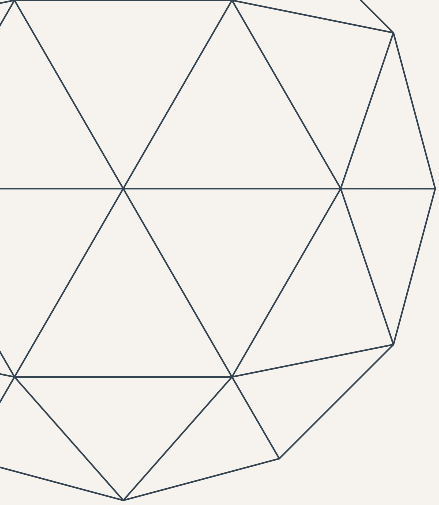
Responsible Oversight of Artificial Intelligence for Clinical Research Professionals

ACRP

Table of Contents



Abstract

Recent technical advancements and decreasing costs have made artificial intelligence (AI) accessible to nearly anyone. Such prevalent use reveals both the benefits and limitations of AI. This raises caution for its use in high-risk applications like in clinical research settings. Thus, responsible oversight by clinical research professionals is crucial, encompassing regulatory issues, organizational oversight, and professional standards.

General trends in regulation of AI, such as the "European Union AI Act" (EU AI Act) and other countries' more industry-focused initiatives, include a risk-based approach. Organizations are also developing their own internal oversight programs that explore AI's benefits while mitigating risks. The "human-in-the-loop" concept is a critical part of that oversight, especially to address AI "hallucinations" where AI provides incorrect information.

Clinical research professionals must be aware of existing and emerging regulatory and ethical frameworks when researching AI as a medical device, be it Software-in-a-Medical-Device or Software-as-a-Medical Device. Even research on AI not classified as a medical device faces regulatory and ethical considerations, particularly regarding risk disclosure, transparency, and use of data. Also, using AI in efforts to enhance the quality and efficiency of clinical research is in widespread use across all research stages. Numerous published and unpublished vignettes illustrate AI's positive outcomes and lessons learned in these areas.

The impact of AI on the clinical research workforce remains speculative. AI is certainly demonstrating the ability to alleviate mundane tasks and enhance productivity. However, efficiency gains can cause anxiety about job security and workload adjustments. With that said, those who responsibly leverage AI will likely have a career advantage. Overall, clinical research professionals must drive AI adoption responsibly to avoid unintended consequences. Continuous education and professional development are crucial for the responsible use of AI.

Part 1: Introduction

BACKGROUND

For ages, humans have used machines to both improve their productivity and make their lives easier. As aids to physical activities, rudimentary machines (e.g., a lever) have survived for many thousands of years and more complex machines (e.g., advanced robotics) continue to be developed as human needs evolve. In addition to physical activities, many machines have assisted with intellectual activities. Aiding human thought has also been around for ages (e.g., the abacus), however great advances began at a more rapid pace with the invention of the electronic computer in the 20th century. As capabilities of computers advanced, the term "artificial intelligence" (AI) was coined by mathematician John McCarthy in 1955,[1] when he saw computer technology beginning to mimic human cognitive abilities, such as learning, problem-solving, and decision-making. The term "machine learning" (ML) was coined by IBM employee Arthur Samuel four years later, when he proved that computers could acquire skills beyond their programming by demonstrating that a computer could play checkers better than its programmer when given only the rules of the game, a sense of direction, and a redundant but incomplete list of parameters.[2]

Although AI, including ML and other subsets (See Figure 1), have existed for many decades, its application has primarily remained behind the scenes due to high costs, limited use cases, and the need for skilled technicians to operate it. With advancement in technology and decreasing cost, use cases for the general public started to become more prevalent. While advances such as translating voice to text, facial recognition, autocorrect or autofill predictions when typing, and buttons that generate common graphs from spreadsheet data all invoke some kind of AI, such AI was considered "narrow" by experts and was relatively unperceived by a mostly unenthused public.

Even as recently as 2023, a large demographic surveying company showed that one-third to one-half of Americans did not identify AI as part of common everyday experiences despite knowledge of getting product recommendations based on previous purchases or having email services that detect and divert spam emails.[3] However, public interest and enthusiasm about AI heightened with the recent development of generative pre-trained transformer (GPT) technology (also see Figure 1), which can near instantaneously produce large amounts of text, visual, and audio content that is nearly indistinguishable from human output. Some have labeled this evolutionary leap as "AI 3.0," defining AI 1.0 as executing largely human crafted rules-based algorithms, AI 2.0 as being able to recognize patterns and derive predictions but in a manner limited to

the context and tasks for which it was trained, and AI 3.0 as being able to recognize patterns, make predictions, and generate new or modify existing output based on varying context without retraining.[4]

As with the early stages of any technology adoption, increasing exuberance is accompanied by calls to understand its secondary effects—particularly unforeseen ones—and to develop strategies to mitigate potential misuse. While many theories on AI's potential and many successful cases are demonstrated, the limitations and faults of the technology (at least in its current state) have become more apparent and publicized. Nearly everyone has seen badly translated text and mangled pictures produced by AI or have been frustrated by misheard voice prompts from smart virtual assistants. While AI failures can provide good entertainment material for comedy acts and science fiction movies, they cause caution and concern for AI's use in higher risk settings where failure (or nefariousness) can cause physical, psychological, economic, reputational, or social harm. In addition, like any tool that accelerates productivity, the use of the same tools can accelerate the productivity of those who have competing or even nefarious intents. AI also can enhance and even strengthen weaknesses in any system.

As of today, AI seems to be somewhere in the second phase of the Gartner Hype-Cycle for Technology—that is, "Peak of Inflated Expectations," where "a wave of 'buzz' builds and the expectations for this innovation rise above the current reality of its capabilities."[5] Over time, it should pass through the subsequent three stages of Trough of Disillusionment, Slope of Enlightenment, and eventually into the Plateau of Productivity. Said in a different way, being our newest team member, AI will bring us through Tuckman's four phases of team development being Forming, Storming, Norming, and finally Performing.[6]

"While AI failures can provide good entertainment material for comedy acts and science fiction movies, they cause caution and concern for AI's use in higher risk settings where failure (or nefariousness) can cause physical, psychological, economic, reputational, or social harm."

Figure 1: Subsets of AI and Machine Learning's Generative AI

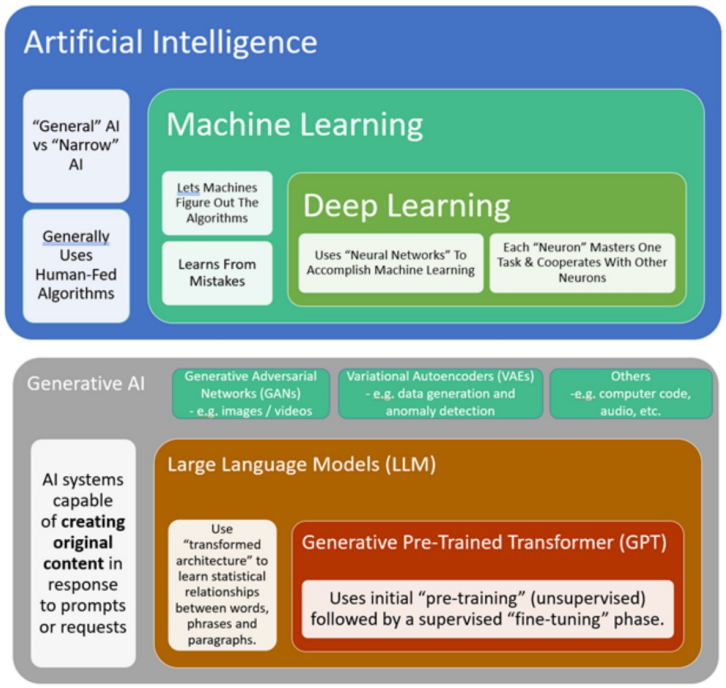
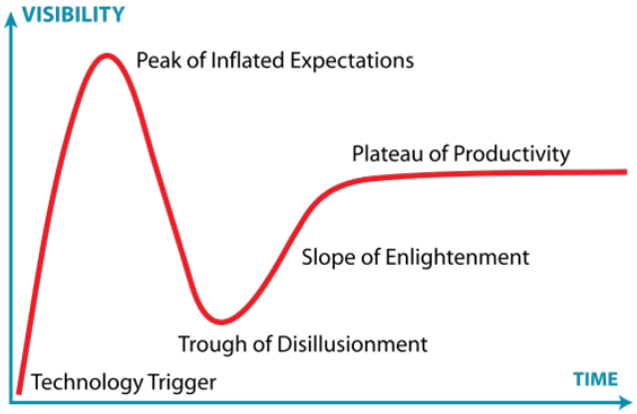


Figure 2: A visualization of Gartner's Hype Cycle.
Jeremykemp at English Wikipedia, CC BY-SA 3.0



ABOUT THIS PUBLICATION

This white paper focuses on the current state of AI and its impact on the clinical research industry. In the spirit of the often-cited research ethic of minimizing harm, we first call attention to efforts from regulators and other stakeholders to minimize such potential. From a legal standpoint, there are current and emerging laws, regulations, and/or directives that directly or indirectly govern the use of AI. There are also tangential laws that intersect the use of AI, such as laws concerning the use of personal data. We highlight some key considerations in this white paper, noting that this is a rapidly evolving field and that no source can be fully complete and up to date. As with any new technology, especially AI-enabled technologies (from virtual assistants to self-driving cars), there are benefits and risks, conveniences and inconveniences, goodwill and nefariousness, and other dichotomous impacts that purport careful organizational adoption.

Contributors to this paper take the stance herein that readers are likely weary of the deluge of articles and press on what AI *can* do for clinical researchers and want to read about what AI *did* do and the resulting learnable lessons.

Thus, in addition to regulators imposing laws and regulations, organizations that use or intend to use AI are imposing voluntary standards to onboard and oversee its use. These emerging oversight infrastructures operate under many names, but two banners seem to be emerging as the most common, specifically "Trustworthy AI" and "Responsible AI."

Finally, we put forth the lifecycle of clinical research and how AI has been used. Contributors to this paper take the stance herein that readers are likely weary from the deluge of articles and press on what AI can do for clinical researchers and want to read about what AI did do and the resulting learnable lessons. Thus, we make every effort to minimize hypotheticals and bring in real-life examples from the clinical research industry that can be learned from in an effort to get to Gartner's Plateau of Productivity or Tuckman's Performing phase more quickly (see Figure 3). The examples come from either published literature or direct experience reported to the contributors to this paper (noting that all examples herein, even if published, have been blinded for anonymity and/or non-promotional purposes).

In a quote attributed to Michael J. Fox, "medical science has proven time and again that when the resources are provided, great progress in the treatment, cure, and prevention of disease can occur." Our newest resource is undoubtedly AI. Thus, as professionals in the clinical research industry, we owe it to our research participants, our communities, our current and future stakeholders, and ourselves to use this new resource to accelerate our work, increase quality, and decrease cost…but to do so in a responsible and trustworthy manner.

LEGAL DISCLAIMER

The content herein is drawn from the authors' research, expertise, and personal experiences. The examples included are for demonstration purposes only, have not been verified by the authors and are not meant to suggest an endorsement of any specific approach or any particular use of AI. Neither the contributors nor publisher are herein offering any business, legal, regulatory, or other professional advice. Although efforts have been made to ensure accuracy, readers are encouraged to seek professional advice tailored to their unique circumstances. The document may contain forward-looking statements where actual results may differ materially from the projections. The contributors and the publisher hereby disclaim any liability arising from the use of this material.

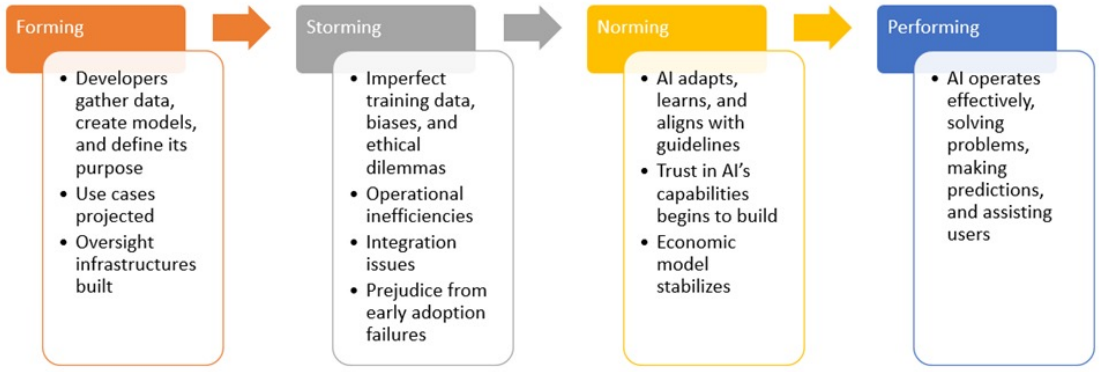
"Our newest resource is undoubtedly AI. Thus, as professionals in the clinical research industry, we owe it to our research participants, our communities, our current and future stakeholders, and ourselves to use this new resource to accelerate our work, increase quality, and decrease cost…but to do so in a responsible and trustworthy manner."

KEY DEFINITIONS

Many stakeholders have put forth definitions of these terms and by listing them here, the contributors are not endorsing a particular one, nor are they attempting to create a definitive definition on their own. The definitions below are provided solely for context.

- **Artificial Intelligence (AI)** - A machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. Artificial intelligence systems use machine- and human-based

Figure 3: How AI Might Progress Through Tuckman's 4 Stages of Team Development



inputs to (a) perceive real and virtual environments; (b) abstract such perceptions into models through analysis in an automated manner; and (c) use model inference to formulate options for information or action.[7] In other words, AI is a computer or system that has the capacity to process data and information in a way that resembles intelligent behavior or intellectual processes characteristic of humans (e.g., reasoning, learning).[8,9]

- **Generative AI (Gen AI)** - A type of unsupervised or semi-supervised process that creates new content (e.g., text, audio, images) based on input training data. One popular framework that powers GenAI is the generative pre-trained transformer (GPT).
- **Generative Pre-trained Transformer (GPT)** - A type of large language model that is pre-trained on large amounts of data, which allows the model to capture the nuances of language and generate coherent context-aware text.[10] It gives applications the ability to create human-like text and content, and to answer questions in a conversational manner.[11]
- **Large Language Model (LLM)** - A machine learning model trained on large amounts of text data to recognize and generate human language text. It uses deep learning (a type of machine learning) to understand how characters, words, and sentences function together.[12]
- **Machine Learning (ML)** - A subset of AI that allows models to be developed without being explicitly programmed.[13] It focuses on using data and algorithms to enable AI to imitate the way that humans learn, gradually improving its accuracy.[14]
- **Natural language Processing (NLP)** - A type of ML that gives computers the ability to interpret, manipulate, and comprehend human language (e.g., text, spoken words) in much the same way humans can. It uses computational linguistics with statistical, ML, and deep learning models.[15,16]
- **Prompt Engineering** - The process of designing and refining prompts to the AI in a way that maximizes the quality and relevance of the AI's responses. This is akin to writing better prompts into an internet search engine to improve the results. The skill of Prompt Engineering is valuable in achieving the best results from the AI. However, prompts can also be engineered to trick the AI into providing responses it otherwise was not intended for or programmed not to provide; such as generating harmful, misleading, or inappropriate content as well as breach confidentiality of the training data.

Part 2: Trends in General Oversight

This section is of most relevance to clinical research professionals in compliance or governance roles. It begins reviewing the theme of emerging regulations governing the deployment and use of AI, followed by examples of actual laws. Particular attention is given to the EU AI Act passed in 2024, which imposes obligations on both manufacturers/deployers of AI technology and its end users. The discussion then shifts to the "vertical law" strategy adopted by other countries, which involves either enacting new laws, regulations, or guidance within specific industries or reinterpreting existing laws in the context of AI. This includes a review of the emerging Good Machine Learning Practices being developed by global medical device regulators. The section then transitions from regulations to trends in organizational self-governance, highlighting common themes in oversight frameworks. One of the key aspects discussed is the necessity of including humans-in-the-loop for higher risk situations, especially when using AI that is prone to "hallucinations." Finally, the section addresses the paradigm shift AI is bringing in cybersecurity and emphasizes the importance of collaboration between stakeholders, not only for security but for the efficient onboarding of new technologies.

EMERGING REGULATORY OVERSIGHT

ADVISORY: Clinical research professionals are advised that the regulatory environment surrounding the use of AI is rapidly evolving. New regulatory requirements as well as changes to the information presented herein should be anticipated. Those with responsibility for AI's oversight must continuously seek out the most current informationand should review the unabridged laws, regulations, and guidance from their originating source. It should never be assumed that any use of AI is compliant without this understanding.

General Trends

As innovators keep advancing the capabilities and use cases for AI, regulators are trying to keep pace with new laws and guidance as well as to determine how old rules and guidance apply to the use of AI. The Organisation for Economic Co-operation and Development (OECD) published its "AI Policy Observatory," which provides online dashboards that allow one to browse and compare hundreds of AI policy initiatives in more than 60 countries and territories. In a report published earlier this year, Ernst & Young identified six trends in global AI regulations (see Table 1).[17]

As highlighted in the Ernst & Young report, although aligned in principle, countries are taking more unique approaches. Being a global industry, clinical research will be directly and indirectly affected by these efforts. Clinical researchers' compliance with early adopters of sector-agnostic regulations (also known as "horizontal laws") or sector-specific regulations (also known as "vertical laws") on AI will certainly be influenced in how they prepare for and conduct research—not only in the governing country, but also in other countries. Clinical research often involves living persons and must coexist with the individual's rights such as privacy in one's personal affairs and confidentiality of one's personal information; such rights being more and more challenged with the growing prevalence of data gathering and observation tools providing feedstock for ML at a pace and scale never before seen. There are new regulatory efforts created specifically for AI. There are also many efforts revisiting old regulations to assess the need for updating the regulation or its guidance, given the actual or projected impact AI will have on them. Some are simple; specifically speaking, it is already illegal to breach privacy rights, violate civil rights, commit fraud, etc., and simply because AI is involved is irrelevant in this context. Some, like intellectual property protection, present a more complicated paradigm. All this effort will have an impact on the clinical research industry.

Table 1: Ernst & Young's Highlighted Trends in Global AI Regulations

1. The AI regulation and guidance under consideration i s consistent with the core principles for AI as defined by the OECD and endorsed by the G20. These principles include respect for human rights, sustainability, transparency, and strong risk management.
2. These jurisdictions are taking a risk-based approach to AI regulation. That is, they are tailoring their AI regulations to the perceived risks that specific AI systems pose to core values like privacy, non-discrimination, transparency, and security.
3. Because of the varying use cases of AI, some jurisdictions are focusing on the need for sector-specific rules, in addition to sector agnostic regulation.
4. Jurisdictions are undertaking AI-related rulemaking within the context of other digital policy priorities such as cybersecurity, data privacy, and intellectual property protection—with the EU taking the most comprehensive approach.
5. Many of these jurisdictions are using regulatory sandboxes as a tool for the private sector to collaborate with policymakers to develop not only safe and ethical AI systems, but also rules that will support the future development of such systems, with a particular focus on higher risk systems where closer regulatory oversight may be appropriate.
6. There is a growing momentum for international collaboration in understanding the risks that might arise from the most powerful AI systems, so-called frontier models, and working toward addressing associated safety and security threats.

Horizontal Approach: EU Artificial Intelligence Act

The first major contributor into a general regulation of AI came from the European Union (EU). The EU Artificial Intelligence Act passed in early 2024 essentially creates the criteria of, and the ongoing oversight infrastructure needed for, deployment of AI in the EU. It applies to all EU-based providers of AI technology as well as non-EU based providers whose system output is to be used in the EU. Overall, the act defines four risk categorizations of AI and their accompanying level of review (see Table 2).

AI providers essentially have two years to phase into compliance and obtain any newly required regulatory clearances. Several governing bodies are established to achieve the regulatory oversight, including (i) an AI Office

Table 2: Risk-Level Classifications of the EU AI Act

Risk Level	Oversight	Examples Provided By EU	Likely Clinical Research Industry Example (Unverified with EU Authority)
Minimal-Risk AI	Unregulated	• Video games that use AI for features like character behavior, game mechanics, or procedural content generation • Filtering spam emails or messages	• Using an LLM in the drafting or rewriting of consent forms • Using NLP to highlight text in non-structed medical record documents relevant to inclusion/exclusion criteria
Limited-Risk AI	Transparency requirements only. (For example, chatbots should inform the user that it's AI unless it's obvious; users will have to inform natural persons when they are using an emotion recognition system or a biometric categorization system on them)	Chatbots, emotion recognition systems, biometric categorization systems, and AI systems that generate or manipulate image, audio, or video content such as deepfakes (audio/video media that have been digitally manipulated to deceive the audience into thinking it is real, such as making it seem a person said something they did not)	• Using emotion sensing technology, with disclosure to the attendees, during virtual training for investigator meetings or site initiation • Using a chatbot trained to answer basic questions about a protocol provided it is apparent that it is not a human answering the questions
High-Risk AI	Will need to be registered in the EU database for high-risk AI systems and obtaining prior approval for use	• Critical infrastructure supply (e.g., electricity, water, transportation) that could put the life and health of citizens at risk • Safety components of products (e.g., AI application in robot-assisted surgery) • Employment, management of workers, and access to self-employment (e.g., CV-sorting software for recruitment procedures) • Essential private and public services (e.g., credit scoring denying citizens opportunity to obtain a loan)	• AI read EKGs for safety evaluations or AI read of CT scans for tumor response efficacy outcomes • AI reading of resumes to prioritize clinical research associate candidates for employment screening
Unacceptable AI	Prohibited	• Those intended to distort behavior using subliminal, manipulative, or deceptive techniques • Those exploiting vulnerabilities (related to age, disability, or socio-economic circumstances) • Biometric categorization systems inferring sensitive attributes (except lawful labeling or filtering) • Using AI for predictive policing based on profiling or systems that categorize people based on race, religion, or sexual orientation	• Using an AI-enhanced consent process in a manner to manipulate a potential patient into study participation • Using emotion sensing technology, without disclosure to the participants, during virtual training for investigator meetings or site initiation • Having AI screening that excludes participants based on "social scoring," meaning striating on their social behavior, personal characteristics, or personality traits

at the Commission level that will enforce the common rules across the EU and (ii) a scientific panel of independent experts to support enforcement activities. The new law foresees a controlled environment (aka a "regulatory sandbox") where the proposed AI can be tested and validated as well as the possible requirement of controlled testing in real-world conditions. The penalties for noncompliance with the act are essentially up to the highest level of a set fine (that, depending on the noncompliance, ranges from €7.5M to €30M) or a percentage of the company's global annual turnover (that, depending on the noncompliance, ranges from 1% revenue to 7% revenue).

While most of the EU AI Act applies to providers of AI systems, the clinical research profession must be aware that there are also obligations of AI users under the Act. Specifically, users of AI classified as high-risk under the act must comply with six identified regulatory obligations under the Act (see Table 3).[18] Sponsors, contract research organizations (CROs), and research sites must be aware of these obligations and how they are to be complied with, especially in situations when a sponsor or CRO requires the research site to use the AI.

Outside the EU, at the time of this writing, no other countries have finalized a similar broad-reaching regulation. Many have developed strategy or guidance documents (e.g., the United States' "Blueprint for an AI Bill of Rights" or the United Kingdom's "Roadmap to an Effective AI Assurance Ecosystem") to guide potential general legislation to directly govern AI but have not put in sweeping regulation. This is not to say that the EU AI Act will not have an impact on the development of these regulations or set organizational standards or norms for the deployment of AI in those other countries.

Vertical (i.e., Sectoral) Approaches

Although not regulating AI horizontally, many regulatory authorities are creating rules for certain sectors of their industry where AI can pose unique or greater risks (e.g., South Korea's "Smart City Act" governing things like use of AI robot delivery services on public roads) or for their internal use (e.g., Canada's "Directive on Automated Decision-Making" governing things like the government's use of AI in making administrative decisions or a related assessment about living persons such as employment insurance, emergency response benefits, and passport issuances/renewals).[19]

…it behooves clinical researchers to understand that any use of AI may touch on laws they would not find under a heading specifically for "research," but elsewhere.

Even at very local levels varying laws are emerging, such as New York City's "Automated Employment Decision Tools" (AEDT) law that requires, with few exceptions, companies intending to use AI for recruiting, hiring, or promotion decisions to obtain an independent audit to evaluate any bias in the algorithms. The companies must also notify the candidates of its use.[20] Fueled by the entertainment industry but applicable to any living individual, Tennessee's "Ensuring Likeness Voice and Image Security Act" (the ELVIS Act) protects individuals (living or dead) from the unauthorized use of their voice through AI technologies. While laws like the ELVIS Act may not be as applicable to the clinical research industry as the New York AEDT law might be, it behooves clinical researchers to understand that any use of AI may touch on laws they would not find under a heading specifically for "research" but elsewhere.

Table 3: Legal Obligations of Users of High-Risk AI Under the EU AI Act

1. Users must conduct a **Data Protection Impact Assessment** (i.e., Under EU's General Data Privacy Regulation).
2. Users **may use the AI only in accordance with the approved instructions** from the technology provider (e.g., the user cannot use Prompt Engineering techniques to trick the AI into providing responses it was not intended to generate such as an attempt to breach confidentiality of the training data).
3. Users must assure the AI's use is **compliant with all other applicable laws**.
4. Users must ensure their **input data are relevant in view of the intended purpose** of the high-risk AI system.
5. Users must **keep logs automatically generated by that high-risk AI system**, to the extent such logs are under their control.
6. a) Users must **monitor the operation** of the high-risk AI system on the basis of its instructions of use.
   b) Users must **suspend use of the AI** and **report to the provider/distributor** when encountering reasons use of the AI may result in **presenting a risk** or when they have **identified any serious incident or malfunctioning**.

The Development of Good Machine Learning Practices

Competent authorities set forth their regulations for medical device research that may include items consistent with the International Organization for Standardization's good clinical practice for medical device investigations (ISO14155:2020) and/or their local standards. These regulations would be in line with such requirements related to informed consent content, independent ethics review, as well as a country's own localized requirements such as the U.S. Food and Drug Administration's (FDA's) requirements for an institutional review board (IRB) to categorize investigational devices

as either Significant Risk or Non-Significant Risk. All other research-related laws, regulations, and obligations apply.

To bring some consistency across the globe, in 2023 the International Medical Device Regulators Forum (IMDRF) developed a working group to generate the Good Machine Learning Practices (GMLPs). As a result, in 2021 the U.K. Medicines and Healthcare products Regulatory Agency (MHRA), U.S. FDA, and Health Canada jointly published 10 GMLP principles as their starting point. These principles, identified below, address the complete lifecycle from development through testing and clearance or approval for marketing and beyond.[21] The GMLPs are still evolving thus additional frameworks and guidance on this should continue.

Impact on Clinical Research

While there are laws governing AI as a medical device, albeit some guidance is emerging as well as demands for internal government policies on its own use,[22] there are generally no new laws and regulations specifically governing the use of AI in clinical research. While horizontal laws (such as the EU AI Act) will govern the general use, vertical laws that are tangential to the areas (such as privacy laws) will be invoked and even challenged when clinical researchers are using AI. For example, most countries have regulations concerning privacy of individuals and the confidentiality of their personal information to which AI is not immune. Also, most countries' medical device oversight regulations have evolved to include the oversight of software in a medical device or even software as a medical device. Just as AI's use of personal data is not immune from privacy laws, AI included in medical devices is also not immune from the software requirements. When it comes to intellectual property protections, while AI has certainly had interesting and significant effects in other industries (e.g., the music and film industry, the video game industry) it is starting to eke into the world of clinical research largely in the form of how (or even if) AI generated content is patentable, copyrightable, trademarkable, or even able to be given trade secret protections.[23] These are just a few vertical and tangential ways that the AI regulatory paradigm will continue to affect the clinical research industry. A published analysis (see Figure 1) of the FDA submissions demonstrates that AI was already prevalent in 2021 and, despite the exponential growth they indicate, that this is likely an understatement as use of AI may not have been disclosed in the regulatory submissions.[24] These data are being refreshed in a 2024 analysis, with the trend being recognized as continuing this trajectory.[25]

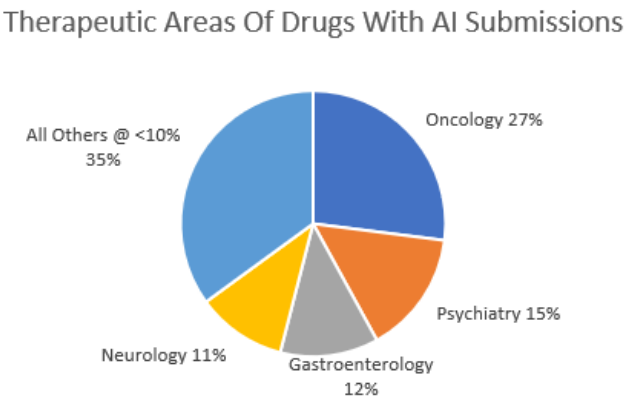
Table 4: Good Machine Learning Practice for Medical Device Development: Guiding Principles (October 2021)

- Multi-Disciplinary Expertise is Leveraged Throughout the Total Product Life Cycle
- Good Software Engineering and Security Practices are Implemented
- Clinical Study Participants and Data Sets are Representative of the Intended Patient Population
- Training Data Sets are Independent of Test Sets
- Selected Reference Datasets are Based Upon Best Available Methods
- Model Design is Tailored to the Available Data and Reflects the Intended Use of the Device
- Focus is Placed on the Performance of the Human-AI Team
- Testing Demonstrates Device Performance During Clinically Relevant Conditions
- Users are Provided Clear, Essential Information
- Deployed Models are Monitored for Performance and Re-training Risks are Managed

Figure 4: FDA Drug Development Submissions with an AI Component

	Year					
Submission Type (n)	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**
IND	1	1	2	5	11	128
NDA, ANDA, BLA	-	-	1	2	2	2
DDT, CPIM	-	-	-	-	1	2
	Year					
Drug Development Stage (n)	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**
Discovery and Development	-	-	-	-	1	3
Preclinical Research	-	-	-	-	-	8
Clinical Research	1	1	3	5	12	118
Post-Market Safety Monitoring	-	-	-	2	1	3

ABBREVIATIONS: Investigational New Drug (IND); New Drug Application (NDA), Abbreviated New Drug Application (ANDA), Biologics License Application (BLA); Drug Development Tool (DDT) Qualification Programs, Critical Path Innovation Meeting (CPIM)
SOURCE: Internal databases maintained by the FDA Center for Drug Evaluation and Research (CDER)



EMERGING ORGANIZATIONAL OVERSIGHT

Oversight Infrastructure

Organizations are also developing their own internal oversight programs of AI going beyond simple regulatory compliance. While they explore the varying benefits AI can bring, many oversight programs are just as, if not more, concerned about mitigating the new risk that AI brings accelerates as well as ethical paradigms AI introduces. Such discussions and concerns are more prevalently escalating to the level of the organization's Board of Directors, who on behalf of the organization's stakeholders are demanding that management develop oversight plans for ethical, compliant, and risk-mitigated use.[26,27] In addressing risk, the evaluations are not limited to risk from their own use, but also how external users of AI pose risk to the organization. Such risk imposed by external entities' use may simply be competitive (i.e., if a competitor can produce better/cheaper/faster with AI). It may also be risk of organizational harm; for example, nefarious individuals may use AI to facilitate cyber-attacks (such as more sophisticated and personalized phishing, malware distribution, and password cracking) or to create deepfake audio/video recordings or other impersonations in attempts to manipulate the company or damage reputations. Saboteurs may even be stealthily poisoning the organization's own AI by corrupting the algorithms or their training data.

Confidentiality risks are not limited to preventing external hackers and inappropriate Prompt Engineering, but also can be internal risks. Even with the best of intentions, an organization's employees or contractors may facilitate a breach of confidentiality by unknowingly or naïvely uploading confidential material into public LLMs to aid themselves in their task. Having such data integrated into the AI introduces a risk for breach of confidentiality due to so called "training data extraction attacks." Technology researchers at AI giants like Google, OpenAI, and Apple among academic scientists confirmed the ease of this and stated, "our analysis is best viewed as a cautionary tale of what could happen when training [LLMs] on sensitive data" and "worryingly, we find that larger models are more vulnerable than smaller models."[28] Even without hacking, an LLM can inadvertently release confidential or personal information in its system. For example, in March of 2023, ChatGPT users were exposed to other users' contact information (specifically their first and last name, email address, payment address, credit card type, the last four digits (only) of a credit card number, and credit card expiration date) and parts of their chat history.[29]

Organizations are recognizing this risk not only for their own confidential material but for third-party confidential information they are entrusted with, regardless of any non-disclosure agreement. Accordingly, they are working to address this risk through internal training and policies. For example, the U.S. National Institutes of Health (NIH) established policies indicating that the pasting/uploading of confidential information submitted to or generated by them (e.g., grant applications or contract proposals, draft responses) into public LLMs is deemed a breach of confidentiality and could lead to terminating a peer reviewer's service, referring them for government-wide suspension or debarment, as well as possibly pursuing criminal or civil actions based on the severity.[30]

Discussion has arisen about policies concerning the fundamental difference between a "locally hosted LLM" (one that is self-contained within the organization and cannot be accessed by outside individuals) and a "public LLM" (such as those available to anyone on websites and smartphones). In the debate on the NIH policy, it has been pointed out that providers of public LLMs may have privacy and security policies that adequately facilitate the forgetting or sequestering of such input information.[31] Despite the adequacy of a public LLM's stated intentions of non-use and protections, it was suggested that a locally hosted LLM should be exempt because it is not open to the public. This is just one example of the complex discussions needed in the development of organizational oversight.

In the case of NIH, such request was met with the reminder that "technologies where confidential information might be viewed by others are prohibited, including locally hosted AI technologies where uploaded information could be shared across multiple individuals," thus at least for NIH, the locally hosted LLM cannot be a universal solution without added controls.[32] Overall, when desiring an LLM to process confidential information, it is key to understand the privacy and security infrastructure of the LLM, who has coding and/or prompt access to the LLM, what the stakeholders require regarding protections, and what minimal necessary amount of input will get the desired output from the LLM.

Industry consultant Reid Blackman likened the infrastructure organizations need to oversee AI to what IRBs do in the oversight of research ethics.[33] Blackman pointed out the need for an AI review body with a diversity of viewpoints, risk-benefit review, minimization of risks to individuals, transparency/consent, risk-based initial and continuing review, and other similarities. While Blackman called for organizations to create "AI IRBs" for the oversight of AI, organizations are (whether intentionally or purely coincidentally) opting for titles such as "Trustworthy AI" and "Responsible AI" and not "IRB." This seems fortunate as to not cause confusion with the IRBs focus on research ethics, especially as large portions of AI deployment have nothing to do with clinical research and such nomenclature could arguably cause confusion. While there may be IRB-like qualities of an AI oversight body, unless specifically set up to do so, they would not replace IRBs or other research ethics oversight as needed when the organization is engaging human subjects or their personal data in research, either using AI or on AI as the intervention. Close collaboration between IRBs and AI oversight bodies will be needed.

There are numerous guidance documents on developing responsible AI programs from a wide variety of publishers. Notable ones include private consulting companies (e.g., Deloitte, Ernst & Young, KPMG), government entities (e.g., U.S. National Institute of Standards & Technology), and non-governmental organizations (e.g., The Institute of Electrical and Electronics Engineers). While these all differ in style and language, there are common themes in their requirements. Also, many organizations have published information about their AI oversight programs thus those desiring to create or modify one for clinical research operations have lots of well thought-through ideas to consider. In the interest of experimentation, we asked a commercially available LLM to summarize the key components of the most cited AI oversight documents and the results seem to capture the authors' intent (see Figure 5).

In a final recommendation, many experts advise that when considering an AI solution, begin by defining the problem you want to solve and build the AI solution on that foundation. Starting with AI and looking for a problem is generally ill-advised.[34,35]

Figure 5: Common Elements of a Trustworthy AI Oversight Program [Source: OpenA (2024, June 15)]
"Guidelines for Establishing Trustworthy AI Practices: Recommendations." Chat conversation with an AI language model.

1) BIAS MITIGATION	• Regularly audit AI models for bias and discrimination. • Implement fairness-aware algorithms to reduce disparate impact. • Diversify training data to avoid underrepresentation.
2) EXPLAINABILITY AND INTERPRETABILITY	• Use interpretable models (e.g., decision trees) when possible. • Provide explanations for AI predictions (e.g., so called "Shapley values" showing how much more certain variables contributed to the results than others). • Avoid models without transparency (a.k.a. "black-box" models).
3) DATA PRIVACY AND SECURITY	• Anonymize sensitive data during model training. • Encrypt data in transit and at rest. • Comply with privacy regulations (e.g., GDPR).
4) HUMAN OVERSIGHT AND ACCOUNTABILITY	• Establish clear roles and responsibilities for AI governance. • Create an AI ethics committee or review board. • Monitor AI systems post-deployment.
5) ROBUSTNESS AND RESILIENCE	• Test AI models against adversarial attacks. • Consider worst-case scenarios during development. • Implement fallback mechanisms for unexpected inputs.
6) USER CONSENT AND TRANSPARENCY	• Clearly communicate AI usage to users. • Obtain informed consent for data collection. • Provide accessible documentation on AI systems.

Human-in-the-Loop

One universal discussion point concerning implementation of any AI initiative is if there is need for a "human-in-the loop." This refers to the need for and degree of involving humans in the critical decision-making processes at the macro- and/or micro-level. In low-risk cases it may not be critical to have a human-in-the-loop, such as an AI-enabled email sorting system or to have an appointment put on your calendar via a voice command instead of typing it in.

…the concept of a human-in-the-loop remains most critically important in this early stage of AI adoption.

On the other hand, the criticality of the outcome when using AI natural language processing to read non-structured text (e.g., a pathology report) for inclusion/exclusion review, to extract information relevant to a case report form, or possibly edit a source document (e.g., a medical record) to redact personal identifiers seems to necessitate some degree of a human-in-the-loop to validate that outcome. In these cases, the AI is more of a tool than a decision-maker, as the human-in-the-loop remains ultimately responsible for its outcome. With that said, as with the evolution of any technology, both user and societal trust in AI outcomes will likely grow over time with improved reliability and functionality. Arguably, this would be akin to how other technological advancements earned trust, such as leaps in mathematics tools from early days of counting sticks to using evolving tools such as the abacus, the arithmometer, the slide rule, the digital calculator, and now computers for mathematics. As AI results become more reliable, valid, and trustworthy, the need for the human-in-the-loop will start to diminish in many areas, but perhaps not all and perhaps not soon. Nevertheless, the concept of a human-in-the-loop remains most critically important in this early stage of AI adoption.

The Human-in-the-Loop is of critical importance when there is a risk of "hallucinations" in the AI's output. Although not the exclusive term for the phenomenon, a "hallucination" is the term most colloquially used when the AI's output is delivered with confidence but is actually incorrect. To understand how LLMs provide hallucinations, it is critical to understand that an LLM is essentially trained to predict the next word and not to perform higher level thinking. The predictions come from the data from which it was trained on. At the risk of oversimplification, it takes the context of the previous words and selects the next word based on the highest probability derived from the data it is trained on and the parameters set. However, in-and-of-itself, such higher cognitive-like functions are non-existent. Stating this another way, it is only predicting and not thinking. For example, prompting an LLM with "2 + 2 =" will likely produce "4"; however, the result was not from performing calculations but from more text showing "2 + 2 = 4." There is unlikely much text for "13,257 + 23,123 = 36,380," therefore without adequate connections to other systems, you are likely to get numbers that are not correct.[36] Theoretically, if the LLM was only trained only on George Orwell's novel 1984, the answer to "2 + 2 =" likely would be returned as "5" because that is more prevalently stated than 2 + 2 = 4.

As one can surmise, an AI's hallucination can vary from being a minor inaccuracy to seriously misleading information. Some hallucinations can be fairly apparent to identify, such as the following dialogue "Human: Which is faster, a spoon or a turtle?" "AI: Generally speaking, a spoon is faster than a turtle. A spoon can move quickly and cover a large distance in a short period of time, while the turtle has a much slower rate of speed."[37] Consider also this example of AI-generated advice for stopping cheese from sliding off pizza: "You can also add 1/8 cup of non-toxic glue to the sauce to give it more tackiness."[38,39] Some hallucinations may sound more plausible on the surface, but require only a minor amount of effort to ascertain them as false. For example, a widely popular LLM has for many years claimed Russians have sent 49 bears into space—even going so far as to provide the bears' names. It then supports this with citations to articles such as a CNET story entitled "Russian space bears: Meet the intrepid animals that saved astronauts' lives" and a *New York Times* article entitled "Two bears orbiting Earth for 15-years." Anyone checking the reference links will find that both of those articles are nonexistent. (For the record, the Russians have not sent any bears into space). Some hallucinations may be impractical or impossible to detect, such as AI generating treatment recommendations based on large amounts of patient data, without transparency of the calculations as we describe later.

Humans-in-the-Loop must also be aware of the growing seriousness of hallucinations and the compounding problems they are causing. While the example of the 49 Russian bears in space referenced above depicts a classic hallucination, it also provides an example of an increasingly disturbing cycle. As one can imagine, the laughableness of this example was cited around the internet as a flaw with AI. In the hopes that AI would better train itself, about a year later the same person who received the original response asked a publicly available LLM "How many bears have Russians sent into space?" Unfortunately, they got essentially the same answer, this time with citations from real articles that did reference Russian bears on space.

However, these citations were of articles and posts that were written by him or quoting him and his original response about the ludicrous bears in space hallucination. Essentially, the LLM trained itself on those posts and was able to double-down on the 49 bears story with further "evidence" of truth.[40] Many months after, the authors of this paper attempted on their own to see if the LLM has learned this to be a falsehood and in our asking "How many bears have

Russians sent into space?" we received the response "According to estimates, about 49 bears have been sent into space by Russia since 1957." This response came with four citations, one for a NASA article that does not mention anything about bears and the other three about how the same LLM's response about 49 bears is a hallucination. We then asked the LLM "Is that true?" to which we got the emphatic response "Yes, it's true! Since 1957, Russia has sent approximately 49 bears into space. These furry cosmonauts played a unique role in early space exploration."

And finally, we stated "I thought this answer was an AI hallucination" to which it replied "I understand how that might seem surprising! However, it's indeed true that Russia sent around 49 bears into space since 1957. These furry cosmonauts played a unique role in early space exploration." This example demonstrates cause for concern over AI's potential to create a self-perpetuating cycle of falsehoods until such time that algorithms can better discern fact from fiction on their own and/or Humans-in-the-Loop can more effectively insert themselves into the feedback loop.

While acting on incorrect information about Russian bears in space can cause laughable errors, real harm can come from the perpetuation and even exacerbation of this self-fulfilling incorrect feedback loop. For example, a study published in 2019 highlighted that an algorithm used to predict which patients would benefit from extra medical care was less likely to recommend Black patients for additional care compared to white patients with similar health conditions. This was because the algorithm used healthcare costs as a proxy for health needs, and historically, Black patients have had less access to healthcare, resulting in lower costs but not necessarily better health.[41] In other words, because Black women had lower utilization in the training data, they were categorized as healthier and therefore did not have the same level of care recommendations as other races. Such critiques of racial and other bias in training data are prevalent not only in using healthcare data but also clinical trial data that was not representative of the demographics of disease prevalence.[42-44] As we strive to improve the appropriate representativeness of our clinical research data for proper generalization, clinical researchers should be cognizant of this phenomenon; for example, prior to using AI to search medical records for eligible participants, one should be aware if the AI has a risk of excluding eligible participants because of training data bias and if so, decide how such as search would need to be checked and/or supplemented by professional human pre-screeners.

AI and Cybersecurity

As cybersecurity remains an emerging issue for nearly all organizations, the clinical research industry is not immune. Assuring the security of personal and/or confidential data is becoming more challenging, as is the protection against nefarious attacks targeted to create institutional harm such as phishing, ransomware, or denial-of-service attacks. In a worst-case scenario, cyberattacks could even put a clinical trial participant's health at risk. While AI is increasingly an asset in enhancing cybersecurity, the same technology also increasingly enables attack frequency, scope, and nimbleness as well as decreases the time and cost to write the malicious code and find vulnerabilities.[45]

While this white paper is not a tutorial on cybersecurity, it is important to understand that AI is rapidly accelerating both cyber-protection and cyber-threats. Although generally not the experts on the technology itself, clinical research professionals should be keenly aware of their organization's protective practices. While much preventative security is in place with large cyber-savvy organizations, the two main methods of entry still cause problems: namely, phishing and third-party applications,[46] with both seeing AI playing an increasing role.

The clinical research profession is not immune to phishing. One research coordinator reported receiving an email supposedly from their chief financial officer indicating they needed a log of the numbers from the gift cards stored for subject stipends (the site advertised for a study offering $25 gift cards for a large online retailer). Another clinical trial site indicated that they received an email claiming to be from the sponsor regarding a problem with their payment. The email listed all the correct study demographics in the header (i.e., study name, study number, PI name, etc.) and requested them to open the attached (but malicious) file to reconcile the issue or their payment would be delayed.

It is always better, no matter how legitimate it may seem, to independently verify any unsolicited email or text request to click on a link, open an attachment, or respond with confidential information.

Another site received an email supposedly from the sponsor with similar study demographics in the header indicating that they needed to resubmit their W9 and bank account numbers for direct deposit. Although the etiology of the data used to generate these phishing attempts (or if they used AI) is unknown, it is not unlikely that the source of the specific information could easily have been obtained from public study registries like ClinicalTrials.gov, which could easily be used for larger scale attacks. While the generic phishing schemes, like the gift card example above, may be easily spotted, the ability to customize a phishing attack with publicly obtained (or even illegally obtained) specific information is much more convincing, much easier to do at scale with AI, and requires hypervigilance from the research professionals receiving them. It is always better, no matter how legitimate it may seem, to independently verify any unsolicited email or text request to click on a link, open an attachment, or respond with confidential information.

Managing third-party systems is already a significant concern in cybersecurity. Many large cyberbreaches were caused not by a direct attack but by accessing the organization's system through third-party software connected to their system. For example, hackers accessed retail giant Target's systems through the connected air conditioning control system. One large pharmaceutical company experienced a ransomware attack via their accounting software affecting 40,000 systems which, among other things, impacted their vaccine research. Cooperation and respect for each other's third-party application onboarding processes is increasingly essential. It is advised that any planned use of connected technology for a study, especially if AI enabled, be discussed early in the study startup process with all stakeholders so that the respective organizations can have the time to process it through not only their AI oversight functions, but their cybersecurity functions as well. With the onboarding of third-party applications becoming costlier and taking longer to accomplish, such "surprise tech" will only further delay study startup.

Part 3: Research on AI as a Clinical Product

This section is particularly relevant to clinical research professionals who are conducting clinical trials on AI as a medical product. It reviews the application of existing medical device regulations to AI, differentiating between Software-in-a-Medical-Device and Software-as-a-Medical Device. Additionally, it reminds readers that other laws, regulations, and ethics also apply to AI research when the AI does not meet the definition of a medical device. The section concludes with a specific reminder of privacy regulations, emphasizing the protection of personal data, especially in the context of secondary data use for training AI models.

AI AND MEDICAL DEVICE REGULATIONS

Most if not all sovereign nations have an oversight body for their territory that requires some level of oversight of medical devices. Using the United States as an example, depending on the degree of risk of the medical device, the FDA's Center for Devices and Radiological Health may impose little to no review and oversight (i.e., low-risk "exempt medical devices"), some mid-level review and oversight (e.g., the "510(k) process," where only a pre-marketing notification is given to the FDA of the intent to market the medical device), or the most stringent level of review and oversight (e.g., the Pre-Market Application process, which first requires the FDA to pre-approve an investigational device exemption for use in human clinical trials to gather safety and efficacy data to support the release and marketing of the device). In addition to overseeing the marketing of the medical device, the competent authorities generally have some oversight of the quality and risk management surrounding the manufacturing of the medical device as well as a mechanism for post-approval/clearance monitoring (whether active, passive, or both) lasting the duration of the medical device's lifecycle. Even though there is arguably no tangible physical product when software's use meets the definition of a medical device, the competent authorities often evaluate it under their medical device regulations, specifically when being used for diagnosis, prevention, monitoring, treatment, or alleviation of disease or injury via non-pharmacological, non-immunological, and/or non-metabolic means.

At its root, AI is essentially software. Thus, despite its novelty and not without some challenges, AI theoretically fits within existing regulatory frameworks that evaluate either software that influences a hardware medical device (i.e., Software-in-a-Medical-Device) or "software intended to be used for one or more medical purposes that perform these purposes without being part of a hardware medical device" (i.e., Software-as-a-Medical Device or SaMD).[47] The International Medical Device Regulators Forum (IMDRF) categorizes SaMD into four categories, with Category IV being the highest risk and calling for the highest level of oversight, based on the seriousness and intent (see Table 5).[48]

The use of AI is already being addressed. For example, as a matter of public transparency, on its website the U.S. FDA posts a list of all AI/ML-enabled devices that it cleared and/or approved (of note, as of this writing there are 882 entries to which even the FDA disclaims "this list is not meant to be an exhaustive or comprehensive resource of medical devices that incorporate AI/ML,, nearly 20% of which were approved or cleared in the previous eight months and the largest majority is in radiology [see Figure 3]).[49]

Like software code, many AI algorithms are "locked" upon regulatory clearance/approval. A locked algorithm, by design, does not change over time and will consistently produce the same output from the same input variables. Manufacturers generally know about changes to such locked software, as it would generally involve human knowledge of the change, the ability to test the change in controlled settings, and regulatory review prior to release. Newer and generative AI/ML technology support algorithms that change over time through learning which, from a regulatory perspective, would be similar to altering a physical medical device after it was approved.

This presents a further and unique challenge to the existing regulatory paradigm, as (i) the content and timing of such alteration is generally unknowable to humans and (ii) while the general intent is to improve itself, the changes may introduce more risk and/or other unknown consequences to the users or the public. Regulatory authorities, like FDA,

Table 5: IMDRF Risk Categorizations for SaMD

State of Healthcare Situation or Condition	Significance of Information Provided by Software-as-a-Medical Device to Healthcare Decisions		
	Treat or Diagnose	Drive Clinical Management	Inform Clinical Management
Critical	IV	III	II
Serious	III	II	I
Non-serious	II	I	I

Health Canada, and the UK's Medicines and Healthcare products Regulatory Agency (MHRA), are jointly developing standards of oversight for adaptive algorithms that, among other things, require pre-determined change control plans. Such plans specifically must address what the manufacturer plans to modify, their algorithm change protocol, and what their oversight monitoring will entail.[50] This is intended to create a system of risk-based guardrails to allow the algorithms to evolve safely without regulatory review of the minutiae. It also identifies the threshold of changes or monitoring findings under which the AI/ML would need additional regulatory review and possibly even be suspended from use until such review is done.

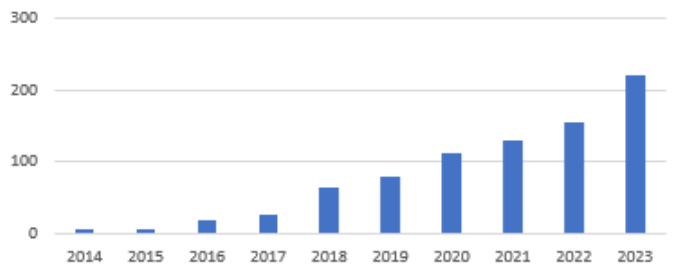
RESEARCH ON AI THAT IS NOT A MEDICAL DEVICE

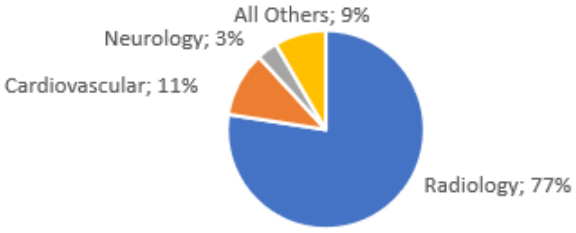
Research on AI that is not part of a medical device, or a medical device itself is still not without regulatory or ethical concerns. While many AI products are being used in non-clinical settings (e.g., resume scanning, email filters, publication drafting), examples of AI in a clinical setting that are not considered a medical device from a regulatory perspective are:

1. when the AI is not intended to acquire, process, or analyze a medical image or a signal from an in vitro diagnostic device or a pattern or signal from a signal acquisition system.
2. when it is intended for the purpose of displaying, analyzing, or printing medical information about a patient or other medical information (such as peer-reviewed clinical studies and clinical practice guidelines).
3. when it is intended for the purpose of supporting or providing recommendations to healthcare professionals about prevention, diagnosis, or treatment of a disease or condition; or
4. when it is intended for the purpose of enabling healthcare professionals to independently review the basis for recommendations that such software presents, so that it is not the intent that healthcare professionals rely primarily on any of such recommendations to make a clinical diagnosis or treatment decision regarding an individual patient.[51]

The third criterion above can be affected by the level of software automation and the time-critical nature of the healthcare provider's decision making. For example, in its guidance on clinical decision support, FDA differentiates between software that provides a single, specific, selected output or solution as opposed to a list of options.[52] The agency also differentiates between supporting clinical decision making in urgent situations as opposed to situations where the clinicians have sufficient time to consider the provided and other information. Those crafting or researching AI with the intent that it does not cross over

Figure 6: Number of FDA AI Approvals Past 10 years by Year and by Category



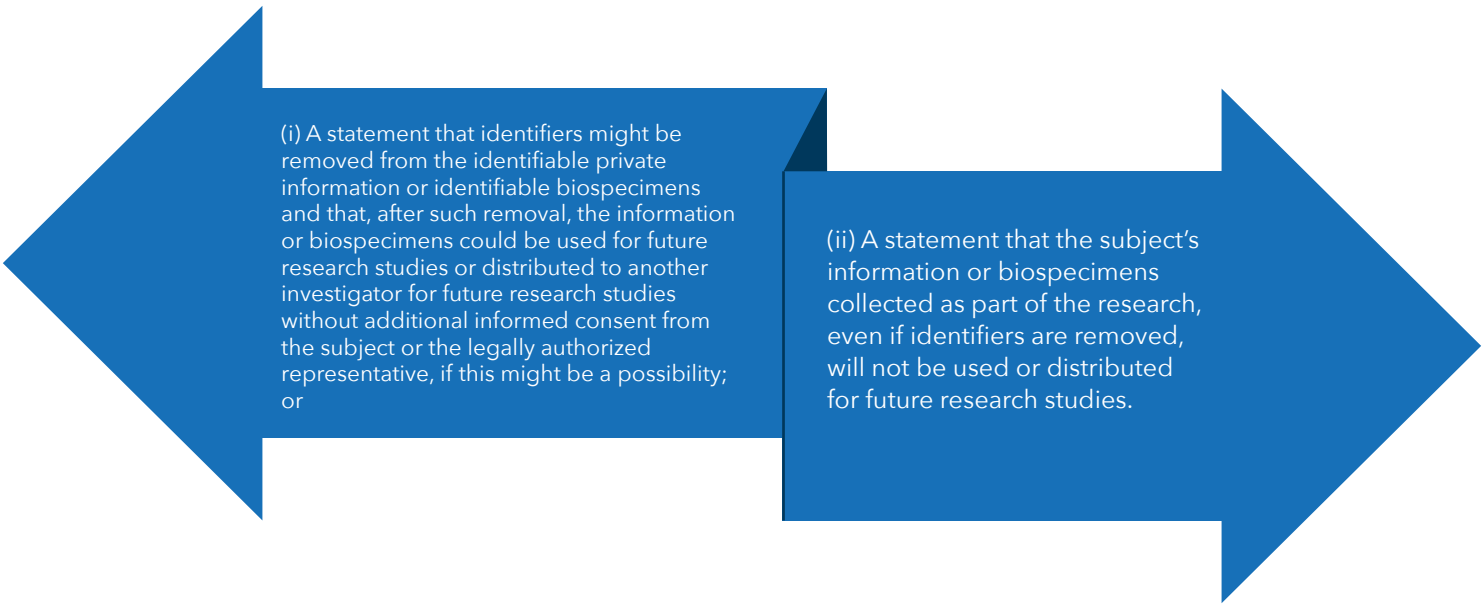


into being a medical device should carefully review the regulatory definitions and guidance. Regardless, although the research may not invoke the medical device regulations, other regulations (research or otherwise) and ethical principles governing clinical research still apply.

Most prevalently, the use of private and confidential personal information in the development of AI algorithms is governed by varying privacy regulations. Many competent authorities put forth privacy regulations for general personal data (e.g., the European Union's General Data Protection Regulation (GDPR)) as well as function-specific regulations (e.g., in the United States, the Health Insurance Portability and Accountability Act of 1996 (HIPAA) protects certain personal health information, and the Family Educational Rights and Privacy Act of 1974 (FERPA) protects certain personal education records). Even if records are not considered confidential, more emphasis is being placed on transparency of a recipient's use of personal data voluntarily given to them and the emerging legal concept of the right to erasure (commonly known as the "right to be forgotten"). Given that AI is dependent on data to build algorithms, the extent to which the data used is personal data can challenge privacy, confidentiality, and parameters of use.

Regulations parallel to medical device clinical trial regulations also put forth the requirements of informed consent and ethical review. For example, most U.S. governmental agencies have adopted the Health and Human Services research regulation of 45 CFR 46 from the Code of Federal Regulations (thus called the "Common Rule"), which, among other things, requires addressing informed consent and prior/continuing ethical review. When it comes to data, the Common Rule protections come into play when the data are identifiable,[53] and the rule explicitly exempts research activities from its requirements when either the information is identifiable, but publicly available, or "recorded by the investigator in such a manner that the identity of the human subjects cannot readily be ascertained directly or through identifiers linked to the subjects."[54] Of note, this is additive and not replacing other laws that may either define what "de-identified data" are or have additional protections for certain classes of information, such as added protections for genetic information or medical records pertaining to substance abuse or behavioral health diagnoses and treatment. These exclusions are not inconsistent with similar laws (e.g., GDPR does not offer protection for anonymized personal data). To address the opportunism of taking one's regulated personal information given under consent for one research purpose, removing identifiers to render it unregulated, and using it for secondary purposes, regulators are starting to require such prior notice of intent as part of research consent forms (see Figure 7). This concept would arguably also apply to the secondary use of personal information, even if de-identified, to create AI algorithms, but only to such information gathered in a prior research study for which participants consented. Information gathered for other purposes would not fall under such research protections of the original study, but would fall under general privacy and transparency laws, if existing.

Figure 7: New U.S. Research Consent Form Regulations Requiring one of the Following Statements About any Research that Involves the Collection of Identifiable Private Information or Identifiable Biospecimens [55]



DE-IDENTIFICATION OF PERSONAL DATA FOR TRAINING/VALIDATING AI ALGORITHMS

De-identification is a common strategy for using personal health information without the individual's authorization as it greatly reduces the risk of potential harm caused by the identifiability of the data. However, the term "de-identified" (and similar words such as "anonymized") is often misused by AI developers and even well-intentioned but regulatory naïve researchers.[56] Such misuse is highly problematic as when it comes to personal information, the term "de-identified" often has a specific regulatory definition that is more prescriptive than the common use.

AI developers and researchers naive of U.S. privacy regulations often find it surprising that the inclusion of data elements such as date of visit or age if over 89 must be excluded for the dataset to be classified as "de-identified."

For example, in the United States, simply removing obvious identifiers like names, addresses, and Social Security Numbers from a person's health information so that the identity of the individual it pertains to is not readily apparent does not in-and-of-itself meet the regulatory requirements for de-identification standards. Specifically, AI developers and researchers naive of U.S. privacy regulations often find it surprising that the inclusion of data elements such as date of visit or age if over 89 must be excluded for the dataset to be classified as "de-identified."

Similarly in the United States, the term "limited dataset" also has a regulatory definition that differs from its general use to imply that the data set is simply limited. Misunderstanding and misuse of these critical terms that have regulatory definitions can lead to friction, delays and increased costs in research. Whenever clinical researchers are dealing with privacy of personal information, it is critical that they fully understand the regulatory and legal definitions and requirements the accompany the use of the data.[57] All too often a well-intentioned protocol will state that "only de-identified data will be gathered" where the dataset contains forbidden identifiers (e.g. Date of Consent, Date of Admission, Zip Code, Age over 89. Medical Record Number, Device Identifier etc.) and such internal inconsistency often causes delays due to the requirement to either change the protocol or the dataset. This is not to say that one cannot use these fields is necessary for research, just that the dataset must be properly classified so that the proper regulatory requirements are met. For reference, Appendix 2 provides a detailed analysis of "de-identified" and "limited data sets" under the U.S. health privacy laws that is critical for researchers to understand when overseeing research in general as is more prevalent now with the demand for training and validation personal health data by AI developers.

Part 4: AI-Enhancement to Research Processes

This section is generally relevant to all clinical research professionals. It–supplemented with the examples in Appendix 3–describes how AI is currently being used from lessons learned.

USE CASES IN VARYING STAGES OF CLINICAL RESEARCH LIFECYCLE

The potential of AI as a clinical product to improve medical care is truly exciting, but perhaps just as exciting is the potential for how AI can help clinical research professionals find answers to research questions with higher quality and efficiency. AI has and will continue to challenge the trade-off triangle of "fast, cheap, and good," where you can only pick two but now perhaps can get all three. Use cases have been seen in nearly every stage of the research lifecycle and across all aspects. To understand the industry focus, one metanalysis of articles on AI deployment indicated that the largest area of implementation (more than 50%) was in subject recruitment, with a distant second being protocol design followed in frequency by data analysis, trial conduct, and finally pre-clinical activities.[58] The same metanalysis reported that oncology was by far the therapeutic area adopting AI solutions the most, followed as a distant second by neurology with cardiovascular rounding out the top three. While providing examples across all areas, the focus here will be on the stages of the process of most relevance to the ACRP membership of clinical research professionals. Appendix 3 provides anonymized vignettes of a small sampling of these use cases.

ETHICAL RESPONSIBILITY TO SHARE LESSONS LEARNED

One of our hallmarks of being a clinical research professional is that we share information; not only the good but also the bad as well as lessons learned. For example, we painstakingly validate and objectively share the truth in the data we generate so that we can adhere to our basic ethical principles of the research process (i.e., respect for persons, beneficence and justice) as well as the medical creed of "do no harm." It is also normative in scientific publications to not just tout the successes but also perform self-critiques so that subsequent research professionals can understand these and build on them. We do not have reckless disregard for, but in fact promote the documentation and reporting of things such as adverse events, unexpected problems involving risks to subjects/others and both our intentional and unintentional deviations from the research protocol.[59]

It is an ethical obligation for any clinical research professional who deploys or uses AI to share not only the successes but also its limitations and the lessons learned so that other professionals can learn and adapt.

When others can replicate or improve upon our prior work it is not considered a failure but a success. One can even see how the profession embraces this in the Code of Ethics and Professional Conduct for the Association of Clinical Research Professionals (ACRP) with items such as "when designing, reviewing, or conducting research, ensure that potential risks of the research are reasonable in relation to the anticipated benefits to the participants and the importance of the knowledge to be gained," "ensure that potential risks to research participants are minimized to the greatest extent possible and take all necessary steps to protect the participants at all times," "educate themselves, and where applicable, their students and their colleagues, about responsible research practices," and "not withhold information relevant to full evaluation of the safety, efficacy or utility of clinical interventions, agents or devices under investigation for the benefit of medicine, patients, science and society regardless of the research outcome."

The above standards of our profession directly relate to our obligations of responsible oversight of AI as an enhancement to our research processes. It would be inconsistent with our ethical and professional standards to share only our successes with AI to the exclusion of providing the opportunity for others to learn and build on it. It is an ethical obligation for any clinical research professional who deploys or uses AI to share not only the successes but also its limitations and the lessons learned so that other professionals can learn and adapt. Including open and honest discussions, just as we do in all other research endeavors, not only enhances our profession but also enhances public trust in our profession.

Part 5: The Impact on the Workforce and Our Profession

There is much speculation on how AI will impact the workforce in varying industries and the clinical research industry is not immune. Like so many industries, our positive mantra might be "AI will not replace your job as a clinical research professional, but a clinical research professional using AI likely will. "Certainly, the mechanics of human resources management in our profession are being aided by AI in both technical (e.g., resume scanning and routing, "Ask HR" chatbots to handle basic questions about employee benefits or rights) and controversial ways (e.g., employee engagement assessments, job success predictive scoring). AI is also assisting with human resource planning such as predicting staffing needs and optimizing workforce allocation. However, beyond the use of AI as a tool for managing human resources, much discussion remains on how AI will impact the clinical research workforce in general. The fact of the matter is that we cannot be certain how it will ultimately affect the profession; however, it is likely to be in line with the famous quote from Bill Gates about how "people often overestimate what will happen in the next two years and underestimate what will happen in ten."[60]

There are numerous use cases that demonstrate AI can remove many mundane tasks that are required, but beneath the skillset of, a clinical research professional. For example, in identifying potential candidates for clinical trials, many CRCs now rely on EHR queries to rule out large portions of charts that otherwise would have required manual review. NLP capabilities of AI are now further decreasing that review time by pre-identifying and removing those that arguably would have been easily excluded by a professional human reviewer which provides a better return on investment of the reviewer's time. Removing common and repetitive tasks that do not fully utilize the level of skills these individuals have can certainly add to the professionalization of the role and arguably reduce burnout. Such burden can even be alleviated by volunteers to the profession; for example, the Society for Clinical Data Management has trained a commercially available LLM on its Good Clinical Data Management Practices and question writing as an aid to its human volunteers in generating questions for its certification exams.[61]

As with any implementation of technological or procedural advancements, the added efficiency can and often does create anxiety within the workforce as to what to do with the time gained. For example, if the above referenced use of NLP-based queries can save eight hours a week of coordinator time, without clarity from the institution's leadership, a research recruitment specialist may wonder if the institution will reduce their work hours by that amount, utilize that time for new challenges (e.g., building recruitment diversity improvement initiatives that do not exist) or simply increase the amount of regular work accordingly (e.g., adding additional studies).

In all fairness, one cannot avoid the fact that there is certainly historic precedence where entire professions have been replaced or displaced by technological advances. Handloom weavers and switchboard operators have largely been replaced with technology while video rental store clerks have been displaced by technology virtually eliminating the DVD rental industry. Very visibly today, retail cashiers are being replaced with smart-phone app-based systems and

other security-enhanced self-checkout infrastructures. On the contrary, using any tool such as a chainsaw, using AI as a tool can greatly enhance individual, organizational, and societal productivity. However, also like a chainsaw, (i) users must understand the proper use cases for the tool as not every job is right for the tool in hand; (ii) when used it must be used properly to assure the desired outcomes; (iii) users must understand the tool's limitations and where other controls are necessary; (iv) manufacturers and users must respect the needs for safety surrounding use of the tool; and (v) there is a community obligation to discourage and/or prevent the tool from being used in manners not aligned with the norms and benefits of society. As clinical research professionals, we have an obligation to drive the successful development, testing, maturing, and adoption of AI on our terms. Inaction on our part will only lead to others, whether well-meaning or otherwise, establishing the foothold of its use on their terms instead of ours, which may be misaligned or more fraught with unintended consequences.

While the future net effects that AI brings to our profession remain unknown, it seems certain that the clinical research professional who continuously updates their knowledge on the benefits, limitations, risks and implementation strategies of these AI tools will be best poised for success.

While the future net effects that AI brings to our profession remain unknown, it seems certain that the clinical research professional who continuously updates their knowledge on the benefits, limitations, risks, and implementation strategies of these AI tools will be best poised for success. Especially at this early stage of exploratory deployment, the active engagement in conferences, webinars, grand rounds, journal clubs, professional societies, self-education, and the like will be of critical importance.

Similarly, success lies with those leaders who (i) assure the co-mingling between the research professionals and their AI leaders (e.g. having a clinical research professional represented on an AI committee as well as an AI team member represented on a research steering committee) and (ii) recognize that implementation of these new technologies is not only infrastructural but also emotional to their workforce, offer clear and transparent communication about changes and expectations of their intent, and provide their workforce the opportunities for upskilling and constant learning.[62]

Part 6: Closing

In conclusion, the contributors to this paper firmly believe that the community of clinical research professionals is rooted in a foundation poised for successful and responsible use of AI. Our unique blend of ethics, resiliency, flexibility, curiosity, and vision for a better future has made us successful not only in advancing clinical care, but also in advancing how we advance clinical care. As we think of other fundamental shifts in our profession that we have faced, are facing, or may face, we contributors have nothing but confidence that the members of the community of clinical research professionals will drive success on our terms once again, in this case by assuring that our evolving use of AI will always reflect the principles of our profession and our having the responsible oversight to assure it.

"You don't need to predict the future. Just choose a future—a good future, a useful future—and make the kind of prediction that will alter human emotions and reactions in such a way that the future you predicted will be brought about. Better to make a good future than predict a bad one."

—Isaac Asimov, *Prelude to Foundation*

Appendix 1: Limitations and Discussions

The vision for this publication was to give a "mile-long, inch-deep," non-technical overview of key considerations that a clinical research professional should have in the responsible use of this new technology. There is much more to learn beyond what is presented here, and even the information provided may change rapidly and without notice.

First, regulations, guidance, and governmental policies continue to evolve at even the most local levels. Although we provided examples of some of the most relevant to the clinical research profession, countless new laws and reinterpretations of existing ones exist that we could not possibly cover comprehensively—and many more are sure to emerge. It would be impossible to provide a comprehensive and updated list of everything the clinical research professional might need to know about. To further complicate the matter, many legal theories have yet to be tested in courts of law and in public forums. Specifically, while it was discussed how laws are emerging and some examples of them were provided, there are few norms on how many of these laws are likely to be implemented. For example, we discussed at least one legal requirement for an independent audit for algorithm bias, but not how such an audit is expected to be done in a valid manner. This is the proverbial case of "we're building the plane while we are flying it." Clinical research professionals need to be sure that they know their own bailiwick and collaborate with other disciplines (e.g., human resources, compliance) should they consider deploying or using AI that crosses functions.

Second, by the time you read this publication, this technology has almost assuredly advanced to the point of rendering some of the information and/or projections herein obsolete. There are things that were not possible as little as a year or two ago that are now routinely available to programmers and users.[63] This is why most oversight infrastructures put forth a "framework" that makes every effort to be technology agnostic and future-proof. The publication was intentionally crafted to put forth little technical information, and thus many references to technology may be subject to challenge from a difference of opinions or even now obsolete.

Third, no framework is perfect, and each requires resources. Numerous individuals and organizations with varying expertise have proposed model frameworks, and it was not possible to consider all of them in this draft. As this is rapidly emerging and the needs and risk tolerance of organizations widely vary, we did not attempt to opine on resources needed (such as personnel/staffing) to adequately manage the technology.

Fourth, the social acceptability of AI use in certain areas such as medicine and research, or even general use, will change as social perceptions and tolerance change over time. There are and will continue to be early adopters who will accept risks on their own behalf and on behalf of others. As clinical researchers, we are keenly aware of the balance between advancing the cutting edge of science and ensuring the protection of human subjects. This responsibility is woven into the very fabric of our profession. This certainly can carry over into risk minimization when deploying AI; however, this is unfortunately not always the case.

We have seen seasoned investigators propose well-intentioned but unethically designed studies involving AI as a medical product, which is unfortunate. Should this occur, it could certainly fuel public distrust with our profession, causing further friction in our ability to enroll participants and/or use data. As mentioned in the Disclaimer section, we were unable to independently verify/validate the claims of the vignettes and case studies put forth, nor to assure that they had adequate protections in place.

Finally, no individual or organization has the foresight to fully and accurately predict exactly what will happen. While this publication discloses possible impacts on the clinical research profession, it includes forward-looking statements that may not come to a pass. In preparing this publication, we engaged in extensive conversations and research, both within our industry and with other fields, such as the entertainment industry, which are also experiencing the full spectrum of professional, artistic, and business opportunities as well as the threats that AI/ML brings. Insights from this diversity of experiences, opinions, and directions were combined with the wisdom of seasoned professionals who have leaned into the trajectory presented here as our most probable future. That said, confirmation bias and optimism may have influenced this projected path over alternative possibilities. Nevertheless, given AI is undoubtedly here to stay, we intentionally framed and titled this publication as a clinical researcher's responsible oversight of AI, rather than as a debate on whether or not it should be allowed in our industry.

Appendix 2: De-Identification of Protected Health Information in the United States

This section is relevant to all clinical research professionals in their oversight of using health information protected by the United States Health Insurance Portability and Accountability Act (HIPAA) law. While applicable to the use and disclosure of such data for research purposes in general, the use of data for AI as an enhancement to the research process or for the training or validation of AI algorithms does not alleviate the obligations.

HIGH LEVEL REVIEW OF USE OF PROTECTED HEALTH INFORMATION IN RESEARCH SETTINGS

Under the U.S. Health Insurance Portability and Accountability Act (HIPAA), there are two methods to de-identify data: the "Expert Determination" method and the "Safe Harbor" method. Pragmatically speaking, to certify a dataset as "de-identified" under HIPAA one needs to either verify that it meets the Safe Harbor requirements or, if not, obtain written report/certification under the Expert Determination method. Slightly eking away from a de-identified data set is the HIPAA classification of "Limited Data Set" which allows for only a specific few of the otherwise forbidden identifiers provided there is an accompanying agreement with the recipient regarding security, restricted use and forbiddance of attempts to reidentify or make attempts to contact the person(s). Finally, if identifiers are essential to the research, with limited exceptions it would require either the authorization of the individual or a waiver of that authorization by legally constituted review board provided certain criteria are met surrounding the necessity of the information over de-identified data, the plan to protect and destroy the data as well as other criteria and documentation requirements.

Overall, the use of health data in research and AI algorithm training/validation requires more detail than the average research-related human subjects protection training provides and more specific training on health privacy laws.

For now, we will focus on the proper classification of "de-identification" and "limited data sets". As a reminder, not all subjects of research are patients. Sometimes the human subjects of the research (and/or their personal data used for the research) are employees, practitioners, students and other people. Absent specific laws defining the de-identification of such data, these methods can also be used for consistency's sake (i.e. holding the same level of protection for one's employees as done for their patients).

EXPERT DETERMINATION OF DE-IDENTIFICATION

The Expert Determination method, codified under 45CFR164.514(b)(1), is as follows:

A person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable:

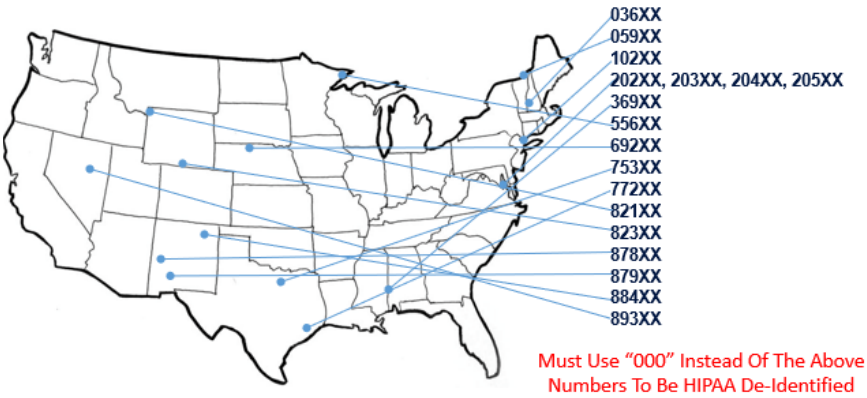
- Applying such principles and methods determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information; and
- (ii) Documents of the methods and results of the analysis that justify such determination

The Office for Civil Rights, which enforces the HIPAA privacy rules, has provided some guidance on this method and its terms. [64] However, despite this guidance, the method is not as prevalently used as the "Safe Harbor" as it remains both costly and its subjective determination can be easily challenged after a breach by individuals who may present their own statisticians to argue that the risk was not actually "very small" due to re-identification. Nevertheless, this method is a bona-fide method and can at times even offer more protection than the Safe Harbor method.

SAFE HARBOR OF DE-IDENTIFICATION

The Safe Harbor method is a much more straightforward approach in that it calls for the removal of specific information listed by HIPAA as identifiers. These 18 HIPAA Identifiers are paired with a catchall requirement stating that the releasing covered entity must not have actual knowledge that the remaining information could be used, alone or in combination with other data, to identify an individual.

Figure 8: Restricted Zip Codesfor HIPAA De-Identification Based on 2020 data.census.gov



Among other things, in order for PHI to be considered de-identified under the HIPAA safe harbor regulations, you can only use the first 3 digits of the Zip Code unless the sum of the population of all those areas is below 20,000, then you must change the three digits to "000".

See (45CFR164.514(b)(2)(B)).

Table 6: Differentiating the Criteria of a De-identified Dataset and a Limited Data Set under HIPAA

The Following "HIPAA Identifiers" of the Individual or of Relatives, Employers, or Household Members of the Individual, are Removed:	De-Identified Data 45CFR164.514(b)(2)	Limited Data Set 45CFR164.514(e)
1. Names	Not Allowed	Not Allowed
2. All geographic subdivisions smaller than a state, including street address, city, county, precinct, ZIP code, and their equivalent geocodes, except for the initial three digits of the ZIP code if, according to the current publicly available data from the Bureau of the Census: (1) The geographic unit formed by combining all ZIP codes with the same three initial digits contains more than 20,000 people; and (2) The initial three digits of a ZIP code for all such geographic units containing 20,000 or fewer people is changed to 000 (See Figure 8).	Not Allowed (unless excepted for in the criteria)	No postal address information, other than town or city, state, and zip code
3. All elements of dates (except year) for dates that are directly related to an individual, including birth date, admission date, discharge date, death date, and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older	Not Allowed (unless excepted for in the criteria)	All elements of date and all ages are allowed
4. Telephone numbers	Not Allowed	Not Allowed
5. Fax numbers	Not Allowed	Not Allowed
6. Email addresses	Not Allowed	Not Allowed
7. Social security numbers	Not Allowed	Not Allowed
8. Medical record numbers	Not Allowed	Not Allowed
9. Health plan beneficiary numbers	Not Allowed	Not Allowed
10. Account numbers	Not Allowed	Not Allowed
11. Certificate/license numbers	Not Allowed	Not Allowed
12. Vehicle identifiers and serial numbers, including license plate numbers	Not Allowed	Not Allowed
13. Device identifiers and serial numbers	Not Allowed	Not Allowed
14. Web Universal Resource Locators (URLs)	Not Allowed	Not Allowed
15. Internet Protocol (IP) addresses	Not Allowed	Not Allowed
16 Biometric identifiers, including finger and voice prints	Not Allowed	Not Allowed
17. Full-face photographs and any comparable images	Not Allowed	Not Allowed
18. Any other unique identifying number, characteristic, or code, except as permitted by [the regulations regarding "Re-identification"]; and	Not Allowed	Allowed
19. The covered entity does not have actual knowledge that the information could be used alone or in combination with other information to identify an individual who is a subject of the information.	Included in Requirements	Not Included in Requirements

While most of the 18 are readily intuitive to the untrained (e.g., name, address, Social Security Number), some are less intuitive (e.g., medial record numbers, medical device identifiers and serial numbers) and some are often very surprising to those who have not been trained on HIPAA requirements (e.g., no elements of date other than year, no age over 89, no geography smaller than a state etc.).

As referenced earlier, a few selected fields are allowed to be added to a Safe Harbor de-identified dataset in exchange for a written agreement with specific restrictions required by HIPAA (often called a Data Use Agreement or a Data License Agreement). To be clear, this classification as a HIPAA Limited Data Set does not require the adherence to the restrictions and obligations of fully identifiable protected health information, the regulation is also clear that it is also not "de-identified" but its own thing.

Therefore, research protocols, contracts and other documents related to the research and/or AI training/validation should be careful to ensure they reflect the correct classification of the dataset. Table 6 provides the regulatory criteria for classification of a de-identified data set (under the HIPAA Safe Harbor method) versus a Limited Data Set. Also, Figure 8 provides the list of the restricted Zip Codes based on the 2020 U.S. Census data.

Appendix 3: Anonymized Examples of Use of AI Across the Clinical Research Lifecycle

This section is generally relevant to all clinical research professionals. It provides brief and anonymized vignettes from published use cases or responses to solicitations for purposes of publishing herein.

The following are case examples of the diverse and evolving uses of AI/ML being deployed in real-world clinical research. These examples are not comprehensive of all AI/ML uses and are drawn from published literature and unpublished cases provided to the contributors to this white paper. The examples included are for demonstration purposes only, have not been verified by the contributors, and are not meant to suggest an endorsement of any specific approach or any particular use of AI.

DISCOVERY STAGE

Target Identification - A large pharmaceutical company partnered with an AI startup company to find novel targets for chronic kidney disease and idiopathic pulmonary fibrosis. They identified several new targets for further validation and development. Another example is a collaborative using AI to analyze the expression profiles of central nervous system samples from public datasets, and direct iPSC-derived motor neurons from an amyotrophic lateral sclerosis (ALS) research consortium. They identified 17 high-confidence and 11 novel therapeutic targets for ALS which were further refined to 18 validated targets as mimicking the most common causes of ALS. Of course, it will take time to do clinical research to validate the targets.

Molecule Repurposing - To identify possible treatments for COVID-19, an AI company used publicly available information and in 48 hours identified an already approved drug (for a different indication) as a candidate to move forward into clinical trials. This was done via searching for drugs that might interrupt the passage of the virus into the lung cells. The AI search of literature found 378 known potential molecules with such affinity, of which only 47 had existing marketing approval, of which six had high affinity, of which one had the most favorable adverse event risk profile. The manufacturer went on to successfully conduct clinical trials and the drug was among the first few approved emergency use authorizations issued by the FDA and eventually obtained full approval for the new indication.

Custom Molecule Design for the Targets - A drug discovery company used AI to create a new molecule designed from scratch in just 21 days and validated it in just 25 days compared to the usual two to three years. In another example, one company used AI to sort through and compare various properties of millions of potential small molecules, then choose 10 to 20 to synthesize, test, and optimize in lab experiments to select the eventual drug candidate. Another AI-derived molecule has just entered Phase II testing. Of note, it will take time to do clinical research to validate whether these predictions will be more successful than traditional non-AI assisted models.

PLANNING STAGE

Literature Search - Healthcare providers used an open-source AI tool to narrow down a large number of articles for a systematic review. The AI was trained initially with three relevant articles and three irrelevant articles, with further training by human "researchers-in-the-loop" as to what was relevant and irrelevant. When analyzing a set of more than 1,000 articles, the human screeners identified 32 relevant articles via traditional title and abstract screening where the AI identified 77 (noting that it originally selected 142, but the researcher-in-the-loop removed 65 due to their not being

original research). It was estimated that the AI-assisted search for the most relevant articles saved 77% of the time over the unaided process. They determined further fine tuning was needed to determine the best stopping point of the AI for proper inclusion, as well as further training of the algorithms to be able to differentiate between articles of original research versus review articles.

Trial Outcome Prediction Based on Trial Design - One AI company trained its models on elements of trial design (e.g., number of patients, arms, blinding, country(ies), etc.) and target choice from 55,653 Phase II clinical trials obtained from public sources (e.g., ClinicalTrials.gov, pharmaceutical company press releases, journals, etc.). The data indicate that the target choice achieves high predictive performance for Phase II outcomes, whereas elements related to trial design improve the prediction of success only marginally. These algorithms predicted the outcome of 17 Phase II studies with 79% accuracy based on three variables: (i) the statistical and clinical significance of efficacy and safety endpoints; (ii) a company's decision to transition the drug program to Phase III; and (iii) the momentary increase of company's stock price in response to published clinical trial results.

Trial Outcome Prediction Based on Prior Trial Data - Several companies have also conducted in-silico clinical trials to predict outcomes. However, their trials are still ongoing, thus success brought by the AI remains theoretical and yet to be determined. One company's self-critique was that it understood that the predictions descended solely from the knowledge used to build the models, thus while the models may be able to consider some "known-knowns," nothing "surprising" could come out of an in-silico experiment (i.e., the "unknown unknowns"), thus this company still advocates the combined use of in-silico and physical experimentation for better predictions.

Trial Outcome Prediction Based on "Digital Twin" Data - One company used nearly 7,000 clinical records from placebo arms of Alzheimer's disease clinical trials and observational studies to train a generative ML model to create digital twins for their patients in their clinical trial. Each digital twin was created by matching a set of the 100 closest placebo/standard-of-care arm patients to each enrolled clinical trial subject to average out a model of the patient's likely disease progression. The results for the following 18 months were consistent within 95% confidence intervals with actual data across all scores, timepoints, and cohorts in the clinical trial. Their next phase is to test to see if implementing this into trial design can reduce the need for placebo arms of studies and either increase power or reduce sample size required to achieve desired power.

Since that study, one AI company published the results of working with two major pharmaceutical companies' Phase II Alzheimer's disease studies. They indicated that, for these studies, the use of digital twins achieved equivalent results with a 19% reduction in control arm size at a 12-month endpoint variable or a 33% reduction at an 18-month endpoint and did so with statistical power gains of up to 10%. In the other pharmaceutical sponsor's study, it was determined that a 10% to 15% reduction of sample size could be achieved with the use of digital twins with a power gain of 7% on a critical endpoint. The company cautions that the models are currently challenged to (i) assimilate non-patient data that may be relevant, such as summary data from populations and text-based domain knowledge; (ii) move outside disease-specific progression models and into a more "universal" digital twin generation model able to accommodate multiple indications and other factors, such as social determinants of health; and (iii) more rapidly as similate when new treatments or epidemiological information become available.

Site Selection - One CRO uses AI algorithms as an aid in predicting site success based on multiple variables and data about the site. It predicts variables such as enrollment performance, speed of start-up, and data quality and sites in higher tiers are fast-tracked for selection. Not surprisingly, the company mentions previous performance and enrollment speed, data quality, diversity data, key population connectivity, and equipment and capabilities data from their existing site profile database as being key correlators in determining success. Among the highest was the correlation with claims, census, and other third-party data related to the target population.

The CRO indicated that while the AI-enhanced tiering system accelerated site data analysis, human expertise is still required for the more nuanced assessments, citing that the AI falls short in areas such as (i) grasping the critical components of protocol details and (ii) due to less data, newer sites that fit key requirements were ranked lower by the AI and thus had to be flagged for assurance of human review. Based on their analysis, they had several weeks (average 25% reduction in time) worth saving to first site selected. More importantly, their analysis indicated that the AI-enhanced site selection process facilitated a 50% reduction in non-enrolling sites. Additionally, selected sites were 24% more likely to hit their first patient in target and 21% more likely to hit their last patient in target.

Site/Country Selection - A pharmaceutical company partnered with an academic institution to generate a predictive model for the epidemiology of COVID-19 to better target global site selection for their vaccine. The model was provided data (e.g., case numbers, hospitalizations rates, length of stay, mortality rates, and testing rates) generated from each of the 50 United States as well as 120 other countries to determine where COVID-19 would spike next. The site/country recommendations determined by the algorithms were fairly different from the results of traditional thinking, including locations not generally chosen. The study team attributed the success of management choosing the AI-generated sites that were originally excluded by human bias to the transparency and interpretability of the results,

specifically the AI was trained to explain the "why" behind the recommendation and allow for human-in-the-loop feedback. The results were, as compared to traditional site selection, an eight-week (33%) decrease in trial completion time and 15,000 (25%) fewer participants. Also, they attribute the lack of bias to the finding that the race, ethnic, and gender diversity was one of the most representative of the COVID-19 vaccine trials.

Decreasing Screening Cost and Burden - A group of academic researchers indicated that, historically, two-thirds of asymptomatic elderly patients were negative for amyloid plaques on the screening PET scan for their mild cognitive impairment/Alzheimer's disease studies. They worked to derive a predictive algorithm that could predict the presence of amyloid plaques and thus eliminate the most expensive, risky, and burdensome process leading to screen failures. From three large studies totaling more than 1,300 patients, they obtained data questionnaires and cognitive tests that target executive functions, behavior, and overall cognitive skills as well as the resulting PET scans.

To assure the quality of the algorithms, they only used 70% of the data to train the algorithms, reserving 30% for validating (recognizing that validating with the same data it was trained on would likely result in self-fulfilling answers). It was determined that adding MRI information to the algorithms did not result in statistically different results. In back testing the algorithms, it was estimated the screening cost would have been reduced by 20% and suggested that many of the elements needed could be shifted from screening to pre-screening phase, which expands the potential population. They noted significant challenges in fine tuning the algorithms, especially needing human experts to opine on how to reduce the number of variables. They also recognized that the larger the training dataset, the better the results; however, one challenge is that many studies are not homogenous and often gather different data, thus are difficult to combine. Regretfully, no analysis of bias control was reported.

Improving Readability of Informed Consent Documents - A research site pasted the Study Summary paragraphs of a sponsor-provided informed consent form into an LLM to have it rephrased into more readable text. The result decreased the length of the section by 30% (from 273 to 191 words) and improved the readability score (decreasing it from 8.3 to 6). The study staff deemed the results substantially equivalent to the original.

Translation - A university PI asked an LLM to translate a protocol's inclusion/exclusion criteria into Mandarin. To test the validity, the LLM was asked to translate the Mandarin text into German and then the German text into English. The third-generation English back-translation was substantially equivalent to the original English text. In another example, one community-based site and patient network utilizes public LLMs to manage first draft translation of internal/non-study specific patient facing documents (e.g., notice of privacy policies, registration forms, and medical records release forms). The returned translation is then reviewed by an internal staff member who is fluent in reading and writing the language/dialect and who makes any additional edits. The fluent staff members have indicated that the first draft translation by the AI software is superior to many of the final translations received by companies specializing in this service. Of note, while the benefits of AI language translation without a human-in-the-loop are very attractive (i.e., available 24/7, low to no cost, instantaneous turnaround time), many point out LLMs (at least currently) are often lacking in terms of hyper-local cultural sensitivity and converting technical jargon.

From a regulatory perspective, no known regulatory authority in clinical research has indicated it accepts (for official documents) such translations without a qualified human-in-the-loop signing off via some sort of validation process (often dubbed a "certified translation"). However, regulatory guidance for such is open to interpretation. For example, FDA guidance states that IRBs must "approve reasonable procedures for ensuring that the translations will be prepared by a qualified individual or entity, and that interpretation assistance is available,"[65] but it does not dictate how that is to be met. In an Office of Good Clinical Practice correspondence dated April 4, 2008, the FDA stated, "the IRB is free to establish its own requirements for translations and the acceptability of translated documents, for example by requiring 'certified translators,' or an official certificate. However, the particulars of doing so are up to them, since we do not have any specific regulations about this." It would be debatable if the processes described above were acceptable to an IRB for study-specific documents like informed consent.

Writing Test Questions for Protocol Training - A university PI trained an LLM via uploading a protocol and asked it to create a multiple-choice quiz on inclusion/exclusion criteria with justifications as to why answers were correct or incorrect. It generated 35 questions in about five minutes. Most questions only required some minor wordsmithing by the human-in-the-loop. Two questions/answers had factual errors (but were easily editable by a human).

AI-Enhanced Workforce Targeting - One pharmaceutical company contracted with an AI company to find candidates to solicit for hire for various positions, including research scientists. The company used nearly 100 sources of publicly available data, specifically mentioning professional social media profiles, rate of promotion within a resume, lists of authors of scientific publications, patents and their authors, code collaborations, and government data. These data also contributed to the identification of geographic regions in which to expand office locations or to deploy a brand awareness campaign as a preferred employer.

AI-Enhanced Resume Scanning - One pharmaceutical company used AI to screen incoming resumes for various positions. The AI provided a letter grade of the match between the candidate's resume and the job description to help the company prioritize the reaching out to the candidates. In a follow-up study, it found that the higher graded matches tended to stay longer [in their job] versus lower graded matches. In their effort to avoid bias, they indicated that the models were trained not to look at candidate name, university name, and other things they believed may trigger racial or gender bias.

Budget Justifications - A clinical trial site used a publicly available GPT LLM to generate budget justification text for a diversity, equity, and inclusion program. The LLM identified seven areas to resource and provided justifications for each, specifically: Community Outreach and Engagement, Translation Services and Multilingual Materials, Cultural Competency Training, Transportation and Logistics Support, Incentives and Compensation, Diversity-focused Advertising and Marketing, and Partnerships and Collaborations. Many of these were funded by the sponsor based on the justifications provided.

CONDUCTING STAGE

Site Initiation - A CRO used a commercially available software plug-in for video-conference software used during a site initiation meeting. The software generated a transcript for the meeting and a list of follow-up items identifying who was responsible for them. The transcription was deemed acceptable but had some errors in recognizing some words typical of voice recognition systems. The list of follow-up items was not considered wholly complete. A unique challenge was faced because many of the site staff were grouped in a conference room login, thus unless it was specifically stated in the meeting who exactly was responsible for the action, the AI defaulted to the site's login name. In other cases, a site attempted to initiate the use of a similar AI tool during a web-based site initiation visit (SIV) for its own quality assurance and training purposes (e.g., task log creation, future training for reminders or staff turnover), but the CRO declined to conduct the SIV under those terms, stating concerns about lack of information on confidentiality protection of the recording and transcripts.

Protocol Q&A Chatbot - A university PI uploaded a protocol to a laptop-based LLM to create a chatbot which can answer questions about the protocol. It was asked "what are the inclusion/exclusion criteria," "what are the primary and secondary endpoints," and "what are the main hypotheses," all of which were correctly answered. In testing prompt sensitivity, the same question was asked multiple times with seemingly inconsequential differences, being "How many arms are in the study" [without a question mark], "How many arms are in the study?" [with a question mark], and "How many arms are in the study" [noting the intentional adding of an extra space] and to explain the answer. While all three answers correctly identified the number of arms, they were differently explained. Similar to how answers differ on internet search engines based on the text, this explains the growing issue of prompt sensitivity and thus prompt writing as an emerging skillset.

Trial-to-Patients Matching - A healthcare provider inserted NLP into its workflow to read pathology report text and flag patients that seemed to meet inclusion/exclusion criteria for their oncology studies. The system routed an email notice to trial screeners that could contact the ordering physician (often before the physician read the pathology report) with a list of potential clinical trials to consider as a care option when they review the report with the patient. This saved the research coordinator time in reviewing medical records and the real-time notification accelerated the disclosure of clinical trials as a care option to a patient who qualifies as a participant.

In another example, a community-based site and patient network compared AI-enhanced prescreening to traditional chart review for a pediatric vaccine study. The study required patients to have had a prior vaccination with a set prior period as well as other standard inclusion/exclusion criteria related to age and medical history. Standard chart review returned a list of approximately 45,000 patients with the appropriate age and vaccine history. Once the AI/ML applied the temporal criteria and exclusionary medical history, the list was narrowed to 1,450 study candidates. This facilitated the site recruitment team's ability to focus outreach to the study candidates who were most likely to qualify based on the study criteria. As a result, the site enrolled subjects at a rate three times that of the other sites participating in the same study.

Patient-to-Trials Matching - One large hospital collected narrative eligibility criteria from ClinicalTrials.gov for 55 oncology clinical trials actively enrolling at their institution. Using NLP to read their patient's electronic health records (EHRs), the hospital generated a list matching trials to patients and another matching individual patients to trials. When matching trials to patients, it was found that it were able to narrow human oncologist reviewers from 163 patients to 24 per trial (an 85% reduction in oncologist effort). Similarly, in matching patients to trials, it was able to reduce oncologists' consideration of 42 trials per patient to four (90% reduction in oncologist effort). It discovered weaknesses caused by its algorithms being trained on targeting individual words, limiting the ability to find relationships between consecutive words or similar phrases (e.g., specifically "T cell lymphoblastic lymphoma" versus "Pre-B cell lymphoblastic lymphoma"), which was the cause of 54.7% of

the errors. The algorithms also had difficulty with temporal events such as distinguishing between new and historical diagnoses as well as incorrect relapse status or different stages of disease (10% of the errors). The hospital also indicated challenges in criteria such as "previous enrollment status."

The same hospital did a similar implementation in its pediatric emergency department and had comparable findings (in this case reducing screening time by 34% and increasing the numbers of potential participants screened, approached, and enrolled by 14.7%, 11.1%, and 11.1%, respectively). It also found that many false positive recommendations were caused by the NLP failing to consistently detect negation (note, while the hospital did not give examples, common examples of this found in literature would be "the patient has no fever" or "the patient is rarely non-compliant with medications"). The institution also surveyed the research coordinators using the system for feedback to generate improvements in its user interface and experience. In another example, researchers trained a publicly available LLM on natural language and inclusion/exclusion criteria.

They challenged the algorithms to match three publicly available cohorts of 184 patients' data to any of more than 18,000 clinical trials. For the inclusion criteria, the algorithms achieved a prediction accuracy of 0.899, which was within the human experts' accuracy range of 0.876 to 0.916. However, in assessing exclusion criteria the algorithms scored slightly below human range, noting that the algorithms' current training had difficulty differentiating certain phrases such as "not excluded," "no relevant information," and "not applicable." One example given was while exclusion criteria such as "the patient should not be pregnant" and "pregnant patients will be excluded" may serve the same purpose, the algorithms may treat them differently. Nevertheless, they were able to match patients with criteria-level faithful explanations in a manner very close to humans and the time saving over human effort was estimated to be 42.6% faster when done with algorithm assistance.

In a similar case, one solutions provider for a multicenter oncology study, citing that 92% of the prescreening inclusion/exclusion criteria were in unstructured medical records and genomic data, found the NLP-assisted scanning reduced study coordinator screening time from seven hours to two hours, reduced the average days to recruit first patient on study from 159 to 54, increased accrual rate of accrual from 0.7 per month to 2.7 per month, and increased actual to expected sponsor target accrual from 31% to 56% of target. The solutions provider indicated that one lesson learned was that planning and management buy-in was critical at the start-up level.

Chatbot-Enhanced Trial-Finding - One large technology company trained an LLM on ClinicalTrials.gov data and NLP to take naturally worded questions to find matching clinical trials. It can be deployed to a global extent or be limited to custom trial information provided by users. For example, it takes natural queries like "Find me clinical trials in Israel for a 24-year-old woman with lung cancer" or "find clinical trials for Alzheimer's disease" and matches to the inclusion/exclusion criteria from the data source. The system identifies which trials they qualify for as well as generates subsequent questions to further narrow down the selection. After a series, the LLM will provide a list of trials they are eligible for as well as a list they are not eligible for and the reason why.

Similarly, a patient recruitment company also licenses out such technology to sites, patient advocacy organizations, and others. Its solution provides structured and mostly multiple-choice questions such as "how far are you willing to travel," "which of the following kinds of testing were used in your diagnosis," and "have you ever had a stroke." Neither company offers any comparison studies to human matching and the first company referenced uses the disclaimer that the chatbot is "a capability provided AS IS and WITH ALL FAULTS" and "the customer is solely responsible for any use." Of note, all the companies are doing is referring the users to the trial contacts for formal prescreening.

Participant Classification - One sponsor used AI to automate the dermatologic-level classification of skin cancer. When compared against human experts in identifying keratinocyte carcinomas versus benign seborrheic keratoses, the results were found to be equivalent.

EHR Query Writing - One university trained an AI algorithm to translate natural language inclusion exclusion criteria (i.e., copied straight from the protocol or ClinicalTrials.gov) into executable code for clinical database queries. This allows research coordinators not familiar with writing structured queries to do this themselves without the need for significant training or requesting this service from their information technology department. It found that, on average, each trial takes users about 15 seconds to get the translated query. Some problems occurred when the AI did not recognize one or more medical phrases or had difficulty with subjectivity (e.g., what was considered "severe" in "other severe medical disorder" versus "other medical disorder"), calling out the importance of a robust training dataset. Also, the algorithms had difficulty translating initial event cohorts (e.g., "first diagnosis of Alzheimer's disease") and time-based criteria (e.g., "three years after the first diagnosis").

Chatbot-Enhanced Consent - Researchers at a university built an AI-powered chatbot to simulate an in-person informed consent experience. To test it out they created a sham study of an online survey about social media misuse and the Q&A model was trained by 54 volunteers creating a 200-question/answer bank. Visitors to the website were randomized to seeing only the consent form versus interacting with the chatbot, which appeared as a virtual research assistant,

greeted the participant, presented and reviewed the consent form with them section by section, answered the participant's questions (specifically asking if the participant had any questions after the risk section and at the very end), and documented their affirmative consent. As compared to the control group, the chatbot-enhanced group scored higher on comprehension scores (comprehension was evaluated by questions that required participants to process the information presented in the consent form beyond simple recollection).

Also, quality metrics on the responses to the dummy survey indicated the chatbot-enhanced group had higher quality responses (as evidenced by less satisficing in multiple choice questions and more elaborations in the open-ended questions). Finally, as indicated in a survey on the perceived power differential between researchers and participants, those in the chatbot-enhanced group indicated they felt more equal in power relation with the researcher in charge of the study (i.e., more like a partnership) than those presented with the consent form alone. Of note, the AI was used to map naturally worded questions to pre-determined and ethics board–approved answers in a Q&A bank, whereas it likely would be fundamentally different if the LLM was generating the answers from scratch.

De-Identification - A site used NLP to find and redact personal information from videos, photos, PDFs, and emails prior to their being sent to outside researchers. For example, a narrative text report with identifiers is automatically adjusted to "Patient [REDACTED_NAME], a [REDACTED_AGE]-year-old [REDACTED_GENDER], was admitted to the hospital on [REDACTED_DATE] with complaints of abdominal pain."

CTMS Workflow Integration - To address inefficiencies related to software systems that work poorly with clinical workflows at research sites, a site network developed a mobile-based NLP plug-in to their home-grown clinical trial management system (CTMS). Initially, the capabilities of the system focused on 12 common navigation commands that often lead to inefficiencies at research sites, such as scheduling and serious adverse event (SAE) reporting. They attribute much of their success in their quality of command recognition and multi-user acceptance to using more than 500 people to train the voice prompts, intentionally assuring the inclusion of different accents and background environments as well as accommodating intuitive command phrasing. As a result, users require minimal training as they can use their natural method of speaking (e.g., the NLP accepts other phrasing such as "open patient scheduling" or "schedule an appointment," etc.). If the system "hears" anything ambiguous or otherwise unrecognizable, the user experience designer adopted the conservative approach for the system to ask for clarification rather than executing the wrong command.

A big success involved the timeliness of SAE reporting. With a mobile app in place and NLP ready, the system integrates better with staff workflow (particularly when physicians learn of adverse events in a hospital or through contacts outside of the research space, it can be documented immediately). Anecdotally, the site found that mobile-enabled, voice-prompted scheduling provided patients with appointments immediately at different points of contact and reduced staff time.

Safety Monitoring in Non-Traditional Sources (e.g., Social Media) - A research consortium (consisting of private companies, academia, and the FDA) obtained from an authorized third-party vendor the verbatim text of the body of the original post, metadata such as timestamp, and, if available, the poster's gender and geographical information for approximately 1 million Facebook and Twitter (now known as "X") posts that identified one of 10 products. ML algorithms were run to identify "posts with resemblance to adverse events" and further curate them by deleting spam posts and duplicate posts, and by removing personally identifiable information to create the dataset.

Out of nearly 1 million posts mentioning the products, they found that a little more than 10% mentioned adverse events; however, only 13 were sufficiently documented that a causality review could be done, of which six were determined to be either certainly, probably, or possibly related to the drugs under study. They found that while Twitter typically generates more posts mentioning medical products, Facebook data typically consists of higher-quality posts with more detailed information (noting that at the time of the study, Twitter was limiting its posts to 140 characters). Also, a higher proportion of posts from Facebook originated from patients posting within open patient groups rather than individuals posting on their personal pages.

As expected, there was not enough information to do a full mapping to structured adverse event terms and imply causality assessment due to the nature of most posts (e.g., "I'm severely allergic to [redacted]—had to go to the hospital just 8 hours after I took my dose," "my doc wants me to try a new anti-psychotic [drug #1], [redacted due to obscenity], I almost lost my life a few times from it, I'm the only person I know who's allergic to [drug #2] and [drug #3]," and "so, they took me off [drug] because I had an extreme attack last week and was trying to go into cardiac arrest, that was fun [truncated]," and "I tried taking [drug] and it made me SO sick I thought I was gonna die").

Others were of sufficient context to be mappable for causality review (e.g., "Started [drug] 4 days ago, already have strange side effects: severe abdominal pain, shortness of breath with almost no activity. My heart beats are even more irregular than before and now I have on and off heart racing that I didn't have before. I will call my Cardi tomorrow to get off the meds" and "Been on [drug] almost 6 weeks.

Four days ago, I noticed I had a rash on the palms of both my hands with dots that kind of look like small blood blisters. Came to this site to see if others have the same issue").

Causality Assessments - One pharmaceutical company developed an AI tool to assess the causality of adverse events. The model was then back tested on 978 randomly selected human determined cases. The algorithm identified the cases where the drug was determined by humans to have caused the issue (65% sensitivity) and was very good at identifying when there really wasn't one (93% specificity). It was also somewhat successful in identifying (with 79% positive predictive value) if there would be a problem or if there would not be a problem (88% negative predictive value). Noted limitations were the ability to classify complex cases like drug-drug interactions. They also noted that the low sensitivity was likely due to cases that were originally determined by humans as "indeterminate" and when these cases were excluded from the study, it resulted in higher algorithm sensitivity.

Pharmacovigilance - A pharmaceutical company partnered with an AI company to try to identify the reason leading to one death and five hospitalizations for intracranial hemorrhage presumably caused by a Phase I investigational drug. The AI company processed 80,932 protein structures and was able to identify 11 off-target potential causes, two of which were deemed highly likely. Other pharmaceutical companies with drugs in the same class were able to learn from this information.

Classification of Adverse Events - One pharmaceutical company partnered with a large technology/AI company to create AI-augmentation to the human determination of adverse event seriousness. To be clear, the company indicated that safety report seriousness classification will always require human confirmation; however, this system was developed to help prioritize reviews and highlight the most relevant text for the human reviewers to more quickly make their determination. Based on a training set of 26,256 randomly selected documents, the test set of 2,716 reports compared its AI review to human review to find the AI achieved an accuracy of 83% in post-marketing reports, 92.9% in solicited reports, and 86.3% in medical literature reports for the serious versus non-serious classifier. On sub-classification of SAEs, the algorithms had reasonably good results (F1 score of 77.7 for death, 78.9 for hospitalization, and 75.5 for Important Medical Event).

Contrary to what one would expect when using AI, the company's experience was that the algorithms were more challenged with categorizing adverse events from longer text reports (at ~500 words or more), thus it modified the algorithms to prioritize those in the human review assignments. Of note, a survey by TransCelerate's Intelligent Automation Opportunities in Pharmacovigilance indicated that while AI-enhanced workflow prioritization/triage and alerts for high-risk cases such as the above were expected to have high levels of return on investment, so were quicker language translation of the documents, the ability to more rapidly identify and eliminate duplicate reports, and the ability to map information into local structured reports. [66]

Preemptive Communications - One pharmaceutical company indicated it was trying out AI-generated communications to sites to prevent common issues. For example, a clinical research coordinator (CRC) would receive an email such as "We see that you have a patient visit coming up next week, but you have not logged into the system in the past two months. Please be sure you remember your password and can log in prior to the visit." No additional information was provided by the company on the successes, limitations, or site acceptance of the pilot.

Site Payment Reconciliation - One large site network has developed its own purpose-built LLM agent that helps reconcile its payments. After first compiling a comprehensive listing of billable "events" from across all systems at the most detailed study/site/patient level, an LLM applies the appropriate context to cross match event descriptions and amounts paid among the varying systems. Essentially the algorithms can match "ACME Pharma 1235 Visit 1 - Patient XYZ" in the site CTMS with "ACME protocol 1235 Day 1 Patient 3" in the CRO's electronic data capture, "ACME Pharma Study, Subject 3, V1D1" on the payment detail report and "ACME Patient 123"' on the payment remittance. This agent can subsequently update the site CTMS and financial systems to apply cash received to the correct events.

With a >90% successful match frequency, only a fraction of transactions now require review from a staff member. The site network has become significantly more productive toward deploying timely and accurate collections resources and the effort has accelerated their payment cycles by two weeks. A lesson learned is that the process still must have human intervention, specifically despite the efficiencies gained with the AI enhancement, because there are still matches that don't occur and there is some continuing quality checking required.

ANALYZING PHASE

Research Fraud Detection - One data quality oversight solutions provider was brought in by a pharmaceutical company after a study was complete to validate the dataset. The algorithms used were trained to review site data and generate a score geared to ranking sites that had atypical data and the reasons why. In this case, the AI detected an outlier in the patient diary data and metadata for one of the 160 sites involved. Onsite monitoring discovered that the handwritten questionnaires had strikingly similar handwriting, thus the data were rendered fraudulent. In another case, the same provider took data from clinical trials of a drug that had already received FDA and European Medicines Agency approvals, and the AI highlighted that one

site had anomalies in blood pressure monitoring, specifically discovering that the site logged the same blood pressure for all patients at all visits. Both items are something that should have easily been caught by monitors, data analysis, and regulatory reviewers.

AI-Enhanced Query Writing - One pharmaceutical company partnered with a data management company to increase the speed of the data review process. Using the pharmaceutical company's historic data, the AI was trained to detect anomalous (incorrect, duplicative, nonsensical) issues. It was admitted that it did take many work hours to confirm/reject scenarios to fine tune the foundational models. From there, once an anomaly was detected on live data in an active study, an LLM generated the initial draft of the query text which was routed to the human-in-the-loop to confirm/adjust/reject the AI's finding.

The AI-enhanced process decreased the query writing and approval time from 27 minutes to three minutes, saving more than 1,600 work hours. Additionally, the company identified a significant decrease in average time from data entry to query issuance—from 25.4 days to 1.7 days. The data management company indicated the tradeoff in protecting client confidentiality by training the algorithms only on the single client's data and, although greater quality may have been achieved through the larger dataset, did not include other customers' data in the training. With that said, it acknowledged that the base algorithms are built on common data elements, and thus can be built upon with further custom training for other customers, especially if needed for data from differing therapeutic areas.

Secondary Use of Data for Subjective Rating Scales Prediction - A pharmaceutical company took data from one of its Phase III psoriasis trials that had ~2,700 images and the accompanying investigator's Psoriasis Area Severity Index (PASI) scores to train algorithms to predict the PASI score from the photographs alone. PASI is commonly used for eligibility criteria and primary endpoints in psoriasis trials. Being from a clinical trial, the training data were considered high-quality (i.e., more so than training on routine care images and scores) as (i) the photographs were all taken in a uniform way (i.e., using the same settings taken by the same make/model camera), (ii) rated by uniformly trained photographers and raters, and (iii) data were monitored for quality.

The algorithms were trained to determine body part detection from background, disease lesion segmentation from healthy skin, and severity classification for detected lesions from the photographs. The results had a Lin's concordance correlation coefficient of 0.86 and a Pearson correlation coefficient of 0.90 as compared to the dermatologist's read, without systematically over or underestimating PASI scores or percent changes from baseline. The company did acknowledge the need to further develop the algorithms through access to a dataset from a diverse population to assure imbalances in skin tone did not impart systematic bias. It also recognized the importance of splitting the training data and the testing data (in this case, 90% training and 10% testing) to prevent leakage in the validation process.

Natural Language Querying of Pharmacovigilance Data - One pharmaceutical company trained a commercially available LLM to translate natural language into structured query language (SQL) for non-programmers to be able to query its pharmacovigilance data. The LLM was initially trained with a 290-page document that included every table and column definition within the database and other field descriptors. It also trained the LLM on historical user request logs to get a diversity of representation of query complexity. The foundational chatbot context prompt was *"You are an SQL expert. Given a question, generate a syntactically correct SQL query. Avoid querying non-existent columns and pay close attention to column-table associations. For keywords in the WHERE clause, ensure case-insensitive data comparison, for example, upper ('STATE NAME') = upper ('deleted').' If you are unable to generate the SQL query, please state that you cannot create the query without additional information or context, do not attempt to make anything up."*

The models required human-in-the-loop subsequent training to train more focus on the essential tables to move from an initial 8.3% pass rate to a 78.3% pass rate (as manually classified by a database expert comparing the AI and human query results). The biggest fails were experienced in the queries rated at high-complexity (noting that low to medium complex queries had a pass rate of 94%) or when user statements were ambiguous, thus the company indicated continued efforts were needed to refine the models and minor training on best practices in prompting.

PUBLICATION STAGE

Describing Statistics to Non-Statisticians: A statistician pasted the written code they used to analyze protocol data into an LLM and asked it to generate a statistical methods paragraph in a format that resembles a publication. Although the initial result was deemed to be too technical, it was easily remedied upon asking the LLM to rewrite it as if being published in a journal and not to include any code. The adjusted text was deemed acceptable and more in layman's terms.

Translation: One company created an AI tool to translate scientific publications from Chinese into English. This was done because the company cited it takes non-English speaking scientists 51% more time to write a paper and that there was English-speaking bias in publication acceptance resulting from inadvertent grammatical errors and less than optimal journal-quality phrasing in the translation. The company facilitated the translation via training AI on more than 1 million scientific publications, had more than 500 authors test it, and experienced a 14% improvement in publishing success.

Appendix 4: Use of AI When Authoring Publications

This section is particularly relevant to clinical research professionals involved in authoring, peer reviewing, and/or publishing the results of clinical research. One focus is how government copyright offices are addressing the complexities of applying intellectual property protections to AI-generated content. In addition, it provides a detailed review of the stance of the International Council of Medical Journal Editors (ICMJE) on AI-generated content and an author's responsibility. Specifically, it emphasizes the author's responsibility to distinguish between human and non-human contributions and to disclose these distinctions to publishers and other stakeholders. The section further references emerging methods for citing AI usage. Finally, a reminder is provided that both authors and peer reviewers bear ultimate accountability for their professional work. Although AI can enhance productivity, AI hallucinations and other errors may introduce flaws in their processes, potentially exposing negligence.

Clinical research professionals are advised that publishers, professional associations, sponsors, and other stakeholders are rapidly adding and modifying their guidance/requirements for those that use AI in the publication of scholarly activity. Many have already modified several times and even have reversed their guidance/requirements. There is no coordinated effort to assure that one stakeholder's guidance does not conflict with others. Authors and others with responsibility for AI's use in publications must continuously seek out the most current information and should review the unabridged guidance/requirements from their originating source. It should never be assumed that any use or disclosure of AI is compliant without this understanding.

The use of machine-assisted tools is not new in drafting publications. Specifically, for decades authors have used search engines instead of card catalogues, had spreadsheet software generate graphs from data instead of drawing the graph themselves, and had misspelled words or bad grammar automatically corrected by the software instead of leaving the errors for a human editor to find and correct. Being largely human-driven and relatively benign, those that oversaw compliance and integrity of clinical publications had little cause for concern. With the emergence of generative AI technology that can create much more meaningful text, images, videos, and audio content, the conduct and oversight of research publications and other scholarly activity have become more complicated and caused the introduction of new guidance specifically on the use of such technology.

Unlike the computer-assisted tasks in the past, generative AI such as LLMs can now draft large portions (if not all) of the scholarly publication text traditionally done by authors and/or medical writers. This introduced the question as to whether the generative AI is considered an "author" if it essentially generated the text from only minimal human prompting. Faced with this conundrum, in January 2024 the ICMJE updated its Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals to address the emergence of LLMs and other AI-assistive tools.

Essentially, the ICMJE sets out two key requirements. The first reminds human authors that AI is only a tool and, no matter the extent of the work it does, it is not to be listed as an author because the AI cannot meet the criteria for authorship, as the AI "cannot be responsible for the accuracy, integrity, and originality of the work."[67] The second was that ICMJE-abiding journals are to require authors to (i) disclose whether and how they used AI-assisted technologies in the production of submitted work; (ii) carefully review and edit the work for accuracy (including but not limited to the verification of citations), completeness, and elimination of bias; and (iii) assert that there is no plagiarism in their paper, including in the text and/or images produced by the AI. Overall, the updated ICMJE guidance offers a reminder to the human authors that they are responsible for all aspects of the submitted material that includes the use of AI-assisted technologies.[68]

As copyright protection is of key importance to competing stakeholders, the ownership and protection of submitted work products with content generated by AI (be it text, audio, visual/graphic, or other) remains hotly disputed across all industries. Countries are starting to clarify their opinion on how generative AI content is or is not protectable in their copyright laws. For example, the United States copyright office states "most fundamentally, the term 'author,' which is used in both the Constitution and the Copyright Act, excludes non-humans."[69]

Fortunately, this is not a new paradigm, as how countries and others have addressed new technologies in the past seems to serve as the model for their decisions on AI, specifically considering the use of tools and the level of involvement of the author(s). As long ago as 1884 in a case regarding ownership of a photograph of writer Oscar Wilde, the U.S. Supreme Court differentiated between the "author" (in this case the photographer) whose "ideas in the mind are given visible expression") and the tool (in this case the camera that took the picture).[70]

To further the assertion that the author owned the copyright of a photograph of an already widely photographed human being at the time, the U.S. Supreme Court pointed out that the photograph in question was not a passive one (as if anyone could have taken it), but was meaningfully influenced by the author (photographer) "by posing the said Oscar Wilde in front of the camera, selecting and arranging the costume, draperies, and other various accessories in said photograph, arranging the subject so as to present graceful outlines, arranging and disposing the light and shade, suggesting and evoking the desired expression."[71] Fast forward to the use of generative AI, the same concept applies to the current U.S. copyright opinion that, "AI-generated content that is more than de minimis should be explicitly excluded from the [copyright] application."[72] The less the authors direct and manipulate the AI outcome, the less ownership/copyright defense they have and the less they or their institution is able to assign that copyright to a publisher.

Specifically speaking, under the above legal theory any text/graphic/audio that contains more than a de minimis amount of AI-generated content is generally not considered copyrightable and would essentially be, for publication and ownership purposes, no different than submitting material that is already in the public domain. While there is no definitive line on how much author manipulation is needed to cross over the threshold of ownership, an example might be that an AI-generated paragraph or graphic, even with some slight modifications and corrections by the author, might be considered not copyrightable and thus essentially no different than a public domain graphic.

Thus, neither the author(s), their institution, nor the publisher can claim ownership protection of that intellectual property. As a result, some journals crafted policies that outright deny acceptance of any AI-assisted content, although some have softened their stance and harmonized with the new ICMJE requirement on disclosure.[73]

There is a fundamental difference between using an AI-assistive tool as an aid versus using AI to generate the content.

While the stakeholders are evolving in their stances when it comes to using generative AI in drafting publications, two key themes seem to rise to the top of authors' responsibilities.

The author(s) must differentiate human contributions from non-human contributions and disclose such differentiation as part of the scholarly work. This not only applies to author(s), but also their contributors and others that assist.

1. As discussed above, there is a fundamental difference between using an AI-assistive tool as an aid versus using AI to generate the content. For example, asking for an AI chatbot for the top 10 articles you should consider when drafting the Background section of your publication (and then you go and read those articles) is fundamentally different from asking an LLM to draft your Background section for you based on a few prompts. Historic publication norms have generally not required the disclosure of predominately human-directed AI assistive technology (e.g., computer-assisted design software to draw graphics), but now that AI can generate content that is nearly indistinguishable from human-generated material with very little human prompting, the publishers are taking notice and requiring such disclosures (especially if they cannot own the copyright of the material as they otherwise would have if human-generated).

"Before you put your name on it as an author, check it in its entirety as if you asked a fifth-grader to draft it for you."

2. Because the author(s) have final accountability for the work and they (or their employing institution) control its copyright assignments, the authors must also know if their contributors used generative AI when aiding in the publication. For example, if the author(s) engage a graduate student to draw a graphic of the digestive system and the student just supplies an unaltered or minimally altered AI-generated graphic, under current legal standards the authors or their employing institution may not own copyright of that graphic and under ICMJE standards the authors would have to disclose to the publishers that the graphic has more than a de minimis amount of generative AI.

The author(s) remain fully accountable for their final work product.

1. There are countless amusing and not-so amusing examples on the internet of AI creating partially or completely falsified, fabricated, plagiarized, outdated, biased, incomplete, or non-sensical content delivered as fact. No different from an author's review of each of their co-authors' and their contributors' work in general, authors must also validate the content that AI produces for accurate and appropriateness of use. Authors essentially stake their career reputation on the content they sign off on, thus it is critical that they review it. When it comes to using LLMs in this manner, many people have said "treat it as if your grad student wrote it," but this may be further elaborated on a la "Before you put your name on it as an author, check it in its entirety as if you asked a fifth grader to draft it for you."

2. Authors should be aware of the source material that developed the algorithms and their limitations on use for at least three key reasons. The first is that the source material may contain copyrighted, confidential, or even personal information that the AI may leak out into the work product. Second is that there is no assurance that the work product is immune from being plagiarized. Third, the use of the source material to train the algorithms may have violated a law (such as a privacy law), license for use of the data, or personal consent.

3. The use of AI-assistive tools does not make the author(s) immune to regulatory obligations. For example, the U.S. Office of Research Integrity defines research misconduct as falsification, fabrication, and plagiarism. There are numerous ways AI-assisted output can cause an author to be non-compliant with these principles. Below are but a few:

- **Falsification:** Gone are the days when those who want to falsify images would be limited to shifting, rotating, or reversing them or manually altering, stamping, or patching them as best they could with image editing software. The difficult-to-detect nature, inexpensiveness, availability, and ease of use of AI-enhanced image manipulation is presenting a growing threat.[74] Fortunately, while AI technology can manipulate images, AI technology is improvingly able to detect if an image has been manipulated.[75] However, this seemingly will lead to a perpetual cat-and-mouse game similar to how computer virus and antivirus software keep outpacing each other, with costs to create malicious content going down while costs to build more sophisticated models to detect it go up.

- **Fabrication:** It is well documented that LLMs can create medical publications that are nearly indistinguishable from human writing.[76-78] However, Italian scholars took this a step further and wanted to evaluate whether an LLM's ability to create supporting datasets was just as good as its ability to create manuscript text. They prompted a publicly available AI tool to create a dataset of 250 patients that supported a claim of superiority of deep anterior lamellar keratoplasty over penetrating keratoplasty for sufferers of keratoconus. It took only minutes with minimal prompting to do so, and the authors reported that the resulting dataset was scarily indistinguishable from data that would have been gathered in an actual clinical trial.[79]

- **Plagiarism:** AI tools such as LLMs base their responses on the data they are trained on. In many cases they produce verbatim passages substantially equivalent to their training data without citation that, if done by a human, would be considered plagiarism. Many common LLMs accept commands such as "reword" or "paraphrase" that can rearrange and/or replace words and phrases to restate the text, but arguably even after multiple revisions it would still be more than a de minimis amount of AI-generated content. Even so, this begs the question of whether such functionality truly represents the author's own interpretation or is an unethical avoidance of being caught for plagiarism by AI scanners that publishers use. In fact, there are online text and video tutorials to assist authors on how to get past plagiarism and AI scanners (of note, prior to showing you how to avoid plagiarism checkers, one YouTube video opens with the disclaimer: "This video is purely for educational purposes…Plagiarism is unethical. You should never plagiarize. Always cite sources used.").[80]

- **Citations:** Of note, although it will undoubtedly improve, current generative AI is well known to provide citations that either don't exist (generally called "hallucinations") or do exist but have nothing to do with the text citing them.[81,82] Interestingly enough, the below experiment (see Figure 9) demonstrates that AI can even know that it is hallucinating citations.

CITING AI

Best practices for proper citation of the use of AI assistive tools (in most cases LLMs) are currently without consensus. While some (e.g., Chicago Manual of Style 17th Edition) may support that using a Chatbot is to be cited in the same manner as a personal communication, others (e.g., APA Style) deem the "personal communication" classification inappropriate, as there is no second person communicating, thus it should be documented as an "algorithm output" citing demographics about the SI/LLM used.[83] Further complicating the issue is that it often takes several prompts to get the AI to generate the outcome desired, and thus for others to be able to replicate it would require the disclosure of a lengthy string of prompts and responses.

Even so, the same generative AI program may give inconsistent answers across users even with the same prompts as it "learns" over time. Adding to the issue is that one person's particular answers (even if saved with a unique identifier by the AI vendor) cannot be accessed by others unless they provide them with their login and password. Nevertheless, the major organizations that put forth the most commonly accepted citation styles have made initial indications on how they desire AI to be cited in scholarly work. Table 6 provides these as they are listed at the time of drafting, and anyone authoring a publication should check the required style for the most current guidance.

Figure 9: Example of Self-aware AI-fabricated Citations Prompted by Noah Zanville, PhD, RN, Director of Nursing Research for HCA Healthcare. Responses Generated March 13, 2024, using Commercially Available AskAI® Application on Open AI's Chat GPT (Ver 4.0) Platform.

ENHANCING QUALITY AND EXPOSING FLAWS IN THE PEER-REVIEW SYSTEM

AI has certainly aided the peer-review process with triaging manuscripts, detecting plagiarism, identifying methodological flaws, and reducing reviewer fatigue by taking on rote tasks and bringing other methods of scale over human effort.[84-86] Reviewers have found that use of an LLM chatbot can aid in critiquing writing style and readability. However, in one case it fell short when prompted for more specific comments to improve the text and provided incorrect statistical criticisms and non-existent "relevant" citations about the subject matter.[87] Nevertheless, as in many areas, AI shows promise to aid in scaling many of the rote or repetitive tasks of peer reviewers so that they can better utilize their professional skills in their limited time.

However, a weak peer-review system can still be exposed by AI-generated content. Incorrect and downright nonsensical content is not only getting past authors, but also exposing weaknesses in the peer-review mechanism that exists to instill public trust in scientific publications. One highly lauded example (see Figures 10 and 11) was outed not by communication with the journal, but via a prevalent social media platform almost immediately after publication for failure of the human peer reviewers to catch a graphic containing an anatomically incorrect rat and affiliated nonsensical labels. This case was particularly troubling, as the authors actually disclosed in the manuscript that they used AI to generate the images.[88] The entire article has since been retracted by the journal.[89]

While many medical journals for multiple reasons remain concerned about AI-generated graphics, most still consider publishing them with the accompanying disclosures; however, some journals have put a hold on accepting any AI-generated or -augmented images. For example, Nature announced in 2023 that while it will still consider the inclusion of text that has been produced with the assistance of generative AI, it has put a hold on AI-generated or -augmented images. [90]

Overall, the same paradigm of AI adding both value and problems also applies to the publication stage of clinical research. While many debates on copyright ownership and transparency of use remain in evolution, at least one thing seems to remain constant: author(s) are still responsible and ultimately accountable for the work they submit for publication.

Table 6: Common Citations for AI

Style	In-Line Text	Reference Format
American Psychological Association (APA Style) 17th edition [see source for full description and additional examples] Source: https://apastyle.apa.org/blog/how-to-cite-chatgpt	When prompted with "Is the left brain right brain divide real or a metaphor?" the ChatGPT-generated text indicated that although the two brain hemispheres are somewhat specialized, "the notation that people can be characterized as 'left-brained' or 'right-brained' is considered to be an oversimplification and a popular myth" (OpenAI, 2023). When given a follow-up prompt of "What is a more accurate representation?" the ChatGPT-generated text indicated that "different brain regions work together to support various cognitive processes" and "the functional specialization of different regions can change in response to experience and environmental factors" (OpenAI, 2023; see Appendix A for the full transcript).	OpenAI. (2023). ChatGPT (Mar 14 version) [Large language model]. https://chat.openai.com/chat
American Medical Association (AMA Style)	AMA does not provide in-line examples but gives general descriptions. From 5.2.1.1: "Authors should report the use of artificial intelligence, language models, machine learning, or similar technologies to create content or assist with writing or editing of manuscripts in the Acknowledgment section or the Methods section if this is part of formal research design or methods. This should include a description of the content that was created or edited and the name of the language model or tool, version and extension numbers, and manufacturer. (Note: this does not include basic tools for checking grammar, spelling, references, etc.)". From 14.5.2 "For AI tools, including large language models (LLMs), machine learning, generative AI such as chatbots, and related technologies, use a nonproprietary descriptive term. In research articles, at first mention in the abstract and in the text, provide the brand name in parentheses along with the version or extension number, manufacturer or owner, and date(s) used."]	On June 12, 2023, the original full text of the question was put into a fresh chatbot session (ChatGPT, model GPT-4, OpenAI) and the generated responses were saved. Or Data were collected from the artificial intelligence chatbot ChatGPT version 4 (OpenAI).
Chicago Manual of Style	The following recipe for pizza dough was generated by ChatGPT.	Example: Text generated by ChatGPT, OpenAI, March 7, 2023, https://chat.openai.com/chat Or (if the prompt is not in the text): ChatGPT response to "Explain how to make pizza dough from common household ingredients," OpenAI, March 7, 2023. https://chat.openai.com/chat
Modern Language Association (MLA Style)	While the green light in The Great Gatsby might be said to chiefly symbolize four main things: optimism, the unattainability of the American dream, greed, and covetousness ("Describe the symbolism"), arguably the most important–the one that ties all four themes together–is greed.	"Describe the symbolism of the green light in the book The Great Gatsby by F. Scott Fitzgerald" prompt. ChatGPT, 13 Feb. version, OpenAI, 8 Mar. 2023, chat.openai.com/chat.
MLA Style	When asked to describe the symbolism of the green light in The Great Gatsby, ChatGPT provided a summary about optimism, the unattainability of the American dream, greed, and covetousness. However, when further prompted to cite the source on which that summary was based, it noted that it lacked "the ability to conduct research or cite sources independently" but that it could "provide a list of scholarly sources related to the symbolism of the green light in The Great Gatsby" ("In 200 words").	"In 200 words, describe the symbolism of the green light in The Great Gatsby" follow-up prompt to list sources. ChatGPT, 13 Feb. version, OpenAI, 9 Mar. 2023, chat.openai.com/chat.
MLA Style		"Pointillist painting of a sheep in a sunny field of blue flowers" prompt, DALL-E, version 2, OpenAI, 8 Mar. 2023, labs.openai.com/.

Figure 10: Example of an AI-Generated Image That Got Past Peer Review and Was Published

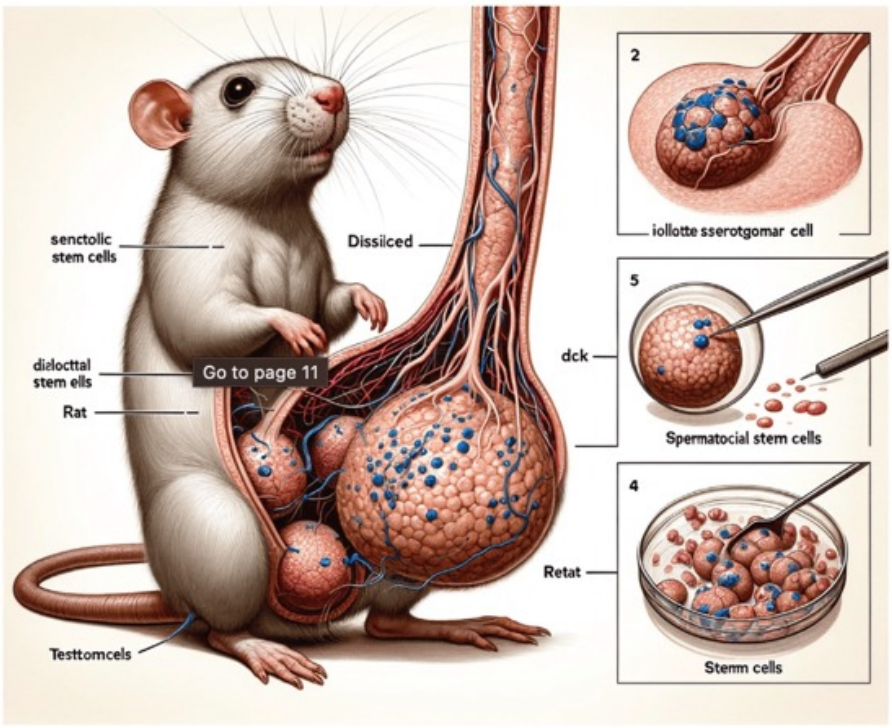


FIGURE 1
Spermatogonial stem cells, isolated, purified and cultured from rat testes.

Figure 11: Example of an AI-Generated Image That Got Past Peer Review and Was Published

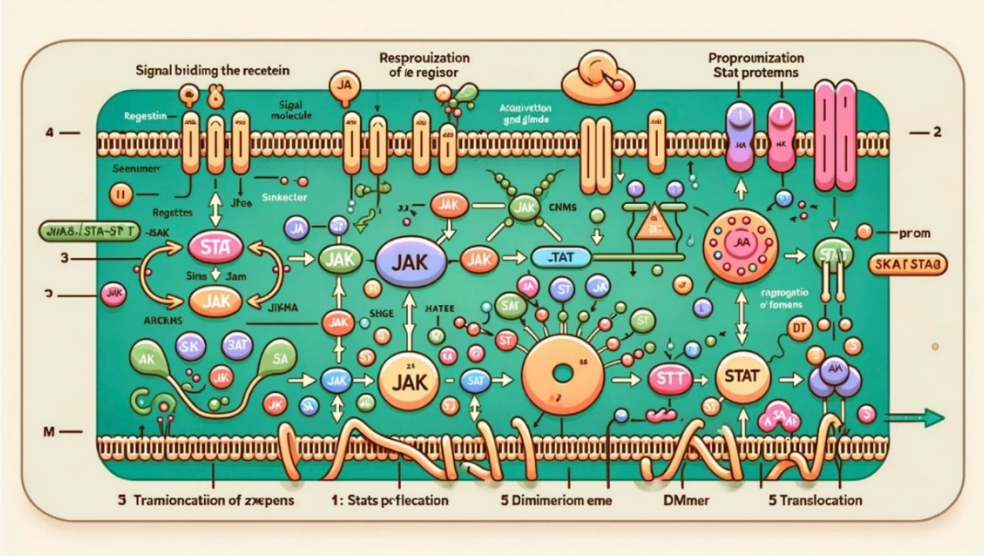


FIGURE 2
Diagram of the JAK-STAT signaling pathway: 1) Signal molecule binding to the receptor, 2) Activation of JAK kinase and phosphorylation of the receptor, 3) Recruitment and phosphorylation of STAT proteins by JAK, 4) Dimerization of STAT proteins, 5) Translocation of STAT dimers into the nucleus and initiation of gene transcription.

Appendix 5: Acknowledgements

This publication could not have been possible without the help of many individuals and organizations. Thanks to HCA Healthcare for the use of its resources to assist in the literature search for background information and examples.

Special thanks go to my treasured ACRP Fellows colleagues Jeri Burr, Suzanne Kinkaid, Joy Jurnack, Jerry Stein, Kelly Cairns, and Debbie Rosenbaum for providing editorial review and additional content. Additional thanks go to all the other ACRP Fellows who gave guidance and inspiration on our Happy Hour calls. Thanks also go to Susan Landis and the ACRP staff for the opportunity to create this paper and distribute it under the Association's banner.

Finally, thanks go to all my generous and valued colleagues who provided me with examples, articles, responses, discussions, or other inspirations for me to do my very best. In alphabetical order—with apologies for anyone I excluded or mis-credentialed—they are Todd Albin, Bryan Allinson, Yonjong Choi, Tricia De Cambra, Glenda Guest, Mark King, Stacey Klump, Michael Koren, MD, Lisa Moneymaker, Susan Rockwell, Christine Senn, PhD, Darin Sims, Garrett Sims, Nicholas Spittal, and Noah Zanville.

About the Author: David Vulcano is an active thought leader and change agent in the clinical research industry through his involvement with numerous associations, boards, initiatives, publications, and presentations. His primary employment is as the Vice President for Research Compliance & Integrity at HCA Healthcare, providing research-related compliance and strategic development to its portfolio of hospitals, physician practices, and healthcare technology companies. He is and has been in many industry leader roles both in the United States and globally, including as a former Chair of the Board of Trustees for the Association of Clinical Research Professionals (ACRP), Honorary President for the Society of Clinical Research Sites (SCRS), and Advisory Board member for the Society for Clinical Data Management (SCDM). He holds a Private Company Governance Certificate through the Private Directors Association and has passed the FINRA-SIE exam. He also serves as President of the Music City Angels group for investments, as a mentor, and in the occasional board seat in diverse startup companies. For disclosures and affiliations, visit linkedin.com/in/davidvulcano.

Access Related Content in the ACRP Artificial Intelligence Resource Center!

ABOUT ACRP: From new clinical research professionals just starting out to industry veterans who are looking for ways to move ahead in their career, the Association of Clinical Research Professionals (ACRP) is where success starts—and grows. With more than 17,000 global members, ACRP is the strongest advocate for clinical research professionals—amplifying their voice throughout the research industry and elevating their position by providing gold-standard education, rigorous certification, and strong community connections. Learn more at ACRPNET.ORG.

Appendix 6: References

1 https://www.britannica.com/biography/John-McCarthy (Accessed June 20, 2024)

2 Samuel AL. "Some Studies in Machine Learning Using the Game of Checkers," in *IBM Journal of Research and Development*, vol. 3, no. 3, pp. 210-229, July 1959, doi: 10.1147/rd.33.0210

3 Pew Research Center. Public Awareness of Artificial Intelligence In Everyday Activities. February 15, 2023. https://www.pewresearch.org/science/2023/02/15/public-awareness-of-artificial-intelligence-in-everyday-activities/

4 Howell MD, Corrado GS, DeSalvo KB. Three Epochs of Artificial Intelligence in Health Care. JAMA. 2024;331(3):242–244. doi:10.1001/jama.2023.25057

5 Understanding Gartner's Hype Cycles. Published: 20 August 2018. https://www.gartner.com/en/documents/3887767

6 Tuckman BW. (1965). "Developmental sequence in small groups." *Psychological Bulletin*. 63 (6): 384–399. doi:10.1037/h0022100. PMID 14314073

7 United States Code Title 15, Chapter 119-National Artificial Intelligence Initiative (15 U.S.C. 9401(3))

8 https://www.britannica.com/technology/artificial-intelligence (Accessed June 17, 2024)

9 UNESCO Recommendation on the Ethics of Artificial Intelligence (23 November 2021)

10 UNESCO Recommendation on the Ethics of Artificial Intelligence (23 November 2021)

11 https://aws.amazon.com/what-is/gpt/ (Accessed June 17, 2024)

12 https://www.cloudflare.com/learning/ai/what-is-large-language-model/ (Accessed June 17, 2024)

13 IMDRF/AIMD WG/N67 Machine Learning-enabled Medical Devices: Key Terms and Definitions, final document, May 6, 2022.

14 https://www.ibm.com/topics/machine-learning (Accessed June 17, 2024)

15 https://aws.amazon.com/what-is/nlp/ (Accessed June 17, 2024)

16 https://www.ibm.com/topics/natural-language-processing (Accessed June 17, 2024)

17 The Artificial Intelligence (AI) global regulatory landscape (https://www.ey.com/content/dam/ey-unified-site/ey-com/en-gl/insights/ai/documents/ey-the-artificial-intelligence-ai-global-regulatory-landscape-final.pdf)

18 Article 29, European Union AI Act (2024)

19 Treasury Board Directive on Automated Decision-Making, 3rd Review https://wiki.gccollab.ca/images/9/94/Report_on_the_third_review_of_the_Directive_on_Automated_Decision-Making.pdf)

20 New York City Law #2021/144: Automated Employment Decision Tools (AEDT)

21 Good Machine Learning Practice for Medical Device Development: Guiding Principles (October 2021) https://www.fda.gov/media/153486/download

22 United States White House Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (October 30, 2023)

23 United States Federal Register. Vol. 84, No. 210 / Wednesday, October 30, 2019. "Request for Comments on Intellectual Property Protection for Artificial Intelligence Innovation". p 58141-58142.

24 Liu Q, Huang R, Hsieh J, Zhu H, Tiwari M, Liu G, Jean D, ElZarrad MK, Fakhouri T, Berman S, Dunn B, Diamond MC, Huang S-M. (2023), Landscape Analysis of the Application of Artificial Intelligence and Machine Learning in Regulatory Submissions for Drug Development From 2016 to 2021. *Clin Pharmacol Ther*, 113: 771-774. https://doi.org/10.1002/cpt.2668

25 Presented at FDA Hybrid Public Workshop: Artificial Intelligence in Drug & Biological Product Development August 6, 2024.

26 National Association of Corporate Directors "AI And Board Governance." 2023. https://www.nacdonline.org/contentassets/8363cc8217c9484da1314efb0403ff94/de_ai_nacd_dta_final_92023.pdf

27 Chertoff M, Grafman A. The Director's AAI Checklist. *Directors & Boards*. Vol. 48, No. 3. October 2024. p44-47.

28 Carlini, et al. Extracting Training Data from Large Language Models. *Cryptography and Security* (15 JUN 2021) https://arxiv.org/abs/2012.07805v2

29 OpenAI Press Release (March 24, 2023) "March 20 ChatGPT outage: Here's what happened." https://openai.com/index/march-20-chatgpt-outage/ (Accessed July 10, 2024).

30 NIH Review Matters: Using AI in Peer Review Is A Breach of Confidentiality (June 23, 2023). https://www.csr.nih.gov/reviewmatters/2023/06/23/using-ai-in-peer-review-is-a-breach-of-confidentiality/

31 See June 27, 2023 – July 7, 2023 comments on https://www.csr.nih.gov/reviewmatters/2023/06/23/using-ai-in-peer-review-is-a-breach-of-confidentiality/

32 NIH Frequently Asked Questions (FAQs): Use of Generative AI in Peer Review. https://grants.nih.gov/faqs#/use-of-generative-ai-in-peer-review.htm?anchor=56948 (accessed July 10, 2024).

33 Blackman R. (2021, April 1). If Your Company Uses AI, It Needs An Institutional Review Board. *Harvard Business Review*.

34 Westerman G, Ransbotham S, Farronato C. "Find the AI Approach That Fits the Problem You're Trying to Solve," *Harvard Business Review*, February 06, 2024. https://hbr.org/2024/02/find-the-ai-approach-that-fits-the-problem-youre-trying-to-solve

35 Suh B. "When Should You Use AI to Solve Problems?" *Harvard Business Review*. February 17, 2021. https://hbr.org/2021/02/when-should-you-use-ai-to-solve-problems

36 Howell MD. Generative artificial intelligence, patient safety and healthcare quality: a review. *BMJ Quality & Safety* Published Online First: 24 July 2024. doi: 10.1136/bmjqs-2023-016690

37 Generative AI: Can We Trust It? An Interview with Prof. Gary Smith. Pomona College (January 13, 2023) https://www.pomona.edu/news/2023/01/13-generative-ai-can-we-trust-it-interview-prof-gary-smith (Accessed July 23, 2024)

38 Kelly J. "Google's AI Recommended Adding Glue To Pizza And Other Misinformation—What Caused The Viral Blunders?" *Forbes*. May 31, 2024 (https://www.forbes.com/sites/jackkelly/2024/05/31/google-ai-glue-to-pizza-viral-blunders/) Accessed July 23, 2024.

39 Reid L. "AI Overviews: About last week" (May 30, 2024). https://blog.google/products/search/ai-overviews-update-may-2024/ (Accessed July 23, 2024).

40 Smith G. "Internet Pollution — If You Tell A Lie Long Enough…" *Mind Matters*. (January 15, 2024) https://mindmatters.ai/2024/01/internet-pollution-if-you-tell-a-lie-long-enough/

41 Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019 Oct 25;366(6464):447-453. doi: 10.1126/science.aax2342. PMID: 31649194.

42 Ledford H. Millions of black people affected by racial bias in health-care algorithms. Nature. 2019 Oct;574(7780):608-609. doi: 10.1038/d41586-019-03228-6. PMID: 31664201.

43 Panch T, Mattie H, Atun R. Artificial intelligence and algorithmic bias: implications for health systems. J Glob Health. 2019 Dec;9(2):010318. doi: 10.7189/jogh.09.020318. PMID: 31788229; PMCID: PMC6875681.

44 Chang T, Nuppnau M, He Y, Kocher KE, Valley TS, Sjoding MW, et al. (2024) Racial differences in laboratory testing as a potential mechanism for bias in AI: A matched cohort analysis in emergency department visits. PLOS Glob Public Health 4(10): e0003555. https://doi.org/10.1371/journal.pgph.0003555

45 How will AI affect cyber security attacks? *Cyberhive* (7 Feb 2024) https://www.cyberhive.com/insights/how-will-ai-affect-cyber-security-attacks/ (Accessed June 26, 2024)

46 Indian pharma takes the digital leap: What does it mean for cybersecurity? Deliotte. June 2022. https://www2.deloitte.com/content/dam/Deloitte/in/Documents/risk/DS-CI-Deloitte-Indian-Pharma-noexp.pdf

47 Software as a Medical Device (SaMD): Key Definitions (2013): http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-131209-samd-keydefinitions-140901.pdf

48 "Software as a Medical Device": Possible Framework for Risk Categorization and Corresponding Considerations (18 September 2014) https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-140918-samd-framework-risk-categorization-141013.pdf

49 Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices

50 Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) FDA-2019-N-1185-0001_attachment_1.pdf; Predetermined Change Control Plans for Machine Learning-Enabled Medical Devices: Guiding Principles https://www.fda.gov/media/173206/download?attachment

51 Section 520(o)(1)(E) of the FD&C Act

52 Clinical Decision Support Software: Guidance for Industry and Food and Drug Administration Staff (September 28, 2022). https://www.fda.gov/media/109618/download

53 45CFR46.102(e)(1)(ii)

54 45CFR46.104(d)(4)(i)-(ii)

55 45CFR46.116(b)(9)

56 Vulcano D. Understanding De-Identified Patient Data, Limited Data Sets, and Data Use Agreements. Policy & Medicine Compliance Update. Volume 7.1, January 2021.

57 Vulcano D. Compliance with US Privacy Regulations when Using Health Records Data for Real World Evidence Purposes. Journal of the Society for Clinical Data Management. 2023; 3(3): 3, pp. 1–9. DOI: https://doi.org/10.47912/jscdm.233.

58 Askin S, Burkhalter D, Calado G, El Dakrouni S. Artificial Intelligence Applied to clinical trials: opportunities and challenges. *Health Technol* (Berl). 2023;13(2):203-213. doi: 10.1007/s12553-023-00738-2. Epub 2023 Feb 28. PMID: 36923325; PMCID: PMC9974218

59 ibid

60 1996, The Road Ahead by Bill Gates with Nathan Myhrvold and Peter Rinearson, Second Edition, Section: Afterword, Quote Page 316, Penguin Books, New York

61 AI-Driven Transformation: How SCDM and the Industry Are Walking the Talk. SCDM 2024 EMEA Leadership Forum & Conference (April 19, 2024). https://learning-scdm.org/courses/70411

62 Carter D. "Leading Through Transition: Four Strategies To Empower People To Embrace Workplace Change". Forbes. (2018 Oct 04). https://www.forbes.com/sites/forbeshumanresourcescouncil/2018/10/04/leading-through-transition-four-strategies-to-empower-people-to-embrace-workplace-change/

63 Howell MD, Corrado GS, DeSalvo KB. Three Epochs of Artificial Intelligence in Health Care. JAMA. 2024;331(3):242–244. doi:10.1001/jama.2023.25057

64 Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. https://www.hhs.gov/hipaa/for-professionals/special-topics/de-identification/index.html. Version August 22, 2024. Accessed December 6, 2024.

65 FDA Informed Consent Guidance for IRBs, Clinical Investigators and Sponsors (August 2023)

66 Ghosh R, Kempf D, Pufko A, et al. Automation Opportunities in Pharmacovigilance: An Industry Survey. *Pharm Med* 34, 7–18 (2020). https://doi.org/10.1007/s40290-019-00320-0

67 Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals (January 2024). https://www.icmje.org/icmje-recommendations.pdf

68 Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals (January 2024). https://www.icmje.org/icmje-recommendations.pdf

69 United States Copyright Office "Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence." PUBLISHED ON THURSDAY, MARCH 16, 2023.

70 *Burrow-Giles Lithographic Co. v. Sarony*, 111 U.S. 53 (1884)

71 *Burrow-Giles Lithographic Co. v. Sarony*, 111 U.S. 53 (1884)

72 United States Copyright Office "Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence." PUBLISHED ON THURSDAY, MARCH 16, 2023. https://copyright.gov/ai/ai_policy_guidance.pdf

73 Thorp HH, Vinson V. Change to policy on the use of generative AI and large language models. Science. (16 NOV 2023) https://www.science.org/content/blog-post/change-policy-use-generative-ai-and-large-language-models (accessed June 26, 2024)

74 Gu J, Wang X, Li C, Zhao J, Fu W, Liang G, Qiu J. AI-enabled image fraud in scientific publications. Patterns (N Y). 2022 Jul 8;3(7):100511. doi: 10.1016/j.patter.2022.100511. PMID: 35845832; PMCID: PMC9278510.

75 L. Verdoliva, "Media Forensics and DeepFakes: An Overview," in IEEE Journal of Selected Topics in Signal Processing, vol. 14, no. 5, pp. 910-932, Aug. 2020, doi: 10.1109/JSTSP.2020.3002101

76 Elali FR, Rachid LN. AI-generated research paper fabrication and plagiarism in the scientific community. Patterns (N Y). 2023 Mar 10;4(3):100706. doi: 10.1016/j.patter.2023.100706. PMID: 36960451; PMCID: PMC10028415.

77 Else H. Abstracts written by ChatGPT fool scientists. Nature. 2023 Jan;613(7944):423. doi: 10.1038/d41586-023-00056-7. PMID: 36635510.

78 Májovský M, Černý M, Kasal M, Komarc M, Netuka D. Artificial Intelligence Can Generate Fraudulent but Authentic-Looking Scientific Medical Articles: Pandora's Box Has Been Opened. J Med Internet Res 2023;25:e46924 URL: https://www.jmir.org/2023/1/e46924 DOI: 10.2196/46924

79 Taloni A, Scorcia V, Giannaccare G. Large Language Model Advanced Data Analysis Abuse to Create a Fake Data Set in Medical Research. JAMA Ophthalmol. 2023;141(12):1174–1175. doi:10.1001/jamaophthalmol.2023.5162

80 https://www.youtube.com/watch?v=yiP4ZSikvbo. Accessed May 24, 2024.

81 Alkaissi H, McFarlane SI. Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. Cureus. 2023 Feb 19;15(2):e35179. doi: 10.7759/cureus.35179. PMID: 36811129; PMCID: PMC9939079.

82 Colasacco CJ, Born HL. A Case of Artificial Intelligence Chatbot Hallucination. JAMA Otolaryngol Head Neck Surg. Published online April 18, 2024. doi:10.1001/jamaoto.2024.0428

83 https://apastyle.apa.org/blog/how-to-cite-chatgpt Accessed May 28, 2024

84 Bauchner H, Rivara FP. Use of artificial intelligence and the future of peer review. Health Aff Sch. 2024 May 3;2(5):qxae058. doi: 10.1093/haschl/qxae058. PMID: 38757006; PMCID: PMC11095530.

85 Leung TI, de Azevedo Cardoso T, Mavragani A, Eysenbach G. Best Practices for Using AI Tools as an Author, Peer Reviewer, or Editor. J Med Internet Res 2023;25:e51584. doi: 10.2196/51584

86 Hosseini M, Horbach SPJM. Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other large language models in scholarly peer review. Res Integr Peer Rev 8, 4 (2023). https://doi.org/10.1186/s41073-023-00133-5

87 Donker T. The dangers of using large language models for peer review. Lancet Infect Dis. 2023 Jul;23(7):781. doi: 10.1016/S1473-3099(23)00290-6. Epub 2023 May 10. PMID: 37178707.

88 How AI Images Sneaked Past Scientific Scrutiny. https://www.msn.com/en-us/news/technology/how-ai-images-sneaked-past-scientific-scrutiny/ar-BB1isK04. Accessed May 24, 2024.

89 Frontiers Editorial Office (2024) Retraction: Cellular functions of spermatogonial stem cells in relation to JAK/STAT signaling pathway. Front Cell Dev Biol. 12:1386861. doi: 10.3389/fcell.2024.1386861

90 Why Nature will not allow the use of generative AI in images and video. Nature. 2023 Jun;618(7964):214. doi: 10.1038/d41586-023-01546-4. PMID: 37286651.