, *Gaming the System. Dodging the Rules, Ruling the Dodgers*, 151 JAMA 443, 443–47 (1991); *see also* Dena Rifkin, *Community Considerations: The Many Effects of Miscoding*, 282 JAMA 1676 (1999); Gregory Webster, *Serving Two Masters: Medical Practice vs Administrative Ethics*, 282 JAMA 1678 (1999).

83 Morreim, *supra* note 82.

84 Morreim, *supra* note 82; *See generally* Rifkin, *supra* note 82; *See generally* Webster, *supra* note 82.

85 Dennis H. Novack et al., *Physicians' Attitudes Toward Using Deception to Resolve Difficult Ethical Problems*, 261 PUB. MED. 2980, 2983 (1989).

86 *Id.* at 2983.

87 Victor G. Freeman et al., *Lying for Patients Physician Deception of Third-Party Payers*, 159 ARCH INTERN MED. 2263, 2266 (1999). Prior studies showed, e.g., that "nearly 70% of respondents would knowingly misdocument a screening test as a diagnostic service to ensure coverage of service. Primary care physicians have historically misdocumented diagnoses to reduce the risk of patient stigmatization or insurance and employment discrimination." *See id.* at 2269 (citing Mark Siegler, *Confidentiality in Medicine: A Decrepit Concept*, 307 NEW ENG. J. MED. 1518 (1982)); *see also* Katja Rost et al., *The Deliberate Misdiagnosis of Major Depression*

cases, in 2000, Wynia et al. documented physicians' actual behavior. In a study providing strong protections for respondents' anonymity, researchers found that nearly 40% reported actually having exaggerated the severity of patients' condition in order to gain insurance coverage.[88]

Fast forward a couple decades. Recent legal circumscriptions on abortion and on medical interventions for transgender minors have opened additional avenues for miscoding, e.g. to miscode a medical or surgical abortion as a miscarriage.[89] Similarly, given that Medicare and

---

in Primary Care, ARCH FAM. MED. 333, 333 (1994); Kathryn L. Anderson, *Deceptive Documentation in Home Healthcare Nursing*, 10 HOME HEALTHCARE NURSE 31, 33 (1992); Julia E. Connelly & Steven DalleMura, *Ethical Problems in the Medical Office*, 260 JAMA 812, 815 (1988). More broadly, regarding physicians and deception: "What is known about the use of deception by physicians in other contexts, however, is cause for concern. Lying to patients about their diagnoses was ubiquitous in the United States until only a few decades ago, and is still common in many other countries. Physicians commonly use deception to resolve ethical dilemmas, particularly when dealing with insurance companies and when attempting to act in their patient's interests. Some physicians even participate in outright fraud. Cheating before and during medical school occurs frequently, and numerous physicians falsely represent their qualifications when applying to specialty fellowships. Many physicians in training say they have covered up mistakes in the medical record and lied to attending physicians about things they neglected to do, and lying about one's role in authoring articles is commonplace." Michael J. Green et al., *Lying to Each Other: When Internal Medicine Residents Use Deception With Their Colleagues*, 160 ARCH INTERN MED. 2317, 2317–18 (2000) (citations omitted). Green and colleagues continue, regarding their study of medical residents' willingness to deceive: "Particular responses warrant further comment. Nineteen percent of responding residents said they would misrepresent information in the medical record if it would protect a patient's privacy about genital herpes. While there is nothing new about the use of deception by physicians to promote a patient's welfare, the relatively common practice of falsifying medical records, even to protect patient privacy, is ethically troubling for several reasons." *Id.* at 2321 (citing DeWitt C. Baldwin et al*., Unethical and unprofessional conduct observed by residents during their first year of training*, 73 ACAD. MED. 1195 (1998)).

88 Matthew. K. Wynia et al., *Physician Manipulation of Reimbursement Rules for Patients: Between a Rock and a Hard Place*, 283 JAMA 1858, 1863 (2000); More precisely, "[o]f the 720 physicians who responded to their survey (64% response rate), 39% reported that they had 'sometimes,' 'often,' or 'very often' 'exaggerate[d] the severity' of a patient's condition, 'change[d] a patient's official (billing) diagnosis,' or 'report[ed] signs or symptoms that a patient did not actually have.' Of these physicians, 54% reported 'using deception of third-party payers to obtain needed benefits' more often 'now' (in 1998) than five years before, while only 12% said they did so less often. More than a quarter (28.5%) of all respondents agreed that 'it is necessary to game the system to provide high-quality care.'" M. Gregg Bloche, *Fidelity and Deceit at the Bedside*, 283 JAMA 1881, 1882 (2000) (summarizing Wynia et al. study). Many commentators have expressed concerns regarding the ethics of gamesmanship. *See also, e.g.,* Rifkin, *supra* note 82; *see also, e.g.*, Webster, *supra* note 82; *see also, e.g.,* Bloche, *supra*.

89 Sravya Chary et. al., *Abortion Miscoding—Legal Risks for Clinicians and Hospital Systems*, 329 JAMA 1911 (2023); *see also* Melissa Brown & Kelly Puente, *Vanderbilt turns over*

many employers currently decline to cover treatments for obesity,[90] and that the same drugs are often used to treat diabetes mellitus type 2 (DMT2),[91] it would not be surprising to see some EMRs begin to state falsely that a patient has DMT2, who in fact does not have the condition.

A relatively new impetus toward gaming under FFS comes from the emergence of private equity (PE). PE ownership of physician practices and other healthcare entities is different from a typical large group medical practice, or hospital acquisition of a medical group. In PE, investor groups purchase practices, aggressively enhance revenues and cut costs, consolidate into larger entities, and sell within 4 to 7 years with an expected return of at least 20%. This profit model heightens the focus on short-term profitability at the expense of

---

transgender patient records to state in attorney general probe, THE TENNESSEAN, https://www.tennessean.com/story/news/health/2023/06/20/vanderbilt-university-m-turns-over-transgender-patient-medical-records-to-tennessee-attorney-general/70338356007/ (June 21, 2023, 2:00 PM); Anita Wadhwani & Sam Stockard, *AG made sweeping demands for Vanderbilt transgender clinic records as part of fraud inquiry*, TENN. LOOKOUT (June 22, 2023, 6:00 AM), https://tennesseelookout.com/2023/06/22/ag-made-sweeping-demands-for-vanderbilt-transgender-clinic-records-as-part-of-fraud-inquiry/.

90 Khrysta Baig et al., *Medicare Part D Coverage of Antiobesity Medications—Challenges and Uncertainty Ahead*, 388 NEW ENG. J MED. 961 (2023); Amanda Seitz, *New weight loss drugs are out of reach for millions of older Americans because Medicare won't pay*, AP NEWS (Dec. 28, 2023, 4:23 PM), https://apnews.com/article/wegovy-ozempic-zepbound-medicare-obesity-weight-loss-02d4500e737d30d070d70907521a4fe0; *see also* Brenda Goodman, *Insurance denials for popular new weight loss medications leave patients with risky choices*, CNN (Jan. 8, 2024, 6:40 AM) https://www.cnn.com/2024/01/08/health/weight-loss-drug-insurance-denials/index.html; States, in contrast, have somewhat greater flexibility for Medicaid coverage of medications, hence some do cover anti-obesity medications. *See, e.g.*, Benjamin Y. Liu & Benjamin N. Rome, *State Coverage and Reimbursement of Antiobesity Medications in Medicaid*, 331 JAMA 1230 (2024). In the aftermath of a trial showing that the weight loss drug Wegovy reduced by 20 percent the risk of adverse cardiovascular events in overweight patients. *See Novo Nordisk A/S: Semaglutide 2.4 mg Reduces the Risk of Major Adverse Cardiovascular Events by 20% in Adults with Overweight or Obesity in the SELECT trial*, NOVO NORDISK, (Aug. 8, 2023), https://www.globenewswire.com/news-release/2023/08/08/2720343/0/en/Novo-Nordisk-A-S-Semaglutide-2-4-mg-reduces-the-risk-of-major-adverse-cardiovascular-events-by-20-in-adults-with-overweight-or-obesity-in-the-SELECT-trial.html. The FDA expanded the approved medical indications for Wegovy to include reducing cardiovascular risks in overweight persons. Thereafter, CMS announced Medicare would cover Wegovy for beneficiaries with a history of cardiovascular disease. *See* Daniel Gilbert, *Medicare Plans Can Cover Wegovy for Patients with Heart Conditions*, WASH. POST (Mar. 21, 2024), https://www.washingtonpost.com/business/2024/03/21/wegovy-medicare-part-d/. Note that, at press time, coverage for obesity drugs remains in flux.

91 Goodman, *supra* note 90.

longer-term investment in the success of the practice. A 2022 study of 578 dermatology, gastroenterology, and ophthalmology physician practices acquired by PE companies found that there was a 20% increase in charges per claim. Private equity acquisition of anesthesia practices was similarly associated with a 26% greater increase in cost trajectory between 2012 and 2017, compared with control practices.[92]

When a PE firm controls 30% or more of a given specialty in a market, costs are likely to rise significantly.[93] More important for present purposes, some of PEs' tactics for building return on investment appear to involve misrepresentations in the EMR. Allegations include pressuring physicians to perform the most lucrative procedures[94] (which in turn can require exaggerating symptoms in the EMR to "justify" the procedure for insurers); gaming reimbursement formulas;[95] "overlooked diagnoses, lost patient biopsies, questionable quality control in the company-owned lab, and overbooking of patients without

---

92 Francis J. Crosson et al., *Private Equity in US Health Care—Now Cradle to Grave?*, 183 JAMA INTERNAL MED. 511, 511–12 (2023); *see also* Christopher Cai & Zirui Song, *A Policy Framework for the Growing Influence of Private Equity in Health Care Delivery*, 329 JAMA 1545 (2023) ("[P]rivate equity firms operate on a shorter timeline, often acquiring and selling a clinical entity within approximately 5 to 10 years. This emphasis on rapid returns on investment is thought to be consistent with the finding that private equity firms increase prices, increase volume, shift to higher-margin services, and reduce labor costs, including physician staff shortly after acquisition."); Yashaswini Singh et al., *Association of Private Equity Acquisition of Physician Practices With Changes in Health Care Spending and Utilization*, 3 JAMA HEALTH FORUM 1, 1 (2022) ("Compared with the 2874 control practices, the 578 PE-acquired physician practices exhibited an average increase of $71 (+20.2%) charged per claim and $23 (+11.0%) in the allowed amount per claim. The PE-acquired practices increased their numbers of unique patients seen by 25.8% compared with control practices, driven by a 37.9% increase in visits by new patients. In aggregate, their volume of encounters increased by 16.3% compared with the control group, with a 9.4% increase in the share of office visits for established patients that were billed as longer than 30 minutes.").

93 Richard M. Scheffler et al., AM. ANTITRUST INST., MONETIZING MEDICINE: PRIVATE EQUITY AND COMPETITION IN PHYSICIAN PRACTICE MARKETS 6 (2023), https://www.antitrustinstitute.org/wp-content/uploads/2023/07/AAI-UCB-EG_Private-Equity-I-Physician-Practice-Report_FINAL.pdf; *see also* Yashaswini Singh et al., *Geographic Variation in Private Equity Penetration Across Select Office-Based Physician Specialties in the US*, 3 JAMA HEALTH FORUM 1, 2 (2022); Cai & Song, *supra* note 92, at 1; Kevin Tyan, *Private Equity Acquisition of Oncology Clinics in the US From 2003 to 2022*, 183 JAMA INTERNAL MED. 621 (2023); Jane M. Zhu et al., *Private Equity Acquisitions of Physician Medical Groups Across Specialties 2013-2016*, 323 JAMA 663, 665 (2020).

94 SCHEFFLER, *supra* note 93, at 13.

95 SCHEFFLER, *supra* note 93, at 16.

sufficient support staff";[96] and engaging in medically dubious practices such as permitting unsupervised allied clinicians to provide care, while documenting that they were appropriately supervised.[97] Recent research indicates that PE-acquired hospitals appear to experience higher-than-expected rates of adverse events and poorer patient outcomes.[98] It would not be surprising to see under-reporting of these events in the associated EMRs.

B. Capitation

Although prepaid healthcare has existed for a very long time, the Health Maintenance Organization Act of 1973[99] accelerated the revitalization of healthcare provided via a flat fee, per member per month, to cover all the care each person needs, usually for a year, with care ordinarily delivered by a specified group of providers.[100] Currently, the most prominent example of capitation is Medicare Part C, also known

---

[96] Ryan Basen, *Doc Fired After Standing Up to Private Equity; RFK Jr.'s Anti-Vax Machine*, MEDPAGE TODAY (Dec. 22, 2021), https://www.medpagetoday.com/special-reports/exclusives/96346; *see also* Kara Grant, *Is Private Equity a Dangerous Employer?—Firms' investments in healthcare continue to expand, at what cost?*, MEDPAGE TODAY (Oct. 14, 2021), https://www.medpagetoday.com/special-reports/exclusives/95022.

PE-acquired dermatologists, for instance, "reported pressure to meet production numbers for procedures, sell products (eg, acne creams and antiaging products), and refer patients to affiliated specialists, laboratory technicians, and estheticians. Other concerns included upcharging in billing offices and significant reliance on physician assistants in unsupervised settings." Suhas Gondi et al., *Potential Implications of Private Equity Investments in Health Care Delivery*, 321 JAMA 1047, 1048 (2019).

[97] Gondi et al., *supra* note 96, at 1047.

[98] Sneha Kannan et al., *Changes in Hospital Adverse Events and Patient Outcomes Associated With Private Equity Acquisition*, 330 JAMA 2365, 2374 (2023).

[99] 42 U.S.C. § 300(e).

[100] MARJORIE SMITH, DIV. OF ECON. AND LONG-RANGE STUD., HEALTH MAINTENANCE ORGANIZATION ACT OF 1973 35 (1974); In a reversal from FFS, which can provide excessive services to enhance profits, a well-known potential problem of capitated health plans is the risk that the plan will provide too little care, enhancing profits by keeping more of the flat fee. *See, e.g.,* Reed Abelson, *Insurers Deny Medical Care for the Poor at High Rates, Report Says*, N. Y. TIMES (July 19, 2023), https://www.nytimes.com/2023/07/19/health/health-insurance-medicaid-denials.html. Although potentially problematic, the issue is not discussed in this Article because it may be less likely to prompt misdocumentation in the EMR and thus far has not been documented to cause EMR inaccuracies. Nevertheless, if under-service prompts EMR entries that understate the severity of the patient's condition, that sort of error could corrupt the EMR and thereby the AI dataset, right alongside up-coding.

as Medicare Advantage, now enrolling half or more of Medicare beneficiaries.[101]

In Medicare Advantage, per-patient fees are risk adjusted so that the Medicare Advantage Organizations (MAOs) that offer these plans are not financially penalized for accepting patients with co-morbidities or complex conditions.[102] The more diagnoses a patient has, and the more complex those diagnoses are, the more the insurer is paid.[103] Each year, MAOs can adjust their patients' diagnoses upward or downward as patients' medical conditions evolve, to ensure correct payment.[104]

Although every diagnosis must be based on face-to-face encounters between patients and providers and documented in the medical record,[105] MAOs can add or subtract diagnoses without directly providing EMR evidence.[106] Rather, they can submit data via (a) retrospective chart reviews the MAOs themselves conduct, or (b) overall encounter data they enter into Center for Medicare and Medicaid Services (CMS) databases.[107] Historically, these retrospective reviews need not actually be directly linked to patients' medical records.[108] Also important, "[w]ith data for millions of people being submitted

---

101 Lagasse, *supra* note 79. *See also* Lanlan Xu et al., *Medicare Switching: Patterns Of Enrollment Growth In Medicare Advantage, 2006-22*, 42 HEALTH AFFS. 1203 (2023); Nancy Ochieng et al., *Medicare Advantage in 2023: Enrollment Update and Key Trends*, KAISER FAM. FOUND. (Aug. 9, 2023), https://www.kff.org/medicare/issue-brief/medicareadvantage- in-2023-enrollment-update-and-key-trends/.

102 Lagasse, *supra* note 79.

103 Medicare uses "hierarchical condition category" (HCC) codes, identifying some conditions as more severe, hence meriting higher risk adjustments than other conditions. *See* JOANNE M. CHIEDI, U.S. DEP'T. OF HEALTH AND HUM. SERV, BILLIONS IN ESTIMATED MEDICARE ADVANTAGE PAYMENTS FROM CHART REVIEWS RAISE CONCERN, 1, 1 (2019), https://oig.hhs.gov/oei/reports/oei-03-17-00470.pdf; *See also* U.S. *ex rel.* Poehling v. UnitedHealth Grp., Inc., No. CV 16-8697 MWF (SSx), 2018 WL 11350603, at *2 (C.D. Cal. Oct. 23, 2018); David J. Meyers & Amal N. Trivedi, *Medicare Advantage Chart Reviews Are Associated with Billions in Additional Payments for Some Plans*, 59 MED CARE 96, 97 (2021).

104 Chiedi, *supra* note 103, at 2–3.

105 *See* 42 C.F.R. § 422.300 (2023); *see also* 42 U.S.C. §§ 1395l(e) (2022), 1395y(a)(1)(A); 42 C.F.R. § 422.310(d) (2024); CTRS. FOR MEDICARE & MEDICAID SERV., MEDICARE MANAGED CARE MANUAL 1, 4 (2014); CHIEDI, *supr*a note 103, at 3.

106 CHIEDI, *supra* note 103, at 5.

107 CHIEDI, *supra* note 103, at 4.

108 CHIEDI, *supra* note 103, at 6.

each year, CMS is unable to confirm diagnoses before calculating capitation rates. Instead, the agency accepts the diagnoses as submitted, and then audits some of the self-reported data a few years later to ensure that they are adequately supported by medical documentation."[109]

In its 2019 report the Office of Inspector General (OIG), Department of Health and Human Services (DHHS), found that MAO chart reviews that were not based on service records added $6.7 billion in risk-adjusted payments for 2017, and another $2.7 billion in risk-adjusted payments that had no link to any face-to-face visit.[110] "In 2016 alone, audits of the data submitted by Medicare Advantage insurers to CMS showed that CMS paid out an estimated $16.2 billion for unsupported diagnoses, equal to 'nearly ten cents of every dollar paid to Medicare Advantage organizations.'"[111]

It appears these escalations are not simply the product of MAOs' better documentation. Rather, a raft of False Claims Act (FCA) allegations details a wide array of incorrect recordings in medical records. Overall, at least eight of the ten largest participating MAOs have now been accused of elevating their populations' risk assessments, including to push physicians and nurses "to document a range of diagnoses, including some—vertebral fractures, pneumonia and cancer—they lacked the equipment to detect."[112]

A brief dive into some of these FCA allegations will illustrate the problem. Note that because FCA cases are often settled rather than adjudicated, these illustrations are allegations from filed complaints, rather than proven facts.[113] Still, they need to be taken seriously. FCA

---

109 U. S. *ex rel.* Silingo v. WellPoint, Inc., 904 F.3d 667, 672 (9th Cir. 2018) (citing 42 C.F.R. §§ 422.310(e), 422.311).

110 CHIEDI, *supra* note 103, at Report in Brief; *see also* Meyers, *supra* note 103.

111 United Healthcare Ins. Co. v. Bacerra, 16 F.4d 867, 872 (D.C. Cir. 2021) (citing U. S. *ex rel.* Silingo v. WellPoint, Inc., 904 F.3d 667, 673 (citing JAMES COSGROVE, U.S. GOV'T ACCOUNTABILITY OFF., GAO-17-761T, MEDICARE ADVANTAGE PROGRAM INTEGRITY: CMS'S EFFORTS TO ENSURE PROPER PAYMENTS 1 (2017)).

112 *See* Reed Abelson & Margot Sanger-Katz, *'The Cash Monster Was Insatiable': How Insurers Exploited Medicare for Billions,* N.Y. TIMES (Oct. 8, 2022, 9:52), https://www.nytimes.com/2022/10/08/upshot/medicare-advantage-fraud-allegations.html. By next year, half of Medicare beneficiaries will have a private Medicare Advantage plan. Most large insurers in the program have been accused in court of fraud. *Id.*

113 *Why Do So Many False Claims Settle Before Trial?*, GOLDBERG KOHN (Mar. 11, 2019)

claims, as a species of fraud, must be pled with particularity—detailing the circumstances of "'the who, what, when, where, and how of the misconduct charged,' including what is false or misleading about a statement, and why it is false."[114]

United States ex rel. Ormsby v Sutter Health[115]

In *United States ex rel. Ormsby v Sutter Health*,[116] relator Kathy Ormsby worked as a risk adjustment project manager for Palo Alto Medical Foundation (PAMF), a provider affiliate of Sutter Health, which in turn furnished physician healthcare services under gainsharing (financial incentive) arrangements for three MAOs: UnitedHealth, Health Net, and Humana.[117] Per Ormsby's allegations, Sutter and PAMF engaged nonphysicians to act as liaisons between physicians and the people extracting diagnosis codes from EMRs (coders).[118] Tactics allegedly included:

> * retroactively characterizing various conditions as current and chronic rather than as individual acute episodes now resolved;[119]
>
> * using a non-physician pit crew to add diagnosis codes to patients' medical records, e.g. to change a diagnosis from acute bronchitis to (more lucrative) pneumonia, whether or not the EMR data justified such additions;[120]

---

https://www.whistleblowersattorneys.com/blog/why-do-so-many-false-claims-cases-settle-before-trial/; *See, e.g.*, *Justice Department's False Claims Act Settlements and Judgments Exceed $5.6 Billion in Fiscal Year 2021*, U.S. DEP'T JUST. (Feb. 1, 2022), https://www.justice.gov/opa/pr/justice-department-s-false-claims-act-settlements-and-judgments-exceed-56-billion-fiscal-year; *see also False Claims Act Settlements and Judgments Exceed $2 Billion in Fiscal Year 2022,* U.S. DEP'T JUST. (Feb. 7, 2023), https://www.justice.gov/opa/pr/false-claims-act-settlements-and-judgments-exceed-2-billion-fiscal-year-2022.

114 United States v. United Healthcare Ins. Co., 848 F.3d 1161, 1163 (9th Cir. 2016), (citing Fed. R. Civ. P. 9(b)); Ebeid *ex rel.* United States v. Lungwitz, 616 F.3d 993, 998 (9th Cir. 2010) (quoting Vess v. Ciba–Geigy Corp. USA, 317 F.3d 1097, 1106 (9th Cir. 2003) and Decker v. GlenFed, Inc., 42 F.3d 1541, 1548 (9th Cir. 1994) (en banc)).

115 United States *ex rel.* Ormsby v Sutter Health, 444 F.Supp.3d 1010 (N.D. Cal. 2020).

116 *Id.*

117 U. S. *ex rel.* Ormsby v Sutter Health, 444 F.Supp.3d 1010, 1026 (N.D. Cal. 2020).

118 *Id.* at 1027.

119 *Id.*

120 *Id.* at 1028.

> * pressuring, incentivizing and rewarding physicians for adding diagnosis codes to patients' medical records;[121]
>
> * requiring one-on-one meetings between auditors and physicians to suggest addenda physicians should make to capture missed diagnoses;[122]
>
> * pre-populating patients' medical records with diagnosis codes, prior to actual physician-patient encounters and regardless of the specific health conditions the physician was evaluating or treating during actual encounters;[123]
>
> * retroactively inserting diagnosis codes as an addendum to EMRs, via "data mining" software;[124]
>
> * limiting auditors to removing "unsupported diagnosis codes only from the 'billing' side of electronic medical records, not from the 'encounter data' that was submitted to CMS for payment."[125]

In October 2012, a UnitedHealth review of Sutter's coding found that 90% of patient records containing a diagnosis of heart attack were "erroneous, invalid, unsupported, or otherwise false."[126] An April 2014 audit by UnitedHealth and Health Net "identified over 8,000 false diagnosis codes for MA Plan patients that Sutter and its affiliates needed to delete based on 'overcod[ing]' and diagnoses 'not supported in documentation.'"[127]

Ormsby herself conducted various audits and found, e.g. for one patient, that eighteen of twenty diagnosis codes were false.[128] In December 2014 her audit team presented a final tally of risk adjustment codes for cancer, strokes, and fractures.[129] "For cancer, 164 of 182 patient records (90%) were erroneous, invalid, unsupported, or

---

121 *Id.* at 1028–29.

122 *Id.* at 1032.

123 *Id.* at 1053.

124 *Id.* at 1030–31.

125 *Id.* at 1052.

126 *Id.* at 1033.

127 *Id.* at 1033.

128 *Id.* at 1037.

129 *Id.* at 1038.

otherwise false.[130] For stroke, 162 of 169 patient records (96%) were erroneous, invalid, unsupported, or otherwise false.[131] For fracture, fifty-seven of eighty-six patient records (66%) were erroneous, invalid, unsupported, or otherwise false."[132]

Per the complaint, the facts go downhill from there.[133]

United States ex rel. Osinek v. Permanente[134]

While details differ somewhat, a similar story emerges in *United States ex rel. Osinek v. Permanente*, in the government's Amended Complaint in Intervention.[135] According to the complaint, Kaiser and its affiliated Permanente Medical group used a variety of tactics designed to augment diagnosis codes and thereby raise CMS's risk-adjusted payments. Allegedly Kaiser implemented a data-mining program called "refresh," where Kaiser would routinely mine patient medical files to find old diagnoses that had not been diagnosed in the current service year. If a physician failed to address any of these old diagnoses at a patient visit, Kaiser provided the physician with a list of these "missed opportunities"—i.e., opportunities for risk-adjustment payment—to create an addendum to retrospectively add these diagnoses to the medical record.[136]

Some of these additions were accomplished via "queries" inserted into the EMR, suggesting that the physician add a new diagnosis or renew an older diagnosis that, in fact, was since medically resolved. Indeed, "many times Kaiser would query physicians to add conditions whose existence at the time was contradicted by the medical record,

---

130 *Id.*

131 *Id.*

132 *Id.*

133 "Initially, Sutter and PAMF management ignored her and continued the RAF [risk adjustment] Campaign unabated. However, as Ormsby and her auditing team deleted false diagnosis codes that mapped to HCCs and negatively impacted the reimbursement from CMS, Sutter and PAMF management took steps to impede her efforts and stop her ability to delete false codes." *Id.* at 1040. The complaint then details ten patients and the specific ways in which their diagnosis codes were falsified. *See id.* at 1045-48.

134 U.S. Amended Complaint-in-Intervention, U. S. *ex rel.* Osinek v. Kaiser Permanente, No. 13-cv-03891-EMC et al. (N.D. Cal. filed Dec. 12, 2022).

135 *Id.*

136 *Id.* at ¶ 133–134.

but without even alerting the physicians to this contradictory information."[137]

To enhance physicians' compliance with these suggestions, Kaiser allegedly meticulously tracked and monitored these metrics across physicians, facilities, and regions. Physicians who scored high were praised and rewarded. Those who did not would often be required to meet with supervisors about their risk-adjustment performance and could face financial consequences. As each year drew to a close, some employees referred to Kaiser's rush to capture as many diagnoses as possible as the "dash for cash." Kaiser allegedly employed numerous other tactics, such as "coding parties" at which it would gather physicians in a room and expect them to work through lists of diagnoses and add these diagnoses to the records of their patient visits.[138]

Many of these diagnoses were added via "retrospective addenda to patient medical records with diagnoses that were unrelated to the medical visit and many times were contradicted by the patient's own medical record."[139] Moreover, Kaiser's "data mining" programs focused on identifying brand-new diagnoses, that is, diagnoses relating to conditions that no physician had ever diagnosed the patient as having. The programs identified these diagnoses using various algorithms that mined the patient's electronic medical records for key words, lab results, medications, clinical indicators, and other items that Kaiser

---

137 *Id.* at ¶ 10.

138 *Id.* at ¶ 11. These coding parties were "group coding sessions where data-quality trainers, and other similar Kaiser employees, would meet with physicians while the physicians coded their refresh lists. At these sessions, physicians would be expected to sit together, perhaps at lunch or after work with food and beverages provided by Kaiser, and work through their lists of specified diagnoses to add to patient visits." *Id.* at ¶ 206.

139 *Id.* at ¶ 13. In the end, allegedly "Kaiser fraudulently added hundreds of thousands of false diagnoses to the medical records of unrelated patient visits. Kaiser's internal analyses reflected that although Medicare accounted for roughly 10% of Kaiser's members, Medicare accounted for more than 30% of Kaiser's total revenue. And risk-adjustment payments (i.e., CMS payments based upon risk-adjustment diagnoses) accounted for more than half of all of Kaiser's Medicare revenue." *Id.* at ¶¶ 103, 105. Purportedly, many of these diagnostic addenda "fail[ed] to properly account for contradictory information in a patient's medical file, especially with respect to the patient visit at issue. Consequently, even if the medical record indicated the condition was historical or otherwise resolved, or documented clinical indicators that contradicted the current existence of the condition, Kaiser would often still query the physician after a visit to create an addendum to add the diagnosis." *Id.* at ¶ 137.

believed might be suggestive of potential diagnoses that would increase risk-adjustment payments.[140]

United States ex rel. Silingo v. WellPoint, Inc.[141]

A third FCA action adds a twist. *United States ex rel. Silingo v. WellPoint, Inc.*[142] does not directly implicate an MAO, but rather a third party, Mobile Medical Examination Services, Inc. (MedXM), that "contracted with the defendant [MAOs] to provide up-to-date diagnosis codes and medical documentation for enrollees who otherwise may not have had an eligible medical encounter during a calendar year,"[143] and whose risk designations were about to expire and re-set to an unadjusted risk average.[144]

According to relator Anita Silingo, a former Compliance Officer and Director of Provider Relations for the company, MedXM sent out "mostly nurse practitioners and physician assistants [who] were not legally authorized to make conclusive medical diagnoses, so their examinations could not support the risk adjustment data that was submitted,"[145] then had physicians fraudulently endorse these providers without having properly supervised their work.[146]

As part of the alleged scheme,

> "MedXM systematically fabricated complex diagnoses that its medical examiners could not have possibly confirmed during an in-home assessment. The complaint identifies a variety of ailments—such as chronic obstructive pulmonary disease, hepatitis, and inflammatory bowel disease—that allegedly cannot be diagnosed without a spirometry test, biopsy, follow-up blood test, or other invasive procedure that MedXM's examiners were unequipped and unauthorized to perform in a person's home. Instead, Silingo alleges, MedXM's medical examiners and coders

---

140 *Id.* at ¶ 140. Another dimension to Kaiser's program was an "initiative in 2012 to focus on four 'key conditions': protein calorie malnutrition, diabetes with neurological manifestations, aortic atherosclerosis, and chronic kidney disease. . . . Kaiser expected each facility in the region to hit a specified prevalence rate for each condition." *Id.* at ¶ 150.

141 U. S. *ex rel.* Silingo v. WellPoint, Inc., 904 F.3d 667 (9th Cir. 2018).

142 *Id.*

143 U. S. *ex rel.* Silingo v. WellPoint, Inc., 904 F.3d 667, 674 (9th Cir. 2018).

144 *Id.*

145 *Id.*

146 *Id.*

> simply recycled prior diagnoses and medical histories in the updated health assessment reports."[147]

Moreover, much of this "diagnostic information" was produced in the absence of the face-to-face encounters required by CMS.[148] And although home assessments typically took about forty-five minutes outside of travel time, MedXM's staff consistently reported over fifteen assessments per day, sometimes as many as twenty-five.[149] Worse, MedXM allegedly edited, even forged, medical records to incorporate more lucrative diagnoses.[150]

The foregoing comes from just three cases. As noted, eight of the ten largest participating insurers have now been accused of engaging in dubious practices such as the above, to elevate their populations' risk assessments.[151]

V. IMPLICATIONS FOR AI IN HEALTHCARE

In sum, errors enter EMRs in diverse ways, both intentional and unintentional. Our next question must be, "so what?" It may be tempting to assume that the EMR errors described above somehow don't matter, that they are random and will eventually "all come out in the wash," or that somehow the various AI applications will sort through the "chaff" to find the proverbial "wheat."

Aside from fallaciously begging the question, such optimism would fly in the face of current experience. Although machine learning as applied to static images such as CT or MRI scans is becoming more accurate, the LLMs that we hope will provide clinical decision support and relieve documentation burdens do not yet appear capable of doing so. "[T]he essence of efficient knowledge retrieval is to ask the right

---

147 *Id.* at 674–75.

148 *Id.* at 675.

149 *Id.*

150 *Id.*

151 Abelson & Sanger-Katz, *supra* note 112. The report continues: MedXM's "doctors and nurses were pushed to document a range of diagnoses, including some—vertebral fractures, pneumonia and cancer—they lacked the equipment to detect, according to a whistle-blower lawsuit. According to the lawsuit, employees who drew patients' blood often were not provided with a centrifuge or cooler; spoiled blood analyzed a day later produced strange results that could be used to justify valuable diagnoses, including kidney disease and leukemia." *Id.*

questions, and the art of critical thinking rests on one's ability to probe responses by assessing their validity against models of the world. LLMs can perform none of these tasks."[152]

More fundamentally, "machine learning methods can only be as good as the information in the training set."[153] "[T]he quality of diagnostic data in the EHR will directly determine the accuracy of the AI-generated diagnosis."[154] "[A] prediction rule is only as good as the administrative data used in its development."[155] Moreover, because much of AI functions in a "black box,"[156] finding patterns that do not necessarily conform to human reasoning, the actual impact of EMR errors may never be known, or even knowable, to us.

Accordingly, we must conclude that most likely it does matter whether the EMRs on which AI applications rely as a major database, are error-laden. Unfortunately, adequately spotting and correcting errors would likely be difficult if not impossible to do, and enormously costly.

---

152 Harrer, *supra* note 6, at 3.

153 Liu, *supra* note 6, at 1812.

154 Kulkarni & Singh, *supra* note 6, at 318; *See also* Chan, *supra* note 6, at 364 ("In using black box clinical AI systems, physicians and hospitals 'place trust not only in the equation of the model, but also in the entire database used to train it and, in the handling (e.g. labelling) of that database by the designers.'"); *See also* Harrer, *supra* note 6, at 2 ("[I]f models have been trained on a vast corpus of internet data with limited filtering (as for example is the case for ChatGPT or stable diffusion), they have ingested facts as much as misinformation, biased content as much as fair content, harmful materials as much as harmless ones. Without a means to assess any of these criteria before answering a prompt, LLMs are at risk—and numerous examples have demonstrated they have fallen to it—of reproducing, amplifying, and disseminating problematic content and misinformation.").

155 Teno, *supra* note 6, at 1.

156 Claudia E. Haupt & Mason Marks, *AI-Generated Medical Advice—GPT and Beyond,* 329 JAMA 1347, 1349–50 (2023); *see also* Pranay Rajpurkar & Matthew Lungren, *The Current and Future State of AI Interpretation of Medical Images*, 388 N. ENGL. J. MED. 1981, 1985 (2023); *see also* Duffourc & Gerke, *supra* note 17; *see also* Chan *supra* note 6, at 356; *see also* Schiff & Borenstein, *supra* note 20, at 139–40; *see also* Hannah R. Sullivan & Scott J. Schweikart, *Are Current Tort Liability Doctrines Adequate for Addressing Injury Caused by AI?*, 21 AMA J. ETHICS E160; *see also* Chen et al., *supra* note 14 at 709; Hunter & Holmes, *supra* note 30, at 1215.

A. Correcting EMR errors

1. Unintentional errors

Where errors stem from inadequate clinical skills—the clinician's ability to do proper history, physical and diagnosis[157]—errors will not ordinarily be caught without a second pair of eyes on both the patient and the documentation.[158] This might happen via an autopsy or a second opinion, adverse outcome, or malpractice suit. However, such redundancy is not the stuff of routine clinical care, hence these are limited as a means for spotting and correcting errors. Additionally, corrections via these avenues presuppose that the second opinion is correct, which may or may not be the case.

When we turn from clinical mistakes to documentation errors — as when a patient's HSV (herpes simplex virus) is somehow recorded as "HIV," or a scribe mis-hears the patient's words, or where copy/paste entries have become inaccurate over time—it is difficult to imagine any vehicle for exhaustively finding and reliably correcting such errors. Where patients identify errors in their own medical record, for example, it will often be difficult to discern, in a given instance, whether the clinician or the patient (or neither) is right. While patients can make important, often valid observations, they also may simply misunderstand what happened, or mis-recall, or may not have been adequately informed about their diagnoses and treatments. And patients sometimes balk at making corrections out of concern not to irritate the physicians on whom they depend for care, or in a belief that the effort would be unlikely to produce any real change.[159] Moreover, if given a chance to correct the record, both patients and providers may have their own ulterior agendas, such as to influence insurance payments or to aid in tort litigation.

Scribe and AI note-writing may make the problem even worse. Despite recent efforts to streamline documentation, medical records are more bloated than ever.[160] At some point the sheer time required

---

157 Kulkarni & Singh, *supra* note 6, at 317.

158 *See id.*

159 Lear et al., *supra* note 66, at 2, 13, 15.

160 Kersten Bartelt et al., *Two Years After Coding Changes Sought to Decrease Documentation, Notes Remain 'Bloated'* EPIC RSCH. (July 6, 2023), https://www.epicresearch.org/articles/two-years-after-coding-changes-sought-to-decrease-documentation-notes-remain-

to read a given EMR for purposes of detecting and correcting errors becomes prohibitive, particularly when a change is solely to correct documentation slip-ups that are unlikely to harm the patient's care.[161] Moreover, for any given note describing history, physical, progress or consult, discerning which entries are errors and which version is correct will often be impossible, given the difficulties of retrospectively verifying whether the physician misidentified a physical finding, the patient misreported his illness's history, a scribe misunderstood the physician's or patient's statement, or an AI "scribe" could not decode a patient's thick accent.

2. Intentional errors

Even if we suppose that unintentional errors are randomly distributed and that somehow they will not adversely affect AI applications, it would be entirely too facile to assume, without evidence, that dishonesty is likewise distributed randomly, or that its effects on AI datasets would somehow be auto-nullified. Gaming, lying, and cheating are not random. They are deliberate. And the tactics a payor system, PE corporation or physician group might use to maximize revenues, or even to help patients, cannot be presumed to be identical from one physician to the next, or one organization to the next, nor even static over time within an organization.

Neither can we assume that AI (currently still producing hallucinations) will reliably "figure out" truth from fiction.[162] Although an AI application may be able to discern internal contradictions within an EMR, it may not have adequate basis for discerning which version—if either—is correct. Moreover, when an MAO infuses factually unsupported diagnoses throughout its members' EMRs year after year, an AI

---

bloated; *see also* Hannah Nelson, *EHR Clinical Note Length Continues to Grow Despite Coding Changes*, EHR INTEL. (July 17, 2023), https://ehrintelligence.com/news/ehr-clinical-note-length-continues-to-grow-despite-coding-changes.

161 Anecdotally, clinicians have reported that one of these technologies, as implemented in an Emergency Department setting, produces numerous pages of "garbage," topped off with a somewhat useful one-paragraph summary. Private communication on file with the author. Given the time-pressures on physicians, there is room to doubt whether many will have or take the time to read "pages and pages" and correct every error, where it is unlikely those errors will directly hurt that patient's care.

162 *See A Recent Study Sheds Light on AI Hallucinations and What Society Thinks of Them*, DIGIT. INFO. WORLD (Sept. 13, 2023, 11:00 PM), https://www.digitalinformationworld.com/2023/09/a-recent-study-sheds-light-on-ai.html.

algorithm could "learn" to associate bogus diagnoses with whatever else is in the EMR, even though in reality those other factors may be medically noncontributory. AI could thus deem such noncontributory factors to be causal rather than incidental.

Also, as shown above, MAOs' alleged tactics vary widely for ramping up diagnoses in members' EMRs. Hence, even if AI could somehow decode one MAO's discrepancies to sort fact from fiction, that learning will not likely apply elsewhere. And even when an MAO's unlawful risk amplifications are caught in a False Claims Act (FCA) suit, the MAO may refund overpayments, pay a penalty and promise to sin no more. Even then, many if not most of the earlier erroneous entries will remain, still corrupting the EMRs. Settlements are typically based on extrapolations from reviewing a limited number of medical records, not a fine-toothed scrutiny of every patient in the MAO.[163] Thus, even a successful FCA action will do little or nothing to correct erroneous EMRs.

Finally, we cannot assume that these errors will be harmless, or somehow will all "come out in the wash." We already know that "dataset shift," the "mismatch between the data set with which an AI system was developed and the data on which it is being deployed,"[164] can cause significant degradation of AI performance.[165] Surely, we could not expect the problems to be any less when the dataset not merely contains errors, but where those errors vary widely from one group of patients to the next, whether by inadvertence or by deliberate contrivance.

B. Consequences of EMR errors

As we have seen, AI can be used to as clinical decision support (CDS) to assist diagnosis, prognosis and treatment.[166] If based on error-laden EMRs, there may be little way to figure out where and how the

---

163 United Healthcare Ins. Co. v Bacerra, 16 F.4th 867, 878 (D.C. Cir. 2021).

164 Andrew Beam et al., *Artificial Intelligence in Medicine,* 388 NEW ENG. J. MED. 1220, 1221 (Mar. 29, 2023) (citing Finlayson et al., *The Clinician and Dataset Shift In Artificial Intelligence*, 385 NEW ENG. J. MED. 283, 283–86 (July 14, 2021); *see also* Parikh et al., *supra* note 27, at 2377.

165 *See* Finlayson et al., *The Clinician and Dataset Shift in Artificial Intelligence*, 385 NEW ENG. J. MED. 283 (July 14, 2021); *see also* Parikh et al., *supra* note 27, at 2377.

166 Sobhan Moazemi et al., *Artificial Intelligence for Clinical Decision Support for Monitoring Patients in Cardiovascular ICUs: A Systematic Review*, 10 FRONT. MED., 1, 8, 20–21 (2023).

AI algorithm may have gone awry, until patients suffer unexpectedly poor outcomes, or when the AI recommendation is so obviously off-the-beam that the physicians rule it out *ab initio*.[167]

If EMR errors threaten AI's value for clinical care, they likewise portend problems on the payment side.[168] Increasingly, insurers rely on algorithms for issuing prior authorization and retrospective coverage decisions.[169] Indeed, "CMS aims to transition to a fully electronic submission and initial-determination system for prior authorization, with separate user interfaces for patients and clinicians and a payer-to-payer exchange."[170]

Such electronic formatting does not, of itself, entail that AI will then be applied to payment and coverage decisions. However, the transition is now under way. AI-governed pre-authorization decisions, for instance, have already sparked litigation. A proposed class action suit filed in late 2023 accuses UnitedHealth Groups of using an AI system to deny elderly patients access to post-hospital nursing facility stays.[171] "The lawsuit centers on an AI algorithm known as nH Predict developed by NaviHealth Inc, a company acquired by UnitedHealth in 2020. It claims that UnitedHealth relies on nH Predict to evaluate claims for so-called post-acute care, which includes stays in skilled nursing facilities and in-home care."[172]

---

167 *See* Teno, *supra* note 6, at 1.

168 *See* Daniel Payne, *AI's big test: Making Sense of $4 trillion in Medical Expenses,* POLITICO (Dec. 31, 2023, 7:00 AM) https://www.politico.com/news/2023/12/31/ai-medical-expenses-00132557.

169 *See id.*

170 Michael Kyle & Zirui Song, *The Consequences and Future of Prior-Authorization Reform*, 389 NEW. ENG. J. MED. 291 (2023).

171 Plaintiff's Class Action Complaint at ¶ 1, Estate of Gene B. Lokken v. United Health Grp., Inc., 0:23-cv-03514 (D. Minn. filed Nov. 14, 2023); *see also* Brendan Pierson, *Lawsuit claims UnitedHealth AI Wrongfully Denies Elderly Extended Care*. REUTERS, (Nov. 14, 2023), https://www.reuters.com/legal/lawsuit-claims-unitedhealth-ai-wrongfully-denies-elderly-extended-care-2023-11-14/; *see also* Ken Alltucker, *Is Your Health Insurer Using AI to Deny You Services? Lawsuit Says Errors Harmed Elder*s, USA TODAY (Nov. 20, 2023, 11:19 AM), https://www.usatoday.com/story/news/health/2023/11/19/unitedhealth-artificial-intelligence-denies-claims-in-error-lawsuit-says/71579822007/; *see also* Michelle Mello & Sherri Rose, *Denial—Artificial Intelligence Tools and Health Insurance Coverage Decisions*, 5(3):e240622 JAMA HEALTH FORUM 1, 1-3 (2024).

172 Pierson, *supra* note 171; Nona Tepper*, Latest Medicare Advantage AI lawsuit targets Humana,* MOD. HEALTHCARE (Dec. 13, 2023, 10:46 AM) https://www.modernhealthcare.com/legal/humana-ai-medicare-advantage-lawsuit-unitedhealth-cigna (noting a

For after-the-fact payment decisions, Cigna has been sued in a class action because, among other allegations, its coverage decisions relied on an algorithm rather than a clinician.[173] Allegedly during

> "two months in 2022, Cigna doctors denied over 300,000 requests for payments using this method, spending an average of just *1.2 seconds* 'reviewing" each request . . .The Cigna Defendants utilize the PXDX system, which employs an algorithm to identify discrepancies between diagnoses and what the Cigna Defendants consider acceptable tests and procedures for those ailments and automatically deny claims on those bases. After the PXDX system denies claims, Cigna doctors then sign off on the denials in batches without opening each patient's files to conduct a more detailed review . . ."[174]

We also see MAOs, like FFS insurers, using algorithms to decide what coverage to offer for the following year, as well as for current patients. For example, UnitedHealth Group, which offers one of the largest MAOs, uses its subsidiary naviHealth's proprietary

> "'nH Predict' tool [which] sifts through millions of medical records to match patients with similar diagnoses and characteristics, including age, preexisting health conditions, and other factors. Based on these comparisons, an algorithm anticipates what kind of care a specific patient will need and for how long . . . . [Coincidentally, t]he tool often predicts a patient's date of discharge, which coincides with the date their insurer cuts off coverage, even if the patient needs further treatment that government-run Medicare would provide."175

Nursing home care is these plans' most frequently denied service,[176] and the problem has become sufficiently concerning that the Federal government has issued new rules, effective January 2024, that will

> "rein in [MCOs'] use of algorithms in coverage decisions. Insurance companies using such tools will be expected to 'ensure that they are making medical necessity determinations based on the circumstances of

---

similar suit has been filed against Humana).

173 Class Action Complaint at ¶ 1, Kisting-Leung v. Cigna, 2:23-cv-01477-DAD-KJN, (E.D. Cal. filed July 24, 2023).

174 *Id.* at ¶¶ 1, 19.

175 Susan Jaffe, *Feds Rein In Use of Predictive Software That Limits Care for Medicare Advantage Patients,* KAISER FAM. FOUND. HEALTH NEWS (Oct. 5, 2023), https://kffhealthnews.org/news/article/biden-administration-software-algorithms-medicare-advantage/.

176 *Id.*

> the specific individual,' the requirements say, 'as opposed to using an algorithm or software that doesn't account for an individual's circumstances.'"[177]

And yet even if AI, rather than potentially corruptible humans, produces such coverage guidance, it can go seriously awry if based on error-laden EMRs.

Finally, we also now see proposals to use AI for basic risk adjustment,[178] again perhaps permitting AI rather than financially incentivized humans to design the system. Here, too, we inquire whether AI based on erroneous EMRs will produce acceptable results.

VI. EFFORTS TO ADDRESS THE ISSUES

Notwithstanding the hype and optimism surrounding AI in healthcare,[179] the watchword at this point surely must be "caution." As noted by Hunter and Holmes, "the very features that make AI such a valuable additional tool for data analysis are the same ones that make it vulnerable from a statistical perspective."[180] Moreover, "[t]he limitations of LLMs are systemic: simply boosting training data sizes and the number of model parameters to create future versions of the same model architectures will not overcome their shortcomings but rather amplify them."[181]

As discussed in Part II, many commentators have already voiced a host of challenges associated with using AI in clinical healthcare. This article supplements those discussions by highlighting a problem heretofore inadequately recognized: errors throughout EMRs, both

---

177 *Id.*; Medicare Program; Contract Year 2024 Policy and Technical Changes to the Medicare Advantage Program, Medicare Prescription Drug Benefit Program, Medicare Cost Plan Program, and Programs of All-Inclusive Care for the Elderly, 88 Fed. Reg. 22120, 22195 (Apr. 12, 2023) (to be codified at 42 C.F.R. pts. 417, 422, 423, 455, and 460).

178 *See* Gary Weissman & Karen Maddox, *Guiding Risk Adjustment Models Toward Machine Learning Methods*, 330 JAMA 807 (2023).

179 *See id.* (optimism regarding better risk adjustment); *see also* Harris, *supra* note 17, at 801–02 (optimism regarding AI's ability to enhance EMRs); *see also* Gottlieb & Silvis, *supra* note 15.

180 Hunter & Holmes, *supra* note 30, at 1211–12 (cautioning that "Given the opaqueness of generative AI foundation models, additional caution is needed for their use in health applications).

181 Harrer, *supra* note 6, at 4.

intentional and unintentional. They are not rare. Neither can they be presumed harmless.

That said, it is encouraging that at least some remediation efforts are now afoot. The problem of flawed AI datasets is gaining recognition outside of healthcare. In October 2022, the federal Office of Science and Technology issued a policy entitled *Blueprint for an AI Bill of Rights*.[182] Discussing AI hazards across a wide range of non-healthcare areas—from policing, to finance, housing, employment practices, national security, transportation, public services and beyond—the Policy expressly notes that significant problems can arise from inappropriate, low-quality or irrelevant data,[183] as well as from data entry errors.[184]

Additionally, in light of the fact that AI innovation appears to be outpacing regulatory oversight, the Food and Drug Administration (FDA) and the Office of the National Coordinator for Health Information Technology (ONC) have been issuing updated guidance, with additional updates expected.[185]

In sum, we may currently be getting too far out over our skis. Physicians, hospitals, and payors who use AI implicitly "place trust not only in the equation of the model, but also in the entire database used to train it and, in the handling (e.g. labelling) of that database by the

---

182 WHITE HOUSE OFF. OF SCI. AND TECH. POL'Y, BLUEPRINT FOR AN AI BILL OF RIGHTS: MAKING AUTOMATED SYSTEMS WORK FOR THE AMERICAN PEOPLE (2022), https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf.

183 *Id.* at 19–20.

184 *See id.* at 42.

185 Deixler et al., *supra* note 19, at 1–2; Goodman et al., *supra* note 19, at 484; *See, e.g.*, U.S. FOOD DRUG ADMIN., GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF: SOFTWARE AS A MEDICAL DEVICE (SAMD): CLINICAL EVALUATION 1, 6–7 (2017); *see also, e.g.*, U.S. FOOD DRUG ADMIN., GUIDANCE FOR INDUSTRY AND FOOD AND DRUG ADMINISTRATION STAFF: DECIDING WHEN TO SUBMIT A 510(K) FOR A SOFTWARE CHANGE TO AN EXISTING DEVICE (2017); *see also, e.g.*, *Public Workshop – Evolving Role of Artificial Intelligence in Radiological Imaging*, U.S. FOOD DRUG ADMIN., https://www.fda.gov/medical-devices/workshops-conferences-medical-devices/public-workshop-evolving-roleartificial-intelligence-radiological-imaging-02252020-02262020 (last visited Mar. 1, 2024); *see also*, *e.g.*, U.S. FOOD DRUG ADMIN., ARTIFICIAL INTELLIGENCE/MACHINE LEARNING (AI/ML)-BASED SOFTWARE AS A MEDICAL DEVICE (SAMD) ACTION PLAN 1, 2 (2021), https://www.fda.gov/media/145022/download; *see also*, *e.g.*, Health Data, Technology, and Interoperability: Certification Program Updates, Algorithm Transparency, and Information Sharing, 88 Fed. Reg. 23746, 23775 (Apr. 18, 2023) (to be codified at 45 C.F.R. pts. 170, 171).

designers."[186] Clearly, that trust is not yet warranted. "Nevertheless, predictive models have frequently been deployed in health care settings without transparency or independent validation . . ."[187] After all, enormous amounts of money stand to be made in developing and selling AI, and Silicon Valley is showing a hearty appetite.[188]

Recognizing the problem aloud, as this Article does, is a crucial first step. Failure to do so risks avoidable harms to patients, avoidable financial waste, and potential liability not just for those who develop and recommend—or sell—AI-developed algorithms, but for health care providers who implement them uncritically.[189]

---

[186] Chan, *supra* note 6, at 364.

[187] Lu et al., *supra* note 24, at 2.

[188] Erin Griffith & Cade Metz, *A New Area of A.I. Booms, Even Amid the Tech Gloom*, N.Y. TIMES (Jan. 7, 2023) https://www.nytimes.com/2023/01/07/technology/generative-ai-chatgpt-investments.html.

[189] Chan, *supra* note 6, at 352 (citing Casey Ross & Ike Swetlitz, *IBM Pitched Its Watson Supercomputer as a Revolution in Cancer Care. It's Nowhere Close*, STAT (Sept. 5, 2017), https://www.statnews.com/2017/09/05/watson-ibm-cancer/); Price et al., *Potential Liability for Physicians Using Artificial Intelligence*, 322 JAMA 1765, 1765–66 (2019); Joann G. Elmore & Christoph I. Lee, *Artificial Intelligence in Medical Imaging—Learning From Past Mistakes in Mammography*, 3 JAMA HEALTH FORUM 1, 2 (2022); Liu et al., *How to Read Articles That Use Machine Learning Users' Guides to the Medical Literature*, 322 JAMA 1806, 1815 (2019).