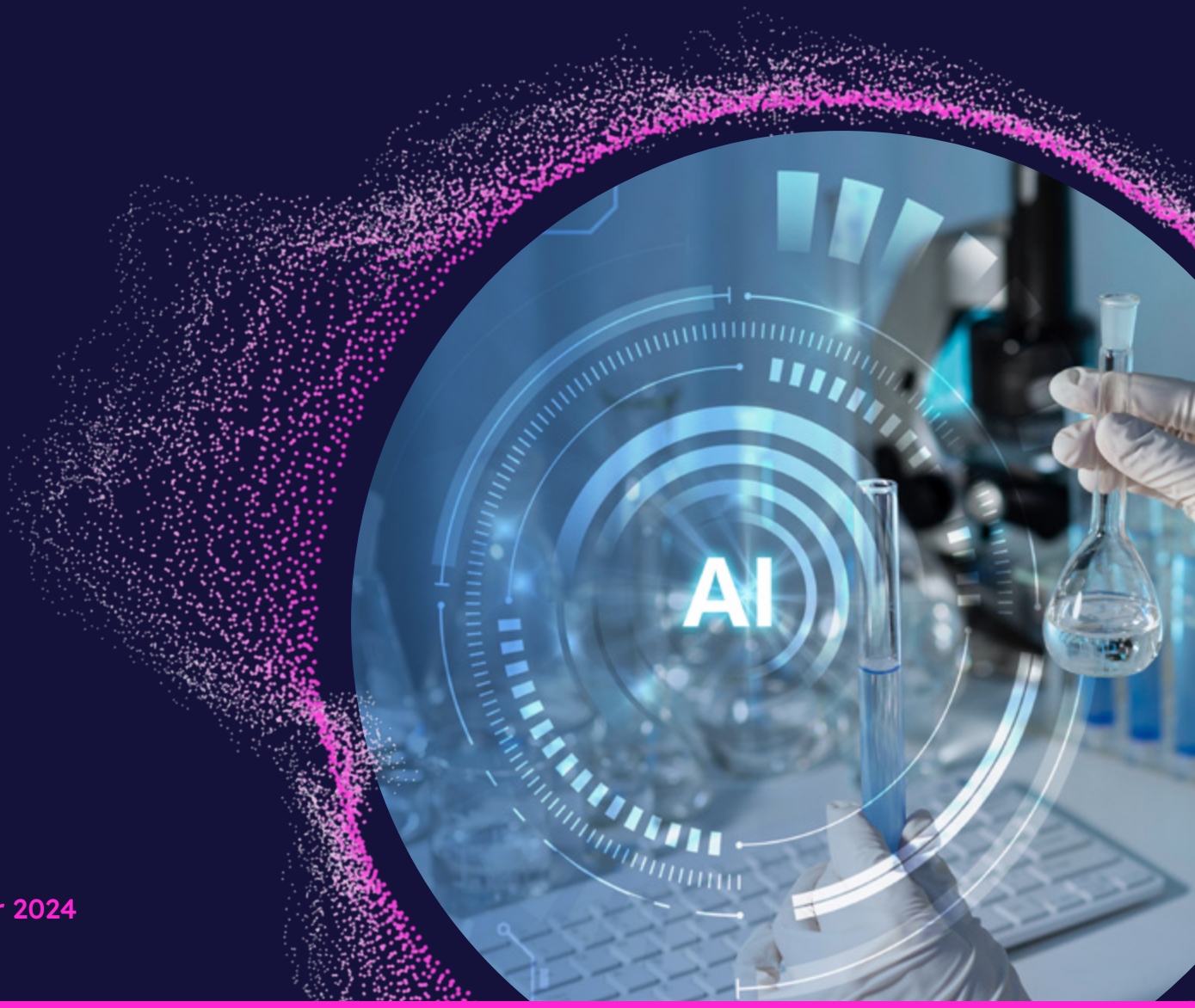
CITELINE
REGULATORY

White Paper

Intelligent Compliance: Transforming Clinical Trial Disclosure with AI



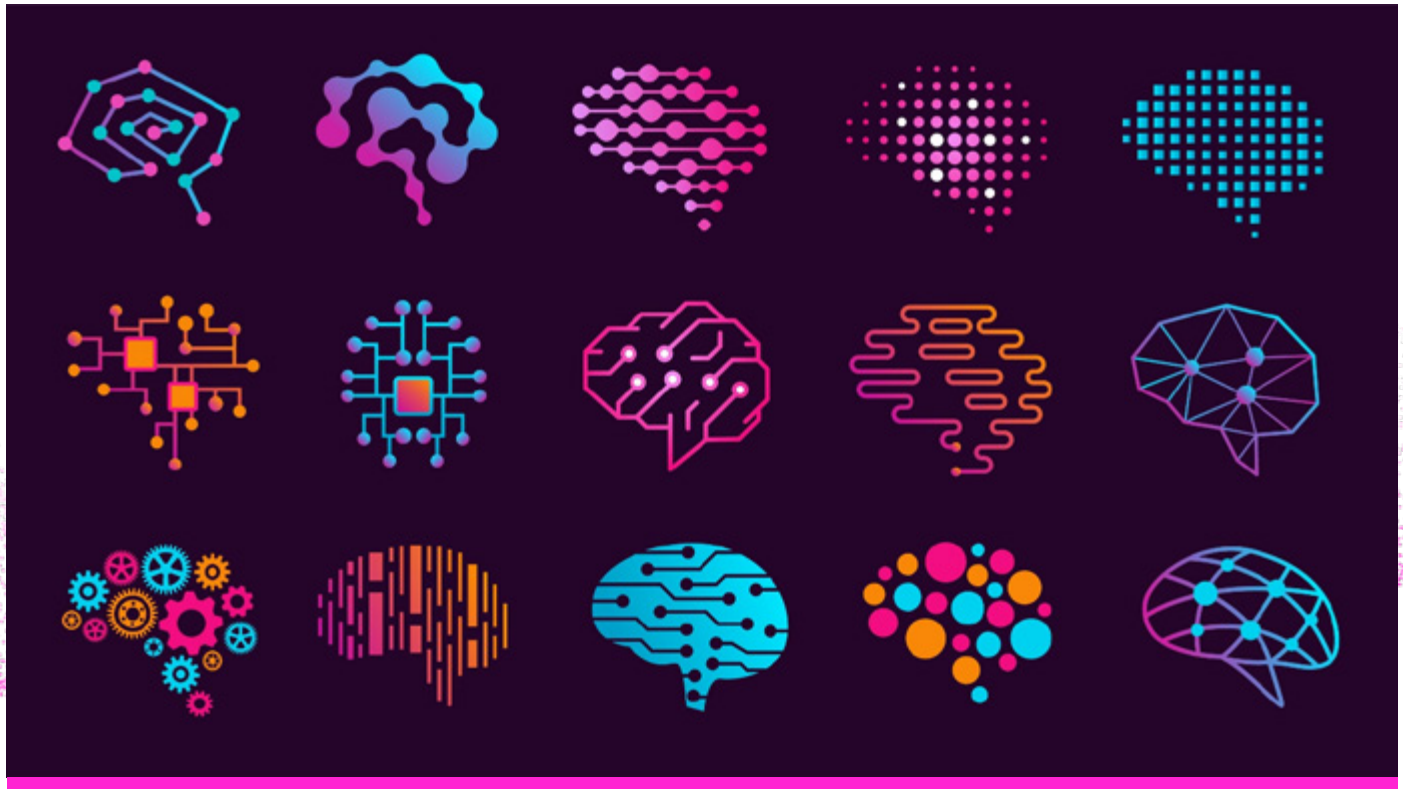
October 2024

Executive summary

The clinical trial disclosure landscape is evolving rapidly, with increasing regulatory complexity, expanding disclosure requirements, and growing demands for transparency. This presents significant challenges for sponsors, who must navigate a complex web of global regulations and ensure trial information is disclosed accurately, consistently, and efficiently.

Artificial intelligence (AI) offers a transformative solution to these challenges. By automating tasks, improving accuracy, and enhancing transparency, AI can help sponsors streamline their clinical trial disclosure processes and meet the growing demands for timely and accessible information. These tools can also help with data extraction and organization, ensuring conformance to registry requirements, generating plain language summaries, and safeguarding sensitive data.

While AI presents significant potential, it's necessary to approach its adoption responsibly. This involves addressing the associated challenges and mitigating risks related to data privacy, bias, and overreliance. Responsible AI adoption necessitates robust data protection measures, careful model training, and ongoing human oversight. By embracing AI strategically and ethically, the clinical research community can harness its power to drive efficiency, compliance, and patient centricity in clinical trial disclosure.



The evolving landscape of clinical trial disclosure

The clinical trial disclosure landscape is increasingly complex, with a patchwork of evolving regulations across different countries and registries. Sponsors and investigators must manage a constant influx of new and updated requirements, making it challenging to stay abreast of the latest guidelines. For example, over 130 regulatory guidance documents and laws related to clinical trial disclosure were published in the past year alone, with approximately 30% already superseded. This rapid pace of change necessitates continuous monitoring and adaptation, placing a significant burden on organizations. Noncompliance with these regulations can have serious consequences, including financial penalties, legal repercussions, loss of investors' trust, and damage to an organization's reputation.

Moreover, the scope of disclosure itself is expanding, encompassing a wider range of clinical documents and data, including full protocols, clinical study reports (CSRs), and anonymized patient data. This broadening scope adds another layer of complexity to an already challenging process. Navigating these requirements demands a proactive and strategic approach to ensure compliance and avoid potential penalties.

The call for greater transparency in clinical trials has intensified in recent years, driven by heightened scrutiny from regulators, patients, investors, and the public. Stakeholders demand access to clinical research data with an emphasis on open science and collaborative research. This shift towards transparency aligns with the broader societal movement towards patient empowerment and informed decision making in healthcare. Patients increasingly seek access to comprehensive and understandable trial information, including plain language summaries of protocols and results. Meeting these evolving expectations requires a patient centric approach to disclosure, where information is clear, concise, and accessible.

AI technologies offer a promising solution to some of these challenges. They have the potential to streamline processes, enhance compliance, and improve the quality and accessibility of clinical trial information. By automating tasks, facilitating data analysis, and enabling intelligent decision making, AI can help organizations navigate the complexities of the disclosure landscape while meeting the growing demands for transparency and patient engagement.

Unlocking disclosure efficiencies with AI: key use cases

Intelligent data ingestion

Clinical trial disclosure involves extracting, organizing, and submitting significant amounts of data from disparate sources, including protocols, statistical analysis plans, clinical study reports (CSRs), and various clinical systems. Traditionally, this task has been labor intensive and time consuming, prone to error and inefficiencies. AI-powered intelligent data ingestion can start to automate these processes, significantly improving efficiency, accuracy, and compliance.

AI algorithms can identify and extract relevant data points from documents, such as protocols and CSRs, transforming them into structured, machine-readable formats. This greatly reduces the need for manual data entry, saving valuable time and resources while minimizing the risk of transcription errors. Additionally, AI can tag and categorize extracted data for efficient organization and retrieval. This enables clinical and regulatory teams to access specific information quickly, streamlining workflows and supporting informed decision making.

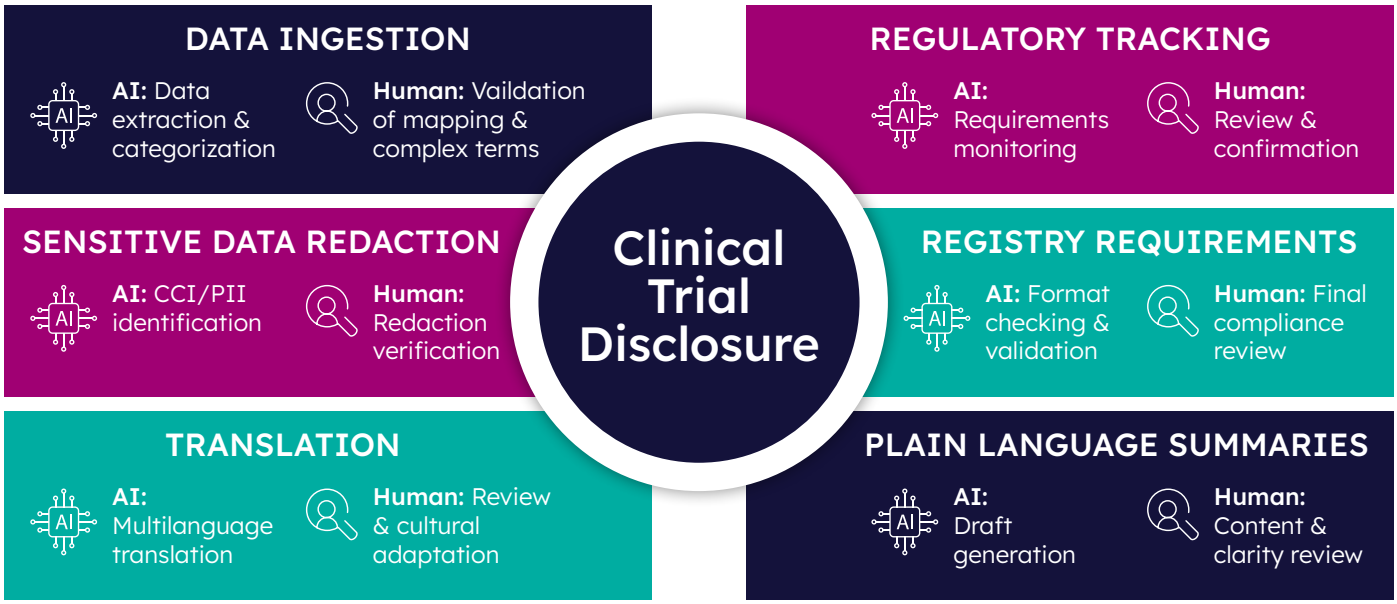
However, while AI promises to streamline data ingestion in disclosure templates or systems, it's important to recognize that current AI models, particularly those operating without human supervision, may still encounter challenges in accurately interpreting and extracting information from complex and nuanced clinical trial documents. For instance, an unsupervised AI might misinterpret ambiguous medical terminology, struggle to differentiate between primary and secondary endpoints or recognize subtle variations in data formats across different sources. These limitations underscore the importance of human oversight and validation of the imported data, ensuring that AI-extracted information is accurate, complete, and fit for purpose in clinical trial disclosure.

Language translation

As clinical trials are increasingly conducted in multiple countries, requiring the disclosure of information in multiple languages, AI-powered translation is becoming more valuable. It enables reasonably accurate translation of complex clinical terms and patient-facing documents. This facilitates global trial registration and results disclosure and ensures crucial information is accessible to diverse patient populations and international stakeholders.

However, while AI translation tools have significantly advanced in recent years, they are not without limitations. Nuances in language, cultural context, and medical terminology can still pose challenges for automated translation systems. For example, automated systems may inadvertently alter the intended reading level by incorporating more complex medical terms when translating plain language results summaries. Additionally, AI can struggle to adapt culturally specific analogies or idioms used in the original version to explain concepts and might not recognize when an alternative example would be more appropriate for the target audience's cultural context. Human review is still required to ensure the accuracy, clarity, and cultural appropriateness of translated materials.

By combining the efficiency of AI with the subject matter expertise and cultural sensitivity of human translators, organizations can achieve high-quality multilingual disclosure that meets both regulatory requirements and the needs of diverse audiences.



Conformance to registry requirements

Clinical trial disclosure involves submitting to over 50 global registries, each with its guidelines, data formats, and submission requirements. Ensuring accurate and consistent disclosure across these diverse platforms can be challenging. Errors can lead to inconsistent postings and the public disclosure of confidential data, making later redaction impossible and potentially risking the rejection of patent applications. Each registry's unique requirements for information content, formatting, structure, and language make maintaining consistent disclosure across global registries more challenging, preventing verbatim content reuse across platforms.

For example, ClinicalTrials.gov requires that all primary and secondary outcomes be listed, each with a single time point and a title of no more than 254 characters. Other registries may require only key secondary outcomes, allow multiple time points per outcome measure, and accept up to 500 characters, not to mention that they may require information in a language other than English.

Preparing the information to meet the requirements of multiple registries while remaining as consistent as possible is a significant writing and process management challenge. Added to this complexity is that some registries subject the submissions to a review process that may trigger requests for information (RFI) or quality control (QC) comments, increasing the disclosure effort and costs, especially when edits to address RFIs and QC comments in one registry require similar updates in others.

Custom AI agents can be trained to check data against each registry's rules, flagging potential errors or inconsistencies before submission. Additionally, AI can help harmonize information across registries while ensuring consistency. This creates a unified representation of trial information across all platforms, which improves efficiency and enhances transparency and trust among stakeholders who rely on consistent and reliable trial data. Ultimately, this reduces the risk of rejection or delays due to noncompliance and streamlines the disclosure process, saving valuable time and resources.

Finally, while AI agents cannot yet fully replace comprehensive regulatory intelligence systems, they can be valuable tools for accessing and understanding complex regulations. For example, AI-powered chatbots can provide a more interactive way to explore regulatory requirements than traditional methods like reading lengthy legal documents. These trained chatbots can answer specific questions, clarify ambiguities, and guide users through relevant regulations sections, improving comprehension and accessibility. However, these chatbots must be built upon a robust and well-structured regulatory intelligence library to ensure accuracy and reliability. Although future advancements may enable AI to manage regulatory intelligence independently, sponsors should take a more conservative approach and rely on established regulatory intelligence systems, perhaps augmented by a trained chatbot.

Plain language summaries made easy

AI will undoubtedly change how companies manage clinical trial disclosure, even if some limitations remain. One area where pharmaceutical companies are currently experimenting is the creation of plain language summaries (PLS).

- **Generating clear and accessible summaries:** AI can assist in generating draft PLS that are clear, concise, and easy for patients to understand, promoting patient engagement and informed decision making.
- **Enhancing the value of registry information:** By translating complex protocols and results into accessible language, AI can increase the value of information posted on clinical trial registries, making it understandable to a broader audience.
- **Improving patient engagement:** PLS can help improve patient engagement by clearly explaining the purpose, risks, and benefits of participating in a clinical trial and communicating the trial results at the end. AI can help create PLS tailored to different patient populations' needs.
- **Facilitating machine analysis:** Although PLS are primarily designed for human readers, they can also be valuable for AI/large language models (LLMs). Clear and simple language can improve machine understanding of clinical trial data, facilitating automated summarization and analysis. With the increasing use of AI to search and interpret publicly available trial information, this clarity can help mitigate misinterpretation by these tools and minimize the spread of distorted information.

Challenges in using AI for plain language summaries

While LLMs offer the potential to automate the summarization of clinical trial results in plain language, they exhibit biases that can hinder the creation of accurate and objective summaries. These biases stem from the nature of their training data and the inherent limitations of current AI technology.

- **Bias towards positive framing:** LLMs, trained on data that often include promotional material, may overemphasize a treatment's benefits in clinical trials while downplaying potential risks or side effects. For instance, an LLM might describe treatment with marginal efficacy as "highly effective" or downplay serious adverse events as "minor side effects."
- **Oversimplification of scientific details:** LLMs may omit crucial details or nuances when summarizing complex clinical trial data, potentially leading to an incomplete or misleading understanding of the results. This can be particularly problematic when conveying statistical significance, effect sizes, or a study's limitations.

- **Stylistic biases that hinder clarity:** LLMs may employ a conversational tone, overuse adjectives, or resort to sensationalism, making the summary less formal and potentially introducing subjectivity. They may also struggle to distill complex information into concise language, resulting in lengthy and convoluted summaries.
- **Inconsistent reading levels and tone:** The output from LLMs can fluctuate between different levels of complexity and formality, making it challenging to ensure consistency for the target audience.
- **Default assumptions and filler content:** LLMs might introduce assumptions or filler content based on patterns in their training data, potentially leading to inaccuracies in the summary.
- **Confirmation bias:** Due to the prevalence of favorable results in published literature and media, LLMs may underrepresent negative or inconclusive findings from clinical trials.

These inherent biases pose challenges to creating accurate, objective PLS of clinical trial results. However, strategies to address these issues include providing explicit instructions on the summary's tone, style, and reading level. Incorporating these instructions into structured prompts and providing PLS examples with an outline template can result in solid first drafts for review by a plain language specialist and refinement through edits and additional prompts.

Leveraging data standards for AI-driven disclosure

Data standards such as the ICH M11 protocol and CDISC Unified Study Data Model (USDM) provide a useful foundation for practical AI applications in clinical trial disclosure. These standards promote consistency and interoperability by establishing clear guidelines for structuring, formatting, and exchanging clinical trial information. By adhering to these standards, sponsors can ensure that their data is machine-readable and readily interpretable by AI algorithms, facilitating a wide range of automated processes.

One key benefit of data standards is the disambiguation of information. By using controlled vocabularies, standardized terminology, and predefined data structures, these standards ensure data are consistently represented across different systems and platforms. This helps AI applications that rely on accurate and unambiguous data for automated data extraction, language translation, and cross-registry comparisons.

Data standards aid in developing tools to access and analyze data from multiple sources, providing a more comprehensive view of clinical trial information. By including relevant standardized metadata such as study phase, therapeutic area, and study population, sponsors can further enhance the categorization of trial information, enabling AI/LLM models to understand and interpret the data better. This ultimately leads to more accurate and reliable insights, supporting informed decision making and accelerating the clinical trial process.

By adhering to standards, clinical trial disclosures become more interoperable and accessible to AI technologies. This standardization:

- **Facilitates data integration:** Harmonized data can be more easily integrated from multiple sources, enhancing the scope and depth of AI analyses.
- **Supports regulatory compliance:** Consistent reporting aligns with regulatory expectations, ensuring AI models work with data that meet necessary legal and ethical standards.
- **Minimizes misinterpretation:** With the rapidly increasing use of AI to search and analyze publicly available clinical data, disclosures incorporating data standards can help prevent these systems from making incorrect assumptions. For example, searches for trials on ClinicalTrials.gov yield wildly different results depending on the wording of the indication, so using Medical Subject Headings (MeSH) terms to describe the indication or including them in the keywords is beneficial. A search for “carcinoma, squamous cell of head and neck” trials shows over 1,000 fewer trials than a search for “squamous cell carcinoma of the head and neck” and less than half the number of trials using the plain language term “head and neck cancer.”

  NOTE: Since ClinicalTrials.gov will attempt to add MeSH terms under “additional relevant MeSH terms” to each study, sponsors can improve the quality of this process by adding the applicable keywords in their submissions.

Addressing challenges and mitigating risks

Safeguarding sensitive data

- Protecting confidential commercial information (CCI) and personally identifiable information (PII) is necessary for clinical trial disclosure. Tools incorporating AI can automate the redaction of CCI/PII, ensuring privacy and compliance with data protection regulations. This is vital to prevent accidental disclosure of sensitive information that could harm patients.
- These tools can also help protect trade secrets by confirming whether CCI is already in the public domain. This information can be used to justify its redaction, ensuring that companies are not unnecessarily disclosing confidential information.

Bias in AI algorithms

AI models can perpetuate existing biases if not carefully designed and trained, leading to inaccurate or unfair outcomes. Organizations must proactively address this risk by using diverse and representative datasets, implementing bias detection and mitigation techniques, and ensuring human oversight in critical decision-making processes.

Overreliance on AI and potential for deskilling

While AI can automate many tasks, organizations should avoid overreliance on AI and maintain human expertise in areas requiring nuanced judgment and interpretation. Excessive dependence on AI tools can lead to declining expertise and professional development within clinical trial disclosure teams. As these tools improve, it will become increasingly important to implement a plan for developing new talent and expertise, especially as AI becomes able to perform tasks typically assigned to entry-level staff.

While AI holds immense promise, it's important to maintain realistic expectations. AI's capabilities are often overstated, and practical applications in clinical trial disclosure are still evolving. Organizations should experiment with AI solutions but avoid hype-driven adoption, focusing on tools addressing specific needs that can be tested and implemented with ongoing human review until the system has proven highly reliable and consistent.

Citeline approach to AI

Citeline's AI-powered solutions for clinical trial disclosure are developed in collaboration between industry experts and a dedicated team of data scientists. This approach ensures that our tools are robust and rooted in practical realities and regulatory complexities of clinical trial disclosure. By combining domain expertise with AI capabilities, Citeline creates advanced and reliable systems. These solutions deliver repeatable results without the "black box" ambiguity often associated with AI.

Given the regulatory compliance requirements of clinical trial disclosure, any technology used in this process must be validated. This can be challenging for AI/LLM solutions because the output of these models may vary across repeated cycles, even with the same input. To address this challenge, Citeline takes a risk-based approach to validating its AI/LLM solutions. This involves a methodical risk management process that:

- **Identifies and prioritizes potential risks:** We focus on areas where the AI/LLM's output could have the greatest impact on clinical trial disclosure.
- **Guides design and implementation choices:** Risk assessment informs how we build and deploy our solutions to mitigate identified risks.
- **Documents risk-based decisions and their rationale:** We maintain clear records of our risk assessments and the reasons behind our design and validation choices

Citeline prioritizes areas with the highest potential impact and verifies data integrity and semantic consistency across process cycles. This approach ensures that our tools meet the requirements of the life sciences industry while delivering consistent and reliable results. Citeline verifies data integrity and semantic consistency across process cycles with a "human-in-the-loop" approach to ensure accuracy and repeatability.

For example, Citeline is developing an AI-based protocol importer to revolutionize how sponsors set up trials in TrialScope Disclose and register their protocols. This innovative tool automates extracting key information from protocol documents and pre-populating registration forms, significantly speeding up disclosure workflows and reducing the workload for clinical trial teams. By eliminating manual data entry, the importer also minimizes the risk of errors, ensuring greater accuracy and consistency in submitted information.

The importer leverages preconfigured data mappings to connect the protocol document with the disclosure data schema. However, recognizing that protocols can vary in structure and content, Citeline incorporates an AI agent trained on clinical documents to suggest additional data mappings. A disclosure expert then verifies the suggested mappings before they are incorporated into the system for use with future protocols. This iterative process ensures accuracy and consistency, allowing the importer to generate reliable and reproducible outputs while continuously learning and expanding the capabilities of the AI-powered importer. By incorporating human review and validation, Citeline ensures that AI augments, rather than replaces, human expertise and builds trust in these solutions.

Another example of Citeline's commitment to responsible AI development is our Disclose Intelligence chatbot. This tool uses AI to provide an interactive way to access and understand clinical trial disclosure requirements

in TrialScope Intelligence. Trained on our curated and best in class intelligence library, the chatbot allows users to ask questions in natural language and receive accurate and relevant answers. This simplifies the process of navigating complex regulations and ensures that users have access to the most up-to-date information. By grounding the chatbot in our rigorously researched and validated regulatory intelligence, Citeline ensures that the information provided is reliable and trustworthy.

Navigating the road ahead: trends and future outlook

Advancements in natural language processing (NLP), machine learning, and deep learning constantly expand AI's capabilities, enabling more sophisticated applications such as automated data extraction, language translation, and predictive analytics. As AI becomes more accurate and trustworthy, it will become a growing part of automating tasks and supporting decision making in clinical trial disclosure.

Realizing the full potential of AI solutions requires close collaboration among different organizational functions, including clinical, regulatory, data management, and IT. Through collaboration and knowledge sharing, organizations can ensure that these tools are integrated into existing workflows and that stakeholders are aligned on the goals and objectives of AI adoption.

However, it's necessary to approach AI adoption responsibly, recognizing the potential challenges and mitigating risks related to data privacy, bias, and overreliance. By embracing AI strategically and ethically, the clinical research community can harness its transformative power to drive efficiency, compliance, and patient centricity in clinical trial disclosure.

CITELINE
REGULATORY

Contact our Clinical Disclosure Experts to learn more

Citeline, a Norstella company, powers a full suite of complementary business intelligence offerings to meet the evolving needs of life science professionals to accelerate the connection of treatments to patients and patients to treatments. These patient-focused solutions and services deliver and analyze data used to drive clinical, commercial, and regulatory-related decisions and create real-world opportunities for growth.

Our global teams of analysts, journalists, and consultants keep their fingers on the pulse of the pharmaceutical, biomedical, and medtech industries, covering them with expert insights: key diseases, clinical trials, drug R&D and approvals, market forecasts, and more. For more information on one of the world's most trusted life science partners, visit Citeline.com.