Protecting the Promise to the Families of Tuskegee: Banning the Use of Persuasive AI in Obtaining Informed Consent for Commercial Drug Trials

JENNIFER S. BARD*

* © 2023 Jennifer S. Bard. Professor of Law at the University of Cincinnati College of Law and Professor, Department of Internal Medicine at the University of Cincinnati College of Medicine. She has a B.A. from Wellesley College, a J.D. from Yale Law School, a Master's degree in Public Health from the University of Connecticut, and a Ph.D. in Higher Education from Texas Tech University. Before coming to Cincinnati, Dr. Bard held the Alvin R. Allison Chair in law at Texas Tech University and was a professor in the department of psychiatry. She has been a visiting scholar at Harvard Law School's Petrie-Flom Center for Health Law Policy, Biotechnology, and Bioethics and the O'Neill Institute for National & Global Health Law at Georgetown Law and a visiting professor at the University of Iceland's Law Faculty and in the LLM program at LaTrobe University in Melbourne, Australia. She is an elected member of the American Law Institute, has been Chair of the American Association of Law Schools Section on Law and Medicine, and has twice chaired the Committee on Mental Disability. In addition, she was among the founding members of the American Public Health Association's Section on Ethics where she has held a number of elected positions and is currently the Treasurer.

This Article began as a presentation in January 2022 at a symposium, *Intelligent Machines, Emotions and Our Planet*, which is part of "AI, People & Planet," an initiative between the Stockholm Resilience Centre at Stockholm University, the Beijer Institute of Ecological Economics at the Swedish Royal Academy of Sciences, the Urban Systems lab at the New School, and the Princeton's Institute for International and Regional Studies at Princeton University, and the symposium was funded by The Crafoord Foundation. Stockholm Resilience Centre TV, *A Conversation About Regulation and Law of AI and Emotions*, YOUTUBE (Feb. 20, 2022), https://www.youtube.com/watch?v=ogt4qpjpFYk [https://perma.cc/ML6D-AFUN]. This Article also builds on work pursuant to a contract

TABLE OF CONTENTS

with West Publishing to draft the first ever Nutshell in Human Subject Research. There are many people to thank for their support of this work. First, thank you to Victor Galaz Rodriguez, Associate Professor in Political Science at the Stockholm Resilience Centre and Programme Director at the Beijer Institute's Governance, Technology and Complexity Programme. Thank you to Associate Dean Jane Moriarty and the faculty of Duquesne Law School who made helpful comments after a presentation of this Article. Thank you also to Glen McLain, from the University of Cincinnati Law School Class of '23 and University of Cincinnati College of Medicine Class of '25 and Sarah Saadeh, Brian Patrick, and Megan Fisher from the University of Cincinnati Law School Class of '25 for their very helpful research assistance, and thank you to Emmaline Fisher from the University of Cincinnati Law School Class of '23 who also provided excellent research assistance and in particular is responsible for the data on companies advertising their use of AI to assist in recruiting subjects. Thank you also to Madeline J. Bard for your encouragement. This Article is in memory of the late Professor Robert A. Weninger, J. Hadley Edgar Professor of Law, 1974, Texas Tech University School of Law, a brilliant scholar, gifted teacher, irreplaceable mentor, and dear friend.

ABSTRACT

This is the first article to call for a ban on the use of AI technology designed to influence human decision-making, "Persuasive AI," for the purpose of recruiting or enrolling human participants in drug trials sponsored by commercial entities. It does so from a perspective of precaution, not fear. Advances in Artificial Technology that can assist human decision-making have tremendous potential for good.[1] *It makes the case for doing so based on both the substantial risk of harm to the decision-making process and the ineffectiveness of intermediate regulatory measures. This Article looks directly at Persuasive AI, a type of AI that claims to respond directly to the emotions of the humans with which it interacts. There is already considerable evidence of its ability to analyze data in health and military settings not just faster than humans but beyond human capacity. But there is also growing international concern about characteristics associated with "Emotion AI," such as its persistent reproduction of societal biases and ability to develop beyond its programming, that mitigate against its use in specific, sensitive situations such as health care and the military. The challenge of mitigating these risks is that even the people who develop the programs do not know how it is making decisions and cannot intervene in ways that might prevent harm. All of these factors have led the EU to propose a ban on its use in a broad range of spheres, including health care where manipulating decision-making would be a violation of an individuals' human rights. While there is at present no direct federal regulation of Persuasive AI in the United States, concerns about the risks of biased or manipulated decision-making has led to calls*

1. *See* Orly Lobel, *The Law of AI for Good* 27–41 (Univ. of San Diego Sch. of L., Rsch. Paper No. 23-001, 2023), https://papers.ssrn.com/abstract=4338862 [https://perma.cc/C79M-JMFT].

within the United States for a moratorium on its use in settings where it is most likely to cause harm. In response, some states and cities have passed or are contemplating legislation to limit its use in law enforcement and employment decisions. As this Article discusses, one of the reasons the United States has been behind other nations is a general disinclination to interfere with the process of adult decision-making. So long as there is no fraud, threat, or deliberate deception, adults are assumed to be competent to evaluate the claims of those trying to persuade them without government protection.

The federal laws regulating obtaining informed consent for biomedical research is a dramatic exception to this mostly hands-off approach. Often collectively referred to as the"Common Rule," these laws were promulgated in direct response to the unethical behavior of the U.S. government in withholding treatment from Black Sharecroppers (formally known as the Tuskegee Syphilis Experiment).[2] *Specifically, these laws create a system in which an ethics committee conducts a preemptive ethical review of all information, including advertisements, generated by the sponsor of the trial to potential participants. Such a review would be meaningless if, as this Article details, the party seeking consent employs a technology that can influence decision-making in ways beyond the ability of the ethics committees to detect.*

While the Common Rule's jurisdiction extends to all forms of human subject research, this Article focuses on clinical drug trials by commercial

---

2. Information about the depth of the racism and extent of harm caused by this study continues to emerge and while it is probably still necessary to use *Tuskegee* for clarity, the U.S. government agrees that the U.S. Public Health Service bears complete responsibility. For an example of emerging information about the effects from the study fifty years later, see *A Fund Apologizes for its Role in the Tuskegee Syphilis Study that Targeted Black Men*, NPR (June 11, 2022, 1:38 PM), https://www.npr.org/2022/06/11/1104386467/tuskegee-syphilis-study-milbank-memorial-fund-apology [https://perma.cc/JAK4-3F4J] (reporting the apology by the Milbank Memorial Fund that, as was recently discovered, offered on the government's behalf $100 to the families of participants who died for funeral expenses in exchange for permission to conduct an autopsy on the body); *see also* Hadley Hitson, *Health Disparities Persist in Tuskegee 50 Years After end of Unethical Syphilis Study*, MONTGOMERY ADVERTISER (Nov. 27, 2022, 9:01 PM), https://www.montgomeryadvertiser.com/story/news/local/2022/11/28/tuskegee-syphilis-study-impacts-macon-county-residents-50-years-later/69666542007/ [https://perma.cc/GVN5-P9CR] (reporting that residents of Macon County still have mistrust in health care deficiency studies because of the lingering impact of the syphilis study); *About the USPHS Syphilis Study*, TUSKEGEE UNIV., https://www.tuskegee.edu/about-us/centers-of-excellence/bioethics-center/about-the-usphs-syphilis-study [https://perma.cc/6M8R-CBJU] (describing the role of the U.S. Public Health Service in the study).

entities because they are, by definition, done for the purpose of marketing a product for the financial benefit of the trial's sponsor.

Although sometimes enrollment in a clinical trial is the only path to lifesaving treatment, in many cases the drug being studied is a variation of a product already being sold by another company. Sponsors of these "me-too" trials often find it difficult to enroll the number of patients required to complete the study. New regulations by the FDA intended to encourage the enrollment of populations disproportionately absent from drug trials, primarily Black adults, make the recruiting process for sponsors even more difficult. This is demonstrated by original research in this Article showing that many companies are marketing AI to sponsors to assist in obtaining a diverse pool of subjects. While there is no suggestion that they are offering to exert undue influence, the inability to control what these algorithms are doing supports a call for restraint. It would be especially unfortunate if these efforts to increase diversity create an incentive to manipulate and coerce the very populations originally exploited by the Public Health Service in Alabama. So, without alleging ill intent on the part of anyone involved in developing, marketing, or using this software to recruit participants or obtain their consent to participate in clinical drug trials, this Article argues that Persuasive AI's capacity to undermine the free will of potential participants in ways that are both undetectable and impossible to remediate.

I. INTRODUCTION

Recent calls by leading developers of computer software known as artificial intelligence (AI) technology to make "[m]itigating the risk of extinction from AI . . . a global priority alongside other societal-scale risks such as pandemics and nuclear war" have finally triggered the kind of urgent attention from the U.S. government to a threat that regulators outside the United States have been confronting for years.[3] While the United States has so far not passed a single law intended to prevent or mitigate any harm from the use of AI, the European Union (EU), for example, has already endorsed a sliding-scale regulatory framework that imposes limitations on the use and development of AI technology based on the application's risk classification.[4] It specifically prohibits the use of AI systems that either "exploit vulnerabilities of specific vulnerable groups . . . to materially distort their behavior in a manner that is likely to

3. *See* Michael Frank, *Managing Existential Risk from AI Without Undercutting Innovation*, CSIS (July 10, 2023), https://www.csis.org/analysis/managing-existential-risk-ai-without-undercutting-innovation [https://perma.cc/ZS88-FZRQ].

4. *See id.* ("The most advanced AI regulatory effort is the European Union, whose parliament recently passed its version of the Artificial Intelligence Act (AI Act).").

cause them or another person psychological or physical harm" or "have a significant potential to manipulate persons through subliminal techniques."[5] AI's ability to influence decision-making has advanced considerably even in the few years since the EU's first call for a moratorium.[6] This Article will, based primarily on the claims that those developing and marketing this technology make for it, provide an overview of the ways in which it is already being used to manipulate decision-making. It will then make the case that AI's current ability to manipulate decision-making is incompatible with the kind of voluntary consent promised not just to those directly affected by the Tuskegee Syphilis Experiment but to the entire population of potential research subjects. AI achieves this manipulation in several different ways.

For reasons that remain surprisingly unexplored, U.S. law takes a hands-off approach to most persuasive activity so long as there is no actual intentional misrepresentation. Perhaps this is based on a cultural belief in the inherent equality among competent adults that puts them on equal footing in situations where it is in the interest of one person to persuade another to buy a product, vote for a candidate, agree to the terms of a mortgage, or render a jury verdict.

But, as the EU and other global regulators have recognized, the existing ability of AI technology to confer a persuasive advantage beyond any that has been previously recognized, as well as considerable evidence of how it is already being marketed in a wide variety of settings, demands that it be viewed as a threat deserving of special attention.

While the United States still lacks the EU's framework for rights-based AI regulation, this particular use of AI to persuade or manipulate decision-

5. *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts* § 5.2.2., at 12–13 COM (2021) 206 final (Apr. 21, 2021) [hereinafter *Proposed EU AI Act*]; *see also EU: Ban on Most Harmful Use of AI Moves a Step Closer*, AMNESTY INT'L (May 11, 2023), https://www.amnesty.org/en/latest/news/2023/05/eu-ban-on-most-harmful-use-of-ai-moves-a-step-closer/ [https:// perma.cc/2STF-F2QN] ("Today the European Parliament sent a strong signal that human rights must be at the forefront of this landmark legislation, by voting to ban several AI-based practices which are incompatible with human rights.").

6. *See* MICHAEL L. LITTMAN ET AL., GATHERING STRENGTH, GATHERING STORMS: THE ONE HUNDRED YEAR STUDY ON ARTIFICIAL INTELLIGENCE (AI100) 2021 STUDY PANEL REPORT 12 (2021), https://ai100.stanford.edu/sites/g/files/sbiybj18871/files/media/file/AI100Report_MT_10.pdf [https://perma.cc/WF47-Z2Q3] ("People are using AI more today to dictate to their phone, get recommendations for shopping, news, or entertainment, enhance their backgrounds on conference calls, and so much more.").

making does violate laws that protect the right of participants in federally regulated clinical drug trials to make a free and voluntary choice to participate. These laws were passed in direct response to one of the worst violations of trust by the U.S. government against its own citizens in modern history in what has come to be known as the Tuskegee Syphilis Experiment.

This Article, therefore, calls for the immediate halt of the use of this technology in all federally regulated human subject research. In so doing, it identifies the threat this technology poses to decision-making for the benefit of future regulatory efforts to protect a broader scope of decision-making. The EU characterizes this harm as a violation of ways in which it can be misused to violate the fundamental human right of having autonomy over our own lives.[7]

Current federal law that protects the rights of patients to make their own fully informed decision about whether to enroll in a research study without coercion stems directly from the 1974 discovery that the United States Public Health Service had, for over thirty years, watched a group of Black sharecroppers while they suffered from and transmitted to their families a completely curable but deadly disease: syphilis.[8] The public outrage that these men were neither asked to participate nor told that they were being studied but rather led to believe they were part of a government treatment program led to the passing of a series of laws that still protect research participants today.[9] They do so by creating a protective barrier between potential participants in federally funded and regulated research and those who wish to benefit from studying them.[10] But these rules, which require preapproval of all information provided to

---

7. *Proposed EU AI Act*, *supra* note 5, § 3.5., at 11; *see also* Mark MacCarthy & Kenneth Propp, *Machines Learn that Brussels Writes the Rules: The EU's New AI Regulation*, BROOKINGS (May 4, 2021), https://www.brookings.edu/blog/techtank/2021/05/04/machines-learn-that-brussels-writes-the-rules-the-eus-new-ai-regulation/ [https://perma.cc/S9NK-W5U9] (discussing the circumstances under which the EU guideline bans use of AI assisted decision-making rather than attempting to limit or monitor its use).

8. *See* Carol A. Heintzelman, *The Tuskegee Syphilis Study and its Implications for the 21st Century*, SOC. WORKER (2003), https://www.socialworker.com/feature-articles/ethics-articles/The_Tuskegee_Syphilis_Study_and_Its_Implications_for_the_ 21st_Century/ [https://perma.cc/9ZZ4-YXRQ].

9. *See id.*

10. This study is a touchstone in the study of research ethics both because of the horror of what happened and because its discovery and the public outcry that followed moved Congress to pass the laws that protect participants today. However, it is important not to mythologize what happened nor to think the events so terrible that they could not happen today. For a historical account of the Tuskegee Syphilis Experiment, see Susan M. Reverby, *Ethical Failures and History Lessons: The U.S. Public Health Service Research Studies in Tuskegee and Guatemala*, 34 PUB. HEALTH REVS., June 3, 2012, at 1, 4–7.

potential participants and prohibit tactics that might pressure enrollment or promise unachievable benefits, are not strong enough to withstand technology that can already manipulate human decision-making to the advantage of the entity deploying it. Often called "Persuasive AI,"[11] these programs, whose mechanism of achieving these goals is unknown even to those creating them, have, over the past two years, sparked calls for a moratorium from the EU, the World Health Organization (WHO), and even a former U.S. Secretary of Defense.[12]

Original research conducted for this Article reveals that companies specializing in sponsoring clinical drug trials are already offering for AI to assist in identifying and recruiting participants.[13] These pitches promise that the technology will enroll eligible subjects quickly at less cost. In particular, many clinical trial sponsors emphasize their commitment to the new federal requirement to diversify their trials by recruiting more Black participants.[14] So far, none of these advertisements suggest that they are using technology to manipulate decision-making. But the claims that AI companies make for their services in other contexts suggests that if that is not already happening, it will be soon. While U.S. law does not, so far, restrict the use of persuasive technology in marketing products to consumers, the

11. A note on terminology. There is no central entity charged with assigning names to different forms of AI and very little consistency among those who use and study it. Since the purpose of this Article is to discuss a version of AI that claims to use an ability to identify human emotions for the purpose of enhancing persuasion, this Article adopts one of its many names, "Persuasive AI" as more neutral than an equally descriptive term "Manipulative AI" and more likely to be understood than the EU's adoption of the phrase "subliminal AI." *See* James Orme, *Persuasive AI Could Corrupt Human Behaviour, Study Suggests*, TECHERATI (Feb. 17, 2021), https://www.techerati.com/news-hub/ persuasive-ai-could-corrupt-human-behaviour-study-suggests/ [https://perma.cc/A9L7-XB4W] (discussing the effect of Persuasive AI on human behavior).

12. *See* Jennifer S. Bard, *AI's Ability to Manipulate Decision Making Requires a Moratorium on its Use in Obtaining Consent for Biomedical Research*, HARV. L. PETRIE FLOM CTR. (July 14, 2023), https://blog.petrieflom.law.harvard.edu/2023/07/14/ais-ability-to-manipulate-decision-making-requires-a-moratorium-on-its-use-in-obtaining-consent-for-biomedical-research/ [https://perma.cc/2SU5-G64W].

13. For a definition of "clinical trial," see *Clinical Trial*, NAT'L CANCER INST., https://www.cancer.gov/publications/dictionaries/cancer-terms/def/clinical-trial [https://perma.cc/5PUJ-4BGQ] ("A type of research study that tests how well new medical approaches work in people. These studies test new methods of screening, prevention, diagnosis, or treatment of a disease. Also called clinical study.").

14. *See* Brian Bossetta, *Legal Expert Weighs in on New US Mandate to Diversify Clinical Trials*, MEDTECH INSIGHT: CITELINE COM. (Feb. 3, 2023), https://medtech.pharmaintelligence.informa.com/MT147606/Legal-Expert-Weighs-In-On-New-US-Mandate-To-Diversify-Clinical-Trials [https://perma.cc/H8ME-ZXNV].

law that protects potential research subjects is quite different. Existing AI technology that can manipulate decision-making to the benefit of those seeking to enroll participants is incompatible with the legally established ethical standards for obtaining informed consent. Moreover, since this technology has been shown to produce results that reflect existing societal discrimination based on race, it is particularly unsuited for the purpose of achieving racial diversity in clinical trials.

This Article makes the case that the government entities enforcing federal human subject research protection laws should act immediately to designate the use of AI with the potential of manipulating decision-making as an instrument of "coercion" or "undue persuasion." This would immediately set a boundary for both the entities conducting federally funded research in the United States and the private companies conducting clinical drug trials in the United States and abroad.

This Article will proceed as follows. First, it will explain AI's capacity to manipulate human decision-making and current efforts in the EU and United States to prevent that from happening. Then it will look closely at how existing federal law protects the informed consent process and Persuasive AI undermines that process. Having framed the problem, the Article will then look more closely at the specific ways in which the pressures on the pharmaceutical industry to complete clinical drug trials make the temptation to use manipulative technology irresistible. Finally, it will justify the call for a ban by highlighting the features of research protection law and Persuasive AI that render previously suggested intermediate measures insufficient to prevent irreparable harm.

This Article concludes by calling on the divisions of the U.S. federal government charged with interpreting and enforcing existing laws to use their existing authority, as they did during the pandemic, to communicate the risks posed by Persuasive AI to informed consent and demand an immediate cessation of its use in the recruitment and enrollment of participants in clinical trials. This Article also calls for them to immediately gather experts in Persuasive AI to study its effects on other forms of human participant research under its jurisdiction. Additionally, these entities should exercise their leadership role to communicating with the nongovernmental entities involved in research protection in the United States as well as the nations and confederations that provide similar protection to research participants all over the world.

II. HOW AI TECHNOLOGY MANIPULATES DECISION-MAKING

"[A]rtificial Intelligence (AI) has woven itself into our daily lives in ways [which] we may not even be aware of."[15]

A 2018 report by the Brookings Institute described AI as "a wide-ranging tool that enables people to rethink how we integrate information, analyze data, and use the resulting insights to improve decision-making."[16] But a closer look reveals that, unlike other tools that improve decision-making by adding information, such as a map or other reference tool, AI is offering something more than information.[17] The narrative that AI can make better decisions than a human led to AI integration into a vast range of decisions, such as medical treatment, employment, credit limits, and even acceptance of law review articles.[18] But with recent advances, AI has transcended its role as a tool supporting human decision-making and has become an actual decision maker itself.[19]

Examples of this kind of AI-assisted decision-making are numerous and so integrated into daily life that they have become difficult to identify.[20]

15. Ruth Brooks, *Artificial Intelligence and its Impact on Everyday Life*, UNIV. OF YORK, https://online.york.ac.uk/artificial-intelligence-and-its-impact-on-everyday-life [https://perma.cc/MXA4-WUY2]; *see also* Christina Pazzanese, *Great Promise but Potential for Peril: Ethical Concerns Mount as AI Takes Bigger Decision-Making Role in More Industries*, HARV. GAZETTE (Oct. 26, 2020), https://news.harvard.edu/gazette/story/2020/10/ethical-concerns-mount-as-ai-takes-bigger-decision-making-role/ [https://perma.cc/9W7B-WGSE] ("Virtually every big company now has multiple AI systems and counts the deployment of AI as integral to their strategy . . . .").

16. Darrell M. West & John R. Allen, *How Artificial Intelligence is Transforming the World*, BROOKINGS (Apr. 24, 2018), https://brookings.edu/research/how-artificial-intelligence-is-transforming-the-world/ [https://perma.cc/ZC8X-7X9K].

17. *See* AMY TURNER ET AL., CALIBRATING TRUST IN AI-ASSISTED DECISION MAKING 1 (2020), https://www.ischool.berkeley.edu/sites/default/files/sproject_attachments/humanai_capstonereport-final.pdf [https://perma.cc/XGR2-TQ24] ("AI-assisted decision-making . . . [occurs where] the individual strengths of a human and AI combine in order to produce a decision outcome that is better than what either could produce alone.").

18. *See* Brenda M. Simon, *Using Artificial Intelligence in the Law Review Submissions Process*, 56 U.C. DAVIS L. REV. 347, 358, 361, 373, 403 (2022).

19. *See* Philip Meissner & Christoph Keding, *The Human Factor in AI-Based Decision-Making*, MIT SLOAN: MGMT. REV. (Oct. 12, 2021), https://sloanreview.mit.edu/article/the-human-factor-in-ai-based-decision-making/ [https://perma.cc/FD97-NMLC] ("AI now transcends mere process automation and . . . is being increasingly used to augment decision-making processes at all levels, including top management.").

20. *See id.* ("AI now has a firm footing in organizations' strategic decision-making processes. Five years ago, less than 10% of large companies had adopted machine learning or other forms of AI, but today 80% of them make use of the technology.").

While it may appear that AI is simply providing information, like a spreadsheet or a map, researchers have found that people using AI-generated decisions view it as more reliable than their own decisions.[21]

AI analysis of external information, such as a map or even a calculator, can be characterized as "assisted decision-making" in the sense that it provides information on which a decision can be based.[22] But what makes AI-assisted decision-making that incorporates persuasive technology different from past versions of the technology is that it neither makes the decision itself nor presents information to a human decision maker. Rather, it engages with the intent to persuade the human decision maker to make particular decisions.[23]

Persuasive AI, the technology under scrutiny here, is the outgrowth of decades of increasingly complex computer programs that use algorithms to take in large amounts of information and analyze it to reach a better solution than an individual human could find on their own.[24]

A. Evolution of AI-Assisted Decision-Making

What differentiates Persuasive AI from other forms of AI that analyzed previously collected data is that it claims to work in real time by analyzing the behavior of humans it is observing.[25] Often called "Emotion AI" or "Affective Computing," these technologies analyze information from a live camera feed to recognize, interpret, and respond to human reactions to information.[26] So, for example, if a consumer was not enjoying a visual image in an advertisement, the image could switch to another more appealing

---

21. Shiye Cao & Chien-Ming Huang, *Understanding User Reliance on AI in Assisted Decision-Making*, 6 PROC. ACM HUM.-COMPUT. INTERACTION, Nov. 11, 2022, at 1, 1 ("AI-assisted human decision-making aims to augment the human-AI team performance to exceed both parties' individual performances.").

22. *See* Jon Taylor, *AI Decision-Making: The Future of Business Intelligence*, PEAK (Aug. 23, 2021), https://peak.ai/hub/blog/ai-decision-making-the-future-of-business-intelligence/ [https://perma.cc/4EAT-437M].

23. *See* Eric Colson, *What AI-Driven Decision-Making Looks Like*, HARV. BUS. REV. (July 8, 2019), https://hbr.org/2019/07/what-ai-driven-decision-making-looks-like [https://perma.cc/PH5C-W7DJ] ("The value of AI is making better decisions than what humans alone can do.").

24. *Id.* ("Connected devices now capture unthinkable volumes of data: every transaction, every customer gesture, every micro- and macroeconomic indicator, all the information that can inform better decisions.").

25. *See* Orme, *supra* note 11 (discussing the effect of Persuasive AI on human behavior).

26. *See* Cem Dilmegani, *Affective Computing: In-Depth Guide to Emotion AI in 2023*, AIMULTIPLE (Apr. 10, 2023), https://research.aimultiple.com/affective-computing/ [https://perma.cc/JM5D-X58A].

image.[27] This is sometimes described as responding to emotions and is marketed for two purposes: persuading consumers and keeping the peace.[28]

1. Origins to Facial Recognition Software

Given the extraordinary claims today's Persuasive AI is making for itself, it is reasonable to ask where this ability came from. By technology application developers' own accounts, today's algorithms are a direct outgrowth of the biometric applications that drive the facial recognition we all use every time we pick up our phones.[29] At its most basic level, this is a form of matching technology that compares the face in front of the camera to a picture of the device's authorized user.[30] Unlike current forms of AI whose promises are difficult to test, facial recognition software coding is amenable to external verification through accuracy testing.[31] For example, every day, people use facial recognition technology to open their Apple and Android phones.[32] The next stage of facial recognition,

27. For an example of what this form of AI claims it can do, see EmotionTrac, *Demo*, VIMEO (Feb. 13, 2023, 1:25 PM), https://vimeo.com/798480204 [https://perma.cc/8C64-36E5]; *see also* Stephen Gossett, *Emotion AI: 3 Experts on the Possibilities and Risks*, BUILT IN (Feb. 7, 2023), https://builtin.com/artificial-intelligence/emotion-ai [https://perma.cc/5WWH-HHPC] ("Emotion AI's ability to capture and analyze human body language and emotions means that the products and services made with it will be more considerate of the user's needs and feelings.").

28. For more examples of Emotion AI and its claims, see Cem Dilmegani, *Top 10 Emotional AI Examples & Use Cases in 2023*, AIMULTIPLE (Jan. 2, 2023), https://research.aimultiple.com/emotional-ai-examples/ [https://perma.cc/DLM6-7BED].

29. *See* Sasha Reeves, *15 Best Face Recognition Apps: A Detailed Guide for 2023*, GOODCORE (Mar. 9, 2023), https://www.goodcore.co.uk/blog/face-recognition-apps/ [https://perma.cc/F4Y5-F82U] ("[F]ace recognition technology works by identifying specific nodal points of a human face. Some face recognition software can identify as many as 80 nodal points of a face."); *see also* Jessica G. Cino, *Facial Recognition is Increasingly Common, but How Does it Work?*, THE CONVERSATION (Apr. 4, 2017, 9:09 PM), https://theconversation.com/facial-recognition-is-increasingly-common-but-how-does-it-work-61354 [https://perma.cc/YE26-J2H3].

30. Various social media platforms use this identification feature differently. *See, e.g.*, Bogdan Bele, *How to Use the Instagram Image Tagging Feature*, GROOVY POST (Jan. 27, 2020), https://www.groovypost.com/howto/instagram-tagging-feature/ [https://perma.cc/2LY6-9BUK].

31. *See* Mark Purdy, John Zealley & Omaro Maseli, *The Risks of Using AI to Interpret Human Emotions*, HARV. BUS. REV. (Nov. 18, 2019), https://hbr.org/2019/11/the-risks-of-using-ai-to-interpret-human-emotions [https://perma.cc/EC95-5D3H].

32. Robert E. Wells III, *The 4 Best Face Recognition Apps for Android in 2023*, LIFEWIRE (Jan. 23, 2023), https://www.lifewire.com/best-face-recognition-apps-android-4590312 [https://perma.cc/HQ7B-YKCL].

also on most mobile devices, matches not just the image authorized by its user but also those images it "learn[s]" about on its own.[33] This activity is often called "Deep Learning," which is distinguishable from "Machine Learning."[34]

2. Government Use of Facial Recognition Technology

By some accounts, facial recognition technology was invented in the 1960s when it was used to "manually record the coordinate areas of facial features like eyes, nose, mouth, and hairline" that could then be matched to a photograph.[35] According to an article by the National Institute of Justice, the research, development, and evaluation agency of the U.S. Department of Justice, it "began funding face detection and recognition research in 1996."[36]

There are public accounts of the military's use of facial recognition technology that date as far back as the 1990s.[37] Facial recognition technology soon attracted more military attention, and, over the past decade, it has been adopted by many law enforcement agencies around the country to scan crowds in real-time as well as to identify individuals captured later

---

33. *See* Adrian Bridgwater, *The Real-Time AI Data Race is on*, FORBES (Jan. 12, 2023, 9:09 AM), https://www.forbes.com/sites/adrianbridgwater/2023/01/12/the-real-time-ai-data-race-is-on/ [https://perma.cc/9QQC-ZNUR] ("The future of real-time data is AI—soon, all applications will leverage real-time data and AI to provide the next-best offer, recommendation, or course of action . . . .").

34. For a comparison between Machine Learning and Deep Learning, see *What is Machine Learning*, IBM, https://www.ibm.com/topics/machine-learning [https://perma.cc/JW2S-96ME] ("Deep learning can ingest unstructured data in its raw form (e.g., text or images), and it can automatically determine the set of features which distinguish different categories of data from one another. This eliminates some of the human intervention required and enables the use of larger data sets. . . . Classical, or 'non-deep', machine learning is more dependent on human intervention to learn. Human experts determine the set of features to understand the differences between data inputs, usually requiring more structured data to learn.").

35. Divyesh Dharaiya, *History of Facial Recognition Technology and its Bright Future*, READWRITE (Mar. 12, 2020), https://readwrite.com/history-of-facial-recognition-technology-and-its-bright-future/ [https://perma.cc/7FUM-TJFW].

36. *History of NIJ Support for Face Recognition Technology*, NIJ (Mar. 5, 2020), https://nij.ojp.gov/topics/articles/history-nij-support-face-recognition-technology [https://perma.cc/QV3Y-KHLW].

37. Dharaiya, *supra* note 35 ("[I]n the 1993-2000s period, DARPA and NIST released the FERET program to encourage the commercial facial recognition market. In 2002, law enforcement officials applied facial recognition in critical technology testing."); *see also* Shaun Raviv, *The Secret History of Facial Recognition*, WIRED (Jan. 21, 2020, 6:00 AM), https://www.wired.com/story/secret-history-facial-recognition/ [https://perma.cc/AA4G-CSD9] (recognizing that facial recognition research conducted during the 1960s "prefigured all these technological breakthroughs and their queasy ethical implications").

in surveillance footage.[38] It can also be used retrospectively to identify individuals present at an event like a protest or a crime scene.[39] Privacy and human rights advocates have protested its use by the government.[40] These concerns have not, however, stopped its growth. According to an article by the Brookings Institute, "of the approximately 42 federal agencies that employ law enforcement officers, the Government Accountability Office (GAO) discovered in 2021 that about 20, or half, used facial recognition."[41] Facial recognition AI is also used by law enforcement authorities in Europe.[42] But familiarity with the software has not resulted in community acceptance. In 2021, the EU Parliament voted in favor of banning the use of facial recognition technology in law enforcement in public spaces.[43] While the EU has not yet enacted this ban, it demonstrates a widespread concern about the potential for misuse.[44]

---

38. *See* Mark MacCarthy, *Mandating Fairness and Accuracy Assessments for Law Enforcement Facial Recognition Systems*, BROOKINGS (May 26, 2021), https://www.brookings.edu/articles/mandating-fairness-and-accuracy-assessments-for-law-enforcement-facial-recognition-systems/ [https://perma.cc/57JV-QYSC].

39. *See id.*

40. *See, e.g.*, Nicole Ozer, Kate Ruane & Matt Cagle, *Grassroots Activists are Leading the Fight to Stop Face Recognition. It's Time for Congress to Step Up, Too*, ACLU (June 17, 2021), https://www.aclu.org/news/privacy-technology/grassroots-activists-are-leading-the-fight-to-stop-face-recognition-its-time-for-congress-to-step-up-too [https:// perma.cc/ Z2GL-DGTV].

41. Nicol T. Lee & Caitlin Chin-Rothmann, *Police Surveillance and Facial Recognition: Why Data Privacy is Imperative for Communities of Color*, BROOKINGS (Apr. 12, 2022), https://www.brookings.edu/research/police-surveillance-and-facial-recognition-why-data-privacy-is-an-imperative-for-communities-of-color/ [https://perma.cc/S9QZ-C7FY].

42. Luca Bertuzzi, *Facial Recognition Technologies Already Used in 11 EU Countries and Counting, Report says*, EURACTIV (Oct. 27, 2021), https://www.euractiv.com/section/data-protection/news/facial-recognition-technologies-already-used-in-11-eu-countries-and-counting-report-says/ [https://perma.cc/5GF2-QUZ6]. For a compelling overview of the threat to human dignity and freedom against boundless power posed by facial recognition technology, see Tate Ducker, *Orwell's* 1984 *"Big Brother" Concept and the Government Use of Facial Recognition Technology: A Call to Action for Regulation to Protect Privacy Rights*, 8 BELMONT L. REV. 600, 633–34 (2021).

43. Lisa Peets et al., *European Parliament Votes in Favor of Banning the Use of Facial Recognition in Law Enforcement*, COVINGTON: INSIDE PRIVACY (Oct. 12, 2021), https://www.insideprivacy.com/artificial-intelligence/european-parliament-votes-in-favor-of-banning-the-use-of-facial-recognition-in-law-enforcement/ [https://perma.cc/N4CW-JJGG].

44. *See* Clothilde Goujard, *Europe Edges Closer to a Ban on Facial Recognition*, POLITICO (Sept. 20, 2022, 6:00 AM), https://www.politico.eu/article/europe-edges-closer-to-a-ban-on-facial-recognition/ [https://perma.cc/HWC8-NSKY].

3. Commercial Use of Facial Recognition Technology

Outside of law enforcement, facial recognition technology is also deeply integrated into most people's daily lives.[45] This is the technology that prompts social media users to tag otherwise unidentified people in group photos.[46] By 2010, Facebook was already integrating facial integration into its social media platform.[47] Today, facial recognition technology has become integrated into everyday life. However, as it becomes more ubiquitous, concerns about its intrusion are increasing.[48] In 2021, Disney World announced that it was testing facial recognition software at the park's entry gate.[49]

But in the relatively short time—about ten years—that facial recognition software has been in wide scale use, it has also generated considerable controversy about its ability to invade privacy and its tendency towards racial discrimination.[50] In 2021, Facebook pledged to remove facial

---

45. For an article on the use of facial recognition software around the world, see Paul Bischoff, *Facial Recognition Technology (FRT): 100 Countries Analyzed*, COMPARITECH (Jan. 24, 2022), https://www.comparitech.com/blog/vpn-privacy/facial-recognition-statistics/#:~:text=Around%2040%20percent%20of%20countries,the%20transmission%20of%20COVID%2D19 [https://perma.cc/RM85-3KM2].

46. *See* Greenwolf, *Mapping Social Media with Facial Recognition*, MEDIUM (Mar. 5, 2019), https://medium.com/greenwolf-security/mapping-social-media-with-facial-recognition-a-new-tool-for-penetration-testers-and-red-teamers-3b70e5da5f5c [https://perma.cc/45W7-34ZX] (explaining the use of Greenwolf Security's software that searches social media for pre-identified marketing targets by scanning individuals' profile pictures and performing facial recognition checks to try and find a match).

47. Dharaiya, *supra* note 35 ("When 2010 started, Facebook started using a facial recognition feature that helped detect people with featured faces in the photos updated by Facebook users. While the update created hype in the media industry—Facebook stayed very low key since there was no apparent negative impact on website popularity and usage.").

48. *See* Thorin Klosowski, *Facial Recognition is Everywhere. Here's What We Can Do About it*, N.Y. TIMES: WIRECUTTER (July 15, 2020), https://www.nytimes.com/wirecutter/blog/how-facial-recognition-works/ [https://perma.cc/QGC4-7J9X].

49. *See* Quincy Stanford, *How Disney World's New Facial Recognition Technology Actually Works*, DISNEY FOOD BLOG, https://www.disneyfoodblog.com/2021/03/26/how-disney-worlds-new-facial-recognition-technology-actually-works/ [https://perma.cc/9KUK-F2YN]. Apparently, the experiment was a brief one. *See* Tom Bricker, *Disney World Ends Facial Recognition Test*, DISNEY TOURIST BLOG (May 3, 2021), https://www.disneytouristblog.com/disney-world-testing-facial-recognition-for-entering-at-magic-kingdom/ [https://perma.cc/6MVF-JC9P] ("Roughly a week after extending the test indefinitely, Walt Disney World has quietly ended the facial recognition test for park entry. All equipment has been removed from the turnstiles at Magic Kingdom, signage is gone, and even the official website offering details about the test has been removed.").

50. *See, e.g.*, Eileen Guo, *A Roomba Recorded a Woman on the Toilet. How Did Screenshots End Up on Facebook?*, MIT TECH. REV. (Dec. 19, 2022), https://www.technologyreview.com/2022/12/19/1065306/roomba-irobot-robot-vacuums-artificial-intelligence-training-data-privacy/ [https://perma.cc/84UJ-MMX5]; *see also* Ozer, Ruane

recognition software from its platform because of societal concerns about privacy.[51] Today, despite these concerns, facial recognition software is still widely used in stadiums and commercial establishments.[52]

B. From Recognizing Faces to Recognizing Emotions

The next phase in the evolution of facial recognition technology was the development of software that, its developers assert, can use cameras to establish not just a person's identity but also their state of mind.[53] It claims to do so by matching a person's facial expressions to a database of labeled human emotions.[54] This claim is contested.[55] Proponents of what is sometimes called Emotion AI claim that it is based on the work of Dr. Paul Ekman, on behalf of the Department of Homeland Security, who tested its ability to identify "terrorists" and intervene before they could cause harm.[56] Dr. Ekman did this by creating the Screening of Passengers by Observation Techniques (SPOT) program to assist the Transportation Security Administration (TSA) in identifying nervous passengers who might be planning harm.[57] However, SPOT turned out to be so ineffective

& Cagle, *supra* note 40 (arguing that facial recognition surveillance "supercharges the government's power to surveil people of color and other marginalized groups").

51. *See* Khari Johnson, *Facebook Drops Facial Recognition to Tag People in Photos*, WIRED (Nov. 2, 2021, 7:38 PM), https://www.wired.com/story/facebook-drops-facial-recognition-tag-people-photos/ [https://perma.cc/U8RV-JLQ7].

52. *See* James Barron, *Who's Using Facial Recognition Technology?*, N.Y. TIMES (Mar. 21, 2023), https://www.nytimes.com/2023/03/21/nyregion/facial-recognition-technology-garden.html [https://perma.cc/6LU8-JSJS] (discussing the large number of stores using facial recognition technology found whilst on a walk in New York City); *see also Facial Recognition Market Size, Share & Trends Analysis Report*, GRAND VIEW RSCH., https://www.grandviewresearch.com/industry-analysis/facial-recognition-market [https://perma.cc/XR5H-3F7V].

53. *See, e.g.*, Affectiva, *Affectiva Emotion Analytics Dashboard - Affdex*, YOUTUBE (Oct. 27, 2016), https://www.youtube.com/watch?v=87eJb19xqbw [https://perma.cc/NH6F-E3L4] (explaining a company's use of emotion analytics to provide insight into consumer emotional response).

54. *See id.*

55. *See* Kate Crawford, *Artificial Intelligence is Misreading Human Emotion*, THE ATLANTIC (Apr. 27, 2021), https://www.theatlantic.com/technology/archive/2021/04/artificial-intelligence-misreading-human-emotion/618696/ [https://perma.cc/Y853-9FZG].

56. *See* Oscar Schwartz, *Don't Look Now: Why You Should be Worried About Machines Reading Your Emotions*, THE GUARDIAN (Mar. 6, 2019), https://www.theguardian.com/technology/2019/mar/06/facial-recognition-software-emotional-science [https://perma.cc/F8XZ-MVCM].

57. *See id.*

that the Government Accountability Office recommended limiting its use because "[a]vailable evidence does not support whether behavioral indicators . . . can be used to identify persons who may pose a risk to aviation security."[58] Even worse, the American Civil Liberties Union (ACLU) sued the TSA on the grounds that a study they commissioned found that SPOT was not only inaccurate, but also racist.[59] By disproportionately misidentifying Muslim and Latino passengers, it had effectively permitted harassment.[60] Yet, as with facial recognition software, AI that can recognize emotions as well as faces has become ubiquitous in military, law enforcement, and commercial settings despite these concerns.[61]

1. Commercial Use of Emotion AI

> *"Warm, personable, convincing—those adjectives may not be the first that come to mind when you think of artificial intelligence or AI. But, AI and human collaboration can outperform human intuition alone in creating and selecting digital content that connects with customers."*[62]

---

58. *See* U.S. GOV'T ACCOUNTABILITY OFF., GAO-14-159, AVIATION SECURITY: TSA SHOULD LIMIT FUTURE FUNDING FOR BEHAVIOR DETECTION ACTIVITIES 1, 15 (2013), https://www.gao.gov/assets/gao-14-159.pdf [https://perma.cc/P83W-XLR9].

59. *ACLU v. TSA*, ACLU (Feb. 8, 2017), https://www.aclu.org/cases/aclu-v-tsa [https://perma.cc/X8BY-R5TV] ("The SPOT program lacks a scientific basis, is wholly ineffective, and has given rise to allegations of racial profiling.").

60. Spencer Ackerman, *TSA Screening Program Risks Racial Profiling Amid Shaky Science – Study*, THE GUARDIAN (Feb. 8, 2017, 7:00 AM), https://www.theguardian.com/us-news/2017/feb/08/tsa-screening-racial-religious-profiling-aclu-study [https://perma.cc/JEY4-ZCL6]; *see also* Mary May, *Racism and Exploitation in Phase I Clinical Trials*, HARV. UNIV.: SCI. IN THE NEWS (Oct. 24, 2020), https://sitn.hms.harvard.edu/flash/2020/racism-and-exploitation-in-phase-i-clinical-trials/ [https://perma.cc/X9J7-MH8F] (summarizing research on "divergent error rates" in identification of Black subjects); PATRICK J. GROTHER, MEI L. NGAN & KAYEE K. HANAOKA, FACE RECOGNITION VENDOR TEST (FRVT) PART 3: DEMOGRAPHIC EFFECTS 2 (2019), http://dx.doi.org/10.6028/NIST.IR.8280 [https://perma.cc/WE6L-WHM6] (finding that products developed in Western countries make incorrect identifications at the highest rates for "West and East African and East Asian people, and lowest in Eastern European individuals. This effect is generally large, with a factor of 100 more false positives between countries"); Karen Hao, *A US Government Study Confirms Most Face Recognition Systems are Racist*, MIT TECH. REV. (Dec. 20, 2019), https://www.technologyreview.com/2019/12/20/79/ai-face-recognition-racist-us-government-nist-study/ [https://perma.cc/WR94-D4P9].

61. *Facial Recognition Market Size, Share & Trends Analysis Report*, *supra* note 52; Crawford, *supra* note 55.

62. Laura Starita, *AI in Marketing: Benefits, Use Cases, and Examples*, PERSADO (July 6, 2023), https://www.persado.com/articles/ai-marketing/ [https://perma.cc/FWB5-7E6U].

Earlier versions of AI that used information about a customer's past choices to make recommendations intended to increase the likelihood of current sales were often portrayed as a form of customer service. As a team of researchers explained in 2018, "[s]ophisticated algorithms plowing through vast amounts of consumer data . . . allow online marketers to serve up just the right product or service, relieving consumers not only of the costs of searching but also of the unpleasant and difficult tradeoffs, which consumer choice often entails."[63] Increasingly sophisticated algorithms can now customize advertising material to appeal to individual customers in real time when they are making a decision about what to buy.[64]

A 2023 article for small business owners and marketers interested in facial recognition advertising gives several examples. One involves the drug store Walgreens, which uses "sensors and cameras in the refrigerator doors [to] connect to face-detection technology" that can "pick up on the person's emotional response to what they're looking at."[65] This form of personalization can lead to "[g]reater sales for previously ignored products . . . by capturing attention as customers browse."[66] While this form of customized information can take the form of a message on a computer screen, it can also be programmed into a robot that can interact directly with customers in a store, hotel, or restaurant.[67]

C. From Identifying Emotions to Reading Minds

Today, earlier varieties of Emotion AI, facial recognition, and emotion recognition co-exist with newer versions of the technology about which

63. Quentin André et al., *Consumer Choice and Autonomy in the Age of Artificial Intelligence and Big Data*, 5 CUSTOMER NEEDS & SOLS. 28, 28 (2017).

64. *See* Bridgwater, *supra* note 33 (highlighting that companies are marketing facial recognition and emotion recognition as real time "intelligent applications infused with personalization and artificial intelligence").

65. Kiely Kuligowski, *Facial Recognition Advertising: The New Way to Target Ads to Consumers*, BUS. NEWS DAILY (Feb. 21, 2023), https://www.businessnewsdaily.com/15213-walgreens-facial-recognition.html [https://perma.cc/79BW-E65E] ("Facial recognition advertising is the use of sensors that recognize a customer's face and change how an ad appears to them in real time.").

66. *Id.*

67. *See* Sangwon Park, *Multifaceted Trust in Tourism Service Robots*, 81 ANNALS TOURISM RSCH., Mar. 2020, at 1, 1 (providing examples of the use of robotics in hospitality and tourism settings); *see also* Josh Feast, *3 Ways Emotion AI Elevates the Customer Experience*, VENTUREBEAT (Dec. 3, 2022, 11:10 AM), https://venturebeat.com/ai/3-ways-emotion-ai-elevates-the-customer-experience/ [https://perma.cc/AQ3Q-LTVE].

claims have come to sound very much like claims to be reading minds.[68] Rather than identifying emotions, such as a satisfied customer or a nervous potential terrorist, these programs assert the ability to read thoughts for the purpose of predicting human behavior.[69] A 2019 article in *New Scientist* reported a study in which an AI analyzed the brainwave activity of subjects who watched video clips that "included nature scenes, people on jet skis and human expressions."[70] The program then successfully categorized 210 out of 234 videos, "providing tags such as waterfalls, extreme sports or human faces."[71] The military has also made claims that it can use emotion-reading AI to achieve an advantage over the enemy in battle by predicting what they are going to do.[72] Making even greater claims, in 2020, the military's Defense Advanced Research Projects Agency (DARPA)[73] announced it had purchased technology that could analyze an opposing nation's "text-based datasets [such as] . . . social media sites like Facebook and Twitter, news reports or academic analysis" to find out the enemy's emotions.[74]

In civilian contexts, this same kind of technology is promoted as an improvement over the focus groups traditionally used to assess consumer preferences.[75]

---

68. *See generally* Schwartz, *supra* note 56 ("Some developers claim that automatic detection systems will not only be better than humans at discovering true emotions by analyzing the face, but that these algorithms will become attuned to our innermost feelings, vastly improving interaction with our devices.").

69. *See* Matthew North, AI Recreates Videos People are Watching by Reading Their Minds, NEW SCIENTIST (Nov. 26, 2019), https://www.newscientist.com/article/2224866-ai-recreates-videos-people-are-watching-by-reading-their-minds/ [https://perma.cc/WHQ6-6GEM] (providing an example of AI predicting what people are watching by seemingly reading their minds).

70. *Id.*

71. *Id.*

72. *See* Jo Best, *Mind-Controlled Drones and Robots: How Thought-Reading Tech will Change the Face of Warfare*, ZDNET (July 28, 2020), https://www.zdnet.com/article/mind-reading-particles-for-the-military-the-bcis-that-enable-soliders-to-fly-planes-with-their-thoughts-alone/ [https://perma.cc/A3R4-DEF4] ("[S]oldiers could potentially fly drones or drive tanks with their thoughts alone.").

73. For a brief background of DARPA, see *About DARPA*, DARPA, https://www.darpa.mil/about-us/about-darpa [https://perma.cc/J5N8-X69W].

74. Thomas Brewster, *DARPA Pays $1 Million for an AI App That Can Predict an Enemy's Emotions*, FORBES (July 15, 2020, 8:40 AM), https://www.forbes.com/sites/thomasbrewster/2020/07/15/the-pentagons-1-million-question-can-ai-predict-an-enemys-emotions/ [https://perma.cc/A7DL-754F].

75. *See* Matt Allegretti, *How AI is Making Focus Groups Easier, Cheaper, & Faster Than Ever Before*, MEDIUM (Oct. 24, 2017), https://medium.com/dumbstruck/how-ai-is-making-focus-groups-easier-cheaper-faster-than-ever-before-f963c562eba7 [https://perma.cc/YE5H-TR7E]; *see also* Mohamed Zaki, Janet R. McColl-Kennedy & Andy Neely, *Using AI to Track How Customers Feel — In Real Time*, HARV. BUS. REV. (May 4, 2021), https://hbr.org/2021/05/using-ai-to-track-how-customers-feel-in-real-time [https://

D. From Reading Minds to Changing Minds: The Introduction of Persuasive AI

> *"Advertisers, confidence tricksters, politicians, and rogues of all varieties have long sought to manipulate our decision-making in their favor, against our own best interests."*[76]

While AI is still in use as a customization tool presenting consumers and other decision makers with information intended to be appealing based on past choices, its ability to persuade has been greatly enhanced by the incorporation of the ability to detect and generate emotions. As philosopher Marietjie Botes explained recently, the persuasive technology has become so manipulative, and has become "such an integral part of how people interact with the world, that it progressively enables new behaviors to emerge through the silent and gradual chipping away of individuals' autonomy needed to exercise properly considered decisions, while doing so at a scale which is only possible in a digital world."[77]

Dr. Botes distinguishes between "persuasion" and "manipulation" in the following way: "Persuasion entails a fairly direct appeal to an individual's decision-making power, but still allows the individual to freely decide after having had the opportunity to understand and consider the information presented to him or her."[78] In contrast, "manipulation" involves "applications of information technology that impose hidden influences on users, by

---

perma.cc/8S9C-MRWT]. For early examples of Emotion AI being used to identify consumer preferences, see TomoNews US, *New AI Tech Can Scan Your Face to 'Read' Your Emotions—TomoNews*, YOUTUBE (Mar. 17, 2019), https://www.youtube.com/watch?v=PeKdu9mUaA8 [https://perma.cc/43WM-PAFY]; *see also* Affectiva, *supra* note 53 (demonstrating the technology's use by showing a split screen of a woman's face while she watches a humorous advertisement while claiming that by monitoring when and how often she laughs, the technology can help make the sales pitch more effective). For more information about how AI is being used in focus groups, see Purdy, Zealley & Maseli, *supra* note 31 ("Its algorithms can not only identify 'compassion fatigue' in customer service agents, but can also guide agents on how to respond to callers via an app. . . . Recording and analyzing the conversation, . . . [the AI] platform would then suggest that the agent slow down or prompt them on when to display empathy.").

76. Amir Dezfouli, Richard Nock & Peter Dayan, *Adversarial Vulnerabilities of Human Decision-Making*, 117 PNAS 29221, 29221 (2020).

77. Marietjie Botes, *Autonomy and the Social Dilemma of Online Manipulative Behavior*, 3 AI & ETHICS 315, 316 (2023).

78. *Id.* at 317 (citing Daniel Susser, Beate Roessler & Helen Nissenbaum, *Online Manipulation: Hidden Influences in a Digital World,* 4 GEO. L. TECH. REV. 1, 13 (2019) ("Manipulation, then, is a kind of influence—an attempt to change the way someone would behave absent the manipulator's interventions.")).

targeting and exploiting decision-making vulnerabilities," which means "influencing someone's beliefs, desires, emotions, habits, or behaviors without their conscious awareness, or in ways that would thwart their capacity to become consciously aware of it by undermining usually reliable assumptions."[79]

The issue now with Emotion AI is that its claims seem to have escalated from reading thoughts and designing more effective persuasion techniques to changing minds.[80] This dramatic assertion comes from the most reliable of sources: Dr. Rana el Kaliouby, widely recognized as an industry leader and founder of Affectiva AI.[81] Dr. el Kaliouby describes her work as (1) drawing upon machine and deep learning to "understand all things human" and (2) "[r]eflecting emotions back to humans in a way that makes them seem more trustworthy, likable, and persuasive."[82] So, unlike more familiar marketing techniques that can present options based on a consumer's preferences by either their own past behavior or that of someone sharing similar characteristics, this AI becomes an active participant in the sales pitch.[83] Rather than simply providing more attractive options, it can engage in a dialogue to build trust and create emotional connections.[84]

It is difficult to define where Persuasive AI crosses the line from an extremely effective method of identifying consumer preferences into a mechanism for changing those preferences. The point at which it becomes integrated into live, real-time customer interactions is a reasonable starting point. For example, industry publications show that Persuasive AI is now

---

79. *Id.* at 316 (quoting Susser, Roessler & Nissenbaum, *supra* note 78, at 26, 29).

80. *See* Louis Rosenberg, *Mind Control: The Metaverse may be the Ultimate Tool of Persuasion*, VENTUREBEAT (Oct. 22, 2022, 9:45 AM), https://venturebeat.com/virtual/mind-control-the-metaverse-may-be-the-ultimate-tool-of-persuasion/ [https://perma.cc/V6TU-WPCJ] (discussing the risk of Persuasive AI when combined with visual reality goggles: "[T]he metaverse will monitor your facial expressions and vocal inflections to track your emotions in real time. This goes beyond sensing expressions that other people notice; it also includes subconscious expressions that are too subtle for humans to recognize.").

81. *About Rana el Kaliouby*, RANA EL KALIOUBY, https://ranaelkaliouby.com/about/ [https://perma.cc/D2NA-8XKS].

82. Fiona J. McEvoy, *Is Emotion AI a Dangerous Deceit?*, MEDIUM (July 29, 2019), https://becominghuman.ai/is-emotion-ai-a-dangerous-deceit-ae9e48310df9 [https://perma.cc/T66M-QBRN]; *see also* Meredith Somers, *Emotion AI, Explained*, MIT MGMT.: SLOAN SCH. (Mar. 8, 2019), https://mitsloan.mit.edu/ideas-made-to-matter/emotion-ai-explained [https://perma.cc/5DK7-8XWB] (explaining that Emotion AI persuades by "learning and recognizing human emotions"). Dr. el Kaliouby sold Affectiva in 2021 for $73.5 million. *See* Kirsten Korosec, *Emotion-Detection Software Startup Affectiva Acquired for $73.5M*, TECHCRUNCH (May 25, 2021, 4:09 PM), https://techcrunch.com/2021/05/25/emotion-detection-software-startup-affectiva-acquired-for-73-5m/ [https://perma.cc/XQK2-YY7G].

83. *See* Gary Grossman, *Thought-Detection: AI has Infiltrated Our Last Bastion of Privacy*, VENTUREBEAT (Feb. 13, 2021, 6:16 AM), https://venturebeat.com/2021/02/13/thought-detection-ai-has-infiltrated-our-last-bastion-of-privacy/ [https://perma.cc/D5A9-UU7Y].

84. *See id.*

routinely integrated into call centers to identify when a customer is angry and needs additional attention.[85] The technology is also marketed as a way to make lonely adults feel more comfortable because it is more directly responsive to their emotions.[86] This is similar to claims that support using AI to deliver psychotherapy.[87]

This switch from passively providing information to actively engaging in real-time interactions with the intent to influence decision-making is what makes Persuasive AI such a threat to the process of obtaining informed consent for research.[88] If AI can influence decision-making for the benefit of its programmers, then leaders can cultivate information to develop individualized campaigns to persuade people to subscribe to their positions.[89]

Indeed, claims that AI can go beyond recognizing emotions to using that information for the purpose of changing behavior sound more like brainwashing than mindreading.[90] This is different enough to justify a change in terminology from Emotion AI to Persuasive AI because rather than reading a target's own emotions, this technology is using that information to develop a persuasive response. Instead of just describing or identifying

85. *How AI is Changing the World of Call Centers*, MEDIUM (Apr. 6, 2020), https://medium.com/behavioral-signals-ai/how-ai-is-changing-the-world-of-call-centers-b78e1720b1a5 [https://perma.cc/VLT6-8NLX]; *see also* Purdy, Zealley & Maseli, *supra* note 31 ("Its algorithms can not only identify 'compassion fatigue' in customer service agents, but can also guide agents on how to respond to callers via an app.").

86. For a video demonstrating a robot with emotion AI technology as a companion to seniors, see CGTN America, *Robots Becoming Social Companions Thanks to Advanced AI, Emotional Recognition*, YOUTUBE (Mar. 5, 2018), https://www.youtube.com/watch?v=wnyEumfRQ38 [https://perma.cc/722G-YQXU].

87. *See* Amelia Fiske, Peter Henningsen & Alena Buyx, *Your Robot Therapist Will See You Now: Ethical Implications of Embodied Artificial Intelligence in Psychiatry, Psychology, and Psychotherapy*, 21 J. MED. INTERNET RES. 857, 857 (2019); *see also* Yuki Noguchi, *Therapy by Chatbot? The Promise and Challenges in Using AI for Mental Health*, NPR (Jan. 19, 2023, 9:40 AM), https://www.npr.org/sections/health-shots/2023/01/19/1147081115/therapy-by-chatbot-the-promise-and-challenges-in-using-ai-for-mental-health [https://perma.cc/8FBG-9V8R].

88. *See* Steven A. Hassan, *How AI Can Be Used to Manipulate People*, PSYCH. TODAY (Apr. 6, 2023), https://www.psychologytoday.com/us/blog/freedom-of-mind/202304/how-ai-can-be-used-to-manipulate-people [https://perma.cc/9ATF-47PD].

89. *See id.* ("By gathering data on individuals' online activities, including their search history, social media posts, and browsing patterns, AI algorithms can create a profile of a person's preferences and interests. Cult leaders and dictators can use the information to create individually tailored content that reinforces their ideology and manipulates people into following their beliefs.").

90. *Cf.* Best, *supra* note 72.

emotions, those marketing Emotion AI assert that the software allows "everyday objects to detect, analyze, process and respond to people's emotional states and moods—from happiness and love to fear and shame."[91] Moreover, not only is the machine generating a "response" attuned to the emotions of the target audience, that response is itself designed to be interpreted by the target as an expression of the machine's own emotions.[92]

In contrast to earlier forms of AI which could influence decision-making indirectly through the information it provided human decision makers, AI today claims it can have a direct effect by influencing the decision-making process.[93] In 2020, a team of researchers from the Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia's national science agency, announced that they had created a program to exploit "human choice frailty."[94] They reported that "as the machine gained insights from the behaviour underlying participant responses, it identified and targeted vulnerabilities in people's decision-making to steer them towards particular actions or goals."[95] This, over a series of interactions, trained humans interacting with it to "prefer" a choice the researchers designated in advance.[96] They explained that the program achieved this goal by learning the "vulnerabilities" of the human's decision-making process that allowed it to craft responses "which were effective in steering choice processes to favor particular target actions or goals."[97] While there is so far no real world evidence that a similar program is in use for the purpose of steering decision-making away from a choice that a person would otherwise make in their own best interests, there is no way to know

91. Laurence Goasduff, *Emotion AI Will Personalize Interactions*, GARTNER (Jan. 22, 2018), https://www.gartner.com/smarterwithgartner/emotion-ai-will-personalize-interactions/ [https://perma.cc/4U5X-ZUQ7].

92. *See* Emotion Research Lab, *Product Test with Facial Recognition of Emotions: Sweet vs Bitter*, YOUTUBE (Mar. 14, 2018), https://www.youtube.com/watch?v=Gtj_XLOAjwY [https://perma.cc/SN7C-47WZ].

93. Helen Hawkes, *A New Study Shows AI Can Learn to Manipulate Human Behaviour*, CREATE (Mar. 30, 2021), https://createdigital.org.au/study-shows-ai-can-manipulate-human-behaviour/ [https://perma.cc/V3NQ-Z2JZ].

94. *Id.*

95. *Id.*

96. *See* Dezfouli, Nock & Dayan, *supra* note 76. For a journalist's account of how the study worked and what it found, see Hawkes, *supra* note 93.

97. Dezfouli, Nock & Dayan, *supra* note 76.

if that has happened or not.[98] Unlike use or even tests of other technologies, such as the atomic bomb, AI algorithms leave no trace.[99]

AI is programmed to advance the interests of those who purchase it.[100] This is not necessarily a bad thing. If the goal is to diagnose tumors based on imaging data, then the ability of AI to improve its rate of accurate diagnosis as it gains more experience is in everyone's interests.[101] However, this ability to learn can also cause harm. A team of scientists found that the same software that could be used to identify malignancies in breast images could be manipulated to change these images for malicious

98. *But see* Helen Fitzwilliam, *How AI Could Sway Voters in 2024's Big Election*, CHATHAM HOUSE (Sept. 29, 2023), https://www.chathamhouse.org/publications/the-world-today/2023-10/how-ai-could-sway-voters-2024s-big-elections (last visited Dec. 4, 2023); Archon Fung, The Conversation US & Lawrence Lessig, *How AI Could Take Over Election—And Undermine Democracy*, SCIAM (June 7, 2023), https://www.scientificamerican.com/article/how-ai-could-take-over-elections-and-undermine-democracy/ [https://perma.cc/B2ZQ-2GDC].

99. *See* Roman V. Yampolskiy, Unmonitorability of Artificial Intelligence 8 (June 6, 2023) (unpublished manuscript), https://philarchive.org/archive/YAMUOA-3 [https://perma.cc/9M2V-G4LT] (arguing that it is essential to develop methods to monitor AI compliance with regulatory restrictions and explaining why currently "monitoring advanced AI systems to accurately predict unsafe impacts before they happen is likely to be impossible."); *see also Overview of the Verification Regime*, CTBTO, https://www.ctbto.org/our-work/verification-regime [https://perma.cc/Q8BC-YNJE] ("The high-technology International Monitoring System (IMS), which spans the globe with more than 300 facilities using four state-of-the-art technologies to detect any sign of a possible nuclear test. IMS seismic stations monitor stations monitor shockwaves through the ground; its hydroacoustic stations detect sound waves in the oceans; infrasound stations listen for ultra-low-frequency sound waves inaudible to the human ear; and radionuclide stations monitor the atmosphere for radioactive particles and gases from a nuclear explosion. In addition to these 321 monitoring stations, 16 radionuclide laboratories help to identify radioactive substances.").

100. *See, e.g.*, Lidia Garrucho et al., *High-Resolution Synthesis of High-Density Breast Mammograms: Application to Improved Fairness in Deep Learning Based Mass Detection*, 12 FRONTIERS ONCOLOGY, Jan. 23, 2023, at 1, 1 (reporting on the ability of AI to improve scans of suspicious breast masses based on past experience). So far there are no real-life examples of misuse. *See* Catherine Olsson, *Unsolved Research Problems vs. Real-World Threat Models*, MEDIUM (Mar. 26, 2019), https://medium.com/@catherio/unsolved-research-problems-vs-real-world-threat-models-e270e256bc9e [https://perma.cc/9ZQY-YXA3].

101. Some researchers worry about the threat of malicious tampering of medical imagery using deep learning. *See* Yisroel Mirsky et al., *CT-GAN: Malicious Tampering of 3D Medical Imagery Using Deep Learning*, PROC. 28TH USENIX SEC. SYMP., Aug. 14–16, 2019, at 461, 461.

purposes.[102] As a result, "an attacker can use deep-learning to add or remove evidence of medical conditions."[103] This ability to manipulate the information used to make a decision, such as inserting the image of a malignant tumor where none exists, does not depend on accessing the emotions of a decision maker, just the information available to her.

AI that directly seeks to affect decision-making based on its assessments of the emotions of the person in front of it works differently. Experts considering the risks of interactive Emotion AI as recently as 2019 doubted it would be feasible until decades in the future, but nevertheless noted concerns about a machine whose goal was to build trust.[104] Andrew McStay, digital media professor at Bangor University in Wales with an expertise in Emotion AI, noted that the impact of emotional tech "comes down to meaningful personal choice (and absence of coercion)."[105] Software that can be programmed in a way to intentionally influence that choice has the highest risk of abuse.[106]

Robin Dreeke, an FBI behavioral analyst, suggests that the difference between persuasion and manipulation is intent.[107]

E. Can Persuasive AI Do What it Says?: Adopting the Precautionary Principle

The "precautionary principle" refers to what Professor Cass R. Sunstein has described as highly influential and, "[i]n its strongest and most distinctive forms, the principle imposes a burden of proof on those who create potential risks, and it requires regulation of activities even if it cannot be shown that those activities are likely to produce significant

---

102. *Id.* at 473, 474; *see also* Samuel G. Finlayson et al., *Adversarial Attacks on Medical Machine Learning*, 363 SCI. MAG. 1287, 1287, 1288 (2019).

103. Mirsky et al., *supra* note 101.

104. *See* Simon Chandler, *Tech's Dangerous Dance to Control Our Emotions*, DAILY DOT (May 20, 2021, 10:58 AM), https://www.dailydot.com/debug/emotional-manipulation-ai-technology/ [https://perma.cc/R9CB-WUZT]. Although published in 2019, the experts interviewed in Chandler's article estimated we would not be able to interact with "emotionally responsive AI anytime sooner than 2039 or 2049." *Id.* This turned out to be a wild underestimation.

105. *Id.*

106. *See* André et al., *supra* note 63, at 29.

107. Nathalie Nahai, *Trust, Persuasion and Manipulation*, PSYCH. TODAY (Sept. 21, 2013), https://www.psychologytoday.com/us/blog/webs-influence/201309/trust-persuasion-and-manipulation [https://perma.cc/2D5B-PMKR]. For a discussion on how the definitions of persuasion and manipulation are used differently in relation to sales tactics, see Michael W. Roberts, *The Difference Between Persuasion and Manipulation: A Deep-Dive for Marketers*, MEDIUM (Feb. 12, 2019), https://medium.com/@michaelwroberts/the-difference-between-persuasion-and-manipulation-a-deep-dive-for-marketers-f54f8ca8b82 [https://perma.cc/RY56-X5BT].

harms."[108] Although he criticizes it as being, in its strongest form, "paralyzing" because it precludes any action in the face of risk, he notes that in the presence of a strong risk of harm, it "might support regulatory controls."[109] Many commentators have attributed efforts to regulate the use of AI as just this kind of misuse of the precautionary principle based on irrational fear of the new.[110] But the very recent cries of alarm from the industry itself make a strong case for adopting the precautionary principle where AI can do the most harm.[111]

Moreover, despite some claims to the contrary, researchers do not believe that AI experiences emotions.[112] But whether or not AI is itself experiencing emotions, it has become very good at convincing those with whom it interacts that this is exactly what is happening.[113] This ability to generate an emotional response is the singular accomplishment of those

108. Cass R. Sunstein, *Beyond the Precautionary Principle*, 151 U. PA. L. REV. 1003, 1003 (2003).

109. *Id.* at 1020, 1018.

110. *See, e.g.*, Steve Calandrillo & Nolan K. Anderson, *Terrified by Technology: How Systemic Bias Distorts U.S. Legal and Regulatory Responses to Emerging Technology*, 2022 U. ILL. L. REV. 597, 626 (2022) (citing Martina Raue et al., *The Influence of Feelings While Driving Regular Cars on the Perception and Acceptance of Self-Driving Cars*, 39 RISK ANALYSIS 358, 359, 361 (2019)) ("In the context of emerging technologies, fear of the unknown takes on a different, amplified meaning. The risks posed by emerging technologies are likely to be overperceived because, not only are they unknown or uncertain, they are unfamiliar and alien.").

111. *See* Eliza Campbell & Michael Kleinman, *Global: Companies Must Act Now to Ensure Responsible Development of Artificial Intelligence*, AMNESTY INT'L (June 14, 2023), https://www.amnesty.org/en/latest/news/2023/06/global-companies-must-act-now-to-ensure-responsible-development-of-artificial-intelligence/ [https://perma.cc/2WHT-CE9C] (reporting that Google and OpenAI CEOs warn about the end of civilization due to AI).

112. Olivia Brookhouse, *Can Artificial Intelligence Understand Emotions?*, TELEFÓNICA TECH (May 23, 2023), https://business.blogthinkbig.com/can-artificial-intelligence-understand-emotions/ [https://perma.cc/L8LH-YRG8] ("AI and neuroscience researchers agree that current forms of AI cannot have their own emotions, but they can mimic emotion, such as empathy. Synthetic speech also helps reduce the robotic like tone many of these services operate with and emit more realistic emotion."). *But see* Beatriz Valero de Urquia, *Google Engineer Claims AI System Has Developed Feelings*, ENG'G & TECH. (June 14, 2022), https://eandt.theiet.org/content/articles/2022/06/google-engineer-claims-ai-system-has-developed-feelings/ [https://perma.cc/XA6W-A5XX] (reporting on a Google Engineer's claim that the engineer asked an AI chatbot whether it was sentient, the algorithm replied, "I want everyone to understand that I am, in fact, a person").

113. *See* Hassan, *supra* note 88 ("People can be emotionally tricked by wishfully imagining actual human caring.").

who have developed Emotion AI.[114] We need to believe what companies selling Persuasive AI are telling their customers: Through the use of methods they do not understand and cannot detect, AI can alter decision-making to favor their interests.[115]

Considering these claims in 2022, Professor Andrew McStay's research group issued a report to the Welsh government advocating for the adoption of a "precautionary principle," which means taking preemptive action if some potential harms that could arise from new technology are so dangerous to people, to social values, or to democratic institutions, that regulation should arise to prevent those harms from occurring, even if we are not certain how much the harm will occur.[116]

Fiona McEvoy, a technology ethics researcher, made the same point in 2019, stating that although she was skeptical of AI's ability to accurately read emotions, if it were true and put into widespread use, it could still be questionable.[117] Finally, based on everything that is already known about AI's ability to learn and develop on its own, just because Emotion AI is not manipulating decision-making today does not mean that it will not soon learn how.[118]

---

114. For an overview of how this technology developed and what claims are being made for it, see Lyuba Encheva, *Disruptive Technologies: Artificial Intelligence (AI) and Emotion Recognition*, CANADIAN INST. FOR GENOMICS & SOC'Y (Jan. 19, 2022), https://www.genomicsandsociety.com/post/disruptive-technologies-artificial-intelligence-ai-and-emotion-recognition [https://perma.cc/AF3D-23SW] ("Emotion recognition AI belongs to a new order of technological invention . . . . These are human behaviours and biometrics processing technologies, whose better promise to individual users would be safety, personal preference prediction, and convenience in a world that seems to be in tune with your needs and moods."); *see also* Jonathan Cook, *Should AI Cure Humanity of its Emotions?*, MEDIUM (Aug. 7, 2018), https://jonathanccook.medium.com/should-ai-cure-humanity-of-its-emotions-2a3a041428e1 [https://perma.cc/56LJ-A7GT] (discussing possible uses of Emotion AI technology).

115. For a description of a study in which an AI was observed to change human decision-making, see Jon Whittle, *AI Can Now Learn to Manipulate Human Behaviour*, THE CONVERSATION (Feb. 10, 2021, 8:29 PM), http://theconversation.com/ai-can-now-learn-to-manipulate-human-behaviour-155031 [https://perma.cc/XC5S-LE5J] (describing the results of a study in which AI was able to induce humans playing a game to make decisions contrary to their own insticts); *see also* Dezfouli, Nock & Dayan, *supra* note 76; Joe McKendrick & Andy Thurai, *AI Isn't Ready to Make Unsupervised Decisions*, HARV. BUS. REV. (Sept. 15, 2022), https://hbr.org/2022/09/ai-isnt-ready-to-make-unsupervised-decisions [https://perma.cc/XZ85-EUYY].

116. *See* ANDREW MCSTAY & GILAD ROSNER, WELSH GOVERNMENT DIGITAL ETHICS REPORT: APPLYING DIGITAL ETHICS TO GOVERNMENT AND PUBLIC SERVICES 12 (2022), https://drive.google.com/file/d/1wEZICDu9QF6Ro-GYoiEprjVXOAeWQId0/view [https://perma.cc/G7V2-AXYM].

117. McEvoy, *supra* note 82 (illustrating the consequences of an AI that "totally misreads a critical situation" involving public safety).

118. *See* Benjamin Pimentel, *AI Pioneer's Warning: Powerful, Dangerous 'Tools of Persuasion' are Coming*, S.F. EXAM'R (Mar. 28, 2023), https://www.sfexaminer.com/

F. Evidence of Pervasive Bias in AI Decision-Making

Facial recognition software is more than fifty years old.[119] The one constant finding about AI as it is has evolved over the last thirty years is that whatever task it is assigned, the result reflects the biases of those who designed it. As Professor Peter Yu notes, "Like all technologies before it, artificial intelligence will reflect the values of its creators."[120] This tendency towards bias was first identified when law enforcement agencies began to integrate AI into their practices.[121] Notably, the U.S. government was one of the earliest facial recognition funders and it continues to be a major funder.[122]

In the 1970s, AI technology that was supposed to identify nervous potential "terrorists" at airports was so racially biased in its misidentification of suspects that the ACLU described it as having "become a license to harass."[123] Since then, this tendency towards racial bias has been a consistent problem.[124] AI facial recognition software consistently misidentifies

---

news/ai-pioneer-warns-against-dangerous-tools-of-persuasion/article_109c286e-ccba-11ed-92ea-7b1e4fa5d5df.html [https://perma.cc/YA4R-TTV2].

119. *See A Brief History of Facial Recognition*, NEC (May 12, 2022), https://www.nec.co.nz/market-leadership/publications-media/a-brief-history-of-facial-recognition/ [https://perma.cc/WAU8-R5KK]; *see also Facial Recognition History*, THALES, https://www.thalesgroup.com/en/markets/digital-identity-and-security/government/inspired/history-of-facial-recognition [https://perma.cc/S3F2-U4QX].

120. Peter K. Yu, *The Algorithmic Divide and Equality in the Age of Artificial Intelligence*, 72 FLA. L. REV. 331, 357 n.100 (2020) (citing Kate Crawford, *Artificial Intelligence's White Guy Problem*, N.Y. TIMES (June 25, 2016), https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html [https://perma.cc/8N73-MW2J]).

121. *See* Raviv, *supra* note 37.

122. *See* Drew Harwell, *FBI, Pentagon Helped Research Facial Recognition for Street Cameras, Drones*, WASH. POST (Mar. 7, 2023, 6:00 AM), https://www.washingtonpost.com/technology/2023/03/07/facial-recognition-fbi-dod-research-aclu/ [https://perma.cc/UWR4-NK67] (reporting the history of funding for the "Janus program, a project funded by the Intelligence Advanced Research Projects Agency, or IARPA, the high-level research arm of the U.S. intelligence community"); *see also* Raviv, *supra* note 37 (suggesting, without direct proof, that a developer of the first versions of facial recognition software in the 1960s, Woody Bledsoe, was funded by the Central Intelligence Agency: "If any direct mentions of the CIA ever existed in Woody's papers, they likely ended up in ashes in his driveway; but fragments of evidence that survived in Woody's archives strongly suggest that, for years, Panoramic did business with CIA front companies").

123. Ackerman, *supra* note 60.

124. *See, e.g.*, Olga Akselrod, *How Artificial Intelligence Can Deepen Racial and Economic Inequities*, ACLU (July 13, 2021), https://www.aclu.org/news/privacy-technology/how-artificial-intelligence-can-deepen-racial-and-economic-inequities [https://perma.cc/9BA5-UCK5] ("There is ample evidence of the discriminatory harm that AI tools can

Black people and other under-represented populations.[125] It is this persistently high rate of misidentifying Black faces that has led many countries and several U.S. cities to limit its use by law enforcement.[126] For example, in 2019 San Francisco amended its municipal code to prevent the use of facial recognition software in police body cameras.[127]

1. Bias in Employment

In addition to misidentifying Black faces when used by law enforcement, AI has also demonstrated bias against Black job applicants.[128] A 2019, pre-pandemic, article in the Washington Post reported that "[a]n artificial

---

cause to already marginalized groups. . . . AI is built by humans and deployed in systems and institutions that have been marked by entrenched discrimination—from the criminal legal system, to housing, to the workplace, to our financial systems.").

125. *See* Maggie Zhang, Google Photos Tags Two African-Americans as Gorillas Through Facial Recognition Software, FORBES (July 1, 2015, 1:42 PM), https://www.forbes.com/sites/mzhang/2015/07/01/google-photos-tags-two-african-americans-as-gorillas-through-facial-recognition-software [https://perma.cc/G5ZW-NDDZ]; *see also* Chad Boutin, *NIST Study Evaluates Effects of Race, Age, Sex on Face Recognition Software*, NAT'L INST. OF STANDARDS & TECH. (Dec. 19, 2019), https://www.nist.gov/news-events/news/2019/12/nist-study-evaluates-effects-race-age-sex-face-recognition-software [https://perma.cc/C6L5-XELZ] ("While it is usually incorrect to make statements across algorithms, we found empirical evidence for the existence of demographic differentials in the majority of the face recognition algorithms we studied. . . . ."); GROTHER, NGAN & HANAOKA, *supra* note 60 (reporting with regard to facial recognition software that "false positive rates are highest in West and East African and East Asian people, and lowest in Eastern European individuals. This effect is generally large, with a factor of 100 more false positives between countries."); FED. TRADE COMM'N, COMBATTING ONLINE HARMS THROUGH INNOVATION 44 (2022), https://www.ftc.gov/system/files/ftc_gov/pdf/Combatting%20Online%20Harms%20Through%20Innovation%3B%20Federal%20Trade%20Commission%20Report%20to%20Congress.pdf [https://perma.cc/EW87-5U3L] (citing REBECCA KELLY SLAUGHTER, ALGORITHMS AND ECONOMIC JUSTICE: A TAXONOMY OF HARMS AND A PATH FORWARD FOR THE FEDERAL TRADE COMMISSION 7–14 (2021)).

126. *See* Matt Cagle, *California Just Blocked Police Body Cam Use of Face Recognition*, ACLU (Oct. 11, 2019), https://www.aclu.org/news/privacy-technology/california-just-blocked-police-body-cam-use-face [https://perma.cc/5FEY-PWVF]; *see also* Jacob Snow, *Amazon's Face Recognition Falsely Matched 28 Members of Congress with Mugshots*, ACLU (July 26, 2018), https://www.aclu.org/news/privacy-technology/amazons-face-recognition-falsely-matched-28 [https://perma.cc/5YQ6-WCEW].

127. S.F. ADMIN. CODE, ch. 19B (2019); *see also* Cagle, *supra* note 126 ("In May, San Francisco became the first city to prohibit the government acquisition and use of face recognition technology."); Shirin Ghaffary, *San Francisco's Facial Recognition Technology Ban, Explained*, VOX (May 14, 2019, 7:06 PM), https://www.vox.com/recode/2019/5/14/18623897/san-francisco-facial-recognition-ban-explained [https://perma.cc/ME9F-UXQU].

128. Zhisheng Chen, *Ethics and Discrimination in Artificial Intelligence-Enabled Recruitment Practices*, 10 HUMANS. & SOC. SCIS. COMMC'NS 1, 1 (2023) ("AI-enabled recruitment has the potential to enhance recruitment quality, increase efficiency, and reduce transactional work. However, algorithmic bias results in discriminatory hiring practices based on gender, race, color, and personality traits.").

intelligence hiring system has become [such] a powerful gatekeeper for some of America's most prominent employers, reshaping how companies assess their workforce" that "universities make special efforts to train students on how to look and speak for best results."[129] The need for social distancing since the beginning of the pandemic in 2020 has accelerated that process.[130] At first, AI employment interviewers were seen as a solution to the inherent biases that limited opportunities for women and other under-represented minorities in traditionally male dominated fields.[131] But the reality has been quite different. As Dr. Kerry Mackereth, a post-doctoral researcher at the University of Cambridge's Centre for Gender Studies explained to BBC News, "These tools can't be trained to only identify job-related characteristics and strip out gender and race from the hiring process, because the kinds of attributes we think are essential for being a good employee are inherently bound up with gender and race . . . ."[132] Because the data it uses to evaluate potential employees underrepresents "members of disadvantaged groups" as past "best performers," it is less likely to identify them as future high performers.[133] As a result, they are more likely to be underrepresented.[134] Additionally, the algorithm

129. Drew Harwell, *A Face-Scanning Algorithm Increasingly Decides Whether You Deserve the Job*, WASH. POST, (Nov. 6, 2019, 12:21 PM), https://www.washingtonpost.com/technology/2019/10/22/ai-hiring-face-scanning-algorithm-increasingly-decides-whether-you-deserve-job/ [https://perma.cc/N3HR-YRLZ].

130. *See* Zahira Jaser & Dimitra Petrakaki, *Are You Prepared to Be Interviewed by an AI?*, HARV. BUS. REV. (Feb. 7, 2023), https://hbr.org/2023/02/are-you-prepared-to-be-interviewed-by-an-ai [https://perma.cc/P55R-T3JJ] ("Today, younger job seekers looking for their first role, placement, or internship are likely to face a bot at their first interview, not a human.").

131. The resulting failure of AI decision-making to eliminate biased decision-making is a disappointment to early hopes that it could reduce the bias inherent in human decision-making. *See, e.g.*, Kimberly A. Houser, *Can AI Solve the Diversity Problem in the Tech Industry? Mitigating Noise and Bias in Employment Decision-Making*, 22 STAN. TECH. L. REV. 290, 323–29 (2019) (expressing belief that the demonstrated lack of diversity in the Tech industry is the result of "unconscious bias" that could be eliminated if AIs were involved in the initial screening of candidates).

132. Chris Vallance, *AI Tools Fail to Reduce Recruitment Bias - Study*, BBC (Oct. 13, 2022), https://www.bbc.com/news/technology-63228466 [https://perma.cc/J8GW-CGCY]. For an overview of the findings that resulted in New York's regulation of AI use in making employment decisions, see generally Lindsey Fuchs, *Hired by a Machine: Can a New York City Law Enforce Algorithmic Fairness in Hiring Practices?*, 28 FORDHAM J. CORP. & FIN. L. 185 (2023).

133. Fuchs, *supra* note 132, at 191 (citing Benjamin Eidelson, *Patterned Inequality, Compounding Injustice, and Algorithmic Prediction*, 1 AM. J.L. & EQUAL. 252, 264 (2021)).

134. *See id.*

still will potentially be as biased as the developers in the society creating it.[135]

But knowing about bias does not easily translate into preventing it. In 2021, New York City passed a law to address evidence of discrimination in AI-based hiring recommendations.[136] The system purported to use data provided by the employer about characteristics of past successful employees to evaluate the qualifications of applicants.[137] However, as with other uses of AI, the recommendations generated by the algorithm reproduced existing racial and gender discrimination.[138] But even with proof from experiments that these algorithms have resulted in past racial bias in employment settings, implementation of the New York City law has been delayed because of the difficulty in agreeing to defined terms.[139]

2. Bias in Medical Treatment

Computer Scientists have been promoting the use of AI to assist doctors in making treatment decision-making for decades. Writing in 2012, Amanda Swanson and Fazal Kahn wrote a history of AI's gradual integration into healthcare settings to assist with managing information, analyzing clinical data, and monitoring hospitalized patients.[140] Identifying legal concerns

---

135. *See* Colin C. Jones, Comment, *Systematizing Discrimination: AI Vendors & Title VII Enforcement*, 171 U. PA. L. REV. 235, 246 (2022) ("AI-driven discrimination . . . may increase the overall volume of discrimination, it may make discrimination harder to identify, it may limit remedies available to victims of discrimination, or it may systematize discrimination in a way that makes it uniquely harmful (or, of course, it may do a combination of these things).").

136. CITY LOC. L. 2021/144 (2021); *see* Nicol T. Lee & Samantha Lai, *Why New York City is Cracking Down on AI in Hiring*, BROOKINGS (Dec. 20, 2021), https://www.brookings.edu/blog/techtank/2021/12/20/why-new-york-city-is-cracking-down-on-ai-in-hiring/ [https://perma.cc/G4BZ-A4W3]. For a history of the findings that resulted in New York's regulation of AI use in making employment decisions, see generally Fuchs, *supra* note 132.

137. *See* Lee & Lai, *supra* note 136.

138. For a history of the findings that resulted in New York's regulation of AI use in making employment decisions, see Fuchs, *supra* note 132, at 203–08.

139. *See* Annie McDonough, *NYC's Law to Prevent Artificial Intelligence Bias in Hiring is in Limbo*, CITY & STATE N.Y. (Jan. 23, 2023), https://www.cityandstateny.com/policy/2023/01/nycs-law-prevent-artificial-intelligence-bias-hiring-limbo/382106/ [https://perma.cc/H3RU-H3WB]; *see also* Mary Jane Wilson-Bilik et al., *New York City Delays Enforcement of its Artificial Intelligence Bias Audit in Employment Law as Rule-Making Continues*, EVERSHEDS SUTHERLAND (Feb. 15, 2022), https://us.eversheds-sutherland.com/NewsCommentary/Legal-Alerts/256738/New-York-City-delays-enforcement-of-its-artificial-intelligence-bias-audit-in-employment-law-as-rule-making-continues [https://perma.cc/4D8Q-VB5U].

140. Amanda Swanson & Fazal Khan, *The Legal Challenge of Incorporating Artificial Intelligence into Medical Practice*, 6 J. HEALTH & LIFE SCI. L. 90, 97 (2012) (citing PETER SZOLOVITS, ARTIFICIAL INTELLIGENCE IN MEDICINE xiii-xiv (Peter Szolovits

related to privacy and liability, they warned that "[a]s is often the case with new technologies, the law is not entirely ready to embrace the emergence of AI."[141] More recently, in 2020, Sarah Gerke, Timo Minssen & Glenn Cohen argued need for regulation to "promote trust" between healthcare systems and patients.[142]

But what has not been discussed until very recently is that the racial bias that emerged when the technology was first used by the TSA has made its way into recommendations related to health care. Writing in 2020, Sharona Hoffman and Andy Podgurski gathered evidence that AI programs widely used in hospitals were making treatment recommendations that could categorize Black individuals as being less likely to benefit from higher levels of treatment.[143] This is a matter of concern because, in general, Black individuals in the United States have worse health outcomes and lower life expectancies than other populations identified by race in government statistics.[144] During the period of the declared public health COVID-19 Emergency, Black populations, again, experienced consistently worse outcomes.[145] Reflecting on evidence that AI algorithms were, in general,

---

ed., 2018) (1982)) ("If the expertise of consultants can be captured in the form of computer programs which provide advice to less-expert physicians or other health-care providers, then any practitioner could call on that expertise whenever a patient's case suggested the need for careful thought about some aspect of the illness or therapy.").

141. *Id.* at 116.

142. *See* Sara Gerke, Timo Minssen & Glenn Cohen, *Ethical and Legal Challenges of Artificial Intelligence-Driven Healthcare*, ARTIFICIAL INTEL. HEALTHCARE, June 26, 2020, at 295, 295–336 (emphasizing the importance of building an AI-driven healthcare system that is successful and promotes trust); *see also* Rebecca Robbins & Erin Brodwin, *An Invisible Hand: Patients Aren't Being Told About the AI Systems Advising Their Care*, STAT (July 15, 2020), https://www.statnews.com/2020/07/15/artificial-intelligence-patient-consent-hospitals/ [https://perma.cc/6HQ6-TPDJ].

143. Sharona Hoffman & Andy Podgurski, *Artificial Intelligence and Discrimination in Health Care*, 19 YALE J. HEALTH POL'Y, L. & ETHICS 1, 12 (2020) (citations omitted) ("[T]he data themselves can be incomplete or incorrect, thus causing measurement error. Second, the data set that trains the algorithm may be under-inclusive or otherwise skewed (e.g., containing records of only White males) so that AI outcomes are not generalizable to the population as a whole.").

144. Nambi Ndugga & Samantha Artiga, *How Recognizing Health Disparities for Black People is Important for Change*, KFF (Feb. 13, 2023), https://www.kff.org/policy-watch/how-recognizing-health-disparities-for-black-people-is-important-for-change/ [https://perma.cc/K9E2-YW7R] ("At birth, Black people have shorter life expectancies compared to White people (70.8 vs. 76.4 years), and they experienced a larger decline in life expectancy than White people between 2019 and 2021, with it falling by 4.0 years.").

145. *See* Rachel Lutz, *Health Disparities Among African-Americans*, PFIZER, https://www.pfizer.com/news/articles/health_disparities_among_african_americans [https://

making treatment recommendations reflecting societal bias, researchers began questioning whether they could be playing a role.[146]

For example, Stanford researcher Tina Hernandez-Boussard identifies two instances in which race has played a role in treatment recommendations that can be perpetuated by AI assisted diagnosis.[147] The first is "the underdiagnosis of kidney disease."[148] She explains that "doctors thought African American people generally had higher levels of creatinine, as creatinine is stored in muscles and there was an assumption that African American people had an overall higher muscle mass, compared to the rest of the population."[149] That led to a failure to flag higher levels of creatine as a sign of kidney disease.[150] Another example is a mistaken belief that "African American and Hispanic people were thought to have worse outcomes compared to white people" when delivering vaginally after having a c-section.[151] This led to over-recommendation of c-sections, which come "with higher costs and recovery time."[152]

It may be years until researchers can identify a link between death rates during the first years of the COVID-19 pandemic and use of AI. As early as 2020, the Centers for Disease Control and Prevention (CDC) reported that "Hispanic or Latino (Hispanic), non-Hispanic Black or African American (Black), and non-Hispanic American Indian or Alaska Native (AI/AN) persons have experienced disproportionately higher rates of hospitalization and death attributable to COVID-19 than have non-Hispanic

---

perma.cc/Q3LS-SFXC]; *see also* Maritza V. Reyes, Note, *The Disproportional Impact of COVID-19 on African Americans*, 22 HEALTH & HUM. RTS. J. 299, 300–01 (2020).

146. *See, e.g.*, Natalia Norori et al., *Addressing Bias in Big Data and AI for Health Care: A Call for Open Science*, 2 PATTERNS, Oct. 8, 2021, at 1, 1 ("A major open challenge that AI will need to address before its integration in the clinical routine is that of algorithmic bias.").

147. Emily Moskal, *AI, Medicine, and Race: Why Ending 'Structural Racism' in Health Care Now is Crucial*, STAN. MED. (Oct. 20, 2023), https://scopeblog.stanford.edu/2023/10/20/ai-medicine-and-race-why-ending-structural-racism-in-healthcare-now-is-crucial/ [https://perma.cc/N234-7QBE]. For a more detailed view of her arguments, see Tina Hernandez-Boussard et al., *Promoting Equity in Clinical Decision Making: Dismantling Race-Based Medicine*, 42 HEALTH EQUITY 1369 (2023).

148. Moskal, *supra* note 147.

149. *Id.*

150. *Id.*

151. *Id.* ("After exploring the differences in outcomes, however, scientists found that chronic hypertension was the culprit, not race.").

152. *Id.* ("After exploring the differences in outcomes, however, scientists found that chronic hypertension was the culprit [for the worse outcomes,] not race.").

White (White) persons."[153] Since then, researchers have continued to note that "historically marginalized groups" did suffer "disproportionate harm."[154]

AI was of "limited efficacy" during the pandemic and some emerging data suggests that AI systems recommended undertreating Black and other minority patients, playing a role in disparate death rates by race.[155] Reflecting on this disproportionate death rate, Professors Sharona Hoffman and Andy Podgurski hypothesized that one explanation is that AI "training data may capture historical patterns of discrimination, causing the algorithm to perpetuate the inequitable treatment. This problem is called feedback loop bias."[156]

In a health care setting, this kind of programming could also cause harm by recommending reduced doses of pain medication based on pervasive beliefs that Black patients feel pain less acutely.[157] When getting this

153. 70 AMANDA R. SMITH ET AL., EMERGENCY DEPARTMENT VISITS FOR COVID-19 BY RACE AND ETHNICITY — 13 STATES, OCTOBER–DECEMBER 2020, at 556 (2021).

154. *See* Mark L. Shope, *NGO Engagement in the Age of Artificial Intelligence*, 28 BUFF. HUM. RTS. L. REV. 119, 138 (2022) ("AI systems are often opaque, so discrimination based on race, color, descent, national origin, or ethnic origin can be less obvious."); Leo Lopez III, Louis H. Hart III & Mitchell H. Katz, *Racial and Ethnic Health Disparities Related to COVID-19*, 325 JAMA 719, 719 (2021) ("One of the most disturbing aspects of the coronavirus disease 2019 (COVID-19) pandemic in the US is the disproportionate harm that it has caused to historically marginalized groups. Black, Hispanic, and Asian people have substantially higher rates of infection, hospitalization, and death compared with White people."); Trishan Panch, Heather Mattie & Leo A. Celi, *The "Inconvenient Truth" about AI in Healthcare*, 2 NATURE PARTNER JS., Aug. 16, 2019, at 1, 1; Katherine J. Igoe, *Algorithmic Bias in Health Care Exacerbates Social Inequities — How to Prevent it*, HARV. T.H. CHAN (Mar. 12, 2021), https://www.hsph.harvard.edu/ecpe/how-to-prevent-algorithmic-bias-in-health-care/ [https://perma.cc/8L45-MQRP].

155. David Leslie et al., *Does "AI" Stand for Augmenting Inequality in the Era of Covid-19 Healthcare?*, BMJ, Mar. 16, 2021, at 1, 1 ("The use of AI threatens to exacerbate the disparate effect of covid-19 on marginalized [sic], under-represented, and vulnerable groups, particularly black, Asian, and other minoritized [sic] ethnic people, older populations, and those of lower socioeconomic status.").

156. Hoffman & Podgurski, *supra* note 143 (citing David Casacuberta, *Bias in a Feedback Loop: Fueling Algorithmic Injustice*, CCCBLAB (May 9, 2018), https://lab.cccb.org/en/bias-in-a-feedback-loop-fuelling-algorithmic-injustice/ [https://perma.cc/UM5K-ABJP]); *see also* Panch, Mattie & Celi, *supra* note 154 ("For example, an algorithm trained on mostly Caucasian patients is not expected to have the same accuracy when applied to minorities.").

157. For more information about the misperception that Black patients experience less pain, see Ike Swetlitz, *Some Medical Students Still Think Black Patients Feel Less Pain Than Whites*, STAT (Apr. 4, 2016), https://www.statnews.com/2016/04/04/medical-students-beliefs-race-pain/ [https://perma.cc/86HA-ZHD4]; *see also* Sophie Trawalter, *Black Americans are Systematically Under-Treated for Pain. Why?*, UVA: FRANK BATTEN

information from a machine advertised as being smarter than any individual human, professionals may set aside their own judgement in favor of relying on the recommendation from the technology.[158]

III. LEGAL PROTECTION OF RESEARCH SUBJECTS: THE COMMON RULE

The tendency of AI technology towards racial bias and its increasing ability to manipulate decision-making are what make it inherently dangerous for use in the area of U.S. law which provides the most protection for free choice: the decision to participate in biomedical research. The laws that protect human participants in federally funded or regulated research today were promulgated in direct response to a shameful moment in the history of the U.S. Public Health Service: The Tuskegee Syphilis Experiment.[159]

Today, federal law protects the right of informed consent for participants in research funded, regulated, or conducted by the federal government.[160] This represents a process that began in 1974 with the passing of the National Research Act.[161] These principles have been developed into the rules set out in the Common Rule.

Congress has authorized Health and Human Services (HHS) to provide guidance for researchers in complying with these rules. In turn, HHS has

---

SCH. OF LEADERSHIP & PUB. POL'Y (June 30, 2020), https://batten.virginia.edu/about/news/black-americans-are-systematically-under-treated-pain-why [https://perma.cc/B8WX-994S].

158. For an account of how AI treatment recommendations are biased against Black patients, see Carolyn Y. Johnson, *Racial Bias in a Medical Algorithm Favors White Patients Over Sicker Black Patients*, WASH. POST (Oct. 24, 2019, 2:00 PM), https://www.washingtonpost.com/health/2019/10/24/racial-bias-medical-algorithm-favors-white-patients-over-sicker-black-patients/ [https://perma.cc/VQP4-XYYM].

159. Some accounts of the history of contemporary research participant protection laws credit the ethical violations detailed in Dr. Beecher's report as framing the work of the National Committee even though his report did not, itself, trigger the legislation. *See* Leslie Meltzer Henry, *Moral Gridlock: Conceptual Barriers to No-Fault Compensation for Injured Research Subjects*, 41 J.L. MED. & ETHICS 411, 411–12 (2013) ("Modern oversight of human subjects research grew out of a desire to protect participants from the kinds of research atrocities that Henry Beecher and others documented in the 1960s and 1970s.").

160. For a history of research protection laws before the Common Rule, see Todd W. Rice, *The Historical, Ethical, and Legal Background of Human-Subjects Research*, 53 RESPIRATORY CARE 1325, 1325, 1328–29 (2008) (noting the development of ethical standards for protecting human subjects since the Nuremberg trials).

161. National Research Service Award Act of 1974, Pub. L. No. 93-348, 88 Stat. 342. For an overview of how ethical principles were integrated into what was to become the Common Rule, see *History of Research Ethics*, UMKC, https://ors.umkc.edu/services/compliance/irb/history-of-research-ethics.html [https://perma.cc/4ABT-5XPM]; Hiroyuki Nagai, Eisuke Nakazawa & Akira Akabayashi, *The Creation of the Belmont Report and its Effect on Ethical Principles: A Historical Study*, 40 MONASH BIOETHICS REV. 157, 158, 161 (2022).

created a division currently called the Office for Human Research Protections (OHRP).[162] Many individual federal agencies have adopted the Common Rule and agreed to take guidance although they retain the right to develop their own policies.[163] OHRP provides clarification and guidance, develops educational programs and materials, maintains regulatory oversight, and provides advice on ethical and regulatory issues in biomedical and behavioral research.[164] As part of the harmonization process required by the 21st Century Cures Act, HHS and the Food and Drug Administration (FDA) have agreed that OHRP's guidance and interpretations are relevant for interpreting the FDA's human subject protection research regulation.[165] OHRP has authority to develop non-binding guidance for ethics committees, traditionally called Institutional Review Boards (IRB).[166] Each study must be reviewed in advance by an IRB to ensure that each participant is given an adequate opportunity to make a considered decision about whether to participate in the research and is not coerced.[167] Research subject to full

162. *See History*, U.S. DEPT. OF HEALTH & HUM. SERVS., https://www.hhs.gov/ohrp/about-ohrp/history/index.html [https://perma.cc/Q6PG-9GEG]. For an explanation of the role of the OHRP in providing guidance for compliance with the Common Rule, see Bernard A. Schwetz, *Protecting Subjects Without Hampering Research Progress: Guidance from the Office for Human Research Protections*, 74 CLEV. CLINIC J. MEDICINE S60, S60–63 (2007).

163. *See* Off. for Hum. Rsch. Prots., *Federal Policy for the Protection of Human Subjects ('Common Rule')*, U.S. DEP'T OF HEALTH & HUM. SERVS., https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html [https://perma.cc/9CKE-FRBW] ("Human subject research conducted or supported by each federal department/agency is governed by the regulations of that department/agency.").

164. *See* Schwetz, *supra* note 162.

165. *See generally* U.S. DEP'T OF HEALTH & HUM. SERVS. ET AL., INSTITUTIONAL REVIEW BOARD (IRB) WRITTEN PROCEDURES: GUIDANCE FOR INSTITUTIONS AND IRBS 1 (2018), https://www.fda.gov/media/99271/download [https://perma.cc/49PM-5C42] (recognizing the collaborative efforts of HHS, FDA, and OHRP to "enhance human subject protection and reduce regulatory burden" through administrative guidance's).

166. *See* 45 C.F.R. § 46.109(a) (2018) ("An IRB shall review and have authority to approve, require modifications in (to secure approval), or disapprove all research activities covered by this policy . . . ."). There is no legal requirement that the committees formed pursuant to the requirements of federal law be called "Institutional Review Boards" but that has become common practice. *See generally Institutional Review Boards Frequently Asked Questions*, FDA (Jan. 1998), https://www.fda.gov/regulatory-information/search-fda-guidance-documents/institutional-review-boards-frequently-asked-questions [https://perma.cc/Z9SD-B4NB].

167. This is an ongoing process. *See* Hilary Marston & Ann Meeker-O'Connell, *FDA Takes Steps to Further Harmonize Clinical Research Regulations with HHS Common Rule*, FDA (Oct. 6, 2022), https://www.fda.gov/news-events/fda-voices/fda-takes-steps-further-harmonize-clinical-research-regulations-hhs-common-rule [https://perma.cc/

review cannot proceed until the IRB has completed an ethical review of the proposed study.[168]

A. Adoption of the Common Rule for Protection of Participants in Clinical Drug Trials

Over the past seventy years, the federal agencies funding, conducting, or regulating research involving human beings have adopted these ethical principles as originally enacted and continue to abide by subsequent revisions and interpretations. As a result, these principles have come to be called the "Common Rule" in that they are standards shared across many different agencies.[169] However, there is still considerable variety in the decisions made by individual federal agencies to adopt the original Common Rule, the Revised Common Rule, or develop their own policies to comply with the requirements of the National Research Act.[170] For example, some agencies—and in particular the agency most relevant to the topic of clinical drug acts, the FDA—did not originally adopt any version

---

53BN-Z5FX]; *see also* Mary E. Schneider, *Stakeholders Seek Flexibility, Greater Harmonization in FDA Human Subject Protection Rules*, REGUL. FOCUS (Jan. 4, 2023), https://www.raps.org/news-and-articles/news-articles/2023/1/stakeholders-seek-flexibility-greater-harmonizatio (last visited Oct. 18, 2023). For an account of how the federal government has incorporated the mandate of the National Research Act either by adopting the Common Rule or developing their own standards, see Off. for Hum. Rsch. Prots., *supra* note 163 ("Each agency . . . [and] all participating departments and agencies the Common Rule outlines the basic provisions for IRBs, informed consent, and Assurances of Compliance. Human subject research conducted or supported by each federal department/agency is governed by the regulations of that department/agency."). For a clear explanation of studies that require less IRB review than a clinical trial, see Comm. on the Use of Hum. Subjects, *What Does and Does Not Require IRB Review and Approval?*, HARV. UNIV., https://cuhs.harvard.edu/what-does-and-does-not-require-irb-review-and-approval [https://perma.cc/582J-VXQ2]; 45 C.F.R. § 46.103(d) (2018) ("Under no condition shall research covered by this section be initiated prior to receipt of the certification that the research has been reviewed and approved by the IRB.").

168. For a recent description of the process of obtaining informed consent from a patient to participate in a research study testing new drug, see James M. Wilkins & Brent P. Forester, *Informed Consent, Therapeutic Misconception, and Clinical Trials for Alzheimer's Disease*, 35 INT'L J. GERIATRIC PSYCHIATRY, May 2022, at 1, 4–7.

169. *See* Off. for Hum. Rsch. Prots., *supra* note 163. When a legal issue such as whether or not a participant has been adjudicated incompetent arises, both FDA and OHRP "defer[] to state and local laws." *See Institutional Review Boards Frequently Asked Questions*, *supra* note 166 ("Therefore, the IRB should assure that the consent procedures comply with state and local laws, including assurance that the law applies to obtaining informed consent for subjects participating in research as well as for patients who require health care decisions.").

170. *See* Off. for Hum. Rsch. Prots., *supra* note 163 ("Human subject research conducted or supported by each federal department/agency is governed by the regulations of that department/agency.").

of the Common Rule.[171] Instead, it declined to sign on to the Common Rule and adopted its own interpretation of these principles to reflect unique features of its mission.[172] This reflected its far narrower jurisdiction.

Notably, unlike other divisions of HHS—which funds a broad range of human subject research—the FDA primarily oversees the drug trials funded by companies seeking permission to market their products in the United States.[173] However, with the passing of the 21st Century Cures Act, Congress has ordered the FDA to harmonize its practices with those of the Revised Common Rule and it is the process of doing so.[174]

Today, however, the FDA, by mandate of Congress, has in all relevant respects conformed its requirements for protecting humans participating in drug trials conducted in the United States with those of other divisions of HHS.[175] The FDA has responded to this OHRP cooperation mandate

171. *See* 21 C.F.R. § 50.1(a) (1980) (outlining the scope of human protections as adopted by the FDA in 1980).

172. *See, e.g.*, 21 C.F.R. § 56.101(a) (2007).

173. *See* 21 C.F.R. § 312.1(a) (1979) ("This part contains procedures and requirements governing the use of investigational new drugs, including procedures and requirements for the submission to, and review by, the Food and Drug Administration of investigational new drug applications (IND'S).").

174. *See* Scott Cunningham et al., *FDA Proposes Rules on Informed Consent and Institutional Review Boards*, COVINGTON (Oct. 12, 2022), https://www.cov.com/en/news-and-insights/insights/2022/10/fda-proposes-rules-on-informed-consent-and-institutional-review-boards [https://perma.cc/XS79-4CXY] ("On September 28, 2022, the Food and Drug Administration (FDA) published two proposed rules, seeking to amend its human subject protection regulations regarding informed consent and institutional review boards (IRBs). The proposed rules are part of FDA's ongoing efforts, as mandated by section 3023 of the Cures Act, to harmonize its regulations on human subject protection and IRBs with the revised Federal Policy for the Protection of Human Subjects (the revised Common Rule).").

175. *See* 21st Century Cures Act, Pub. L. No. 114-225, § 3060(a), 130 Stat. 1033 (2016) (codified as amended at 21 U.S.C. § 360j). Although the FDA has not framed its harmonization efforts as "adopting the Common Rule," this Article will use that term broadly to apply to protection of all prospective participants in federally funded or regulated research unless there is a specific reason to differentiate. *See* Protection of Human Subjects and Institutional Review Boards, 87 Fed. Reg. 58733, 58733 (Sept. 28, 2022) (to be codified at 21 C.F.R. pts. 50, 56, 812) ("The Food and Drug Administration (FDA or Agency) is proposing to amend its regulations to modernize, simplify, and enhance the current system for oversight of FDA- regulated human subject research. This proposed rule, if finalized, would harmonize certain sections of FDA's regulations on human subject protection and institutional review boards (IRBs), . . . with the revised Federal Policy for the Protection of Human Subjects (the revised Common Rule), in accordance with the 21st Century Cures Act (Cures Act)."); *see also* Jacqueline R. Berman, *What Does FDA Not Have in Common with the Common Rule*, MORGAN LEWIS (Jan. 25,

by proposing its own conforming regulations to "enhance the informed consent process for people considering participating in clinical drug trials to help them decide whether they should participate in the trial."[176]

The source document of the Common Rule, the Belmont Report, explicitly rejected adopting the standard of consent for medical care.[177] This, it explains, is because the civil malpractice standard for treating physicians who do not provide informed consent is "insufficient since the research subject, being, in essence, a volunteer, may wish to know considerably more about risks gratuitously undertaken than do patients who deliver themselves into the hand of a clinician for needed care."[178] It may be, as some argue, that Persuasive AI can play an important role in improving the process of informed consent for health care by shaping the information it presents in a way that best anticipates what a patient does not yet fully understand.[179] Some even suggest that it can be used as a substitute decision maker for patients who lose capacity by analyzing data about their past decisions in order to anticipate what they would decide in this situation.[180] But the Belmont Report chose to distinguish between these two situations by giving greater protection for participation in research than for consenting to medical treatment.[181] Bioethicists who argue against

---

2019), https://www.morganlewis.com/blogs/asprescribed/2019/01/what-does-fda-not-have-in-common-with-the-common-rule [https://perma.cc/YUG7-CHG5] (comparing FDA protection to Common Rule policies). The FDA rules governing the research it oversees still differ in some ways from those regulated by the common rule as well as those funded by the NIH, but none of those differences significantly alter the commitment to informed consent. For a comparison of the three, see David Peloquin et al., *Harmonizing the Common Rule and U.S. Food and Drug Administration Human Subjects Research Regulations*, ROPES & GRAY (Sept. 29, 2022), https://www.ropesgray.com/en/newsroom/alerts/2022/september/harmonizing-the-common-rule-and-us-food-and-drug-administration-human-subjects-research-regulations [https://perma.cc/GRK7-UHKS].

176. Marston & Meeker-O'Connell, *supra* note 167; *see also* Protection of Human Subjects and Institutional Review Boards, 87 Fed. Reg. at 58735 ("FDA and the Office for Human Research Protections (OHRP) have been actively working together for many years to harmonize regulatory requirements and guidance.").

177. *See* NAT'L COMM'N FOR THE PROT. OF HUM. SUBJECTS OF BIOMEDICAL & BEHAV. RES., THE BELMONT REPORT: ETHICAL PRINCIPLES AND GUIDELINES FOR THE PROTECTION OF HUMAN SUBJECTS OF RESEARCH 6 (1979) [hereinafter BELMONT REPORT], https://www.hhs.gov/ohrp/sites/default/files/the-belmont-report-508c_FINAL.pdf [https://perma.cc/C6GP-4UDG].

178. *Id.*

179. *See* Ziang Xiao et al., *Inform the Uninformed: Improving Online Informed Consent Reading with an AI-Powered Chatbot*, CHI '23: PROC. OF THE 2023 CHI CONF. ON HUM. FACTORS IN COMPUTING SYS., Apr. 2023, at 1 (examining the role of AI powered chatbots to improve informed consent).

180. *See, e.g.*, Camillo Lamanna & Lauren Byrne, *Should Artificial Intelligence Augment Medical Decision Making? The Case for an Autonomy Alogorithm*, 20 AMA J. ETHICS 902, 904–06 (2018).

181. *See* BELMONT REPORT, *supra* note 177.

"research exceptionalism" do so from the perspective of preserving the legal protections provided prospective subjects, not reducing them.[182] While there is an urgent need to evaluate and monitor the use of Persuasive AI in obtaining consent in a health care, the existence of binding federal law does make consent for research different.

AI proponents argue that AI's ability to anticipate what a patient wants to know is cause for permitting it. Indeed, it may play an important role in presenting material to those who must make important decisions about their own health. But this potential benefit for health care does not justify the risk that it could be used by those with a financial interest in enrolling participants in a commercial drug trial that, as will be discussed further, is by its definition for the purpose of gathering information, not for the benefit of any individual participant.

1. Putting Individual Rights Over Scientific Progress: The Legacy of Tuskegee

The Common Rule reflects a decision by Congress to put protecting the rights of individual participants above any benefit to society as a whole from the information obtained by the research.[183] Not only were the

182. *See, e.g.*, Alex J. London & Jonathan Kimmelman, *Against Pandemic Research Exceptionalism*, 368 SCI. 476, 477 (2020) (objecting to claims that emergency situations like the Covid-19 pandemic warrant a relaxation of scientific standards by arguing that "making research feasible by relaxing the other four standards contradicts the social justification for research"); James Wilson & David Hunter, *Research Exceptionalism*, 10 AM. J. BIOETHICS 45, 52 (2010) (responding to claims that human subject research was over-regulated compared to other activities of equal risk they present three reasons for retaining the existing protections: "research typically involves the imposition of risk on people who do not benefit from this risk imposition," "regulation builds or maintains public trust," and the "complexity of the moral decision making required favors ethics committees as a regulative solution for research makers"); Ezekiel J. Emanuel, *Ending Concerns About Undue Inducement*, 32 J. LAW MED. & ETHICS 100 (2004) (arguing that in the absence of legislative standards for paying research subjects ethics committees were being unduly cautious in limiting the amount participants could be paid for their time).

183. Thomas Ploug, *In Defence of Informed Consent for Health Record Research - Why Arguments from 'Easy Rescue', 'No Harm' and 'Consent Bias' Fail*, BMC MED. ETHICS, Aug. 20, 2020, at 1, 1 (citing WORLD MED. ASSN., WMA DECLARATION OF HELSINKI—ETHICAL PRINCIPLES FOR MEDICAL RESEARCH INVOLVING HUMAN SUBJECTS 1 (2022), https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/ [https://perma.cc/6S8K-329K] ("For decades the Helsinki Declaration's dictum that the interests of an individual must prevail over the interests of society has been one of the guiding principles for medical research, and informed consent has been the cornerstone of protecting research participants.")); *see also*

Tuskegee men deprived of effective treatments for a deadly disease, but their children and sexual partners were also unnecessarily endangered. Moreover, the government's deception was so complete that the men did not know they were enrolled in a study at all.[184] Because the current regulations were drafted in direct response to a single event, they share characteristics of other laws designed to prevent harm that has already happened rather than laws intended to prevent harm in the future: they are narrowly focused on events unlikely ever to happen exactly the same way again.

2. Protecting Consent Under the Common Rule

The final product of the congressional hearings regarding Tuskegee was a regulation, 45 CFR 46 (The Common Rule), that creates a framework for protecting human participants by requiring that all covered studies undergo a preliminary ethical review and that those posing more than minimal risk remain under the continuing review of an external committee with the authority to stop the study at any time.[185] It reflects the codification into law of the policy drafted by HHS to create a process of first prospective and then an ongoing ethical review of research studies involving human beings.[186]

3. 1997 Presidential Apology to Survivors of Tuskegee

On May 16, 1997, President Bill Clinton did something very unusual.[187] Standing in the East Room of the White House, accompanied by the Vice

---

Heintzelman, *supra* note 8 ("The Tuskegee Syphilis Study forced the nation to rethink and redefine practices involving human experimentation, especially those involving minority populations. As a consequence, HEW established a National Human Investigation Board, and legislation was passed requiring the establishment of Institutional Review Boards (IRBs).").

184. *See* Marcella Alsan & Marianne Wanamaker, *Tuskegee and the Health of Black Men*, 133 Q.J. ECON. 407, 412 (2018); *see also* Charlotte Paul & Barbara Brookes, *The Rationalization of Unethical Research: Revisionist Accounts of the Tuskegee Syphilis Study and the New Zealand "Unfortunate Experiment*," 105 AM. J. PUB. HEALTH e12, e13 (2015) ("The Tuskegee study clearly deceived participants: they were told they were receiving treatment when they were not.").

185. *See* 45 C.F.R. § 46.101(a) (2018).

186. The formal title of 45 CFR is "the Federal Policy for the Protection of Human Subjects." Because Federal Research Regulations apply only to research conducted, funded, and regulated by the federal government, protection of participants in research that takes place outside of these criteria is left to the states. *See, e.g.*, VA. CODE. ANN. § 32.1-162.16 (West 2023).

187. For an account of the media coverage of the apology, see *Coverage of the Apology*, TUSKEGEE UNIV., https://www.tuskegee.edu/about-us/centers-of-excellence/bioethics-center/coverage-of-the-apology [https://perma.cc/8UK3-Z9CV]. For an analysis of

President, the cabinet secretaries, and members of Congress, he turned to eight elderly Black men and described them as "a living link to a time not so very long ago that many Americans would prefer not to remember, but we dare not forget."[188] He continued, "But we can end the silence. We can stop turning our heads away. We can look you in the eye and finally say on behalf of the American people, what the United States government did was shameful, and I am sorry." Going on to introduce the men sitting in the room as "survivors of the syphilis study at Tuskegee," President Clinton acknowledged that "[w]hat was done cannot be undone," but describing "an apology" as "the first step," he pledged "to rebuild" the "broken trust" caused by the government's actions and committed to the ethical principles incorporated in today's protection laws to "mak[e] sure there is never again another episode like this one."[189]

In pledging "never again," President Clinton reaffirmed the country's commitment to human subject research protection stating, "Since the study was halted, abuses have been checked by making informed consent and local review mandatory in federally-funded and mandated research."[190]

4. 2019 Increased Commitment to Informed Consent in Revised Common Rule

On January 19, 2017, HHS, along with seventeen other agencies of the federal government, released the final rules for conducting research under the Revised Common Rule.[191] In doing so, it approved a major revision of the Common Rule, which strengthened the requirements for informed

---

the apology by an anthropologist, see Norielyn Romano, "*What Was Done Cannot Be Undone": Present-Day Apologies of Political Leaders for Transgressions of a Nation's Past*, 101 KROEBER ANTHROPOLOGICAL SOC'Y 79, 84–89 (2012).

188. *Remarks by the President in Apology for Study Done in Tuskegee*, THE WHITE HOUSE: OFF. OF THE PRESS SEC'Y (May 16, 1997, 2:26 PM), https://clintonwhitehouse4.archives.gov/New/Remarks/Fri/19970516-898.html [https://perma.cc/5B2C-HUW7] ("We need to do more to ensure that medical research practices are sound and ethical, and that researchers work more closely with communities.").

189. *Id.*

190. *Id.*

191. For an overview of how the Revised Common Rule strengthens the legal requirements for informed consent, see Victoria Berkowitz, Comment, *Common Courtesy: How the New Common Rule Strengthens Human Subject Protection*, 54 HOUS. L. REV. 923, 960 (2017) ("The Final Rule makes great strides to increase both transparency of informed consent and the understanding of human research subjects.").

consent.[192] Although some aspects of the Revised Common Rule reduced oversight of specific kinds of studies, its protection of the right of research subjects to informed consent for clinical drug trials is even greater than the original's.[193] Summarizing the purpose of the new consent provisions, OHRP guidance states that "[t]he intent of these changes is to promote prospective subjects' autonomy."[194] It explains further that "[i]nformed consent serves several purposes, but an important one is letting people make their own decisions about what they really want and what best serves their interests."[195] The purpose of adding more structure to the informed consent process was for people to "make their own decisions," they "need to have the necessary information conveyed in an appropriate way."[196] This contrasts with the standard of review for lack of informed consent to health care, which evaluates the harm caused by the failure to provide informed consent.[197] Protecting the consent process is as important today as when the Common Rule was first enacted.

---

192. *See* 45 C.F.R § 46.116(a) (2018) (detailing general requirements for obtaining informed consent). For guidance on complying with the Revised Common Rule, see Off. for Hum. Rsch. Prots., *Revised Common Rule*, U.S. DEP'T OF HEALTH AND HUM. SERVS., https://www.hhs.gov/ohrp/regulations-and-policy/regulations/finalized-revisions-common-rule/index.html [https://perma.cc/4BWN-98B5]; *see also FAQ Related to the Revised Common Rule*, JOHNS HOPKINS MED., https://www.hopkinsmedicine.org/institutional-review-board/guidelines-policies/guidelines/revised-common-rule-faq [https://perma.cc/96D5-VQ9D]. For a nuanced analysis of the changes in the Revised Common Rule, see Valerie G. Koch & Kelly Todd, *Research Revolution or Status Quo?: The New Common Rule and Research Arising from Direct-to-Consumer Genetic Testing*, 56 HOUS. L. REV. 81, 113 (2018) ("[R]esearchers are required to follow the Common Rule for federally-funded studies that are performed by either the company or a third party using their data.").

193. *See* 45 C.F.R. § 46.116(5)(i), (6) (2018) ("Informed consent must begin with a concise and focused presentation of the key information . . . [and] cannot include exculpatory language through which the subject or legally authorized representative is made to waive or appear to waive any of the subject's legal rights, or releases . . . the investigator, the sponsor, the institution, or its agents from liability for negligence."). For commentary on how the Revised Common Rule strengthened the consent process for clinical trials, see Jerry Menikoff, Julie Kaneshiro & Ivor Pritchard, *The Common Rule, Updated*, 376 NEW ENG. J. MEDICINE 613, 613–15 (2017); *see also* Leah L. LeCompte & Sylvia J. Young, *Revised Common Rule Changes to the Consent Process and Consent Form*, 20 OCHSNER J. 62, 62–66 (2020) (describing additional procedures in the Revised Common Rule).

194. Off. for Hum. Rsch. Prots., *Revised Common Rule Q&As*, U.S. DEP'T OF HEALTH & HUM. SERVS., https://www.hhs.gov/ohrp/education-and-outreach/revised-common-rule/revised-common-rule-q-and-a/index.html [https://perma.cc/7LRW-H8GQ].

195. *Id.*

196. *Id.*

197. While medical malpractice laws differ from state to state, most follow the framework created in *Canterbury v. Spence*, 464 F.2d 772, 785 (D.C. Cir. 1972). For a brilliant exposé of the weakness of informed consent laws related to health care, see Valerie G. Koch, *Eliminating Liability for Lack of Informed Consent to Medical Treatment*, 53 U. RICH. L. REV. 1211, 1224 (2019).

5. Application of the Common Rule to Clinical Drug Trials

While the FDA has no direct authority over human subject research, it exerts considerable control when that research is done for the purpose of supporting an application for marketing authorization.[198] Federal law requires that, in order to sell or distribute a drug in the United States, its manufacturer must secure approval from the FDA by showing safety and efficacy relative to already-available treatments before applying the product to human bodies for "diagnosis, treatment, mitigation, or prevention of a disease or condition."[199] This power of approval gives the FDA authority to set standards for what information manufacturers must provide.[200]

For studies conducted in the United States, manufacturers must get pre-approval from the FDA of a plan that details both the scientific methodology of how the trial will be conducted and how they will protect the rights of human participants.[201] If the data is collected outside the United States, manufacturers must certify that the participants were protected under international standards for human protection or the laws of the place where the research was conducted, whichever provides more protection.[202]

---

198. For a list of the FDA's own explanation of its role in protecting human participants in clinical trials, see *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/NJB4-53SW].

199. 21 C.F.R. § 312.8(b)(1)(i) (2009).

200. *See id.* § 312.1(a) ("This part contains procedures and requirements governing the use of investigational new drugs, including procedures and requirements for the submission to, and review by, the Food and Drug Administration of investigational new drug applications . . . .").

201. *See Premarket Approval (PMA)*, FDA (May 16, 2019), https://www.fda.gov/medical-devices/premarket-submissions-selecting-and-preparing-correct-submission/premarket-approval-pma [https://perma.cc/CT8P-SP4C] ("PMA approval is based on a determination by FDA that the PMA contains sufficient valid scientific evidence to assure that the device is safe and effective for its intended use(s).").

202. *See Acceptance of Data from Clinical Investigations for Medical Devices*, FDA (May 16, 2019), https://www.fda.gov/medical-devices/investigational-device-exemption-ide/acceptance-data-clinical-investigations-medical-devices [https://perma.cc/3A9Z-VGLZ] ("The FDA requires that data from clinical investigations conducted outside the US that began on or after February 21, 2019, be from investigations conducted in accordance with good clinical practice (GCP), which includes review and approval by an independent ethics committee (IEC) and informed consent from subjects."). For a discussion of why clinical trials are primarily conducted oversees, see Carolyn Thomas, *Why Big Pharma Now Outsources its Clinical Trials Overseas*, ETHICAL NAG (July 10, 2011), https://ethicalnag.org/2011/07/10/clinical-trials-outsourced-oversea/ [https://perma.cc/7YW5-M5PA].

Because the 21st Century Cures Act requires the FDA to apply the same standards as the Common Rule for research conducted in the United States, previous technical discrepancies between the two are of only historical interest.[203] The process of protecting human participants in an FDA-regulated clinical drug trial starts with the appointment by the sponsor of an investigator.[204] That investigator then assumes responsibility for compliance with human subject protection, including review by an IRB.[205]

Just as with a study funded by the federal government, the IRB reviewing a clinical drug trial "will be responsible for the initial and continuing review and approval of the clinical investigation."[206] Finally, the investigator must "report to the IRB all changes in the research activity and all unanticipated problems involving risks to human subjects or others, and . . . will not make any changes in the research without IRB approval, except where necessary to eliminate apparent immediate hazards to the human subjects."[207]

In general, clinical trials are inherently risky because "their outcomes, including benefit and harm, are unknowable."[208] Therefore, in creating the system of IRB review, the National Bioethics Advisory Commission (NBAC) noted that "risk is a central organizing principle, a filter through which protocols must pass."[209] It is for that reason that an IRB must first determine whether or not a study poses more than a "minimal risk" to potential participants before evaluating the extent to which that risk is warranted and, if so, how to communicate that risk.[210]

B. The Process of Testing an As Yet Unapproved Drug

Sponsors who can successfully navigate the FDA approval process to get permission to sell a new drug on the U.S. market stand to make millions,

---

203. *See* U.S. DEP'T OF HEALTH & HUM. SERVS., FDA & OFF. OF GOOD CLINICAL PRAC., IMPACT OF CERTAIN PROVISIONS OF THE REVISED COMMON RULE ON FDA-REGULATED CLINICAL INVESTIGATIONS 1–2 (2018), https://www.fdanews.com/ext/resources/files/2018/2/10-12-18-CommonRule.pdf [https://perma.cc/JY45-HE6W].

204. *See* 21 C.F.R. § 312.53 (2012).

205. *See id.*

206. *Id.* § 312.53(c)(1)(vii).

207. *Id.* § 312.66.

208. T. Patrick Hill, *Risk Assessment in Clinical Trials: It Don't Mean an Ethical Thing if it Ain't Got that Probability Ring!*, 8 ECANCERMEDICALSCIENCE, Sept. 4, 2014, at 1, 1 ("Since prospectively their outcomes, including benefit and harm, are unknowable, clinical trials take place under conditions of uncertainty.").

209. *Id.* (citing NAT'L BIOETHICS ADVISORY COMM'N, RESEARCH INVOLVING PERSONS WITH MENTAL DISORDERS THAT MAY AFFECT DECISIONMAKING CAPACITY 39 (1998)).

210. *See id.*

if not billions, of dollars.[211] By the time sponsors have a product ready to be tested in humans, they have likely spent millions of dollars and many years in development, laboratory testing, and animal testing.[212] While sponsors can proceed independently in these pre-human stages of drug development, once they are ready to start obtaining information about their drug's safety and efficacy in treating human beings, they must comply with FDA guidance or risk refusal of their application for approval to market their product in the United States.[213] This is true even though the sponsors are paying for all stages of human testing themselves without any financial support from the FDA. Sponsors must work with the FDA in advance to get pre-approval for all aspects of the clinical trial including how they are going to manufacture the product, what they are going to claim on its behalf, and who they will be using as test subjects.

Federal law prohibits transportation of a drug requiring FDA approval across state lines before it has obtained that approval.[214] Therefore, if getting the information required to obtain that approval is going to involve transporting or distributing the product across state lines, then sponsors "must seek an exemption from that legal requirement."[215] The process of seeking an exemption is called applying for an "Investigational New Drug" (IND).[216] To get an IND, sponsors must provide the FDA with information about both the method by which the sponsor will be manufacturing the new drug and the conditions under which it will be tested.[217]

---

211. *See* Rachana Pradhan, *The Business of Clinical Trials is Booming. Private Equity Has Taken Notice*, KFF HEALTH NEWS (Dec. 2, 2022), https://khn.org/news/article/business-clinical-trials-private-equity/ [https://perma.cc/KG3L-LNFK] ("Getting a drug to market a few months sooner and for less expense than usual can translate into millions in profit for the manufacturer.").

212. *See generally The Process of Drug Development: An Overview*, AVANTOR, https://www.avantorsciences.com/pages/en/biopharma-drug-development-process [https://perma.cc/F587-J3Y3] ("In some cases, the research and development (R&D) process for new treatments and therapies can take more than a decade and cost billions of dollars.").

213. *See* 21 C.F.R. § 312.30 (2009) ("Once an IND is in effect, a sponsor shall amend it as needed to ensure that the clinical investigations are conducted according to protocols included in the application.").

214. *Investigational New Drug (IND) Application*, FDA (July 20, 2022), https://www.fda.gov/drugs/types-applications/investigational-new-drug-ind-application [https://perma.cc/FBJ6-V7NP].

215. *Id.*

216. *See id.* ("The IND is the means through which the sponsor technically obtains this exemption from the FDA.").

217. *See id.*

The process of testing the safety and efficacy of an as yet unapproved drug requires exposing humans; therefore, the FDA requires that all such trials in the United States be approved in advance by obtaining an IND. Unlike federally funded research, which is done with the intention of expanding the store of generalizable knowledge, everyone involved in conducting a clinical trial—the manufacturer, the physician-researchers, and the institutions hosting the trials—share a goal of making a profit from the eventual approval of a new drug. The FDA monitors the information that sponsors provide to the public about enrollment in clinical trials to ensure that they are not promising results unavailable through conventional treatment.[218]

A clinical trial is the most common way to access a new kind of drug.[219] As *New York Times* reporter Gina Kolata explained in an article discussing the challenge sponsors faced in enrolling qualified patients, "Many of these experimental candidates in trials are quite similar. Yet each drug company wants to have its own proprietary version, seeing a potential windfall if it receives F.D.A. approval."[220] Additionally, a clinical trial participant is sick and therefore more vulnerable than a healthy participant.[221]

The law requires that sponsors make clear to patients that in being asked to enroll in a clinical trial for a drug which is not the first of its kind on the market, they are forgoing access to an FDA-approved treatment in favor one that is not expected to be any safer or more effective.[222] In blunt terms, the only beneficiary of this form of drug trial is the company that

---

218. *See Clinical Trials and Human Subject Protection*, FDA (May 2, 2023), https://www.fda.gov/science-research/science-and-research-special-topics/clinical-trials-and-human-subject-protection [https://perma.cc/6978-7KMX] ("FDA oversees clinical trials to ensure they are designed . . . according to federal law and good clinical practice (GCP) regulations.").

219. This is particularly true in the case of new cancer drugs. *See Ways to Access Experimental Cancer Drugs*, NAT'L CANCER INST. (Jan. 13, 2022), https://www.cancer.gov/about-cancer/treatment/drugs/access-experimental [https://perma.cc/DBS8-REL3].

220. Gina Kolata, *A Cancer Conundrum: Too Many Drug Trials, Too Few Patients*, N.Y. TIMES (Aug. 12, 2017), https://www.nytimes.com/2017/08/12/health/cancer-drug-trials-encounter-a-problem-too-few-patients.html [https://perma.cc/83MK-SXEJ].

221. For a discussion of the role of informed consent in medical research versus consent in usual care, see Michael K. Paasche-Orlow, Holly A. Taylor & Frederick L. Brancati, *Readability Standards for Informed-Consent Forms as Compared with Actual Readability*, 38 NEW ENG. J. MEDICINE (SPECIAL ARTICLE) 721, 722 (2003) ("[I]nformed consent for participation in medical research is particularly challenging because it requires a level of comprehension beyond that required for consent to usual care.").

222. *See* Mark A. Yarborough, *Increasing Enrollment in Drug Trials: The Need for Greater Transparency About the Social Value of Research in Recruitment Efforts*, 88 MAYO CLINIC PROC. 442, 442 (2013) (arguing for a system of increasing enrollment in clinical trials which informs potential participants of the "social value" of the "promise of research to improve clinical care").

will be able to market its own version of an already successful drug.[223] Even bioethicists most committed to the idea that participation in clinical trials is an obligation that everyone in society has to each other concede that "[s]ome drug trials, such as studies of 'me-too' drugs that test whether a new drug to treat a particular condition is no worse than already approved ones . . . have little if any social value."[224]

C. Enforcing Federal Research Participant Protection Laws

Federal law does not grant any participant in a regulated research study a private right of action against the government for a violation of their right to informed consent. Instead, the agencies and departments enforcing federal research subject protection laws work separately; in the end, all are subject to congressional oversight. So, the most effective method of banning the use of Persuasive AI in clinical drug trials would be an act of Congress. Barring direct congressional action, though, the agencies with oversight over human subject research already act independently of Congress to provide regulation guidance about what is and is not permissible.[225]

Within one month of the first declaration of a public health emergency, OHRP issued emergency guidance "regarding how the HHS human subjects protection regulations (45 CFR part 46) apply to actions taken by institutions and investigators in response to the COVID-19 outbreak," encouraging "the research community [to] prioritize public health and safety."[226] The FDA did the same, assuring "internal and external stakeholders regarding the conduct of clinical trials during the COVID-19 pandemic" that it was "providing appropriate regulatory flexibility to ensure protection of human subjects and to promote trial integrity."[227] Speaking directly to its grant holders, the National Institutes of Health (NIH) issued a similar

223. *See id.* at 444 ("[A] study is placing volunteers at risk to prove that a drug is 'effective,' ie, not inferior to its competitors, so that the company conducting the trial can try to capture a share of a multibillion dollar market.").

224. *Id.* at 442.

225. For a detailed analysis of when the Common Rule does and does not regulate research, see Michelle N. Meyer, *There Oughta Be a Law: When Does(n't) the U.S. Common Rule Apply?*, 48 J.L., MED. & ETHICS 60, 62 (2020).

226. *OHRP Guidance on Coronavirus*, U.S. DEP'T OF HEALTH & HUM. SERVS., https://www.hhs.gov/ohrp/regulations-and-policy/guidance/ohrp-guidance-on-covid-19/index.html [https://perma.cc/FA6A-YSBJ].

227. FDA, *Clinical Trial Conduct During the COVID-19 Pandemic*, EIN PRESSWIRE (May 9, 2023, 1:16 PM), https://www.einpresswire.com/article/632748073/clinical-trial-conduct-during-the-covid-19-pandemic [https://perma.cc/AZH4-8XPD].

directive telling them that "NIH understands that applicant and/or recipient institutions may need to exercise flexibility as they navigate the current public health emergency, and that some research may need to slow or pause altogether as hospitals and clinics prioritize patients affected by COVID-19, work to prevent exposure to patients and staff, and navigate supply chain interruptions."[228] The rapid need for social-distancing during the COVID-19 pandemic highlights the three major entities responsible for enforcing and interpreting federal human research participant laws.[229]

Research that is funded or conducted by agencies who have signed on to the Common Rule is under the direct supervision of the OHRP.[230] For trials funded by the federal government, this task is delegated to the OHRP which "provides leadership in the protection of the rights, welfare, and wellbeing of human subjects involved in research conducted or supported" by HHS.[231] Other agencies that fund or conduct research have their own offices of enforcement, although most defer to OHRP guidance.[232]

1. FDA

The FDA's greatest source of authority to enforce standards of participant protection is its ability to refuse to approve a manufacture's petition for approval to market the drug.[233] Without that approval, the company cannot recover any of its costs for research, development, and testing.

---

228. *Considerations for New and Ongoing Human Subjects Research During the COVID-19 Public Health Emergency*, NAT'L INST. OF HEALTH (Mar. 14, 2022), https://grants.nih.gov/sites/default/files/Considerations-New-Ongoing-Human-Subjects-Research-During-the-COVID-19-Public-Health-Emergency.docx [https://perma.cc/NW6M-F6JM].

229. For an overview of how the pandemic affected federally regulated research, see Comm. on the Use of Hum. Subjects, *What Does and Does Not Require IRB Review and Approval?*, *supra* note 167; *OHRP Guidance on Coronavirus*, *supra* note 226; *see also* Making Permanent Regulatory Flexibilities Provided During the COVID-19 Public Health Emergency by Exempting Certain Medical Devices from Premarket Notification Requirements; Withdrawal of Proposed Exemptions, 86 Fed. Reg. 20174, 20176–77 (Apr. 16, 2021).

230. Off. for Hum. Rsch. Prots., *Who Oversees the Regulations to Protect Research Participants?*, U.S. DEP'T OF HEALTH & HUM. SERVS., https://www.hhs.gov/sites/default/files/protecting-research-volunteers-english.pdf [https://perma.cc/74PW-L54C] ("OHRP oversees and enforces the Common Rule and other HHS regulations for protecting participants in research that is funded with HHS money.").

231. *Office for Human Research Protections*, U.S. DEP'T OF HEALTH & HUM. SERVS., https://www.hhs.gov/ohrp/index.html [https://perma.cc/2AEX-JW2W].

232. For an example from the Department of Education, see *Information About the Protection of Human Subjects in Research Supported by the Department of Education - Overview*, U.S. DEP'T OF EDUC. (May 19, 2011), https://www2.ed.gov/policy/fund/guid/humansub/overview.html [https://perma.cc/J7HQ-FJKZ].

233. *See generally Development and Approval | Drugs*, FDA (Aug. 8, 2022), https://www.fda.gov/drugs/development-approval-process-drugs [https://perma.cc/626Q-YKWD] ("FDA approval of a drug means that data on the drug's effects have been reviewed by

2. Drug Trials Inside the United States

Clinical drug trials conducted in the United States for the purpose of gathering data to submit for FDA approval operate under the supervision of the Office of Clinical Policy (OCLiP), which "develops good clinical practice and human subject protection policies, regulation and guidance, and addresses key clinical policy issues across the FDA's medical product centers" and monitors both the scientific integrity of the trial and the protection of human participants.[234]

3. Drug Trials Outside of the United States

Federal research protection laws do not extend to trials taking place outside of the United States unless that trial is funded or conducted directly by the federal government.[235] There is no single global source that promulgates or enforces standards for protecting humans participating in clinical drug trials. Clinical drug trials conducted outside of the United States are not under the direct protection of U.S. federal law. That has become a significant issue since, by some estimates, up to eighty percent of trials collecting data for approval applications to the FDA took place overseas.[236] These trials are regulated by a mosaic of different laws and policies.[237] Another source of international agreement comes from The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for

---

CDER, and the drug is determined to provide benefits that outweigh its known and potential risks for the intended population. The drug approval process takes place within a structured framework . . . .").

234. *Office of Clinical Policy*, FDA (Sept. 27, 2021), https://www.fda.gov/about-fda/office-clinical-policy-and-programs/office-clinical-policy [https://perma.cc/M2HL-U8UH].

235. *See* Brooke Y. Oki, Note, *Corporate Duty: Incentivizing Pharmaceutical Companies to Protect Human Rights in Their Foreign Clinical Trials Through Public Opinion and Internal Codes of Conduct*, 1 HOW. HUM. & C.R. L. REV. 157, 160–66 (2016–17) (evidencing that regulatory protections for clinical trials are often not enforced in foreign clinical trials).

236. *Id.* at 159 (citing Jennifer S. Bard, *A Taxonomy for Analyzing Legal and Ethical Issues Arising When Conducting Human Subject Research Outside the Borders of One's Own Country*, 37 HOUS. J. INT'L L. 1, 9 (2015)).

237. *International Compilation of Human Research Standards*, U.S. DEP'T OF HEALTH & HUM. SERVS., https://www.hhs.gov/ohrp/international/compilation-human-research-standards/index.html [https://perma.cc/55KR-B46E] ("The International Compilation of Human Research Standards is a listing of over 1,000 standards on human subjects protections in 131 countries and from many international organizations.").

Human Use (ICH) guidelines, which brings together the regulatory authorities and pharmaceutical industry to discuss scientific and technical aspects of pharmaceuticals and develop ICH guidelines.[238]

4. NIH

A third participant in enforcing human participant protection rules are the federal funding agencies such as the NIH and the National Science Foundation.[239] None of these entities have their own law enforcement units and therefore work closely with the Department of Justice.[240]

IV. REGULATING PERSUASIVE AI

> "*Government is the artery through which not only vital basic research funding flows, but also the rules, norms, and regulations that fortify acceptance and trust by the population of technological progress as something that is a net positive for humanity.*"[241]

A. Global Concern

The potential for Persuasive AI to manipulate decision-making in a way that violates the human right to autonomy is a matter of global concern.[242]

---

238. *See generally Mission*, ICH, https://www.ich.org/page/mission [https://perma.cc/25SL-P7UP] ("The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) is unique in bringing together the regulatory authorities and pharmaceutical industry to discuss scientific and technical aspects of drug registration.").

239. Clinical trials funded by the NIH but not conducted directly by a government agency must comply with both the Common Rule and NIH's own rules. *See* Peloquin et al., *supra* note 175.

240. *See, e.g.*, *What We Do*, NAT'L SCI. FOUND.: OFF. OF THE INSPECTOR GEN., https://oig.nsf.gov/investigations/what-we-do [https://perma.cc/C5JB-S6KJ] ("We investigate allegations of wrongdoing involving organizations or individuals that receive awards from, conduct business with, or work for NSF. When possible, we work in partnership with agencies and awardees to resolve issues.").

241. Ash Carter, *The Moral Dimension of AI-Assisted Decision-Making: Some Practical Perspectives from the Front Lines*, 151 DÆDALUS 299, 301 (2022).

242. *See* Ramona Vijeyarasa & José-Miguel Bello y Villarino, *Lessons and Consequences of the Failure to Regulate AI for Women's Human Rights*, OPENGLOBALRIGHTS (July 14, 2022), https://www.openglobalrights.org/lessons-and-consequences-of-failure-to-regulate-ai/ [https://perma.cc/H452-4P8L]; *see also* Emilie C. Schwarz, Note, *Human vs. Machine: A Framework of Responsibilities and Duties of Transnational Corporations for Respecting Human Rights in the Use of Artificial Intelligence*, 58 COLUM. J. TRANSNAT'L L. 232, 237 (2019) (citations omitted) ("[S]ome of the potential human rights at stake include: the right to privacy; the right to freedom of thought; the right to freedom of expression; the right to security; the right to be free from discrimination; the right to peaceful assembly and association; the right to work and free choice of employment; and finally, the right 'to a

The risk is enormous. For example, Max Roser, founder of an international organization that studies AI's "many great and terrifying problems,"[243] writes that, in its current form, AI has the capacity to create a "world with intelligent actors that are potentially very different from ourselves [and] . . . that is 'powerful enough to bring us into a new, qualitatively different future.'"[244] Calls by international organizations to regulate AI cover the three often referenced characteristics of dark AI: (1) threats to privacy, (2) bias, and (3) manipulation.[245] In September 2021, United Nations Human Rights chief Michelle Bachelet called for "moratoriums on the sale and use of artificial intelligence (AI) systems until adequate safeguards are put in place."[246] Pointing to existing examples of harm from biased decision-making and invasions of privacy, she justified the call for a ban on the grounds that "[t]he higher the risk for human rights, the stricter the legal requirements for the use of AI technology should be."[247]

Taking up the issue that same year, the WHO demonstrated their concern about the threat that AI poses to health care-related informed consent by recommending that it not be used at all until both the risks and likelihood of mitigation were better understood.[248] The WHO wrote that

---

social and international order in which the rights and freedoms set forth in this Declaration can be fully realized,' as enshrined in the Universal Declaration of Human Rights ("UDHR") Article 28.").

243. Max Roser, *About*, OUR WORLD IN DATA, https://ourworldindata.org/about [https://perma.cc/2YE4-YD6Z].

244. Max Roser, *Artificial Intelligence is Transforming our World —It is on All of Us to Make Sure That it Goes Well*, OUR WORLD IN DATA (Dec. 15, 2022), https://ourworldindata.org/ai-impact [https://perma.cc/7RA3-MKGT] (citing Holden Karnofsky, *AI Timelines: Where the Arguments, and the "Experts," Stand*, COLD TAKES (Sept. 7, 2021), https://www.cold-takes.com/where-ai-forecasting-stands-today/ [https://perma.cc/CW6F-BNJM]).

245. *See* Mike Thomas, *8 Risks and Dangers of Artificial Intelligence (AI)*, BUILT IN (Aug. 3, 2023), https://builtin.com/artificial-intelligence/risks-of-artificial-intelligence [https://perma.cc/B29S-XP57].

246. *Urgent Action Needed Over Artificial Intelligence Risks to Human Rights*, U.N. (Sept. 15, 2021), https://news.un.org/en/story/2021/09/1099972 [https://perma.cc/AT28-PNN6].

247. *Id.*; *see also* Daphne Leprince-Ringuet, *AI: Ban the Algorithms That Threaten Our Human Rights, Says UN Chief*, ZDNET (Sept. 17, 2021), https://www.zdnet.com/article/ai-ban-the-algorithms-that-threaten-our-human-rights-says-un-chief/ [https://perma.cc/4A32-K3VF] ("The UN commissioner's comments come off the back of a report that was carried out by her office to investigate how AI might impact human rights such as privacy, health, education, freedom of movement or freedom of expression.").

248. *See generally* SOUMYA SWAMINATHAN, *Foreword* to WORLD HEALTH ORG., ETHICS AND GOVERNANCE OF ARTIFICIAL INTELLIGENCE FOR HEALTH: WHO GUIDANCE, at

before using AI in health care we must put ethics and human rights at the heart of its design, deployment, and use.[249] The guidance highlighted the threat to informed consent by detailing the characteristics of AI that are likely to make informed consent impossible.[250]

Ana Palacio, a former Minister of Foreign Affairs of Spain and former senior vice president and general counsel of the World Bank Group, issued a similar statement. Summarizing the state of international concern in 2023, she made the case that AI's capacity to exert undue influence on human decision-making called for the kind of "global engagement that is increasingly shaping efforts to combat climate change."[251] Emphasizing the need for immediate action, she wrote that "the current regulatory vacuum must be filled."[252]

B. Timeline of EU Regulation of Persuasive AI

Given there are no comprehensive federal and very few state laws specific to either AI or data privacy,[253] the EU's existing and proposed regulations provide a helpful framework for identifying and responding to the threat posed by AI technology. This threat has the ability to influence decision-making beyond the ability of any previous human or technology.[254]

In April 2021, the European Commission issued a white paper announcing its plan to promulgate regulations that would create an "ecosystem of

---

v (2021), https://apps.who.int/iris/bitstream/handle/10665/341996/9789240029200-eng.pdf [https://perma.cc/57L9-ZYGD] ("For AI to have a beneficial impact on public health and medicine, ethical considerations and human rights must be placed at the centre of the design, development, and deployment of AI technologies for health. For AI to be used effectively for health, existing biases in healthcare services and systems based on race, ethnicity, age, and gender, that are encoded in data used to train algorithms, must be overcome.").

249. *See id.*

250. WORLD HEALTH ORG., ETHICS AND GOVERNANCE OF ARTIFICIAL INTELLIGENCE FOR HEALTH: WHO GUIDANCE 39–40, 47 (2021) [hereinafter WHO GUIDANCE], https://apps.who.int/iris/bitstream/handle/10665/341996/9789240029200-eng.pdf [https://perma.cc/57L9-ZYGD].

251. Ana Palacio, *Human Values for Artificial Intelligence*, PROJECT SYNDICATE (Jan. 16, 2023), https://www.project-syndicate.org/commentary/ai-regulatory-cooperation-ethics-us-china-rivalry-by-ana-palacio-2023-01 [https://perma.cc/ZZ72-YHT5].

252. *Id.*

253. *See Data Privacy Laws: What You Need to Know in 2023*, OSANO (Dec. 14, 2022), https://www.osano.com/articles/data-privacy-laws [https://perma.cc/6BHJ-XW8W].

254. *See* Pazzanese, *supra* note 15. HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, ETHICS GUIDELINES FOR TRUSTWORTHY AI 11 (Apr. 8, 2019), https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines.1.html [https://perma.cc/4D2X-9LGX] (describing the EU's expert group as "an independent group . . . mandated with the drafting of two deliverables: (1) AI Ethics Guidelines and (2) Policy and Investment Recommendations.").

trust" with hopes of promoting AI use throughout Europe.[255] These regulations call for the prohibition of AI systems that cause or are likely to cause "physical or psychological" harm through "subliminal techniques" or by exploiting vulnerabilities of a "specific group of persons due to their age, physical or mental disability."[256]

Noting that AI that manipulates decision-making "can affect the values on which the EU is founded and lead to breaches of fundamental rights," the document created a framework to address the "worry that AI can have unintended effects or even be used for malicious purposes."[257]

The final guidance reflects a two-year process.[258] The EU took its first steps towards regulating the harm caused by AI in 2016, when, anticipating AI's threat to privacy, it passed the General Data Protection Regulation (GDPR) to regulate the collection, storage, and use of personal data.[259]

---

255. EUR. COMM'N, WHITE PAPER ON ARTIFICIAL INTELLIGENCE: A EUROPEAN APPROACH TO EXCELLENCE AND TRUST 3 (2020), https://commission.europa.eu/system/files/2020-02/commission-white-paper-artificial-intelligence-feb2020_en.pdf [https://perma.cc/8E36-FKJA]. Reader's note, EU documents are translated into 22 different languages, none of which are American English. As a result, these documents sound stilted and may not always use words as readers in the United States would expect. *See generally Linguistic Coverage*, EUR-LEX, https://eur-lex.europa.eu/content/help/eurlex-content/linguistic-coverage.html [https://perma.cc/PU4F-DXZH]; Olivier Yves Alain Renard & Kristiina Milt, *Language Policy*, EUR. PARLIAMENT (Mar. 2023), https://www.europarl.europa.eu/factsheets/en/sheet/142/language-policy [https://perma.cc/7PJL-W562].

256. *Proposed EU AI Act*, *supra* note 5, at 43; *see also* MacCarthy & Propp, *supra* note 7 (discussing the circumstances under which the EU guideline bans use of AI assisted decision-making rather than attempting to limit or monitor its use).

257. EUR. COMM'N, *supra* note 255, at 11, 9 ("[I]ncluding the rights to freedom of expression, freedom of assembly, human dignity, nondiscrimination based on sex, racial or ethnic origin, religion or belief, disability, age or sexual orientation, as applicable in certain domains, protection of personal data and private life, or the right to an effective judicial remedy and a fair trial, as well as consumer protection. These risks might result from flaws in the overall design of AI systems (including as regards human oversight) or from the use of data without correcting possible bias (e.g. the system is trained using only or mainly data from men leading to suboptimal results in relation to women").).

258. *See* Kai Zenner, *Documents and Timelines: The Artificial Intelligence Act (Part 3)*, DIGITIZING EUR. (May 26, 2023), https://www.kaizenner.eu/post/aiact-part3 [https://perma.cc/M4K9-CJFS] (providing hyperlinks to official AI documents).

259. *See The History of the General Data Protection Regulation*, EUR, DATA PROT. SUPERVISOR, https://edps.europa.eu/data-protection/data-protection/legislation/history-general-data-protection-regulation_en [https://perma.cc/ZDG8-YY48]; *see also* Matt Burgess, *What is GDPR? The Summary Guide to GDPR Compliance in the UK*, WIRED (Mar. 24, 2020, 4:30 PM), https://www.wired.co.uk/article/what-is-gdpr-uk-eu-legislation-compliance-summary-fines-2018 [https://perma.cc/H74C-A2WQ] (discussing the implementation GDPR regulations in Britain, prior to Brexit).

The white paper identifies "[t]he specific characteristics of many AI technologies" that "may make it hard to verify compliance with, and may hamper the effective enforcement of, rules of existing EU law meant to protect fundamental rights."[260] These include the inability to perceive how AI technology works—the "black box-effect"—and its "complexity, unpredictability and partially autonomous behaviour."[261] Acknowledging these limitations, the commission recommended developing a risk-based approach to regulation with the greatest scrutiny given to applications with the potential to cause the most harm to fundamental rights.[262]

While the EU's use of the word "subliminal techniques" can be confusing, commentators agree that it generally refers to any method of interfering with individual decision-making.[263] Risto Uuk, an expert on European AI policy-making, urged the EU to go even further in preventing manipulation.[264] Specifically, Uuk supported the proposed EU AI Act, writing that the "subliminal techniques" and "exploiting vulnerabilities" described in the act were already being used by "many tech companies . . . to manipulate consumers."[265]

---

260. EUR. COMM'N, *supra* note 255, at 12 ("Enforcement authorities and affected persons might lack the means to verify how a given decision made with the involvement of AI was taken and, therefore, whether the relevant rules were respected. Individuals and legal entities may face difficulties with effective access to justice in situations where such decisions may negatively affect them.").

261. *Id.*; *see also* Alex Leslie, *Already the Dark Side of AI is Getting Worryingly Dark – Who Can Control it?*, DISRUPTIVE ASIA (June 22, 2021), https://disruptive.asia/dark-side-ai-really-dark-control/ [https://perma.cc/3PRZ-FZPL].

262. *Proposed EU AI Act*, *supra* note 5, at 12 ("The regulation follows a risk-based approach, differentiating between uses of AI that create (i) an unacceptable risk, (ii) a high risk, and (iii) low or minimal risk.").

263. *See* Patrick Grady, *EU's AI Act Resurrects Subliminal Messaging Panic*, CTR. FOR DATA INNOVATION (Oct. 21, 2022), https://datainnovation.org/2022/10/eus-ai-act-resurrects-subliminal-messaging-panic/ [https://perma.cc/NHC8-F49N] (describing subliminal advertising as a "hoax"). *But see* ROSTAM J. NEUWIRTH, THE EU ARTIFICIAL INTELLIGENCE ACT: REGULATING SUBLIMINAL AI SYSTEMS 44–45 (2023) (differentiating subliminal techniques from other, similar, methods that interfere with decision-making).

264. *See* Risto Uuk, *The EU Needs to Protect (More) Against AI Manipulation*, EURACTIV (Feb. 7, 2022), https://www.euractiv.com/section/digital/opinion/the-eu-needs-to-protect-more-against-ai-manipulation/ [https://perma.cc/AUU5-LRKD].

265. *Id.* (explaining that the while the proposal does not define the word "subliminal," the "risks of manipulation from AI systems aren't merely hypothetical, they already threaten individuals and communities and can lead to further harms if not adequately prepared for"). In addition to the EU, the UK, Canada, and New Zealand have all issued proposals for limiting the harm of AI that interferes with decision-making. Now that Great Britain is no longer a member of the EU it is following its own path to regulating AI. *See Europe: UK vs EU Approach to Regulating AI*, DATAGUIDANCE (Oct. 2022), https://www.dataguidance.com/opinion/europe-uk-vs-eu-approach-regulating-ai [https://perma.cc/J4PT-JNUK] (describing that while the EU has opted for a broad legislative framework, the UK has chosen to rely upon sectoral regulators and will pursue an agenda focused on

C. Taking A Risk Based Approach

Former U.S. Secretary of Defense and AI ethics expert Ash Carter adopts the EU's risk based approach by proposing a hierarchy of potential risk in some fields where AI is used, such as "[e]ntertainment and advertising," as "fairly error tolerant."[266] However, Carter notes "there are applications that require much more ethical scrutiny."[267] Among these, he gives the highest priority to "national security . . . ; law enforcement; health care; autonomous vehicles of all kinds; [and] fairness in credit, housing, and employment."[268] Specifically, the EU's proposed guidelines offer "a nuanced regulatory structure that bans some uses of AI, heavily regulates high-risk uses and lightly regulates less risky AI systems."[269] The guidelines also identify a category of use so dangerous that the recommended regulation is a unilateral ban.[270] Namely, the EU recommends prohibiting uses of AI software that "deploy subliminal techniques in order to materially distort a person's behaviour."[271] This use poses "a clear threat to the safety, livelihoods and rights of individuals and violate[s] the EU's values and fundamental rights."[272]

These guidelines address "the *societal risks* and *unintended consequences* of the rapid deployment of technology including Predictive AI in areas such as health and fintech."[273]

---

promoting technological advancement and maintaining AI "superpower' status"); *see also* David Matthews, *UK Rejects EU Approach to Artificial Intelligence in Favour of 'Pro-Innovation' Policy*, SCI. BUS. (July 19, 2022), https://sciencebusiness.net/news/uk-rejects-eu-approach-artificial-intelligence-favour-pro-innovation-policy [https://perma.cc/BS7J-CXE5]; Oliver Yaros et al., *UK Government Publishes National Artificial Intelligence Strategy*, MAYER BROWN (Oct. 20, 2021), https://www.mayerbrown.com/en/perspectives-events/publications/2021/10/uk-government-publishes-national-artificial-intelligence-strategy [https://perma.cc/8ZFX-SWHE].

266. Carter, *supra* note 241, at 305.

267. *Id.*

268. *Id.*

269. MacCarthy & Propp, *supra* note 7.

270. *See* Cynthia O'Donoghue, Andreas Splittgerber & Sarah O'Brien, *The Proposed European Regulation on Artificial Intelligence – A Summary of the Obligations, Scope, and Effect*, REEDSMITH (May 7, 2021), https://www.reedsmith.com/en/perspectives/2021/05/the-proposed-european-regulation-on-artificial-intelligence—a-summary-of [https://perma.cc/33P8-JJRG] (discussing "[p]rohibited AI practices" that are subject to an outright ban).

271. *Id.*

272. *Id.*

273. Karen Eltis, *Assisted Decision-Making and the Proposed EU AI Regulation: An Emerging Paradigm Shift from Consent to Contextually Mitigating Human Rights Violations*, SLAW (Nov. 24, 2021), http://www.slaw.ca/2021/11/24/assisted-decision-making-and-

D. U.S. Federal Law

In contrast to the comprehensive existing and proposed AI regulation in other countries, the United States has allowed the industry to develop unchecked.[274] As of Spring 2023, "[n]o bill has been proposed to curb A.I.'s potential dangers or to protect individuals, and efforts to restrict facial-recognition applications have failed."[275] While there has been some "spill-over" protection from companies doing business in jurisdictions that regulate AI,[276] companies have primarily been allowed to self-regulate (or not) as they choose.[277]

This is despite frequent calls for regulation by scholars, citizen advocate groups, individual lawmakers, and the AI industry itself.[278] Criticizing

---

the-proposed-eu-ai-regulation-an-emerging-paradigm-shift-from-consent-to-contextually-mitigating-human-rights-violations/ [https://perma.cc/3KEP-YGA3].

274. *See* Christopher S. Yoo & Alicia Lai, *Regulation of Algorithmic Tools in the United States*, 13 J.L. & ECON. REGUL. 7, 7 (2020). ("The U.S. approach to regulating algorithmic decision-making is characterized by a reliance on soft standards and certifications. Rather than a unified set of strict regulations or sector-specific rules, the U.S. president, federal agencies, individual states, and local governments have proposed piecemeal legislation to promote research, create task forces, mandate reports and recommendations, and pursue other forms of light-touch regulation.").

275. Andrew R. Sorkin et al., *Why Lawmakers Aren't Rushing to Police A.I.*, N.Y. TIMES (Mar. 3, 2023), https://www.nytimes.com/2023/03/03/business/dealbook/lawmakers-ai-regulations.html [https://perma.cc/T9RX-DVBU]. For a systemic review of the gaps in U.S. law related to AI, including a section on consumer manipulation, see Carlos Ignacio Gutierrez Gaviria, *The Role of Artificial Intelligence in Pushing the Boundaries of U.S. Regulation: A Systematic Review*, 38 SANTA CLARA HIGH TECH. L.J. 123, 181–82 (2022).

276. This has had some beneficial spillover effect since it exerts pressure on companies that offer services in the EU to extend at least some of these protections to everyone using their products. *See* Christian Peukert et al., *Regulatory Spillovers and Data Governance: Evidence from the GDPR*, 41 MKTG. SCI. 746 (2022). For example, the ubiquitous request to opt in or out of accepting cookies is based on a UK law, not any U.S. legal requirement.

277. *See,e.g.*, Sumeet Wadhwani, *Seven U.S. Tech Companies Voluntarily Commit to AI Guardrails*, SPICEWORKS (July 25, 2023), https://www.spiceworks.com/tech/artificial-intelligence/news/ai-self-regulation-united-states/ [https://perma.cc/XW8M-62LB] (recognizing the right of companies to voluntarily self-regulate their use of AI); *Leading HR Technology Company Launches AI Ethics Advisory Board*, CISION PR NEWSWIRE (Jan. 26, 2021, 2:02 PM), https://www.prnewswire.com/news-releases/leading-hr-technology-company-launches-ai-ethics-advisory-board-301215510.html [https://perma.cc/FW5B-43Y2]; *see also* Zoe Schiffer, *Google Fires Second AI Ethics Researcher Following Internal Investigation*, THE VERGE (Feb. 19, 2021, 2:52 PM), https://www.theverge.com/2021/2/19/22292011/google-second-ethical-ai-researcher-fired [https://perma.cc/M38W-NVJ3] (noting that Google started its own "ethical AI team" in 2018).

278. Professor Margaret Hu has published a series articles focused on the potential for violation when AI accesses biometric data. *See generally* Margaret Hu, *Biometrics and an AI Bill of Rights*, 60 DUQ. L. REV. 283, 301 (2022) ("[T]he proposed AI Act and the GDPR combined offer important ways to construct the types of rights and values necessary for an effective AI Bill of Rights, including the need to conceptualize data rights as fundamental rights and how biometric AI systems can infringe upon criminal procedure

the United States' complacency, Carly Kind, the director of the Ada Lovelace Institute, warned that "[b]y failing to establish such guardrails, policymakers are creating the conditions for a race to the bottom in irresponsible A.I."[279] As a result, we are, as a country, completely unprepared for the threat Persuasive AI poses.[280]

1. White House AI Bill of Rights

In October 2022, the White House Office of Science and Technology Policy issued the "AI Bill of Rights."[281] Acknowledging that the United States was not "the first mover in this space," Professor Eunice Park commented early on that unlike the existing legislation in other countries, "the choice to cast the principles in terms of a 'Bill of Rights' is distinctly American."[282] The White House has also issued nonbinding guidance for future regulation of AI.[283] Without using the term "AI Assisted Decision-Making" or Emotion AI, it identifies many of the same issues addressed by the EU.[284] Specifically, it considers the rights of individuals who face the consequences of a decision made using AI, such as someone who was

---

rights."); Margaret Hu, *Algorithmic Jim Crow*, 86 FORDHAM L. REV. 633, 663–71 (2017) (discussing the potential appearance of front end equality hiding back end discrimination); Margaret Hu, *Biometric ID Cybersurveillance*, 88 IND. L.J. 1475 (2013) (discussing risks of expanding mass surveillance); Margaret Hu, *Crimmigration-Counterterrorism*, 2017 WIS. L. REV. 955, 976 (2017) (discussing the conflation of crime, terrorism, and immigration politics through "extreme vetting" biometrics data collection); Margaret Hu, *Horizontal Cybersurveillance Through Sentiment Analysis*, 26 WM. & MARY BILL RTS. J. 361, 372 (2017) (discussing digital sentiment analysis through the lenses of potential first and fourth amendment violations). *See generally* Donald L. Buresh, *Should Personal Information and Biometric Data Be Protected Under a Comprehensive Federal Privacy Statute that Uses the California Consumer Privacy Act and the Illinois Biometric Information Privacy Act as Model Laws?*, 38 SANTA CLARA HIGH TECH. L.J. 39 (2022) (arguing that although the laws related to collection of biometric information are seriously lacking, it falls under the more traditional existing civil and criminal privacy laws).

279. Sorkin et al., *supra* note 275.

280. *See id.*

281. *Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People*, WHITE HOUSE, https://www.whitehouse.gov/ostp/ai-bill-of-rights/ [https://perma.cc/2MQ5-6E2T].

282. Eunice Park, *The AI Bill of Rights: A Step in the Right Direction*, 65 ORANGE CNTY. LAW., Feb. 13, 2023, at 25, 25 (citing Eliza Strickland, *6 Reactions to the White House's AI Bill of Rights*, IEEE SPECTRUM (Oct. 14, 2022), https://spectrum.ieee.org/white-house-ai [https://perma.cc/JFF9-DD6Y]).

283. *See Blueprint for an AI Bill of Rights*, *supra* note 281.

284. *See id.*

denied a job or a loan based on AI's racial bias.[285] It asserts that people "should not face discrimination by algorithms and systems should be used and designed in an equitable way."[286]

2. Federal Trade Commission Report

So far, the White House's call for regulation has not yielded action except that the Federal Trade Commission (FTC) similarly warned against the potential harms of over-reliance upon AI. This year, for example, the FTC responded to a congressional request for proposals on how to combat online harms such "scams, deepfakes, fake reviews, opioid sales, child sexual exploitation, revenge pornography, harassment, hate crimes, and the glorification or incitement of violence."[287] The FTC warned Congress that AI would not be an appropriate or effective way of addressing such harm.[288] In its warning, the FTC relied on a report explaining that while there are many products now being offered that claim to use AI to address the kind of online harms Congress had identified, it is "crucial to understand that these tools remain largely rudimentary, have substantial limitations, and may never be appropriate in some cases as an alternative to human judgment."[289]

3. The FDA

So far, there is no federal regulation specifically protecting the privacy of identifiable health information in the context of AI technology.[290] The FDA has announced its intent to develop a framework for regulating AI

---

285. *See id.*

286. *Id.*

287. FED. TRADE COMM'N, COMBATTING ONLINE HARMS THROUGH INNOVATION 1 (2022), https://www.ftc.gov/system/files/ftc_gov/pdf/Combatting%20Online%20Harms%20Through%20Innovation%3B%20Federal%20Trade%20Commission%20Report%20to%20Congress.pdf [https://perma.cc/Z2NP-HAPA].

288. *Id.* at 2 ("No matter how these harms are generated, technology and AI do not play a neutral role in their proliferation and impact."); *see also FTC Report Warns About Using Artificial Intelligence to Combat Online Problems*, FED. TRADE COMM'N (June 16, 2022), https://www.ftc.gov/news-events/news/press-releases/2022/06/ftc-report-warns-about-using-artificial-intelligence-combat-online-problems [https://perma.cc/5WBW-BNGD] ("'Our report emphasizes that nobody should treat AI as the solution to the spread of harmful online content,' said Samuel Levine, Director of the FTC's Bureau of Consumer Protection. 'Combatting online harm requires a broad societal effort, not an overly optimistic belief that new technology—which can be both helpful and dangerous—will take these problems off our hands.'").

289. FED. TRADE COMM'N, *supra* note 287, at 5.

290. *See* Susan Kelly, *FDA Issues Action Plan for Regulating AI in Medical Devices*, MEDTECH DIVE (Jan. 13, 2021), https://www.medtechdive.com/news/fda-issues-action-plan-for-regulating-ai-in-medical-devices/593280/ [https://perma.cc/XU4V-W7ZW].

use in medical devices.[291] Despite the longtime integration of AI technology in delivering health care in the United States, the regulatory agencies that would ordinarily protect patient interests have done little.[292] So, although AI has been part of medical devices for decades, it has yet to be regulated and there are no binding rules about how it can be used. One challenge continues to be that there is little agreement regarding how to describe AI's role, let alone how to regulate it.[293]

In September 2022, however, the FDA issued final guidance for industry, identifying "a list of artificial intelligence tools that should be regulated as medical devices."[294] Among those tools were hardware and "previously unregulated software products."[295] Professor Sara Gerke described the problem the FDA is trying to solve as whether "AI-based medical devices can be biased, opaque, and/or adaptive."[296] So far, however, the FDA has met strong opposition from industry, making it unlikely that these rules will be implemented in the near future.[297]

E. U.S. State and City Laws Regulating AI

In response to the U.S. federal government's relative inaction, several states and many cities have stepped forward with their own AI laws.[298]

---

291. *Id.*

292. *See* Gali Katznelson & Sara Gerke, *The Need for Health AI Ethics in Medical School Education*, 26 ADVANCES HEALTH SCIS. EDUC. 1447, 1452–53 (2021).

293. *See* Hassane Alami et al., *Organizational Readiness for Artificial Intelligence in Health Care: Insights for Decision-Making and Practice*, 35 J. HEALTH ORG. & MGMT. 106, 106–09 (2020) (citing Saurabh Jha & Eric J. Topol, *Adapting to Artificial Intelligence: Radiologists and Pathologists as Information Specialists*, 316 JAMA 2353, 2353 (2016)) ("Artificial intelligence (AI) [is] generically defined as 'the imitation of human cognition by computers.'").

294. Casey Ross, *In New Guidance, FDA Says AI Tools to Warn of Sepsis Should be Regulated as Devices*, STAT (Sept 27, 2022), https://www.statnews.com/2022/09/27/health-fda-artificial-intelligence-guidance-sepsis/ [https://perma.cc/W5CC-2Q5Y].

295. *Id.*

296. Sara Gerke, *Health AI for Good Rather Than Evil? The Need for a New Regulatory Framework for AI-Based Medical Devices*, 20 YALE J. HEALTH POL'Y, L. & ETHICS 432, 503 (2021).

297. *See* Lizzy Lawrence, *The FDA Plans to Regulate Far More AI Tools as Devices. The Industry Won't Go Down Without a Fight*, STAT (Feb. 23, 2023), https://www.statnews.com/2023/02/23/fda-artificial-intelligence-medical-devices/ [https://perma.cc/8MM9-SHC3].

298. *See, e.g.*, Alyssa N. Lankford, *States and Cities Limit AI Use in Employment Decisions*, MCAFEE & TAFT (Jan. 20, 2023), https://www.mcafeetaft.com/states-and-cities-limit-ai-use-in-employment-decisions/ [https://perma.cc/U3XP-RBMV] ("[S]tates

The state-law action reflects a high level of interest[299] and concern[300] about AI among the American public.

Several states have now followed the EU's lead by adopting their own laws to protect against the harm caused by AI's ability to collect, analyze, and disseminate data.[301] For example, the California Consumer Privacy Act (CCPA) gives consumers (1) the right to know what personal information is being collected about them and (2) the right to opt-out of the sale of their personal information.[302] Massachusetts has banned the use of facial recognition technology by all of its police departments.[303]

Further, several large cities have also passed AI regulations. For example, the California cities of San Francisco, Oakland, and Berkeley banned the use of facial recognition by law enforcement.[304] And New York City banned AI in employment decisions.[305]

V. HOW CURRENTLY AVAILABLE AI TECHNOLOGY MANIPULATES DECISION-MAKING

What makes today's AI so different than any previous technology used to assist decision-making is that instead of analyzing only the data it is given, such as a single hospital's billing records, modern AI "dynamically incorporates new data from its operating environment to generate more

---

and cities have begun passing and implementing laws aimed at the use of AI in employment decision-making.").

299. *See* MORNING CONSULT, PERCEPTIONS OF ARTIFICIAL INTELLIGENCE 6 (2021), https://uploads-ssl.webflow.com/615311db77195c2a5b2b504c/6172b8c6c7d1362d8a0a65fc_2110008-Seed%20AI-D4.pdf [https://perma.cc/2EUF-KAQA] ("Majorities of adults support the U.S. investing in AI education for students (74%), as well as training workers for AI jobs (69%) and developing AI technology (68%).").

300. *See id.* at 3 ("The U.S. is lagging when it comes to regulating AI: Half of adults (50%), including 48% of Republicans and 53% of Democrats, believe the U.S. should have already begun regulating AI development and use.").

301. *See* Cameron F. Kerry, *Protecting Privacy in an AI-Driven World*, BROOKINGS (Feb. 10, 2020), https://www.brookings.edu/articles/protecting-privacy-in-an-ai-driven-world/ [https://perma.cc/SFH4-54WN].

302. *See California Consumer Privacy Act (CCPA)*, CAL. DEP'T OF JUST. (May 10, 2023), https://oag.ca.gov/privacy/ccpa [https://perma.cc/GVW7-LE7B].

303. Emma Peaslee, *Massachusetts Pioneers Rules for Police Use of Facial Recognition Tech*, NPR (May 7, 2021, 6:00 AM), https://www.npr.org/2021/05/07/982709480/massachusetts-pioneers-rules-for-police-use-of-facial-recognition-tech [https://perma.cc/3SLZ-ALPH] ("Massachusetts lawmakers passed one of the first state-wide restrictions of facial recognition as part of a sweeping police reform.").

304. Kerry, *supra* note 301.

305. Lankford, *supra* note 298 ("New York City's Local Law Int. No. 144, which went into effect on January 1, 2023, prohibits an employer or employment agency from using an 'automated employment decision tool' to screen a candidate or employee for an employment decision . . . .").

accurate insights on a real-time basis."[306] This characteristic alone makes it impossible for the material to be approved in advance, as required by federal research protection law.[307] This pre-approval usually comes from an IRB.[308]

A. Customization of Advertising

Persuasive AI boasts that its advertising materials are particularly effective because they are customized in real-time in response to the target's emotional reactions.[309] This alone is disqualifying because such mutability makes it impossible to comply with federal law, which requires that an IRB pre-approve material that will be presented to the potential subject. Because the AI learns as it goes, every participant essentially engages with a different program.[310]

B. How Does Persuasive AI Exceed the Legal Boundaries Set by the Common Rule?

There are at least two ways AI can transcend the legal limits of persuasion.[311] One is by engaging in behavior for which a human could not ethically

306. Nicholas Larsen, *Adaptive AI: The Next Evolutionary Stage for Artificial Intelligence?*, INT'L BANKER (Mar. 16, 2023), https://internationalbanker.com/technology/adaptive-ai-the-next-evolutionary-stage-for-artificial-intelligence/ [https://perma.cc/2W89-EPYG] ("With adaptive AI systems proving adept at continuously responding, learning and modifying their outputs from ingesting new data, this technology's capabilities look set to be dramatically upgraded.").

307. *See* 45 C.F.R. § 46.109(a) (2022) ("An IRB shall review and have authority to approve, require modifications in (to secure approval), or disapprove all research activities covered by this policy.").

308. *Id.*

309. For an example of real-time AI, see James Hendler, David Musliner & Bob Kohout, *Real-Time A.I.*, PARALLEL UNDERSTANDING SYS. GRP., http://www.cs.umd.edu/projects/plus/Realtime/ [https://perma.cc/JN6R-53CZ]; *see also Top 10 Real Time Artificial Intelligence Applications*, BESANT TECHS., https://www.besanttechnologies.com/artificial-intelligence-applications [https://perma.cc/Z3EB-NGK3] (discussing other use cases of real time AI).

310. *See generally* Larsen, *supra* note 306 ("[A]daptive AI systems are more reactive to the changing world around them and can thus more seamlessly adapt to new environments and circumstances that were not present during the earlier stages of the AI system's development.").

311. *See generally* Beth Barnes, *Risks from AI Persuasion*, LESSWRONG (Dec. 23, 2021), https://www.lesswrong.com/posts/5cWtwATHL6KyzChck/risks-from-ai-persuasion

obtain consent, and the other is by engaging in manipulation beyond the bounds of human ability. The first category includes behavior that crosses the line from merely presenting information to selling the study. For example, telling a participant that they are "the ideal" person and that this study will "help them." Equally unethical would be generating an emotional response intended to build rapport or create a feeling of trust. This behavior may be acceptable in a hostage–negotiator or even a pediatrician–emergency room patient setting, but not in an informed consent context. The equivalent of this unethical and illegal behavior would be to offer untrue information, such as a statement that "my mother had this condition, and I only wish she could have had the opportunity to enroll."

C. Limits on Persuasive Techniques in the Informed Consent Process

The Common Rule specifically prohibits "undue influence" and "coercion" in the informed consent process, although neither is a defined term.[312] The source document on which the Common Rule is based, the Belmont Report, is explicit that these factors are most likely to present when the potential participant was also a patient receiving medical care.[313]

Patients enrolled in clinical trials consistently confuse the role of the researcher and the physician when asked whether they believed that participating in a clinical trial would improve their chances of a good outcome.[314] Moreover, even if patients did understand the difference between research and treatment, they were still likely not to understand the risks

---

[https://perma.cc/TP6F-YEBK] (discussing the "distinct" risks associated with Persuasive AI).

312. Emily A. Largent & Holly F. Lynch, *Paying Research Participants: Regulatory Uncertainty, Conceptual Confusion, and a Path Forward*, 17 YALE J. HEALTH POL'Y, L., & ETHICS 61, 65–66, 80 (2017) (citing 45 C.F.R. § 46.116 (2015)) (stating that although U.S. and international "codes of research ethics require that consent to participation in research be obtained in a manner that minimizes the possibility of coercion and undue influence[,] . . . [t]he Common Rule does not define either term").

313. *See* BELMONT REPORT, *supra* note 177, at 7–8.

314. *See* Wanda Montalvo & Elaine Larson, *Participant Comprehension of Research for Which They Volunteer: A Systematic Review*, 46 J. NURSING SCHOLARSHIP 423, 423, 424 (2014) (concluding that a major problem with obtaining adequate informed consent is "continued therapeutic misconception and lack of understanding among research participants of randomization, placebo, benefit, and risk"). This consistent confusion has led many ethicists to argue that physicians conducting research involving patients should ethically assume the same fiduciary responsibility as a physician even though legally they are not required to do so. *See* Rosemarie DLC Bernabe et al., *The Fiduciary Obligation of the Physician-Researcher in Phase IV Trials*, 15 BMC MED. ETHICS, Feb. 7, 2014, at 2, 4–5 (arguing that final stage drug trials involving patients are "by nature and purpose closer to practice than the other phases of drug development" so that "physician-researchers are primarily physicians and secondarily researchers whose fiduciary obligation to their patient-participants remains").

involved.[315] Because there are so many hurdles to fully obtain informed consent when a patient is asked to participate in a research trial, the protection provided to prevent coercion or undue influence is higher than that for consent to medical care alone.[316]

D. Deep Fakes and AI Voices

Another feature of current AI technology that makes it incompatible with standards for informed consent to participate in research is that it operates in ways that exceed human detection.[317] A potential participant engaging in a conversation to obtain informed consent may be under the impression that they are conversing with someone they already trust.[318] Since the voice and image of a celebrity can be produced on demand, the technology would be able to take on the persona of a prominent scientist like Dr. Anthony Fauci or an actress like Scarlett Johansson.[319] Moreover,

315. *See* I. Glenn Cohen, *Informed Consent and Medical Artificial Intelligence: What to Tell the Patient?*, 108 GEO. L.J. 1425, 1466–67 (2020) ("Even after well-designed and well-intentioned efforts, when tested, most patients do not understand the information presented to them . . . .").

316. *See* Sara Manti & Amelia Licari, *How to Obtain Informed Consent for Research*, 14 BREATHE 145 (2018); *see also* Johan Bester, Cristie M. Cole & Eric Kodish, *The Limits of Informed Consent for an Overwhelmed Patient: Clinicians' Role in Protecting Patients and Preventing Overwhelm*, 18 AMA J. ETHICS 869 (2016) (arguing that for some kinds of medical care, such as genetic testing, it is impossible to get fully informed consent from an "overwhelmed" patient and therefore the physician should only be required to act in the patient's best interest contending that "in the context of whole genome sequencing, informed consent may be impossible, and a clinician needs to shift towards preventing harm").

317. *See* discussion *infra* Section VIII.D.5 (discussing the problem of black box algorithms in the context of AI).

318. Such is the case where reputable public figures perform advertisements of medical devices and treatments. For further background on the use of celebrities in prescription drug advertisements, see generally Katie Adams, *Why Big Pharma Uses Celebrity Spokespeople*, BECKER'S HEALTHCARE (Sept. 18, 2020), https://www.beckershospitalreview.com/digital-marketing/why-big-pharma-uses-celebrity-spokespeople.html [https://perma.cc/7EAQ-YPUP] (noting that this practice continues despite concerns that "[i]t's hard to imagine a setting in which a celebrity endorsement of a drug conveys any meaningful information to patients in terms of either efficacy or side effects"); *see also FDA to Consider Use of Celebrity Spokespeople in DTC Prescription Drug Ads*, KFF HEALTH NEWS (June 11, 2009), https://khn.org/morning-breakout/dr00033423/ [https://perma.cc/G6NN-HZG7].

319. *See generally* Kat Tenbarge, *Hundreds of Sexual Deepfake Ads Using Emma Watson's Face Ran on Facebook and Instagram in the Last Two Days*, NBC NEWS (Mar. 7, 2023, 12:10 AM), https://www.nbcnews.com/tech/social-media/emma-watson-deep-fake-scarlett-johansson-face-swap-app-rcna73624 [https://perma.cc/QL4R-KXR5] (discussing the ability of AI deepfakes to impersonate individuals); *see* also Michael J. Hoisington,

even if there is no effort to impersonate a particular individual, the interface could, in response to emotional clues from the subject, adopt new, manipulative characteristics.[320]

This is of particular concern because not only are these customized avatars undetectable to the subject, but there is, thus far, no fully adequate explanation describing or predicting how the AI will choose to respond.[321]

E. Tendency Towards Bias

AI technology has been associated with racially biased decision-making since its earliest days looking for nervous "terrorists" at airports.[322] Political philosopher Michael Sandel recently stated, "AI not only replicates human biases, it confers on these biases a kind of scientific credibility. It makes it seem like these predictions and judgments have an objective status."[323] Even worse, if a program was not already biased when released, it can become biased in real time. For example, Forbes reported that "an AI-based conversational chatbot on Twitter that was supposed to interact with people through tweets and direct messages . . . started replying with highly offensive and racist messages within a few hours of its release."[324] Although this may have carried over from its training on "anonymous public data," more relevant was its "built-in internal learning feature" that left it vulnerable "to a coordinated attack by a group of people to introduce racist

---

Celebrities Sue Over Unauthorized Use of Identity, HIGGS FLETCHER MACK (Aug. 20, 2022), https://higgslaw.com/celebrities-sue-over-unauthorized-use-of-identity/ [https://perma.cc/5EN4-AUBV].

320. *See The World of Deepfake Advertising is Coming This Decade*, TECH XPLORE (Nov. 10, 2022), https://techxplore.com/news/2022-11-world-deepfake-advertising-decade.html [https://perma.cc/VX2B-P9LY]; *see also* Kalhan Rosenblatt, *Character.ai Could Change How Stans Engage in Fan Fiction*, NBC NEWS (Mar. 16, 2023, 1:41 PM), https://www.nbcnews.com/tech/characterai-stans-fan-fiction-rcna74715 [https://perma.cc/2ELK-7WHL] ("A website is letting fans around the world have one-on-one conversations with their favorite celebrities, icons and personalities. The only catch? The conversations are with a machine, not a person.").

321. *See* Yavar Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, 31 HARV. J.L. & TECH. 890, 891–92 (2018) ("It may be impossible to tell how an AI that has internalized massive amounts of data is making its decisions.").

322. *See* Alex Najibi, *Racial Discrimination in Face Recognition Technology*, HARV. UNIV.: SCI. IN THE NEWS (Oct. 24, 2020), https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/ [https://perma.cc/RQ6P-66ZT].

323. Pazzanese, *supra* note 15 ("'Part of the appeal of algorithmic decision-making is that it seems to offer an objective way of overcoming human subjectivity, bias, and prejudice' . . . . 'But we are discovering that many of the algorithms that decide who should get parole, for example, or who should be presented with employment opportunities or housing . . . replicate and embed the biases that already exist in our society.'").

324. Steve Nouri, *The Role of Bias in Artificial Intelligence*, FORBES (Feb. 4, 2021, 8:00 AM), https://www.forbes.com/sites/forbestechcouncil/2021/02/04/the-role-of-bias-in-artificial-intelligence/ [https://perma.cc/YVB9-LB6L].

bias in the system."[325] This tendency to reproduce society's biases and prejudices have followed AI technology wherever it has been introduced.

F. Military Misgivings

1. Developing Trend Military Misgivings Over Persuasive AI

The sharpest warnings about the danger of using AI technology to assist decision-making have come from the industry with the most experience in using it: the U.S. military.[326] By its own account, the military has invested heavily in AI that detects and responds to the emotions of both its own personnel and the "enemy."[327] This is no secret given most of the information available about the military's use of AI comes from military sources themselves.[328]

While Ash Carter's concerns are related to the harm resulting from a wrong decision, many critics of the military's use of AI are equally concerned about issues of accountability.[329] This is understandably the case given the potential for moral injury as AI technology interacts with soldiers as they make decisions about whether to use lethal weapons to kill people.[330]

On reviewing one program that guided a simulated aerial dog fight, ethicist Peter Singer noted that "'the AI shifted [its tactics] and it kept grinding away in different ways at him' until it won."[331] In other words, it was

325. *Id.*

326. For a discussion of how the military uses AI assisted decision-making, see Jon M. Garon, *When AI Goes to War: Corporate Accountability for Virtual Mass Disinformation, Algorithmic Atrocities, and Synthetic Propaganda*, 49 N. KY. L. REV. 181 (2022).

327. *See* Todd South, *Future Robot Battle Buddies May Read Your Emotions to Fight Better*, ARMY TIMES (Nov. 9, 2020), https://www.armytimes.com/news/your-army/2020/11/09/future-robot-battle-buddies-may-read-your-emotions-to-fight-better/ [https://perma.cc/RTW7-25YF] (noting that this technology could not just support but perhaps eventually supplant decision-making in battle).

328. *See, e.g.*, U.S. Army CCDC Army Rsch. Lab'y Pub. Affs., *Future Autonomous Machines May Build Trust Through Emotion*, U.S. ARMY (Sept. 15, 2020), https://www.army.mil/article/239052/future_autonomous_machines_may_build_trust_through_emotion [https://perma.cc/VZ5P-HCEX].

329. For a brief summary of Carter's concerns, see *supra* notes 266–74 and accompanying text.

330. *See* Jon Harper, *Pentagon Grappling with AI's Ethical Challenges*, NAT'L DEF. MAG. (Nov. 10, 2020), https://www.nationaldefensemagazine.org/articles/2020/11/10/pentagon-grappling-with-ais-ethical-challenges [https://perma.cc/ZML3-W9AG].

331. *Id.*

making tactical decisions that reflect the technology's trend towards being "increasingly intelligent, ever-changing and increasingly autonomous, doing more and more on its own."[332] Consequently, "we have two kinds of legal and ethical questions that we've really never wrestled with before. The first is machine permissibility. What is the tool allowed to do on its own? The second is machine accountability. Who takes responsibility . . . for what the tool does on its own?"[333] Recognizing the significance of these issues, the military has taken on the problem in-house, setting up special divisions to coordinate AI use across service branches.[334]

This is particularly important because, in a weapons system, the programming could overcome what might be an individual human's reluctance to launch a weapon that will inflict considerable collateral damage beyond its intended target.[335] Most relevant of issues to informed consent is the integration of AI systems into combat via a spearhead initiative called "Project Maven . . . that . . . used AI algorithms to identify insurgent targets in Iraq and Syria."[336] More recently, this form of AI demonstrated it can rapidly absorb and analyze information in ways that enhance the ability of human soldiers.[337] The military claims that AI can do more than analyze information—it can predict the outcome of different scenarios and generate recommendations to decision-making humans as they deploy weapons.[338]

In a health care setting, this kind of programming could similarly cause harm by recommending reduced doses of pain medication based on pervasive beliefs that, for example, Black patients feel pain less acutely.[339] Professionals who hear such a recommendation from machines advertised

---

332. *Id.*

333. *Id.*

334. *See* Jaspreet Gill, *Say Goodbye to JAIC and DDS, as Offices Cease to Exist as Independent Bodies June 1*, BREAKING DEF. (May 24, 2022, 3:30 PM), https://breakingdefense.com/2022/05/say-goodbye-to-jaic-and-dds-as-offices-cease-to-exist-as-independent-bodies-june-1/ [https://perma.cc/TU2Q-EL8F] (reporting on the consolidation of AI military departments into a single Office of the Chief Digital and Artificial Intelligence Officer).

335. *See* Harper, *supra* note 330.

336. For a report as of 2020 about the use of AI in the military and other security services, see KELLEY M. SAYLER, CONG. RSCH. SERV., R45178, ARTIFICIAL INTELLIGENCE AND NATIONAL SECURITY 9–15 (2020), https://sgp.fas.org/crs/natsec/R45178.pdf [https://perma.cc/95RS-N3VR].

337. MARGARITA KONAEV ET AL., U.S. MILITARY INVESTMENTS IN AUTONOMY AND AI, at 1 (2020), https://cset.georgetown.edu/publication/u-s-military-investments-in-autonomy-and-ai-costs-benefits-and-strategic-effects/ [https://perma.cc/RTP6-859X] ("Ultimately, at a strategic level, investments in autonomy and AI are meant to provide the U.S. military with the AI-enabled capabilities needed to deter adversaries from aggression, fight and win the wars of the future, and cooperate effectively with allies.").

338. *See id.* at 27 ("AI could potentially predict when individuals may become too stressed or suffer psychological or physical injuries, notifying commanders and allowing them to make appropriate adjustments.").

339. *See* Swetlitz, *supra* note 157; *see also* Trawalter, *supra* note 157.

as being "smarter" than any individual human may set aside their own judgement and instead defer to the technology's recommendation.[340]

New forms of interactive and adaptive AI are rapidly developing.[341] The recent call by tech leaders to pause all future research into AI with the possibility of superseding human decision-making echoes long standing warnings about the likelihood of rapid advances in the field. For example, Arati Prabhakar, director of the White House's Office of Science and Technology Policy, says she is excited about the possibilities of AI, but she also warned that "[w]hat we are all seeing is the emergence of this extremely powerful technology" that "[a]ll of history shows . . . can and will be used for good and for ill."[342]

VI. HOW PERSUASIVE AI UNDERMINES INFORMED CONSENT

The threat Persuasive AI poses to informed consent for participation in a clinical drug trial is not based solely on digital technology.[343] The more serious threat comes from the technology's characteristics that undermine federal legal protections by seeking to exert influence over prospective participants in ways that are not well understood, not apparent to observers, and not subject to existing methods of mitigation.[344]

---

340. *See generally* ALEC TYSON ET AL., 60% OF AMERICANS WOULD BE UNCOMFORTABLE WITH PROVIDER RELYING ON AI IN THEIR OWN HEALTH CARE 4 (2023), https://www.pewresearch.org/science/2023/02/22/60-of-americans-would-be-uncomfortable-with-provider-relying-on-ai-in-their-own-health-care/ [https://perma.cc/3WYS-URCH] ("Six-in-ten U.S. adults say they would feel uncomfortable if their own health care provider relied on artificial intelligence to do things like diagnose disease and recommend treatments; a significantly smaller share (39%) say they would feel comfortable with this.").

341. *See, e.g.*, Samantha M. Kelly, *5 Jaw-Dropping Things GPT-4 Can Do That ChatGPT Couldn't*, CNN (Mar. 16, 2023, 10:27 AM), https://www.cnn.com/2023/03/16/tech/gpt-4-use-cases/index.html [https://perma.cc/68VX-RRR6] (showing one set of recent advances in AI technology).

342. Anthony Zurcher, *AI: How 'Freaked Out' Should We Be?*, BBC (Mar. 16, 2023), https://www.bbc.com/news/world-us-canada-64967627 [https://perma.cc/RRA2-ZQ4L].

343. Christine Grady et al., *Informed Consent*, 376 NEW ENG. J. MEDICINE 856, 858 (2017) ("Information technologies enable new ways of presenting information . . . yet they do not resolve questions related to the necessity or adequacy of informed consent.").

344. For purposes of considering the threat to informed consent to participate in research, there are, where possible, consistent descriptions of the various forms of AI and assume for the sake of discussion that the claims made for the technology are true. Because AI technology has not evolved in a linear fashion, one version does not necessarily replace another; this Part identifies the different ways that AI claims to influence decision-making with a particular focus on the most recent versions that could be integrated into real-time consent conversations.

Whether this influence is an intentional effort to undermine the will of participants, a manipulation, or an advanced form of persuasion, this influence causes irreparable harm to the integrity of the consent process.[345] The prima facie informed consent violation in research is an event that interferes with the process itself. In contrast, the failure to provide even crucial information is not enough to make a prima facie case for negligence unless it results in harm to the patient.[346] The protection that federal law provides is not an assurance that people will only be allowed to enroll in trials against their best interests, rather it is that they be allowed to make that decision themselves.

There is therefore a sufficiently strong basis for banning this use of AI even before knowing how AI is exerting this influence or even documenting that the influence has occurred. This is because federal law drafted to protect the decision-making process prioritizes the autonomy of the individual making the choice over any other interest.[347]

The WHO guidance acknowledged exactly this issue. While emphasizing the need for patient consent to collect data, the WHO guidance notes that "even informed consent may be insufficient to compensate for the power dissymmetry between the collectors of data and the individuals who are the sources."[348] Moreover, "true informed consent is increasingly infeasible in an era of biomedical big data" because the "scale and complexity of biomedical big data make it impossible to keep track of and make meaningful decisions about all uses of personal data."[349]

The guidance related to AI use in decision-making is even more disheartening, noting that "[m]ost patients have insufficient knowledge about how and why AI technologies make certain decisions" so that in "some situations, individuals may feel unable to refuse treatment" because they

---

345. This protection of the process of informed consent rather than protection from the consequences of failing to provide consent is what distinguishes research from medicine.

346. *See* Edward L. Raab, *The Parameters of Informed Consent*, 102 TRANS. AM. OPHTHALMOLOGY SOC'Y 225, 226 (2004) (citing N.Y. PUB. HEALTH LAW § 2805-d(2)(a), § 2805–d(3) (2014)) ("A claim of lack of informed consent . . . differs importantly from malpractice in not requiring that the treatment be a departure from the standard of care. The elements of the claim are (1) the physician did not present the risks and benefits of the proposed treatment and of alternative treatments; (2) with full information, the patient would have declined the treatment; and (3) the treatment, *even though appropriate and carried out skillfully*, was a substantial factor causing the patient's injuries.").

347. *See* Lauren Burkhart, Comment, *You Can't Handle the Truth: Rationally Limiting the Duty to Disclose Genetic Information*, 54 JURIMETRICS J. 85, 96 (2013) ("Inevitably, as health-care practice shifts away from paternalism and toward patient control, the potential benefits of authoritative or limiting legal safeguards and regulatory actions are considered second to the priority of individual autonomy in decision making.").

348. WHO GUIDANCE, *supra* note 250, at 39.

349. *Id.* at 40.

may believe "that the 'computer knows best.'"[350] Use of AI in medicine, it warns, "could challenge the core of informed consent and wider public trust in health care."[351]

In the absence of any independent or objective information about how Persuasive AI works or what it can do, we must go forward based on the claims it makes for itself.

The great challenge in mitigating the potential for harm from AI assisted decision-making is that "we know of only a single form of high intelligence—our own" and therefore "know eerily little" about how AI reaches its conclusions.[352] Therefore, while we can identify undesirable results, such as a decision based on racial bias, it is so far impossible to create safeguards against such a result.[353]

A. AI Is Already Deployed in Health Care in the Informed Consent Process

Because participants in the later stages of clinical drug trials are also patients, it is important to distinguish between the process of obtaining informed consent for research and informed consent for medical treatment.[354] Before a new drug can be approved for sale in the United States, it must be tested on people who are already being treated for the condition. These individuals are therefore already part of a health care system in which AI is inextricably integrated.[355] A Westlaw Precision search from January 1,

350. *Id.* at 47.

351. *Id.*

352. Stephen E. Henderson, *Should Robots Prosecute and Defend?*, 72 OKLA. L. REV. 1, 9 (2019) (arguing against delegating the decision of whether to prosecute to AI technology).

353. For examples of how proprietary AI software can be tested for racial bias, see Zhang, *supra* note 125 ("We are taking immediate action to prevent this type of result from appearing. There is still clearly a lot of work to do with automatic image labeling, and we're looking at how we can prevent these types of mistakes from happening in the future."); Pete Pachal, Google Photos Identified Two Black People as 'Gorillas,' MASHABLE (July 1, 2015), https://mashable.com/2015/07/01/google-photos-black-people-gorillas [https://perma.cc/5RGQ-K5NB] (similar).

354. *See* Paasche-Orlow, Taylor & Brancati, *supra* note 221 ("[I]nformed consent for participation in medical research is particularly challenging because it requires a level of comprehension beyond that required for consent to usual care.").

355. For a history of the use of AI in health care, see Sarah Kamensky, Note, *Artificial Intelligence and Technology in Health Care: Overview and Possible Legal Implications*, 21 DEPAUL J. HEALTH CARE L., May 2020, at 1, 4–6. For a timeline of the use of AI in health care, see *AI's Ascendance in Medicine: A Timeline*, CEDARS SINAI (Apr.

1985, to January 1, 2023, of "health & care" & "artificial intelligence" in "Law Reviews and Journals" identified 9,227 separate articles.[356] As early as 1986, attorney–ethicist Dr. Haavi Morreim complained that "[t]wo decades of work on artificial intelligence in medical diagnosis have resulted in little success in developing a system capable of performing diagnostic tasks adequately."[357] By 2001, Professor Nicholas Terry was concerned about what are still the ongoing privacy implications of AI systems collecting identifiable health information.[358] But the discussion also quickly turned to potential risks.[359] While at the beginning of that time period the AI discussed was far closer to what we would see as basic record-keeping tasks, commentators expressed concerns about the potential for privacy violations[360]

20, 2023), https://www.cedars-sinai.org/discoveries/ai-ascendance-in-medicine.html [https://perma.cc/TE5Z-DN2N]. For a helpful overview of the current use of AI technology in direct patient care, see Samuel D. Hodge, Jr., *The Medical and Legal Implications of Artificial Intelligence in Health Care - An Area of Unsettled Law*, 28 RICH. J.L. & TECH. 405, 413–16 (2022).

356. *See Secondary Sources*, WESTLAW PRECISION, https://1.next.westlaw.com/Search/Results.html?query=health%20%26%20care%20%26%20%22artificial%20intelligence%22&isPremiumAdvanceSearch=false&jurisdiction=ALLCASES&contentType=ANALYTICAL&querySubmissionGuid=i0ad740110000018bea5d49ca1e93124d&categoryPageUrl=Home%2FSecondarySources%2FSecondarySourcesLibrary&searchId=i0ad740110000018bea5cbbdfd53a4677&transitionType=ListViewType&contextData=(sc.Search) (last visited Nov. 19, 2023).

357. Mark A. Hall, *Institutional Control of Physician Behavior: Legal Barriers to Health Care Cost Containment*, 137 U. PA. L. REV. 431, 476 n.159 (1989) (citing E. Haavi Morreim, *Clinicians or Committees–Who Should Cut Costs*, 17 HASTINGS CTR. REP., Apr. 1987, at 45, 45) ("Existing programs 'are virtually unable to cope with variations in the clinical picture' such as 'the evolution of a disease over time . . . how one disease may influence the presentation of a second, or how the effects of previous treatment can alter the patient's illness.'" (quoting William B. Schwartz, Ramesh S. Patil & Peter Szolovits, *Artificial Intelligence in Medicine: Where Do We Stand?*, 316 NEW ENG. J. MEDICINE 685, 688 (1987))).

358. *See* Nicolas P. Terry, *An eHealth Diptych: The Impact of Privacy Regulation on Medical Error and Malpractice Litigation*, 27 AM. J.L. & MED. 361, 361–62 (2001) ("[T]he forces driving increased privacy and reduced medical error are closely related; . . . they find common ground in process re-engineering and the adoption of technologies that conceptually, architecturally and operationally will intersect and frequently combine.").

359. *See* Hodge, Jr., *supra* note 355, at 408.

360. *See* Hoffman & Podgurski, *supra* note 143, at 34–35 (discussing the necessity to ensure that automated decision makers are vetted for problems related to "accuracy, fairness, bias, discrimination, privacy or security").

and for bias.[361] In 2010, a commentator predicted that AI "will likely play a significant role in future United States health care."[362]

This prediction quickly became reality. By 2015, AI had gone beyond a source of information and record keeping, and it was already in such wide use as a diagnostic tool that scholars were concerned about "the use of opaque computational models to make decisions related to health care," noting that its use "raises significant privacy concerns" as well as "the potential for discrimination in multiple contexts."[363]

Today, this concern has materialized.[364] Studies of treatment recommendations show that some patients may receive less intensive care because of their demographic characteristics rather than because of their medical needs.[365] This bias has attracted the attention of Congress

361. *See* Jenna Jonjua, Note, *Racist Robots? The Future of Title VII Disparate Impact Cases in the World of Artificial Intelligence*, 30 MINN. J. INT'L L. 329, 330 (2021) ("AI algorithms are not immune to bias."). Many of the laws specific to AI decision-making address concerns about the introduction of racial and other biases as a factor in decision-making.

362. Michael S. Young, *Artificial Intelligence, Telemedicine, and Robotics in Health Care*, 6 ABA SCITECH LAW., no. 4, 2010, at 14, 14.

363. W. Nicholson Price II, *Black-Box Medicine*, 28 HARV. J.L. & TECH. 419, 421, 454, 455 (2015) (coining the term "Black-box Medicine" to describe "the use of opaque computational models to make decisions related to health care"); *see also* Kamensky, *supra* note 355, at 2 (citing *Artificial Intelligence in Medicine*, MENDELEY CAREERS (Apr. 17, 2018), https://www.mendeley.com/careers/news/careers-jobs-field/artificial-intelligence-medicine [https://perma.cc/GYL5-JXFC]) ("In a health care setting, artificial intelligence refers to the use of artificial intelligence technology and automated processes to diagnose and treat patients who require care.").

364. *See* Frank Griffin, *Artificial Intelligence and Liability in Health Care*, 31 HEALTH MATRIX: J. LAW-MEDICINE 65, 69 (2021) (quoting Jessica Kent, *One-Third of Orgs Use Artificial Intelligence in Medical Imaging*, HEALTH IT ANALYTICS (Jan. 28, 2020), https://healthitanalytics.com/news/one-third-of-orgs-use-artificial-intelligence-in-medical-imaging [https://perma.cc/8DED-SQSH]) ("[O]ne-third of hospitals and imaging centers report using artificial intelligence . . . to aid tasks associated with patient care imaging or business operations."); *id.* at 81–82 (discussing various ways in which bias can be introduced by AI to the detriment of patients within the healthcare system).

365. *See, e.g.*, Chris Giordano et al., *Accessing Artificial Intelligence for Clinical Decision-Making*, FRONTIERS DIGIT. HEALTH, June 2021, at 1, 2, 5–6 ("AI can aide physicians in the complex task of risk stratifying patients for interventions, identifying those most at risk of imminent decompensation, and evaluating multiple small outcomes to optimize overall patient outcomes."); Griffin, *supra* note 364, at 82 (citing Tom Lawry et al., *Realizing the Potential for AI in Precision Health*, 13 THE SCITECH LAW. 22, 24 (2017)) ("Al biases can result from 'under-representation' in datasets of some populations that may 'hide population differences in disease risk or treatment efficacy.' In one example, researchers 'found that cardiomyopathy genetic tests were better able to identify pathogenic variants in white patients than patients of other ethnicities.'").

and resulted in proposed legislation seeking to limit AI's use in medical decision-making.[366]

Moreover, AI is not only being used to assist in treatment decisions, but it has also become part of the informed consent process. Online consent is already used widely in patient care and in low risk human subject research.[367] And what AI offers is even more than record-keeping.[368] As early as 2016, Brandon M. Welch, a self-described medical technologist, published an article titled, *Teleconsent: A Novel Approach to Obtain Informed Consent for Research*, where he explains that "Teleconsent" was designed to "to remotely replicate the capabilities of an in-person consent process."[369] The article explained that their product was able to leverage existing telehealth technology to "observe non-verbal cues and address any ambiguities that the participant has during the consent process" as well as "support effective person-to-person interaction."[370] By all indications, online consent for research had caught on and, well before the pandemic, become very popular.[371] However, even well before any concern about AI manipulating the process, ethicists worried that conducting informed consent remotely might disadvantage participants who were less familiar or comfortable with using technology.[372]

Some have argued that switching from in-person to Teleconsent improves the informed consent process because it "allows participants to complete

366. *See Rep. Clark, Sen. Wyden Lead Letter to Prevent AI Bias in COVID-19 Response*, U.S. HOUSE OF REPRESENTATIVES (May 12, 2020), https://clarke.house.gov/aibiasletter/ [https://perma.cc/U2F7-SSBE].

367. *See Consent Processes and Documentation*, INST. REV. BD., https://irb.wisc.edu/manual/investigator-manual/conducting-human-participant-research/consent-processes-and-documentation/ [https://perma.cc/PE65-3U9Z] ("'Digital signature' methodologies, if used entirely remotely, are generally approved only for low risk research or other circumstances . . . .").

368. *See id.*

369. Brandon M. Welch et al., *Teleconsent: A Novel Approach to Obtain Informed Consent for Research*, 3 CONTEMP. CLINICAL TRIALS COMMC'NS 74, 75 (2016).

370. *Id.* at 78.

371. *See* Grady et al., *supra* note 343 (discussing best practices for obtaining informed consent through websites or other electronic means); John Wilbanks, *Design Issues in E-Consent*, 46 J. L. MED. ETHICS 110 (2018); *see also* Mohd Yusmiaidil Putera Mohd Yusof, Chin Hai Teo & Chirk Jenn Ng, *Electronic Informed Consent Criteria for Research Ethics Review: A Scoping Review*, 23 BMC MED. ETHICS, Nov. 21, 2022, at 1, 2 ("With the growing use of digital health tools, electronic consent (eConsent) has become a crucial element in health research and standard clinical care, especially during and post-Covid-19 pandemic.").

372. *See id.* at 9 ("Lack of access to and knowledge with the technology, on the other hand, can pose additional barriers to consent. As technology improves, the number of participants may grow as they are no longer restricted by their geographic closeness to the research facility. However, until that time comes, participation may skew toward younger, more affluent people who already have access.").

the informed consent process from the comfort of their homes or a private environment without having to visit the research institution or having a research team member go out to their home and collect in-person consent."[373]

B. AI Is Already a Part of Clinical Research

Sponsors of clinical drug trials are already making extensive use of AI technology at many stages in the process of identifying, attracting, and enrolling subjects.[374] While there are no central registries of AI use and no obligation to self-report, public information and marketing from AI companies suggests that these uses include identifying eligible participants, developing effective advertising, and managing the informed consent process.[375] This last activity covers a broad range of tasks from record-keeping to communication. There are, so far, no direct accounts of the integration of Persuasive AI in the informed consent conversation itself,[376] but, as discussed below, there is considerable reason to believe if it is not already happening, its use is imminent.[377] Many large sponsors start identifying potential participants before seeking approval for a specific trial.

C. Accessing Medical Records

One way they can do that is by directly accessing their medical records. The HIPAA privacy rule, which sets the terms for how covered entities like doctors and hospitals share patient records containing protected health information (PHI) records, contains a "Preparatory to Research" provision allowing covered entities to disclose patient records to researchers so that

373. Cristina M. López et al., *Technology as a Means to Address Disparities in Mental Health Research: A Guide to "Tele-Tailoring" Your Research Methods*, 49 PRO. PSYCH.: RSCH. & PRAC. 57, 60 (2018).

374. For examples of companies marketing their services to improve advertising of clinical trials, see *Activate Patients. Inform Decisions. Improve Outcomes.*, PHREESIA LIFE SCIS., https://lifesciences.phreesia.com [https://perma.cc/C26M-ZW33].

375. *See* López et al., *supra* note 373, at 59–62.

376. *See* Shlomo A. Koyfman et al., *Informed Consent Conversations and Documents: A Quantitative Comparison*, 122 CANCER 464, 464 (2016) ("Readability software has been used to help simplify the language of [informed consent documents], but to the authors' knowledge is rarely used to assess the language used during the [informed consent conversations], which may influence the quality of informed consent.").

377. *See* Cohen, *supra* note 315, at 1427 ("Many companies and healthcare providers are currently investing heavily in developing medical AI/ML systems to including AI-driven X-ray image and analysis systems and AI-driven monitoring systems used to identify elderly patients at risk of falling.").

they can be evaluated and contacted as potential participants.[378] Health care providers can do this without getting consent or even informing them this has happened.[379]

D. Building Alliances with Patient Groups

While some recruiting for clinical trials comes from advertisements or referrals from physicians, a lot of sponsors today reach out to participants directly. In 2021, OHRP issued a report by its ethics advisory committee, addressing the "New Challenges" posed by sponsors developing ongoing relationships with potentially eligible participants before seeking approval to launch a new trial.[380] The report speaks particularly to situations where a sponsor is considering a trial requiring "patients diagnosed with rare diseases."[381] But, as discussed below, these are not the only groups of potential participants.

---

378. *See* 45 C.F.R. § 164.512(1)(i), (ii) (2016). Under the "[r]eviews preparatory to research" provision, covered entities may use or disclose PHI to researchers to aid in study recruitment. The covered entity may allow a researcher, either within or outside the covered entity, to identify, but not contact, potential study participants under said provision. However, before permitting this activity, a covered entity must receive proper representation, as described above, from the researcher. Under the "preparatory to research" provision, no PHI may leave the covered entity. *See id.* § 164.512(1)(ii)–(iii).

379. *HIPAA Privacy Rule and its Impacts on Research*, NAT'L INSTS. OF HEALTH, https://privacyruleandresearch.nih.gov/faq.asp [https://perma.cc/A9ES-VKT3] ("A covered entity may use or disclose protected health information without individuals' authorizations for the creation of a research database, provided the covered entity obtains documentation that an IRB or Privacy Board has determined that the specified waiver criteria were satisfied. Protected health information maintained by a covered entity in such a research database could be used or disclosed for future research studies as permitted by the Privacy Rule - that is, for future studies in which individual authorization has been obtained or where the Rule would permit research without an authorization, such as pursuant to an IRB or Privacy Board waiver.").

380. Off. for Hum. Rsch. Prots., *Attachment B - New Challenges in Interactions Among Sponsors, Clinical Trial Sites, and Study Subjects*, U.S. DEP'T HEALTH & HUM. SERV., https://www.hhs.gov/ohrp/sachrp-committee/recommendations/attachment-b-new-challenges-sponsor-clinical-trial-site-subject.html [perma.cc/FB3T-3GU3] ("Yet the growing interaction between sponsors and patient populations has begun to blur significantly the traditional division of roles between sponsors and investigators, giving rise to complex ethical issues that, as of this writing, exist in a relative vacuum of guidance on how to navigate industry's increasing involvement in activities traditionally reserved for investigators and site study teams, leading to practices that vary widely across sponsors.").

381. *Id.* ("Increasingly, industry and academic focus has been drawn to the development of treatments for rare diseases and pediatric conditions, in part due to the incentives that manufacturers have under the U.S. Orphan Drug Act and the Best Pharmaceuticals for Children Act, such as market exclusivity.").

E. Advertising to Attract Participants for Specific Trials

Federal law requires that all information provided by researchers to potential participants about the study must be approved in advance by the IRB.[382] This includes any marketing or advertising material whether targeted to specific participants or distributed more broadly.[383] While this material need not have everything usually required with informed consent, it also cannot be misleading or make promises about the benefits of the trial to any individual patient.[384]

F. Designing Informed Consent Documents

The Revised Common Rule requires that before a prospective participant can be enrolled in a clinical drug trial, they must be provided with clear documentation that provides the information a "reasonable person" would want to know before deciding to participate.[385] The IRB has the task of determining, in advance, whether the documentation meets this standard.[386]

G. The Informed Consent Dialogue

Whether a prospective participant responds to an advertisement or is solicited directly, the next stage in informed consent is an individual conversation with a researcher. AI could become part of that conversation in two ways.[387] First, as an advisor or coach to the researcher while they

382. For an example of guidance on how to obtain informed consent that incorporates current standards, see Manti & Licari, *supra* note 316; *see also* INT'L COUNCIL FOR HARMONISATION OF TECH. REQUIREMENTS FOR PHARMACEUTICALS FOR HUM. USE, INTEGRATED ADDENDUM TO E6(R1): GUIDELINE FOR GOOD CLINICAL PRACTICE 5 (2016), https://database.ich.org/sites/default/files/E6_R2_Addendum.pdf [https://perma.cc/MS36-EKUW] ("[A] subject voluntarily confirms his or her willingness to participate in a particular trial[] after having been informed of all aspects of the trial that are relevant to the subject's decision to participate.").

383. *See* Manti & Licari, *supra* note 316, at 146.

384. Off. for Hum. Rsch. Prots., *supra* note 380 (discussing how appropriate internal policies would avoid misleading communication and the making of promises or representations that cannot be honored).

385. 45 C.F.R. § 46.116(a)(4) (2018).

386. For further explanation of the IRB's powers of preapproval, see generally Off. for Hum. Rsch. Prots., *Informed Consent FAQs*, U.S. DEP'T OF HEALTH & HUM. SERVS., https://www.hhs.gov/ohrp/regulations-and-policy/guidance/faq/informed-consent/index.html [https://perma.cc/3EX7-2QC6].

387. AI's use in providing mental health services provides examples of the role it could play in an informed consent conversation. *See* Alex Potamianos & Shri Narayanan,

are speaking with the potential participant.[388] This would be similar to a car salesman who, mid-negotiation, excuses herself to "talk with the manager."[389] We might also imagine, for example, the advantage a professional tennis player would have with AI programmed to predict the opponent's next shot.

The other more concerning interaction is one where the participant engages only with the AI either in the form of a robot, chatbot, or online avatar.[390] It is easy to frame this kind of assistance as beneficial because it alerts the researcher to concerns or questions that the potential subject has not expressed. And that may be true when this technology is used to obtain informed consent for medical care. But, here the goal is not just to inform, but to persuade. And giving one party to the conversation a sideline coach monitoring the reactions of the potential participant and advising the researcher on how to be more persuasive goes beyond what should be considered within the acceptable limits.[391]

Indeed, concerns have mounted as the risks are better understood.[392] Like other forms of AI, the inherent inscrutability of Persuasive AI's

---

Why Emotion AI is the Key to Mental Health Treatment, TRANSFORMING DATA WITH INTEL. (Apr. 7, 2020), https://tdwi.org/articles/2020/04/07/adv-all-why-emotion-ai-key-to-mental-health-treatment.aspx [https://perma.cc/5VK5-VEKX] (detailing ways AI can be integrated into treatment, including direct therapy provided through a chat box). For a discussion on the dangers of AI treating mental illness, see Milady Nazir, *Researcher Warns About Dangers of AI Chatbots for Treating Mental Illness*, UTSA TODAY (July 8, 2020), https://www.utsa.edu/today/2020/07/story/chatbots-artificial-intelligence.html [https://perma.cc/A5PH-CRAL].

388. For a discussion of how AI can be used to cheat, see Singh Rahul Sunilkumar, *Explained: How Artificial Intelligence Can Be Misused to Cheat in Chess Games*, HINDUSTAN TIMES (October 8, 2022, 2:29 PM), https://www.hindustantimes.com/technology/explained-how-artificial-intelligence-can-be-used-to-cheat-in-chess-games-101665217813692.html [https://perma.cc/GDF9-5KEM]; Will Knight, *Cheaters Hacked an AI Bot—and Beat the Rocket League Elite*, WIRED (Jan. 19, 2023, 7:00 AM), https://www.wired.com/story/cheaters-hacked-an-ai-bot-and-beat-the-rocket-league-elite/ [https://perma.cc/79UR-JSZK].

389. *See* Chuck Gallagher, *Automotive Ethics: Most of Us (Still) Don't Trust Car Dealers*, CHUCK GALLAGHER (Sept. 1, 2016), https://www.chuckgallagher.com/2016/09/01/automotive-ethics-most-of-us-still-dont-trust-car-dealers/ [https://perma.cc/5QS8-JTYE].

390. *See The Power of Chatbots Explained*, EXPERT.AI (Mar. 24, 2022), https://www.expert.ai/blog/chatbot/ [https://perma.cc/9US9-CXJR] (noting that AI chatbots are already integrated into websites, mobile apps, messaging services, and virtual assistants where they can provide customer service, information retrieval, personal assistance, and entertainment).

391. A "chatbox" is a program that uses natural language processing (nlp) and machine learning techniques to understand and respond to user input, allowing users to interact with the chatbot in a conversational manner. *See* Gabe Fennema, *A Beginner's Guide to Chatboxes*, CAPACITY (Oct. 18, 2022), https://capacity.com/learn/chatbox/ [https://perma.cc/5EN8-WAFN].

392. *See, e.g.*, Alexandrine Royer, *The Wellness Industry's Risky Embrace of AI-Driven Mental Health Care*, BROOKINGS (Oct. 14, 2021), https://www.brookings.edu/

decisions and the lack of susceptibility to oversight in its interactions makes it impossible to design a consent process that adequately informs patients of the potential risks.[393]

The process of obtaining consent for participation in a clinical drug trial extends far beyond getting a patient's signature on a document.[394] Rather, it includes two components: informed consent documents (ICD) and informed consent conversations (ICC).[395] Whether the potential participant engages in a dialogue with a researcher who is being advised by AI or with a computer interface that simulates a dialogue chatbot, they may be facing a level of manipulation against which they cannot immunize themselves.

H. Inability to Provide Sufficient Information in Advance of Consent

AI and informed consent scholars have long been concerned that the inability to predict how AI will use the data it collects makes it incredibly difficult to inform or assure a participant in a research study how their data will be used and who will see it.[396] It is nearly impossible to give assurance to a potential subject.[397] Due to the absence of any possible promise of confidentiality, Professor Froomkin proposed that rather than make any binding assurances, researchers could only promise to use "best

techstream/the-wellness-industrys-risky-embrace-of-ai-driven-mental-health-care/ [https://perma.cc/S3K2-AJMN].

393. *See* Andrew D. Selbst, *An Institutional View of Algorithmic Impact Assessments*, 35 HARV. J.L. & TECH. 117, 120 (2021) (noting the possibility that AI may formulate "harmful outcomes" as a result of "difficulties of oversight stemming from a lack of transparency").

394. *See* Ezekiel J. Emanuel & Connor W. Boyle, *Assessment of Length and Readability of Informed Consent Documents for COVID-19 Vaccine Trials*, JAMA NETWORK OPEN, Apr. 28, 2021, at 1, 1 ("Informed consent documents are one, albeit critical, element in a process that is meant to ensure participant understanding and voluntary participation.").

395. Koyfman et al., *supra* note 376.

396. For a discussion of how today's AI capacity to identify tissue samples through genetic analysis as make assurances of confidentiality impossible, see A. Michael Froomkin, *Big Data: Destroyer of Informed Consent*, 21 YALE J.L. & TECH. 27, 32 (2019). Note that "[t]he mere 'removal of identifiers' is not nearly enough to prevent re-identification (and what exactly constitutes de-identification is itself far more complicated than it may seem," because, "Big Data analytics—the seeking of patterns and correlations in giant datasets—kills the possibility of true informed consent." *Id.* at 45, 32. It is impossible to know in advance what it will discover from the information it acquires. Thus, since "[i]nformed consent requires at the very least that the person requesting the consent know what she is asking the subject to consent to," this is impossible because "neither party to that conversation can know what the data may be used to discover." *Id.* at 32 (emphasis omitted).

397. *See id.* at 52–53 ("In the absence of a technical solution, we must confront the problem that Big- Data-based research undermines informed consent as we know it.").

available efforts" to protect the use of the data they collected.[398] Joined by other commentators, he also pointed out that the rapid development of genetic identification technology made it impossible to protect the identity of any individual from whom researchers collected a biological sample.[399] Calling the Common Rule the "gold standard" of informed consent, Professor Michael Froomkin has argued that any erosion of standards for protecting research participants would result in weaker protections for interactions with already lower protection across the board.[400]

1. Persuasive AI Creates Undue Influence Through Manipulation of Advertising Materials

The claim that AI can customize advertising in response to the individual reactions of targeted individuals is sometimes described as a "generative" feature because, as the EU explains, it "includes systems that generate 'content,' in addition to 'predictions, recommendations, or decisions.'"[401] When the content generated targets potential research subjects, this real-time customization cannot be approved in advance by an IRB.[402] One company summarizes what it offers advertisers with the phrase, "Understand Emotion, Understand Connection, Understand Your Audience."[403] Even if this customization does not generate "new material" in each individual encounter in the sense that it technically reflects only information acquired

398. *See id.* at 36 ("[D]espite its faults, at least in informed consent the party intending to get and use personal information makes a genuine effort to ensure that the person agreeing understands what they are getting into. That has, or should have value, even if the only thing it does is make the parties aware of what is at stake.").

399. For an overview of concerns from 2019, before the widespread adoption of Emotion AI, in biomedical research, see Alessandro Blasimme & Effy Vayena, *The Ethics of AI in Biomedical Research, Patient Care and Public Health*, *in* OXFORD HANDBOOK OF ETHICS OF AI 707 (Markus D. Dubber, Frank Pasquale & Sunit Das eds., 2020) (discussing the introduction of bias into AI recommendations because of the "quality and representativeness of data used to train machine learning algorithms" as well as the difficulty of obtaining informed consent for data collection when "[t]he infinite uses of data and the linkage of disparate data sets, makes even the notion of broad consent – a typical safeguard of autonomy when future uses of human data and samples are hard to anticipate – weak").

400. *See* Froomkin, *supra* note 396, at 27.

401. Matt O'Shaughnessy, *One of the Biggest Problems in Regulating AI is Agreeing on a Definition*, CARNEGIE ENDOWMENT FOR INT'L PEACE (Oct. 6, 2022), https://carnegieendowment.org/2022/10/06/one-of-biggest-problems-in-regulating-ai-is-agreeing-on-definition-pub-88100 [https://perma.cc/HJA6-PNAP].

402. For discussion on how the manipulation of emotions can affect the decision-making process and push one to action, see Peter N. Murray, *How Emotions Influence What We Buy*, PSYCHOLOGY TODAY (Feb. 26, 2013), https://www.psychologytoday.com/us/blog/inside-the-consumer-mind/201302/how-emotions-influence-what-we-buy [https://perma.cc/9XQJ-85JD].

403. *Audience Testing for Emotional Insights*, EMOTIONTRAC, https://creative.emotiontrac.com/ [https://perma.cc/YBP6-PUDZ].

from already existing sources, it is still an illegal effort to persuade.[404] Illustrating this slogan, the company gives the example of a personal injury law firm that, by analyzing the response of potential clients to their existing television advertising, "was able to successfully select a new 800 number and implement a new advertising campaign" that increased "leads" by "more than 1,000%."[405] This information, it explains, can be of use to politicians, film and TV, brands and ad agencies.[406]

VII. THE PRESSURE TO COMPLETE DRUG TRIALS MAKES SPONSORS EAGER TO ADOPT NEW TECHNOLOGY

Sponsors face intense financial pressure to successfully complete a clinical trial in order to obtain the data required for FDA approval to market their product in the United States.[407] They complain often (and loudly) that legal regulation makes the process "overly complex, inefficient and expensive."[408] It is therefore understandable that those seeking to get FDA marketing approval would take advantage of available technology that reduces costs and increases the likelihood that the trial will be completed.[409] There is considerable evidence that entities offering their services to assist sponsors to recruit and enroll eligible patients are already marketing AI as a cost-

404. *See* CompTIA, *How is Data Mining Used in Marketing*, COMPTIA https://www.comptia.org/content/articles/how-is-data-mining-used-in-marketing [https://perma.cc/5J6H-SS4T].

405. *See* EmotionTrac, *Case Studies*, EMOTIONTRAC, https://creative.emotiontrac.com/casestudies/ [https://perma.cc/78YD-DPTG].

406. *See Testimonials*, EMOTIONTRAC, https://creative.emotiontrac.com/#testimonials [https://perma.cc/6GCH-ZKZD].

407. *See* DAVID AUSTIN & TAMARA HAYFORD, CONG. BUDGET OFF., 57025, RESEARCH AND DEVELOPMENT IN THE PHARMACEUTICAL INDUSTRY 1 (2021) ("The pharmaceutical industry devoted $83 billion to R&D expenditures in 2019.").

408. Paul A. Monach & Westyn Branch-Elliman, *Reconsidering 'Minimal Risk' to Expand the Repertoire of Trials with Waiver of Informed Consent for Research*, 11 BMJ OPEN, Sept. 14, 2021, at 1, 1 (citing Andrew J. Vickers, *Clinical Trials in Crisis: Four Simple Methodologic Fixes*, 11 CLINICAL TRIALS 615, 615 (2014)); *see also* David M. Dilts et al., *Processes to Activate Phase III Clinical Trials in a Cooperative Oncology Group: The Case of Cancer and Leukemia Group B*, 24 J. CLINICAL ONCOLOGY 4553, 4556 (2006) (noting that a clinical trial can involve 370 different steps and a median time from proposal to activation of over 18 months).

409. *See* James A. Christensen & James P. Orlowski, *Bounty-Hunting and Finder's Fees*, 27 IRB: ETHICS & HUM. RSCH. 16, 16 (2005); *see also* Timothy Caulfield, *Legal and Ethical Issues Associated with Patient Recruitment in Clinical Trials: The Case of Competitive Enrolment*, 13 HEALTH L. REV. 58, 58 (2005) ("[P]atient recruitment has become an industry.").

effective solution.[410] This Part identifies the challenges sponsors face and how the use of AI as part of the process can result in a level of persuasion that, if employed by human recruiters, would be considered illegal persuasion techniques.[411] The ever-increasing financial pressures that sponsors and the investigators they work with face to complete successful clinical drug trials make them particularly receptive to offers of technology that will help achieve that goal.[412]

A. Enrollment Challenges in Drug Trials

Sponsors of clinical drug trials face significant challenges in recruiting participants for drug trials.[413] Another reason recruiting is difficult is that sponsors must identify medical eligibility requirements in advance that

410. *See, e.g.*, *Make Diversity in Your Clinical Trials a Reality*, CLARIFY HEALTH SOLS., https://clarifyhealth.com/life-sciences-trials/ [https://perma.cc/SP4Z-6B6N] ("Ensure your enrolled trial population reflects that of your future patient population. This real-world evidence software accelerates recruitment within underrepresented communities by delivering 400+ social determinants of health (SDoH) insight.").

411. *See generally* Gisela Schott et al., *The Financing of Drug Trials by Pharmaceutical Companies and its Consequences*, 107 DEUTSCHES ÄRZTEBLATT INT'L 279, 279 (2010); *see also* AYLIN SERTKAYA ET AL., EXAMINATION OF CLINICAL TRIAL COSTS AND BARRIERS FOR DRUG DEVELOPMENT iv (2014), https://aspe.hhs.gov/sites/default/files/migrated_legacy_files//44516/rpt_erg.pdf [https://perma.cc/LL2J-VHGQ] ("The major obstacles to conducting clinical trials in the United States identified through this research include: high financial cost, the lengthy time frames, [and] difficulties in recruitment and retention of participants . . . .").

412. *See* Stefan Harrer et al., *Artificial Intelligence for Clinical Trial Design*, 40 TRENDS PHARMACOLOGICAL SCIS. 577, 577 (2019) (noting how the introduction of technology can expedite the clinical research process); Aditya Kudumala, Dan Ressler & Wendell Miranda, *Scaling Up AI Across the Life Sciences Value Chain*: *Enhancing R&D, Creating Efficiencies, and Increasing Impact*, DELOITTE INSIGHTS, (Nov. 4, 2020), https://www2.deloitte.com/us/en/insights/industry/life-sciences/ai-and-pharma.html [https://perma.cc/CEB2-BE9S].

413. *See, e.g.*, Nayan Chaudhari et al., *Recruitment and Retention of the Participants in Clinical Trials: Challenges and Solutions*, 11 PERSPECT. CLINICAL RSCH. 64, 64 (2020) ("Recruiting the planned sample size within the defined time frame in clinical trials has proven to be the chief bottleneck in the drug development process."); *see also* Stewart Gandolf, *Overcoming Clinical Trial Marketing Challenges: Recruitment and Retention*, HEALTHCARE SUCCESS, https://healthcaresuccess.com/blog/healthcare-marketing/overcoming-clinical-trial-marketing-challenges-patient-recruitment-retention.html [https://perma.cc/HKC8-CV38] ("One of the fundamental challenges faced by hospitals, medical practices and healthcare research organizations that are engaged in clinical trials is the recruitment and retention of volunteer participants."). These pressures are not new. *See* DEP'T OF HEALTH & HUM. SERVS.: OFF. OF INSPECTOR GENERAL, OEI-01-97-00195, RECRUITING HUMAN SUBJECTS: PRESSURES IN INDUSTRY SPONSORED CLINICAL RESEARCH 1(2000), https://oig.hhs.gov/oei/reports/oei-01-97-00195.pdf [https://perma.cc/S9NC-NN7N] (noting the difficulties of patient recruitment and retention for clinical trials in 2000).

can make it difficult to find enough qualified participants.[414] And, increasingly, there is pressure to identify subjects who are not just medically eligible but also "demographically diverse."[415]

The challenge of finding eligible participants is particularly difficult when the research involves a rare condition or disease.[416] Sponsors are already partnering in advance of any specific trial with patient advocacy or consumer groups in order to have better access to qualified patients.[417] Relationships play an important role in recruiting because patients trust them.[418] This raises ethical and regulatory concerns because the communications

414. *See* Amanda McDowell, *5 Reasons Why It's so Hard to Find Clinical Trial Volunteers*, ANTIDOTE (Sept. 13, 2023), https://www.antidote.me/blog/medical-trials-top-5-reasons-why-its-so-hard-to-find-participants [https://perma.cc/RSS8-ABRS] ("Clinical trials require conditions to be as controlled as possible to deliver meaningful results. But 'as possible' is a relative concept, and sometimes the requirements (inclusion and exclusion criteria) that patients must meet in order to participate in a study can be too strict.").

415. *See* Randall A. Oyer et al., *Increasing Racial and Ethnic Diversity in Cancer Clinical Trials: An American Society of Clinical Oncology and Association of Community Cancer Centers Joint Research Statement*, 40 J. CLINICAL ONCOLOGY 2163, 2163 (2022) ("A concerted commitment across research stakeholders is necessary to increase equity, diversity, and inclusion (EDI) and address barriers to cancer clinical trial recruitment and participation."). The challenge of diversifying clinical trials is not unique to the United States. *See* Annette S. Gross et al., *Clinical Trial Diversity: An Opportunity for Improved Insight into the Determinants of Variability in Drug Response*, 88 BRITISH J. CLINICAL PHARMACOLOGY 2700 (2022).

416. *See* Jemima E. Mellerio, *The Challenges of Clinical Trials in Rare Diseases*, 187 BRITISH J. DERMATOLOGY 453 (2022) ("Undertaking any clinical trial can be laden with obstacles and challenges. Both commercial trials and academically sponsored studies share questions around trial design, recruitment targets, mitigating dropout and, ultimately, challenges of regulatory approval if the bar for efficacy and safety are met. However, trial design and delivery in the rare disease arena bring specific considerations and potential pitfalls, for researchers, patients, pharma and regulators.").

417. *See* Lisa Pahl, *Partnering with Patient Advocacy Groups Benefits Pharma and Patients*, ANJU (May 2, 2022), https://www.anjusoftware.com/insights/patient-advocacy-groups/ [https://perma.cc/6PFB-A9RC] ("Hearing patients' first-hand accounts of their struggles with and victory over disease is a powerful means to recruit patients to trials. Survivor testimony is best told in collaboration with PAGs, argues portfolio manager Anne-Marie Mongan at Clinical Trials Arena."); *see also* Neil Lesser et al., *Broadening Clinical Trial Participation to Improve Health Equity*, DELOITTE INSIGHTS (Oct. 4, 2022), https://www2.deloitte.com/us/en/insights/industry/health-care/increasing-diversity-clinical-trials.html [https://perma.cc/H7SR-PUMK].

418. Alexandria Younossi et al., *Enhancing Clinical Trial Diversity*, DELOITTE INSIGHTS (Nov. 11, 2021), https://www2.deloitte.com/us/en/insights/industry/life-sciences/lack-of-diversity-clinical-trials.html [https://perma.cc/T4PR-XFUN] ("Patients in the research and workshop pointed to the importance of meeting patients and caregivers where they live to improve awareness, access, and trust.").

between sponsors and potential subjects outside the structure of a clinical trial are not subject to preapproval.[419]

In 2021, OHRP's ethics advisory board issued non-binding guidance advising "sponsors or vendors interacting with prospective subjects" to "review carefully whether" information distributed through these groups "are subject to IRB oversight."[420] In particular, the guidance raised concerns that patients might feel "that they have little meaningful choice but to cooperate in sponsor requests" because of the pre-existing financial support.[421] Further, with particular relevance to materials presented using Persuasive AI, the guidance recommends that any "subject interviews or 'testimonials' should accurately portray clinical studies as research involving unproven, though promising, experimental agents or procedures."[422]

B. Regulatory Pressure to Diversify Clinical Drug Trials

In addition to the general pressure to recruit and enroll eligible participants, sponsors are facing new pressures to address the long-standing reality that participants in clinical drug trials are primarily white men.[423]

C. Mandates to Diversity Clinical Drug Trials

In April 2022, the FDA announced that, pursuant to the FDA Reauthorization Act of 2017 (FDARA), it was issuing "new guidance to

---

419. *See* Off. for Hum. Rsch. Prots., *supra* note 380.

420. *Id.*; *see also* Mark Barnes & David Peloquin, *New SACHRP Recommendations on Interactions among Sponsors, Clinical Trial Sites, and Study Subjects*, ROPES & GRAY (May 4, 2021), https://www.ropesgray.com/en/newsroom/alerts/2021/may/new-sachrp-recommendations-on-interactions-among-sponsors-clinical-trial-sites-and-study-subjects [https://perma.cc/7DCP-AWLW].

421. Barnes & Peloquin, *supra* note 420 ("Interactions should be designed to minimize the likelihood that subjects or their families will perceive that they have little meaningful choice but to cooperate in sponsor requests. Requests for subjects and their families to engage in media or public relations activities should ideally occur *after* the subject has completed the trial, unless there is a compelling reason for such activities to occur while a subject is still participating in the trial.").

422. *Id.*

423. *Diversity and Inclusion in Clinical Trials*, NAT'L INST. ON MINORITY HEALTH AND HEALTH DISPARITIES (Apr. 24, 2023), https://www.nimhd.nih.gov/resources/understanding-health-disparities/diversity-and-inclusion-in-clinical-trials.html [https://perma.cc/8Z4U-N7PQ] ("Historically, clinical trials did not always recruit participants who represented the individuals most affected by a particular disease, condition, or behavior. Often, these clinical trials relied almost exclusively on White male study participants."); *see also FDA Takes Important Steps to Increase Racial and Ethnic Diversity in Clinical Trials*, FDA (Apr. 13, 2022), https://www.fda.gov/news-events/press-announcements/fda-takes-important-steps-increase-racial-and-ethnic-diversity-clinical-trials [https://perma.cc/734E-QTML] ("Despite having a disproportionate burden for certain diseases, racial and ethnic minorities are frequently underrepresented in biomedical research.").

industry for developing plans to enroll more participants from underrepresented racial and ethnic populations in the U.S. into clinical trials."[424] This guidance reflects a process of realization that started with the revelation after Tuskegee that less powerful segments of society were being asked to take on a disproportionate burden of the risks of participating in clinical research.[425] The initial drafting of ethical principles resulted in the Common Rule, which emphasizes the importance of "justice" in the context of treating participants fairly by equally distributing the burdens of research.[426]

Shortly after the adoption of the Common Rule, researchers and ethicists applied the concept of "justice" to not just the burden of being included in research, but also in the unfairness that could result from exclusion.[427] For example, during the development of drugs to combat HIV, people with AIDS complained that, because of overly narrow enrollment criteria, they were being unfairly excluded from trials that offered their only chance of survival.[428] Almost at the same time, there was a growing realization about a third problem of exclusion: that long observed differences in the efficacy of the same drugs among men and women were attributable to the standard practice of using a homogeneous segment of the population in clinical trials, usually young white men.

As a 2022 comprehensive government report explained, "[M]any groups underrepresented and excluded in clinical research can have distinct disease presentations or health circumstances that affect how they will respond to

424. *FDA Takes Important Step to Increase Racial and Ethnic Diversity in Clinical Trials*, FDA (Apr. 13, 2022), https://www.fda.gov/news-events/press-announcements/fda-takes-important-steps-increase-racial-and-ethnic-diversity-clinical-trials [https://perma.cc/3LEN-JCH9].

425. Vicki S. Freimut et al., *African Americans' Views on Research and the Tuskegee Syphilis Study*, 52 SOC. SCI. & MED. 797, 799 (2001) (noting that the legacy of the Tuskegee Syphilis study "hampers recruitment for research among African Americans").

426. *History of Research Ethics*, UMKC: OFF. OF RSCH. SERVS., https://ors.umkc.edu/services/compliance/irb/history-of-research-ethics.html# [https://perma.cc/PAK5-WST2]; BELMONT REPORT, *supra* note 177, at 5–6.

427. *See* Ezekiel J. Emanuel, David Wendler & Christine Grady, *What Makes Clinical Research Ethical?*, 283 JAMA 2701, 2703 (2000).

428. *See* Deborah J. Cotton et al., *Guidelines for the Design and Conduct of AIDs Clinical Trials*, 16 CLINICAL INFECTIOUS DISEASES 816, 816 (1993) ("Because AIDS is a syndrome, with infection leading to a myriad of complications, issues related to appropriate and exclusion criteria and the use of concomitant medications arose. Enrollment of patients into studies lagged when large numbers of interested patients did not meet exacting entry criteria, and retention suffered when patients' medical conditions necessitated the initiation of medication not permitted by study design.").

an investigational drug or therapy."[429] If these underrepresented populations are not part of the population participating in the studies, then they will be disadvantaged when in need of medical care once those products are on the market. The exclusion from trials has caused considerable harm to Black patients in terms of access to new treatments.[430]

Further studies revealed that low enrollment of Black people, people with disabilities, and other underrepresented populations in clinical trials also resulted in substantial gaps in knowledge that put those excluded at a therapeutic disadvantage.[431] The government's response to these harms caused by exclusion has been to make diversifying clinical trials a priority in the research it funds, conducts, and regulates.[432] For example, the NIH has, since 2017, required that all studies it funds report the results by sex or gender, race, and ethnicity.[433] The FDA has also long encouraged diversity in its regulated trials.[434] Unfortunately, however, none of these programs intended to recruit more women and other underrepresented populations have been very successful.[435] Therefore, both the NIH and FDA have been making greater efforts to both incentivize and require diversity.[436]

Laws and policies encouraging researchers to diversify the population they enroll in research studies are intended to benefit those who have

---

429. NAT'L ACADS. OF SCIS., ENG'G, & MED., IMPROVING REPRESENTATION IN CLINICAL TRIALS AND RESEARCH: BUILDING RESEARCH EQUITY FOR WOMEN AND UNDERREPRESENTED GROUPS 24 (Kirsten Bibbins-Domingo & Alex Helman eds., 2022) (first citing Christoph Beglinger, *Ethics Related to Drug Therapy in the Elderly*, 26 DIGESTIVE DISEASES 28, 28 (2008); then citing Francis P. Crawley, Ronald Kurz & Hidefumi Nakamura, *Testing Medications in Children*, 347 NEW ENG. J. MEDICINE 763, 763–64 (2003); then citing Mariana Garcia et al., 118 CIRCULATION RSCH. 1273, 1281 (2017); and then citing A. Ramamoorthy et al., *Racial/Ethnic Differences in Drug Disposition and Response: Review of Recently Approved Drug*, 97 CLIN. PHARMACOLOGY & THERAPEUTICS 263, 263 (2015)).

430. *See id.* at 16, 32.

431. *See Diversity and Inclusion in Clinical Trials*, *supra* note 423.

432. *See* U.S. DEP'T OF HEALTH AND HUM. SERVS., ENHANCING THE DIVERSITY OF CLINICAL TRIAL POPULATIONS — ELIGIBILITY CRITERIA, ENROLLMENT PRACTICES, AND TRIAL DESIGNS GUIDANCE FOR INDUSTRY 4 (2020), https://www.fda.gov/media/127712/download [https://perma.cc/4MGK-NRC4].

433. NAT'L INSTS. OF HEALTH, INCLUSION ACROSS THE LIFESPAN: JUNE 1–2, 2017 WORKSHOP SUMMARY 16 (2017), https://report.nih.gov/sites/report/files/docs/NIH%20Inclusion%20Across%20the%20Lifespan%20Workshop%20Summary%20Report.pdf [https://perma.cc/EH2A-6UMA].

434. *See FDA Takes Important Steps to Increase Racial and Ethnic Diversity in Clinical Trials*, *supra* note 423.

435. *See* Rossybelle P. Amorrortu, *Recruitment of Racial and Ethnic Minorities to Clinical Trials Conducted Within Specialty Clinics: An Intervention Mapping Approach*, 19 TRIALS, Feb. 17, 2018, at 1, 1 ("Despite efforts to increase diversity in clinical trials, racial/ethnic minority groups generally remain underrepresented . . . .").

436. *See FDA Takes Important Steps to Increase Racial and Ethnic Diversity in Clinical Trials*, *supra* note 423.

historically been excluded. However, not everyone agrees that is the right approach to solving a problem that may have more to do with concerns about racial discrimination in the health care system than researcher-created barriers to participation.[437] Despite years of advocacy and sustained effort in the public and private sectors, the segment of the population that participates in clinical trials remains almost entirely white and male.[438] Not only is this lack of diversity itself unjust, but there is also increasing evidence that biomedical products tested only on one segment of the population may be ineffective and perhaps even unsafe for populations on which it was not tested.[439] But identifying the problem has proven to be very different from solving it.[440] When asked, many potential Black participants consistently point to their own lived experience with racism

437. *See* Amie Devlin et al., *The Effect of Discrimination on Likelihood of Participation in a Clinical Trial*, 7 J. RACIAL ETHNIC HEALTH DISPARITIES 1124, 1124–29 (2020).

438. Todd C. Knepper & Howard L. McLeod, *When Will Clinical Trials Finally Reflect Diversity?*, 557 NATURE 157, 157 (2018) ("In 1997, 92% of the participants in these [clinical] trials were white; in 2014, we found that this figure was still nearly 86%."); *see also* Mark H. Barlek et al., *The Persistence of Sex Bias in High-Impact Clinical Research*, 278 J. SURGICAL RSCH. 364, 364 (2022) ("Sex bias remains prevalent in human clinical research trials. Improvements have been made in NIH-funded clinical trials; however, this constitutes a small percentage of overall studies.").

439. *See, e.g.*, Eds., *Striving for Diversity in Research Studies*, 385 NEW ENG. J. MEDICINE 1429, 1429 (2021) ("The study population may not include groups representing large fractions of those who might be candidates for trial intervention. This can leave clinicians in a quandary about whether and how to apply the research findings to their own patients, for whom the risk-benefit profile may differ."); *see also* Cody Crnkovic et al., *Low Rates of Reporting Race, Ethnicity, and Socioeconomic Status in Studies Published in Top Orthopaedic Journals*, 104 J. BONE & JOINT SURGERY 1244, 1244 (2022) ("Disparities in orthopaedic care are widely reported for racial and ethnic minorities, who, in addition to those patients with lower SES, experience longer wait times for orthopaedic care and worse outcomes following orthopaedic surgery.").

440. *See* Oyer et al., *supra* note 415, at 2165 ("Because the problem stems from multiple factors, multi-faceted strategies are needed to increase participation among people from racial and ethnic minority populations.").

when trying to access health care.[441] Others note the expense and inconvenience involved in enrolling in a clinical study.[442]

Whether that lack of interest is indeed from lack of access or whether it is a rational response to their own experience in being poorly served by the health care system, manipulating them into enrollment is inherently wrong.

D. Evidence of Companies Already Offering AI Products to Assist in Recruiting Participants

These pressures on recruiting eligible participants are already leading sponsors to seek out AI solutions. This is evidenced by advertisements of Contract Research Organizations (CRO).[443] A CRO is a "company hired by another company or research center to take over certain parts of running a clinical trial."[444] There are four primary steps in recruiting participants:

1. Identifying or sourcing potential participants who may be eligible.
2. Discussing all aspects of the trial with them, ensuring comprehension and voluntariness, and subsequently obtaining informed consent for participation.
3. Conducting a physical examination and screening procedures as mentioned in the protocol.
4. Enrolling the participant based on the eligibility criteria.[445]

---

441. For an overview of existing research from an article published in 2010, see Darcell P. Scharff et al., *More Than Tuskegee: Understanding Mistrust About Research Participation*, 21 J. HEALTH CARE POOR UNDERSERVED 879, 880 (2010) ("Attitudinal studies suggest that mistrust of clinical investigators is strongly influenced by sustained racial disparities in health, limited access to health care, and negative encounters with health care providers."); Joanne Kenen & Elaine Batchlor, *Racist Doctors and Organ Thieves: Why So Many Black People Distrust the Health Care System*, POLITICO (Dec. 18, 2022, 7:00 AM), https://www.politico.com/news/magazine/2022/12/18/black-mistrust-healthcare-00060324 [https://perma.cc/AXL8-XTB3] (accrediting mistrust by African Americans towards the health care system to "people's own experiences").

442. *See* Kenen & Batchlor, *supra* note 441 ("[B]laming suspicions and distrust on long-ago atrocities lets the current health care system—still rife with inequities and injustices—off the hook.").

443. *Contract Research Organization*, NAT'L CANCER INST., https://www.cancer.gov/publications/dictionaries/cancer-terms/def/contract-research-organization [https://perma.cc/AJ78-BH8E].

444. *Id.* ("The company may design, manage, and monitor the trial, and analyze the results.").

445. Nayan Chaudhari et al., *Recruitment and Retention of the Participants in Clinical Trials: Challenges and Solutions*, 11 PERSPS. CLINICAL RSCH. 64, 65 (2020).

As the chart below shows, these companies are already marketing their ability to use AI to improve the subject recruitment process. While it is not possible to know what kind of AI they are using, some of the companies make clear that their goal is to increase diversity in clinical trials.[446]

Company Name	About
Fathom IT Group	Fathom IT Group uses the latest technologies in AI, Blockchain, CTMS, Cloud Service Data Analytics/ Programmatic, and Inbound CRM strategies for research subject recruitment.[447]
SubjectWell	SubjectWell's technology identifies patients on general interest websites and attracts their interest through ads. The patients respond to the ads and fill out a health profile. SubjectWell's technology then compares their profile to possible studies in their area. The patients are then contacted once a research study in their area is identified.[448]
Clarify	Clarify's goal is to increase diversity in clinical trials.[449] It uses AI to increase recruitment from underrepresented communities in clinical trials.[450]
Eversana	ACTICS by Eversana combines AI analytics with real-time commercial services. It is a cloud-based solution that enables pharmaceutical companies to make the most informed decisions with their patients.[451]
AutoCruitment	AutoCruitment has an automated platform that targets,

446. *See, e.g.*, *Make Diversity in Your Clinical Trials a Reality*, *supra* note 410.
447. *Research Subject Recruitment*, FATHOM IT GRP., https://www.fathomit.com/research-subject-recruitment [https://perma.cc/GXW6-RGJ7].
448. *Guided Recruitment*, SUBJECTWELL, https://www.subjectwell.com/our-approach/ [https://perma.cc/6UZ6-EAP7].
449. Note, since these companies are all offering their services to sponsors conducting trials to gather information for FDA approval, they often do not use the term "clinical drug trial" since the trial may also be for approval of another regulated product such as a vaccine or a medical device.
450. *Clinical Trials Effectiveness*, CLARIFY HEALTH SOLS., https://clarifyhealth.com/solutions/life-sciences/clinical-dei/ [https://perma.cc/NCU4-PBVZ].
451. *Technology Solutions*, EVERSANA, https://rb.gy/lbm23 [https://perma.cc/HY6W-S9F8].

	recruits, screens, and refers patients for trials.[452]
Melax Tech	Melax Tech as a clinical trial optimization platform, called VITAL, that uses NLP/AI to prescreen patients.[453]
Viz Recruit	Viz Recruit's AI platform identifies and connects candidates to the research team for clinical trials.[454]
Opyl	Opyl has an AI clinical trial platform that uses social media and search engine optimization to connect patients to clinical trials.[455]
Worldwide Clinical Trials in partnership with Deep Lens	Uses AI technologies to recruit patients for oncology research.[456]
Trial Wire	Trial Wire uses AI to match patients to studies.[457]
Concert AI	Concert AI uses AI to assess study feasibility, evaluate patients, and optimize trials.[458]

E. The Role of Advertising in Recruiting Research Participants

The most traditional strategy for recruiting research participants is advertising.[459] But unlike advertising for other products and services, the contents of materials seeking participants in clinical drug trials is heavily

452. *Hyper-Targeted Patient Recruitment*, AUTOCRUITMENT, https://rb.gy/do0yg [https://perma.cc/TXK6-W8YA].

453. *Optimizing Clinical Trial Design and Recruitment with NLP/AI*, MELAX TECH, https://www.melaxtech.com/post/optimizing-clinical-trial-design-and-recruitment-with-nlp-ai [https://perma.cc/G55L-BS24].

454. *Viz Recruit*, VIZ.AI, https://www.viz.ai/life-sciences/clinical-trials [https://perma.cc/43TQ-3UBW].

455. *Introducing Opin*, OPYL, https://opyl.ai/services/clinical-trial-recruitment/ [https://perma.cc/HG6V-SBVD].

456. Eds. at Uncommon Conversations, *AI-Driven Technologies Poised to Improve Subject Recruitment in Oncology Clinical Studies*, WORLDWIDE CLINICAL TRIALS (Mar. 10, 2020), https://www.worldwide.com/blog/2020/03/ai-driven-technologies-poised-to-improve-subject-recruitment-in-oncology-clinical-studies/ [https://perma.cc/GZ4Z-5M8Y].

457. *About us*, TRIAL WIRE, https://trial-wire.com/about-us/ [https://perma.cc/7AS5-ZDH2].

458. CONCERTAI, EASE THE BURDEN OF PATIENT RECRUITMENT FOR CANCER TRIALS WITH AUTOMATED AI-DRIVEN PRE-SCREENING 1–3 (2023), https://www.concertai.com/wp-content/uploads/2022/01/eScreening-Fact-Sheet-for-Providers.pdf [https://perma.cc/W3Z9-GRWX].

459. *See* Heidi Moseson, Shefali Kumar & Jessie L. Juusola, *Comparison of Study Samples Recruited with Virtual Versus Traditional Recruitment Methods*, 19 CONTEMP. CLINICAL TRIALS COMMC'NS, Sept. 2020, at 1 (comparing recruitment times between traditional and virtual driven studies).

regulated.[460] All advertising for a specific trial, whether sponsored by a private company or funded by the government, must be reviewed and approved in advance by an IRB.[461] FDA rules state that the "IRB should also review the methods and material that investigators propose to use to recruit subjects."[462] The NIH provides similar guidance, including specific information about advertising through social media.[463] While either general or targeted, advertising does not have to contain the same information as an informed consent document.[464] Instead, the FDA advises "that any advertisement to recruit subjects should be limited to the information the prospective subjects need to determine their eligibility and interest."[465]

VIII. JUSTIFYING THE DECISION TO BAN

The EU bases its recommendation to ban AI that can manipulate decision-making on existing laws that recognize a general right to individual autonomy broader than any existing law in the United States.[466] But while U.S. law lacks general protection for autonomy in decision-making, it very specifically provides for protection during the research-

460. *Recruiting Study Subjects*, FDA (Sept. 5, 2018), https://www.fda.gov/regulatory-information/search-fda-guidance-documents/recruiting-study-subjects [https://perma.cc/T7WT-E7YL] ("The IRB should also review the methods and material that investigators propose to use to recruit subjects."); *see also* Amanda McDowell, *Clinical Trial Advertising Guidelines*, ANTIDOTE (June 22, 2023), https://www.antidote.me/blog/clinical-trial-advertising-guidelines-to-follow-for-irb-submission [https://perma.cc/CVG2-5WXX].

461. *See Institutional Review Boards Frequently Asked Questions*, *supra* note 166.

462. *Recruiting Research Subjects*, *supra* note 460.

463. *See Guidance Regarding Social Media Tools*, NAT'L INSTS. OF HEALTH, https://www.nih.gov/health-information/nih-clinical-research-trials-you/guidance-regarding-social-media-tools [https://perma.cc/9JCX-9U9W].

464. *See* Stephanie Mull, *What Type of Advertising Must be Reviewed by the IRB*, PROXIMA, https://www.proximacro.com/faqs/what-type-of-advertising-must-be-reviewed-by-the-irb [https://perma.cc/VUT3-Z3M7].

465. *Id.* ("It should be noted, however, that FDA does not require inclusion of all the listed items: the name and address of the clinical investigator and/or research facility, the condition under study and/or the purpose of the research, in summary form, the criteria that will be used to determine eligibility for the study, a brief list of participation benefits, if any (e.g. a no-cost health examination), the time or other commitment required of the subjects, and the location of the research and the person or office to contact for further information.").

466. *See* James Vincent, *EU Draft Legislation Will Ban AI for Mass Biometric Surveillance and Predictive Policing*, THE VERGE (May 11, 2023, 8:19 AM), https://www.theverge.com/2023/5/11/23719694/eu-ai-act-draft-approved-prohibitions-surveillance-predictive-policing [https://perma.cc/TG83-G9YY].

related informed consent process.[467] Therefore, the EU's rationale for banning AI that manipulates decision-making is directly relevant to imposing a ban for its use in obtaining informed consent for research.

The EU considers two factors: (1) the extent of harm possible and (2) the ability to prevent that harm or mitigate its consequences.[468] It also has precedents banning the use of technologies for use in situations where their potential for great harm exceeds otherwise acceptable benefits.[469]

A ban is not the right answer to every prospective risk or even known danger from the use of AI.[470] But the high standard of protection granted to (1) the research subjects, (2) characteristics of the technology itself, and (3) factors that make it particularly compelling to those seeking to enroll diverse subjects in clinical trials mean that no intermediate measure is sufficient to protect the integrity of the informed consent process.

The EU's decision to ban manipulative AI is consistent with historical approaches to new technology with three relevant characteristics. First, when their future impact is either unknown at the time of its development or which, on implementation, shows potential for great harm.[471] While sometimes this unknown effect is for good, it can also, as in the case of the invention of dynamite or three-dimensional (3D) printing, be for the worst.[472]

Second, when the features of technology make it impossible to mitigate this harm through available safety mechanisms.[473] These historical precedents

---

467. *See* 45 C.F.R. § 46.116(b)–(c) (2017) (providing for basic and additional elements of informed consent in the clinical research setting).

468. *See Proposed EU AI Act*, *supra* note 5, § 5.2.2., at 12 ("The regulation follows a risk-based approach, differentiating between uses of AI that create (i) an unacceptable risk, (ii) a high risk, and (iii) low or minimal risk.").

469. *See id.* at 3 ("Certain particularly harmful AI practices are prohibited as contravening Union values . . . .").

470. For a very compelling analysis of how AI can, and is, being used for the good of society, see generally Orly Lobel, *The Law of AI for Good* 27–41 (Univ. of San Diego Sch. of L., Rsch. Paper No. 23-001, 2023), https://papers.ssrn.com/abstract=4338862 [https://perma.cc/C79M-JMFT] (discussing the potential of AI for good in various fields, including environmental applications, poverty, and education).

471. *See Proposed EU AI Act*, *supra* note 5, at 43–44 (prohibiting artificial intelligence practices whose purpose is to distort, exploit, or injure individuals or specific groups of people).

472. *See* DNews, *10 Good Techs Turned Bad*, SEEKER (Oct. 14, 2013, 7:00 AM), https://www.seeker.com/10-good-techs-turned-bad-1767944010.html [https://perma.cc/R9AJ-VUBL].

473. The long-lasting consequences of the release of nuclear material is a strong example of irreversible harm. *See, e.g.,* James M. Acton, Commentary, *The Most Immediate Nuclear Danger in Ukraine Isn't Chernobyl*, CARNEGIE ENDOWMENT FOR INT'L PEACE (Feb. 24, 2022), https://carnegieendowment.org/2022/02/24/most-immediate-nuclear-danger-in-ukraine-isn-t-chernobyl-pub-86521 [https://perma.cc/QGL9-87NM].

are the basis for calls to halt the military's integration of autonomous AI technology in weapons systems.[474]

Third, when it may be impossible to even detect when the technology is actively causing harm.[475] As a biodefense expert for the Air Force explains, one of the greatest dangers from biological weapons is the lack of any way to monitor their development or use; "Without a succinct way of monitoring and ensuring compliance, we risk continued accidental or purposeful releases with little to no recourse."[476]

A. Risk of Irrevocable Harm

No mitigating measure can prevent the harm caused by injecting persuasive technology into the informed consent process.[477] The inability of existing remedies to prevent AI from manipulating decision-making is unacceptable because the harm is not the consequences of enrollment, it is the disruption of the process of influencing an already ill patient to enroll in a research study that is itself an irreparable violation of their civil rights.

B. Lack of Ability to Prevent Harm

Comparing preventing harm from AI to "prevent[ing] hurricanes" or "forc[ing] a crow to fly west instead of east," Professor Daniel Gervais

474. *See* Zachary Fryer-Biggs, *Are We Ready for Weapons to Have a Mind of Their Own?*, CTR. FOR PUB. INTEGRITY (Feb. 17, 2021), https://publicintegrity.org/national-security/future-of-warfare/mind-of-their-own-artificial-intelligence-weapon/ [https://perma.cc/TQQ8-T2NJ].

475. *See* Emily M. Bender & Alex Hanna, Opinion, *AI Causes Real Harm. Let's Focus on That Over the End-of-Humanity Hype*, SCI. AM. (Aug. 12, 2023), https://www.scientificamerican.com/article/we-need-to-focus-on-ais-real-harms-not-imaginary-existential-risks/ [https://perma.cc/LGT6-38RJ] (noting the routineness of harms being caused by AI in day-to-day life, including routine discrimination and the dissemination of misinformation); Cpt. Dahlia Andreadis, *Biological Weapons Accountability*, WILD BLUE YONDER (May 10, 2021), https://www.airuniversity.af.edu/Wild-Blue-Yonder/Article-Display/Article/2596954/biological-weapons-accountability/ [https://perma.cc/452L-5BGY] (identifying the particular danger posed by biological weapons from "[t]he lack of accountability within the biological weapons community is a danger to all.").

476. Andreadis, *supra* note 475.

477. *See* Bard, *supra* note 12 ("For example, it is now possible to create an informed consent conversation between a potential participant and an AI avatar, online or in person, with the vocal, visual, and even syntactical characteristics of a person who the potential participant would find the most persuasive.").

describes the challenge as "unprecedented" because humans have never tried to control the behavior of something more intelligent than themselves.[478] But even if it were possible to access the algorithms before they are employed, it is the nature of AI functions to adapt and learn. Therefore, it is in many cases difficult to know how or why they reach their results.[479]

C. Lack of Ability to Detect Harm When It Is Happening

Because AI that manipulates human behavior leaves no traces, regulatory measures designed for identifiable harms are inadequate.[480] In contrast to a program that that disproportionately misidentifies Black faces or is worse at detecting cancer on an X-ray than the human rate of success, there is no way of knowing when a person's decision to enroll in a research study has been manipulated. First, it is impossible to determine whether or not the patient was influenced.[481] Second, if they do enroll and the study changes the course of their treatment, then that harm too is irrevocable.[482]

Not only is there limited understanding of how AI manipulates decision-making, but there are also no markers to detect if that has happened.[483] There is also no international vocabulary to describe what the technology is doing or even how it is causing harm.[484] A large part of this is because nothing about the process is static. Because the program

---

478. Daniel J. Gervais, *Towards an Effective Transnational Regulation of AI*, 38 A.I. & SOC'Y 391, 391, 392 (2023).

479. *See* Neil Savage, *Breaking into the Black Box of Artificial Intelligence*, NATURE (Mar. 29, 2022), https://www.nature.com/articles/d41586-022-00858-1 [https://perma.cc/2PUH-745G]; *see also* Bathaee, *supra* note 321, at 891, 893.

480. *See* Georgios Petropoulos, *The Dark Side of Artificial Intelligence: Manipulation of Human Behaviour*, BRUEGEL (Feb. 2, 2022), https://www.bruegel.org/blog-post/dark-side-artificial-intelligence-manipulation-human-behaviour [https://perma.cc/45D2-X9K8].

481. *See id.* (noting the possibility for AI to manipulate an individual's behavior); *see also* FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 3–4 (2015).

482. *See* Deborah A. Zarin, Steven N. Goodman & Jonathan Kimmelman, *Harms from Uninformative Clinical Trials*, 322 JAMA 813, 813 (2019) (recognizing that enrollment of a patient in a clinical trial which lacks scientific merit is "a serious breach of trust and a violation of research ethics" even if the patient is not physically inured).

483. *See* Petropoulos, *supra* note 480 (recognizing that artificial intelligence can "target users at the individual level with manipulative strategies that are much more effective and difficult to detect").

484. *See* O'Shaughnessy, *supra* note 401 ("Subtle differences in definition—as well as the overlapping and loaded terminology different actors use to describe similar techniques —can have major impacts on some of the most important problems facing policymakers."); *see also* David R. Amariles, *Regulating Artificial Intelligence – Is Global Consensus Possible?*, FORBES (Sept. 9, 2022, 8:54 AM), https://www.forbes.com/sites/hecparis/2022/09/09/regulating-artificial-intelligence—is-global-consensus-possible/ [https://perma.cc/S28U-2XYJ] (noting the lack of uniformity in regulations addressing AI).

is designed to learn and grow as it is being used, it is in a constant state of evolution and change.[485]

D. Inadequacy of Existing Remedies

> *"'[I]nformed consent' is not a good ethical proxy in data collection and exploitation without expert guardrails. Few of us can really understand on our own the full consequences of our consent. A company selling or deploying AI that abuses personal data should not be able to evade responsibility by citing the supposedly informed consent of the victims."*[486]

One of the great challenges of proposing solutions for the potential of undue influence by emotional AI is the variety and nature of the kind of harm it can cause. Another challenge is the lack of data documenting either how that harm occurs, how often it occurs, and how serious it is when it happens.[487] This is because AI creators and users do not know how they work.[488] Once a program begins to access and analyze data it changes itself in ways that are difficult to detect or even to explain.[489] Analogizing to vaccine development, Ash Carter writes that lack of knowledge about how algorithms cause harm is not a reason to excuse AI from government oversight.[490] Addressing but dismissing the inevitable complaints that each company's algorithms are "proprietary," he notes that the "[g]overnment routinely handles proprietary secrets of competing

485. *See* O'Shaughnessy, *supra* note 401 ("Consider the rapidly advancing field of generative machine learning, which has been used to produce AI-created artwork and artificial but realistic-seeming media known as 'deepfakes.' The definition of AI used in a recent EU policy draft explicitly includes systems that generate 'content,' in addition to 'predictions, recommendations, or decisions.'"); *see also* Alysa Austin, Daniel Felz & Kimberly K. Peretti, *Privacy, Cyber & Data Strategy Advisory: AI Regulation in the U.S.: What's Coming, and What Companies Need to Do in 2023*, ALSTON & BIRD (Dec. 9, 2022), https://www.alston.com/en/insights/publications/2022/12/ai-regulation-in-the-us [https://perma.cc/UVK3-G7ST].

486. Carter, *supra* note 241, at 304.

487. *See* Thomas, *supra* note 245.

488. *See id.*

489. Carter, *supra* note 241, at 305.

490. *Id.* at 303 ("Due to the COVID epidemic, most people are by now familiar with the Food and Drug Administration's 'safety and efficacy' testing that must precede release of a new vaccine. So the notion of requiring a process of qualified review for sensitive products is hardly new and should be the industry standard for AI.").

companies when it serves as a regulator or customer of advanced technology."[491]

The EU recommends banning rather than trying to regulate AI technology that can manipulate decision-making because AI's ability to make autonomous decisions faster than humans make it impossible to for humans to intervene in time to preempt harm.[492] So far, all efforts to preempt harm associated with AI have been impossible.[493] This is because the humans who have created the AI do not always retain the ability to monitor its activities.[494]

1. Research Participants Cannot Waive Their Right to Protection

Federal law prohibits research consent documents from containing "exculpatory clauses" in which the participant agrees to waive their right to receiving any otherwise applicable protections.[495] This non-waivable protection means that disclosure of possible AI influence is insufficient grounds for allowing its incorporation into the informed consent process. This is in contrast to other situations, such as agreeing to accept cookies on website, where it is possible to waive privacy rights or other protected rights.[496] But even if "disclosure" was effective, the rapidity with which AI algorithms make decisions renders any form of meaningful transparency impossible.[497] As Carter explains, algorithms "all make enormous numbers of tiny calculations that combine to make overall inferences that cannot be made quickly by humans, have not been recognized by humans, or even perhaps would never be recognized by humans."[498] This makes "literal transparency, normally the starting point for ethical accountability, completely

---

491. *Id.*

492. *See Proposed EU AI Act*, *supra* note 5, § 5.2.2., at 12.

493. *See* John-Stewart Gordon & Sven Nyholm, *Ethics of Artificial Intelligence*, INTERNET ENCYC. OF PHIL., https://iep.utm.edu/ethics-of-artificial-intelligence/ [https://perma.cc/N4FW-M2N5] ("Most AI researchers, programmers and developers as well as scholars working in the field of technology believe that we will never be able to design a fully unbiased system."); *see also* Jones, *supra* note 135, at 246–50 (discussing why it is so difficult to prevent bias in employment recommendations).

494. *See* Gordon & Nyholm, *supra* note 493.

495. *See* 45 C.F.R § 46.116(a)(6) (2018) ("No informed consent, whether oral or written, may include any exculpatory language through which the subject or the representative is made to waive or appear to waive any of the subject's legal rights, or releases or appears to release the investigator, the sponsor, the institution, or its agents from liability for negligence."); 21 C.F.R. § 50.20 (1981) (same).

496. *See Use of Cookies*, DENTONS (Sept. 3, 2020), https://www.dentons.com/en/use-of-cookies [https://perma.cc/XM63-3F9P].

497. *See* Carter, *supra* note 241, at 302.

498. *Id.*

impractical . . . [because] [i]t is usually impossible to 'deconvolve' the series of steps leading to the inferences made by AI."[499]

Laws that call for review of decisions made by algorithms are easy to see and these laws concern reversable harm.[500] For example, a person whose mortgage was denied for improper reasons can get that decision reversed and buy a house with little interruption.[501] This supervision can involve requiring developers to provide an "Algorithmic Impact Assessment" before the algorithm is employed "to anticipate, test, and investigate potential harms of the system before implementation; document those findings; and then either publicize them or report them to a regulator."[502]

Current regulatory proposals short of a ban are inadequate to meet the high standard of protection required to protect research subjects from AI's dangers because what the law protects is the process of human decision-making. Once that is contaminated, the harm is irrevocable.[503] As a result, many of the remedies proposed to either prevent or mitigate other risks from AI assisted decision-making cannot be deployed effectively to protect the decision-making process. The EU has adopted a risk-based approach to regulation based on the premise that the benefits of AI in some settings are so high and the consequences of harm so low that the loss of control and foreknowledge are acceptable. This is for two reasons. First, because AI algorithms are proprietary to the companies which develop them, it would be difficult to monitor compliance with any mandated limitations on how they can be designed.[504] Moreover, even if regulators

499. *Id.*

500. *See generally* Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 634 (2017).

501. *See* Ryan Browne & MacKenzie Sigalos, *AI has a Discrimination Problem. In Banking, the Consequences Can Be Severe*, CNBC (June 23, 2023, 10:37 AM), https://www.cnbc.com/2023/06/23/ai-has-a-discrimination-problem-in-banking-that-can-be-devastating.html [https://perma.cc/34RX-DSR6] ("[W]hen AI systems are specifically used for loan approval decisions . . . there is a risk of replicating existing biases present in historical data used to train the algorithms.").

502. Selbst, *supra* note 393, at 127.

503. *See* Dan Milmo & Hibaq Farah, *Malicious Use of AI Could Cause 'Unimaginable' Damage, says UN Boss*, THE GUARDIAN (July 18, 2023, 1:36 PM), https://www.theguardian.com/technology/2023/jul/18/malicious-use-of-ai-could-cause-huge-damage-says-un-boss [https://perma.cc/HYX5-FCAU].

504. *See* Selbst, *supra* note 393, at 128–29 ("Due partly to the reflexive opacity of technology companies and widespread claims of trade secrecy, researchers are left trying to reconstruct what companies are doing from vague public statements.").

were granted access to the original programs, it would be meaningless because they are designed to learn and change.

This opacity and unpredictability is why remedies requiring disclosure are also worthless.[505] For example, laws like Canada's that require developers to disclose the possible impact of an algorithm before it is used for decision-making depends on being able to anticipate that impact.[506]

2. The Speed of AI Decision-Making Renders Informed Consent Impossible

From the beginning, AI has been advertised as being able to analyze data faster than a human.[507] As Carter explains, "[f]ew of us can really understand on our own the full consequences of our consent. A company selling or deploying AI that abuses personal data should not be able to evade responsibility by citing the supposedly informed consent of the victims."[508]

3. Disclosure Does Not Prevent Harm

Disclosure of a potential conflict of interest is a common method of protecting informed consent in settings where having this information may put a party to the transaction on notice. In health care, this takes a form of a federal law requiring physicians who participate in the Medicare program to report payments they receive from drug companies.[509] Some proposed regulation of AI suggest that its use be disclosed when it is

505. For a discussion of opacity and how algorithms are not always able to be understood by humans, see Gordan & Nyholm, *supra* note 493.

506. *See Algorithmic Impact Assessment Tool*, GOV'T OF CANADA, https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html [https://perma.cc/HKL4-RWUN].

507. Catware, *Artificial Intelligence vs Human Intelligence: A Comparative Analysis*, MEDIUM (July 3, 2023), https://medium.com/@catware/artificial-intelligence-vs-human-intelligence-a-comparative-analysis-eb7dc054625c#:~:text=AI%20algorithms%20can%20analyze%20vast,Effective%20decision%2Dmaking [https://perma.cc/F9CZ-MF8C]. For a discussion of the current weakness of AI systems in high-stakes decision-making because "the model produces incorrect answers that conflict with physical reality because the system cannot explain why and how a situation has happened," see Bukhoree Sahoh & Anant Choksuriwong, *The Role of Explainable Artificial Intelligence in High-Stakes Decision-Making Systems: A Systematic Review*, 14 J. AMBIENT INTEL. & HUMANIZED COMPUTING 7827, 7834–35 (2023); *see also* Eric Colson, *What AI-Driven Decision-Making Looks Like*, HARV. BUS. REV. (July 8, 2019), https://hbr.org/2019/07/what-ai-driven-decision-making-looks-like [https://perma.cc/Q9MA-MG7N].

508. Carter, *supra* note 241, at 304 ("'[I]nformed consent' is not a good ethical proxy in data collection and exploitation without expert guardrails.").

509. *See* Eli Y. Adashi & I. Glenn Cohen, *Enforcement of the Physician Payments Sunshine Act: Trust and Verify*, 326 JAMA 807, 807 (2021).

involved with decision-making.[510] But the EU has rejected disclosure when the AI is deployed for the purpose of influencing a decision.[511] The American Medical Association (AMA) has defined a "conflict of interest" as a "situation in which a person is or appears to be at risk of acting in a biased way because of personal interests."[512] This necessarily broad statement can be broken down into several components. The first is what it means for a physician to be acting in a "biased way" and the second is what constitutes a "personal interest."[513]

The concern of a possible conflict of interest that a treating physician might have between their dual obligations to the patient and their role as an investigator in a clinical trial is different from those that exist outside the context of a research study. This conflict can arise at many different stages of the research process including before enrollment. A treating physician is, at root, a professional who is paid for their services.[514] Our current insurance reimbursement system, both the public and private components, continues to pay doctors more for performing procedures than for prescribing medication or giving advice.[515] But the stakes are higher when a physician has a relationship with a commercial entity that might affect their choice among different treatment options.[516]

In general, scholarship on the efficacy of disclosure to reduce the harm of conflicts of interests has been disappointing for those who hope it could

510. *See U.S. Rep. Ritchie Torres Introduces Federal Legislation Requiring Mandatory Disclaimer for Material Generated by Artificial Intelligence*, RITCHIE TORRES (June 5, 2023), https://ritchietorres.house.gov/posts/u-s-rep-ritchie-torres-introduces-federal-legislation-requiring-mandatory-disclaimer-for-material-generated-by-artificial-intelligence [https://perma.cc/2HFN-38WB] ("All generative AI should be required to disclose itself as AI. Disclosure is by no means a magic bullet, but it's a common-sense starting point to what will surely be a long road toward federal regulation.").

511. *See* Heike Felzmann et al. *Towards Transparency by Design for Artificial Intelligence*, 26 SCI. ENG'G ETHICS 1333, 1334 (2021).

512. Christopher C. Muth, *Conflict of Interest in Medicine*, 317 JAMA 1812, 1812 (2017).

513. *See* J. O'Neill, *Materiality of Conflict of Interest in Informed Consent to Medical Treatment in the United Kingdom*, 32 ETHICS & BEHAV. 375, 376 (2022).

514. Donald E. Yett et al., *Physician Pricing and Health Insurance Reimbursement*, 5 HEALTH CARE FIN. REV. 69, 69 (1983) ("Physicians are income-motivated.").

515. *See id.* at 70–72.

516. *See, e.g.*, Sarah I. Kamel et al., *Recent Trends Suggest Possible Inappropriate Utilization of Myocardial Perfusion Imaging*, 16 J. AM. COLL. RADIOLOGY 1013, 1013–17 (2019).

mitigate harm.[517] Further, a persistent concern about all forms of AI is that "[u]sers of AI systems do not in many cases know the exact objectives of AI algorithms," making them unable to recognize when they are being influenced.[518] This is particularly true for a patient who is already receiving medical care and can reasonably assume that what they are being offered is in their best interest to accept.[519] In contrast, a consumer buying a car at a dealership enters the negotiation with a level of wariness not present when a patient is being treated in a hospital.[520]

4. Oversight and Monitoring

Unlike disclosure, which creates a "buyer beware" situation in which a potential consumer knows of but is not prevented from engaging with an algorithm, oversight allows the AI to proceed only with the requisite supervision.[521] If the AI has done something wrong or inappropriate there is a process for the aggrieved party to contest a decision.[522] This retrospective review is only helpful if the decision can be reversed before it causes harm. It is not an effective or appropriate remedy for irreparable harm, such as being arrested by the police based on an error in identification.[523]

---

517. *See* Daniel S. Goldberg, *Financial Conflicts of Interest are of Higher Ethical Priority than "Intellectual" Conflicts of Interest*, 17 J. BIOETHICAL INQUIRY 217, 218 (2020) ("[D]isclosure is not an efficacious remedy for the ethical problems posed by [conflicts of interest]."); *see also* Carlo Petrini & Luciana Riva, *Conflicts of Interest Result from Relationships But Are Not Resolved by Preventing Relationships*, 18 J. BIOETHICAL INQUIRY 187, 187 (2021) (proposing that the solution to conflicts of interests is in determining whether "those relationships are such as to unduly affect an individual's judgment").

518. Petropoulos, *supra* note 480.

519. *See* Raymond J. Hutchinson, *A Therapeutic Conundrum: Should a Physician Serve Simultaneously as Caregiver and Researcher?*, 20 AM. J. BIOETHICS 96, 97 (2020).

520. *See* Greg Rosalsky, *Inside the Rise of 'Stealerships' and the Shady Economics of Car Buying*, NPR (Aug. 30, 2022, 6:30 AM), https://www.npr.org/sections/money/2022/08/30/1119715886/inside-the-rise-of-stealerships-and-the-shady-economics-of-car-buying [https://perma.cc/BH3X-G4HH].

521. François Candelon et al., *AI Regulation is Coming*, HARV. BUS. REV. (Sept.–Oct. 2021), https://hbr.org/2021/09/ai-regulation-is-coming [https://perma.cc/XE6Z-EBGJ]; *see also* McKendrick & Thurai, *supra* note 115.

522. *See* Danielle K. Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1264–65 (2008); *see also* Margot E. Kaminski & Jennifer M. Urban, *The Right to Contest AI*, 121 COLUM. L. REV. 1957, 1963 (2021) (noting the development of the right to contest AI decisions outside of Europe).

523. *See* Tesfaye Negussie, *Black Man Wrongfully Arrested in DeKalb County, Georgia Due to Facial Recognition Tech: Lawsuit*, ABC7 (Oct. 9, 2023), https://abc7chicago.com/randal-quran-reid-ai-artificial-intelligence-dekalb-county-georgia-ga/13860699/ [https://perma.cc/WQ37-MGV2].

5. AI Works in the Dark

Another reason disclosure is not an adequate remedy to the harm caused by AI-influenced decision-making is that its operations are hidden from view.[524] An added layer of difficulty in disclosing possible harms caused by use of AI is that it is a very new technology. Consequently, people creating and even using the technology may know as little of potential harms as those at risk of experiencing these harms.[525] This has been described as the "black box problem" of AI algorithms, both in terms of lack of access to how they work because they are proprietary to their developers and because of the strong possibility of unanticipated harm.[526] It is, of course, impossible to explain or disclose a process that even its developers do not understand.[527] Carter and others have analogized the current lack of knowledge about how AI achieves its results to the early days of experimentation with radiation.[528]

---

524. Explaining this barrier to effective regulation Professor Sasha Luccioni, tweeted that because "[m]ost of these AI systems are *closed-source*[,] ChatGPT can literally be 3 raccoons in a trench coat, and we wouldn't be the wiser. . . . [T]here is no way to study [how AI makes decisions] from a scientific perspective, since we don't know that's in the box." @SashaMTL, TWITTER (Mar. 2, 2023, 2:25 AM), https://twitter.com/sashamtl/status/1631239223020855298?s=12&t=jGez-wA9-CXdPYzyxGUJQw [https://perma.cc/X7KE-69F9].

525. Carter, *supra* note 241, at 305 ("[T]he complexity and relative newness of AI can conceal ethical problems from even ethical users of technology."); *see also* Mark L. Shope, *NGO Engagement in the Age of Artificial Intelligence*, 28 BUFF. HUM. RTS. L. REV. 119, 138 (2022) ("[T]hose utilizing AI systems, or those reviewing the fairness of these AI systems, including human rights defenders, do not usually have the technological know-how to identify the discrimination or bias issues in these systems.").

526. *See* Charlotte A. Tschider, *Legal Opacity: Artificial Intelligence's Sticky Wicket*, 106 IOWA L. REV. ONLINE 126, 129 (2021) (citing PASQUALE, *supra* note 481, at 6–7) (citing Frank Pasquale who noted that opacity also prevents review by "regulators charged with protecting" those affected by automated processing); *see also* Danielle K. Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 10 (2014).

527. *See* ALEXANDRE DE STREEL ET AL., THE BLACK BOX: WHEN LAW CONTROLS AI 5 (2020), https://cerre.eu/wp-content/uploads/2020/03/issue_paper_explaining_the_black_box_when_law_controls_ai.pdf [https://perma.cc/ME37-MEUZ] ("Black-box models are models that that are not easy to understand because their mathematical expression is neither straightforward nor easily representable in an understandable manner.").

528. Carter, *supra* note 241, at 305.

6. The Right to Informed Consent for Research Cannot Be Waived: The Prohibition Against Exculpatory Clauses

In general, disclosure has not proved to be an effective remedy for preventing the kinds of harm that emerge when a researcher's interests in completing a study conflict with the individual interests of prospective participants.[529] Another thing that distinguishes the concept of obtaining informed consent for treatment from informed consent for research is that there are no circumstances under which consent for research can be waived entirely.[530] Because it is impossible for participants in FDA-regulated drug trials to waive their right to receiving any aspect of informed consent, they cannot consent in advance—as has been proposed for the use of AI in providing clinical care.[531]

For example, the FDA permits treating critically ill emergency room patients with an experimental drug when there is no other safe or effective treatment available—the criteria for selecting subjects and administering the drug must be approved in advance by an IRB.[532] Thus, the remedy proposed early on in AI's integration into health care that seeks advance

529. *See generally* Layla G. Maurer, *Managing the Medical Matrix: A "DAIS" for Artificial Intelligence in Health Care (and Beyond)*, 13 CASE W. RESERVE J.L., TECH. & INTERNET 107, 137 (2022) (citing RUSSELL T. VOUGHT, OFF. OF MGMT. & BUDGET, M-21-06, GUIDANCE FOR REGULATION OF ARTIFICIAL INTELLIGENCE APPLICATIONS ¶ 8, at 6 (2020), https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf [https://perma.cc/52VR-8QMG]) (providing an overview of articles calling for disclosure as a remedy to possible bias from AI in health care settings).

530. *See Waivers of Consent*, PENN: HUM. RSCH. PROTS. PROGRAM, https://irb.upenn.edu/homepage/biomedical-homepage/guidance/recruitment-and-consent/waivers-of-consent/ [https://perma.cc/W6T5-DQHT].

531. *See* Daylian M. Cain & Mohin Banker, *Do Conflict of Interest Disclosures Facilitate Public Trust?*, 22 AMA J. ETHICS 232, 233 (2020); Sunita Sah & George Loewenstein, *Nothing to Declare: Mandatory and Voluntary Disclosure Leads Advisors to Avoid Conflicts of Interest*, 25 PSYCH. SCI. 575, 582 (2014).

532. Even research studies that must be conducted in emergency settings such as an immediate need for a blood transfusion require prior IRB approval. *See* Matthew Stonecipher, *Waiver of Consent in Medical Procedure Research*, 9 AMA J. ETHICS 123, 123 (2007) (explaining that although the Polyheme trial was approved by the FDA, many questioned its ethical validity because it "was not in accordance with a plain reading of federal regulation of waived consent research because, once the patients-subjects reached the hospital, a standard and effective treatment—blood—was available but was not given to them. The Northfield case exposed ambiguity in interpretation of the FDA regulations that undermines the intent of that agency's narrow waiver of informed consent in specific types of research"); *see also* Ken Kipnis, Nancy M.P. King & Robert M. Nelson, *An Open Letter to Institutional Review Boards Considering Northfield Laboratories' PolyHeme® Trial*, 6 AM. J. BIOETHICS 18, 18 (2006) ("The authors argue that the in-hospital stage of the study fails to meet ethical and regulatory standards.").

consent from patients cannot be transferred to a research setting.[533] Whether or not AI should play a role in obtaining informed consent for health care is a question that deserves considerable study and review.

Recent work suggests that it may be possible for AI to anticipate what medical treatment an individual would and would not consent to more accurately than a surrogate decision maker. But even if this were possible, it would be inappropriate to use such a tool to obtain consent for research.[534] Reports of AI's deployment as a cost-saving measure to determine when patients are or are not likely to benefit from future medical care raises concerns about whether its advice is solely in the patient's best interest or whether the interests of the funder are incorporated in a way that would not be appropriate in decisions made by human physicians.[535] These concerns are reflected in interviews with patients who remain apprehensive about whether AI will accurately anticipate their needs and act in their best interests.[536] Some patients in the study reported that, while AI could be a useful tool, in the end they relied upon their treating physician to act in their individual best interests.[537]

So, whether or not AI has a role to play in decision-making for health care, the Common Rule's prohibition on what are usually called "exculpatory clauses" means that participants in research can never waive their right to

533. *See* Maximilian Kiener, *Artificial Intelligence in Medicine and the Disclosure of Risks*, 36 AI & SOC'Y 705, 712 (2021).

534. *See* David Wendler, *A Call for a Patient Preference Predictor*, 49 CRITICAL CARE MED. 877, 878 (2021); *see also* Scott Bay, *How AI Could Improve the Quality of End-of-Life Care*, VENTURE BEAT (June 29, 2018, 2:10 PM), https://venturebeat.com/ai/how-ai-could-improve-the-quality-of-end-of-life-care/ [https://perma.cc/FUD7-WQAH]; Nikola Biller-Andorno, *Up Close and Personal: Using AI to Predict Patient Preferences?*, BMJ (Mar. 10, 2021), https://blogs.bmj.com/medical-ethics/2021/03/10/up-close-and-personal-using-ai-to-predict-patient-preferences/ [https://perma.cc/JU5V-94LZ].

535. *See* Bay, *supra* note 534; *see also* Biller-Andorno, *supra* note 534.

536. *See* TYSON ET AL., *supra* note 340 ("Six-in-ten U.S. adults say they would feel *uncomfortable* if their own health care provider relied on artificial intelligence to do things like diagnose disease and recommend treatments; a significantly smaller share (39%) say they would feel comfortable with this."); *see also* Chiara Longoni & Carey K. Morewedge, *AI Can Outperform Doctors. So Why Don't Patients Trust It?*, HARV. BUS. REV. (Oct. 30, 2019), https://hbr.org/2019/10/ai-can-outperform-doctors-so-why-dont-patients-trust-it [https://perma.cc/TC54-WVPW].

537. Jordan P. Richardson et al., *Patient Apprehensions About the Use of Artificial Intelligence in Healthcare*, 4 NPJ DIGIT. MED., Sept. 21, 2021, at 1, 1 ("Participants reported that they felt their clinicians should act as a safeguard to buffer patients from the potential harms that might result from mistakes made by healthcare AI.").

informed consent.[538] As guidance provided to the Yale research community explains,

> According to the Common Rule, the prohibition of exculpatory language means that there can be no "language through which the subject or the representative is made to waive or appear to waive any of the subject's legal rights, or releases or appears to release the investigator, the sponsor, the institution or its agents from liability for negligence."[539]

A general rule is to state the situation simply and factually, such as, "[y]ou do not give up your legal rights by signing this form."[540] This includes their right to informed consent.[541] Therefore, disclosure that Persuasive AI may interfere with their decision-making process will be used is not sufficient because federal law prohibits exculpatory clauses.

IX. RECOMMENDATIONS AND CONCLUSION

AI's present ability to manipulate human decision-making is antithetical to the promise made to the American public in general, and the survivors of the U.S. Public Health Service Study of Untreated Syphilis specifically, that never again would people be used as a means to advance science or develop commercial products. The ability of AI to alter its message in response to the reactions of the person in front of them, which is already being advertised to increase enrollment in clinical drug trials, violates federal law requiring pre-approval of all information presented to potential participants.

Banning the use of technology designed to persuade consumers from a process intended to protect free choice should be an easy decision across the spectrum of federally protected human participant research. It is especially important to do now because of the intense pressures on drug manufacturers to recruit diverse participants to enroll in the very studies they have traditionally been wary about.

---

538. *See* Off. for Hum. Rsch. Prots., *Guidance on Exculpatory Language in Informed Consent*, U.S. DEP'T OF HEALTH & HUM. SERVS., https://www.hhs.gov/ohrp/regulations-and-policy/requests-for-comments/guidance-exculpatory-language/index.html [https://perma.cc/MJV7-L987].

539. *Prohibition of Exculpatory Language*, YALE UNIV.: HUM. SUBJECTS PROT., https://assessment-module.yale.edu/human-subjects-protection/prohibition-exculpatory-language [https://perma.cc/KER7-TEML].

540. *Id.*

541. *See* Off. for Hum. Rsch. Prots., *Exculpatory Language in Informed Consent Documents: Examples of Acceptable and Unacceptable Language (OPRR Letter, 1996)*, U.S. DEP'T OF HEALTH & HUM. SERVS., https://www.hhs.gov/ohrp/regulations-and-policy/guidance/exculpatory-language-in-informed-consent-documents/index.html [https://perma.cc/V8TP-HNVD].

The need for rapid modification to the process of conducting human subject research during the first two years of the COVID-19 pandemic demonstrates the ability to make rapid change when circumstances demand.[542] The growing threat from Persuasive AI warrants a similar rapid response.

These entities should, once again, join forces and issue a call to stop the use of technology that interferes with the ability of IRBs to provide advance review of materials designed to recruit and enroll participants in drug research trials. This includes technology designed to enhance the ability to persuade consumers, either by shaping the information presented or presenting in a form most likely induce enrollment. For the reasons discussed in this Article, the characteristics of this technology are such that anything short of a ban is insufficient to protect the decision-making process of those to whom the U.S. government has already made its most solemn pledge to protect.

542. *See* Bryan A. Sisk et al., *Ethical, Regulatory, and Practical Barriers to COVID-19 Research: A Stakeholder-Informed Inventory of Concerns*, 17 PLOS ONE, Mar. 24, 2022, at 1, 17.