Global Health & Medicine. 2026; 8(3):161-165.

DOI: 10.35772/ghm.2026.01067

Human-in-the-loop reconsidered: Shadow use and reliance management in drug development

Yusuke Inoue*

Department of Healthcare Ethics, Kyoto University School of Public Health, Kyoto, Japan.

Abstract: This article examines the ethical governance of artificial intelligence (AI) use in drug development through joint principles of good AI practice issued by the U.S. Food and Drug Administration (FDA) and the European Medicines Agency (EMA). It argues that the significance of the principles lies in moving beyond AI exceptionalism: AI should neither be uniformly prohibited nor uniformly permitted but assessed in a risk-based manner according to context, purpose, and potential impact across the drug lifecycle. Among the ethical and governance risks associated with AI, this study focuses on two organizational risks that are particularly relevant to implementation. The first is shadow use, in which AI involvement remains insufficiently visible, documented, or reviewed. The second is reliance management. Once AI is integrated into research and regulatory workflows, some degree of reliance is inevitable; however, such reliance must remain conscious, proportionate, reviewable, and supported by meaningful human oversight. Overreliance and deskilling are risks associated with poorly managed reliance. Ethical governance should therefore make AI use visible and reviewable while preserving the practical ability to question, verify, escalate, or set aside AI-assisted outputs.

***Keywords*:** regulatory science, drug development, shadow use, reliance management, human oversight, risk-based governance

1. Introduction

Use of artificial intelligence (AI) in medicine and research has expanded beyond diagnostic support to include document drafting, translation, clinical trial documentation management, data analysis, pharmacovigilance, and manufacturing control. The American Medical Association has framed medical AI not as "artificial intelligence" intended to replace human judgment but as "augmented intelligence" designed to support physicians' judgment and work (*1,2*). Although this framework was developed primarily for clinical medicine, it is also relevant to drug development, where AI increasingly supports document preparation, evidence generation, data interpretation, and regulatory process. This framing indicates a shift in the ethical focus of AI from the abstract question of whether AI should be used to more practical questions of implementation: in what contexts, by whom, to what extent, and under what governance and accountability arrangements AI should be used.

2. The U.S. Food and Drug Administration (FDA)/ European Medicines Agency (EMA) joint principles as an implementation framework

As a basis for examining ethical governance of AI in drug development, this article focuses on the "Guiding Principles of Good AI Practice in Drug Development" issued in 2026 by the FDA and EMA (*3,4*). These principles are based on recognition that AI can generate evidence relevant to regulatory decision-making across the entire drug lifecycle, including nonclinical research, clinical trials, manufacturing, and post-marketing safety surveillance. Accordingly, such evidence may affect assessments of quality, efficacy, and safety, and inappropriate use of AI may have implications for product approval.

Significance of the FDA/EMA joint principles lies not in treating AI as an exceptional technology requiring uniform regulation but in translating AI governance into concepts already familiar in pharmaceutical development and regulation. These concepts include human-centered design, a risk-based approach, adherence to existing standards, a clear definition of the context of use, multidisciplinary expertise, data governance, model design, risk-proportionate performance assessment, lifecycle management, and clear communication (*3*). In this respect, the principles move ethical discussion away from the binary question of whether AI should be

allowed, and toward more practical questions: whether a given AI use is fit for its intended purpose, whether it has been evaluated in proportion to its risk, whether it remains reliable after deployment, and whether assignment of responsibility is clearly defined.

This implementation-oriented framing is particularly important in drug development, which involves complex, multistage, and often transnational processes. AI may be introduced at different points in the lifecycle, used by different actors, and connected to decisions of varying regulatory significance. Therefore, a uniformly permissive or prohibitive rule is poorly suited to practice. By emphasizing risk-based governance, alignment with existing standards, performance monitoring, and life-cycle management, the FDA/EMA principles provide a practical framework for calibrating oversight according to intended use of AI, its potential impact, and its degree of influence on decision-making.

Joint issuance of these principles by the FDA and EMA is also significant. Substantial divergence in regulatory approaches to AI use across regions increases uncertainty for sponsors, research institutions, and developers involved in international drug development. By articulating a shared regulatory direction for the United States and Europe, the principles may provide a basis for future international regulatory convergence and discussion in forums such as the International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use. Indeed, these principles have been perceived as a step toward facilitating international regulatory harmonization and global implementation (*3,5*).

However, the FDA/EMA joint principles should be understood as high-level guiding principles that provide a shared vocabulary for assessing AI use in a risk-based manner rather than as a complete operational manual or legally binding regulatory framework (*3,6*). This article does not attempt to provide a comprehensive taxonomy of ethical issues raised by AI in drug development, such as data quality, bias, privacy, transparency, validation, and security. Instead, this article focuses on two related but distinct implementation challenges: shadow use and reliance management. These challenges test whether human-in-the-loop is understood as a substantive governance requirement rather than merely as a procedural label. Shadow use raises the question of whether organizations can identify where AI is being used, including uses embedded in ordinary tools and uses not fully recognized by users themselves. Reliance management raises a different question: once AI is integrated into drug development workflows and some reliance on AI-assisted outputs becomes unavoidable, can organizations ensure that such reliance remains conscious, proportionate, reviewable, and resilient? Overreliance and deskilling are risks that may arise when reliance becomes unexamined, excessive, or disconnected from meaningful human judgment.

3. Shadow use as a visibility problem

The first question concerns visibility: whether organizations can identify, document, and review how and where AI is used. The logic of the FDA/EMA principles draws attention to the problem of shadow use, although they do not provide an operational definition of the term. If organizations do not know where and how AI is used, they cannot meaningfully assess risk, allocate responsibility, document decisions, or manage AI systems throughout their lifecycle. In this sense, shadow use is not a peripheral compliance issue but a central governance problem.

Shadow use refers to a situation in which analysts, researchers, or other staff use large language models or similar AI tools in routine work, even though the organization has not explicitly positioned, documented, or governed such use, and management lacks sufficient visibility in actual practice (*5,7*). Shadow use may occur in routine tasks such as document drafting, translation, summarization, searching, code generation, data formatting, and issue framing. It may also occur through embedded AI functions in applications, such as autocomplete, translation assistance, search support, or text revision, even when users do not clearly recognize that they have "used AI".

Therefore, shadow use should not be treated solely as a matter of rule violation or individual negligence. Intentional concealment or inappropriate AI use should not be tolerated, particularly when such use involves confidential information, personal data, proprietary data, or regulatory submissions. However, as AI functions become increasingly embedded in routine work environments, some forms of unrecognized or undocumented AI involvement may arise even without deliberate misconduct. Moreover, AI may influence upstream stages of reasoning, such as organizing issues, considering counterarguments, and discussing analytical strategies, even when no AI-generated text remains in the final document. Therefore, a governance system that merely asks individuals to declare whether they have used AI is unlikely to capture the full extent of their involvement with AI.

One implication is that organizations may need to consider how to make AI involvement more visible within routine workflows rather than relying solely on retrospective self-reporting. This does not require all uses of AI to be treated as posing the same level of risk. A more feasible approach may be to distinguish between low-risk routine uses and uses that may require documentation, human review, or restrictions on entering sensitive or submission-relevant information, with the level of oversight proportionate to associated risk.

Therefore, practical governance measures should be designed at the workflow level. Possible measures include an internal AI-use registry covering approved tools and use cases; context-of-use documentation

incorporated into analysis plans, validation reports, or submission-relevant documents; risk-based thresholds that determine when AI use requires disclosure, validation, or independent review; restrictions on entering confidential information, personal data, clinical trial data, proprietary information, or submission-relevant data into external AI tools; audit trails for AI-assisted codes, summaries, translations, and regulatory text; and internal consultation pathways for borderline cases. The purpose is not merely to detect and punish hidden use but to convert otherwise invisible AI involvement into reviewable, accountable, and proportionately governed use.

Risk-based governance also requires distinguishing among contexts of AI use. Low-risk uses may include language editing, formatting, and preliminary summarization, provided that no confidential or submission-relevant data are entered into unapproved external AI systems. Moderate-risk uses may include internal literature reviews, coding assistance, data formatting, and drafting internal working documents. High-risk uses may include statistical programming, data cleaning that affects dataset analysis, safety signal detection, manufacturing controls, and regulatory submission documents. The highest-risk uses are those in which AI outputs directly influence evidence generation, benefit-risk assessment, quality evaluation, or regulatory decision-making. Although these categories should be adapted to each organization's workflows, they illustrate why governance should be proportionate to context of use and potential regulatory impact.

Such an approach may also help avoid framing shadow use only as individual misconduct. If AI use is treated solely as an exceptional or prohibited act, researchers may be less willing to seek advice or disclose borderline cases. Clearer expectations regarding when AI use should be discussed, documented, or reviewed can make AI involvement easier to identify. The aim is not to eliminate every form of routine AI assistance but to make AI use that is relevant to governance more transparent, explainable, and reviewable.

4. Reliance management and human oversight

The second implementation challenge concerns reliance management, which involves human oversight. In AI-assisted drug development, this goal cannot eliminate reliance on AI outputs. Once AI tools are used for literature reviews, data cleaning, statistical programming, safety signal detection, manufacturing control, and regulatory writing, some degree of reliance is expected and operationally necessary. The governance question is therefore not whether professionals rely on AI but whether such reliance is conscious, proportionate to risk, and supported by conditions that enable critical review.

Accordingly, human-in-the-loop should not be treated as an ethical guarantee. Although the FDA/EMA principles emphasize human-centered design, simply designating a human as final decision-maker is insufficient. If researchers, analysts, or reviewers are expected to use AI-assisted outputs in fast-moving workflows but lack the time, information, expertise, or authority to question those outputs, human oversight may become a procedural label rather than a substantive safeguard.

Distinguishing three related failure modes of poorly managed reliance helps clarify these concerns. Overreliance refers to excessive trust in AI outputs for a particular task, such as accepting AI-generated code, summaries, analyses, or regulatory texts without sufficient verification. Deskilling refers to the gradual erosion of professional judgments or domain-specific expertise through repeated dependence on AI. Merely formal or nominal human oversight refers to situations in which a human decision-maker remains formally responsible but lacks the time, expertise, authority, or information required for critical evaluation of AI-assisted outputs. These concepts are analytically distinguishable but not mutually exclusive. Formal oversight may facilitate task-specific overreliance, while repeated overreliance may contribute to deskilling over time.

A review of AI-induced deskilling noted that AI-based decision support may contribute to erosion of professional skills and reduce opportunities to acquire them (*8*). However, direct empirical evidence of AI-induced deskilling in drug development remains limited. Deskilling in this field should therefore be framed as a plausible organizational risk that requires monitoring and governance, rather than as an already established outcome across all AI-assisted workflows. Likewise, the literature on automation bias helps explain how reliance may become excessive in particular tasks (*9,10*), whereas critiques of human oversight caution that merely placing a human reviewer in the workflow may create only a formal safeguard unless practical conditions for meaningful review are present (*11*).

This framing is consistent with the practical distinction between confidence in an AI tool and appropriate reliance on its output in a specific clinical or research context. A tool's having undergone a certain level of evaluation does not by itself determine how much weight should be placed on its output in a particular situation (*12*). Appropriate reliance should vary with the intended use, data quality, uncertainty, reversibility, regulatory significance, and consequences of errors. Therefore, the same AI output may warrant different levels of scrutiny, depending on whether it is used for preliminary exploration, internal drafting, formal evidence generation, or support for regulatory submission.

Similarly, the World Medical Association emphasized that AI should augment human judgment and that systems should be in place to ensure availability of alternative procedures in the event of AI system failure,

Table 1. Two implementation challenges of AI governance in drug development and possible organizational responses

Implementation challenge	Potential ethical concerns	Possible governance responses: directions and examples
Shadow use	AI involvement may remain invisible to organizations or reviewers.	Make AI involvement visible and traceable: approved-tool registry; context-of-use records.
	Responsibility, documentation, and accountability may become unclear.	Set proportionate documentation and review rules: documentation thresholds; audit trails.
	A structural visibility problem may be treated only as individual misconduct.	Enable safe discussion of borderline cases: consultation pathway; guidance on sensitive or submission-relevant data.
Reliance management	AI outputs may be accepted without scrutiny appropriate to the task's risk.	Calibrate reliance to risk: verification triggers; escalation criteria.
	Human-in-the-loop may become a sign-off exercise rather than critical oversight.	Secure meaningful human oversight: reviewer time, expertise, authority, and information.
	Repeated reliance may weaken skills needed to detect errors and question assumptions.	Maintain professional judgment over time: training; task rotation; periodic competency checks.

Note: The measures listed are illustrative organizational responses and are not intended as formal regulatory requirements.

critical evaluation of AI outputs, and incident reporting (*13*). These requirements are important because meaningful human oversight depends on more than mere presence of a human reviewer. Reviewers must have sufficient information, time, expertise, and authority to question or override AI-assisted output. Without these conditions, human-centered design may become a procedural label rather than an operational safeguard.

Related evidence from radiology suggests that AI may not always reduce professional burden and may instead create additional interpretive, post-processing, or psychological demands in certain settings (*14*). Although this evidence cannot be directly generalized to drug development, it raises a relevant possibility: AI-assisted workflows may shift professional work away from producing outputs and toward verifying, documenting, and explaining them. In this sense, reliance management is not only a matter of individual professional training but also a governance concern for maintaining the credibility of AI-assisted drug development.

Table 1 summarizes the main ethical concerns and illustrative governance responses to the two implementation challenges discussed in this study.

5. Conclusions

The ethical challenge of AI use in drug development is not simply whether AI should be used or whether reliance on it can be avoided. Once AI is integrated into research and regulatory workflows, a certain degree of reliance becomes inevitable. The central question is whether AI involvement can be made visible and reviewable and whether reliance on AI-assisted outputs can remain conscious, proportionate, and resilient. The FDA/EMA joint principles provide a practical starting point by framing good AI practices around human-centered design, risk-based governance, data governance, performance assessment, lifecycle management, and clear communication.

These principles have limited impact unless organizations translate them into workflow-specific governance practices. Shadow use illustrates that AI involvement may remain insufficiently visible, including when AI functions are embedded in ordinary tools and not fully recognized by users. Reliance management illustrates a second challenge: the human-in-the-loop concept is meaningful only when reviewers have sufficient time, expertise, authority, and information to question AI-assisted outputs, and when organizations preserve professional judgment needed to decide when AI should be accepted, verified, escalated, or set aside. The aim is not to discourage responsible AI use but to govern inevitable reliance in ways that remain compatible with scientific and regulatory integrity.

Funding: This study was supported by the Japan Agency for Medical Research and Development (JP 26oa0439001, JP223fa627001).

Conflict of Interest: The author has no conflicts of interest to disclose.

References

1. American Medical Association. 2026 physician survey on augmented intelligence. *https://www.ama-assn.org/system/files/physician-ai-sentiment-report.pdf* (accessed May 30, 2026).
2. American Medical Association. Augmented intelligence in medicine. *https://www.ama-assn.org/practice-management/digital-health/augmented-intelligence-medicine* (accessed May 30, 2026).
3. U.S. Food and Drug Administration; European Medicines

Agency. Guiding principles of good AI practice in drug development. *https://www.fda.gov/about-fda/artificial-intelligence-drug-development/guiding-principles-good-ai-practice-drug-development* (accessed May 30, 2026).
4. European Medicines Agency. EMA and FDA set common principles for AI in medicine development. *https://www.ema.europa.eu/en/news/ema-fda-set-common-principles-ai-medicine-development-0* (accessed May 30, 2026).
5. Health Policy Watch. EU and US regulators reach landmark accord on AI principles in drug development. *https://healthpolicy-watch.news/eu-and-us-ai-principles/* (accessed May 30, 2026).
6. Oualikene-Gonin W, Jaulent MC, Thierry JP, Oliveira-Martins S, Belgodère L, Maison P, Ankri J; Scientific Advisory Board of ANSM. Artificial intelligence integration in the drug lifecycle and in regulatory science: Policy implications, challenges and opportunities. Front Pharmacol. 2024; 15:1437167.
7. Klotz S, Kopper A, Westner M, Strahringer S. Causing factors, outcomes, and governance of shadow IT and business-managed IT: A systematic literature review. International Journal of Information Systems and Project Management. 2019; 7:15-43.
8. Natali C, Marconi L, Dias Duran LD, Cabitza F. AI-induced deskilling in medicine: A mixed-method review and research agenda for healthcare and beyond. Artificial Intelligence Review. 2025; 58:356.
9. Goddard K, Roudsari A, Wyatt JC. Automation bias: A systematic review of frequency, effect mediators, and mitigators. J Am Med Inform Assoc. 2012; 19:121-127.
10. Parasuraman R, Manzey DH. Complacency and bias in human use of automation: An attentional integration. Hum Factors. 2010; 52:381-410.
11. Green B. The flaws of policies requiring human oversight of government algorithms. Comput Law Secur Rev. 2022; 45:105681.
12. NHS AI Lab; Health Education England. Understanding healthcare workers' confidence in AI: Report 1 of 2. *https://digital-transformation.hee.nhs.uk/binaries/content/assets/digital-transformation/dart-ed/understandingconfidenceinai-may22.pdf* (accessed May 30, 2026).
13. World Medical Association. WMA statement on artificial and augmented intelligence in medical care. *https://www.wma.net/policies-post/wma-statement-on-artificial-and-augmented-intelligence-in-medical-care/* (accessed May 30, 2026).
14. Liu H, Ding N, Li X, Chen Y, Sun H, Huang Y, Liu C, Ye P, Jin Z, Bao H, Xue H. Artificial intelligence and radiologist burnout. JAMA Netw Open. 2024; 7:e2448714.

Received June 5, 2026; Revised June 17, 2026; Accepted June 18, 2026.

Released online in J-STAGE as advance publication June 20, 2026.

**Address correspondence to*:
Yusuke Inoue, Department of Healthcare Ethics, Kyoto University School of Public Health, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan.
E-mail: inoue.yusuke.6m@kyoto-u.ac.jp