ELSEVIER

Editorial

Context is everything in regulatory application of large language models (LLMs)

Dear Editor,

Since its inception in 2013, the Global Coalition for Regulatory Science Research (GCRSR) with 10 member countries has been at the forefront of integrating cutting-edge technologies into international regulatory practices (https://www.GCRSR.net). The coalition's efforts are well-documented through a series of scholarly publications summarizing its Global Summit on Regulatory Science conferences that offer a unique platform for dialog among stakeholders to advance regulatory science research.[(p1),(p2),(p3),(p4),(p5),(p6),(p7),(p8),(p9)] In recognition of the transformative impact of large language models (LLMs), the GCRSR established an Interagency LLMs Taskforce, reflecting a proactive vision to assimilate these powerful tools into the regulatory process. With machine learning-based applications becoming a staple in the regulatory framework, it is essential to articulate a clear vision with a specific focus on "context-of-use" to apply LLMs in regulatory settings and emphasizing the foundational principles and the intended scope of AI utilization across a global regulatory landscape.

The rapid advancement of LLMs offers both challenges and opportunities to global regulators. On one side, these agencies are tasked with critically assessing sponsors' applications based on LLMs, which has resulted in an expanding sea of "guidances" and "guidelines" for manufacturers developing new technologies. On the other side, these agencies actively evaluate the utilities of LLMs to improve their routine tasks and augment their institutional knowledge base. The Taskforce was formed to focus on the latter in developing proof-of-concept with a set of criteria aiming at the context-of-use of LLMs across GCRSR members.

One crucial recognition is that the bedrock of regulatory agencies is their vast reservoirs of knowledge, predominantly encapsulated in protected documents. These documents, at times, confidential and laden with specialized terminology, defy simplistic algorithmic solutions. Yet, it is apparent that many regulatory functions display a significant degree of commonality across national frameworks, suggesting a shared set of challenges that are ripe for international collaboration. Although historically reliant on deep human expertise, the growing complexity of these tasks necessitates the integration of LLMs, which are designed to be both generalizable across GCRSR member activities and safe enough to instill confidence in their regulatory task-specific applications. It is this synergy of generalized capability and security, context-specific application that the Taskforce aims to cultivate. In pursuit of this goal, the Taskforce envisions a model where LLMs, once trained, can be deployed universally across GCRSR member states—thereby retaining their generalizability without sacrificing the nuanced understanding of individual regulatory contexts. This would not only streamline deployment but also significantly reduce maintenance costs, advancing the GCRSR's mission to promote excellence in regulatory science through technological synergy.

Consequently, the Taskforce is charged with two key objectives: mapping out key areas of application that are common across global regulatory agencies and then delineating the specific context-of-use for employing LLMs in a global regulatory context. This dual mission is geared towards leveraging the exceptional abilities of LLMs in a manner that is both effective and responsible within the intricate arena of global regulation. The context-of-use in our application is to define a set of criteria by which a LLM will be used by the intended agencies. Currently, the Taskforce is developing a set of context-of-use metrics to assess diverse applications of LLMs in regulatory settings. These metrics, named TREAT, form an acronym encapsulating its core context-of-use consideration: Transparency, Reliability, Explainability, Applicability, and Trustworthiness.

Unlike academic models or other AI paradigms discussed in the broader communities, the TREAT principle stands out due to its specificity in context-of-use for regulatory settings. The metrics are sculpted with a distinct objective: to assess the significance of each metric to bridge the knowledge and practice gap

1359-6446/
https://doi.org/10.1016/j.drudis.2024.103916

between AI innovation and regulatory requirements. The TREAT metrics are designed to synchronize LLM initiatives across global agencies, especially those within the GCRSR consortium. This ensures a cohesive approach, minimizing disparities in LLM application across jurisdictions and regulatory bodies. Moreover, TREAT is not just about internal cohesion; it also emphasizes clear and effective communication with external stakeholders. By presenting a unified front, agencies can relay consistent messages regarding LLM usage, ensuring that all involved parties—from developers to end-users—are on the same page.

A cornerstone of the TREAT framework is its principle of "treating" AI akin to humans, where the human behaviour serves as a reference point. This perspective is both symbolic and pragmatic. Symbolically, it represents the aspiration to hold AI to the same ethical and operational standards that we expect of humans in similar roles. Pragmatically, it underscores the importance of understanding AI decisions, ensuring reliability, and fostering a culture of trust around these systems. By advocating for human-like transparency, reliability, and accountability, these metrics ensure that LLMs, as they become integral to regulatory processes, are not just effective but also held to accepted social expectations and norms.

- Transparency: For LLMs to be effective, they must operate transparently. Just as humans require reasons for their reactions, LLMs must provide clear explanations for their outputs. By monitoring the data and algorithms driving LLM performance, users can better assess and trust the technology.
- Reliability: Human cognition can sometimes be marred by biases or insufficiencies in processing information. Similarly, it is unavoidable for some LLMs that could be skewed by biases as well. The reliability of using these types of models cab only be cemented when the foundational data and logic of the LLMs can be thoroughly examined to define context-of-use.
- Explainability: Humans possess an inherent need to comprehend the rationale behind behaviours and decisions. LLMs should cater to this by being not only interpretable in human terms but also by establishing a scientifically supported link between driving parameters and model performance. In a case of this clarity that is difficult to achieve, concerns about potential discrimination and unjust outcomes can be mitigated with the readily accessible sources that drive the conclusion.
- Applicability: Analogous to humans discerning when and where to apply rules, LLMs should possess a well-defined context-of-use. It is crucial to determine the appropriate application domains, best practices, and whether LLMs should complement or replace existing technologies.
- Trustworthiness: Trust lies at the heart of human interactions, and LLMs should reflect this importance. By defining strong ethical boundaries of application and responsibilities, while managing potential risks, we can ensure the responsible and trustworthy use of AI in regulatory settings.

The TREAT metrics are envisioned as a structured boundary to streamline the deployment of LLMs with context-of-use, and it is organized around several sequential stages. First, it begins by articulating the common problem or challenge that a LLM aims to resolve across the participating regulatory agencies, which define context-of-use. Subsequently, it underscores the significance of each metric in terms of detailing the kind of data that needs to be gathered to cultivate and validate the findings of the LLMs. This is followed by the crafting of a robust LLM platform, initially tested for its viability through a carefully designed pilot study. Finally, the last step encompasses the comprehensive development and widespread deployment of the LLMs, with clearly defined context of use across the global regulatory agencies. Here, it is crucial to continually assess efficiency, revisiting and refining the model in response to fresh data and evolving context of use. The overarching aspiration of TREAT is not just to bolster the deployment of LLMs, but to instil a culture of consistency in the adoption of such technologies across the expanse of the global regulatory landscape.

As the capabilities of LLMs expand and mature, they are poised to redefine the paradigms in regulatory practice across global agencies. Their latent potential, especially in intricate fields like safety evaluation and risk assessment within regulatory frameworks, stands out. Yet, realizing this potential isn't without its hurdles. To genuinely harness the transformative power of LLMs, the challenges largely rest on context-of-use. Collaborative endeavours, coupled with thoughtfully constructed metrics such as TREAT, are instrumental in this journey. With these concerted efforts, the horizon seems promising, illuminating a future where LLMs may potentially revolutionize and elevate regulatory processes to unprecedented heights.

On October 30, 2023, President Biden signed "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence". The executive order underscores the necessity of deploying LLMs technologies in the right context for improved security, equality, and transparency. This is in line with the global regulatory mandate to establish and uphold AI standards and protections. While the Interagency LLMs Taskforce is working within the GCRSR community, the challenges we are facing and the solutions we are proposing are generic and common across diverse communities. We call this collaboration since it requires not only technical acumen, but also an intricate interplay between AI professionals, regulatory bodies, ethical theorists, and legal experts for a synergy aimed at harnessing the potential of LLMs and ensuring context-of-use where its integration is both ethically sound and legally compliant.

Disclaimer

This manuscript reflects the views of its authors and does not necessarily reflect those of the U.S. Food and Drug Administration or Swissmedic. Any mention of commercial products is for clarification only and is not intended as approval, endorsement, or recommendation.

Conflict of interest

The authors declare no conflict of interest.

CRediT authorship contribution statement

Weida Tong: Writing – review & editing, Writing – original draft, Supervision, Conceptualization. **Michael Renaudin:** Writing – review & editing, Writing – original draft, Conceptualization.

Data availability

No data was used for the research described in the article.

References

1. Anklam E et al.. Emerging technologies and their impact on regulatory science. *Exp Biol Med (Maywood)*. 2022;247:1–75.
2. Allan J et al.. *Regul Toxicol Pharmacol*. 2021;122 104885.
3. Thakkar S et al.. Regulatory landscape of dietary supplements and herbal medicines from a global perspective. *Regul Toxicol Pharmacol*. 2020;114 104647.
4. Slikker W Jr et al.. Emerging technologies for food and drug safety. *Regul Toxicol Pharmacol*. 2018;98:115–128.
5. Lambert D et al.. Baseline practices for the application of genomic data supporting regulatory food safety. *J AOAC Int*. 2017;100:721–731.
6. Healy MJ et al.. Regulatory bioinformatics for food and drug safety. *Regul Toxicol Pharmacol*. 2016;80:342–347.
7. Tong W et al.. Genomics in the land of regulatory science. *Regul Toxicol Pharmacol*. 2015;72:102–106.
8. Howard PC, Tong W, Weichold F, Healy M, Slikker W. Global Summit on Regulatory Science 2013. *Regul Toxicol Pharmacol*. 2014;70:728–732.
9. Miller MA, Tong W, Fan X, Slikker W Jr.. 2012 Global Summit on Regulatory Science (GSRS-2012)–modernizing toxicology. *Toxicol Sci*. 2013;131:9–12.

Weida Tong *

Michael Renaudin
, on behalf of the GCRSR Interagency LLMs Taskforce [1]

* *Corresponding author at: 3900 NCTR Rd., Jefferson, AR 72079, USA.*

[1] The full list of co-authors: Swissmedic, Switzerland; Joshua Xu, FDA's National Center for Toxicological Research (NCTR), USA; Leihong Wu, FDA's National Center for Toxicological Research (NCTR), USA; Rebecca Kusko, Cellino Biotech, USA; Liam Childs, Paul Ehrlich Institute, Germany; Monica Carvalho-Soares, Anvisa, Brazil; Flavia Moreira Cruz, Anvisa, Brazil; Hadas Lerner Nussbaum, Institute for Standardization and Control of Pharmaceuticals, Medical Technology Directorate (MTIIR), Israel; Anat Boehm-Cagan, Medical Technology Directorate (MTIIR), Israel; Benjamin Er, Singapore Food Agency, Singapore ; Anna Maria Gerdina Pasmooij, Medicines Evaluation Board (MEB), Netherlands; Norimasa Tamehiro, Food safety Commission (FSCJ), JAPAN.