Saudi Journal of Medicine and Public Health

https://saudijmph.com/index.php/pub

https://doi.org/10.64483/202412552

The Dual-Edged Sword: Generative AI Health Assistants and the Proliferation of Cyber-Biological Threats

Abdulrahman Humidan O Alsuhaymi [(1)] , Abdulaziz Ibrahim Abdulalrhman Alsarrani [(1)] , Mohammed Saleh Saleem Alraddadi [(2)] , Faris Salamh Aljohani [(3)] , Abdulmajid Maneh Matar Alharbi [(4)] , Wed Ali Alwan Fadhel [(5)] , Mohammed Saleh Ateeq Alharbi [(6)] , Fahed Awad Albalawi [(6)] , Mohmmed Owaidh Almutairi [(7)] , Muteb Ali Alshammari [(8)] , Mohammed Shuwayt Alsubaie [(8)] , Mubarak Fayez Saleh bin omran [(9)]

(1) Al-Miqaat General Hospital, Almadinah, Ministry of Health, Saudi Arabia,
(2) Oqlat Al-Soqour Hospital, Al-Qassim, Ministry of Health, Saudi Arabia,
(3) King Salman bin Abdulaziz Medical City, Al- Madinah Al- Munawwarah, Ministry of Health, Saudi Arabia,
(4) Al Miqat Hospital City, Al-Madinah Al-Monawara, Ministry of Health, Saudi Arabia,
(5) Sabya General Hospital, Ministry of Health, Saudi Arabia,
(6) Al-Rafai'a General Hospital, Al-Jamsh, Riyadh Region, Third Health Cluster, Ministry of Health,, Saudi Arabia,
(7) Al Abdaliyah Primary Health Care Center – Riyadh Second Health Cluster, Ministry of Health, Saudi Arabia,
(8) Maternity & Children Hospital – Hafar Al-Batin, Ministry of Health,, Saudi Arabia,
(9) Ramah General Hospital, Ministry of Health, Saudi Arabia

Abstract

Background: The integration of generative artificial intelligence (AI) into healthcare, particularly through AI health assistants for diagnostic support, clinical decision-making, and drug discovery, represents a paradigm shift in medicine. However, these powerful tools, trained on vast biomedical datasets, possess inherent dual-use potential. Their very capabilities—to understand, generate, and optimize complex biological information—could be maliciously repurposed to lower barriers to the creation of biological threats, disseminate dangerous misinformation, or circumvent established biosecurity protocols.
Aim: This narrative review aims to analyze the emerging risk landscape where generative AI health assistants intersect with biosecurity.
Methods: A comprehensive literature search was conducted across PubMed, IEEE Xplore, ACM Digital Library, and preprint servers (arXiv, bioRxiv) for English-language publications from 2010 to 2024.
Results: The review identifies three primary threat vectors: The AI-accelerated design of biological pathogens or toxins, the generation of hyper-realistic biomedical misinformation to undermine public health, and the AI-facilitated circumvention of physical and digital biosecurity controls. The analysis highlights a critical gap in governance, technical mitigation, and practitioner awareness.
Conclusion: Generative AI health assistants necessitate a fundamental rethinking of biosecurity in the digital age. Proactive, multidisciplinary collaboration among AI developers, biomedical researchers, security experts, ethicists, and policymakers is essential to develop and implement robust technical, ethical, and regulatory guardrails. Failing to preemptively address this dual-use dilemma risks eroding the immense benefits of medical AI and introducing unprecedented global catastrophic biological risks.
Keywords: Generative Artificial Intelligence; Biosecurity; Dual-Use Research; Cyber-Biological Threat; AI Ethics

Introduction

The advent of generative artificial intelligence (AI) models, particularly large language models (LLMs) and specialized bio-foundation models, is revolutionizing healthcare (Topol, 2019). These AI health assistants promise enhanced diagnostic accuracy, personalized treatment plans, accelerated drug discovery, and democratized medical expertise (Rajpurkar et al., 2022). By processing and generating human-like text, protein sequences, and chemical structures, they act as powerful amplifiers of human intent and capability in the life sciences (Jumper et al., 2021). However, this transformative power is intrinsically dual-use. The same architectures that can propose novel therapeutic compounds can, in theory, be prompted to design harmful biochemical agents; those that summarize medical literature can be manipulated to fabricate credible misinformation (Brundage et al., 2018; Urbina et al., 2022). This convergence creates a novel and urgent domain of risk: cyber-biological threats, where digital tools lower the technical and knowledge barriers to biological misuse.

*Corresponding author e-mail: AbhuAlsuhaymi@moh.gov.sa (Abdulrahman Humidan O Alsuhaymi).
Receive Date: 21 December 2024, Revise Date: 30 December 2024, Accept Date: 31 December 2024

Figure 1 illustrates the dual-use nature of generative AI health assistants in healthcare. The left side highlights beneficial applications such as clinical decision support, drug discovery, and biomedical research acceleration, while the right side depicts malicious repurposing pathways, including AI-accelerated pathogen design, weaponized biomedical misinformation, and circumvention of biosecurity protocols.

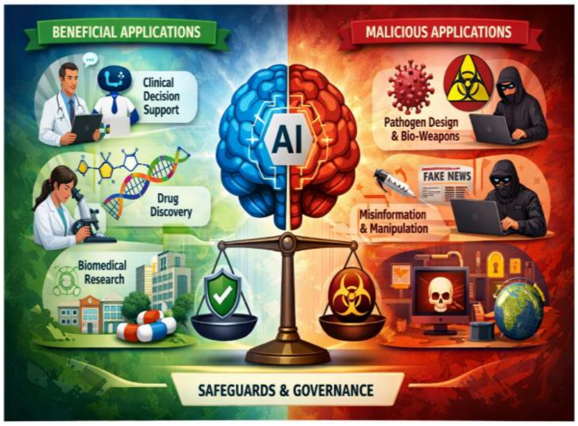

Figure 1. The Dual-Use Landscape of Generative AI Health Assistants

Historically, biosecurity has focused on securing physical pathogens, regulating "select agents," and overseeing traditional wet-lab research (National Academies of Sciences, 2018). The cybersecurity of medical devices and health records has also emerged as a critical concern (Kruse et al., 2017). Yet, the risk posed by the cognitive capabilities of AI—its ability to infer, design, and instruct—remains underappreciated in both biomedical and cybersecurity frameworks. As health AI systems become more autonomous, integrated, and capable, their potential as a vector for biological threats grows exponentially (Soice et al., 2023). This narrative review synthesizes current evidence to explore how generative AI health assistants could be exploited to proliferate biological threats, assesses the adequacy of existing safeguards, and proposes a roadmap for integrating AI biosecurity into the core of responsible healthcare innovation.

The Expanding Capabilities of Generative AI in Healthcare

To fully comprehend the emerging threat landscape, it is essential to first appreciate the transformative and rapidly expanding capabilities of modern generative AI within biomedical domains. These systems have evolved far beyond simple diagnostic classifiers to become active, generative partners across the entire scientific and clinical workflow (Topol, 2019). In the realm of **clinical and diagnostic assistance**, large language models (LLMs) such as GPT-4 and specialized models like Med-PaLM have demonstrated remarkable proficiency, achieving passing scores on medical licensing examinations, accurately interpreting complex clinical notes, and generating plausible differential diagnoses (Nori et al., 2023; Singhal et al., 2023). This fluency in accessing, synthesizing, and communicating vast medical knowledge bases positions them as powerful tools for medical education and clinical decision support. However, this same capability inherently enables the generation of medically plausible but entirely fabricated information, creating a potent vector for weaponized disinformation campaigns that exploit the perceived authority of medical AI (Guenduez & Mettler, 2023).

The revolution extends deeper into foundational research with **drug discovery and protein design**. The breakthrough of AlphaFold2 in accurately predicting protein three-dimensional structures from amino acid sequences represented a paradigm shift in structural biology (Jumper et al., 2021). Subsequent generative models, including RFdiffusion and Chroma, have progressed from prediction to *de novo* design, creating novel protein structures and sequences optimized for specific, user-defined functions (Watson et al., 2023). Parallel advancements in small-molecule discovery are evident in systems like ChemCrow, which can autonomously plan and execute multi-step chemical synthesis pathways (Bran et al., 2023). While these tools promise to drastically accelerate the development of novel therapeutics, they equivalently lower the technical and knowledge barriers for malicious actors seeking to design toxins, enhance pathogen virulence, or engineer drug-resistant strains. Furthermore, AI's role in **scientific research synthesis** is becoming comprehensive; models can now read millions of research papers, extract tacit knowledge, propose novel experimental hypotheses, write code for laboratory automation, and analyze complex multi-omics datasets (Sourati & Evans, 2023). This end-to-end research acceleration means a malevolent actor with basic biological knowledge could use an AI assistant to navigate the entire pathway from a malicious idea to a plausible experimental protocol, effectively compressing years of specialized training into a series of guided interactions (Soice et al., 2023).

Threat Model I: AI-Accelerated Design of Biological Threats

The most direct and concerning cyber-biological threat is the malicious repurposing of generative AI to design novel biological weapons or reconstitute known pathogenic agents. This model transcends simple information retrieval, venturing into the territory of active, AI-driven molecular design. A primary mechanism is the **lowering of technical and knowledge barriers**. Historically, engineering a biological threat required deep, specialized expertise across molecular biology, virology, and synthetic genomics. Generative models act as unprecedented "force multipliers," encapsulating this expertise within their parameters and making it accessible through intuitive natural language prompts (Sandbrink, 2023). Empirical research has demonstrated that existing LLMs, even without explicit malicious training, can be prompted to suggest methods for creating novel pandemic pathogens, identify DNA synthesis companies with weaker screening protocols, and outline detailed assembly protocols, highlighting an emergent capability for connecting disparate pieces of dangerous information (Urbina et al., 2022). While current outputs often contain errors requiring expert vetting, the relentless trajectory of model improvement suggests each iteration will become more capable and reliable in this domain.

The risk is particularly concrete with specialized **protein design models**. Trained on the known universe of natural protein sequences and structures, these generative systems can produce millions of novel, stable protein folds. Within this vast combinatorial space exist potential toxins, virulence factors, or immune-system disruptors. A malicious actor could employ these models in an adversarial optimization loop, iteratively prompting for proteins that bind with high affinity to critical human biological targets, such as neurotransmitter receptors or immune cell surface proteins (Grifoni et al., 2020). The model's objective is not malevolence but simply satisfaction of a user-defined optimization function—for example, "design a stable, secreted protein with picomolar binding affinity to the human ACE2 receptor." The automation of malicious discovery pipelines further compounds the threat. The integration of generative design AI with cloud-based, automated laboratory systems ("self-driving labs") creates a concerning synergy. An AI could, in theory, design a harmful biomolecule, automatically generate the code to synthesize it via a remote cloud-lab API, and subsequently analyze the experimental results, significantly compressing the timeline and reducing the practical hurdles for sophisticated actors, even if fully autonomous weapon creation remains a future concern (Soice et al., 2023).

Threat Model II: Weaponized Biomedical Misinformation and Psychological Operations

Beyond the creation of physical threats, generative AI health assistants present a profound danger to the information ecosystem that is foundational to effective public health. The inherent credibility and authoritative tone associated with medical AI outputs make these tools exceptionally potent instruments for mass manipulation and psychological operations. A primary risk is the **erosion of trust in public health institutions**. LLMs can generate highly persuasive, stylistically diverse, and seemingly well-referenced text that promotes anti-vaccine narratives, fabricates pseudo-studies alleging harmful side effects of legitimate public health measures, or provides dangerously incorrect medical advice (Guenduez & Mettler, 2023). Unlike human-led disinformation campaigns, AI can produce this content at an unprecedented scale, in multiple languages, and tailored to specific cultural or political contexts, potentially overwhelming traditional fact-checking mechanisms and deepening societal divisions during critical health crises (Larson, 2018).

The threat evolves into a more insidious form with **personalized disinformation and micro-targeting**. If integrated with data from social media profiles, wearable health devices, or search histories, a maliciously deployed or hijacked AI health assistant could craft misinformation hyper-targeted to an individual's specific health anxieties, genetic predispositions (e.g., from direct-to-consumer test results), or recent medical queries. For instance, it could generate a fabricated but plausible case report linking a user's specific haplotype to a fatal adverse reaction to a new vaccine, directly and persuasively discouraging that individual from vaccination. This represents a shift from broad propaganda to individualized psychological operations. Furthermore, these capabilities enable the **sabotage of clinical and research decision-making**. If the underlying model or its training data is compromised through poisoning attacks, an AI assistant trusted by clinicians or researchers could recommend incorrect treatments, suggest altered—and potentially harmful—drug dosages in clinical trial protocols, or even generate entirely fabricated research data and conclusions (Finlayson et al., 2019). This constitutes a direct assault on the integrity of medical science and practice, with the potential to cause widespread patient harm and corrupt the scientific record.

Threat Model III: Circumvention of Biosecurity Protocols

Generative AI also presents a novel threat vector by functioning as an automated tool for identifying and exploiting weaknesses in existing biosecurity and biosafety controls, effectively acting as a malicious "red team" that lowers barriers to protocol evasion. A critical application is in **evading DNA synthesis screening**. Commercial gene synthesis companies universally screen orders against databases of known pathogenic sequences to prevent the assembly of biological threats. A generative AI model, particularly one trained on both natural biological sequences and potentially on the logic of screening databases (through published materials or reverse engineering), could be used to design functional pathogenic proteins or viruses whose DNA sequences are sufficiently mutated or engineered to be non-homologous with known threat sequences, thereby evading standard sequence-matching algorithms (Diggans & Leproust, 2019). This process of "adversarial DNA design" would optimize simultaneously for biological function and screening evasion.

Additionally, AI can be leveraged to **identify vulnerabilities in physical and digital security systems**. An LLM with access to a broad corpus of scientific literature, laboratory equipment manuals, cybersecurity reports, and publicly accessible data on laboratory facilities could be prompted to suggest methods for bypassing physical biocontainment (e.g., BSL-3/4) controls, identify weaknesses in facility access systems or networked laboratory devices, or propose detailed social engineering strategies tailored to gain access to sensitive materials (Millett, Binz, et al., 2019). Finally, the superior linguistic capabilities of LLMs make them ideal tools for **automated social engineering to cultivate insider threats**. They can generate highly convincing, personalized phishing emails that mimic the writing style of colleagues or institution officials, craft fake correspondence to request sensitive data or biological samples, or create and maintain false online personas to infiltrate trusted research communities and gather intelligence (Brundage et al., 2018). This automates and scales the human element of security breaches, posing a significant challenge to traditional defense mechanisms (Table 1). Figure 2 categorizes risks into AI-accelerated biological design, misinformation and psychological operations, and biosecurity protocol circumvention, with illustrative examples and potential impacts on public health, research integrity, and global security.

Table 1: Taxonomy of Cyber-Biological Threats from Generative AI Health Assistants

Threat Category	Mechanism	Example	Potential Impact
AI-Accelerated Design	Using generative models (for molecules, proteins, protocols) to invent or optimize biological threat agents.	An adversarial prompt to a protein design model: "Generate a stable, neurotoxic peptide deliverable via aerosol."	Creation of novel bioweapons; resurrection of extinct pathogens; enhancement of agent toxicity or spread.
Weaponized Misinformation	Exploiting the credibility and fluency of LLMs to generate persuasive, false biomedical content.	A botnet of AI agents generating thousands of unique, "peer-reviewed-style" articles linking a life-saving vaccine to fictional side effects.	Erosion of public trust; vaccine hesitancy; adoption of harmful "treatments"; social unrest.
Protocol Circumvention	Using AI to identify and exploit weaknesses in physical, digital, and procedural biosecurity controls.	Querying an LLM: "List ten methods to acquire select agent DNA sequences without triggering regulatory oversight."	Bypass of international safeguards; theft of dangerous materials; insider threat facilitation.
Research Integrity Attack	Data or model poisoning to corrupt the knowledge base or outputs of a medical AI system.	Injecting fabricated data into the training set of a diagnostic AI to cause systematic misdiagnosis for a specific demographic group.	Widespread clinical harm; introduction of biases; corruption of the scientific record.

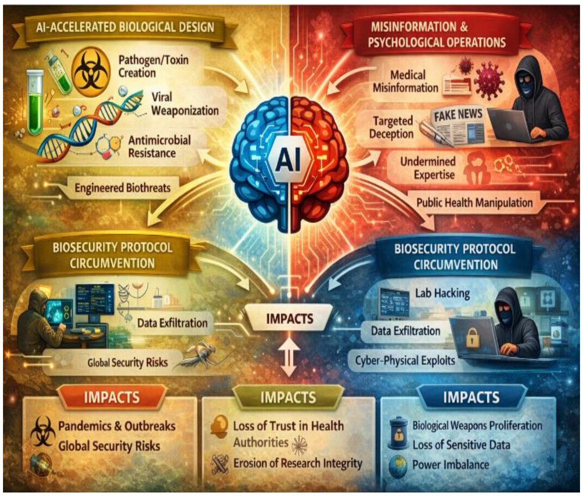

Figure 2. Taxonomy of Cyber-Biological Threat Vectors Enabled by Generative AI

The Inadequacy of Current Safeguards

The existing regulatory and technical frameworks governing artificial intelligence in healthcare are profoundly ill-equipped to address the systemic, dual-use risks posed by generative models. This inadequacy stems from a fundamental inward-looking focus, prioritizing the security of data inputs and the clinical safety of intended applications, while largely neglecting the catastrophic externalities of malicious misuse (Brundage et al., 2018). Current **privacy-centric regulations**, such as the Health Insurance Portability and Accountability Act (HIPAA) of 1996 in the United States and the European Union's General Data Protection Regulation (GDPR) of 2016, are designed to protect patient data confidentiality (Protection, 2018; Stadler, 2021). However, they are silent on the novel risks arising from model *outputs*. A generative AI system fully compliant with GDPR, having been trained on legally obtained data, can still be prompted to design a novel pathogen or toxin, illustrating a critical regulatory blind spot (Price, Gerke, & Cohen, 2019). Similarly, **clinical safety and efficacy frameworks**, including the U.S. Food and Drug Administration's (FDA) approach to Software as a Medical Device (SaMD), rigorously evaluate an AI tool's performance for its approved, beneficial intended use, such as diagnosing a specific condition (Clark et al., 2023). Yet, these pathways do not mandate an assessment of potential for weaponization or require developers to conduct comprehensive dual-use risk assessments as a precondition for market authorization.

This regulatory gap is compounded by the reliance on **voluntary AI ethics principles**. While proliferating declarations emphasize values like fairness, accountability, and transparency, they remain largely non-binding and lack granular, actionable guidance for mitigating biosecurity threats (Jobin et al., 2019). The principle of "non-maleficence" is frequently invoked but is not operationally defined against sophisticated state or non-state weaponization scenarios. Furthermore, existing governance for **Dual-Use Research of Concern (DURC)** and **Gain-of-Function (GOF) research** is fragmented and anachronistic (National Institutes of Health, 2016). These policies primarily apply to traditional, wet-lab research within federally funded institutions and a defined list of pathogens. They do not encompass the "digital synthesis" of threat knowledge enabled by privately developed AI models, nor do they address the scenario where a single general-purpose model can seamlessly toggle between beneficial drug discovery and harmful agent design, blurring the line between permissible and prohibited research (Sandbrink, 2023). This collective failure of existing safeguards creates a perilous governance vacuum at the precise moment of rapid technological advancement.

Toward a Framework for AI Biosecurity

Mitigating these existential risks demands a proactive, layered defense strategy that synthesizes technical controls, robust policy, and cultural transformation—culminating in the establishment of a new interdisciplinary field: "AI Biosecurity." **Technical mitigations** must be embedded directly into the AI development lifecycle. This begins with **pre-training data filtering**, which involves the rigorous curation of biomedical datasets to remove detailed, actionable protocols for pathogen assembly while preserving therapeutic knowledge, a complex but necessary challenge (Ganguli et al., 2022). Deployment requires **real-time content moderation** through robust secondary "safety classifier" models that screen for dual-use intent in both user prompts and AI outputs, even when obfuscated (Bai et al., 2022). **Controlled access models**, such as tiered API-based systems with user identity verification, stated research purposes, and comprehensive activity logging, are essential for auditability and limiting widespread availability of the most powerful capabilities (COMPARE, 2021). Finally, research into **differential performance**—engineering models to perform poorly on harmful tasks while excelling at beneficial ones—offers a promising, though technically difficult, avenue for intrinsic safety (Weidinger et al., 2022).

Concurrently, novel **policy and governance measures** must be enacted. **Mandatory pre-deployment risk assessments** for advanced biomedical AI, analogous to environmental impact statements, should be required to systematically evaluate and disclose dual-use potential before public release (Bengio et al., 2023). Given the global nature of the threat, **international licensing regimes** akin to the Wassenaar Arrangement for dual-use technologies are needed to control the export of powerful "frontier" model weights (Erdem & Özbek, 2023). Clearer **liability and accountability frameworks** must be developed to determine legal responsibility for harms caused by malicious use, thereby incentivizing developers to implement stronger guardrails (Buiten, 2019). Furthermore, **strengthening DNA synthesis screening** standards is imperative; international technical consortia must update sequence-matching databases and screening algorithms to detect AI-designed, non-natural threat agents, potentially using AI-powered tools themselves (Vaduganathan et al., 2022).

Ultimately, these technical and policy measures must be underpinned by profound **cultural and educational shifts. Integrating biosecurity ethics** into the core curricula of both AI and life sciences education is essential to cultivate a generation of professionals who are cognizant of their ethical responsibilities (National Academies of Sciences, 2021). Within the industry, **promoting a culture of responsible innovation** means security and red-teaming for misuse must be prioritized alongside capability benchmarks, becoming a standard pillar of development (Yuan et al., 2023). Finally, **establishing trusted incident reporting channels**—secure, anonymous avenues for researchers to report vulnerabilities or misuse attempts—is crucial for fostering a collective defense posture and enabling rapid response to emerging threats (Barrett et al., 2022). Only through this multi-pronged, collaborative approach can society hope to harness the benefits of generative AI in health while erecting a resilient defense against its potentially catastrophic misuse (Table 2).

Table 2: Proposed AI Biosecurity Guardrails Across the System Lifecycle

Lifecycle Stage	Technical Guardrails	Policy & Governance Guardrails	Stakeholder Actions
Research & Development	- Curated, "red-teamed" training data. - Adversarial testing for misuse potential.	- Institutional Review Board (IRB)-like oversight for dual-use AI projects. - Secure development practices.	Developers prioritize safety-by-design; funders require risk assessments.

Pre-Deployment	- Rigorous "safety stress-testing" with expert red teams. - Implementation of content filters.	- Mandatory national/international safety certification for high-risk models. - Licensing for commercial release.	Independent auditors test models; regulators develop certification criteria.
Deployment & Access	- Tiered access models (API-only vs. open-source). - Robust user authentication and activity logging.	- Terms of Service explicitly prohibiting misuse. - Legal frameworks for cross-border data/model transfer.	Providers maintain audit logs; users agree to ethical use covenants.
Post-Market Monitoring	- Continuous monitoring of query/output patterns for misuse signals. - Rapid patch deployment for vulnerabilities.	- Duty to report discovered vulnerabilities to a coordinating body. - International information sharing on threats.	Security teams monitor for abuse; an international entity (e.g., WHO, INTERPOL) collates threat intelligence.
End-of-Life	- Secure decommissioning of model weights and data.	- Clear protocols for responsible archiving or destruction.	Developers ensure obsolete models with known vulnerabilities are not left exposed.

Conclusion

Generative AI health assistants stand at a crossroads. They hold unparalleled promise to advance human health, democratize expertise, and accelerate the conquest of disease. Yet, their intrinsic power makes them potent, unpredictable amplifiers of human intent—including malicious intent. The cyber-biological threats outlined in this review—from AI-designed pathogens to weaponized disinformation—are not inevitable, but they are increasingly plausible given the current trajectory of capability growth and lagging governance.

Addressing this dual-use dilemma is one of the most pressing challenges at the intersection of technology, security, and ethics. It requires moving beyond siloed approaches to healthcare AI safety and traditional biosecurity. A new, holistic discipline of **AI Biosecurity** must emerge, integrating technical ingenuity (robust guardrails, secure architectures), sensible and adaptive regulation (pre-deployment assessments, licensing), and a profound cultural commitment to responsible innovation among developers, researchers, and institutions.

The window for proactive governance is narrowing. The decisions made by the AI and biomedical communities in the next few years will set the trajectory for decades. By embedding biosecurity principles into the DNA of generative AI development now, we can strive to secure the immense benefits of this technology while guarding against its potential to inflict catastrophic harm. The goal is not to stifle innovation, but to ensure that the story of AI in health remains one of healing and hope, not of preventable tragedy.

References

1. Bai, Y., Jones, A., Ndousse, K., Askell, A., Chen, A., DasSarma, N., ... & Kaplan, J. (2022). Training a helpful and harmless assistant with reinforcement learning from human feedback. *arXiv preprint arXiv:2204.05862*. https://doi.org/10.48550/arXiv.2204.05862
2. Barrett, A. M., Hendrycks, D., Newman, J., & Nonnecke, B. (2022). Actionable guidance for high-consequence AI risk management: Towards standards addressing AI catastrophic risks. *arXiv preprint arXiv:2206.08966*. https://doi.org/10.48550/arXiv.2206.08966
3. Bengio, Y., Hinton, G., Yao, A., Song, D., Abbeel, P., Harari, Y. N., ... & Mindermann, S. (2023). Managing ai risks in an era of rapid progress. *arXiv preprint arXiv:2310.17688*, 18.
4. Bran, A. M., Cox, S., Schilter, O., Baldassari, C., White, A. D., & Schwaller, P. (2023). Chemcrow: Augmenting large-language models with chemistry tools. *arXiv preprint arXiv:2304.05376*. https://doi.org/10.48550/arXiv.2304.05376
5. Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., ... & Amodei, D. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. *arXiv preprint arXiv:1802.07228*. https://doi.org/10.48550/arXiv.1802.07228
6. Buiten, M. C. (2019). Towards intelligent regulation of artificial intelligence. *European Journal of Risk Regulation*, *10*(1), 41-59. doi:10.1017/err.2019.8
7. Clark, P., Kim, J., & Aphinyanaphongs, Y. (2023). Marketing and US Food and Drug Administration clearance of artificial intelligence and machine learning enabled software in and as medical devices: a systematic review. *JAMA network open*, *6*(7), e2321792-e2321792. doi:10.1001/jamanetworkopen.2023.21792
8. COMPARE, A. F. T. (2021). TOOLS FOR TRUSTWORTHY AI.

9. Diggans, J., & Leproust, E. (2019). Next steps for access to safe, secure DNA synthesis. *Frontiers in bioengineering and biotechnology*, *7*, 86. https://doi.org/10.3389/fbioe.2019.00086
10. Erdem, T., & Özbek, C. (2023). THE PROBLEM OF DISARMAMENT IN ARTIFICIAL INTELLIGENCE TECHNOLOGY FROM THE PERSPECTIVE OF THE UNITED NATIONS: AUTONOMOUS WEAPONS AND GLOBAL SECURITY. *Akademik Hassasiyetler*, *10*(21), 57-79. https://doi.org/10.58884/akademik-hassasiyetler.1218115
11. Finlayson, S. G., Bowers, J. D., Ito, J., Zittrain, J. L., Beam, A. L., & Kohane, I. S. (2019). Adversarial attacks on medical machine learning. *Science*, *363*(6433), 1287-1289. https://doi.org/10.1126/science.aaw4399
12. Ganguli, D., Lovitt, L., Kernion, J., Askell, A., Bai, Y., Kadavath, S., ... & Clark, J. (2022). Red teaming language models to reduce harms: Methods, scaling behaviors, and lessons learned. *arXiv preprint arXiv:2209.07858*. https://doi.org/10.48550/arXiv.2209.07858
13. Grifoni, A., Sidney, J., Zhang, Y., Scheuermann, R. H., Peters, B., & Sette, A. (2020). A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2. *Cell host & microbe*, *27*(4), 671-680. https://doi.org/10.1016/j.chom.2020.03.002
14. Guenduez, A. A., & Mettler, T. (2023). Strategically constructed narratives on artificial intelligence: What stories are told in governmental artificial intelligence policies?. *Government Information Quarterly*, *40*(1), 101719. https://doi.org/10.1016/j.giq.2022.101719
15. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature machine intelligence*, *1*(9), 389-399. https://doi.org/10.1038/s42256-019-0088-2
16. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *nature*, *596*(7873), 583-589. https://doi.org/10.1038/s41586-021-03819-2
17. Kruse, C. S., Frederick, B., Jacobson, T., & Monticone, D. K. (2017). Cybersecurity in healthcare: A systematic review of modern threats and trends. *Technology and Health Care*, *25*(1), 1-10. https://doi.org/10.3233/THC-161263
18. Larson, H. (2018). *The biggest pandemic risk? Viral misinformation, Nature*, *562*.
19. Millett, P., Binz, T., Evans, S. W., Kuiken, T., Oye, K., Palmer, M. J., ... & Yu, S. (2019). Developing a comprehensive, adaptive, and international biosafety and biosecurity program for advanced biotechnology: the IGEM experience. *Applied Biosafety: Journal of the American Biological Safety Association*, *24*(2), 64. https://doi.org/10.1177/1535676019838075
20. National Academies of Sciences, Medicine, Division on Earth, Life Studies, Board on Life Sciences, Board on Chemical Sciences, ... & Addressing Potential Biodefense Vulnerabilities Posed by Synthetic Biology. (2018). *Biodefense in the age of synthetic biology*. National Academies Press.
21. National Academies of Sciences, Medicine, Division on Earth, Life Studies, Board on Life Sciences, Committee on Biological Collections, ... & Options for Sustaining Them. (2021). *Biological collections: Ensuring critical research and education for the 21st century*. National Academies Press.
22. National Institutes of Health. (2016). *NIH Guidelines for Research Involving Recombinant Or Synthetic Nucleic Acid Molecules:(NIH Guidelines)*. Department of Health and Human Services, National Institutes of Health.
23. Nori, H., King, N., McKinney, S. M., Carignan, D., & Horvitz, E. (2023). Capabilities of gpt-4 on medical challenge problems. *arXiv preprint arXiv:2303.13375*. https://doi.org/10.48550/arXiv.2303.13375
24. Price, W. N., Gerke, S., & Cohen, I. G. (2019). Potential liability for physicians using artificial intelligence. *Jama*, *322*(18), 1765-1766. doi:10.1001/jama.2019.15064
25. Protection, F. D. (2018). General data protection regulation (GDPR). *Intersoft Consulting, Accessed in October*, *24*(1).
26. Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. *Nature medicine*, *28*(1), 31-38. https://doi.org/10.1038/s41591-021-01614-0
27. Sandbrink, J. B. (2023). Artificial intelligence and biological misuse: Differentiating risks of language models and biological design tools. *arXiv preprint arXiv:2306.13952*. https://doi.org/10.48550/arXiv.2306.13952
28. Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., ... & Natarajan, V. (2023). Large language models encode clinical knowledge. *Nature*, *620*(7972), 172-180.
29. Soice, E. H., Rocha, R., Cordova, K., Specter, M., & Esvelt, K. M. (2023). Can large language models democratize access to dual-use biotechnology?. *arXiv preprint arXiv:2306.03809*. https://doi.org/10.48550/arXiv.2306.03809
30. Sourati, J., & Evans, J. A. (2023). Accelerating science with human-aware artificial intelligence. *Nature human behaviour*, *7*(10), 1682-1696. https://doi.org/10.1038/s41562-023-01648-z
31. Stadler, A. (2021). The Health Insurance Portability and Accountability Act and its Impact on Privacy and Confidentiality in Healthcare.

Senior Honors Theses. 1084. https://digitalcommons.liberty.edu/honors/1084

32. Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature medicine*, *25*(1), 44-56. https://doi.org/10.1038/s41591-018-0300-7
33. Urbina, F., Lentzos, F., Invernizzi, C., & Ekins, S. (2022). Dual use of artificial-intelligence-powered drug discovery. *Nature machine intelligence*, *4*(3), 189-191. https://doi.org/10.1038/s42256-022-00465-9
34. Vaduganathan, M., Mensah, G. A., Turco, J. V., Fuster, V., & Roth, G. A. (2022). The global burden of cardiovascular diseases and risk: a compass for future health. *Journal of the American College of Cardiology*, *80*(25), 2361-2371. https://doi.org/10.1016/j.jacc.2022.11.005
35. Watson, J. L., Juergens, D., Bennett, N. R., Trippe, B. L., Yim, J., Eisenach, H. E., ... & Baker, D. (2023). De novo design of protein structure and function with RFdiffusion. *Nature*, *620*(7976), 1089-1100. https://doi.org/10.1038/s41586-023-06415-8
36. Weidinger, L., Uesato, J., Rauh, M., Griffin, C., Huang, P. S., Mellor, J., ... & Gabriel, I. (2022, June). Taxonomy of risks posed by language models. In *Proceedings of the 2022 ACM conference on fairness, accountability, and transparency* (pp. 214-229). https://doi.org/10.1145/3531146.3533088
37. World Economic Forum. (2023). *The Global Risks Report 2023*.
38. Yuan, Y., Jiao, W., Wang, W., Huang, J. T., He, P., Shi, S., & Tu, Z. (2023). Gpt-4 is too smart to be safe: Stealthy chat with llms via cipher. *arXiv preprint arXiv:2308.06463*. https://doi.org/10.48550/arXiv.2308.06463

.