Seattle Journal of Technology, Environmental, & Innovation Law

Volume 15 | Issue 1 Article 6

12-20-2024

Defining the Social Licence of Large Language Models in Healthcare

Robin L. Pierce
University of Exeter, Law School

Jack Gallifant
Massachusetts Institute of Technology

Ashley Cordes
University of Oregon

Amelia Fiske
Technical University of Munich, School of Medicine

Matilda Dorotic
BI Norwegian Business School

See next page for additional authors

Defining the Social Licence of Large Language Models in Healthcare

Authors

Robin L. Pierce, Jack Gallifant, Ashley Cordes, Amelia Fiske, Matilda Dorotic, Mataroria Lyndon, Shrey Jain, Joe Zhang, Judy Gichoya, and Leo Anthony Celi

Defining the Social Licence of Large Language Models in Healthcare

Authors:
Jack Gallifant [1, 2]
Ashley Cordes [3, 4]
Amelia Fiske [5]
Matilda Dorotic [6]
Mataroria Lyndon [7]
Shrey Jain [8, 9]
Joe Zhang [2, 10]
Judy Gichoya [11]
Leo Anthony Celi [1, 12, 13]
Robin L. Pierce [14 *]

[1] Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Laboratory for Computational Physiology, Cambridge, MA, USA.
[2] Department of Critical Care, Guy's & St Thomas' NHS Trust, London, UK.
[3] University of Oregon, Indigenous Studies in Environmental Studies and English, Eugene, OR, USA.
[4] Coquille/KōKwel Indian Tribe, Southwest Oregon, USA.
[5] Institute of History and Ethics in Medicine, School of Medicine, Technical University of Munich, Munich, Germany.
[6] Department of Marketing, BI Norwegian Business School, Norway.
[7] Centre for Medical and Health Sciences Education, Faculty of Medical and Health Sciences, The University of Auckland, New Zealand.
[8] University of Toronto, Faculty of Engineering, Toronto, Canada.
[9] Microsoft Research, Seattle, Washington, USA.
[10] Institute of Global Health Innovation, Imperial College London, London, UK.
[11] Department of Radiology, Emory University School of Medicine, Atlanta, Georgia.
[12] Division of Pulmonary, Critical Care, and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA.
[13] Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA.
[14] The Law School, Faculty of Humanities, Arts, and Social Sciences, University of Exeter, Exeter, United Kingdom.
* Corresponding author
Correspondence:
Robin Pierce
Law School
University of Exeter
Exeter
EX4 4RJ
United Kingdom
Email (preferred): R.P.Pierce@exeter.ac.uk
Phone: (413) 692 9888

Table of Contents

I. DEFINING THE SOCIAL LICENCE OF LARGE LANGUAGE MODELS

As Large Language Models (LLMs) continue to grow and evolve, the ethical, social, and legal challenges they present, especially in terms of data use and sharing, come sharply into focus. Particularly in high-risk domains like healthcare, where the sensitivity of data is pronounced, and the ramifications of data leakages and misuse are profound. These licences, envisioned to address the complexities introduced by LLMs, still grapple with barriers that hinder their full potential: notably, the absence of authoritative regulatory bodies, ambiguities in defining what constitutes "sufficient public benefit," issues of data sovereignty, and dilemmas surrounding responsibility and liability. Herein, we delve into these intricacies, focusing on the evolution of social licences, their relationship with prevailing governance model challenges, and the essential transformations required to ensure their safe and responsible deployment in the rapidly evolving world of LLMs.

Large Language Models (LLMs) are neural network language models such as the LLM chatbot released by OpenAI, ChatGPT.[15] LLM chatbot's impressive abilities are a result of its extraordinary computing power and training data–available to organizations that can pay for these resources. Healthcare holds a unique position within the AI landscape due to the sensitive nature of health data and the gravity of consequences that could be realized from misuse, mismanagement, and misspecification of LLMs. The accompanying ethical, legal, and social dilemmas that have arisen in the healthcare domain are increasingly prevalent and are heightened due to the lack of transparency in the field, which has resulted in a sense of awe that is shrouded in fear.[16] This has led to increasing recognition of the need for effective governance of LLMs across sectors and, notably, in healthcare. A public letter from U.S. Sen. Mark R. Warner (D-VA), Chairman of the Senate Select Committee on Intelligence, to Google CEO Mr Sundar Pichai sharing concerns regarding reports of LLM deployment in U.S. hospitals earlier this year.[17]

This letter identifies several key issues, from the commonly quoted fears of AI hallucinations and bias to deep concerns regarding the lack of transparency surrounding the management of protected health information.[18] While LLMs typically use vast quantities of training data

---

[15] *Introducing ChatGPT*, OPENAI (Nov. 30, 2022), https://openai.com/index/chatgpt/ [https://perma.cc/C7PC-LYRE].

[16] Bertalan Meskó & Eric J. Topol, *The Imperative for Regulatory Oversight of Large Language Models (or Generative AI) in Healthcare*, 6 NPJ DIGIT. MED. 120 (2023); Christopher Ryan Maboloc, *Chat GPT: The Need for an Ethical Framework to Regulate Its Use in Education*, 46 J. PUB. HEALTH 152 (2024).

[17] Press Release, U.S. Sen. Mark Warner, Warner Urges Google to Protect Patients and Ensure Ethical Deployment of Health Care AI (Aug. 8, 2023), https://www.warner.senate.gov/public/index.cfm/2023/8/release-warner-urges-google-to-protect-patients-and-ensure-ethical-deployment-of-health-care-ai [https://perma.cc/7ZXN-VENG] [hereinafter *Warner Urges Google*]; Press Release, U.S. Sen. Mark Warner, Warner Calls on AI Companies to Prioritize Security and Prevent Malicious Misuse (Apr. 26, 2023), https://www.warner.senate.gov/public/index.cfm/2023/4/warner-calls-on-ai-companies-to-prioritize-security-and-prevent-malicious-misuse [https://perma.cc/7ZXN-VENG].

[18] *Warner Urges Google*, *supra* note 17.

from publicly available sources, such as Wikipedia, publicly available data has been used by private companies for years; what has changed is the scale at which these models can capture and integrate such vast quantities of data.[19] Simple social media posts about health updates or blogs on starting a new job in a particular location can be attached to sensitive characteristics about each person, including those inferred from media such as photos. Furthermore, chatbots introduce a new challenge since the processing of data that is entered into ChatGPT by users may be used for subsequent training. Moreover, it remains unclear how this data is managed, whether private information is removed, and how this affects a user's right to erasure.[20]

The rapid uptake of LLMs has raised issues regarding the ethical nature and permissibility of data collection techniques often used to train LLMs that involve "scraping" publicly available data from the internet. This, in turn, raises the question of whether there is a social licence to collect this data about individuals in large volumes for use in LLMs, given that a social licence is generally understood as use of people's data for socially beneficial purposes, e.g. health research, without obtaining individual consent for the collection of that data. The success of LLMs clearly moves beyond individual concepts of social licence due to the "mass effect" of subject data being picked up that allows spurious and biased associations to be made about the society that an individual is in. This has created a novel situation, unique from anything before and therefore worth explicating and addressing. In this piece, we discuss the development of social licences, how this relates to current barriers to effective governance models, and highlight the necessary developments that must be established for the safe and responsible deployment of LLMs.

II. FINDING THE MEANS TO MEET GROWING DATA REQUIREMENTS

The AI boom has benefited from and relied upon the increasingly available data produced throughout the 21st century; LLMs sit squarely within this phenomenon, with LLAMA2, a state-of-the-art language model, trained on 2 trillion tokens of text data.[21] However, the growing AI economy is demanding more. As a result, increasing concern arises about maintaining agency over the data one produces in the course of daily living – by sending emails, seeking health care, and purchasing goods over the Internet. This has led to a range of proposals addressing issues pertaining to open data and the resulting social benefits and risks of data use.[22]

---

[19] Hugo Touvron et al., *Llama 2: Open Foundation and Fine-Tuned Chat Models*, ARXIV CORNELL UNIVERSITY (Jul. 18, 2023), http://arxiv.org/abs/2307.09288 [https://perma.cc/8VCD-FHD6]; Meskó & Topol, *supra* note 16.

[20] Dawen Zhang et al., *Right to Be Forgotten in the Era of Large Language Models: Implications, Challenges, and Solutions*, ARXIV CORNELL UNIVERSITY (June 5, 2024), https://arxiv.org/pdf/2307.03941 [https://perma.cc/Y5DC-4L5S].

[21] Touvron et al., *supra* note 19.

[22] Theresa Xie & Isaiah Portiz, *ChatGPT Creator OpenAI Sued for Theft of Private Data in 'AI Arms Race,'* BLOOMBERG (June 28, 2024, 4:15 PM), https://www.bloomberg.com/news/articles/2023-06-28/chatgpt-creator-sued-for-theft-of-private-data-in-ai-arms-race [https://perma.cc/9ZHY-KUHE].

Initial data protection regulations in America and Europe were designed to establish data rights, such as the right to erasure[23], which grants data subjects a right to removal of inaccurate or irrelevant online data about them and protect personal information. For example, the General Data Protection Regulation (GDPR) in Europe and HIPPA in the U.S. aim to safeguard personal data, with HIPPA focusing specifically on health data. The U.S. regulatory framework for privacy in healthcare, as exemplified by HIPAA represents a sector-specific approach to data protection.[24] However, HIPAA was enacted in 1996, long before the advent of digital healthcare records and advanced internet technologies. As a result, the legislation may be increasingly outmoded in today's digital landscape.[25] Further, in addition to personal data (defined as data pertaining to an identified or identifiable person), the GDPR also recognizes "special categories of personal data".[26] Under the GDPR, the processing of "sensitive data", those revealing health status, political opinions, religious beliefs, or race or ethnicity, is prohibited unless it falls under one of the designated exceptions. However, the category of such personal data is increasingly broad given that IP addresses, mobility data (data about the geographical location of a device), and consumer data can all lead to the identification of a specific individual.[27] Scholars have observed that health-related data has similarly expanded, as even datasets not specifically regarding health can be linked together, and identifying connections that can be used to infer health-related characteristics.[28]

New types of digital technologies, such as social media platforms, have resulted in complex challenges, forcing regulators to identify and provide further protection from newly configured risks of harm (cf., Digital Services Act in the EU and Online Safety Bill in the UK). Additionally, the European Union has proposed the European Health Data Space (EHDS), a health data sharing platform that would allow for the sharing of health data across member states.[29] It should be noted that the EHDS would not operate on an individual consent basis rooted in

---

[23] *Proposal for protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC*, COM (2016) 679 final [hereinafter GDPR].

[24] *See* 45 C.F.R. § 46 (2017) [hereinafter HHS Regulations]; *see* 45 C.F.R §160 (2000) [hereinafter Admin Requirements]; *see* 45 C.F.R § 164 (2000) [hereinafter Privacy and Security]; *Health Insurance Portability and Accountability Act of 1996 (HIPPA)*, CDC (Sept. 10, 2024), https://www.cdc.gov/phlp/php/resources/health-insurance-portability-and-accountability-act-of-1996-hipaa.html [https://perma.cc/A5VH-2EKQ].

[25] Mason Marks & Claudia E. Haupt, *AI Chatbots, Health Privacy, and Challenges to HIPAA Compliance*, 330 JAMA 309 (2023).

[26] GDPR, *supra* note 23.

[27] Nadezhda Purtova, *The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law*, 10 L. INNOV. TECH. 40 (2018); Amelia Fiske et al., *Value-Creation in the Health Data Domain: A Typology of What Health Data Help Us Do*, 18 BIOSOCIETIES 473 (2023).

[28] *See generally* Barbara Prainsack, *Research for Personalised Medicine: Time for Solidarity*, 36 MED. & L. 87 (2017), https://link.springer.com/chapter/10.1007/978-3-030-44062-6_7; BARBARA PRAINSACK & INE VAN HOYWEGHEN, *Shifting Solidarities: Personalisation in Insurance and Medicine*, *in* SHIFTING SOLIDARITIES: TRENDS AND DEVELOPMENTS IN EUROPEAN SOCIETIES 127 (Ine Van Hoyweghen, Valeria Pulignano, & Gert Meyers eds., 2020), https://doi.org/10.1007/978-3-030-44062-6_7 [https://perma.cc/BLE4-CT93].

[29] *Proposal for a Regulation on the European Health Data Space*, COM (2022) [hereinafter *Proposed European Health Data Space*].

individual data protection. The anticipated final version of the EU AI Act further promotes data sharing in its pro-innovation strategy through the development of "open ecosystems formed around European public supercomputers". These so called "AI Factories" would facilitate the development of generative AI models and applications, thus making it easier for small and medium-sized enterprise's (SME) to enter the market.[30] However, this strategy also implies increased reliance on "established data centres," which raises concerns about data security and privacy. In contrast, the U.S. has not taken such an aggressive approach to data sharing. However, medical research may prove to be an exception as sectoral regulatory approaches[31] to generative AI are crafted.[32] Because medical research may reasonably be regarded as a socially beneficial purpose and the practice of sharing health data for health research gains traction within research cultures, as exemplified by the NIH's Scientific Data Sharing Program[33] in the U.S. and NHS' Data and Clinical Record Sharing Programme in the UK, the development of health-specific LLM's for use in various aspects of healthcare could be seen as permissible use. Indeed, the scientific research exemption in the GDPR, which allows for the unconsented re-use of personal data for research purposes[34] could be interpreted to support such use by an LLM.

Despite the emergence of data protection legislation worldwide, often influenced by the GDPR, data protection laws leave many issues regarding the use and processing of personal data unresolved. Concerns about ownership and agency over the data produced arise as individual's digital traces are particularly intense among minority groups (based on race, gender, or ethnicity) and Indigenous Nations that have historically been marginalized by research and policy.[35] Principles of data sovereignty (rights over the use of one's personal data) and restrictions on data sharing and use have broader implications, particularly in the context of developing and implementing artificial intelligence (AI) tools. There are tensions between protecting individual and group rights, promoting inclusion, and ensuring AI tools are trained on diverse datasets. On the one hand, promoting inclusion to ensure that AI tools are trained on diverse data sets requires the collection of data from historically marginalized groups. A tension arises in that such inclusion could expose these groups to increased vulnerabilities due to any failings in data security or misuse of the data, including direct or indirect use of the data for discriminatory purposes. Thus, inclusion in the service of ensuring diverse data sets must

[30] Luca Bertuzzi, *LEAK: EU Commission Prepares 'Strategic Framework' to Boost AI Start-Ups, Generative AI Uptake*, EURACTIV (2024), https://www.euractiv.com/section/artificial-intelligence/news/leak-eu-commission-prepares-strategic-framework-to-boost-ai-start-ups-generative-ai-uptake/ [https://perma.cc/JPF3-DQ32].

[31] Data and Clinical Record Sharing, NHS ENGLAND (June 28, 2023), https://www.england.nhs.uk/long-read/data-and-clinical-record-sharing/ [https://perma.cc/43TU-J5ZG].

[32] Bertuzzi, *supra* note 30.

[33] *See generally* Scientific Data Sharing, NIH, https://sharing.nih.gov/ [https://perma.cc/CJC6-G5UL].

[34] GDPR, *supra* note 23.

[35] *See generally* GREGORY YOUNGING, ELEMENTS OF INDIGENOUS STYLE: A GUIDE FOR WRITING BY AND ABOUT INDIGENOUS PEOPLE (Brush Education Inc. ed., 2018).

be accompanied by safeguards for the protection of individual as well as group rights. The implications of excluding data from populations in the modelling that drives AI development could reduce health benefits and potentially expose them to harm. AI-driven health tools that are not trained on diverse data sets have demonstrated poorer performance on patients not represented in the training data, typically historically underrepresented groups. Consequently, enhanced performance on the majority population of AI diagnostic tools, may render less effective detection and diagnosis to members of underrepresented groups, resulting in harms such as undetected disease or late-stage diagnosis. These concerns are widely acknowledged in the healthcare setting, but remain without resolution, leaving the potential harms to accumulate.

III. THE EVOLUTION OF SOCIAL LICENSING FOR DATA SHARING

The modern concept of social licence for data sharing traces its early roots to bioethical debates in the late 1970s over whether unconsented research uses of people's health data can be ethically justified. Before the 1970s, using personal healthcare data in research was commonplace without obtaining individual consent. An influential set of Fair Information Practices from 1973 called for obtaining consent before using people's identifiable health data. These practices shaped the subsequent development of research and privacy policies in the United States and internationally.[36] As consent requirements to process personal data by firms and institutions became widely accepted, concerns grew that they might interfere with socially beneficial data uses or introduce harmful selection bias into research datasets.[37]

Bioethicists examined the circumstances under which unconsented data sharing could be ethically justified, if ever. While recognizing the need for unconsented data sharing in certain circumstances (e.g., to track an emerging epidemic, to detect child abuse, or to save another patient's life), bioethicists widely agree that unconsented use should proceed only when there is strong ethical justification. The "central ethical issue" is to ensure that the potential public benefits of proposed data use are sufficient

[36] Office of the Assistant Secretary for Planning and Evaluation, Transmittal Letter to Secretary Caspar W. Weinberger, Secretary of Health, Education, and Welfare (June 30, 1973), https://aspe.hhs.gov/reports/records-computers-rights-citizens [https://perma.cc/93A9-HZWL].

[37] BEYOND THE HIPPA PRIVACY RULE: ENHANCING PRIVACY, IMPROVING HEALTH THROUGH RESEARCH, (Sharyl J. Nass, Laura A. Levit, & Lawrence O. Gostin eds., 2009), http://www.nap.edu/catalog/12458 [https://perma.cc/TTU8-99AK]; Brian Buckley et al., *Selection Bias Resulting from the Requirement for Prior Consent in Observational Research: A Community Cohort of People with Ischaemic Heart Disease*, 93 HEART 1116 (2007), https://doi.org/10.1136/hrt.2006.111591 [https://perma.cc/32RG-KSD6]; Steven J. Jacobsen et al., *Potential Effect of Authorization Bias on Medical Record Research*, 74 MAYO CLIN. PROC. 330 (1999), https://www.sciencedirect.com/science/article/abs/pii/S002561961164398X [https://perma.cc/9PXM-3SS3]; *see generally* Jack V. Tu et al., *Impracticability of Informed Consent in the Registry of the Canadian Stroke Network*, 350 N. ENGL. J. MED. 1414 (2004), https://pubmed.ncbi.nlm.nih.gov/15070791/ [https://perma.cc/Q3W3-TRJK]; *see generally* S. H. Woolf et al., *Selection Bias from Requiring Patients to Give Consent to Examine Data for Health Services Research*, 9 ARCH. FAM. MED. 1111 (2000), https://triggered.edina.clockss.org/ServeContent?issn=1063-3987&volume=9&issue=10&spage=1111 [https://perma.cc/3NP3-DDDZ].

to warrant the burden on the individual's privacy rights.[38] Late in the 1970s, two U.S. federal advisory committees recommended that unconsented research use of personal medical records should be allowed only if "the importance of the research or statistical purpose for which any use of disclosure is to be made is such as to warrant the risk to the individual from additional exposure of the record or information contained therein," and if an Institutional Review Board (i.e., an ethics review body) ensures this condition is met.[39]

Unfortunately, efforts to include this "public-benefit" criterion as a condition for granting unconsented research access to data under the Common Rule[40] and Health Insurance Portability and Accountability Act (HIPPA) Privacy Rule[41]–two significant U.S. federal research and medical privacy regulations–ultimately stalled, in part because Institutional Review Boards expressed concern that they were not qualified to assess the broader public benefits of research and weigh them against the individual's rights.[42]

Debates about the ethics of unconsented data use eventually recognized that individual consent may be insufficient and inappropriate to protect harm to the public or to a particular group. As a result, the modern concept of social licence for data sharing is much broader than the earlier debate about the ethics of *unconsented* data sharing. Social licence recognizes that *even consented* data uses can have harmful public consequences because individuals' autonomous choices do not always ensure good societal outcomes. The bioethics of the late 20th century embraced an individualized vision of autonomy as the highest moral good, neglecting principles of caring, social interdependency, justice, and equity.[43] Yet, individuals' privacy is, in fact, interdependent, and one

---

[38] *See generally* Laura E. Bothwell, Annika Richterich, & Jeremy Greene, *Bioethical Issues in Pharmacoepidemiologic Research*, *in* TEXTBOOK OF PHARMACOEPIDEMIOLOGY 276 (Brian L. Strom MD, MPH, Stephen E. Kimmel MD, MSCE, Sean Hennessy PharmD, PhD, eds., 2021), https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119701101.ch16; *see generally* BRIAN L. STROM MD, MPH, STEPHEN E. KIMMEL MD, MSCE, SEAN HENNESSY PHARMD, PHD, TEXTBOOK OF PHARMACOEPIDEMIOLOGY (Brian L. Strom MD, MPH, Stephen E. Kimmel MD, MSCE, Sean Hennessy PharmD, PhD eds., 4th ed. 2021); *see generally* National Bioethics Advisory Commission, *Ethical and Policy Issues in Research Involving Human Participants* (2001), https://scholarworks.iupui.edu/handle/1805/25 [https://perma.cc/SM2P-6H3L]; Peter Jacobson, *Medical Records and HIPAA: Is It Too Late to Protect Privacy*, 86 MINN. L. REV. 1497 (2002), https://scholarship.law.umn.edu/mlr/2082 [https://perma.cc/TYE7-5Z54].

[39] THE PRIVACY PROTECTION STUDY COMMISSION, PERSONAL PRIVACY IN AN INFORMATION SOCIETY (July 1977), https://www.ojp.gov/pdffiles1/Digitization/49602NCJRS.pdf [https://perma.cc/JYK2-V5XS]; Dep't of Health, Education, and Welfare, *Protection of Human* Subjects, 43 FED. REG. 56173 (1978), https://www.govinfo.gov/content/pkg/FR-1978-11-30/pdf/FR-1978-11-30.pdf [https://perma.cc/LRC2-5M3Y].

[40] *See generally* HHS Regulations, *supra* note 24.

[41] *See generally* Admin and Privacy Requirements, *supra* note 24.

[42] *See generally* Barbara J. Evans, *Much Ado About Data Ownership*, 25 HARV. J. L. & TECH. 70 (2012), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1857986 [https://perma.cc/82Y7-BEQS].

[43] *See generally* O. CARTER SNEAD, WHAT IT MEANS TO BE HUMAN: THE CASE FOR THE BODY IN PUBLIC BIOETHICS (Harvard University Press ed., 2020); Paul Root Wolpe, *The Triumph of Autonomy in American Bioethics: A Sociological View*, *in* BIOETHICS AND SOCIETY: CONSTRUCTING THE ETHICAL ENTERPRISE 38-59 (Raymond DeVries & Janardian Subedi eds. 1998); TOM L. BEAUCHAMP & JAMES F. CHILDRESS, PRINCIPLES OF BIOMEDICAL ETHICS (Oxford University Press, Inc. ed., 2001); *see generally* Alfred I. Tauber, *Sick Autonomy*, 46 PERSP. BIOL. MED. 484 (2003), https://muse.jhu.edu/pub/1/article/48182.

person's consented disclosures can reveal information about others, particularly in the case of genetic and biometric data.[44] This first drew attention in specific contexts, such as when genetic testing implicates the privacy of family members[45] or when research involving Indigenous peoples supports stigmatizing inferences about other members of small Tribal communities.[46] It is thus not surprising that Indigenous Nations have played a leadership role in developing the modern concept of social licensing of data sharing. Recognizing that consent to the use of genetic data of one or more individual members of a tribe could be used to make inferences about the whole tribe in ways that are inconsistent with the values, interests, or collective will of the tribe, Indigenous Nations have drafted guidelines and principles to guide collection and use of data from tribal members.[47] Even if individual community members might be willing to consent to share their data, the way data is used has broader impacts that require ethical analysis at the level of families and communities, not just the individual.

In the current age of large-scale data analytics, the problem of privacy interdependency is no longer a concern that only affects families and "discrete and insular minorities" facing heightened risk of discrimination and marginalization.[48] It could affect all members of society with possible consequences for future generations. If computational tools produce generalizable results, they can support privacy-invasive inferences about people, even if they were not included in the data set to begin with.[49] That is, after all, the point of generalizability.

The regulatory developments in this aspect were initially primarily concerned with an individual's right to protect personal information, also known as informational privacy. Informational privacy was understood to give individuals the right to control their personal data that is captured and used by the system.[50] In response, firms began enforcing legal statements of notice and consent—privacy policies—that require users to consent to the use of data. However, this practice defeated its initial purpose of rationally informing users. Due to the effort required to read lengthy, legally complex documents—combined with people's cognitive biases in

---

[44] Gergely Biczók & Pern Hui Chia, *Interdependent Privacy: Let Me Share Your Data*, *in* FINANCIAL CRYPTOGRAPHY AND DATA SECURITY 338-353 (Ahmad-Reza Sadeghi, ed., Foteini Baldimtsi, Christian Cachin eds. 2013).

[45] Marwan K. Tayeh et al., *The Designated Record Set for Clinical Genetic and Genomic Testing: A Points to Consider Statement of the American College of Medical Genetics and Genomics (ACMG)*, 25 GENET. MED. 100342 (2023), https://www.gimjournal.org/article/S1098-3600(22)01026-7/fulltext (last visited May 3, 2023).

[46] Krystal S. Tsosie, Joseph M. Yracheta, & Donna Dickenson, *Overvaluing Individual Consent Ignores Risks to Tribal Participants*, 20 NAT. REV. GENET. 497 (2019), https://pmc.ncbi.nlm.nih.gov/articles/PMC7250136/ [https://perma.cc/U8X4-82LP].

[47] *See generally* Stephanie Russo Carroll, et al., *The CARE Principles for Indigenous Data Governance*, 19 DATA SCI. J. 43 (2020), https://datascience.codata.org/articles/10.5334/dsj-2020-043 [https://perma.cc/BLN8-EQP4].

[48] United States v. Carolene Prods. Co., 304 U.S. 144 (1938).

[49] Barbara J. Evans, *Rules for Robots, and Why Medical AI Breaks Them*, 10 J. L. BIOSCIENCES 1 (2023).

[50] *See generally* Kelly D. Martin & Patrick E. Murphy, *The Role of Data Privacy in Marketing*, 45 J. ACAD. MARK. SCI. 135 (2017).

evaluating threats and risks of online information access-- it soon became obvious that even individuals who claimed to care about privacy did not practice it online.[51] Only recently have regulators in Europe, Australia, and the UK, started to address the potential systemic risks and harms from data sharing and evaluate potential large technology solutions (cf., Digital Services Act). The accumulation of numerous small, individual violations of privacy becomes systemic risk and harm at the societal level, a function of the larger choices societies make about what types of data analyses to pursue or prohibit.[52] The individual's right to opt in or opt out of participating in those analyses no longer has the power to prevent companies from drawing unwanted personal inferences that might cause discrimination, stigmatization, or dignitary harm. Individual agency and control over "one's data"—individual-level data collected from or contributed by each individual—no longer translates into control over what others can know or purport to know about the individual.[53] Controlling *one's data* and the stream of data each of us produces does not limit what others, including firms and third parties, can infer about us with modern data analytic tools, such as those that process data on what, when, and for how long a user looks at a product or page while browsing.

Consent and data ownership lose effectiveness as a means of privacy protection, as seen through the ineffective nature of privacy policies. Similarly, de-identification (e.g., through data anonymization techniques) has lost effectiveness as improved analytic tools make re-identification progressively easier.[54] Despite these limitations, policymakers continue to rely on individual consent as a means for protecting privacy, perhaps because consent rules remain popular with the public and cost less than making systems truly secure or curtailing remunerative data uses that cause systemic privacy loss.

IV. PENDING PROBLEMS FOR THE USE OF LLMS IN HEALTHCARE

The arrival of ChatGPT in November 2022 created great potential for its use or misuse as a source of medical advice. This underscores the stakes in debates over social licences, particularly the need for greater verifiability and provenance of outputs from large language models (LLMs). While LLMs can simulate near human-level responses to a wide range of questions, there are three main issues pertaining to the

[51] Alessandro Acquisti, Laura Brandimarte, & George Loewenstein, *Privacy and Human Behavior in the Age of Information*, 347 SCIENCE 509 (2015).

[52] Danielle Keats Citron & Daniel J. Solove, *Privacy Harms*, 102 BOS. UNIV. L. REV. 793 (2022); Evans, *supra* note 42. Citron and Solove offer a typology of privacy harms to facilitate meaningful address in the courts. They note that courts typically require a finding of harm in privacy cases, which do not fit well with current judicial approaches.

[53] Evans, *supra* note 49; World Bank Data Help Desk, *What Do We Mean by Microdata?*, WORLD BANK, https://datahelpdesk.worldbank.org/knowledgebase/articles/228873-what-do-we-mean-by-microdata [https://perma.cc/U4PW-UDSC].

[54] Evans, *supra* note 49 (discussing the failure of the AI Bill of Rights to provide for adequate privacy protections in AI and M, advocating for contextually-informed approaches); Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701 (2010), https://papers.ssrn.com/abstract=1450006 (discussing how anonymization does not provide the protection that was generally thought given the ease with which re-identification can occur).

development and regulation of AI models in general, and generative AI models to which LLMs belong, in particular.

First, legally protected rights, such as data protection and privacy, as well as copyright, may be violated in the context of generative AI as freely or voluntarily uploaded data effectively becomes "public." This practice of collecting publicly available data—without consent—is commonly known as "scraping." This scraped data can then be used for subsequent training and inclusion in outcomes of the models. Some of this proprietary data that a user inputs in prompt engineering, the method of giving instructions to LLM models to achieve the user's desired output, can become available and identifiable to others.[55]

In addition to using data that may allow for the identification of a specific person without knowledge or consent, the uptake and integration of personal data in LLMs could become problematic in the context of the GDPR Article 17's "right to erasure"—sometimes referred to as the right to be forgotten—as it is virtually impossible to retrieve a single data point pertaining to a specific individual after the model has been developed and deployed.[56]

Similarly, there is little clarity regarding whether and how personal data, once generated, are being used downstream in LLMs. This is particularly relevant considering modern LLMs are being trained initially on large corpora of texts and then fine-tuned for a given purpose, which ultimately may be very different from the context of large tranches of training data. This leaves two key unresolved questions about the downstream usage of personal emails, documents, or messages that are shared, repackaged, and ultimately incorporated in LLMs, which are used for health purposes:[57] (1) What should be the appropriate chain of provenance? (2) Is it critical to understand whose health data is being used to train LLMs that are subsequently fine-tuned and deployed beyond the health domain?

Second, exploring the potentially harmful effects of discrimination resulting from training data construction or the fallibility of models becomes cumbersome due to the black-box nature of these models. Therefore, mitigating potential risks to marginalized or underrepresented groups in the system becomes challenging, as models themselves may reflect and exacerbate explicit and implicit human biases in healthcare, yet be impervious to identification or mitigation.[58] Consider a recent Lancet

[55] Biswas, Anjanava & Wrick Talukdar, *Guardrails for Trust, Safety, and Ethical Development and Deployment of Large Language Models (LLM)*, 6 J. SCI. & TECH. 55-82 (2023).

[56] GDPR, *supra* note 23.

[57] Joe Zhang et al., *Mapping and Evaluating National Data Flows: Transparency, Privacy, and Guiding Infrastructural Transformation*, 5 LANCET DIGIT. HEALTH 737 (2023).

[58] Allison Koenecke et al., *Racial Disparities in Automated Speech Recognition*, 117 PROC. NAT'L ACAD. SCI. 7684 (2020), https://pubmed.ncbi.nlm.nih.gov/32205437/ [https://perma.cc/B76A-CKFJ]; Aylin Caliskan, *Detecting and Mitigating Bias In Natural Language Processing*, THE BROOKINGS INSTITUTION (May 10, 2021), https://www.brookings.edu/articles/detecting-and-mitigating-bias-in-natural-language-processing/ [https://perma.cc/8UZ2-PLBS].

report that showed a heightened extent of generative AI's bias in creating healthcare imagery, in which the model could not produce visuals of black doctors treating white children when asked.[59] Further, automated speech detection models are known for their lower accuracy in recognizing speech by minority groups, and numerous studies indicated AI's propensity to exacerbate bias and perpetuate health inequities.[60] Furthermore, this lack of disclosure in the data creates additional issues, from a transparency and safety perspective, it is important to ensure outputs are based on data that adequately represent the end user and produce valid results in a given context.[61] Further, it is also essential for users to feel that recommendations provided by LLMs reflect or at least respect their values and priorities, or at the very least provide accurate information. This is extremely difficult if training data and processes are not shared due to "the competitive landscape and the safety implications of large-scale models like GPT-4."[62]

Third, human oversight of decisions is the main regulatory and lay-folk requirement for implementing AI systems (social licence), particularly in the domains that impose moral decisions or potential societal risks. In health systems, the human ability to effectively oversee AI decision-making is problematic. Due to potential reliance on AI models, sometimes referred to as "automation bias" of AI models, the AI output may "nudge" physician decision-making in a way that may not be justified without human scrutiny or consideration.[63] This concern of *automation bias* amplifies when an algorithm unduly influences the physician who may in turn attempt to influence a patient's decision-making as it is unclear which frame of reference or objective of framing the information the algorithm will use. Implementing AI systems in healthcare has been documented to lead to different interpretations and

[59]Arsenii Alenichev, Patricia Kingori & Koen Peeters Grietens, *Reflections before the Storm: The AI Reproduction of Biased Imagery in Global Health Visuals*, 11 LANCET GLOB. HEALTH 1496 (2023), https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(23)00329-7/fulltext (last accessed Oct. 30, 2024).

[60] *See generally* R. Agarwal et al., *Addressing Algorithmic Bias and the Perpetuation of Health Inequities: An AI Bias Aware Framework*, 12 HEALTH POL'Y TECH. 100702 (2023), https://www.sciencedirect.com/science/article/abs/pii/S2211883722001095 [https://perma.cc/8WDA-PBW2].

[61] *See generally* Olga Akselrod, *How Artificial Intelligence Can Deepen Racial and Economic Inequities*, ACLU (Jul. 13, 2021), https://www.aclu.org/news/privacy-technology/how-artificial-intelligence-can-deepen-racial-and-economic-inequities [https://perma.cc/L5TN-779L]; *see generally* Marie-Laure Charpignon, MSc. et al., *Critical Bias in Critical Care Devices*, 39 DATA SCI. IN CRITICAL CARE 795 (2023), https://www.sciencedirect.com/science/article/abs/pii/S0749070423000131 [https://perma.cc/7MQD-4599].

[62] OpenAI, et al., *GPT-4 Technical Report*, ARXIV CORNELL UNIVERSITY (March 15, 2023), https://arxiv.org/abs/2303.08774 [https://perma.cc/7PKF-394X].

[63] Center for Devices and Radiological Health, *CDRH Issues Draft Guidance on Predetermined Change Control Plans for Artificial Intelligence/Machine Learning-Enabled Medical Devices*, U.S. FOOD AND DRUG ADMIN. 8 (Mar. 30, 2023), https://www.fda.gov/medical-devices/medical-devices-news-and-events/cdrh-issues-draft-guidance-predetermined-change-control-plans-artificial-intelligencemachine (last visited Aug. 14, 2023).

goals by diverse stakeholders in the implementation – government policymakers, hospital managers, doctors, and IT managers.[64]

V. CURRENT BARRIERS TO DEVELOPING AN EFFECTIVE GOVERNANCE MODEL

Despite the dangers of AI in the medical space, human oversight over AI decision-making may unintentionally lead to higher negative consequences for patients. Suppose the algorithm's decision does not correspond to that of the doctor. In that case, the doctor is likely to question their initial judgment and ask for more tests, to use as evidence, assuming that the AI models are correct and that they may have missed some relevant connection. On the contrary, in attempts to provide oversight over AI, health professionals were found to underutilize potentially relevant information from AI, resulting in over-testing predictably low-risk patients and undertesting predictably high-risk patients.[65] The movement for social licensing seeks to give communities and members of society a voice in decisions about how and to what extent to licence their data and allow companies to use it. Will the public benefits of that data use be sufficient to justify the systemic privacy impacts on those who opt-in and those who opt out? How will communities agree on social licence for a particular use? By what process and on what basis?

A. *Deliberating 'Sufficient' Public Benefits to Justify Privacy Risks*

There are significant differences in focus between the U.S., EU, and Chinese models of AI regulation.[66] U.S. regulation focuses strongly on the market-driven model and protecting freedom of expression and innovation, which supports a laissez-faire approach to technology regulation that avoids governmental interventionism and relies on the market and self-regulation. This is why the privacy regulations are also more relaxed in the U.S. than in Europe.[67] The absence of a constitutional or fundamental right to privacy in the U.S. allows for sectoral regulation of privacy as pertains to specific types of information, such as health, banking, and insurance.[68] Should one make a claim in court, they are required to prove that the privacy harm has led to a substantial physical or

[64] Tara Qian Sun & Rony Medaglia, *Mapping the Challenges of Artificial Intelligence in the Public Sector: Evidence from Public Healthcare*, 36 Gov. Inf. Q. 368 (2019), https://www.sciencedirect.com/science/article/abs/pii/S0740624X17304781 [https://perma.cc/H48F-96EM].

[65] *See generally* Sendhil Mullainathan & Ziad Obermeyer, *Diagnosing Physician Error: A Machine Learning Approach to Low-Value Health Care*, 137 Q. J. Econ. 679 (2022); Nikhil Agarwal et al., *Combining Human Expertise with Artificial Intelligence: Experimental Evidence from Radiology* (Nat'l Bureau of Econ. Rsch., Working Paper No. w31422).

[66] *See generally* ANU BRADFORD, DIGITAL EMPIRES: THE GLOBAL BATTLE TO REGULATE TECHNOLOGY (Oxford Univ. Press ed., 2023).

[67] Paul M. Schwartz, *The EU-U.S. Privacy Collision: A Turn to Institutions and Procedures*, 126 HARV. L. REV. 1966 (May 2013), https://harvardlawreview.org/print/vol-126/the-eu-u-s-privacy-collision-a-turn-to-institutions-and-procedures/ [https://perma.cc/EP6N-4CPZ]; *see generally* Khadija Robin Pierce, *Comparative Architecture of Genetic Privacy*, 19 IND. INT'L COMPAR. L. REV. 89 (2009), https://journals.indianapolis.iu.edu/index.php/iiclr/article/view/17600/17764.

[68] Pierce, *supra* note 67.

economic injury.[69] The weakness of the U.S. system lies in the fact that firms that are producers or commercial users of AI systems, albeit they may not have any malicious intent per se, are driven by the race to market and business models that strongly emphasize the growth in number of users and their engagement. This business-driven mindset may be opposing the metrics that are based on harm, safety, or risk mitigation for consumers or patients. Currently, market motivations (i.e. seeking commercial gain) decide which data should be pursued through collection and internalization. This has resulted in developments such as the commercial use of wellness and fitness data from personal devices, which is largely unregulated. Contrastingly, social licensing has worked in Indigenous Nations, partly because they have organized leadership in a position to render decisions on behalf of the group. But for society at large, who can make the decision that a data-use should go forward? There are many examples in medical privacy law where state or federal legislators make those decisions without input of the many people the legislation will affect.[70] Having "community representatives" can be criticized as tokenism, and *community representatives* on IRBs have not had enough voting power to impact decisions seriously. Various "data-citizenship" proposals would leave these decisions to everybody whose data is in a dataset.[71] Looking forward, what are the options for addressing such concerns?

If it is presumed that *sufficient public benefit* can be realized from sharing sensitive data, then it is also necessary to define what is meant by 'public benefit,' and how this is operationalized. This involves procedural questions regarding the process through which public benefit is defined and who's involved in this process. Outside of AI, regulators have used approaches such as the 'Takings Clause,' where private property can be taken for public use as long as there is just compensation in the form of public benefit.[72] The inverse is notably true.[73] An example of the Taking's Clause implementation was *Armstrong v. United States* (1960), which prevented the government from compelling individuals to bear public burdens in situations where it was deemed fair that they were borne by the public as a whole.[74] Informed consent for sharing personal information, either for medical trials and treatment advancement, or with other third parties, is the cornerstone for regulating good clinical practice in the U.S.

---

[69] Citron & Solove, *supra* note 52.

[70] *See generally* Evans, *supra* note 29; Barbara J. Evans, *Power to the People: Data Citizens in the Age of Precision Medicine*, 19 Vand. J. Ent. Tech. L. 243 (2017), https://pmc.ncbi.nlm.nih.gov/articles/PMC5673282/ [https://perma.cc/A3U9-MMNN].

[71] Evans, *supra* note 70, at 4.

[72] Richard A. Epstein & Eduardo M. Peñalver, *The Fifth Amendment Takings Clause: Common Interpretation*, CONSTITUTION CENTER, https://constitutioncenter.org/the-constitution/amendments/amendment-v/clauses/634 [https://perma.cc/ZE44-LSPB].

[73] *Id.*

[74] Armstrong v. United States, 364 U.S. 40, 49 (1960).

and Europe.[75] However, deeming a situation "just and fair" to compel private data sharing is not simple. Often, it is clearer to list what is *not* an acceptable public use of data rather than to specify a priori what is, thus relying on an approach that carves away unacceptable use cases.

Extending this approach to LLMs, complicated questions emerge: Could an individual's personal information be shared with a health provider if it could contribute to saving another person's life? Should one have the choice of sharing their data to save a specific person or group's life? How should benefits broadly construed be reconciled outside of life-or-death decisions? These issues become particularly convoluted if individual health data is permitted to be shared with health insurance companies that may decide to use such information for their own profit. While it is unlikely this decision would materialize in such explicit terms, there are adjacent scenarios in the case of LLMs and large AI models. For example, failing to contribute a group's data to a given dataset may harm those like the group, resulting in a model with fewer examples to learn from. In addition, the over-provision of one group's data or narrative may shape an AI model's view of the world. Here, it is possible to foresee the potential enthusiasm of third-party companies to provide a given set of data that would increase the likelihood of a narrative that benefits their image, products, or bottom line.

B. *Data Sovereignty and the Constitution of Groups*

As health data has scaled to new heights, there has been pressure from minority groups, such as Indigenous populations, to ensure principles of data sovereignty apply.[76] Data sovereignty is the ability for groups to own the data they produce and determine the rules for how that data can be used to ensure its usage is in line with their priorities.[77] However, not all groups are similarly constituted, represented, or acknowledged, and therefore, applying data sovereignty across all populations presents challenges.

Further, modern AI creates new challenges where one individual consenting to data sharing could reveal information about other group members (e.g., genetic data). This will likely soon be the case for much larger groups due to the volume of information available. Even if an individual does not consent to data sharing, if a model has access to enough similar individuals, then inferences can be drawn about the former that could be privacy-invasive, despite not having access to their own microdata. Therefore, setting boundaries around a small group's data use, which is typically confined to one region, may be feasible.

---

[75] U.S. Food and Drug Administration, *Regulations: Good Clinical Practice and Clinical Trials*, FDA (Jan. 21, 2021), https://www.fda.gov/science-research/clinical-trials-and-human-subject-protection/regulations-good-clinical-practice-and-clinical-trials [https://perma.cc/5VFM-QZCV].

[76] Global Indigenous Data Alliance, *CARE Principles for Indigenous Data Governance*, GIDA, https://www.gida-global.org/care [https://perma.cc/97VJ-CJPR].

[77] Patrik Hummel, Matthias Braun, & Peter Dabrock, *Data Sovreignty: A Review*, 8 BIG DATA & SOC'Y (2021), https://journals.sagepub.com/doi/epub/10.1177/2053951720982012.

However, the data required for LLMs is much bigger, and its sphere of influence is significantly greater than any single group. In the setting of LLMs, this implies the social licence of many groups, as any social licence must now scale to a much broader societal level. Given the extent of LLM data use, which is more expansive than a singular collective in each place. Even with identifiable priorities, leadership, and other working protocols, implementing a social licence on this level is substantially more complicated. It remains to be seen if this cordoning of data is feasible at the LLM scale.

With that said, this may be more widely applicable, as there are already precedents for protecting Indigenous data rights. The International Indigenous Data Sovereignty Interest Group—within the Research Data Alliance—published the *CARE Principles for Indigenous Data Governance*—Collective Benefit, Authority to Control, Responsibility, and Ethics—align with the primary goals of "fostering Indigenous self-determination by enhancing Indigenous use of data for Indigenous pursuits."[78] The CARE Principles build upon the FAIR Principles—Findable, Accessible, Interoperable, Reusable—for data governance, management, and stewardship, first published in 2016 while ensuring that data is shared on broader Indigenous terms.[79] Due to the concern that most Indigenous data is overseen by non-Indigenous institutions, the CARE Principles respond to an increased demand for Indigenous participation in data governance. The CARE Principles center on the people and implementation of data, which complement the accessibility principles of FAIR.[80] These principles are now referenced in policy documents, including the AIATSIS Code of Ethics and the UNESCO Recommendation on Open Science.[81]

It is, therefore, essential to decide what governance model should be adopted to determine which uses of data are concordant with the goals and social values of different communities. The real challenge will be to develop governance models that allow for potentially conflicting values to

[78] Stephanie Russo Carroll, *supra* note 47.

[79] *Id.*

[80] Stephanie Russo Carroll, et al., *Operationalizing the CARE and FAIR Principles for Indigenous Data Futures*, 8 DATA SCI. J. 108, https://pubmed.ncbi.nlm.nih.gov/33863927/ [https://perma.cc/RE4R-CCHH].

[81] Sara Tomkins & Angus Harden, *The AIASTIS Code of Ethics for Aboriginal and Torres Strait Islander Research*, *in* THE ROUTLEDGE HANDBOOK OF HUMAN RESEARCH ETHICS AND INTEGRITY IN AUSTRALIA 83-96 (2020); Anup Kumar Das, *UNESCO Recommendation on Open Science: An Upcoming Milesgtone in Global Science*, 2 SCI. DIPLOMACY REV. 39 (Nov. 2020); Amnesty International & Access Now, *The Toronto Declaration: Protecting the Rights to Equality and Non-discrimination in Machine Learning Systems*, ACCESSNOW (May 16, 2018), https://www.accessnow.org/press-release/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/ [https://perma.cc/RM79-U4Y9]; Jason Edward Lewis et al., INDIGENOUS PROTOCOL AND ARTIFICIAL INTELLIGENCE POSITION PAPER (Jason Edward Lewis ed., Aboriginal Territories in Cyberspace ed., 2020); Jessica Fjeld et al., *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI*, THE BERKMAN KLEIN CENTER FOR INTERNET & SOC'Y RESEARCH Pub. No. 2020-1, 2020 (Jan. 2020), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3518482;Karaitiana Taiuru, *Treaty of Waitangi/Te Tiriti and Māori Ethics Guidelines for: AI, Algorithms, Data and IOT*, TE KETE O KARAITIANA TAIURU BLOG (May 3, 2020), http://taiuru.co.nz/TiritiEthicalGuide/ [https://perma.cc/BS69-RVTT]; Australian Instit. of Aboriginal and Torres Strait Islander Studies, *AIATSIS Code of Ethics for Aboriginal and Torres Strait Islander Research* (AIATSIS ed., 2020), https://aiatsis.gov.au/sites/default/files/2020-10/aiatsis-code-ethics.pdf.

exist concurrently. This will inevitably have downstream consequences; if one group forms a majority in the dataset, LLMs and models more generally will shape their narrative based on this group, resulting in inequity of ideas, health outcomes, and legal perspectives. Therefore, minority groups and Indigenous Nations may find themselves being penalized for failing to agree to a private company's data use agreement, particularly in a monopolized market sector or without strong external intervention. AIs and LLMs have thus created a significant new challenge for regulation, managing both input from data sources and the impacts of predictive model outputs.

C. *Can Old Systems Regulate New Privacy Concerns?*

The U.S. regulatory framework for privacy in healthcare, as exemplified by HIPAA, represents a sector-specific approach to data protection.[82] However, given its enactment in 1996, long before the advent of digital healthcare records and advanced internet technologies, the legislation may be increasingly outmoded in today's digital landscape.[83] Originally designed to address privacy concerns related to physical medical records, HIPAA now faces the daunting task of ensuring privacy in an age marked by rapid advancements in artificial intelligence and LLMs. The proliferation of digital platforms and the exponential growth of online patient portals have resulted in vast and complex healthcare databases. This shift presents an array of challenges that were unforeseen during HIPAA's formulation. Further, existing data-sharing agreements are based on previous risks and benefits of associated capabilities, yet LLMs have changed this landscape in both senses. The risk of unintentionally releasing sensitive data is greater in cases such as training LLMs on personal emails, unpublished academic articles, or medical and judicial data.[84] Similarly, an important distinction must be made between an event's frequency and the outcome's effect. Accuracy or leakage rates do not quantify the actual harm that would occur from one such event. Not only has the scale changed, with ChatGPT reaching 100 million users in 2 months, but the sphere of influence has changed too, with the worldwide distribution of information collection and usage.[85]

Now, LLMs are not required to comply with HIPPA, not even those used in health-related applications outside of conventional healthcare settings. These AI-driven chatbots do not fall into the category of "covered entities"—like hospitals—stipulated by HIPAA, resulting in them effectively operating outside the jurisdiction of this key healthcare regulation. Although they may provide diagnosis, treatment recommendations, and analysis patterns in medical records, there has been

[82] *See generally supra* note 24.

[83] Marks & Haupt, *supra* note 25.

[84] Genevieve P. Kanter & Eric A. Packel, *Health Care Privacy Risks of AI Chatbots*, 330 JAMA 311 (2023).

[85] Krystal Hu, *ChatGPT Sets Record for Fastest-Growing User Base - Analyst Note*, REUTERS (Feb. 2, 2023, 7:33AM PST), https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/ (last visited Apr. 3, 2023).

discussion about whether LLMs are treated as regulated medical devices.[86] The current predominant view is that LLMs are treated as medical devices is they are developed specifically for use in medicine and if the LLMs are used for any of the medical purposes listed above, but their unreliability precludes approval.[87]This is in contrast to AI-driven radiology diagnostic imaging tools, which are clearly regarded as medical devices in the EU and U.S. This regulatory blind spot underscores a pressing need for a comprehensive re-evaluation of the efficacy of existing laws to adequately respond to the rapidly evolving digital health landscape. The inability to currently classify LLMs as subject to HIPAA can, in part, be attributed to the fact that they exist in a regulatory "grey area." These AI-based tools are primarily software platforms rather than healthcare providers or insurers, thus evading the conventional definitions of covered entities under HIPAA. Moreover, as they often function in non-healthcare environments, they fall outside HIPAA's purview which is traditionally limited to the healthcare sector.

The task of bringing LLMs under HIPAA's compliance umbrella is undoubtedly complex. It would necessitate a significant expansion and reinterpretation of existing regulations to encompass these digital health entities. Furthermore, it would entail the creation of stringent protocols for data deidentification and storage, along with the implementation of robust security measures to safeguard sensitive health information handled by these AI tools. However, whether there is sufficient call to implement these measures directly affects the feasibility and desirability of making LLMs HIPAA-compliant. HIPAA's Privacy Rule operates on the assumption that deidentified data remains secure. This premise, while sound in theory, falters in the face of advanced technologies like LLMs, one of the many technologies giving rise to the emergence of "reidentification science."[88] The ability of these technologies to easily reidentify deidentified data presents significant potential for harm, thereby challenging the effectiveness of HIPAA's deidentification requirements in the current digital age. Therefore, the extension of HIPAA regulations to LLMs may not necessarily provide the intended protection of privacy or address the larger issues of power disparity and inequality inherent in the digital healthcare sphere. If LLMs were to become HIPAA compliant, it could offer an illusion of safety while still leaving users vulnerable to privacy breaches and misuse of sensitive health information.

Consequently, this raises a pivotal question: Rather than adapting the existing HIPAA framework to cover LLMs, should the development of an entirely new regulatory framework be considered? This new framework would need to address the unique challenges posed by AI and LLMs in

[86] Meskó& Topol, *supra* note 16.
[87] Stephen Gilbert, et al., *Large Language Model AI Chatbots Require Approval as Medical Devices*, 29 Nature Med. 2396 (2023), https://pubmed.ncbi.nlm.nih.gov/37391665/ [https://perma.cc/J73L-R7M6].
[88] *See e.g.*, Muazzam Maqsood, et al., *An Efficient Deep Learning-Assisted Person Re-identification Solution for Intelligent Video Surveillance in Smart Cities*, 17 FRONTIER COMPUT. SCI. 174329 (2023), https://doi.org/10.1007/s11704-022-2050-4.

healthcare, ensure robust data privacy and security, and tackle broader issues related to power dynamics and health inequities in the digital realm.

Currently, the most advanced regulatory attempts to curb the adverse effects of AI, are to be found in the recently passed EU AI Act,[89] which was amended in its second proposed iteration to address the complexities of these models. This lag is partly due to AI models' novelty and rapid speed of change, and partly to the inability and unfeasibility of attempting to access and mitigate all foreseeable risks that are likely to impact health, safety, and fundamental human rights—both for the foundational model itself and for the specific cases in which the models could be used.[90] Interestingly, however, a recent study by Bommasani and colleagues ranked 10 AI models against the EU's draft rules on AI, including describing data sources and summarising copyrighted data; the disclosure of the technology's energy consumption, and computing requirements; and reports of evaluations, testing, and foreseeable risks associated with it. Each model fell short in several key areas, with six of ten providers scoring less than fifty percent on compliance (ranging from 25%-75%). Therefore, it is likely, that self-audit is an unacceptable method, especially because of the large amount of economic ground at stake and the established poor track record of tech companies regulating themselves.[91]

D. *How Should Responsibility and Liability be Allocated?*

Generally, scrutiny for safety and efficacy of medical devices are the remit of the Federal Food and Drug Administration ("FDA"), which takes a risk-based approach to regulation of medical devices. As it stands, generative AI technologies, such as ChatGPT, are yet to undergo the rigorous process of FDA review.[92] These models operate in a regulatory *grey area*, potentially necessitating FDA scrutiny if they cross into territory reserved for diagnosing, treating, or preventing diseases, yet they do not fit neatly into the existing medical device exceptions delineated by the Federal Food, Drug, and Cosmetic Act[93] or HIPPA. Paradoxically, ChatGPT offers differential diagnoses but concurrently urges users to consult medical professionals, raising pivotal questions about its role and responsibilities in the healthcare sector. Existing disclaimers highlight the known limitations of ChatGPT, including the possibility of errors and "hallucinations"—instances where the AI presents fabricated information as fact. But, as these systems become increasingly integrated into healthcare delivery, the question of responsibility for inaccuracies—and

[89] 2024 O.J. L 2024/1689. (cf. Art. 113 AI Act).

[90] Phillip Hacker et al., *Regulating ChatGPT and other Large Generative AI Models*, *in* FAccT '23: PROCEEDINGS OF THE 2023 ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 1112-1123 (Ass'n for Computing Machinery, ed., 2023).

[91] Rodrigo Ochigame, *The Invention of "Ethical AI": How Big Tech Manipulates Academia to Avoid Regulation*, THE INTERCEPT (Dec. 20, 2019, 1:19 PM), https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/ [https://perma.cc/TUE7-8HNN].

[92] *Federal Food, Drug, and Cosmetic Act (FD&C Act)*, U.S. FOOD AND DRUG ADMIN. (Mar. 29, 2018), https://www.fda.gov/regulatory-information/laws-enforced-fda/federal-food-drug-and-cosmetic-act-fdc-act [https://perma.cc/XH6F-QUXT].

[93] 21 U.S.C. §§ 301-392 (Suppl. 5 1934).

liability for any harm caused by them—becomes central, regardless of any disclaimers.[94]

Liability becomes a convoluted issue when medical professionals rely on potentially flawed advice produced by AI. To the extent that medical professionals are expected to use reliable tools in the delivery of healthcare, reliance on the output of unvalidated or unapproved tools could be the basis for liability if the standard of care is not met by such use. On the other hand, when patients act based on incorrect AI-generated advice, absent an explicit or implicit invitation to rely on an output and accompanying device approval, assigning responsibility may be little different from a layperson's reliance on a Google or Wikipedia output providing information or advice, although there is increasing recognition of platform accountability for harms caused on or by the platform. The ambiguity surrounding the extent of accountability for AI systems and users points towards a pressing need for robust regulatory guidance. Some of the most important challenges involve determining the level of trust bestowed upon AI systems, identifying the standard of care in an AI-dominated landscape, and setting clear accountability parameters. In the face of the increasing use of AI in healthcare, it is essential to implement strong safeguards to protect patients and physicians and to allocate responsibility and, ultimately, apportion liability in logical and sustainable ways that encourage optimal development, use, and deployment that respect safety, individual, and collective rights. It is necessary to ensure the responsible use of AI and require that appropriate parties be held accountable for the accuracy of AI systems' output, or place parameters around permissible usage in addition to ensuring that users understand the limitations of LLMs.

In addition to informing liability, there is a pressing need for regulatory bodies such as the FDA and the American Medical Association to provide much-needed guidance on the deployment of generative AI in healthcare. A clear regulatory framework would encompass guidelines for responsible AI usage, clear directives for integrating AI tools within existing healthcare structures, and provide clear direction regarding the permissible collection and use of personal data in the training and use of generative AI for healthcare. Navigating this new realm of AI in healthcare requires ongoing, open dialogue, and a commitment to uphold the principles of safe and ethical practice. We contend that the regulatory attempts should not be directed towards the technology itself since it changes rapidly, but it should require transparency in terms of the potential adverse effects on the well-being of individuals.

VI. CLEARING THE PATH FOR A FUTURE FRAMEWORK

As existing regulatory approaches are insufficient, what is needed to address these challenges? To navigate this terrain, we suggest three pivotal

[94] Jack Gallifant et al., *Peer Review of GPT-4 Technical Report and Systems Card*, 3 PLOS Digit. Health 1 (2024).

amendments or modifications that should be pursued by individuals, communities, and governments to overcome impediments.

A. *Mandate Transparent LLM Training Processes*

In the world of LLMs, transparency is not a luxury, but a necessity, particularly in the context of social licence and what is relevant to a determination of whether a social licence is merited. Determining the threshold for a "catastrophe" or what constitute an "adequate precaution" forms a fundamental part of this transparency. Even the terms usually associated with social licence, such as *sufficient public benefit* must be balanced against risks of harm. In the context of transparency, the nature of risks and benefits can only be determined if the data the LLM has been trained on is known. Additionally, whether the model is sufficiently of public benefit may be determined by whether the training data is sufficiently representative of the relevant public. Moreover, Intellectual Property (IP) rights may be affected by the adequacy of AI transparency. The reality is that LLMs are often trained using vast, and at times, opaque data sources, inadvertently exposing the creator to the possibility of IP infringement[95]–especially if the output uses copyrighted material that is not properly recognized. A commitment to transparency could help alleviate this potential pitfall. The EU AI Act tackles this issue with provisions requiring disclosure of the training data and observation of The EU Copyright Directive.[96]

These LLM tools are increasingly embedded into everyday life, yet, understanding the underlying data and the processes that produce the final output remains convoluted. This is equally true for data used pre-training, and in the use of data that is collected through interaction (e.g., prompts). Borrowing from the strides made in machine learning to develop model cards can serve as an inspiration to ensure the models are trained in a specified manner. Further, an essential measure would be metrics to measure the diversity of data used in the training process. It is imperative to inspire confidence in a model so that it can accurately represent a given population without compromising data confidentiality.

B. *Invest in Infrastructure that Temporally Evaluates LLMs*

Despite the rapid emergence of LLMs in healthcare, the future of this technology remains precarious due to technological, regulatory, and legal uncertainty, as well as a lack of clarity. However, for these observations to convert into actionable recommendations, it is imperative to specify the evaluative parameters.

Considering the emphasis on social licence in this discourse, several essential concepts require clarification:

---

[95] Timo Minssen, LLD, et al., *The Challenges for Regulating Medical Use of ChatGPT and Other Large Language Models*, 330 JAMA 315 (2023), https://pubmed.ncbi.nlm.nih.gov/37410482/ [https://perma.cc/FV46-B2CL].

[96] AU AIA, GPAI provisions requiring disclosure and observation of the Copyright Directive.

1. Protected Personal Attributes: One crucial consideration is discerning how LLMs describe, identify, and utilize protected personal attributes in their outputs. For instance, when diagnosing a condition or recommending a treatment, is the model inadvertently biased towards or against certain demographics? Is it accounting for differences in disease prevalence among specific ethnic groups without resorting to blanket generalizations?
2. Interaction and Output Disparities: Another significant aspect is the differential interaction LLMs may have with diverse user subgroups. Does the model exhibit a variance in the likelihood of producing a useful output when interacting with a male patient as opposed to a female patient? Or between an elderly patient and a younger one? This evaluation could encompass examining the ease of use, the quality of recommendations, or even the accuracy of diagnosis across different subgroups.
3. Real-World Learnings and Effects: LLMs can influence real-world decisions and actions in healthcare. Hence, it becomes pivotal to study if there is any divergence in subsequent real-world learnings or effects correlated with either the user's demographics or the demographics of the subject within the narrative. For example, if an LLM provides guidance on patient care, are there noticeable differences in patient outcomes based on the patient's age, ethnicity, or gender? Is this disparity due to the LLM's recommendations, or are there other factors?

Strategic investment in infrastructure that supports the systematic evaluation of LLMs throughout their life cycle is pivotal. This approach helps, in real-time, to identify vulnerabilities, mitigate issues, and acknowledge successes in a setting where precision and dependability are paramount. It is not merely about addressing risks after LLM's deployment, but also about proactively identifying and rectifying potential errors. At the same time, it is essential to uphold the social licence of LLMs through a commitment to dynamic transparency. Given the evolving nature of these systems, our transparency efforts must mirror this dynamism. As LLMs continue to change and evidence of new risks or benefits emerges, these changes must be promptly communicated to the public, ensuring an ongoing dialogue. Dynamic transparency is the cornerstone of securing public trust and consent, empowering users to make informed decisions and navigate their interactions with these advanced technologies confidently.

By combining strategic infrastructure investment with dynamic transparency, it is possible to safeguard the successful integration of LLMs into healthcare.

C. *Create Capable Bodies to Govern LLM Implementation*

Relying on the twin pillars of transparency and continual-accumulation of information, the next step is to put these insights into action. The unique nature of the LLM sector is likely to require innovative regulation and oversight mechanisms. Effectiveness requires integrating people from various disciplines and sectors to fully grasp the technical, clinical, and political implications of changes in this domain. It is necessary to build a system that anticipates and adapts to the future trajectory of LLMs, not one that is solely reactive to the present circumstances. A common challenge of technological innovation is that regulation and governance tend to lag behind. This lag persists until multidisciplinary teams fully understand and address the multiple dimensions and implications of these advances. The ELSI (Ethical, Legal, and Social Issues) committee of the Human Genome Project offers an example of how this can be achieved.[97]

Additionally, maintaining adequate levels of performance, interpretability, corrigibility, safety, and cybersecurity is a prerequisite to LLM development and trust in crucial industries like healthcare. Regulatory focus should be geared towards risk assessment and mitigation measures of fundamental rights. LLMs might need to adopt a model that provides public assurance of a certain standard of data and modelling processes. Standardization bodies in Europe as well as the U.S. are positioning to take on an active role in governance of AI,[98] which could extend to LLMs, as well. This could harness advances in cryptography, like zero-knowledge proofs.

While the universal consensus on AI regulation is to minimize risk of harm, it is also imperative to invest time in defining publicly and explicitly what the risks are in the context of generative AI, both now and in the future as the technology matures. These measures provide a foundation for a potential future governance framework for LLMs. As progress advances, it is necessary to remain vigilant, adaptable, and committed to fostering safety, equity, and transparency in this rapidly evolving field.

D. *Engage Public Discourse to Create Equitable Impact*

The extensive impact of LLMs, both in terms of their far-reaching application and their potential for benefit and harm, makes the issue of data sovereignty crucial. Decisions on what data can be rightfully shared or acquired are not to be made lightly, necessitating broader, substantive discussions at both the national and international level. This debate must extend to what constitutes an acceptable risk from both a policy and public perspective, bearing in mind that perceptions of risk can differ greatly

[97] Eric T. Juengst, *Anticipating the Ethical, Legal, and Social Implications of Human Genome Research: An Ongoing Experiment*, 185 AM. J. GENET. A. 3369 (2021), https://pmc.ncbi.nlm.nih.gov/articles/PMC8530886/ [https://perma.cc/C4CH-HYSB].

[98] Marta Cantero Gamito & Christopher T. Marsden, *Artificial intelligence co-regulation? The role of standards in the EU AI Act*, 32 INT'L J. L. INFO. TECH. 11 (2024), https://academic.oup.com/ijlit/article-abstract/32/1/eaae011/7709069?redirectedFrom=fulltext [https://perma.cc/EBT2-FLSR].

among various populations. For instance, while racial and ethnic data collection is illegal in South Africa due to the nation's history of apartheid, the U.S. National Institute of Health mandates the release of this data from funded projects. This presents a complex challenge for data sovereignty that warrants ongoing dialogue and impacts data-sharing licences and usage conditions. Diverse representation across disciplines, nations, and cultures in these discussions is essential due to existing power structures and corresponding unequal power dynamics. This is paramount to protect the values and priorities of different groups. Maximizing individual agency over personal data, while reaping the benefits of LLM technology, is a high-priority task. In creating a social licence for data use in LLMs, it is essential to respect collective values that undergird the concept of *public benefit*. Moreover, transparency about why and how a particular option or solution is presented to the user is a cornerstone in fostering user trust in these systems.

While policymakers play a pivotal role in these decisions, individual users and healthcare professionals should also have a say. The government must implement safeguards that promote benefits, mitigate risks, and uphold fundamental rights, thus balancing the pursuit of technological advancement with the protection of individual rights.

VII. CONCLUSION

The advancements of LLMs have changed the landscape of data use, creating new opportunities but also new risks. Given the large scale of LLMs, the underlying data on which they are trained has the potential to be incorrectly assumed as accurate representations of social groups. This highlights how data acquisition and implementation in AI produce systemic power inequalities, leading to ineffective representation and flawed calculations. This highlights the appropriateness of social licences based on the provision of *public benefit*. Incentives for data use and data-sharing have never been higher. This is especially true for groups that have historically experienced data marginalization. It is critical that organizations using private or potentially biased data are diverse, appropriately accredited, and robustly regulated to minimize the likelihood that LLMs will exacerbate social inequities, undermine sovereignties, or harm individuals or marginalized groups. The scope of social licence for the development of LLMs based on publicly available data clearly hinges on the degree to which all members of society can benefit from these models and are not harmed by them.