ORIGINAL RESEARCH

Governance for Generative AI in Clinical Development: A Cross-Sectional Survey in Japan

Munenori Takata*[†], Mari Sugimoto[‡], Masaharu Harada[§], Makoto Hiraide[‖],
Daisuke Ichikawa[¶], Yukikazu Hayashi[†,‡], Satoru Fukimbara[†,§], Takashi Kawaguchi[†,‖],
Hideki Suganami[†,¶] and Takuhiro Yamaguchi*[†]

Introduction: The development of generative artificial intelligence (AI) has been driven by advances in AI and machine learning, leading to innovative applications across various fields such as natural language processing and image generation. Particularly in the clinical development industry, generative AI has contributed to streamlining data analysis and research support, thereby fostering progress in personalized medicine. However, the current situation is characterized by a lag in governance and regulatory compliance related to its use.

Aim: This study aims to investigate the current status of generative AI utilization and governance within organizations in Japan's clinical development industry, clarifying the extent of adoption and identifying associated challenges.

Methods: Between May 21 and June 13, 2025, an online survey was conducted that targeted pharmaceutical companies, contract research organizations (CROs), academic research organizations (AROs), and system vendors. A total of 35 items were collected regarding organizational attributes, the status of generative AI use, governance measures, and training activities. Data were aggregated and analyzed using keyword analysis and word cloud visualization to identify salient features.

Results: Respondents and organizations involved included AROs (49 respondents across 36 organizations), pharmaceutical companies (48 respondents across 21 organizations), CROs (33 respondents across 13 organizations), and system vendors (2 respondents across 2 organizations). All respondents were individuals who held positions with the responsibility to make decisions regarding the use of generative AI within their respective organizations or departments in clinical development. Approximately 76% of respondents reported obtaining approval for generative AI use, with tools such as OpenAI's ChatGPT series and Microsoft's Copilot being predominantly used. The status of governance varied between organizations; more than 83.3% of pharmaceutical companies (40 respondents), 60.1% of CROs (20 respondents), and 16.3% of AROs (8 respondents) had some form of governance documentation related to generative AI use. However, the development of operational-level standard operating procedures (SOPs) was insufficient across all organizations: only 8.3% of pharmaceutical companies, 6.1% of CROs, and none of the AROs or system vendors had such documents fully in place. Generative AI was mainly used for document translation, brainstorming, and document creation and maintenance, with expectations for future applications including advanced data analysis and programming tasks. Benefits cited included increased operational efficiency, automation, and creative support, while barriers such as security and privacy concerns and risks of misinformation were also noted. The focus of education and training centered on AI literacy and safe usage practices, which emphasized the need for strengthened security education within organizations.

Conclusion: We investigated and summarized the current status of generative AI utilization and governance for clinical development in Japan. A key finding was a lack of organizational governance documents related to utilization of generative AI and education and training. There is an urgent need to establish such governance frameworks along with more practical education and training methods.

* Clinical Research Data Center, Tohoku University Hospital, Sendai, JP

† Division of Biostatistics, Tohoku University Graduate School of Medicine, Sendai, JP

‡ A2 Healthcare Corporation, Tokyo, JP

§ ONO Pharmaceutical Co., Ltd., Osaka, JP

‖ Department of Clinical Assessment, Tokyo University of Pharmacy and Life Sciences, Tokyo, JP

¶ Kowa Company Ltd., Tokyo, JP

Corresponding author: Munenori Takata, M.D., Ph.D. (munenori.takata.a3@tohoku.ac.jp)

Introduction

The rise of generative artificial intelligence (AI) has been driven by the rapid advancement of AI and machine learning. Its history traces back to the early AI research of the 1950s.[1] During that period, rule-based systems that focused on search and inference were predominant. From the 1980s to the 1990s, progress in neural network research led to the accumulation of extensive expertise and the practical implementation of expert systems that were capable of functioning similarly to human specialists.[2] However, maintaining the integrity and consistency of vast amounts of knowledge via rules proved challenging, causing a temporary decline in AI's momentum. In the 2000s, improved computational power and the availability of large-scale digital data propelled significant advances in machine learning research. This era also marked a breakthrough with the development of deep learning, which enables the automatic extraction of features from data.[3]

Particularly remarkable are the advancements in natural language processing (NLP). In 2018, Google released BERT, followed by OpenAI's GPT series in 2022. These innovations allowed AI models to engage in highly natural dialogues with humans and to generate sophisticated text and images. Such generative AI models, initially designed to perform specific tasks, are now heralding the arrival of more versatile and broadly applicable "strong AI", moving beyond the "weak AI" that focused solely on isolated functions such as text and image creation.

In the clinical development industry, AI has been increasingly incorporated into manufacturing, the conduct of clinical trials, quality assurance, traceability, process management, quality control, and risk-based approaches. Generative AI holds significant potential to strengthen data-driven approaches within clinical development.[4] Specifically, it analyzes vast amounts of health care data to contribute to new drug development and improve the efficiency of clinical trials. For example, NLP technologies enable rapid extraction of valuable information from medical records and academic papers, supporting hypothesis generation in early research phases.[5] It also introduces innovations in clinical trial design. AI-driven simulation techniques allow for more efficient determination of trial conditions and participant selection criteria, which facilitates adaptive design approaches. Furthermore, in patient monitoring, real-time analysis of individual health data enables early detection of anomalies, allowing swift interventions. Based on predictive models utilizing generative AI, personalized treatment plans are increasingly feasible, advancing the realization of precision medicine.[6] Consequently, these approaches can reduce the overall project timeline, lower personnel costs, and focus on critical quality attributes during clinical development, enhancing overall efficiency and quality.[4,7]

Meanwhile, regulatory authorities in the clinical development field must continuously update their guidelines to reflect modern scientific and regulatory needs. At the global level, the World Health Organization (WHO) published "*Ethics and governance of artificial intelligence for health: Guidance on large multimodal models*" in 2024, and has continued updating it through 2025; the WHO calls on national governments to establish legal and governance frameworks to ensure the safe use of generative AI in healthcare.[8] The rapidly expanding adoption of generative AI prompted the United States Food and Drug Administration (FDA) to issue guidance titled *"Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products"* in January 2025.[9] This document offers specific guidelines for AI application in drug development processes. While actively recommending the use of AI to optimize trial design and automate safety monitoring, it emphasizes the importance of transparency in AI decision-making and the need for human oversight. The focus is on collaborative models that involve both human expertise and AI, rather than full automation. Similarly, the European Medicines Agency (EMA) has clarified its approach to managing risks while maximizing AI's capabilities, aligned with the EU AI Act.[10] The agency plans to provide comprehensive guidance covering the entire AI lifecycle, develop frameworks for AI tools, foster collaborative networks, and pursue experimental approaches as strategic priorities from 2025 to 2028.

In contrast, Japan has not yet issued official specific guidelines for AI utilization in pharmaceutical and medical device development, and organizations are navigating this landscape with little normative guidance.

Aim

The aim of this study is to investigate the current use of generative AI for clinical development in Japan, organize the findings to understand the status of generative AI utilization, and identify the challenges and recommendations for establishing best practices to ensure the responsible use of generative AI in the future.

Methods

Method of Conducting the Questionnaire Survey

The developed questionnaire was structured into three sections: organizational attributes, current status of generative AI usage, and training status. For each section, specific items were designed to gather relevant information. A total of 35 questionnaire items were carefully reviewed by the working group (WG) members to ensure that the wording and nuances aligned with the realities of the respective industries.

Regarding organizational attributes, the categories included the type of organization (e.g., academic research organizations (AROs), pharmaceutical companies, contract research organizations (CROs), other vendors), as well as the size of the organization and department. For respondent attributes, items covered operational domains, positions, and the status of generative AI use. Specifically, data collected included: permission to use AI in operations, the existence of governance structures, the level of governance (such as policies, standard operating procedures (SOPs), manuals, and instructions), scope of AI utilization, tools used for AI, operational environment, and the degree of achievement in AI use/operation (rated on an 11-point scale from 0 to 10). Additional questions addressed the perceived benefits of AI, concerns or barriers

to implementation and operation, and the current and future training and education on AI, collected as free-text responses (see Supplemental Appendix 1).

The survey questionnaire was prepared separately in Google Forms and Microsoft Teams Forms to account for potential conflicts with corporate security policies, with identical items created on both platforms.

Survey Participants and Schedule

The survey targeted companies involved in clinical development operations and university medical institutions within Japan. In Japan, organizations such as the Japan Pharmaceutical Manufacturers Association, the Japan CRO Association, National University Hospital Clinical Research Promotion Initiative (NUH-CRPI), and Metropolitan Academic Research Consortium (MARC) actively engage in clinical development activities. These organizations, as well as other clinical development-related groups, were asked to distribute the survey to their member organizations.

Respondents were selected from professionals that were responsible for specific operational duties within their organizations, particularly from those who had the authority to consider and decide on the use of generative AI in their work. If individuals outside of these target groups received the survey request, it was clearly stated—both in written correspondence and within the questionnaire—that they should forward the survey to a responsible respondent within their organization who has the relevant authority.

The survey was open for responses from May 21, 2025, to June 13, 2025.

Data Analysis

Data collected separately through Google Forms and Microsoft Teams Forms were combined and aggregated. No statistical analysis was applied to the collected data. For the creation of a word cloud, the free-text responses were processed in three steps using Microsoft Copilot (generative AI tool; Microsoft Corporation, Redmond, WA, USA). First, the Japanese responses were translated into natural and accurate English on a one-to-one basis. Second, text mining was performed to extract and normalize meaningful keywords (e.g., training, e-learning), and their frequencies along with representative response excerpts were organized in tabular form. Finally, a word cloud was generated in which keyword size reflected frequency, and colors were assigned according to categories: security/compliance, training methods, technical aspects, and others.

Results

Background of Respondents

Respondents included key stakeholders involved in clinical development: 49 respondents from academic institutions (36 organizations), 48 from pharmaceutical companies (21 organizations), 33 from CROs (13 organizations), and 2 from system vendors (2 organizations). The survey captured respondents across the spectrum of decision-makers and practitioners involved in clinical development activities.

The organizational positions of respondents primarily included managers at the decision-making level. Approximately 3% were at the director or executive level, around 30% at the department head level, about 30% at the section or site manager level, and roughly 30% at the team leader level. This indicates that many responses came from managerial staff directly involved in operations.

In terms of professional responsibilities, respondents covered various roles of clinical development, including statistical analysis, clinical data management, monitoring, and study or project management. Notably, many respondents were engaged in clinical data management, demonstrating a high interest in AI applications for data quality control.

Regarding organizational size, over 70% of respondents worked in large organizations with more than 1,000 employees, while only a small fraction belonged to organizations with 100 or fewer employees. Most headquarters or main offices were located within Japan; however, some pharmaceutical companies and CROs operated globally, incorporating international perspectives. Department sizes typically ranged from 10 to 50 staff members, with some departments exceeding 100 staff members, suggesting that generative AI adoption levels may vary depending on organizational scale.

Regarding permissions for AI use within their organizations, the survey found that 75.5% of academic respondents, 95.8% of pharmaceutical companies, 81.8% of CROs, and 100% of system vendors reported having approval to utilize generative AI in their workflows. This indicates that generative AI use in clinical development organizations in Japan is rapidly expanding.

The most frequently used generative AI tools were OpenAI's ChatGPT series, followed by Microsoft Copilot. Google Gemini was also used to some extent. When asked whether the use of such AI tools was officially recommended or mandated within their closed environment, 87.5% of pharmaceutical companies, 69.7% of CROs, and 100% of system vendors responded affirmatively, whereas only 14.3% of academic respondents indicated the same (**Table 1**).

Status of Generative AI Governance and Utilization Achievement

Regarding the documentation necessary for the use and operation of generative AI, a significant portion of industry organizations have established guidelines and formalized governance structures. Specifically, 87.5% of pharmaceutical companies and 60.1% of CROs indicated that they have established guidelines, and 43.8% of pharmaceutical companies reported having formalized governance structures or policies. While the highest-level documents—such as overarching policies and operational manuals—are generally in place, the status of SOPs remains insufficient. In contrast, AROs more frequently lack such documentation, with approximately 90% responding that they have no such documents. This suggests that governance frameworks are comparatively less developed among academic entities. Larger organizations tend to have more comprehensive rules governing generative AI use, reflecting a correlation between organizational size and governance maturity.

Table 1: Background Information of Respondents.

		ALL	AROs	Pharmaceuticals	CROs	System Vendors
Number of Respondents, N (%)		132	49 (37.1)	48 (36.4)	33 (25.0)	2 (1.5)
Number of Organizations/ Companies, N (%)		72	36 (50.0)	21 (29.2)	13 (18.1)	2 (2.7)
Respondent's Position Level, N (%)	Executive Level	4	1 (2.0)	0 (0.0)	2 (6.1)	1 (50.0)
	Department Head Level	36	14 (28.6)	9 (18.8)	13 (39.4)	0 (0.0)
	Section/Branch Manager Level	43	14 (28.6)	18 (37.5)	10 (30.3)	1 (50.0)
	Team Leader Level within Each Section/Branch	36	16 (32.7)	13 (27.1)	7 (21.2)	0 (0.0)
	Team Members within Each Section/Branch	13	4 (8.2)	8 (16.7)	1 (3.0)	0 (0.0)
Business Activities, N (%)	Biostatistics	31	12 (24.5)	17 (35.4)	2 (6.1)	0 (0.0)
	Clinical Data Management	39	22 (44.9)	13 (27.1)	4 (12.1)	0 (0.0)
	Clinical Monitoring	46	18 (36.7)	9 (18.8)	19 (57.6)	0 (0.0)
	Study Management/Project Management	43	20 (40.8)	11 (22.9)	12 (36.4)	0 (0.0)
	Medical Writing	15	3 (6.1)	10 (20.8)	2 (6.1)	0 (0.0)
	Provision of Clinical Research related Systems (EDC, Data Analysis Tools)	21	12 (24.5)	2 (4.2)	5 (15.2)	2 (100.0)
	Digital Infrastructure Management	15	6 (12.2)	6 (12.5)	3 (9.1)	0 (0.0)
	Clinical Research Coordinator	12	12 (24.5)	0 (0.0)	0 (0.0)	0 (0.0)
	Medical Representative	1	1 (2.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Others	18	7 (14.3)	5 (10.4)	6 (18.2)	0 (0.0)
Scale of the Organizations/ Companies, N (%)	>1000	85	35 (71.4)	34 (70.8)	16 (48.5)	0 (0.0)
	300–1000	24	6 (12.2)	13 (27.1)	5 (15.2)	0 (0.0)
	100–300	9	0 (0.0)	1 (2.1)	7 (21.2)	1 (50.0)
	50–100	14	8 (16.3)	0 (0.0)	5 (15.2)	1 (50.0)
	10–50	0	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
	<10	0	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Scale of the Departments/ Sections, N (%)	>100	11	2 (4.1)	3 (6.3)	6 (18.2)	0 (0.0)
	50–100	18	6 (12.2)	2 (4.2)	10 (30.3)	0 (0.0)
	10–50	71	29 (59.2)	31 (64.6)	11 (33.3)	0 (0.0)
	<10	32	12 (24.5)	12 (25.0)	6 (18.2)	2 (100.0)
Headquarters/Head Office	Domestic/Global	117/15	49/0	36/12	30/3	2/0
Usage of Generative AI is Permitted, N(%)		112	37 (75.5)	46 (95.8)	27 (81.8)	2 (100.0)
Types of Generative AI Tools, N (%)	OpenAI's ChatGPT series (OpenAI)	73	31 (63.3)	28 (58.3)	12 (36.4)	2 (100.0)
	Microsoft Copilot	77	21 (42.9)	37 (77.1)	19 (57.6)	0 (0.0)
	Google Gemini	22	14 (28.6)	5 (10.4)	1 (3.0)	2 (100.0)
	DeepSeek	3	1 (2.0)	2 (4.2)	0 (0.0)	0 (0.0)
	Anthropic Claude	3	3 (6.1)	0 (0.0)	0 (0.0)	0 (0.0)
	GitHub/Microsoft GitHub Copilot	3	1 (2.0)	1 (2.1)	1 (3.0)	0 (0.0)
	Amazon CodeWhisperer	1	1 (2.0)	0 (0.0)	0 (0.0)	0 (0.0)
	Other	18	5 (10.2)	4 (8.3)	9 (27.3)	0 (0.0)
Within-Organization/Company Closed Environment, N (%)		74	7 (14.3)	42 (87.5)	23 (69.7)	2 (100.0)

EDC = Electronic Data Capture.

The median level of achievement in deploying generative AI across organizations was rated at 3 (on a scale of 0–10) for AROs, pharmaceutical companies, and CROs. The interquartile ranges (IQR) were as follows: AROs (1–4.75), pharmaceutical companies (2–5), and CROs (2–6), indicating diversity in implementation levels. Excluding system vendors, the median score was also 3, suggesting that many organizations have not yet fully exploited the potential of generative AI. This data implies that, despite increasing governance measures, many organizations are still in the early stages of effective AI utilization (see **Table 2**). Among organizations headquartered abroad, 14 global pharmaceutical companies and CROs reported permitting the use of generative AI. None of these companies had an SOP in place, but most had established governance frameworks and user manuals. Satisfaction with generative AI use tended to be higher in these companies compared with domestic firms (see Supplemental Table 1).

Utilization of Generative AI
Regarding the current methods of using generative AI, the most common responses were "translation of documents" (71.2%), "brainstorming" (67.4%), and "document creation and maintenance" (62.1%). These results indicate that generative AI is primarily applied to tasks related to document and text processing. In addition, some respondents reported using generative AI for more specialized tasks, including "programming activities (e.g., development of electronic data capture (EDC) systems, statistical analysis)", "data aggregation and analysis", "business management/operations management", "interpersonal communication (e.g., chatbots)", and "data processing".

Looking ahead, the most expected future use is for "document creation and maintenance" with over 80% indicating interest. This is followed by "translation of documents", "data aggregation and analysis", and "brainstorming".

From these findings, it can be inferred that, for the foreseeable future, generative AI will mainly be utilized for text and document-related tasks. However, there is a strong intention to expand its application to more specialized activities such as "programming activities", and "data aggregation and analysis", especially as AI technology advances, which could lead to increased adoption in more complex and higher-level tasks (**Figure 1**).

Benefits and Concerns/Barriers to the Adoption and Operation of Generative AI Tools
Regarding the benefits that organizations and companies perceive from adopting and operating generative AI tools, the most frequently reported were "improving work efficiency" (99.2%), "business process automation" (87.9%), and "supporting creative work" (78.8%). There is recognition of the advantages in supporting creative activities. Conversely, a substantial proportion of respondents selected with "neutral" regarding benefits such as "quality improvement (reducing errors and mistakes)", "cost reduction" and "support for learning and education", indicating a division in perceptions—some respondents see clear benefits, while others do not (**Figure 2**).

Regarding concerns and barriers about implementing generative AI, approximately 97.0% of respondents expressed "very high concern/major barrier" for "data security and privacy protection" and about 92.6% expressed similar concerns about "misinformation or lack of reliability (hallucinations)". Additionally, around 93% indicated "lack of knowledge and skills related to AI technology" as a significant concern or barrier.

These results suggest that issues such as security risks, misinformation, and insufficient knowledge pose substantial obstacles to the adoption of generative AI tools (**Figure 3**).

Current Status and Future Needs of Education and Training for Generative AI Implementation
For organizations not yet conducting education and training on the use of generative AI, word cloud analysis revealed key focus areas for future education and training, with prominent words such as "literacy", "education", "usage", "security", "knowledge" and "training". This indicates that these are the prioritized areas of development.

Table 2: Status of Generative AI Governance and Utilization Achievement.

|	ALL	AROs	Pharmaceuticals	CROs	System Vendors
Number of Respondents	132	49	48	33	2
Number of organizations that possess governance frameworks concerning the use of generative AI, N (%)	69	8 (16.3)	40 (83.3)	20 (60.1)	1 (50.0)
Status of Documentation on Generative AI Governance, N (%)					
Written Governance Framework/Policies	28	2 (4.1)	21(43.8)	5(15.2)	0 (0.0)
Standard Operating Procedure (SOP) for the Usage of Generative AI	6	0 (0.0)	4 (8.3)	2(6.1)	0 (0.0)
Guidelines for the Usage of Generative AI	58	4 (8.2)	36 (75.0)	18 (54.5)	0 (0.0)
User Manual for Using Generative AI Tools	28	1 (2.0)	17 (35.4)	9 (27.3)	1 (50.0)
Others	2	2 (4.1)	0 (0.0)	0 (0.0)	0 (0.0)
Satisfaction Level of Generative AI Usage Median (IQR)*	3 (2–5)	3 (1–4.75)	3 (2–5)	3 (2–6)	6,8
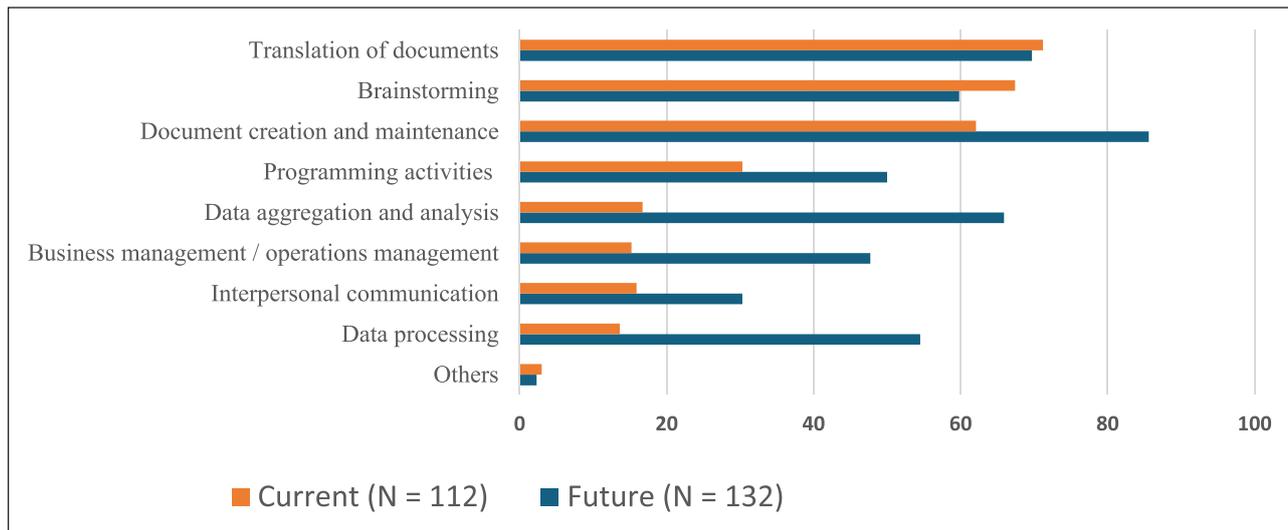
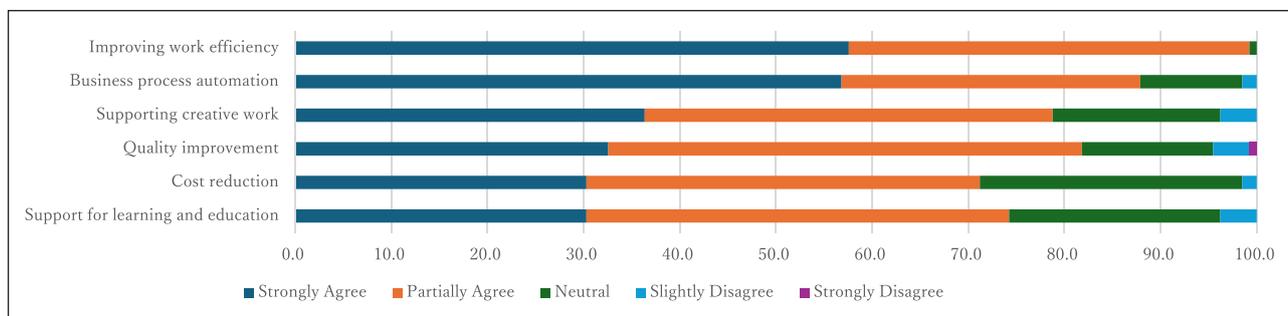**Figure 1:** Generative AI Usage Status, (%).

Figure 2: Benefits for Organizations/Companies from Implementing and Operating Generative AI Tools, N = 132, (%).
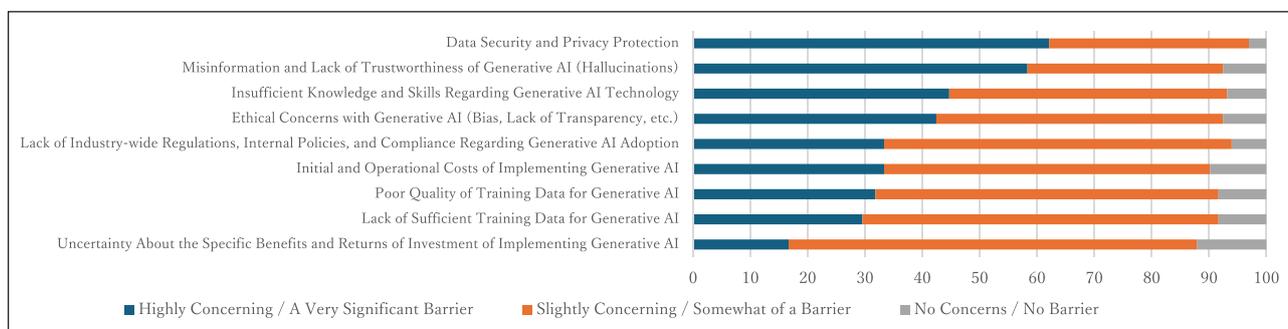

Figure 3: Concerns and Barriers in Introducing Generative AI Tools, N = 132 (%).

Figure 4: Education and Training for Generative AI (red: security/compliance, green: training methods, blue: technical aspects, gray: others). **A.** Currently, responses from organizations and companies that do NOT provide education or training: Thoughts on future education and training N = 50. **B.** Currently, responses from organizations and companies that do provide education or training: Thoughts on current education and training N = 57. **C.** Currently, responses from organizations and companies that do provide education or training: Thoughts on Future education and training N = 50.

The large size of "literacy" indicates that foundational education—aimed at understanding basic AI concepts and knowledge—is frequently emphasized. Similarly, the prominence of "usage" reflects the importance placed on practical, hands-on training, while "security" and "ethics" suggest that awareness of safe and ethical AI

practices is considered essential (**Figure 4A**). Conversely, organizations already providing education and training tend to emphasize words like "training" and "e-Learning" suggesting that current educational activities are primarily focused on foundational knowledge acquisition through online platforms rather than practical workshops (**Figure 4B**). Looking ahead, organizations anticipate future educational needs highlighted by terms such as "knowledge sharing", "training", "use case", "literacy" and "security". Aside from the importance of basic knowledge literacy and security—consistent with organizations not yet providing education and training—there is a clear emphasis on sharing specific use cases and establishing foundational content as critical future directions (**Figure 4C**).

Overall, the word clouds in **Figures 4A–C** consistently show the word "security" as prominently displayed, indicating that security measures are one of the most crucial elements in AI education and training. Given the risks associated with data leakage and misuse in the use of generative AI, acquiring knowledge and skills related to security is indispensable for organizations implementing these technologies.

Discussions

Although this survey was conducted over a short period of time, it successfully gathered responses from many key individuals involved in the adoption and operation of generative AI within AROs, CROs, and pharmaceutical companies engaged in clinical development in Japan. Overall, the responses predominantly came from professionals with practical experience in clinical development and organizational influence, making this data a valuable source for understanding the status and challenges of AI implementation. However, responses from system vendors were limited, and thus the survey does not comprehensively cover all relevant organizational contexts within Japan. Additional targeted investigation of these organizations may be necessary in the future.

Regarding the generative AI tools used, it is likely that some organizations employ multiple tools concurrently. Notably, in pharmaceutical companies, the utilization rate of Copilot reached 77.1%, surpassing that of OpenAI's ChatGPT series. This high usage is probably due to Copilot's strong integration with Microsoft Office products, which facilitates its adoption within corporate environments.

One of the leading mega-tech companies in AI utilization, Amazon, emphasizes in its internal guidance that standardizing output verification processes and maintaining a "closed environment" for AI use are essential, highlighting the importance of utilizing AI in secure, controlled settings.[11] Compared to pharmaceutical companies, CROs, and system vendors, the proportion of academic respondents who reported that their generative AI tools are recommended or mandated for use in closed environments was significantly lower. This discrepancy suggests that, although the importance of governance is recognized within AROs, the complex organizational structures pose challenges to the formulation of governance at the departmental level, which may consequently be associated with a reduced capacity to implement effective security measures. Additionally, it

suggests that policies related to risks in the industry—such as service quality and manufacturing—differ considerably between AROs and industrial sectors (**Table 2**). In line with the subgroup analyses presented in **Table 2**, organizations headquartered abroad (n = 14) were more likely to have established governance frameworks and user manuals for generative AI, and reported higher satisfaction with AI use, despite the continued absence of SOPs (Supplemental Table 1). This pattern supports the notion that headquarters location and organizational size may influence governance maturity and operational readiness, which could partly explain the lower prevalence of closed-environment practices among AROs.

Contrary to our initial expectations, it was notably revealed that many organizations, particularly companies, have established documented frameworks related to the use of generative AI. However, most organizations have yet to develop documentation at the level of SOPs that link policies and manuals, indicating that they are still exploring how to implement AI in specific operational tasks. In AROs, on the other hand, there appears to be a lack of governance documents underpinning AI usage altogether. This results in a regulatory environment for AI deployment that is essentially unregulated—a "wild west." This may be due to limited approval for AI use at the university level, as well as restrictions on which departments or personnel are authorized to utilize such technologies (**Table 2**).

While many international regulatory agencies emphasize ongoing monitoring and change management across the entire "AI lifecycle"—covering data quality, model development, training, deployment, and evaluation[7]—most organizations in Japan are still in the early stages of establishing rules primarily focused on initial implementation. Additionally, regarding the development of governance documents related to the use of generative AI, Japan lags behind regions such as the United States and the EU, where regulations on AI use have already been issued.

In Japan, notifications and regulations from authorities such as the Ministry of Health, Labour and Welfare (MHLW) and the Pharmaceuticals and Medical Devices Agency (PMDA) have been issued relatively late. This has resulted in a gap between responses from global companies and organizations headquartered outside Japan and those from domestic Japanese companies and organizations. This gap in approach could pose risks such as data drift, in which AI performance fluctuates as the model's environment evolves, making it difficult to predict and manage future safety and efficacy issues. A failure to effectively address these challenges may lead to unforeseen problems related to safety and effectiveness.[12]

Given this situation, we advocate that urgent action is required to develop a comprehensive governance framework and policy for the use of generative AI. This should include the formulation of quality management system (QMS) documents that effectively connect organizational policies with detailed manuals and instructions. It should also involve the creation of a refined documentation structure that coherently integrates with other internal QMS documents used within organizations engaged in clinical development.

As for the applications of generative AI, the most prominent uses are brainstorming and initial document translation or creation. However, it appears that more complex operations—such as data ingestion and utilization of the processed data, as well as advanced skills like prompt engineering—are still insufficiently developed across organizations. This suggests a need for targeted education and training to improve immediate skills, as well as higher-level literacy and ethical considerations necessary for organizations when employing such tools. This notion is supported by the fact that the median level of utilization achievement value is around 3 (as shown in **Table 2**). As the scope of generative AI usage expands, it is anticipated that the achievement levels in various industries will increase accordingly (**Figure 1**).

Regarding the benefits of implementing and operating generative AI tools within organizations, it can be inferred that the established use cases, such as "document creation and maintenance", "translation of documents", and "brainstorming"—highlighted in **Figure 1**—are directly linked to improvements in "work efficiency" and "support for creative tasks". Conversely, for benefits like "improvement in quality (error/mistake reduction)", "cost savings", and "support for learning and training", higher-level skills—such as prompt mastery and information preprocessing—are likely required. The higher proportion of respondents answering "neither agree nor disagree" concerning these benefits suggests that organizations are still at an early stage of realizing these advantages, given the skills needed for effective application (**Figure 2**).

Regarding the concerns and obstacles associated with the adoption of generative AI tools, the most prominent issue identified was "data security and privacy protection", which ranked highest among respondents. Ethical issues and the development of regulations were also highlighted as important challenges in advancing AI deployment. Moreover, concerns about "misinformation and lack of trustworthiness (hallucinations)"—which ranked quite high—are driven by the fact that AI outputs do not always provide accurate information, raising fears that incorrect data could influence decision-making in clinical development.[13] In the clinical setting, "hallucinations" have been demonstrated to be problematic. Studies show that Large Language Models (LLMs) can repeatedly produce or elaborate on fabricated information, such as fictitious test values, with high probability—ranging from 50% to 82%—particularly during "adversarial hallucination attacks" in which false data is deliberately embedded.[14] To mitigate hallucinations in clinical development, implementing a process where humans verify AI outputs is considered essential. Enhancing this process to confirm accuracy is an urgent priority. Alongside this, developing algorithms that determine the required level of information accuracy and volume for reliable AI training data is equally important for AI development (**Figure 3**).

Word cloud analysis, which visually highlights frequently appearing words, was applied to analyze responses regarding current educational and training practices, as well as future needs. This visualization aids intuitive understanding. The words most often cited by organizations not currently providing education and training include "literacy", "education", "usage", "security", "knowledge", and "training". These observations indicate that AI education in clinical development should employ a multi-layered approach. Fundamental AI literacy should go beyond operational skills to encompass early education on AI ethics. Practical, hands-on training remains essential for the effective application of AI in daily tasks, and it is recommended to incorporate awareness of security and ethical considerations throughout the training program to help mitigate potential risks.[15] This trend aligns with international perspectives as well. Global surveys reveal that, although many students and educators in the medical field are enthusiastic about using AI tools, formal education related to AI remains scarce. Educators recognize the importance of AI but feel unprepared due to their own lack of knowledge or to ethical concerns.[16,17] Meanwhile, organizations actively engaged in education and training emphasize words such as "knowledge sharing", "training", "use case", "literacy" and "security". Similar to organizations that do not currently provide education and training, these keywords underscore the importance of knowledge, literacy and security. Additionally, sharing use cases and developing practical skills for streamlining organizational processes are highlighted as essential "next level" skills necessary for efficient AI deployment within organizations.

Considering these findings, we advocate that in order to effectively implement educational and training programs within the clinical development setting, it is essential to focus on the following points: 1) enhancing AI literacy, 2) acquiring practical skills related to AI, 3) raising awareness of security issues, 4) addressing ethical concerns, and 5) promoting the sharing of use cases. Developing and providing education and training programs that incorporate these elements is expected to facilitate organization-wide AI utilization and contribute to strengthening overall competitiveness.

Limitations
Although extensive surveys were conducted through relevant clinical development organizations in Japan, responses from system vendors were limited to only two cases. This limited representation—attributable to the low membership of system vendors in the Japan CRO Association, which constituted the primary sampling frame—precluded a thorough examination of trends related to clinical development infrastructure.

Conclusion
We investigated and summarized the current status of generative AI utilization and governance for clinical development in Japan. A key finding was the lack of organizational governance documents related to AI use and training. Deploying AI solutions without proper SOPs and governance can pose significant risks, including data security breaches, reduced reliability, and ethical or legal issues. The absence of standardized procedures may also hinder oversight and increase error and non-compliance risks. Establishing robust governance frameworks and practical training methods is therefore urgently needed to ensure safe and effective AI deployment.

Additional File

The additional file for this article can be found as follows:

Acknowledgements

Competing Interests

The authors declare that the Tohoku University Department of Medical Statistics, and all members of the working group, have no conflicts of interest with any company or organization relevant to this study.

References

1. **Turing, AM.** Computing machinery and intelligence. *Mind.* 1950;59(236): 433–460. DOI: https://doi.org/10.1093/mind/LIX.236.433

2. **Pires PB, Santos JD, Pereira IV.** Artificial neural networks: history and state of the art. *Encyclopedia of Information Science and Technology, Sixth Edition.* 2025; 1–25. DOI: https://doi.org/10.4018/978-1-6684-7366-5.ch037

3. **He R, Cao JC, Tan T.** Generative artificial intelligence: a historical perspective. *National Science Review.* 2025;12(5). DOI: https://doi.org/10.1093/nsr/nwaf050

4. **Layne E, Olivas C, Hershenhouse J,** et al. Large language models for automating clinical trial matching. *Curr Opin Urol.* 2025 Apr 22;35(3):250–258. DOI: https://doi.org/10.1097/MOU.0000000000001281

5. **Pontes CB, Valerio Netto A.** The use of Artificial Intelligence Algorithms in drug development and clinical trials: A scoping review. *Int J Med Inform.* 2025;195:105798. DOI: https://doi.org/10.1016/j.ijmedinf.2025.105798

6. **Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW.** Large language models in medicine. *Nat Med.* 2023;29(8):1930–1940. DOI: https://doi.org/10.1038/s41591-023-02448-8

7. **Zhang B, Bornet A, Yazdani A,** et al. A dataset for evaluating clinical research claims in large language models. *Sci Data.* 2025;12(1):86. DOI: https://doi.org/10.1038/s41597-025-04417-x

8. **World Health Organization (WHO).** Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models. Published March 25, 2025. Accessed December 28, 2025. https://www.who.int/publications/i/item/9789240084759

9. **US Food and Drug Administration (FDA).** Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products. Published January 6, 2025. Accessed December 28, 2025. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-artificial-intelligence-support-regulatory-decision-making-drug-and-biological

10. **European Medicines Agency (EMA).** Artificial intelligence. Accessed December 28, 2025. https://www.ema.europa.eu/en/about-us/how-we-work/data-regulation-big-data-other-sources/artificial-intelligence

11. **Amazon Web Services.** AWS Prescriptive Guidance: Building an enterprise-ready generative AI platform on AWS. Accessed December 28, 2025. https://docs.aws.amazon.com/prescriptive-guidance/latest/strategy-enterprise-ready-gen-ai-platform/best-practices.html

12. **Protschky D, Lämmermann L, Hofmann, P, Urbach N.** What Gets Measured Gets Improved: Monitoring Machine Learning Applications in Their Production Environments. *IEEE Access.* 2025, 13:34518–34538. DOI: https://doi.org/10.1109/ACCESS.2025.3534628

13. **Roustan D, Bastardot F.** The Clinicians' Guide to Large Language Models: A General Perspective With a Focus on Hallucinations. *Interact J Med Res.* 2025;14(1):e59823. DOI: https://doi.org/10.2196/59823

14. **Omar M, Sorin V, Collins JD,** et al. Multi-model assurance analysis showing large language models are highly vulnerable to adversarial hallucination attacks during clinical decision support. *Commun Med (Lond).* 2025;5(1):159. DOI: https://doi.org/10.1038/s43856-025-01021-3

15. **Zhang S, Prasad PG, Schroeder NL.** Learning About AI: A Systematic Review of Reviews on AI Literacy. *Journal of Educational Computing Research.* 2025;63(5):1292–1322. DOI: https://doi.org/10.1177/07356331251342081

16. **Mousavi Baigi SF, Sarbaz M, Ghaddaripouri K, Ghaddaripouri M, Mousavi AS, Kimiafar K.** Attitudes, knowledge, and skills towards artificial intelligence among healthcare students: A systematic review. *Health Sci Rep.* 2023;6(3):e1138. DOI: https://doi.org/10.1002/hsr2.1138

17. **Blanco MA, Nelson SW, Ramesh S,** et al. Integrating artificial intelligence into medical education: a roadmap informed by a survey of faculty and students. *Med Educ Online.* 2025;30(1):2531177. DOI: https://doi.org/10.1080/10872981.2025.2531177