

The future of large language models in toxicological risk assessment: Opportunities and challenges

Ananth R. Kattamreddy¹, Harisrujan Chinnam²

AFFILIATION

¹ University of Eastern Finland, Kuopio, Finland

² Tampere University, Finland

CORRESPONDENCE TO

Ananth R. Kattamreddy. University of Eastern Finland, Yliopistoranta 8, 70210, Kuopio, Finland. E-mail: ananth.kattam@gmail.com

KEYWORDS

artificial intelligence, risk assessment, large language models, regulatory toxicology, RLRS

Received: 16 February 2025, **Revised:** 22 February 2025, **Accepted:** 24 February 2025

Public Health Toxicol 2025;5(1):3

<https://doi.org/10.18332/pht/202241>

Dear Editor,

Toxicological risk assessment (RA) is a systematic, scientific evaluation of the likelihood of harm resulting from exposure to an entity (physical, chemical or biological etc.), considering both hazard and exposure information. Traditionally, risk assessment has been based on single-stressor models, assessing the impact of individual chemicals in isolation. However, with rapid advancements in technology and industrial demand, the field is growing at a faster pace than expected¹. The integration of real-life risk simulation (RLRS), multi-stressor approaches, and dynamic exposure modelling is transforming traditional risk assessment methods, enhancing their ability to reflect complex real-world conditions. Nonetheless, adopting artificial intelligence (AI) is essential for precise and adaptive decision-making in risk assessment (RA)². The growth in data science and artificial intelligence (AI), especially large language models (LLMs), accelerates the process of risk assessment, with AI-powered regulatory assistance becoming a reality soon.

With innovation speeding up in the post-COVID AI era, the pressure to shorten turnaround times for market authorization of new substances, drugs and products has reached a historic high. AI-driven approaches, including machine learning algorithms, predictive modeling, and big data analytics, are being employed already to enhance the precision and efficiency of risk assessment³. The integration of AI in toxicology is inevitable, and the focus should not be on resisting it but on optimizing its incorporation into existing processes. One increasingly accepted AI-driven approach in the toxicology community is Quantitative Structure-Activity Relationship (QSAR) modeling. This technique utilizes methods like regression, random forests, and support vector machines to correlate chemical

descriptors with toxicity endpoints, helping predict potential hazards more efficiently⁴.

Large Language Models (LLMs) can transform toxicological risk assessment (RA) by improving data curation, uncertainty characterization, and evidence integration for regulatory compliance. With their advanced natural language processing (NLP) capabilities, LLMs can efficiently process large and complex datasets. By training them on publicly available repositories such as PubChem, ChEMBL, ACToR, and Tox21/ToxCast, the models can be fine-tuned to better extract and analyze chemical data. In addition, predictive toxicology models like the DeepTox have demonstrated significant augmentation in the process of risk characterization in recent times. Interestingly, the European Food Safety Authority (EFSA) has investigated the potential of LLMs in toxicology within the AI4NAMS project by evaluating their performance in handling data on Bisphenol A (BPA). The study benchmarked the performance of a baseline Generative Pre-trained Transformer (GPT) model against a fine-tuned model, with the latter proving to be more effective in extracting and consolidating relevant toxicological data⁵.

To date, LLMs have achieved even higher performance efficiency beyond mere automation to the extent of successfully completing tasks like extracting, structuring, and summarizing scientific information. The AI4NAMS project led by EFSA explored the application of GPT-based LLMs for risk assessment acceleration by operating on unstructured scientific literature. A fine-tuned Curie (GPT-3) was compared with out-of-the-box models (text-davinci-002 and text-davinci-003) wherein the fine-tuned model was found to be far more competent due to the fine-tuned model's high knowledge of domain along with the structured

response features. The recent LLMs like GPT-4 and GPT-4o have proven to have improved contextual understanding, accuracy, and recall. These models can process larger context windows and are thus even more appropriate for regulatory document processing, risk assessments, and systematic literature reviews. However, one apparent limitation is that even these advanced LLMs could be plagued by performance degradation in processing extremely long documents⁶.

The goal of integrating AI, particularly large language models (LLMs), in toxicology is not to replace or supplement human decision-making, but to focus on their practical utility in tasks such as preparing dossiers, regulatory compliance documents, and analysis reports. LLMs have the potential to significantly streamline these processes, improving efficiency and accuracy². For instance, under regulations like REACH, a company may currently rely on a large team to prepare a dossier. However, with a tailored LLM, the same task could be managed more efficiently with a smaller team, saving both human costs and valuable time. Similarly, regulatory agencies like the European Chemicals Agency (ECHA) could pull in LLMs to streamline the integration, compilation, and presentation of dossiers for tasks such as re-registration and compliance checks⁶.

An industry-specific LLM could streamline the dossier preparation process by automatically extracting and summarizing toxicity data from both structured and unstructured sources, standardizing formatting according to regulatory guidelines, and identifying key studies using weight-of-evidence approaches while flagging gaps for human review. This would significantly reduce the manpower required, allowing a small, specialized team to oversee the AI's output. Similarly, regulatory agencies like ECHA could implement LLMs to automate the back-end review and integration of submitted dossiers, greatly expediting regulatory processes and enhancing overall efficiency⁷. Nonetheless, at all stages, the decision-making process such as determining whether a study should be included in a risk assessment, remains human-led. The LLM role is to collate, compile, extract, and present data in the required format. We cannot deny the fact that toxicology is a highly specialized field where interdisciplinary expert judgment is crucial at every inflection or deflection point. Human decision making remains crucial at every juncture, while AI streamlines the structuring and formatting of information.

A major advancement would be the fine-tuning of LLM models tailored to specific regulatory frameworks – for example, separate models for EFSA (EU), REACH (EU), EPA (US). By developing exclusive AI models for different classes of compounds (e.g. pesticides, pharmaceuticals, industrial chemicals), risk assessment can be customized, context-aware, and aligned with the latest compliance requirements. Such AI-driven automation can improve accuracy, reduce turnaround time, and enhance transparency, allowing human experts to focus on interpretation, critical thinking, and

decision-making rather than tedious administrative tasks.

LLMs trained on toxicological datasets risk leaking sensitive information through privacy attacks like membership inference and data extraction. Centralized training further increases the risk of data breaches and raises issues with GDPR. Federated Learning (FL) mitigates these risks by training models locally without exposing raw data, leveraging encryption and blockchain for secure aggregation. However, data heterogeneity and privacy-preservation trade-offs can impact model accuracy.

Other key challenges include data bias in LLMs, which necessitates diverse datasets and bias-detection tools; lack of standardization, which can be addressed through harmonized ontologies; and the need for multidisciplinary collaboration among toxicologists, data scientists, and software engineers. Additionally, model interpretability and regulatory acceptability remain critical, requiring validation tests and uncertainty quantification.

Combining Federated Learning (FL) with differential privacy, homomorphic encryption, and blockchain-based auditing strengthens data security. Meanwhile, hybrid human-AI workflows enhance compliance and reliability in toxicological risk assessment³.

Investing in the development of fine-tuned LLMs is not a bad business idea at any point in time. A company developing a specialized LLM for dossier preparation and regulatory integration could market it to industries and regulatory agencies at one go. Just as statistical tools like SPSS are used for data analysis, an AI-powered tool tailored for risk assessment and regulatory compliance could become an essential part of toxicological workflows in the coming days.

Fine-tuning LLMs for toxicological risk assessment is much cheaper than traditional methods currently in use. Traditional toxicological assessments are highly labor-intensive, requiring large teams and years to complete, with millions of dollars spent per chemical on testing and regulatory reporting. Conversely, fine-tuning an LLM is comparatively low in cost, ranging from \$0.0004 to \$0.03 per 1000 tokens for training and \$0.0016 to \$0.12 per 1000 tokens for usage, depending on the model and algorithm used. Fine-tuned LLMs enable automated data extraction, organization, and summarization, reducing manpower and time while maintaining regulatory compliance⁷.

The integration of large language models (LLMs) in toxicological risk assessment will streamline processes and enhance transparency. LLMs can automate tasks such as data mining, study identification, and dossier preparation from comprehensive datasets. These models minimize manual effort by extracting key data points, organizing them into regulatory formats, and identifying gaps for human input. Real-time data collection through AI systems ensures traceability and reduces the risk of manipulation, as raw experimental data can be directly integrated into assessments without human handling. AI can democratize access to toxicological data, enabling anonymized sharing

through public repositories. This fosters collaboration, supports informed decision-making, and makes vital information available to policymakers, researchers, and the public. Low-income countries and smaller organizations can leverage open-access AI tools for risk assessments with limited resources. While AI handles routine tasks, human experts remain central to decision-making, ensuring scientific rigor^{8,9}.

In summary, LLMs offer a practical and scalable solution for automating routine tasks in toxicology (RA) while ensuring that expert decision-making remains intact with humans. The future of LLMs in this domain lies not in replacing human expertise but making our job easier. In the near future, hybrid AI-human workflows could become the gold standard, where AI handles data aggregation and pattern recognition, while experts apply their domain knowledge to validate conclusions and provide nuanced risk evaluations. This collaborative AI-assisted approach would optimize efficiency without compromising scientific rigor^{10,11}.

ABBREVIATIONS

ACToR: Aggregated Computational Toxicology Resource (by US EPA), AI4NAMS: Artificial Intelligence for New Approach Methodologies in Safety Assessment, ChEMBL: Chemical Biology Database (by EMBL-EBI), COVID: coronavirus disease, DeepTox: Deep Learning for Toxicity Prediction, EPA: Environmental Protection Agency, GDPR: General Data Protection Regulation (EU), PubChem: Public Chemical Database (by NCBI), Tox21/ToxCast: Toxicology in the 21st Century/Toxicity Forecaster (US EPA and partners), REACH: Registration, Evaluation, Authorization, and Restriction of Chemicals (EU Regulation)

REFERENCES

1. Tsatsakis A, Goumenou M, Liesivuori J, Dekant W, Hernández AF. Toxicology for real-life risk simulation – Editorial preface to this special issue. *Toxicol Lett.* 2019;309:33–34. doi:[10.1016/j.toxlet.2018.12.003](https://doi.org/10.1016/j.toxlet.2018.12.003)
2. Tsatsakis AM, Renieri E. The concept of risk assessment evaluations in the 21st Century: The speedy and slippery progress of science and its political stand. *Public Health and Toxicology.* 2024;4(4):19. doi:[10.18332/pht/196857](https://doi.org/10.18332/pht/196857)
3. Kleinstreuer N, Hartung T. Artificial intelligence (AI)—it's the end of the tox as we know it (and I feel fine). *Arch Toxicol.* 2024;98(3):735–754. doi:[10.1007/s00204-023-03666-2](https://doi.org/10.1007/s00204-023-03666-2)
4. Singh B, Crasto M, Ravi K, Singh S. Pharmaceutical advances: Integrating artificial intelligence in QSAR, combinatorial and green chemistry practices. *Intelligent Pharmacy.* 2024;2(5):598–608. doi:[10.1016/j.iphpa.2024.05.005](https://doi.org/10.1016/j.iphpa.2024.05.005)
5. Sonnenburg A, van der Lugt B, Rehn J, et al. Artificial intelligence-based data extraction for next generation risk assessment: Is fine-tuning of a large language model worth the effort? *Toxicology.* 2024;508(153933):153933. doi:[10.1016/j.tox.2024.153933](https://doi.org/10.1016/j.tox.2024.153933)
6. Amanda D, Christopher G, Changho S, Frederic S. Evaluating language model context windows: A “working memory” test and inference-time correction. *arXiv.* doi:[10.48550/arXiv.2407.03651](https://doi.org/10.48550/arXiv.2407.03651)
7. Tetko IV, Klambauer G, Clevert D-A, Shah I, Benfenati E. Artificial intelligence meets toxicology. *Chem Res Toxicol.* 2022;35(8):1289–1290. doi:[10.1021/acs.chemrestox.2c00196](https://doi.org/10.1021/acs.chemrestox.2c00196)
8. Singh AV, Chandrasekar V, Paudel N, et al. Integrative toxicogenomics: Advancing precision medicine and toxicology through artificial intelligence and OMICs technology. *Biomed Pharmacother.* 2023;163(114784):114784. doi:[10.1016/j.biopha.2023.114784](https://doi.org/10.1016/j.biopha.2023.114784)
9. Da Silveira M, Deladiennee L, Acem K, Freudenthal O. Combining knowledge graphs and LLMs for hazardous chemical information management and reuse. *arXiv.* doi:[10.48550/arXiv.2412.09644](https://doi.org/10.48550/arXiv.2412.09644)
10. Schmeisser S, Miccoli A, von Bergen M, et al. New approach methodologies in human regulatory toxicology - Not if, but how and when!. *Environ Int.* 2023;178:108082. doi:[10.1016/j.envint.2023.108082](https://doi.org/10.1016/j.envint.2023.108082)
11. Hartung T. ToxAIcology - The evolving role of artificial intelligence in advancing toxicology and modernizing regulatory science. *ALTEX.* 2023;559–570 doi:[10.14573/altex.2309191](https://doi.org/10.14573/altex.2309191)

CONFLICTS OF INTEREST

The authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and none was reported.

FUNDING

There was no source of funding for this research.

ETHICAL APPROVAL AND INFORMED CONSENT

Ethical approval and informed consent were not required for this study.

DATA AVAILABILITY

Data sharing is not applicable to this article as no new data was created.

PROVENANCE AND PEER REVIEW

Not commissioned; internally peer-reviewed.

DISCLAIMER

The views and opinions expressed in this article are those of the authors.