

COMMITTEE ON CARCINOGENICITY OF CHEMICALS IN FOOD, CONSUMER PRODUCTS AND THE ENVIRONMENT

Horizon scanning update – November 2025

1. This paper presents an update on the table of topics raised under horizon scanning since 2024, taking on board feedback from the Committee on the table at the July 2025 meeting. Additionally other materials of interest flagged to or identified by the Secretariat are described.

Questions for the Committee

2. Members are invited to consider the draft table and other topics, and:
- i. indicate whether there are any other topics that the Committee should be aware of; and
 - ii. note whether any topics should be prioritised at this time.

Secretariat
November 2025

This is a background paper for discussion.
It does not reflect the views of the Committee and should not be cited.

COC watching brief of horizon scanning topics

Topics for progression

Meeting raised	Topic	Status
November 2024	Chemicals in Hair relaxers	July 2025: Secretariat discussing with OPSS potential consideration November 2025: Secretariat is having ongoing discussion with OPSS; there is interest in progressing the topic, however the scope requires consideration before a request is made of COC.
March 2025	Progression of “A case for change: the challenge to develop a better approach to assessing risk of cancer caused by chemicals”	July 2025: Document published. November 2025: Separate discussion paper (CC/2025/09) to review submissions received to date. Need to further consider dissemination and discuss with the smaller group of Members identified in July

Topics to be kept under a watching brief or not yet prioritised

Meeting raised	Topic	Status
November 2024	Chemicals in air fresheners used in the home	November 2024: Not prioritised to date. Secretariat notes that if prioritised, consideration if required on whether this should sit with COC or other Committees
November 2024	Risk of cancer from the regular use of mouth wash	November 2024: Not assessed further at this time July 2025: Secretariat note – any consideration would require discussion with OPSS
November 2024	AI	July 2025: COT is holding a workshop on AI in risk assessment alongside the October 2025 meeting. November 2025: COT workshop held in October – background paper provided to COT for discussion is attached at Annex A of this paper (to note this is reserved as it contains pre-publication material); workshop report to follow in due course.

This is a background paper for discussion.
It does not reflect the views of the Committee and should not be cited.

Meeting raised	Topic	Status
November 2024	PFAS	July 2025: COT subgroup undertaking review of PFAS; Secretariat will explore involvement of COC when relevant
March 2025	IARC 5 year strategy	July 2025: Secretariat noted IARC have a medium-term strategy document. The current version covers 2021-2025, with a 2026-2030 in development. With respect to the IARC Monographs an Advisory Group report is available recommending priorities for 2025-2029

Topics not considered a priority or outside COC's remit

Meeting raised	Topic	Status
November 2024	The risk evaluation for tris(2-chloroethyl) phosphate (TCEP)	November 2024: Not considered a priority and Secretariat notes this is likely to be within COT's remit if it were to be progressed.
November 2024	A reconsideration of car flame retardants	November 2024: Not considered a priority and Secretariat notes this is likely to be within COT's remit if it were to be progressed.
November 2024	Human risk of exposure to microplastics and microfibres	July 2025: COT has previously published an overarching statement and two substatements on microplastics November 2025: COMEAP has recently published its statement on inhaled nano- and microplastic air pollution . Information on the SETE assessments undertaken to sit alongside this are provided below.

Other information and topics for COCs awareness

3. Recently, an article in the Washington Post was flagged to the Secretariat, which provides a perspective on trends in research and substances of interest to the public: [The Washington Post. What researchers suspect may be fueling cancer among millennials.](#)
4. As noted in the tables above, COMEAP has published a [statement on inhaled nano- and microplastic air pollution](#). As part of this, COMEAP used the COT-COC Synthesis and Integration of Epidemiological and Toxicological Evidence (SETE) approach to consider the evidence base on [nano- and microplastic air pollution](#) and [traffic-related air pollution](#) as a comparator; these are attached at Annex B to this paper. While the topic itself is not currently a priority for COC, the Committee will welcome an assessment using the SETE approach.
5. At the July meeting, the COC had a presentation on the COM QSAR statement, and agreed to take forward a consideration of QSARs in the context of carcinogenicity and where they would complement other assessments. This could be added to the list of topics for progression, if Members agree.

This is a background paper for discussion.
It does not reflect the views of the Committee and should not be cited.

CC/2025/10 Annex A

**COMMITTEE ON CARCINOGENICITY OF CHEMICALS IN FOOD, CONSUMER
PRODUCTS AND THE ENVIRONMENT**

**Horizon scanning update – November 2025 (Reserved
business)**

COT background paper: Artificial Intelligence in Chemical Risk Assessment
(TOX/2025/39)

This paper and its Annex are attached for Members; they are not reproduced here for as they are reserved due to containing pre-publication information. Information from the paper will be available on the [COT website](#) in due course.

**Secretariat
November 2025**

TOX/2025/39

Committee on the Toxicity of Chemicals in Food, Consumer Products and the Environment

Artificial Intelligence in Chemical Risk Assessment (Reserved Business)

1. In February 2025, the [Potential future discussion items – horizon scanning paper](#) was presented to the Members of Committee on Toxicity of Chemicals in Food, Consumer Products and the Environment (COT). The Committee agreed that Artificial Intelligence (AI) would be a suitable topic for the next COT Annual Workshop. It was intended that the proposed workshop be a first step towards reviewing the state of the art of AI technologies relevant to chemical risk assessment as well as discussing the opportunities and the challenges associated with the application of AI in chemical safety assessment.
2. As background, the Secretariat informed the Committee that a scoping paper on AI in risk assessment considering these points would be presented prior to the workshop to compliment it.
3. In addition, this forms part of the [work the COT have been doing on integrating New Approach Methodologies \(NAMs\) in risk assessment](#) and continuing to develop a UK NAMs Roadmap (Osborne et al., 2024).
4. NAMs include *in silico* computer modelling strategies (e.g. AI and machine learning (ML)) for the evaluation of hazard and exposure of chemical in risk assessment. The COT briefly reviewed some *in silico* technologies in 2019 as part of the [NAMs scoping paper](#).

5. *In silico* technologies, driven by AI advancements in the last five years, have rapidly accelerated, coupled with substantial increases in computational resources which have made it feasible to train and deploy these large, complex models.
6. This scoping paper ([Annex A](#)) sets out a brief history of AI, the different AI spaces and application in chemical risk assessment (CRA). It reviews the state-of-the-art AI tools, discuss the opportunities and challenges of harnessing these technologies. Furthermore, the paper will explore the complexity of data ecosystems that they entail towards AI integration in chemical risk assessment in the regulatory setting.
7. This scoping paper will be transformed into a state of the science report.

Questions to the Committee

- i) Are Members content with structure and content of the state of science report?
- ii) Should anything else be included?
- iii) Can Members pull out any / many themes to take forward in this space?
- iv) Any frameworks the COT would like to review or use as guidance going forward in this area?
- v) Are the Members content with the proposed addition to the COT UK NAMs Roadmap?
- vi) Any other comments?

Secretariat

October 2025

Annex A

Committee on the Toxicity of Chemicals in Food, Consumer Products and the Environment

The attached annex is being treated as reserved as it contains pre-publication data.

Secretariat

October

2024

Annex A

Committee on the Toxicity of Chemicals in Food, Consumer Products and the Environment

Artificial Intelligence in Chemical Risk Assessment (**Reserved Business**)

“These are possibilities of the near future, rather than Utopian dreams”

-Alan Turing (1950)

Introduction and Background

1. In February 2025, the [Potential future discussion items – horizon scanning paper](#) was presented the Members of the Committee on Toxicity of Chemicals in Food, Consumer Products and the Environment (COT) The Committee agreed that Artificial Intelligence (AI) would be a suitable topic for the next COT Annual Workshop. It was intended that the proposed workshop be a first step towards reviewing the state of the art of AI technologies relevant to chemical risk assessment as well as discussing the opportunities and the challenges associated with the application of AI in chemical safety assessment.
2. As background, the Secretariat informed the Committee that a scoping paper on AI in risk assessment considering these points would be presented prior to the workshop to compliment.
3. In addition, this forms part of the [work the COT have been doing on integrating New Approach Methodologies \(NAMs\) in risk assessment](#) and continuing to develop a UK NAMs Roadmap (Osborne et al., 2024).
4. NAMs include *in silico* computer modelling strategies (e.g. AI and machine learning (ML)) for the evaluation of hazard and exposure of chemicals in risk

assessment. The COT briefly reviewed some *in silico* technologies in 2019 as part of the [NAMs scoping paper](#).

5. *In silico* technologies, driven by AI advancements in the last five years, have rapidly accelerated coupled with substantial increases in computational resources which have made it feasible to train and deploy these large, complex models.

Definition of AI

6. AI is “a technical and scientific field devoted to the engineered system that generates outputs such as content, forecasts, recommendations or decisions for a given set of human-defined objectives” ([ISO/IEC 22989:2022](#)).

Disclaimer

7. AI is a fast-moving field; therefore, the information in this scoping paper is up to date up until Summer 2025.

The scoping paper

8. This scoping paper sets out a brief history of AI, the different AI spaces and application in chemical risk assessment (CRA). It will review the state-of-the-art AI tools, discuss the opportunities and challenges of harnessing these technologies. The paper will also explore the complexity of data ecosystems that are needed towards AI integration in chemical risk assessment in the regulatory setting.

Brief History of AI

9. Alan Turing is widely considered to be the father of theoretical computer science and AI. His seminal paper in 1950 which proposed the question "*Can machines think?*" was published in *Mind Computing Machinery and Intelligence*. In it he wrote, “We are the more ready to do so in view of the fact that the present interest in 'thinking machines' has been aroused by a particular kind of machine, usually called an 'electronic computer' or 'digital computer'”. He proposed a test of machine

intelligence called The Imitation Game i.e. the Turing test (**Figure 1**) which is a test of a machine's ability to exhibit intelligent behaviour equivalent to that of a human.

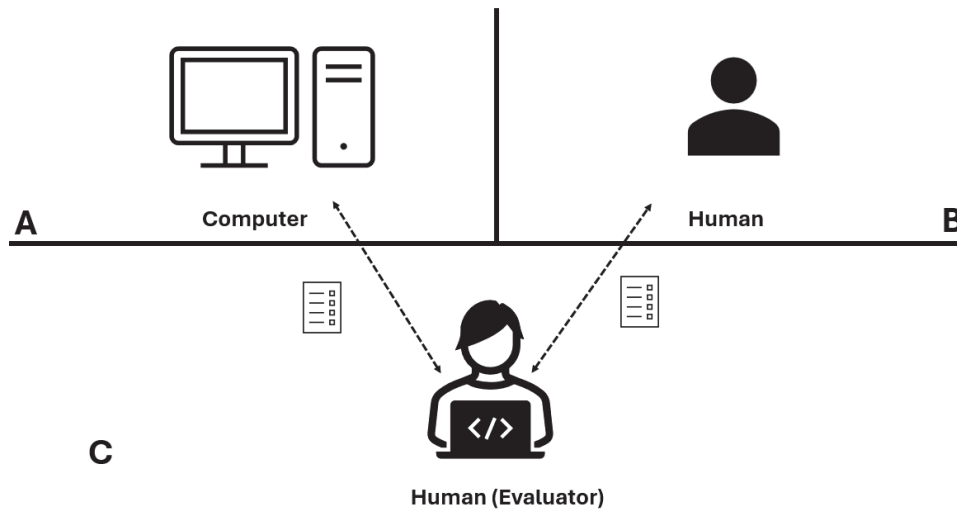


Figure 1. The "standard interpretation" of the Turing test, in which C (the evaluator), is given the task of trying to determine which player A or B, is the computer and which B is the human. The interrogator is limited to using the responses to written questions to make the determination.

10. In 1952, a computer scientist named Arthur Samuel developed a programme to play chequers, which is the first to ever learn the game independently (Samuel., 1959). The Samuel Checkers-playing Program was among the world's first successful self-learning programs using the "tree" decision (**Figure 2**).

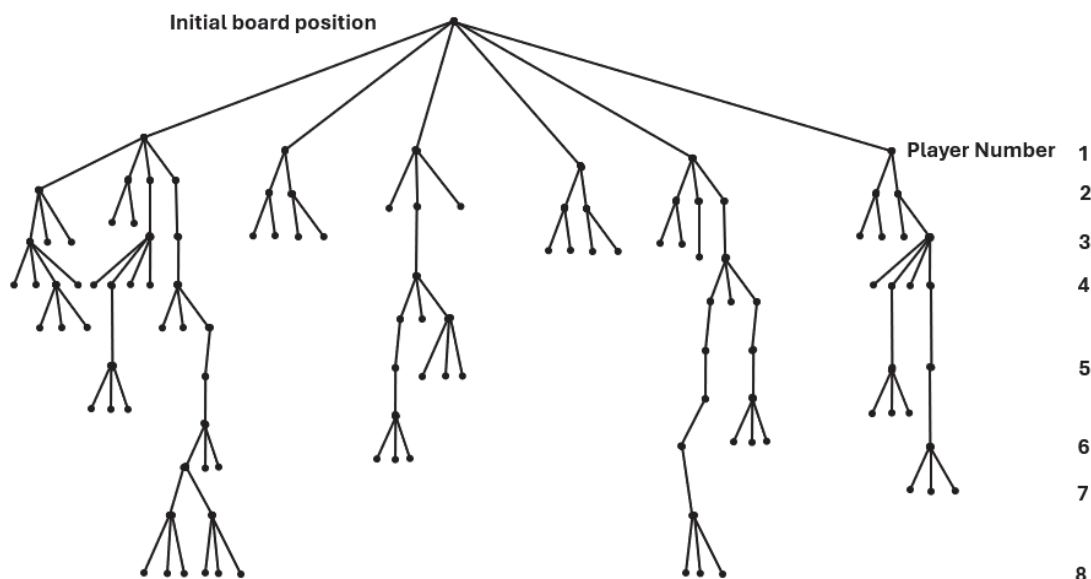


Figure 2. A "tree" of moves which might be investigated during the look-ahead procedure. The actual branchings are much more numerous than those shown, and the "tree" is apt to extend to as many as 20 levels (Figure adapted from Samuel, A.L., 1959).

11. As a result, in 1959, Arthur Samuel created the term "machine learning" (ML) when doing a speech about teaching machines to play chess better than the humans who programmed them.

12. However, the Dartmouth Summer Research Project of 1956 is often taken as the event that initiated AI as a research discipline (Moor, 2006). The project lasted approximately six to eight weeks and was essentially an extended brainstorming session.

13. In it, was a proposal that goes on to discuss computers, natural language processing, neural networks, theory of computation, abstraction and creativity (these areas within the field of AI are considered still relevant to the work of the field). The following are some aspects of the AI problem that was discussed:

Automatic Computers

14. If a machine can do a job, then an automatic calculator can be programmed to simulate the machine. The speeds and memory capacities of present computers (1956) may be insufficient to simulate many of the higher functions of the human brain, but the major obstacle is not lack of machine capacity, but our inability to write programs taking full advantage of what we have.

How Can a Computer be Programmed to Use a Language

15. It may be speculated that a large part of human thought consists of manipulating words according to rules of reasoning and rules of conjecture. From this point of view, forming a generalization consists of admitting a new word and some rules whereby sentences containing it imply and are implied by others. This idea has never been very precisely formulated nor have examples been worked out.

Neuron Nets

16. How can a set of (hypothetical) neurons be arranged so as to form concepts. Considerable theoretical and experimental work has been done on this problem by Uttley, Rashevsky and his group, Farley and Clark, Pitts and McCulloch, Minsky, Rochester and Holland, and others. Partial results have been obtained but the problem needs more theoretical work.

Theory of the Size of a Calculation

17. If we are given a well-defined problem (one for which it is possible to test mechanically whether or not a proposed answer is a valid answer) one way of solving it is to try all possible answers in order. This method is inefficient, and to exclude it one must have some criterion for efficiency of calculation. Some consideration will show that to get a measure of the efficiency of a calculation it is necessary to have on hand a method of measuring the complexity of calculating devices which in turn can be done if one has a theory of the complexity of functions. Some partial results on this problem have been obtained by Shannon, and also by McCarthy.

Self-Improvement

18. Probably a truly intelligent machine will carry out activities which may best be described as self-improvement. Some schemes for doing this have been proposed and are worth further study. It seems likely that this question can be studied abstractly as well.

Abstractions

19. A number of types of "abstraction" can be distinctly defined and several others less distinctly. A direct attempt to classify these and to describe machine methods of forming abstractions from sensory and other data would seem worthwhile.

Randomness and Creativity

20. A fairly attractive and yet clearly incomplete conjecture is that the difference between creative thinking and unimaginative competent thinking lies in the injection of some randomness. The randomness must be guided by intuition to be efficient. In

other words, the educated guess or the hunch include controlled randomness in otherwise orderly thinking.

21. In 1966, the Massachusetts Institute of Technology (MIT) computer scientist Joseph Weizenbaum created the first “chatterbot” (later shortened to chatbot), ELIZA. ELIZA was like a mock psychotherapist intended to simulate therapy, that used natural language processing (NLP), the famous script: DOCTOR, to converse with humans by repurposing the answers users gave into questions that prompted further conversation, also known as the Rogerian argument. Like Turing, Weizenbaum intended the program as a method to explore communication between humans and machines.

22. In 1974, the applied mathematician Sir James Lighthill (Lighthill, 1973) published a critical report on academic AI research, claiming that researchers had essentially over-promised and under-delivered when it came to the potential intelligence of machines. His condemnation resulted in stark funding cuts. A decade later, “AI winter” was coined, that referred to the gap between AI expectations and the technology’s shortcomings.

23. The American Association of Artificial Intelligence which is now known as the [Association for the Advancement of Artificial](#) Intelligence (AAAI) was founded in 1979. The AAAI is “dedicated to advancing the scientific understanding of the mechanisms underlying thought and intelligent behaviour and their embodiment in machines”.

24. AI then experienced several AI winter/ hype cycles until early 2000 with renewed interest in AI, the field experienced significant growth which led to increasingly intelligent machines. Within this period a “social robot” capable of identifying and simulating human emotions named Kismet, created in MIT’s Artificial Intelligence Laboratory and helmed by Dr. Cynthia Breazeal, contained sensors, a microphone, and programming that outlined “human emotion processes” using synthetic nervous system (SNS). Meanwhile, IBM had created a computer system named Deep Blue, a chess-playing computer programme to compete against then-world chess champion Gary Kasparov in a six-game match-up.

25. In the meantime, by the 1990s, ML-based techniques were being used in drug discovery using quantitative structure–activity relationship (QSAR) (described later under [QSAR Modelling](#)). Although Hansch and co-workers seminal work on QSAR, *i.e.* establishing the relationship between structure and activity was first reported in 1962 that Hammett functions and partition coefficients would become significant in that they can extract the chemical, physical, and biological functions embedded in a massive data set of complex molecular structures. Throughout this transformation, QSAR became a vital component of drug discovery, allowing for the highly efficient, low-cost prediction of activities and properties as well as structure-based virtual screening of potentially active hits from chemical libraries composed of millions of drug candidates.

26. It wasn't until 2011, during a presentation about its iPhone product that Apple showcased a new feature: a virtual assistant named Siri. Natural language processing capabilities that could understand a spoken question and respond with an answer. Interestingly, Siri was a spin-out from the Stanford Research Institute's Artificial Intelligence Center and is an offshoot of the US Defense Advanced Research Projects Agency's (DARPA)-funded Cognitive Assistant that Learns and Organizes (CALO project). CALO was an AI project that attempted to integrate numerous AI technologies into a cognitive assistant.

27. Even though computer scientist Geoffrey Hinton began exploring the idea of neural networks (computational systems inspired by the structure and function of the human brain, designed to recognize patterns and make predictions based on data) while working on his PhD in the 1970s; it wasn't until circa 2012 when he and two of his graduate students from University of Toronto displayed their research at the competition ImageNet, that the tech industry saw the ways in which neural networks had progressed. It was then, where funding and computer capabilities synergised to drive deep learning techniques into the spotlight. By 2016, developed by the London-based DeepMind Technologies the AlphaGO a computer program that plays the board game Go beat Lee Sedol, one of the best players in the world. AlphaGo is a combination of neural networks and advanced search algorithms trained to play Go

using a method called reinforcement learning, which strengthened its abilities over the millions of games that it played against itself (Silver et al., 2016).

28. In 2017 with the transformer architecture growth accelerated, [Google Lens](#) image analysis and comparison tool was released with associates of millions of landscapes, artworks, products and species to their text description.

29. This led to the AI boom of the early 2020s, with companies, universities, and laboratories overwhelmingly based in the United States pioneering significant advances in AI and data architecture.

30. Microsoft then introduced its Turing Natural Language Generation (T-NLG), which is the "largest language model ever published at 17 billion parameters".

31. Then OpenAI introduces GPT-3, a state-of-the-art autoregressive language model that uses deep learning to produce a variety of computer codes, poetry and other language tasks exceptionally similar, and almost indistinguishable from those written by humans. Its capacity was ten times greater than that of the T-NLG (Brown et al., 2020).

32. The first version of DALL-E was announced in January 2021. DALL E is a text-to-image models developed by OpenAI using deep learning methodologies to generate digital images from natural language descriptions known as prompts.

33. By 2022, ChatGPT, an AI chatbot developed by OpenAI, debuts. It uses generative pre-trained transformers (GPTs), to generate text, speech, and images in response to user prompts. While it gains considerable praise for the breadth of its knowledge base, deductive abilities, and the human-like fluidity of its natural language responses, it also gains criticism for its tendency to "hallucinate". Hallucination is a phenomenon in which an AI responds with factually incorrect answers with high confidence. This triggers widespread public discussion on and its potential impact of AI on society.

34. In 2024, Google DeepMind unveils [AlphaFold Server](#) a web-service that can generate highly accurate biomolecular structure predictions containing proteins,

DNA, RNA, ligands, ions, and also model chemical modifications for proteins and nucleic acids in one platform.

35. In 2025, Mistral AI releases Le Chat, an AI assistant able to answer up to 1,000 words per second.

The *Here and Now*

36. We are now journeying our way through a vast and rapid technological revolution that will fundamentally alter the way we live, work, and relate to one another. AI has been stated to transform many aspects of our society.

37. It has been discussed that the use of AI and ML will prompt the transition towards a more efficient risk assessment, as it allows more accurate, efficient use of the available data and knowledge.

38. AI is considered *in silico* systems (computational tools) that can champion the 3R principles of Replacement, Reduction, and Refinement of animal testing, and increase human safety and support critical decisions. They are increasingly being cited in the toxicology, academic and regulatory field including guidance documents which are forming a key element of New Approach Methodologies (NAMs).

39. The development and use of AI/ML methods has evolved rapidly, the computational capacity in conjunction with the availability of large data sets, particularly due to an increase in available data (Pawar et al., 2019) and computational capacity. We are at a pivotal point in time to harness these technologies to use the best available science for optimum risk assessment methodologies and output decision making. It is said that rather than just replicating human skills at larger scales, AI should be viewed as a transformative technology to advance and innovative existing systems to creating new capabilities.

40. In this discussion paper we review the state-of-the-art AI tools, discuss the opportunities and challenges of harnessing these technologies. Furthermore, the paper will explore the complexity of data ecosystems that they require to enable AI integration in chemical risk assessment and beyond.

The need-*Chemical Universe is expanding*

41. Given the rapid advancement in emerging technologies across chemical industries, the global regulatory agencies are increasingly required to conduct safety evaluation and chemical risk assessment (CRA) of the growing number of chemicals. It has been stated that AI integration can significantly enhance regulatory efficiency and decision-making (Kleinstreuer et al., 2024).

42. A recent survey of the chemical inventories of 19 countries and regions revealed that about 350,000 chemical substances have been registered for production and large-scale use over the past 30 to 40 years (Wang et al., 2020). Egeghy et al. (2023) found that, out of 547,000 substances, available environmental measurement data were very limited with water having the largest number of unique compounds (1,150 chemicals).

43. Many factors influence the toxicity of chemicals: the route of exposure (oral, dermal, or inhaled), the dose, frequency, and duration of exposure, specific properties related to Absorption, Distribution, Metabolism, and Excretion/Elimination (ADME), interactions between exogenous or endogenous substances, subject characteristics (age, sex, or body mass), and specific physicochemical properties (lipophilicity, solubility, boiling point, among others). All these parameters and tests make it challenging and sometimes not feasible through animal experiments which are often limited, financially costly, time consuming, and ethical considerations.

44. Furthermore, testing such many substances through animal experiments is not feasible. Numerous computer techniques, including read-across, structural warnings have been applied in recent decades to forecast the toxicological effects of compounds.

AI Opportunities- *Deus Ex Machina?*

45. Chemical safety assessments have evolved from relying on observations made in animal studies through alternative assays which are still primarily observation-based through to approaches that seek to more formally contextualise

and integrate knowledge using expert judgement through to defined approaches where decisions can become rule-based (Pawar et al., 2019).

46. In recent years, AI has demonstrated substantial potential in various domains (**Figure 3**), including drug discovery, disease diagnosis, protein structure prediction, chemical synthesis, and the discovery of new materials (Huang et al., 2025).

47. AI-based methods and models are expected to greatly contribute to the efficiency of performing chemical risk assessments and to improve our mechanistic understanding of the toxic outcomes associated with chemical exposure (Sonnenburg et al., 2024). Overall, AI tools have a central role to play in aggregating, visualising, contextualising and analysing complex data, exposing uncertainty, injecting expert knowledge, and ultimately supporting safer and more confident decisions (Barber et al., 2024). Applications include predictive toxicology, cumulative exposure modelling, and the use of high-dimensional data (e.g. omics, imaging, environmental data). AI's scalability is particularly valuable in managing the increasing volume and complexity of regulatory data (Hartung et al., 2025).

48. Interestingly, from the stakeholder interviewees in a recent Food Standards Agency Report on [New Approach Methodologies \(NAMs\) to Support Regulatory Decisions for Chemical Safety | Published in FSA Research and Evidence](#) it was discussed that AI could offer opportunities to reduce the quantity of animals used in testing. In addition, these could accelerate the use of NAMs and that ML and AI could allow the assessment of mechanisms that were previously too complex to model. This is due to the significant advances in large language models and AI in the last couple of years. In order to leverage large datasets to predict toxicological outcomes, machine learning models and AI are being developed to facilitate their interrogation. These models are expected to be able to analyse chemical data, predict toxicity endpoints, and inform regulatory decision-making processes (Myatt et al., 2018).

49. There is widespread acceptance that the success of AI-based methods and models in toxicology will depend on the ability to complement this domain with critical decision support tools and meaningful human oversight (Barber et al., 2024).

50. AI offers numerous benefits across various industries and applications. Some of the most commonly cited benefits include automation of repetitive tasks; more and faster insight from data; enhanced decision-making, fewer human errors and 24/7 availability.

51. Toxicity-related end points have always been challenging to evaluate experimentally with respect to *in vivo* translation due to the required resources for human and animal studies; this has impacted data availability in the field. ML can augment or even potentially replace traditional experimental processes depending on the project phase and specific goals of the prediction.

52. In 2018, the Joint Research Centre (JRC)-organised workshop, titled “AI4CRA – Artificial Intelligence for Chemical Risk Assessment” identified the following topics as prime candidates to be examined more closely when further efforts are undertaken to make AI an enabling technology in CRA: identifying problems; gathering evidence; systematic review; knowledge discovery; supporting evaluation; finding experts; facilitating collaboration; process simulation and cognitive modelling (Wittwehr, et al., 2020).



Figure 3. Some possible examples of artificial intelligence (AI) in chemical risk assessment. PBPK = physiologically based pharmacokinetic, (Q)SAR = (quantitative) structure-activity relationship, AOP = adverse outcome pathway, DL = deep learning, Tox = toxicology, ML = machine learning, qAOP = quantitative AOP, NAMS = new approach methodologies.

Benefits of AI based approaches

53. In a 2025 paper Hartung and Kleinstreuer stated some potential benefits of AI-based approaches:

- **Increased efficiency:** AI can rapidly retrieve, process and analyse vast amounts of data, significantly reducing the time and resources required for toxicity assessments.
- **Improved accuracy:** By identifying complex patterns and relationships in data, AI models can potentially provide more accurate predictions of toxicity than traditional methods.
- **Reduced animal testing:** AI-based NAMS offer the potential to reduce reliance on animal testing by providing reliable alternatives for many toxicity endpoints.

- **Enhanced mechanistic understanding:** Advanced AI techniques can help elucidate mechanisms of toxicity by identifying key molecular pathways and biological processes associated with adverse outcomes.
- **Better integration of diverse data:** AI can synthesize information from multiple sources and data types, providing a more holistic view of chemical toxicity.
- **Personalized risk assessment:** AI models can account for individual and population variability, enabling more tailored risk assessments.
- **Predictive power for novel compounds:** AI-based approaches can potentially predict toxicity for new or untested chemicals, supporting green chemistry initiatives and safer chemical design.
- **Continuous learning:** Unlike static models, AI systems can be designed to continuously learn and improve as new data becomes available.

54. In addition, in regulatory decisions, it might be able to offer objective evaluation where it makes a weight-of-evidence consideration in risk assessment less subjective and more reliable. AI tools could automate and streamline aspects of systematic reviews, making them more rapid and efficient.

Different Spaces of AI

55. The types of AI can be largely understood by examining two encompassing categories: AI capabilities and AI functionalities (**Figure 4**).

Demystifying the AI Landscape

56. These can then be divided into narrow AI, reactive, limited memory and theory of mind.

Artificial Narrow AI

57. Narrow AI also known as “Weak AI” is the only type of AI that exists today. Any other form of AI is theoretical. It can be trained to perform a single or narrow task, often far faster and better than a human mind can.

Reactive

58. AI algorithms that can respond to various inputs but lack memory-based functionality to learn from previous encounters.

Limited memory

59. Limited memory encompasses the capabilities of reactive AI with the added ability to temporarily store data from past experiences, allowing it to leverage historical data for decision making. This powers a wide array of contemporary AI applications which include generative AI tools e.g. ChatGPT.

Theory of mind

60. AI that has the capability of algorithms to attribute mental states to the entities they interact with. These make decisions based on understanding and remembering emotions and adapting their behaviour accordingly during interactions.

Self-aware

61. Self-aware is the most advanced type of AI (and is still being developed). It is AI that has evolved to be on par with human intelligence, so much so, that it's self-aware.

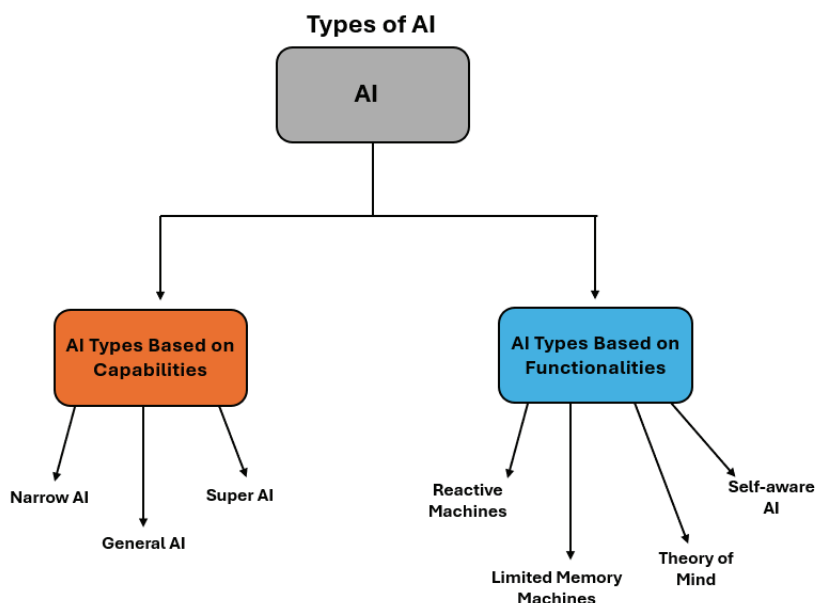


Figure 4. The different types of AI based on capabilities and functionalities. AI types based on capabilities include narrow, general and super AI. AI types of functionalities include reactive machines, limited memory machines, theory of mind and self-aware AI.

62. Within these various spaces of the AI landscape exist various fields of AI (**Figure 5**). AI and machine learning are often used interchangeably, but they're not the same. Within the AI field you have a subset of machine learning, another subset that is deep learning, followed by generative and within that explainable.

63. These all include learning, reasoning, knowledge representation, planning, natural language processing, perception, and support for robotics.

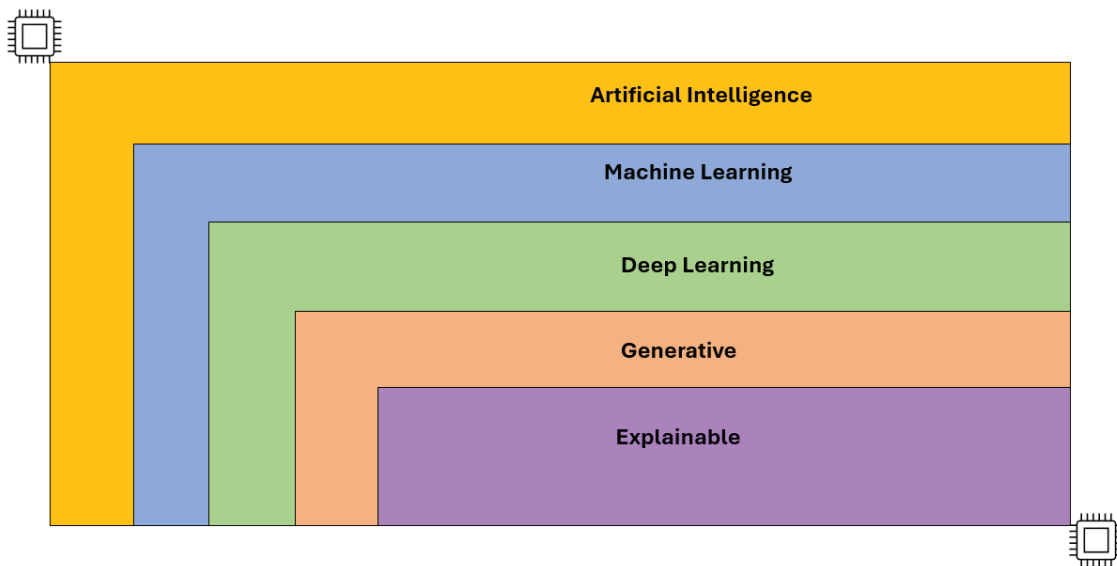


Figure 5. Different fields of AI. Machine learning, deep learning, generative and explainable.

Different fields of AI and applications

Machine Learning (ML)

64. Machine learning (ML) is the discipline of developing and deploying statistical algorithms that extrapolate and learn relationships and patterns from data, enabling statistically derived predictions, clustering and other useful manipulations, where

explicit prior domain knowledge is not necessarily required (Jordan et al., 2015 and Pandey et al., 2019).

65. There are several kinds of machine learning: supervised, unsupervised, semi supervised, self-supervised, reinforcement learning and transfer learning.

Supervised machine learning

66. Supervised machine learning is a type of machine learning where the model is trained on a labelled dataset (i.e., the target or outcome variable is known). It comprises several types of algorithms and is usually used for risk assessment, image recognition, or predictive analytics (Russel et al., 2021).

Unsupervised machine learning

67. Unsupervised learning analyses a stream of data, draws inferences from unlabelled datasets and finds patterns with no guidance which facilitates exploratory data analysis, enabling pattern recognition and predictive modelling.

Semi-supervised learning

68. Semi-supervised learning combines supervised and unsupervised learning by using both labelled and unlabelled data to train AI models for classification and regression tasks.

Self-supervised learning

69. Self-supervised learning generates implicit labels from unstructured data, rather than relying on labelled data sets for supervisory signals.

Reinforcement learning

70. Reinforcement learning is a type of dynamic programming which learns by trial-and-error and reward functions rather than by extracting information from hidden patterns. With repetition, the agent learns the best strategies.

Transfer learning

71. Transfer learning in which knowledge gained through one task or data set is used to improve model performance on another related task or different data set.

Machine Learning in chemical risk assessment

72. In research relevant to CRA, ML is commonly applied to predictive toxicology (Wang et al 2020 and Luechtefeld et al., 2018); models are trained on datasets of chemicals with known features (e.g. physicochemical properties relevant to Lipinski's Rule of Five, molecular fingerprints, molecular SMILES (Simplified Molecular Input Line-Entry System) strings, molecular graphs etc.) against corresponding toxicological properties (e.g. median lethal doses, median inhibitory concentrations for a particular protein target, target organ concentrations, no observed adverse effect level doses etc.) (Wang et al 2020; Luechtefeld et al., 2018 and Sinha et al., 2023). SMILES is a specification in the form of a line notation for describing the structure of chemical species using short American Standard Code for Information Interchange (ASCII) strings. SMILES strings can be imported by most molecule editors for conversion back into two-dimensional drawings or three-dimensional models of the molecules (**Figure 6**).

73. Therefore, the term SMILES refers to a line notation for encoding molecular structures and specific instances should strictly be called SMILES strings. Interestingly, SMILES was developed through funding from the U.S. Environmental Protection Agency, Mid-Continent Ecology Division-Duluth (MED-Duluth) Duluth,

Minnesota to the Medicinal Chemistry Project at Pomona College, Claremont, California and the Computer Sciences Corporation, Duluth, Minnesota.

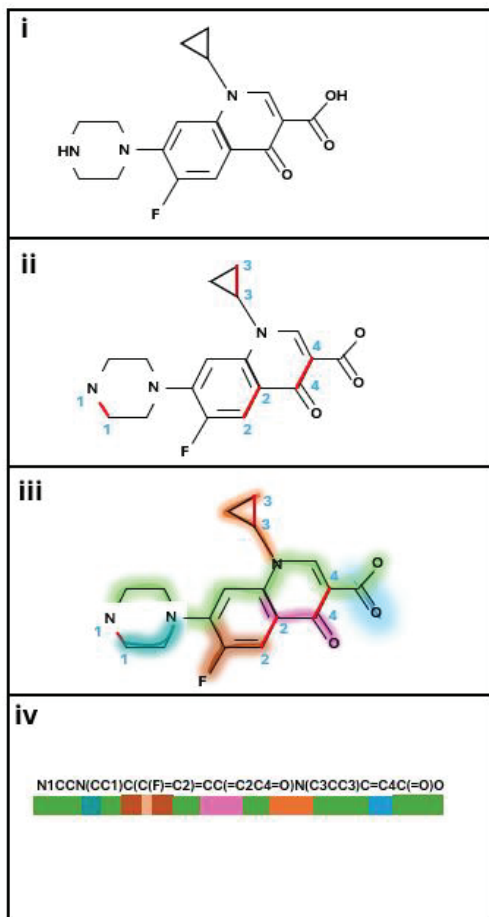


Figure 6. SMILES generation algorithm for ciprofloxacin: break cycles, then write as branches off a main backbone (i-iv).

74. Popular classical ML algorithms include random forests, support vector machines (SVMs), gradient boosting, logistic regression, and k-nearest neighbours, while deep learning driven algorithms have gained traction in more recent years (Wang et al 2020; Luechtefeld et al., 2018 and Sinha et al., 2023). Model accuracy typically depends strongly on careful descriptor selection, as well as inherent quality of underlying data (Sinha et al., 2023).

75. Supervised ML in CRA typically maps descriptors or learned features to hazard or potency data, using these algorithms, with statistically sound validation

(external splits, applicability domain) critical for regulatory use (Wang et al 2020; Luechtefeld et al., 2018 and Sinha et al., 2023). Feature importance and error analysis may be used to guide Weight-of-Evidence (WoE) integration and read-across decisions (Wang et al 2020; Luechtefeld et al., 2018 and Sinha et al., 2023).

76. A ML workflow in the chemical sciences typically includes: (1) Descriptor calculation (converting structures to numerical features); (2) Model training/validation (optimising and testing models using cross-validation or external test sets) and (3) Applicability domain analysis (checking whether predictions for new chemicals fall within the chemical space represented by training data) (Wang et al 2020; Luechtefeld et al., 2018 and Sinha et al., 2023).

77. Industry, academic and regulatory adoption is relatively established, yet still emerging, for ML-driven QSAR models (Wang et al 2020; Luechtefeld et al., 2018 and Sinha et al., 2023) which underpin tools such as the Organisation for Economic Co-operation and Development (OECD) QSAR Toolbox (Bhatia et al., 2015 Schultz et al., 2018). These models predict toxicological endpoints such as mutagenicity, aquatic toxicity, cardiotoxicity, endocrine disruption and others, which may assist regulators in prioritising substances of concern, while reducing reliance on animal testing (Bhatia et al., 2015 Schultz et al., 2018). For instance, ML models report strong performance for freshwater aquatic toxicity (Wang et al 2020). ML models for Ames mutagenicity frequently exceed 80% overall accuracy (Wang et al 2020 and Luechtefeld et al., 2018); and acute oral toxicity prediction continues to improve across industrial chemicals (Wang et al 2020 and Luechtefeld et al., 2018).

78. A 2010 paper outlined ideal practices for QSAR modelling (which applies to ML-based models), including rigorous data curation, descriptor selection, external validation and Applicability Domain (AD) assessment - for models intended for regulatory decisions (Tropsha, 2010). A more recent 2023 field-wide review outlined how modern ML/AI extends QSAR beyond traditional boundaries, emphasizing the need for reproducible workflows, transferability, and regulatory acceptance, across diverse endpoints (Belfield et al., 2023).

Machine learning in computational histopathology

79. Digital histopathological images, high-resolution images of stained tissue samples, are a vital tool for clinicians to diagnose and stage cancers. The visual analysis of patient state based on these images are an important part of oncology workflow. Although pathology workflows have historically been conducted in laboratories under a microscope, the increasing digitization of histopathological images has led to their analysis on computers in the clinic. The last decade has seen the emergence of machine learning, and deep learning in particular, a powerful set of tools for the analysis of histopathological images. Machine learning models trained on large datasets of digitized histopathology slides have resulted in automated models for prediction and stratification of patient risk (Cooper et al., 2023).

Deep Learning (DL)

80. Deep Learning (DL) is a specialised branch of ML, which uses artificial neural networks with multiple layers (**Figure 7**) to automatically learn complex patterns and hierarchical features from data (LeCun et al., 2015 and Goodfellow et al., 2016). In contrast to traditional ML, certain DL algorithms may work directly from raw or minimally processed data, reducing the need for handcrafted features (LeCun et al., 2015 and Goodfellow et al., 2016). DL often outperforms classical methods on complex tasks when sufficient data are available, against classical techniques, for approximating highly complex functions via black box models (LeCun et al., 2015 and Goodfellow et al., 2016).

81. The main DL architectures include: (1) Multilayer Perceptrons (MLPs) for tabular descriptor data (Popescu et al., 2009); (2) Convolutional neural networks (CNNs) for images or structured chemical fingerprints (O'shea and Nash, 2015); (3) Recurrent Neural Networks (RNNs) for sequential data like SMILES strings Medsker (2001); (4) Transformers (self-attention-based) for handling complex and large datasets (Vaswani et al 2007) and; (5) Niche families of architectures, such as Graph Neural Networks (GNNs) - which may directly process molecular graphs and other inherently graph-structure data (Wu et al., 2020).

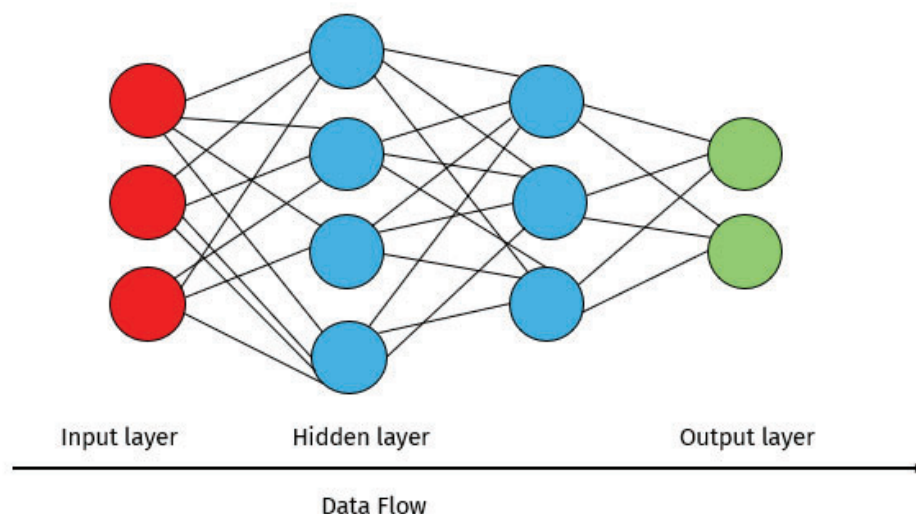


Figure 7. How neural networks work. A neural network consists of layers of nodes or artificial neurons, an input layer, one or more hidden layers, and an output layer. Each node connects to others and has its own associated value and threshold. If the output of any individual node is above the specified threshold value, that node is activated, sending data to the next layer of the network. If not, no data are passed along to the next layer of the network. Neural networks rely on training data to learn and improve their accuracy over time.

Deep learning in chemical risk assessment

82. In a CRA context, DL presents numerous opportunities. High-throughput virtual screening across multiple endpoints is a key opportunity; e.g. DeepTox (a deep neural network) outperformed classical ML in predicting multiple toxicity endpoints in the Tox21 challenge (Mayr et al., 2017), whereas other studies have outlined the potential to virtually screen large databases of known compounds (Pérez Santín et al., 2021 and Swanson et al., 2024). Multimodal toxicity prediction is another promising opportunity, as differing DL architectures can be combined to consider diverse insights into chemical structure (e.g. molecular graphs, SMILES, or even structures represented as images), alongside physicochemical properties, biological assay data, and other relevant information, drawing on unique strengths of different data sources and improving overall predictive performance (Karim et al 2019a; Hong and Kwon, 2025).

83. Deep learning is furthermore versatile enough to not just be applied to the prediction of toxicological properties (Wang et al 2020; (Mayr et al., 2017; Pérez Santín et al., 2021; Swanson et al., 2024; Karim et al 2019a; Hong and Kwon, 2025), but may also be used to rank substances in order of concern (Wassenaar, et al., 2021; Liu et al., 2025), generate novel molecules (e.g. for safer-by-design approaches, or for attempting to explore substitution space for harmful molecules) (Tan et al 2025; Zhang et al 2024), and even predict likely metabolites (Mittal et al., 2022; Tran et al 2023).

84. Further to the studies outlined, a 2023 study constructed DL models to output substructure-informed endocrine disruption and cytotoxicity predictions, enhancing mechanistic interpretability alongside competitive accuracy (Born et al., 2023); this demonstrates the potential for DL models to be interpretable by domain experts. A separate 2023 study combined DL with a variety of ML-based dimensionality reduction algorithms, to predict Ames mutagenicity (Kalian et al., 2023); this demonstrates the flexibility of DL models to be combined in workflows with other ML techniques, where beneficial.

85. Despite their predictive power, DL models present challenges and limitations; such models typically require large and high-quality datasets (Guo et al., 2023), are computationally intensive (Karim et al 2023b), and are often perceived as difficult to interpret "black boxes" (hence posing challenges for regulatory interpretability) (Kleinstreuer et al., 2024; Jia et al 2023). Industry, academic and regulatory uptake is growing but has lagged classical ML approaches, due to these interpretability and data limitations, as well as fundamentally posing as newer approaches in the wider ML and AI spaces (Kleinstreuer et al., 2024; Jia et al 2023). Promising emerging trends include the development of explainable DL methods (Kleinstreuer et al., 2024), transfer learning (adapting pretrained models to new endpoints) (Simões et al 2018), and integrating mechanistic biological knowledge (Singh et al., 2024).

Generative AI (GenAI)

86. Generative AI (GenAI) refers to models that create new, realistic data, which is statistically derived from similar training data (Banh and Strobel 2023; Epstein et

al., 2023). In chemical modelling applications, GenAI tools may propose novel, feasible chemical structures, enabling safer-by-design approaches and innovative chemical discovery (Zeng et al., 2022; Anstine and Isayev 2023).

87. Key GenAI approaches include (but are not limited to): (1) Generative Adversarial Networks (GANs) - where two competing networks (generator and discriminator) improve each other to create new molecules (Zeng et al., 2022; Anstine and Isayev, 2023); (2) Variational Autoencoders (VAEs) - deep neural networks which naturally compress data into spaces governed by statistical distributions, before then reconstructing the data, hence learning to build smooth latent representations of molecules that enable sampling of new, valid structures and; (3) Diffusion models - used for generating molecules by progressively adding statistical noise to training data, which the models then learn to remove, reconstructing the original data, while eventually reaching the ability to create entirely novel molecules from pure statistical noise inputs (Yang et al., 2023).

88. GenAI proposes structures by optimizing learned latent spaces or policies (e.g., VAE/Reinforcement Learning (RL)) subject to property predictors; conditional design and constrained sampling enable safer-by-design screening where toxicophores are avoided before experimental work starts (Yang et al., 2023; Anstine and Isayev, 2023). GenAI may also be combined with toxicological QSAR models as post-hoc filters, ensuring output molecules are statistically predicted to be safe, over covered endpoints.

GenAI in chemical risk assessment

89. GenAI is currently being used in various CRA-adjacent studies. One use-case is for designing safer chemicals e.g. Reinvent 4 (a RNN with reinforcement learning) enables generation of molecules that are guided to statistically hold favourable toxicological properties (alongside other desired properties) (Loeffler et al., 2024). GenAI is also being used to accelerate drug discovery - e.g. GANs and VAEs rapidly generate candidate drugs optimised for efficacy and reduced adverse effects (Loeffler et al., 2024). Predicting metabolites and transformation products is another use-case, with diffusion models having been trained for this purpose, whereas

transformer-based GenAI has outputted metabolites as SMILES strings (Bohde et al., 2025; Zhu et al., 2024; Litsa et al., 2020).

90. A 2018 Science review systematised inverse molecular design with VAEs, GANs and RNNs, enabling property-conditioned sampling and constrained optimisation workflows for de novo molecule generation (Sanchez-Lengeling and Aspuru-Guzik, A, 2018). Prior to this, a 2017 study used adversarial autoencoders to generate molecular fingerprints under property controls, proposing candidate leads with desired activity profiles (Kadurin et al., 2017). A 2020 study introduced constrained Bayesian optimisation in VAE latent space to boost validity and property-targeting during automatic chemical design (Griffiths et al., 2020). A review released in 2020 critically evaluated claims around generative models in medicinal chemistry, highlighting validation and synthesis-feasibility checks are needed before deployment (Walters and Murcko, 2020).

91. Despite these advancements, certain challenges and limitations persist; not all generated molecules are chemically plausible or viably synthesisable (Gao and Coley, 2020; Brown et al., 2020; Polykovskiy et al., 2020); predictive screening models are often still needed to assess safety (Chenthamarakshan et al., 2023), and regulatory acceptance still requires robust validation and traceability (European Medicines Agency, 2023); generative outputs alone are insufficient for decisions (European Medicines Agency, 2023). Emerging solutions involve integrated workflows of separately trained models for design and evaluation, conditional generation for regulatory endpoints, and improved transparency of generative processes (Lim et al., 2018; Yi et al., 2024).

Large Language Models (LLMs)

92. Large Language Models (LLMs) are (typically) transformer-based models scaled to hundreds of billions of trainable parameters and beyond, trained on vast text datasets, to model and statistically emulate complex human language communication - notably for use as AI chatbots (Rae et al., 2021., Chowdhery et al., 2018). LLMs significantly outperform other language models at parsing,

extracting, summarising, and drafting text (Rae et al., 2021., Chowdhery et al.,2018), making them increasingly valuable in biomedical and chemical research applications (including CRA) for mining literature, automating report drafting, and supporting regulatory submissions (Luo et al., 2022; Hatakeyama-Sato et al., 2023; Guo et al., 2023; Meskó and Topol, 2023; Tian et al., 2023).

LLM in chemical risk assessment

93. LLMs/Natural Language Processing (NLP) may support CRA by mining literature, safety reports, and patents to extract mechanistic evidence (Adverse Outcome Pathway (AOP) links) (Canada et al., 2017; Corvi et al., 2025; Chopard et al., 2021), endpoints (Corvi et al., 2025; Chopard et al., 2021), and exposure contexts (Schoene et al 2022). Agentic pipelines that perform chain retrieval, extraction and synthesis, under Subject-Matter Expert (SME) oversight (Luo et al., 2022 Hatakeyama-Sato et al., 2023; Guo et al., 2023), may improve first-pass dossiers and WoE tables (Lai et al., 2025).

94. LLMs such as GPT (Generative Pre-trained Transformer) (Yenduri et al 2024) and BERT (Bidirectional Encoder Representations from Transformers) (Devlin et al., 2019) use self-attention to capture linguistic and contextual information [15]. Two training stages are typical for creation of LLMs: pre-training on generic text and fine-tuning or prompt engineering for specific tasks (Devlin et al., 2019) (e.g. hazard extraction from regulatory dossiers (Raffel et al., 2020).

95. In the CRA space and adjacent domains (i.e. biomedical and chemical research), various LLM-based tools and studies have already been released, alongside earlier related advancements in the wider NLP applications space. For safety-relevant information extraction, the 2017 NLP-based LimTox system mined toxic endpoint associations of chemicals from the literature (Canada et al., 2017).

96. By 2021, language models integrated into deep-learning pipelines were used in studies to extract adverse events from clinical-trial narratives (Chopard et al 2021), with BERT-based models used to classify drug-induced liver injury from drug-labelling text (Wu et al., 2024).

97. In 2022, a study introduced BioGPT (a fine-tuned and specialised version of GPT), for biomedical text generation and mining (Luo et al., 2022), whereas a separate 2022 review surveyed use of NLP for extracting exposure information relevant to risk assessment (Schoene et al., 2022).

98. In the preclinical toxicology domain, a 2025 study (PretoxTM) automated extraction of treatment-related findings from study reports (Corvi et al., 2025), while a separate 2025 study demonstrated language models assisting data extraction and risk-of-bias assessment for evidence synthesis (Lai et al., 2025). Concerning more generic applications to research, a 2023 study benchmarked GPT-class models across eight tasks and separately explored GPT-4 prompt-engineering tactics for scientific use (Guo et al., 2023). Broader 2023 reviews outlined opportunities and challenges of LLMs in biomedical research and urged robust regulatory oversight for healthcare applications - central concerns for potential CRA deployments (Meskó et al., 2023).

99. Despite these opportunities from LLMs, current limitations include the risk of hallucinated (incorrect) content (Huang et al., 2025) challenges in prompt design (Liu et al., 2023; Khatun and Brown, 2023), and the need for human verification (Ong et al., 2024). Regulatory guidance is furthermore actively evolving to address key questions concerning transparency, traceability, and data privacy, associated with use of LLMs (Veale et al 2023; USA FDA).

Graph Neural Networks (GNNs)

100. GNNs are a family of deep learning architectures designed to operate directly on graph-structured data - learning node, edge, and graph-level representations through iterative message passing between connected nodes (Scarselli et al., 2008; Wu et al., 2021; Zhou et al 2020). Unlike traditional ML or DL methods that potentially require heavily modified or pre-calculated descriptors, GNNs may learn from raw molecular graph topology and associated atom or bond features, making them particularly suitable for molecular modelling, where chemical compounds may

be intuitively represented as graphs of atoms (nodes) and bonds (edges) (Scarselli et al., 2008; Wu et al., 2021; Zhou et al 2020).

101. Notable GNN variants include: (1) Graph Convolutional Networks (GCNs), which apply the notion of convolutions (commonly used in CNNs for computer vision tasks) to graph-structured neighbourhoods of nodes (Wu et al., 2021; Zhou et al., 2020; Kipf, 2016); (2) Graph Attention Networks (GATs), which weight neighbour contributions using self-attention (Zhou et al., 2020; Velickovic et al., 2017); (3) Graph Isomorphism Networks (GINs), which use MLPs to govern message passing, in a way which outperforms GCNs and GATs for the Weisfeiler-Leman graph isomorphism test (Wu et al., 2021; Zhou et al., 2020; Xu et al., 2018); (4) Message Passing Neural Networks (MPNNs), which unify various GNN formulations under a flexible message-update framework (Wu et al., 2021; Zhou et al., 2020; Gilmer et al., 2017); and (5) Spectral-based GNNs, which define graph convolution in the Laplacian (graph-Fourier) basis using learned spectral filters (this is distinct from message passing) (Wu et al., 2021; Zhou et al., 2020; Bruna et al., 2013). A significant number of other GNN variants also exist, which are not exhaustively listed in this review, for example, GNN variants which incorporate transformer-based architectures for message passing (Yun et al., 2019), as well as variants which hold scalability advantages for larger graphs (in terms of numbers of nodes and edges) (Hamilton et al., 2017).

GNN in chemical risk assessment

102. In a CRA-relevant context, GNNs have been successfully applied in a variety of advantageous ways. GNN-based QSAR models have been used in studies to predict over a broad range of toxicological endpoints e.g. a 2021 study introduced AmesFormer, which utilised graph transformer networks to classify Ames mutagenicity at state-of-the-art performance (Thompson et al., 2021). Furthermore, a 2023 study introduced GeoDILI, consisting of a GIN-based QSAR model to predict drug-induced liver injury (Wu et al., 2023). A 2024 study introduced AttentiveSkin, which used AttentiveFP (another GNN variant) to predict skin corrosion/irritation (Huang et al., 2024). All models in these studies directly processed molecular graphs, however other GNN-based QSAR models have also used knowledge graphs

that incorporate molecules as nodes, connected to similar other molecules via edges, such as a 2024 study which applied GATs to a molecular knowledge graph to predict points of departure relevant to neurotoxicity, developmental toxicity, and reproductive toxicity (Kalian et al., 2024). While specific choices of GNNs are often arbitrary or poorly justified in this space, a 2025 study compared the ability of GCNs, GATs, and GINs, to predict over 7 different toxicological assay datasets, ultimately concluding that GATs hold unique strengths over data-scarce environments (Kalian et al., 2025).

103. GNN-based models are also emerging in use for prediction of protein-ligand interactions (e.g. binding affinity – predicted using GNNs in a 2021 study (Li et al., 2021) and a 2023 study (Mastropietro et al., 2023), which may aid in outputting more mechanistically interpretable contributions of AI to CRA. Even for models solely operating on molecular graphs, devoid of associated proteins as a mechanistic context, GNNs still pose advantages for mechanistic interpretability, as model weights may be used to directly derive structural alerts and other interpretable findings (Thompson et al., 2021; Wu et al., 2023; Kalian et al., 2025).

104. GNN advantages include their ability to exploit relational inductive bias, learn task-specific chemical representations, and reduce dependence on manual feature engineering (Thompson et al., 2021; Wu et al., 2023; Kalian et al., 2025). However, challenges remain - GNNs are computationally intensive for large graphs (Zeng et al., 2018; Chaing et al., 2019), are prone to over-smoothing (loss of node feature distinction at excessively deep layers) (Rusch et al., 2023; Chen et al., 2020), and conventional GNN-based approaches still face interpretability issues similar to other DL models (Wu et al., 2023; Kakkad et al., 2023). Emerging trends involve explainable GNN frameworks for highlighting influential atoms/bonds (Ying et al., 2019; Luo et al., 2020), hybrid GNN-mechanistic models (Yuan and Nault, 2025; Liang et al., 2022), and scaling GNNs for ultra-large chemical libraries (Méndez-Lucio et al., 2024; Sypetkowski et al., 2024; Kuan et al., 2023).

Explainable AI (XAI)

105. Explainable AI (XAI) covers methods that enable AI model predictions that are more understandable to humans, with the aim of increasing trust, supporting regulatory use, and enabling more robust scientific scrutiny of predictions/outputs (Ying et al., 2024; Luo et al., 2020; Ebers, 2020; Guidotti et al., 2018). This is crucial for regulatory domains, such as CRA, where evidence-based decisions are critical (Kleinstreuer and Hartung, 2024; EMA, 2023 Meskó and Topol, 2023; Veale et al., 2021; US FDA, 2023; Ebers, 2020; Guidotti et al., 2018).

106. XAI spans inherently interpretable models and post-hoc explainers that interrogate trained models via perturbations, gradients, or additive decompositions (Guidotti et al., 2018) and are often used as part of generalised XAI frameworks (Lundberg and Lee, 2017; Ribeiro et al., 2023; Sundararajan et al., 2017). One example of a notable XAI framework is SHAP (SHapley Additive exPlanations), which assigns each feature a Shapley value - the unique additive attribution satisfying local accuracy, missingness, and consistency - yielding both local and global importance summaries (Lundberg and Lee, 2017). LIME (Local Interpretable Model-agnostic Explanations) fits a simple, human-readable surrogate (typically a sparse linear model) on perturbed samples around a specific instance, giving strictly local, model-agnostic explanations whose fidelity depends on the perturbation kernel and neighbourhood size (Ribeiro et al., 2023). Integrated Gradients is model-specific to differentiable networks: it integrates the output gradient along a path from a baseline to the input, producing attributions that satisfy sensitivity and implementation invariance and that sum to the prediction relative to the baseline (Sundararajan et al., 2017).

XAI in chemical risk assessment

107. In CRA-relevant research, XAI has been used to interpret outputs of QSAR models that rely on ML or DL architectures, as well as evaluate feature importance scores for molecular descriptors over associated toxicological endpoints, as well as statistically identify structural alerts (Walter et al., 2024; Jaganathan et al., 2022; Nascimento et al., 2023; Jia et al., Lee et al., 2025). A 2024 study mined structural alerts of mutagenicity and attempted to enable direct model interpretability, via direct processing and analysis of trained weights of DNNs trained to classify mutagenicity (Walter et al., 2024). Conversely, a 2022 study trained a variety of ML-based QSAR

models to predict respiratory toxicity, prior to application of SHAP to assess feature importance scores (Jaganathan et al., 2022). Structural alerts were derived from DNN-based QSAR models trained on the Tox21 dataset, using an implementation of LIME, in a 2023 study (Nascimento et al., 2023). A 2023 review outlined the wider space of using XAI for improved and interpretable workflows in computational toxicology (Jia et al., 2023). A 2025 study constructed GCN-based QSAR models of developmental and reproductive toxicity, while implementing a novel masking process and modification of the loss function, via use of the Louvain community detection algorithm, to detect structural alerts and aid in model interpretability (Lee et al., 2025).

108. XAI poses a promising avenue for increasing transparency of DL models in the toxicological sciences, to aid regulatory acceptance and increase the viability of DL predictive model adoption in industry and academia (Walter et al., 2024; Jaganathan et al., 2022; Nascimento et al., 2023; Jia et al., Lee et al., 2025). Challenges include the complexity and variability of explanations (Jiménez-Luna et al., 2020; Wellawatte et al., 2023) especially for large models (Turpin et al., 2023; Agarwal et al., 2023), and the need for validation against chemical or biological knowledge (Han, and Liu, 2022; Karim et al., 2023).

Next space of AI: Taking superintelligence seriously

109. A superintelligence is a hypothetical agent that possesses intelligence surpassing that of the brightest and most gifted human minds. Nevertheless, a general-purpose superintelligence remains hypothetical, and its creation may or may not be triggered by an intelligence explosion or a technological singularity.

110. Professor Nick Bostrom wrote the book *Superintelligence: Paths, dangers, strategies* in 2014. Superintelligence is roughly organized into three sections, as suggested by its subtitles. Bostrom discusses paths by which superintelligence (a system that vastly exceeds human levels of intelligence in virtually all areas) might be obtained. Next, he argues that the default outcome of developing superintelligence is likely to be catastrophic, motivating the need for substantial care

in such systems' design and governance. Finally, he critically analyzes possible strategies for ensuring that the development of superintelligence, if it does occur, proceeds safely and maximally benefits humanity. Cumulatively, these sections of the book constitute a persuasive demolition of what Bostrom calls the “null hypothesis”, namely that the future risks of AI need not be taken seriously today. Steering AI development toward safe and broadly beneficial outcomes is, to Bostrom, the “essential task of our age.” (Brundage et al., 2015).

AI Tools

111. The following are some examples of AI tools in the biological and chemical space.

Benchmark Dose (BMD) Modelling

112. Benchmark Dose (BMD) modelling is a mathematical machine learning computational method to estimate the dose of a substance that causes a predefined level of effect (the benchmark response) by fitting a mathematical model, or models, to dose-response data. Unlike traditional methods, which typically rely on identifying a single “no-effect” dose, BMD uses all available data to derive a statistically robust and informative reference point, typically taken as the lower confidence bound (BMDL) of the BMD estimate (**Figure 8**). By using the full dose-response dataset, this allows for a more statistically robust analysis, reducing the influence of arbitrary dose selection and increasing the reliability of the derived reference point (RP). BMD modelling also provides a quantitative estimate of uncertainty through the calculation of confidence or credible intervals around the BMD. Both the European Food Safety Authority (EFSA) and the US Environment Protection Agency (EPA) endorse BMD modelling as a scientifically superior method to the use of No-observed-adverse-effect level (NOAEL)/ Lowest-observed-adverse-effect level (LOAEL), particularly for deriving Health Based Guidance Values (HBGVs).

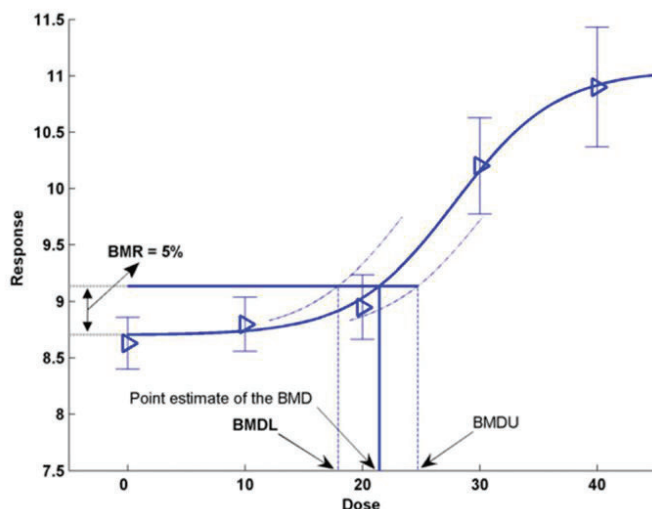


Figure 8. Illustration of the BMD approach using hypothetical continuous data (Figure reproduced from EFSA, 2017). Hypothetical experimental mean response data (triangles) are plotted along with their confidence intervals. The solid curve represents the fitted dose-response model. The curve determines the point estimate of the BMD, generally defined as a dose that corresponds to a low but measurable change in response, and here representing a benchmark response (BMR) of 5%. The dashed curves represent, respectively, the upper and lower 95% one-sided confidence bounds for the effect size as a function of dose. Their intersections with the horizontal line are at the lower and upper bounds of the BMD, denoted BMDL and BMDU, respectively.

113. BMD modelling has gained widespread acceptance among international regulatory bodies, including the EFSA, the U.S. EPA, and the World Health Organization (WHO), as a preferred method for deriving RPs and HBGVs. Both EFSA and EPA endorse BMD modelling as a scientifically superior method to the use NOAEL/ LOAEL.

114. BMD modelling can be applied to both continuous and quantal (binary) data and is suitable for a wide range of study types, including traditional *in vivo* toxicology, *in vitro* assays, and even emerging data streams such as transcriptomics and high-throughput screening. This adaptability makes BMD modelling particularly valuable in the context of NAMs, which aim to reduce reliance on animal testing and incorporate more mechanistic and human-relevant data.

115. BMD modelling can also support the derivation of RPs that are consistent and comparable across substances and studies. By defining the benchmark response (BMR), a predetermined level of change considered biologically relevant, this ensures that the RP corresponds to a known effect size. This means the calculation of relative potency factors (RPFs) and toxic equivalency factors (TEFs) are essential for assessing chemical mixtures and structurally related compounds. As the BMD can interpolate between tested doses, it is not constrained to the specific dose levels used in a study, unlike NOAELs.

116. The development of specialised software tools such as EFSA's [PROAST](#), and the [EPA's BMDS](#), has facilitated the practical application of BMD modelling. These tools offer user-friendly interfaces, model selection guidance, and visualisation capabilities, making BMD analysis more accessible to risk assessors. Recent advancements have also introduced Bayesian approaches and model averaging techniques, which further enhance the robustness of BMD estimates by incorporating prior knowledge and accounting for model uncertainty.

117. Whilst BMD modelling is a more sophisticated approach to dose response data, its use across the field of chemical risk assessment is not without challenges. These include the need for specialised training, variability in the statistical and computational tools, and divergences in international guidance. The development of OECD guidelines and further harmonisation of modelling practices will be essential for the broader adoption of BMD in regulatory toxicology.

COT discussions of BMD modelling

118. In 2021, the COT identified the need for UK-specific guidance on BMD modelling, recognising the importance of consistency in interpretation and implementation. In 2022, [a discussion paper in 2022 was brought to the committee](#) which reviewed international guidance, software tools, and methodological considerations. The paper also included a case study on PFAS to demonstrate the practical application of BMD in deriving HBGVs.

119. Whilst COT members recognised its scientific rigor and potential advantages, they also highlighted several challenges. The COT's review underscored both the promise and the complexity of integrating BMD modelling into UK chemical risk assessment. While supportive of its scientific merits, the committee advocates for a cautious, informed, and harmonised approach to implementation. Continued dialogue, training, and collaboration will be essential to realise the full potential of BMD in protecting public health.

Physiologically based pharmacokinetic modelling (PBPK) modelling

120. PBPK models are mathematical representations of physiological processes affecting a chemical and its metabolite(s)'s *in vivo* toxicokinetics; ADME or, simply put, describe the relationships between external exposure and the concentration-time profile of a chemical within the body (COT, 2021b).

121. The COT has evaluated PBPK modelling on several occasions, the latest being a report published in 2021 following the workshop on "[PBPK for Regulators](#)" (COT, 2003a, 2003b, 2007, 2009, 2021a).

122. To build a PBPK model, the traditional approach consists of empirically measuring the relevant parameters and then estimating the values of those not available by fitting to *in vivo* pharmacokinetic datasets. However, this approach is labour intensive, time-consuming, expensive and cannot keep up with the increasing demand of PBPK models (Lin and Chou, 2022).

123. ML approaches can efficiently predict PBPK parameters based on compounds' physicochemical properties to generate PBPK models for a large number of compounds. In turn, a PBPK model can be used to generate a large amount of simulated data to be analysed with ML approaches to obtain new insight (Lin and Chou, 2022).

124. Kamiya et al (2021) developed an *in silico* model based on a gradient boosting machine (i.e. an ensemble learning technique that builds a strong model by

combining many weak learners, typically decision trees), to predict three key PBPK parameters, including absorption rate constant, volume of distribution, and hepatic intrinsic clearance. This prediction was based on around 14–26 physicochemical properties of 246 compounds obtained from several cheminformatics software tools. These compounds were drugs, food components, and industrial chemicals with a broad range of chemical structures. The results showed that PBPK-predicted concentration values in plasma, liver, and kidney of rats using the *in silico* estimated parameter values were well correlated with those based on traditionally determined parameter values with a correlation coefficient of $r \geq 0.83$.

125. Lin et al., (2022) compared the adequacy of various ML algorithms when analysing the data generated by a generic PBPK model developed to predict the delivery efficiency of different nanoparticles to tumours in mice. The results showed that the deep neural network model (i.e. an artificial neural network with multiple hidden layers that transforms the input through weights, biases and activation functions into an output layer) outperformed all other ML methods and adequately predicted the delivery efficiency of different nanoparticles to different tumours. The corresponding R^2 values in the test dataset were 0.70, 0.46, 0.33 and 0.63 for the maximum delivery efficiency, delivery efficiency at 24 h, at 168 h, and at the last sampling time, respectively.

126. The strengths of using PBPK models in human health risk assessment are their ability to predict target organ dosimetry and the extrapolation capability across species, life stages, and exposure paradigms (Lin et al., 2024). Incorporating ML frameworks into PBPK models is increasingly appealing because these novel approaches not only largely speed up the efficient development of robust PBPK models for xenobiotics but also save substantial resources and have the potential to become an alternative strategy to traditional *in vivo*-data-based PBPK modelling. However, there are still some challenges and practical constraints in this field, e.g. a lack of sufficient interpretability of existing models due to the black-box nature of the ML algorithms, and the need to substantially reduce the number of chemical descriptors in the training set to avoid overfitting the model while ensuring crucial parameters to fit for the best results are retained (Chou and Lin, 2023).

(Q)SAR modelling

127. (Q)SAR is a computational modelling and simulation method for predicting the physicochemical, pharmacokinetic, toxicological, and environmental fate properties of chemicals based on the chemical structures and other relevant information (Lin et al., 2024; Lin and Chou, 2022).

128. Development of a (Q)SAR model typically involves four main steps: (1) collecting a training dataset (i.e. chemicals with known experimentally-derived physical and/or biological properties), (2) encoding chemicals with molecular descriptors, (3) training the model to predict chemical properties based on their molecular descriptors using mathematical algorithms, and (4) evaluating the model performance using a validation dataset. These training algorithms can vary from simple multiple linear regression to ML algorithms (Lin and Chou, 2022).

129. ML based (Q)SAR employs ML algorithms to model the relationship between physical and/or biological properties of compounds and their chemical structures. This approach is much more cost-effective for dealing with large data sets (Cheng and Ng, 2019). These models are also an ideal tool to perform read-across in toxicology (i.e. to predict the bioactivities of new chemicals based on structurally related or similar analogues without doing additional *in vitro* or *in vivo* experimentation) (Lin and Chou, 2022).

130. Cheng and Ng (2019) built a ML based (Q)SAR model to predict the bioactivity of per- and polyfluorinated alkyl substances (PFAS), a large family of ubiquitous environmental contaminants widely used in industrial and consumer products. By examining a number of available molecular data sets, the authors constructed a PFAS-specific dataset to serve as a training data set that contained bioactivity information on 1012 PFAS for 26 bioassays. These bioassays were all binary classification assays (active or inactive) and involved different target receptors or enzymes. Five different ML models (i.e. logistic regression, random forest, multitask neural network, graph convolutional model, and weave model) were evaluated on different assays with a dataset from the OECD containing bioactivity information on 3486 PFAS, and the best model was selected for each assay. The

results showed that the average of the area under the curve (AUC) score for each bioassay was 0.916.

131. Traditionally, the carcinogenic potency of a compound is evaluated with a 2-year carcinogenicity study in at least two rodents, but this process is very time-consuming and resource-intensive (Lin and Chou, 2022).

132. Li et al (2021) developed the DeepCarc model to predict carcinogenicity for small molecules using DL-based model-level representations. The DeepCarc model was developed with a training data set of 692 chemicals and evaluated with a validation data set consisting of 171 chemicals. Both data sets involved studies in rats and mice. The results showed that model predictions from DeepCarc had an accuracy of 0.754, a sensitivity of 0.910, and a specificity of 0.467 in the test dataset.

133. This DeepCarc model provides an early screening nonanimal-based tool to assess potential carcinogenicity of new chemicals and is useful for prioritising chemicals on their potential carcinogenic risk.

134. Emerging ML and AI approaches are now commonly employed to build robust (Q)SAR models to predict bioactivities of a large number of chemicals. However, some ML based (Q)SAR models are used for bioactivity classification only (yes/no for bioactivity) and cannot provide information about intensity of effect or dose–response (Cheng and Ng, 2019). The standardisation of training and validation splits (i.e. the process of dividing a data set into separate subsets for model training and performance evaluation) is also necessary in order to compare different predictive performance results (Vall et al., 2021).

AOP analysis

135. An AOP is a hierarchically structured pathway representation that describes existing knowledge on the connection between a direct molecular initiating event (MIE) and an adverse outcome (AO) at a biological level of organisation that is relevant to human health risk assessment (Lin and Chou, 2022).

136. A typical AOP includes a MIE triggered by an organism's contact with a stressor (e.g. chemical), one or more key events (KEs) that subsequently take place in cells and organs and characterise the progression of toxicity, and an AO that occurs at individual or population levels (Blümmel et al., 2024).

137. One of the most commonly studied AOPs is related to nuclear oestrogen receptor (ER) α and β . Oestrogen receptors play important roles in many biological functions, and multiple xenobiotics have been shown to activate oestrogen receptors, with the potential to result in endocrine disruption and adverse effects on reproductive organs (Lin and Chou, 2022).

138. Traditional approaches to evaluate endocrine disruptors that activate estrogenic signalling requires labour- and resource-intensive *in vitro* or *in vivo* experiments. However, AI-based high-throughput screening (HTS) assays are a more efficient approach to measure KEs of AOPs (Lin and Chou, 2022).

139. Ciallella et al., (2021) developed a deep neural network model to analyse publicly available HTS data to identify compounds with nuclear oestrogen receptor α and β binding potentials. The input layer contained information on 1024 functional connectivity chemical fingerprints plus three known ER α /ER β toxicophores. The output layer of the network was the target activity, i.e. *in vivo* rodent uterotrophic bioactivity. Between the input and the output layers, there were five hidden layers organised using an AOP framework, with each layer representing a higher level of biological organisation than the last. Each neuron included in the five hidden layers represented one *in vitro* HTS assay from the training dataset. The resulting model successfully predicted critical relationships among ER α /ER β target bioassays based on chemical fingerprints. The model used an AOP framework to mimic the signalling pathway initiated by ER α and was able to identify compounds that mimic endogenous oestrogens (AUC = 0.864–0.927).

140. One limitation of AOP analysis is that it does not provide quantitative relationships from chemical exposure to effect timing and magnitude. However, when there are sufficient data on quantitative relationships between chemical exposure and KE (or molecular initiating events or adverse outcomes), a mathematical model can be developed to connect chemical exposure to KEs in a quantitative AOP

(qAOP), which offers potential regulatory applications in chemical risk assessment (Lin and Chou, 2022).

141. Paini et al (2022) proposed a six-step structured framework for developing qAOPs, integrating mechanistic data from *in vitro*, *in vivo*, and *in silico* sources. The workflow begins with defining the problem and selecting relevant AOPs, followed by identifying and extracting reliable data sources. These data are then curated into a machine-readable format and used to build mathematical models, which are iteratively refined to ensure predictive accuracy and regulatory relevance. Three case studies (skin sensitisation, developmental neurotoxicity, and carcinogenicity) successfully demonstrated the application of Bayesian networks and dose–response models to quantify KE relationships.

142. AOP analysis aided by AI ultimately supports the transition towards more efficient, knowledge-driven chemical safety evaluations Paini et al (2022). However, experts in the field emphasise the importance of FAIR data principles (i.e. findability, accessibility, interoperability, and reusability) and curated e-resources to support model development, as well as moving away from traditional black-box models that are lack of mechanistical explainability to increase model credibility and regulatory acceptance (Lin and Chou, 2022; Paini et al., 2022).

143. Most recently a study introduced the novel software platform AOP Network Viewer, designed to address the critical challenges of scalability, interactivity, and deep analysis in toxicological research (Seal et al., 2025). The system is built on a scalable client-server architecture with a graph database backend, supporting sophisticated interactive features such as multi-pathway selection, dynamic path-finding, and advanced filtering. The principal innovation of the AOP Network Viewer is its integrated AI Analysis Panel, an LLM-powered engine that provides a deep, context-aware analysis of selected network components. This AI system utilizes a unique three-layer prompt architecture to deliver domain-specific insights across seven specialized analytical modes. Case studies demonstrate the platform's utility in developing alternative testing strategies for endocrine disruptors, interpreting high-throughput screening data, and generating hypotheses for toxicity assessment of mixtures. The AOP Network Viewer is engineered to translate complex user queries

into focused, interpretable biological networks, transforming AOP network visualization from a passive viewing experience into an interactive and intelligent analytical process, thereby accelerating mechanistic discovery and enhancing evidence-based regulatory decision-making.

GPT4 for chemistry

144. A recent paper (Hatakeyama-Sato et al 2023) evaluated the capabilities and limitations of the Generative Pre-trained Transformer 4 (GPT-4) in chemical research (**Figure 9**). Although GPT-4 exhibited remarkable proficiencies, it is evident that the quality of input data significantly affects its performance. They explore GPT-4's potential in chemical tasks, such as foundational chemistry knowledge, cheminformatics, data analysis, problem prediction, and proposal abilities. While the language model partially outperformed traditional methods, such as black-box optimization, it fell short against specialized algorithms, highlighting the need for their combined use. The paper shares the prompts given to GPT-4 and its responses, providing a resource for prompt engineering within the community, and concludes with a discussion on the future of chemical research using large language models.

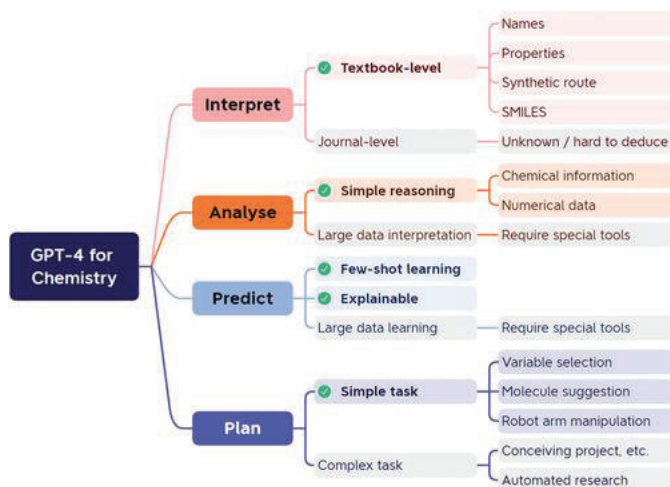


Figure 9. Overview of the capabilities of GPT-4 for chemical research. (Figure taken from Hatakeyama-Sato et al 2023)

ChemCrow

145. ChemCrow, an LLM chemistry agent designed to accomplish tasks across organic synthesis, drug discovery, and materials design (Bran et al., 2023). By integrating 18 expert-designed tools, ChemCrow augments the LLM performance in chemistry, and new capabilities emerge. Our agent autonomously planned and executed the syntheses of an insect repellent, three organocatalysts, and guided the discovery of a novel chromophore. Our evaluation, including both LLM and expert assessments, demonstrates ChemCrow's effectiveness in automating a diverse set of chemical tasks. ChemCrow is an open-source package for the accurate solution of reasoning-intensive chemical tasks.

Built with Langchain, it uses a collection of chemical tools including RDKit, paper-qa, as well as some relevant databases in chemistry, like Pubchem and chem-space.

Alpha Fold

146. An example using neural networks is AlphaFold (Jumper et al., 2021). This was the first computational method that can regularly predict protein structures with atomic accuracy even in cases in which no similar structure is known. [Alpha Fold](#) is a state-of-the-art AI system developed by Google DeepMind, is able to computationally predict protein structures with unprecedented accuracy and speed. Working in partnership with EMBL's European Bioinformatics Institute (EMBL-EBI), they released over 200 million protein structure predictions by AlphaFold that are freely and openly available to the global scientific community. Included are nearly all catalogued proteins known to science – with the potential to increase humanity's understanding of biology by orders of magnitude.

VEGA

147. [VEGA](#) provides tens of QSAR models to predict tox, ecotoxicology, environment, and phys-chem properties of chemical substances. Using the VEGA platform, you can access a series of QSAR models for regulatory purposes or

develop your own model for research purposes.

EFSA OpenFood Tox

148. [OpenFoodTox](#) provides open-source data for the substance characterisation, the links to EFSA's related outputs, background European legislation, and a summary of the critical toxicological endpoints and reference values. In addition, it contains physicochemical properties and pharmacokinetic/toxicokinetic data for more than 850 substances.

Comptox Chemicals Dashboard

149. The CompTox Chemicals Dashboard is a part of a suite of databases and web applications developed by the [U.S. Environmental Protection Agency](#) (EPA). The [CompTox Chemicals Dashboard](#) provides public access to chemical data. It is a widely used resource for chemistry, toxicity, and exposure information for over a million chemicals. The available data includes, but are not limited to, chemical properties, environmental fate and transport, hazard (e.g., point of departure, legacy toxicity values, screening levels, exposure limits), *in vitro* to *in vivo* extrapolation (IVIVE), exposure (e.g., predictions associated with the [ExpoCast](#) effort, use categories, toxic release inventory, monitoring data), bioactivity (e.g., high-throughput data from [ToxCast and Tox21 efforts](#)).

150. In addition, the Dashboard provides information on similar chemicals and related substances, chemical lists, links out to other reputable resources, including Open Source widgets and tools such as [PubChem widgets for Bioactivities, Articles, and Patents](#), and links to tools such as the [Generalize Read-Across tool](#) (GenRA), the web-version of the [Toxicity Estimation Software Tool \(WebTEST\)](#), and a web-version of the [Abstract Sifter](#).

151. [The Distributed Structure-Searchable Toxicity \(DSSTox\) Database](#) provides the chemical and chemistry foundation for the Dashboard. It provides chemical structure information, compiled and expertly quality-control checked chemicals,

which come from public sources such as PubChem. These data have varying levels of reliability and accuracy, with links to the external resources. Expansion, curation, and validation of the content is ongoing.

152. Users can search the Dashboard by chemical identifiers, consumer product categories, and assays/genes associated with high-throughput screening data. The Dashboard provides unique capabilities including real time prediction of physicochemical properties and toxicological endpoints, batch searching for thousands of chemicals at a time, and advanced searching approaches to support non-targeted screening mass spectrometry research effort.

ECHA CHEM

153. [ECHA CHEM](#) is a new public chemicals database with information from all REACH registrations received by the Agency.

AI tools applied in safety assessment examples

AI tools in the assessment of NAMs data (automation)

154. In 2024, the EFSA, under its Science Studies and Project Identification & Development Office (SPIDO), commissioned a project titled “AI4NAMs” to explore the potential of AI in automating the extraction, harmonisation, and integration of data from NAMs into chemical risk assessment workflows. The project was conducted by a consortium comprising d-fine GmbH, the German Federal Institute for Risk Assessment (BfR), and Wageningen University (Blümmel et al., 2024).

155. The primary aim was to evaluate AI tools for their ability to support the integration of NAM-derived data into AOP-like knowledge networks. The project was structured into three work packages:

- Work Package A: Evaluation of AI tools.
- Work Package B: Implementation of six case studies.
- Work Package C: Development of strategic recommendations for EFSA's NAMs and AI roadmaps.

156. A seven-step workflow was developed to guide the integration of NAMs data: initial data collection, initial review, data transformation, full paper/data review, harmonisation, data extraction and integration into AOP-like networks.

157. Tools were assessed using a two-round, weighted scoring framework based on general and workflow-specific criteria. Tools included:

- Natural Language Processing (NLP)/ LLMs: GPT-3, AlexaTM 20B, SciBERT, ScispaCy
- Parsing/Extraction: Grobid, IBM Deep Search, Tabula
- Review Platforms: DistillerSR, Sysrev
- Ontology/Storage: Owlready2, IUCLID Uploader
- Visualisation: Gephi, NetworkX

158. To evaluate the practical application of these AI tools across the seven different workflow steps, six case studies were conducted, each focusing on either a specific chemical group or a toxicological endpoint relevant to EFSA's remit. These included pyrethroids (neurotoxicity), phthalates (endocrine disruption), bisphenols (non-monotonic dose responses), dioxins (aryl hydrocarbon receptor-mediated toxicity), hypothyroidism (as an endocrine endpoint), and liver toxicity (hepatotoxic effects of pesticides).

159. Across six case studies, the integration of structured data (e.g. from ToxCast and PubChem) was generally successful, with tools like Grobid and IBM Deep Search performing well in parsing and extracting relevant information. However, the extraction of data from unstructured sources, such as scientific publications, presented more significant challenges.

160. LLMs, including GPT-3 and AlexaTM 20B, were tested for tasks such as harmonisation, information extraction, and reliability assessment. While these models showed promise, their performance was highly dependent on input formatting and prompt engineering. For example, GPT-3 achieved moderate accuracy in extracting data relevant to hypothyroidism and bisphenol toxicity but also

exhibited issues with error propagation and false positives. Similarly, AlexaTM 20B demonstrated instability in output consistency, limiting its reliability for automated data extraction.

161. NLP tools such as medspaCy and ScispaCy were employed for named entity recognition and relation extraction. These tools showed variable performance, often requiring retraining or manual correction to achieve acceptable precision and recall.

162. Ontology mapping using tools like Owlready2 achieved moderate success, with mapping accuracies ranging from 65% to over 90% depending on the complexity of the data.

163. Visualisation tools such as Gephi and NetworkX were effective in integrating and displaying AOP-like networks, aiding in the interpretation of mechanistic data.

164. Overall, while AI tools enhanced efficiency in several workflow steps, expert oversight remained essential, particularly in interpreting complex toxicological data and ensuring the scientific validity of outputs.

165. The project identified several areas where improvements and strategic developments could enhance the utility of AI in chemical risk assessment. For workflow-specific recommendations, it was advised that structured databases improve their documentation and accessibility through application programming interfaces (i.e. a set of rules and protocols that allows different software systems to communicate with each other, APIs), and that publishers adopt machine-readable formats to facilitate automated data extraction. The development of harmonised templates, such as expanded versions of the OECD Harmonised Template (OHT201), would support more consistent data integration. The use of ontologies and controlled vocabularies should be expanded to improve terminology alignment across data sources.

166. For full-text review and data extraction, the fine-tuning of LLMs was recommended, with careful consideration given to the trade-offs between using pre-trained models and developing domain-specific adaptations. Named entity recognition models should be further developed to improve the identification of key

toxicological concepts, and tools for extracting data from figures and tables should be refined to reduce manual workload.

167. From a broader perspective, the authors encouraged EFSA to adopt agile project management frameworks and invest in the development of AI microservices tailored to regulatory needs. The creation of an internal AI repository and the promotion of training initiatives would support capacity building within EFSA and its partners. Specific project proposals include the development of an enterprise search platform, a chatbot for risk communication, and targeted applications of AI in areas such as toxicokinetics (AI4ADME), endocrine disruption (AOP4ED), and cumulative risk assessment (ReFineCAG).

168. These recommendations aim to support the transition toward next-generation risk assessment (NGRA) by enabling more efficient, transparent, and mechanistically informed evaluations of chemical safety.

DL in food safety

169. Incidents of vegetables containing pesticide residues exceeding safety standards have been reported frequently, posing a significant threat to the safety of agricultural products due to the potential harmful effects of these pesticide residues. Consequently, rigorous monitoring of pesticide residue levels in vegetable produce have become essential measures to ensure the protection of human health (Zhou et al., 2023).

170. Mass spectrometry coupled with pretreatment and separation technologies have emerged as the gold standard for pesticide residue detection in laboratories, owing to their high specificity, sensitivity, and accurate qualitative and quantitative analysis. Due to the limited longevity of agricultural products, it is crucial to achieve rapid on-site detection of pesticide residues in vegetables. This goal has spurred the development of instrument miniaturisation in the field of food safety testing. However, miniature mass spectrometers exhibit decreased detection sensitivity, accuracy, and resolution, compared to laboratory mass spectrometers (Zhou et al., 2023).

171. Zhou et al., (2023) established a DL-enabled rapid qualitative and quantitative analysis method for pesticides on vegetable surfaces that aimed to enhance the on-site system detection capabilities. The authors combined a direct pretreatment by placing vegetables in a flexible ziplock bag with a mass spectrum recognition algorithm coupling one-dimensional convolutional neural network DL approach (i.e. a ML architecture designed to recognise patterns in data that has a temporal or ordered structure). Their method had a qualitative recognition accuracy of 99.62% and a quantitative detection limit of 10 µg/kg for carbendazim in cowpea.

172. Food fraud poses threats to both public health and the international economy. Meat products frequently fall victim to food fraud, owing to their elevated consumer demands and high selling prices. One category of meat fraud involves utilising inexpensive colourants or curing agents to disguise inferior quality meat as a higher-value product. Several of these additives are known to endanger consumer health (Jo et al., 2023).

173. Spectroscopic methodologies are widely acclaimed for addressing food safety concerns, owing to their capacity to provide comprehensive insights into the compositions and traits of various substances. However, their application for detecting counterfeit meat remains underexplored. Consequently, reliable, straightforward spectroscopic methodologies that can directly utilise spectral information for identifying counterfeit meat are urgently required (Jo et al., 2023).

174. Jo et al., 2023 developed a DL-based spectroscopic method for identifying counterfeit beef altered to appear fresh. The experiment involved 60 beef samples, half of which were artificially adulterated using a colouring solution. Imaging analysis confirmed that upon visual examination, the counterfeit group was indistinguishable from the standard group. The method achieved a classification accuracy of 98.84% in the 344–1040 nm spectral range. An external validation set of 70 beef samples was applied to further evaluate the model's robustness and generalisability, and comparative performance analysis revealed that this method significantly outperformed traditional ML models.

175. These studies show a promising application of the proposed deep learning-enabled analytical systems in food safety assessments. However, the authors highlighted the necessity for broader datasets, target chemicals and real-world application testing to ensure the robustness and generalisability of their DL models (Jo et al., 2023; Zhou et al., (2023).

Development of machine learning-based quantitative structure–activity relationship models for predicting plasma half-lives of drugs in six common food animal species

176. QSAR models were developed using data from the Food Animal Residue Avoidance Databank (FARAD) Comparative Pharmacokinetic Database. The deep neural network (DNN) algorithm demonstrated the best prediction ability of plasma half-lives. The DNN model with all descriptors achieved superior performance with a high coefficient of determination (R^2) of 0.82 ± 0.19 in 5-fold cross-validation on the training sets and an R^2 of 0.67 on the independent test set, indicating accurate predictions and good generalizability. The final model was converted to a user-friendly web dashboard to facilitate its wide application by the scientific community. This ML-based QSAR model serves as a tool for predicting drug plasma half-lives and extra label withdrawal intervals in 6 common food animals based on physicochemical properties. It also provides a foundation to develop more advanced models to predict the tissue half-life of drugs in food animals (Wu et al., 2024).

Proof of principle case studies from the Food Standards Agency postdoctoral fellow and PhD student

177. As mentioned, the FSA and COT have been reviewing NAMs to scope the best scientific methodologies available to be used in risk assessment of chemicals in foods and the environment, and to understand how these can be incorporated and accepted in a regulatory context.

178. In 2021, the FSA started funding a 4-year computational toxicology postdoctoral fellow at the University of Birmingham and a three-year PhD Student (London Interdisciplinary Doctoral Program-LIDo-TOX AI) at King's College London.

179. The fellow and PhD student have been working alongside other government departments to understand how NAMs will improve indicative levels of safety in chemical risk assessment.

180. The fellowship and studentship also compliment the work set out in the [COT FSA UK NAMs Roadmap](#) towards using new approach methodologies in chemical risk assessment.

181. The following are the abstracts of published examples (link in title) of AI tools in safety assessment case studies that the FSA, postdoctoral fellow and PhD student have been working on as proof of principle.

[*Exploring dimensionality reduction techniques for deep learning driven QSAR models of mutagenicity \(Kalian et al, 2023\).*](#)

182. "Dimensionality reduction techniques are crucial for enabling deep learning driven quantitative structure-activity relationship (QSAR) models to navigate higher dimensional toxicological spaces, however the use of specific techniques is often arbitrary and poorly explored. Six dimensionality techniques (both linear and non-linear) were hence applied to a higher dimensionality mutagenicity dataset and compared in their ability to power a simple deep learning driven QSAR model, following grid searches for optimal hyperparameter values. It was found that comparatively simpler linear techniques, such as principal component analysis (PCA), were sufficient for enabling optimal QSAR model performances, which indicated that the original dataset was at least approximately linearly separable (in accordance with Cover's theorem). However certain non-linear techniques such as kernel PCA and autoencoders performed at closely comparable levels, while (especially in the case of autoencoders) being more widely applicable to potentially non-linearly separable datasets. Analysis of the chemical space, in terms of XLogP and molecular weight, uncovered that the vast majority of testing data occurred within the defined applicability domain, as well as that certain regions were

measurably more problematic and antagonised performances. It was however indicated that certain dimensionality reduction techniques were able to facilitate uniquely beneficial navigations of the chemical space.” (Kalian et al., 2023)

[Improving accuracy scores of neural network driven QSAR models of mutagenicity \(Kalian et al 2023\)](#)

183. “Multiple QSAR models of mutagenicity were created and compared, using knowledge graph approaches to train and test multi-layer perceptron classifiers, following dimensionality reduction from several thousand dimensions to hundreds of dimensions via principal component analysis. Such knowledge graphs were built in one case using molecular fingerprint based structural similarities, while in another case using molecular fragments found via application of the Girvan-Newman algorithm. A simple hybrid model was also explored. However, both competing QSAR models performed with comparable accuracies, with both sensitivity and specificity scores for each occurring within range of 70%. The predictions of both models were in agreement in an average of 71% of cases, meaning that each could offer a related yet notably different perspective of toxicological space; hence a simple hybrid model was trialed, which only output predictions agreed between both constituent models, which averaged at 78% accuracy.” (Kalian et al., 2023)

[Graph attention networks using knowledge graphs, for predicting novel points of departure for brominated flame retardants \(Kalian et al., 2024\)](#)

184. “Brominated Flame Retardants (BFRs) are present in everyday products and materials, to improve fire safety. Various studies have identified BFRs that are neurotoxic, teratogenic and reprotoxic in animals, yet other BFRs continue to lack relevant in-vivo Points of Departure (PODs). In-vivo studies pose ethical, scalability and validity concerns, which in-silico methods such as Quantitative Structure-Activity Relationship (QSAR) modelling may address. This study hence aims to develop a novel artificial intelligence driven QSAR model, using Graph Attention Networks (GATs) acting on Knowledge Graphs (KGs) of molecules, to predict new BFR PODs relevant to neurotoxicity, teratogenicity and reprotoxicity.

185. Datasets of PODs for each endpoint (over a consistent species, exposure route and POD type) were obtained via curation of the Toxicity Value Database (ToxValDB), each containing 532-2022 molecules. PODs included Median Effective Concentrations (EC50s) for water flea neurotoxicity, No Observed Adverse Effect Levels (NOAELs) for rat teratogenicity, while both No Observed Effect Concentrations (NOECs) and Lowest Observed Effect Concentrations (LOECs) for water flea reprotoxicity. Conflicting POD values for certain molecules, often varying by several orders of magnitude, were averaged over interquartile ranges. All PODs followed log-normal distributions and so were logarithmised to enforce normal distributions. Furthermore, 432 BFR descriptors were aggregated from relevant literature, with a majority absent from the endpoint-specific datasets.

186. KGs were created for each endpoint, encoding dataset chemicals and BFRs as nodes, with shared substructures as edges; substructures of molecular graphs were computed via the Girvan-Newman algorithm and then used in graph isomorphism searches. Node feature vectors were included, containing physicochemical metrics relevant to Lipinski's rule of 5. The QSAR Applicability Domain (AD) was uniquely defined as any molecule connectable to the KG, via the substructure search method. GATs were trained to perform node regression of log-POD values over each KG, implemented in PyTorch Geometric (Python 3), with a Mean Absolute Error (MAE) loss function. MAEs converged to minima on the testing data for all KGs, corresponding to uncertainties of 22%–51% on mean log-POD values for each endpoint. Exponentiation into PODs propagated these into error factors ranging from 3–16; nonetheless significantly lower than the variation of several orders of magnitude in the original PODs data.

187. Overall, the novel QSAR methodology explored was found to be an effective approach for predicting ranges for in-vivo PODs to occur within, over the relevant endpoints. Predicted ranges were sufficiently specific to aid in prioritisation of BFRs of greater concern, for future research. An open-source dataset of novel BFR POD predictions is planned, following enhancements to the model such as inclusion of edge features.” (Kalian et al., 2024)

Comparison of Optimised Geometric Deep Learning Architectures, over Varying Toxicological Assay Data Environments (Kalian et al., 2025)

188. “Geometric deep learning is an emerging technique in Artificial Intelligence (AI) driven cheminformatics, however the unique implications of different Graph Neural Network (GNN) architectures are poorly explored, for this space. This study compared performances of Graph Convolutional Networks (GCNs), Graph Attention Networks (GATs) and Graph Isomorphism Networks (GINs), applied to 7 different toxicological assay datasets of varying data abundance and endpoint, to perform binary classification of assay activation. Following pre-processing of molecular graphs, enforcement of class-balance and stratification of all datasets across 5 folds, Bayesian optimisations were carried out, for each GNN applied to each assay dataset (resulting in 21 unique Bayesian optimisations). Optimised GNNs performed at Area Under the Curve (AUC) scores ranging from 0.728-0.849 (averaged across all folds), naturally varying between specific assays and GNNs. GINs were found to consistently outperform GCNs and GATs, for the top 5 of 7 most data-abundant toxicological assays. GATs however significantly outperformed over the remaining 2 most data-scarce assays. This indicates that GINs are a more optimal architecture for data-abundant environments, whereas GATs are a more optimal architecture for data-scarce environments. Subsequent analysis of the explored higher-dimensional hyperparameter spaces, as well as optimised hyperparameter states, found that GCNs and GATs reached measurably closer optimised states with each other, compared to GINs, further indicating the unique nature of GINs as a GNN algorithm.” (Kalian et al., 2025).

A novel method to derive a human safety limit for PFOA by gene expression profiling and modelling (e Silva, A. D. C. et al., 2024)

189. “Perfluorooctanoic acid (PFOA) is a persistent environmental contaminant that can accumulate in the human body due to its long half-life. This substance has been associated with liver, pancreatic, testicular and breast cancers, liver steatosis and endocrine disruption. PFOA is a member of a large group of substances also known as “forever chemicals” and the vast majority of substances of this group lack toxicological data that would enable their effective risk assessment in terms of

human health hazards. This study aimed to derive a health-based guidance value for PFOA intake (ng/kg BW/day) from *in vitro* transcriptomics data. To this end, we developed an *in silico* workflow comprising five components: (i) sourcing *in vitro* hepatic transcriptomics concentration-response data; (ii) deriving molecular points of departure using BMDExpress3 and performing pathway analysis using gene set enrichment analysis (GSEA) to identify the most sensitive molecular pathways to PFOA exposure; (iii) estimating freely-dissolved PFOA concentrations *in vitro* using a mass balance model; (iv) estimating *in vivo* doses by reverse dosimetry using a PBK model for PFOA as part of a quantitative *in vitro* to *in vivo* extrapolation (QIVIVE) algorithm; and (v) calculating a tolerable daily intake (TDI) for PFOA. Fourteen percent of interrogated genes exhibited *in vitro* concentration-response relationships. GSEA pathway enrichment analysis revealed that “fatty acid metabolism” was the most sensitive pathway to PFOA exposure. *In vitro* free PFOA concentrations were calculated to be 2.9% of the nominal applied concentrations, and these free concentrations were input into the QIVIVE workflow. Exposure doses for a virtual population of 3,000 individuals were estimated, from which a TDI of 0.15 ng/kg BW/day for PFOA was calculated using the benchmark dose modelling software, PROAST. This TDI is comparable to previously published values of 1.16, 0.69, and 0.86 ng/kg BW/day by the European Food Safety Authority. In conclusion, this study demonstrates the combined utility of an “omics”-derived molecular point of departure and *in silico* QIVIVE workflow for setting health-based guidance values in anticipation of the acceptance of *in vitro* concentration-response molecular measurements in chemical risk assessment.” (e Silva A. D. C. et al., 2024).

[Predicting Organ-Specific Toxicity of Selective Androgen Receptor Modulators. using Transfer Learning on Graph Convolutional Networks \(Kalian et al., 2025\)](#)

190. “Novel Quantitative Structure-Activity Relationship (QSAR) models were constructed using Graph Convolutional Networks (GCNs), to predict Drug-Induced Liver Injury (DILI), Drug-Induced Renal Injury (DIRI) and Drug-Induced Cardiotoxicity (DICT) of Selective Androgen Receptor Modulators (SARMs), an emerging class of performance-enhancing drugs. Prior to training on DILI, DIRI and DICT datasets, the GCN QSAR models were first pre-trained on a variety of unrelated biomedical assay datasets, as an attempt to improve model performance via transfer learning. The

success of the transfer learning was mixed; model performances were measurably improved via pre-training on certain datasets, by statistically weak increases. The optimal final QSAR models achieved overall accuracy scores of 68% for DILI (no significant improvement via ensemble modelling), 76% for DIRI (improved to 77% via ensemble modelling) and 65% for DICT (improved to 67% via ensemble modelling). Application of the most optimal singular models to a dataset of 25 SARMs predicted that 21 of the 25 SARMs are either DILI-positive, DIRI-positive, or both – which raises concern, given the rising use of SARMs. All SARMs except for one were predicted as DICT-negative. A novel definition of the Applicability Domain (AD) was used, intended for close relevance to the models, via generating three-dimensional graph embeddings, for each model. Convex hulls were fitted around training data embeddings, with a $\pm 10\%$ buffer, defining the AD as the region of embedded chemical space covered by the convex hull, for a given model. Subsequent analysis found that a majority of DILI, DIRI and DICT testing data lay within the AD, alongside a majority of the SARMs, adding consensus to the reliability of the predictions.” (Kalian et al., 2025).

Challenges of AI integration in chemical risk assessment

191. There are multiples challenges with integrating AI into the current regulatory systems. Trust, data ecosystem, resources, data management, data bias, acquiring people talent (subject matter expertise) to name a few will be discussed below.

“The claim that "machines cannot make mistakes" seems a curious one.”

“The claim that a machine cannot be the subject of its own thought can of course only be answered if it can be shown that the machine has some thought with some subject matter. In this sort of sense, a machine undoubtedly can be its own subject matter. It may be used to help in making up its own programmes or to predict the effect of alterations in its own structure. By observing the result so fits own behaviour it can modify its own programmes as to achieve some purpose more effectively.”

-Alan Turing (1950)

Data Management Challenges

192. In 1998, John Massey stated to “Create, understand, store, move ... or else ...*Drown in Wave of Infrastructure Stress*.” He termed this InfraStress; which are the bad effects of faster change in computer subsystems and usage such as Central Point Units (CPUs), memory, disks, demand than in underlying infrastructure i.e. bandwidths, addressability & naming, scalability of interconnect (i.e. handling large data sets), operating systems, file systems, backup. Symptoms include bottlenecks, odd limits, workarounds, instability, unpredictability, nonlinear surprise, over-frequent releases, multiple versions, hardware obsolete before depreciated. In organizations that grow quickly, there will be stress on management and support infrastructure.

193. Therefore, magnitude of data and management will be one of the biggest challenges in the application of AI.

194. Within the data challenges, the data itself will pose ethical and regulatory hurdles. It has been stated that validation should have the post-validation process and application in mind (Bottini et al., 2008), *i.e.*, the regulatory context-of-use and societal needs for the method in question, and it must adhere to ethical standards as a part of confidence building (Hartung, 2024).

Validation

195. In 2025, Hartung and Kleinstreuer discussed the challenges for validation of AI-based new approach methods (Hartung and Kleinstreuer (2025)):

- **Bias mitigation:** AI models are susceptible to inheriting biases present in the data they are trained on. If the training data contains existing biases, the resulting models may perpetuate these biases, leading to inaccurate or unfair predictions.
- **Transparency:** The “black box” nature of complex AI models, particularly DL models, represents a significant obstacle to their wider acceptance and implementation. Since the inner workings of these models are not easily interpretable, it can be challenging to understand how they arrive at their pre-

dictions. This lack of transparency can lead to distrust, especially in high stakes domains like chemical risk assessment where understanding the reasoning behind a prediction is crucial for decision-making.

- **Balancing innovation and caution:** There is an ethical imperative to advance methods that could reduce animal testing and improve human health protection, but this must be balanced against the need for caution in implementing new, potentially less understood methods.
- **Equity and access:** Ensuring that the benefits of AI-based NAMs are equitably distributed and accessible to researchers and regulators globally, not just in well-resourced institutions, is an important ethical and management consideration.
- **Human oversight:** While AI can automate many tasks, human oversight remains critical in AI-driven toxicology. AI systems should be designed to augment human capabilities, not to re-place human judgment and expertise.

E validation

196. E-validation (Hartung and Kleinstainer, 2025) proposes a five-pronged approach, integrating distinct AI/ML learning modules to streamline and enhance the validation process (**Figure 10**):

- **Smart selection of reference chemicals:** This component utilizes AI, specifically advanced clustering algorithms, to identify the most informative chemicals for testing NAMs.
- **Simulating validation studies:** E-validation leverages the power of simulation to predict potential study outcomes before conducting resource intensive physical experiments.

- **Mechanistic validation:** This component emphasizes moving beyond mere correlations between NAM results and those from traditional methods.
- **AI-enhanced training module:** Recognizing the need for widespread adoption of NAMs, e-validation proposes an AI-powered training module to educate researchers and regulators on these innovative methods.
- **Centralized dashboard interface:** To seamlessly integrate all modules and streamline workflows, e-validation proposes a centralized dashboard interface.

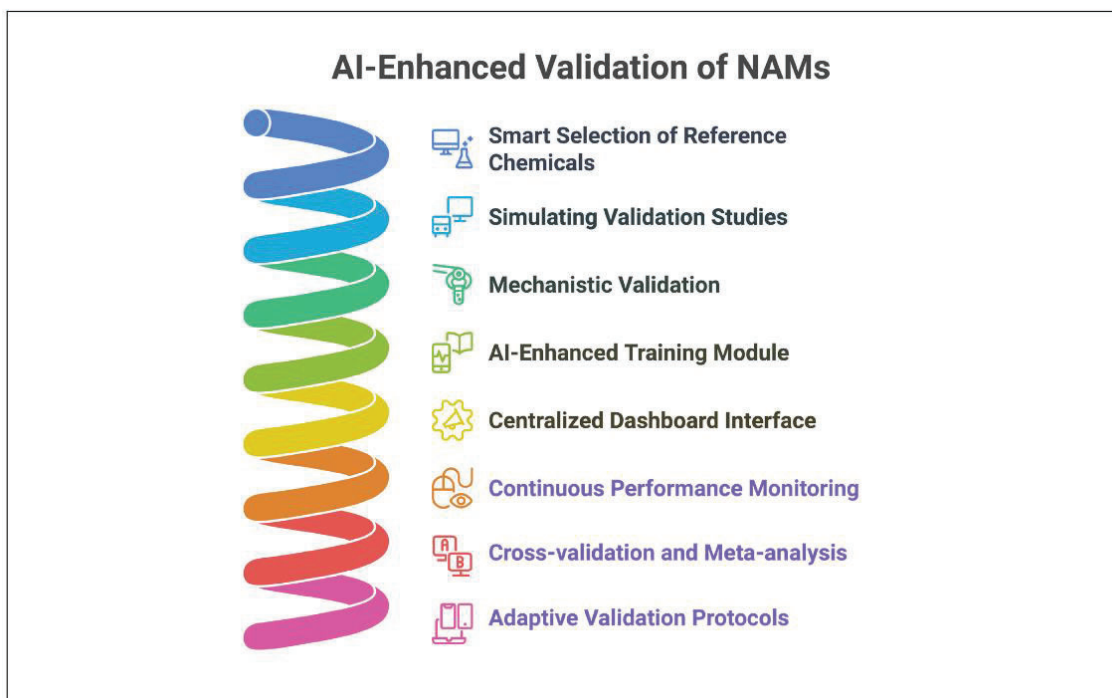


Figure 10. How e-validation concepts can be adapted for AI-based NAMs. The basic ideas of e-validation apply across both AI-based and experimental NAMs. They comprise: smart selection of reference chemicals (identify the most informative and diverse set of reference chemicals) to ensure comprehensive coverage of chemical space and toxicological mechanisms; simulating validation studies (this could involve: (a) predicting potential outcomes of validation experiments, (b) optimizing experimental designs and parameters, (c) identifying potential bottlenecks or issues in the validation process); mechanistic validation (analyse scientific literature and databases to establish mechanistic links between predicted toxicological effects and

known biological pathways for biological plausibility); AI-enhanced training module (AI-powered training systems for researchers and regulators on the proper use and interpretation of NAMs to personalize learning experiences and provide real-time feedback on the application of these methods), and centralized dashboard interface (AI-driven dashboard that integrates (a) real-time monitoring of validation progress, (b) automated data quality checks, (c) performance metrics visualization, and (d) uncertainty quantification displays. New elements for AI-based NAMs (in blue) include continuous performance monitoring (continuously monitor the performance of validated AI-based NAMs as new data becomes available to ensure ongoing relevance and reliability of the methods); cross-validation and meta-analysis (extensive cross-validation of AI-based NAMs across different datasets and to conduct meta-analyses of multiple validation studies, providing a more comprehensive assessment); and adaptive validation protocols (dynamically adjust validation protocols based on ongoing results to ensure that the validation process remains relevant and efficient as the AI-based NAM evolves with uncertainty quantification enhancing the interpretability and reliability). Figure taken from (Hartung and Kleinstainer, 2025).

Data Ownership

197. One the legal challenges is who owns the data once it has been created. Should a regulator build its own AI entity from scratch, or use a model or engine supplied by a supplier? Regulatory agencies may own some data but must obtain proper licenses and consent to use third-party or public data for training. There are also legal ethical reasons as mentioned which helps prevent the misuse of data and the unauthorized use of materials in AI training and outputs. Data ownership ensures control over data integrity, crucial for accurate AI model performance.

Bias

198. Risk of bias is a critical factor influencing the reliability and validity of toxicological studies, impacting evidence synthesis and decision-making in regulatory and public health contexts.

199. The integration of AI-based tools into risk of bias assessments can significantly improve the efficiency, consistency, and accuracy of evaluations. However, AI models are themselves susceptible to algorithmic and data biases, necessitating robust validation and transparency in their development.

200. In the Hartung et al 2025 paper they list the empirical patterns of bias direction which has important implications for AI-based approaches:

- **Training data considerations:** meta-research findings can inform the development of AI models by providing empirical priors about the expected direction and magnitude of bias under different methodological conditions. This knowledge can help calibrate ML models and improve their predictive accuracy.
- **Novel pattern detection:** AI systems might identify previously unknown combinations of study characteristics that predict specific directions of bias. For example, ML analyses of large datasets might reveal interaction effects between different sources of bias that are not apparent in traditional meta-analyses.
- **Validation metrics:** when developing AI tools for bias assessment, performance metrics should consider not only the detection of bias presence but also the accuracy in predicting bias direction and magnitude. This requires careful consideration of ground truth data and validation approaches, but there might not be enough studies to provide this grounding. We have recently suggested that AI tools are used to model validation studies (Hartung et al. 2024), which might help here.
- **Context-specific adjustments:** the empirical evidence suggests that bias patterns may vary across different fields of toxicology and types of outcomes. AI systems should be capable of adjusting their assessments based on these

context-specific patterns, rather than applying one-size-fits-all corrections.

201. The types of bias in AI-based toxicology tools are listed (Hartung et al., 2024) and include data bias, sampling bias, measurement bias, and historical bias (**Figure 11**).

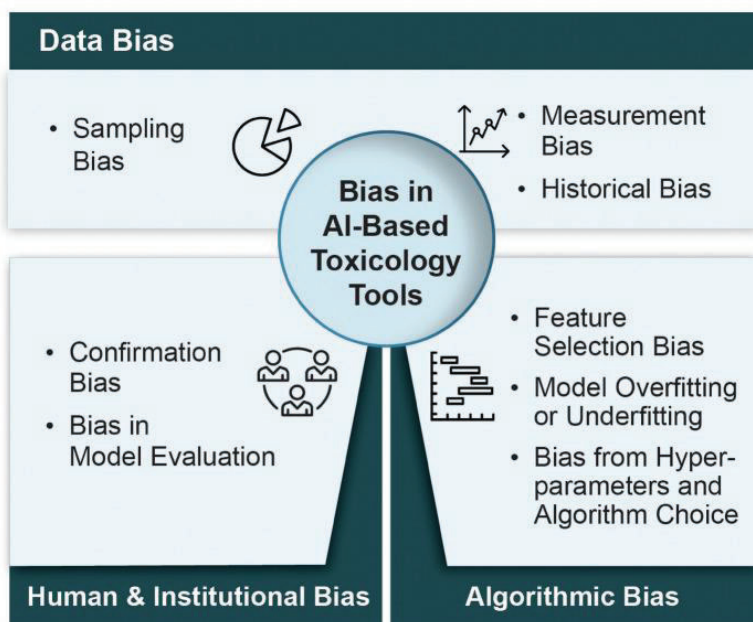


Figure 11. Sources of bias in AI-based toxicology tools. (Figure taken from Hartung et al., 2024).

202. When training and designing the models the AI or ML algorithms themselves introduces bias:

- **Model overfitting or underfitting:** occurs when an AI model is too closely tailored to the training data (overfitting) or when it fails to capture the underlying patterns due to excessive simplification (underfitting). This can result in biased predictions when the model is applied to new data.
- **Feature selection bias:** bias can be introduced through the selection of features used to train the model. If the selected features are not adequately validated or are inherently biased (e.g., reliance on surrogate endpoints), the model's outputs may reflect these biases.

- **Bias from hyperparameters and algorithm choice:** the selection of model hyperparameters (i.e. a parameter that can be set in order to define any configurable part of a model's learning process) and algorithm types can also influence the bias of the model. Certain algorithms may be more prone to specific biases (e.g., decision trees being sensitive to imbalanced data), leading to skewed outputs.

203. They also discuss the human and institutional bias which can distort results:

- **Confirmation bias:** human biases can be encoded into AI models through choices made during data selection, feature engineering, or interpretation of results. For example, researchers may select specific datasets or features that align with preconceived hypotheses, inadvertently introducing confirmation bias.
- **Bias in model evaluation and validation:** the metrics used to evaluate and validate AI models (e.g., accuracy, sensitivity, and specificity) can influence how model performance is perceived. If the validation metrics are not appropriately chosen to address potential biases, the resulting models may appear robust but perform poorly in real-world applications.

204. Most importantly the bias can lead to mistakes in predictions, incorrect labelling of descriptors and inherently not a comprehensive relevant and reliable risk assessment. This in turn undermines the trust and credibility of these models in research and regulatory settings.

Hallucinations

205. Hallucinations are outputs from an AI model that, while appearing plausible or authoritative, are factually inaccurate, nonsensical, or not supported by the training data.

206. These occur for a number of reasons including gaps in training data, model overconfidence, ambiguous prompts, and a lack of grounding in real-world chemical data.

207. These hallucinations could lead to an incorrect risk assessment conclusion which would compromise safety and in turn create distrust in the technological model.

208. These can be mitigated by human oversight which requires domain expertise, structured knowledge integration, clearer prompting and robust data verification.

Frameworks

209. From the JRC 2018 workshop Chemical Risk Assessment proposed a framework to adopt AI in chemical risk assessment (JRC, 2019). The framework includes the following criteria to use AI in Chemical Safety Risk Assessment:

- **Data Availability and Quality:** AI can be employed to analyse large datasets for chemical safety risk assessment, provided that high-quality data is available. This includes data on chemical properties, toxicity, exposure levels, and environmental impact
- **Interpretability and Explainability:** AI algorithms can be developed to provide interpretable and explainable results in chemical safety risk assessment. This involves designing models that offer insights into how predictions are made, allowing for transparency and understanding of the decision-making process
- **Regulatory Compliance:** AI technologies can assist in ensuring regulatory compliance by automating processes related to data management, risk assessment, and reporting. This includes aligning AI-driven assessments with regulatory guidelines such as REACH and TSCA.
- **Uncertainty and Confidence Estimation:** AI models can be enhanced to quantify uncertainty and provide confidence estimates for their predictions. This involves implementing techniques for uncertainty quantification, such as probabilistic modelling and sensitivity analysis, to assess the reliability of model outputs

- **Generalisation and Transferability:** AI algorithms can be designed to generalise across diverse chemical compounds and contexts. This includes developing models that can transfer knowledge and insights learned from one dataset to another, improving their applicability and utility in different scenarios
- **Ethical and Societal Implications:** AI technologies must address ethical and societal concerns related to data privacy, algorithmic bias, and transparency. This involves integrating principles of fairness, accountability, and transparency into the design and deployment of AI systems to ensure responsible and ethical use in chemical safety risk assessment.

210. These parameters outline the potential applications and considerations for utilising AI in chemical safety risk assessment, highlighting the importance of data quality, interpretability, regulatory compliance, uncertainty quantification, generalisation, and ethical considerations in the development and implementation of AI-driven solutions.

211. Overall, the workshop participants highlighted the potential of AI to revolutionise chemical risk assessment by addressing existing challenges, enhancing decision-making processes, and fostering collaboration across disciplines and sectors.

ML tools for toxicity prediction

212. A recent perspective (Seal et al., 2025), it emphasizes the need to enhance the understanding and application of ML models in drug discovery, focusing on well-defined data sets for toxicity prediction based on small molecule structures. We focus on five crucial pillars for success with ML-driven molecular property and toxicity prediction: (1) data set selection, (2) structural representations, (3) model algorithm, (4) model validation, and (5) translation of predictions to decision-making (**Figure 12**).

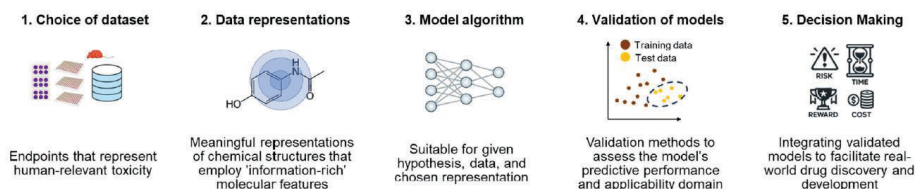


Figure 12. Five critical pillars deserving attention from researchers using ML tools for toxicity prediction.

1. Selecting appropriate data sets that accurately represent the toxicity of interest ensures that model predictions are relevant.
2. Chemical structures must be encoded into relevant representations that capture essential molecular 'information' to generate ML-ready features.
3. Model algorithms must be suitable to learn the signal in the data and representation characteristics mentioned above.
4. Models must be validated to assess their predictive performance, both retrospectively and prospectively, within their applicability domain.
5. In practical model applications, it is crucial to consider project scenarios and desired outcomes to facilitate real world drug discovery and development.

Data

213. AI-driven chemical risk assessment models rely on significant computing power to process vast datasets, enabling faster and more accurate predictions of chemical toxicity. This power allows for the application of complex algorithms like deep neural networks to identify patterns, generate synthetic data, and perform high-throughput screening, which helps reduce animal testing and accelerate the discovery of mechanisms for chemical-induced harm. While the computational demand is high, it facilitates more predictive, efficient, and comprehensive chemical risk assessment.

214. The summation of **all** the world data, whether it is created, captured, or replicated, is called the Global Datasphere, and it is experiencing tremendous growth. International Data Corporation (IDC) predicted that the Global Datasphere

will grow from 45Zettabytes (ZB) in 2019 to 175 ZB by 2025 (Rydning et al., 2018).

Big Data

215. The term Big Data was made famous by John Massey in his [1998 presentation “Big Data...and the Next wave of InfraStress”](#) by explaining the technology waves of data that will ensue and the challenges if we don’t keep up with it.

216. Big data usually includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process the data within a tolerable elapsed time (Snijders et al., 2012). Big data primarily refers to data sets that are too large or complex to be dealt with by traditional data-processing software. Current usage of the term big data tends to refer to the use of predictive analytics, user behaviour analytics, or certain other advanced data analytics methods that extract value from big data.

The V's of big data

217. The "V's of Big Data" are volume, velocity, variety, veracity and value. These are the five characteristics that make big data unique from other kinds of data and these are the attributes that explain how big data differs from traditional datasets (**Figure 13**).

Volume: The quantity of generated and stored data.

Velocity: The speed at which the data is generated and processed.

Variety: The diversity and the many different formats that big data can take.

Value: The worth in information that can be achieved by the processing and analysis of large datasets.

Veracity: Veracity refers to the accuracy and reliability of data.

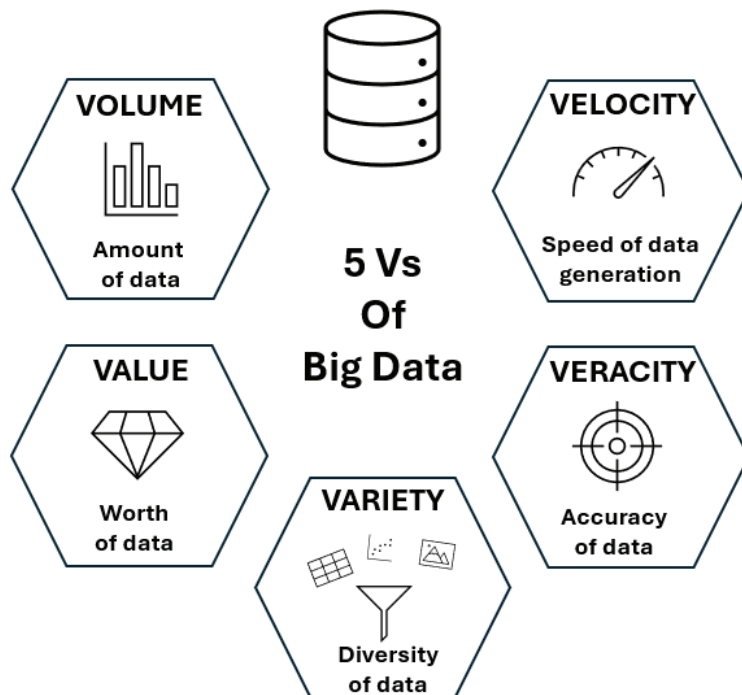


Figure 13. The 5 "V's of Big Data": volume, velocity, variety, veracity and value.

Data capabilities

Computational capacity

218. In 1949 Turing wrote: "Many of these limitations are associated with the very small storage capacity of most machines"; "It is probably not necessary to increase the speed of operations of the machines at all. Parts of modern machines which can be regarded as analogues of nerve cells work about a thousand times faster than the latter. This should provide a "margin of safety" which could cover losses of speed".

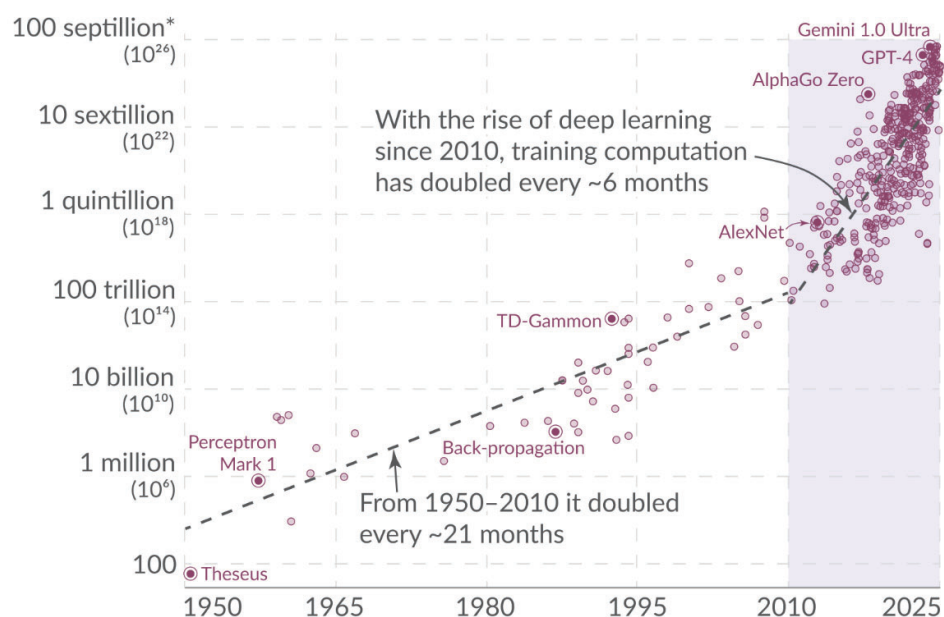
219. Since 2010, the training computation of notable AI systems has doubled every six months (**Figure 14**). We are at a point in technology that the computational capacity of the fastest supercomputers is at 1 billion giga Floating point operations per second (FLOPS) Dongarra et al. (2024). Training computation is predominantly measured using floating-point operations or "FLOP". One FLOP represents a single

arithmetic operation involving floating-point numbers, such as addition, subtraction, multiplication, or division.

The computation used to train notable AI systems has doubled every ~6 months since 2010

Our World
in Data

Training computation is measured in total floating-point operations (FLOP). Each FLOP represents a single arithmetic calculation, such as multiplication. Shown on a logarithmic scale.



*For comparison, 1 septillion (1,000,000,000,000,000,000,000,000) is the estimated number of stars in the universe.

Data source: Epoch (2024)

CC BY

Figure 14. Graph showing computation used to train notable AI systems has doubled every 6 months since 2010.

220. The development and use of AI/ML methods has evolved rapidly in the last decade, particularly due to an increase in available data (Pawar et al., 2019) and computational capacity (OpenAI, 2018).

221. Therefore, the available access to a vast wealth of available data and knowledge is changing the shape of toxicity risk assessment. During the same timeframe, the application of deep learning techniques demonstrated the potential to

create highly performant models but at the cost of transparency and explainability (Figure 15).

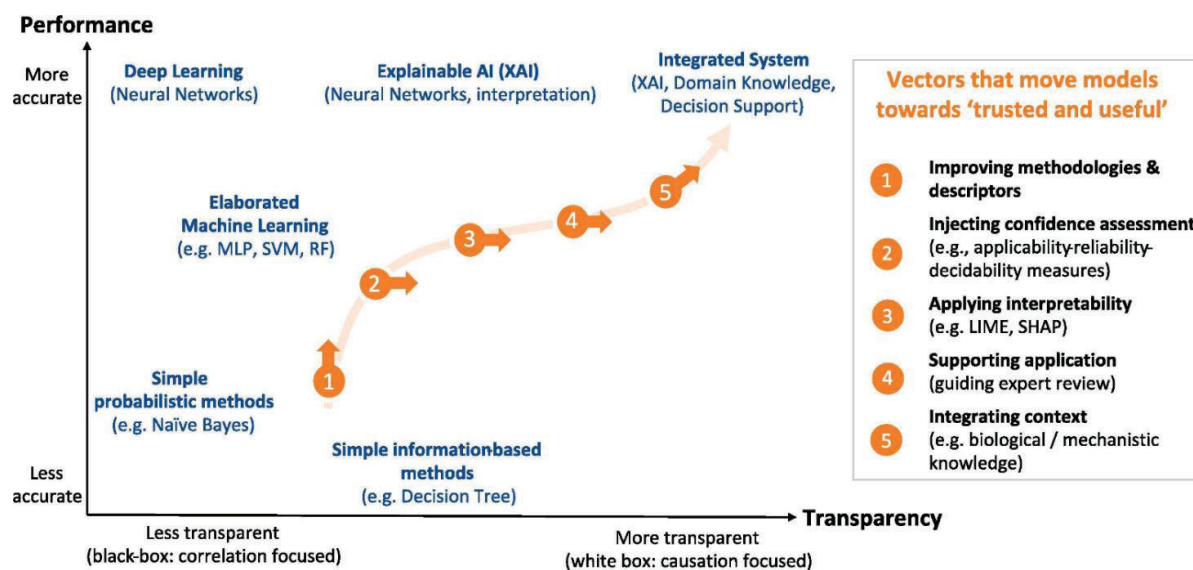


Figure 15. Performance vs Transparency. Machine-learned models for chemical toxicity are becoming more performant and trustworthy over time (orange arrow). Progress is dependent upon data quality and quantity, on modelling methodologies and more recently, upon the ability to integrate domain knowledge into the model (Figure taken from Barber et al., 2024).

222. New technologies must be accepted and integrated because regulations, infrastructure, user practices and maintenance networks are aligned to the existing technology (Geels, 2002). For these new technologies to progress, users have to integrate them into their practices, organisations and routines. This involves learning/training and various adjustments to infrastructures (Lie and Sørensen, 1996). It is known that links between technical and social elements provide stability leading to sociotechnical change.

Emerging & Future Quantum AI:

223. Quantum AI integrates the immense processing power of quantum computing with the advanced algorithms of AI, enabling AI to solve complex problems faster and more efficiently than classical computers (Devadas and Sowmya, 2025). Unlike

classical computers that use bits (0 or 1), quantum computers use qubits, which can be 0, 1, or both simultaneously (superposition).

224. This allows them to perform vast numbers of calculations in parallel, exponentially increasing speed. However, this still at an early nascent stage; significant challenges remain in developing stable and scalable quantum hardware, reducing error rates, and creating new quantum AI frameworks.

Data Integrity

225. Data integrity is the maintenance of, and the assurance of, data accuracy and consistency over its entire life-cycle (Sandhu, 1993). The overall intent of any data integrity technique is the same: ensure data is recorded exactly as intended (such as a database correctly rejecting mutually exclusive possibilities). Moreover, upon later retrieval, ensure the data is the same as when it was originally recorded. In short, data integrity aims to prevent unintentional changes to information.

226. Data integrity is a fundamental aspect of storage security and reliability (Sivathanu et al., 2005). Data integrity is also critical to regulatory compliance, and the fundamental reason for 21 Code of Federal Regulations (CFR) Part 11 published by the U.S. Food and Drug Administration (FDA). FDA published the first guideline in 1963, and since then FDA and European Union (EU) have published numerous guidelines on various topics related to data integrity for the pharmaceutical industry. Regulators wanted to make certain that industry capture accurate data during the drug development lifecycle and through commercialization.

227. In a recent short perspective, they proposed 3 pillars present they believe are three chemical data and quality pillars that are essential to the continued growth and scientific impact of the cheminformatics field to ensure public access and integrity of chemical databases powering cheminformatics (Williams and Richard 2025).

- **Pillar 1:** Government funding and public support for structure-indexed, searchable, downloadable chemical databases.
- **Pillar 2:** clear data licensing, provenance, and the need for FAIRness.

- **Pillar 3:** Coordinated community approaches regarding structure formats, ontologies, and quality curation procedures to ensure accurate association of chemical substances with associated identifiers, including structures, chemical names, and CAS Registry Numbers® (CAS RNs).

228. In terms of regulatory data integration one of the challenges is structural complexity. Unlike a single organisation, the regulatory field is a constellation of institutions, each with its own priorities and objectives. Achieving alignment and sharing data across departments will be a challenge especially with data security measures.

COT and FSA on data

229. When the COT and FSA organized three related workshops: [Exploring Dose Response](#); [PBPK Regulatory Workshop](#); and [Paving the way for the UK Roadmap: Development Validation and Regulatory Acceptance of New Approach Methodologies](#) they published and discussed new insights into how to translate new scientific technologies into future regulatory implementation. One of the areas was data and compliance of new technologies (Osborne et al., 2024). In it, they recommended that data integrity and capacity will be key in the process of integrating these new methodologies into risk assessment (Osborne et al., 2024). In order to manage areas such as “big data”, we will need to enhance our technological capabilities and capacity to support supercomputers and to develop suitable statistical techniques and software to interrogate and visualize such data (Osborne et al., 2024). In addition, we will need to have recognized established data integrity systems (guidelines) much like findability, accessibility, interoperability, and reusability (FAIR) and the Mutual Acceptance of Data (MAD) (Osborne et al., 2024).

AI Worldwide dimensions and initiatives

230. There are multiple international initiatives are working to define the dimensions, develop and deploy AI responsibly. These dimensions address both ethical principles and practical implementation. Below are a few examples.

OECD AI Principles (2021)

231. The OECD AI Principles are the first intergovernmental standard on AI. They promote innovative, trustworthy AI that respects human rights and democratic values. Adopted in 2019 and updated in 2024, they are composed of five values-based principles and five recommendations that provide practical and flexible guidance for policymakers and AI actors.

232. By May 2023, governments reported over 1000 policy initiatives across more than 70 jurisdictions in the OECD which follow the OECD AI Principles.

233. As of 2024, 47 countries, including all OECD members and the European Union, have adhered to these principles, which are widely used to guide national AI strategies and regulatory frameworks. They are centred around five core values:

- Inclusive growth, sustainable development, and well-being.
- Human rights and democratic values, including fairness and privacy.
- Transparency and explainability.
- Robustness, security and safety.
- Accountability.

234. They also list some recommendations for policy makers:

- Investing in AI research and development.
- Fostering an inclusive AI-enabling ecosystem.
- Shaping an enabling interoperable governance and policy environment for AI.
- Building human capacity and preparing for labour market transformation.
- International co-operation for trustworthy AI.

EFSA Roadmap for actions on artificial intelligence for evidence management in risk assessment (2022)

235. In May 2021, [the European Food Safety Authority \(EFSA\) launched the “Roadmap for actions on Artificial Intelligence for evidence management in risk assessment”](#) project to develop an approach for the implementation of AI methods in the evidence management phase of its internal risk assessment process. The main

objective is to identify specific projects to be carried out to increase by 2027; the accessibility and breadth of the body of evidence and enhance the trustworthiness of the risk assessment process by applying human-centric AI in close coexistence with human expertise. The roadmap presents a full understanding of ongoing activities, market readiness, knowledge gaps, and societal interests and concerns, as well as collaboration opportunities in the field of AI application to the evidence management process. Subsequently, based on these findings, it outlines a set of recommendations aimed at developing the EFSA Agency's capabilities to adopt and integrate AI solutions in the evidence management process. Finally, the roadmap contains considerations aimed at identifying potential communication and engagement opportunities in the area of AI applied to evidence management in risk assessment, considering specific needs and concerns of EFSA's stakeholders on the adoption of AI.

FDA Guidance-Considerations for the Use of Artificial Intelligence To Support Regulatory Decision-Making for Drug and Biological Products (2025)

236. [Considerations for the Use of Artificial Intelligence To Support Regulatory Decision-Making for Drug and Biological Products](#) guidance provides recommendations to sponsors and other interested parties on the use of AI to produce information or data intended to support regulatory decision-making regarding safety, effectiveness, or quality for drugs. Specifically, this guidance provides a risk-based credibility assessment framework that may be used for establishing and evaluating the credibility of an AI model for a particular context of use (COU).

237. The risk-based credibility assessment framework is a 7-step process:

- **Step 1:** Define the question of interest that will be addressed by the AI model.
- **Step 2:** Define the COU for the AI model.
- **Step 3:** Assess the AI model risk.
- **Step 4:** Develop a plan to establish the credibility of AI model output within the COU.
- **Step 5:** Execute the plan.

- **Step 6:** Document the results of the credibility assessment plan and discuss deviations from the plan.
- **Step 7:** Determine the adequacy of the AI model for the COU.

ECETOC Integrating AI into chemical safety assessment Workshop (2024)

238. The workshop took place in Sophia Antipolis, France, on October 16-17, 2024. The workshop brought together leading experts to explore the latest advancements in AI technologies and their applications in evaluating chemical risks. Discussions focused on how AI enhances data processing, improves predictive analysis, and increases the efficiency of chemical safety assessments. The [flash report](#) highlights key findings from the European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) workshop on AI in chemical safety assessment. The report summarizes key insights, including opportunities and challenges in leveraging AI for more accurate and comprehensive risk evaluation.

239. The key take home messages were:

- **AI in Chemical Safety:** AI is transforming chemical safety assessment, complementing human expertise while still relying on expert judgment.
- **Data is Key:** High-quality, accessible data in harmonised formats, along with transparent data practices, are essential for AI success. Tools like the IUCLID software, maintained by the European Chemicals Agency (ECHA), could become even more impactful if adapted to FAIR principles
- **Building Trust in AI:** Standardised reporting, validation frameworks, and accessible tools are critical to overcoming trust barriers.
- **Emerging Priorities:** Key focus areas include assessing mixtures, data read-across, exposure, and data quality evaluation transforming chemical safety assessment, complementing human expertise while still relying on expert judgment.

HESI Building a roadmap for AI-enabled human and environmental health protection Workshop (2025)

240. In June 2025, the Health and Environmental Sciences Institute (HESI Global) convened a two-day in-person workshop ([Building a Roadmap for AI-Enabled Human and Environmental Health Protection](#)) in Washington, DC, to explore the strategic use of AI in advancing human and environmental health protection within a One Health framework. The workshop gathered over 150 participants from more than 10 countries, representing a wide range of sectors including academia, government, industry, and nonprofit organizations. A novel element of the workshop was the deployment of a customized generative AI model (Workshop GPT), designed to synthesize live inputs and identify key thematic focus areas, referred to as Guiding Pillars, that serve as strategic destinations on the path toward AI-enabled health and environmental systems. Drawing from expert presentations, moderated discussions, and real-time polling, the Workshop GPT produced five Guiding Pillars: Connected Data; Trust and Validation; Real-World Use Cases; Capacity Building; and Responsible Design. These Guiding Pillars define critical areas for the practical and responsible implementation of AI across biological, ecological, and regulatory domains. The [meeting report](#) presents these outputs, combining AI generated content with extensive review and refinement by the HESI Global scientific team.

UK Government reports, strategies, guidance, white papers and initiatives

241. There are various UK government documents related to AI. Below are some of the key strategies, reports, white papers and initiatives.

Guidance-National AI Strategy (2021)

242. The [National AI Strategy](#) builds on the UK's strengths but also represents the start of a step-change for AI in the UK, recognising the power of AI to increase resilience, productivity, growth and innovation across the private and public sectors.

243. The UK's National AI Strategy aims were to:

- Invest and plan for the long-term needs of the AI ecosystem to continue our leadership as a science and AI superpower.
- Support the transition to an AI-enabled economy, capturing the benefits of innovation in the UK, and ensuring AI benefits all sectors and regions
- Ensure the UK gets the national and international governance of AI technologies right to encourage innovation, investment, and protect the public and our fundamental values.

OPPS Study on the Impact of Artificial Intelligence on Product Safety (2021)

244. The [Study on the Impact of Artificial Intelligence on Product Safety](#) was commissioned by the Office for Product Safety and Standards (OPSS), part of the Department for Business, Energy and Industrial Strategy (BEIS), and carried out by the Centre for Strategy and Evaluation Services (CSES) between January and June 2021.

245. The objective of this study was to examine the current and forecasted future impacts of AI in consumer products, and what this means for product safety. This breaks down into the following three specific objectives:

- **Objective 1:** Analyse the current and likely future applications of AI in the home, highlighting the advantages and disadvantages for consumers and product safety implications / risks.
- **Objective 2:** Assess whether the current product safety framework is sufficient for a new generation of products that incorporate AI.
- **Objective 3:** Examine what factors Regulators should consider when responding to these new challenges to ensure consumer safety and foster product innovation.

AI Safety Summit (2023)

246. [The AI Safety Summit](#) was a major global event that took place on the 1st and 2nd November 2023 at Bletchley Park, Buckinghamshire. This first global AI Safety

Summit sparked a 'Bletchley Effect' as countries build new AI institutions and make international agreements to cooperate on research and development.

247. The summit brought together international governments, leading AI companies, civil society groups and experts in research with the aims considering the risks of AI, especially at the frontier of development and to discuss how they can be mitigated through internationally coordinated action.

248. The 5 objectives discussed at the summit were:

- a shared understanding of the risks posed by frontier AI and the need for action.
- a forward process for international collaboration on frontier AI safety, including how best to support national and international frameworks.
- appropriate measures which individual organisations should take to increase frontier AI safety.
- areas for potential collaboration on AI safety research, including evaluating model capabilities and the development of new standards to support governance.
- showcase how ensuring the safe development of AI will enable AI to be used for good globally.

National Audit Office Use of artificial intelligence in government Report – Value for money (2024)

249. The [Use of artificial intelligence in government](#) report considers how effectively the government has set itself up to maximise the opportunities and mitigate the risks of AI in providing public services.

250. The primary focus of the report is the role of the Cabinet Office and DSIT in supporting the adoption of AI in the public sector.

251. Specifically, the report looks at:

- **Part One:** the government's strategy and governance for AI use in public services.
- **Part Two:** how government bodies are using AI and how government understands the opportunities.
- **Part Three:** central government's plans for supporting the testing, piloting and scaling of AI; and progress in addressing barriers to AI adoption.

252. The conclusions from the report were that AI presents the government with opportunities to transform public services. The centre of government has identified the potential for large-scale productivity gains from the adoption of AI across the public sector.

253. Responsibility for AI rests with DSIT and the Cabinet Office and, while the government is working on a draft strategy for AI adoption in the public sector, it has not yet finalised it or published an implementation plan.

254. The survey of government bodies found that AI was not yet widely used across government, but 70% of respondents were piloting and planning AI use cases.

255. Government departments are required to create AI adoption plans by June 2024.

256. There are risks to value for money if the government does not establish which department has overall ownership and accountability for delivery of the strategy for AI adoption in the public sector and set out appropriate roles and responsibilities for those who need to contribute.

257. Achieving large-scale benefits is likely to require not just adoption of new technology but significant changes in business processes and corresponding workforce changes.

258. To deliver the transformational benefits of AI, the government needs to ensure its overall programme for AI adoption is ambitious and supported by a realistic plan for the skills, funding and wider enablers needed.

259. The government must also maintain focus on addressing other fundamental barriers to AI adoption, such as legacy systems, and data access and sharing, which will otherwise limit the extent to which it can exploit the future potential of AI.

AI regulation: a pro-innovation approach policy paper (2023-2024)

260. The [white paper detailed plans for implementing a pro-innovation approach to AI regulation](#). This white paper detailed plans for implementing a pro-innovation approach to AI regulation. They also had a supporting consultation.

261. In the AI regulation white paper, it was proposed that a regulatory framework was needed to keep pace with a rapidly advancing technology. This would include five cross-sectoral principles for existing regulators to interpret and apply within their remits in order to drive safe, responsible AI innovation.

262. These principles should encompass:

- Safety, security and robustness.
- Appropriate transparency and explainability.
- Fairness.
- Accountability and governance.
- Contestability and redress.

263. The paper mapped out the regulation landscape showing the relationships between the government, regulators, industry, and the wider ecosystem.

MHRA Guidance-Machine learning medical devices: transparency principles (2024)

264. The Medicines and Healthcare products Regulatory Agency (MHRA), US Food and Drug Administration (FDA) and Health Canada have collaborated to identify five guiding principles for the development of predetermined change control

plans (PCCPs). These guiding principles for PCCPs aim to remove the regulatory burden for developers of machine-learning-enabled medical devices (MLMDs), enabling reallocation of resources to improve product performance for patients.

265. The MHRA published guidelines for communicating clear and relevant information about machine learning-enabled medical devices. Building on the [10 guiding principles for Good Machine Learning Practice](#), the five guiding principles for MLMD manufacturers outline that a PCCP must be:

- **Focused and Bounded:** Describing specific changes that a manufacturer intends to implement.
- **Risk-based:** The intent, design, and implementation of a PCCP are driven by a risk-based approach that adheres to the principles of risk management.
- **Evidence-based:** Demonstrating that benefits outweigh the risks throughout the product lifecycle.
- **Transparent:** Provide clear and appropriate information and detailed plans for ongoing transparency to all stakeholders, from patients to healthcare professionals.
- **Total Product Lifecycle Perspective:** Improve the quality and integrity of a PCCP by continually considering the perspectives of all stakeholders.

International Scientific Report on the Safety of Advanced AI: interim report (2024)

266. An interim report on the science of advanced AI safety was published in 2024 a key outcome of the groundbreaking AI Safety Summit held at Bletchley Park in November 2023. It is the publication of the first 'International Scientific Report on the Safety of Advanced AI' which has been produced by a diverse group of 75 AI experts contributed to this report, including an international Expert Advisory Panel nominated by 30 countries, the European Union (EU), and the United Nations (UN).

267. The following are some of the highlights of the executive summary:

- If properly governed, general-purpose AI can be applied to advance the public interest, potentially leading to enhanced wellbeing, more prosperity, and new

scientific discoveries. However, malfunctioning or maliciously used general-purpose AI can also cause harm, for instance through biased decisions in high-stakes settings or through scams, fake media, or privacy violations.

- As general-purpose AI capabilities continue to advance, risks such as large-scale labour market impacts, AI-enabled hacking or biological attacks, and society losing control over general-purpose AI could emerge, although the likelihood of these scenarios is debated among researchers. Different views on these risks often stem from differing expectations about the steps society will take to limit them, the effectiveness of those steps, and how rapidly general-purpose AI capabilities will be advanced.
- There is considerable uncertainty about the rate of future progress in general-purpose AI capabilities. Some experts think a slowdown of progress is by far most likely, while other experts think that extremely rapid progress is possible or likely.
- There are various technical methods to assess and reduce risks from general-purpose AI that developers can employ and regulators can require, but they all have limitations. For example, current techniques for explaining why general-purpose AI models produce any given output are severely limited.
- The future of general-purpose AI technology is uncertain, with a wide range of trajectories appearing possible even in the near future, including both very positive and very negative outcomes. But nothing about the future of AI is inevitable. It will be the decisions of societies and governments that will determine the future of AI.

AI Opportunities Action Plan- Independent report (2025)

268. The report covers 50 Recommendations for the government to capture the opportunities of AI to enhance growth and productivity and create tangible benefits for UK citizens. These include:

- grow the UK's AI sector
- drive adoption of AI across the economy to boost growth
- improve products and services

269. The UK's starting point makes this aspiration plausible but achieving it will require bold and visionary action.

270. The government must:

- **Invest in the foundations of AI- Lay the foundations to enable AI:** We need world-class computing and data infrastructure, access to talent and regulation.
- **Push hard on cross-economy AI adoption- Change lives by embracing AI:** The public sector should rapidly pilot and scale AI products and services and encourage the private sector to do the same. This will drive better experiences and outcomes for citizens and boost productivity.
- **Position the UK to be an AI maker, not an AI taker Secure our future with homegrown AI:** As the technology becomes more powerful, we should be the best state partner to those building frontier AI. The UK should aim to have true national champions at critical layers of the AI stack so that the UK benefits economically from AI advancement and has influence on future AI's values, safety and governance.

DSIT Research and analysis Frontier AI: capabilities and risks – discussion paper (2025)

271. A discussion paper on the capabilities of, and risks from, frontier AI. This report covers many risks, but we wish to emphasise that the overarching risk is a loss of trust in and trustworthiness of this technology which would permanently deny us and future generations its transformative positive benefits. In discussing the other risks, we do so in order to galvanize action to mitigate them, such that we can capture the full benefits of frontier AI.

AI Playbook for the UK Government (2025)

272. The [AI playbook](#) offers guidance on using AI safely, effectively and securely for civil servants and people working in government organisations.

273. The playbook builds on the white paper [A pro-innovation approach to AI regulation](#) that set out [5 principles](#) for regulators to inform AI development in government and public sector organisation. The playbook defines 10 common principles to guide the safe, responsible and effective use of AI in government organisations.

274. The principles are:

- **Principle 1:** You know what AI is and what its limitations are.
- **Principle 2:** You use AI lawfully, ethically and responsibly.
- **Principle 3:** You know how to use AI securely.
- **Principle 4:** You have meaningful human control at the right stage.
- **Principle 5:** You understand how to manage the AI life cycle.
- **Principle 6:** You use the right tool for the job.
- **Principle 7:** You are open and collaborative.
- **Principle 8:** You work with commercial colleagues from the start.
- **Principle 9:** You have the skills and expertise needed to implement and use AI.
- **Principle 10:** You use these principles alongside your organisation's policies and have the right assurance in place.

AI Opportunities Action Plan

275. The actions government for AI opportunities were proposed:

- **Invest in the foundations of AI:** We need world-class computing and data infrastructure, access to talent and regulation.
- **Push hard on cross-economy AI adoption:** The public sector should rapidly pilot and scale AI products and services and encourage the private sector to do the same. This will drive better experiences and outcomes for citizens and boost productivity.
- **Position the UK to be an AI maker, not an AI taker:** As the technology becomes more powerful, we should be the best state partner to those building frontier AI. The UK should aim to have true national champions at critical layers of the AI stack so that the UK benefits economically from AI advancement and has influence on future AI's values, safety and governance.

The Office for Artificial Intelligence

276. The Office for Artificial Intelligence was responsible for overseeing implementation of the National AI Strategy.

277. Its mission was to drive responsible and innovative uptake of AI technologies for the benefit of everyone in the UK. It did this by engaging with organisations and securing broad public trust and support, focused on 3 pillars:

- Investing in and planning for the long-term needs of the AI ecosystem to continue our leadership as a science and AI superpower.
- Supporting the transition to an AI-enabled economy, capturing the benefits of innovation in the UK, and ensuring AI benefits all sectors and regions.
- Ensuring the UK gets the national and international governance of AI technologies right, to encourage innovation, investment, and protect the public and our fundamental values.

The AI Safety Institute

278. The AI Safety Institute's mission is to minimise surprise to the UK and humanity from rapid and unexpected advances in AI. The has now been preceded by the AI Security Institute.

279. The Institute will adjust its activities within the scope of its headline mission to ensure maximum impact in a rapidly evolving field. It will initially perform 3 core functions:

- **Develop and conduct evaluations on advanced AI systems**, aiming to characterise safety-relevant capabilities, understand the safety and security of systems, and assess their societal impacts
- **Drive foundational AI safety research**, including through launching a range of exploratory research projects and convening external researchers
- **Facilitate information exchange**, including by establishing – on a voluntary basis and subject to existing privacy and data regulation – clear information-sharing channels between the Institute and other national and international actors, such as policymakers, international partners, private companies, academia, civil society, and the broader public.

AI Security Institute

280. The [AI Security Institute](#)'s mission is to equip governments with a scientific understanding of the risks posed by advanced AI. They conduct research and develop and test mitigations. They are also working with the wider research community, AI developers and other governments to affect how AI is developed and to shape global policymaking on this issue.

HSE's regulatory approach to Artificial Intelligence (AI) (2025)

281. The Health and Safety Executive (HSE) outlined its regulatory approach to AI. The HSE oversees AI in the design, manufacture, and supply of workplace machinery and equipment under the Product Safety regulatory framework. Its remit includes building safety; chemicals; and pesticides regulation where AI is involved.

AI and HSE's regulatory remit

282. HSE regulates AI in a way that aligns with [our mission and priorities](#).

HSE's role in regulating AI includes:

- regulating the use of AI where it impacts on health and safety in [workplaces where HSE is the enforcing authority](#)
- regulating the use of AI in design, manufacture and supply of workplace machinery, equipment and products for use in the workplace as a [Market Surveillance Authority](#) under the Product Safety regulatory framework.
- where AI impacts on HSE's role to protect people and places, including building safety, chemicals and pesticides regulation.

Developing HSE's regulatory approach to AI

283. The focus of the work HSE is doing is to continue to develop their regulatory approach to AI includes:

- co-ordinating work on AI, sharing knowledge and identifying key issues through an internal AI common interest group, bringing together colleagues from across HSE.
- working with government departments to shape the approach to AI regulation.
- supporting the standards making process, to establish benchmarks for AI interaction with machinery and functional safety by engaging with international standards organisations (BSI, IEC and ISO).
- establishing relationships with industry and academic stakeholders, to share knowledge and learning on AI use cases and the impact on health and safety.
- collaborating with other regulators, through forums including the AI Standards Forum for UK Regulators, Information Commissioners Office AI Regulators Forum and the United Kingdom Health and Safety Regulators Network Innovation Sub-Group, to encourage a consistent regulatory approach.
- identifying AI developments of interest to HSE through horizon scanning activities and monitoring AI developments in Great Britain and around the world, from a practical and regulatory perspective.

- building our capability and experience in AI across specialist and scientific areas of HSE and working with partners as appropriate.
- supporting research bids that align with HSE's [areas of research interest](#) (on GOV.UK) and help develop safe use of AI and the ability to regulate AI use.
- setting up and trialling of an [Industrial Safetytech Regulatory Sandbox](#) (on [discoveringsafety.com](#)) to explore practical barriers to adoption of Industrial Safetytech in construction and how to break them down.

AI in UK government departments Research Briefing (2025)

284. A [briefing paper](#) that gives some examples how AI is being used by UK government departments in various ways.

The UK's Modern Industrial Strategy (2025)

285. The [Modern Industrial Strategy](#) is a 10-year plan to increase business investment and grow the industries of the future in the UK. The Strategy will make it quicker and easier for business to invest and will provide the certainty and stability needed for long-term investment decisions. Under the Sector Plans they also published a Life Sciences Sector Plan and a Digital and Technologies Sector Plan which has aims and actions that cover AI and tool development.

Life Sciences Sector Plan

286. [The Life Sciences Sector Plan](#) sets out a vision and an action plan to drive growth, innovation, and better health outcomes.

287. They have under Action 2b: The Government will, by the end of 2025, publish an Alternative Methods Strategy to support the development, validation, and uptake of alternative models to reduce and, where possible, eliminate the use of animals, ensuring that the full suite of policy levers is deployed in addition to further investment in Research & Development.

Digital and Technologies Sector Plan

288. The [Digital and Technologies Sector Plan](#) sets out to drive and unlock growth in the technologies of the future as part of the UK's Modern Industrial Strategy which includes:

- Strengthening the UK's position as a global hub for AI R&D by committing £1 billion to scale up the UK's AI Research Resource capacity by at least 20 times by 2030, and up to a further £750 million to build a new supercomputer in Edinburgh.
- Accelerating AI-enabled scientific breakthroughs in targeted areas of UK strategic priority. The programme will support AI tool development, interdisciplinary research and strengthen the UK's AI capabilities in science.

Is Regulatory Science Ready for AI?

289. The Food Drug Administration (FDA) hosted the [14th Global Coalition for Regulatory Science Research \(GCRSR\)](#) with a theme on Digital Transformation in Regulatory Science with international participation. GCRSR comprises of 16 regulatory agencies from 11 countries and the European Union, and its mission focuses on adopting emerging technologies and big data science to improve regulatory science research on the safety and efficacy of foods and drugs. The event consisted of seven sessions, one of which was a dialog among panellists (most co-authors of the paper), engaging the audience in discussing the question: "Is Regulatory Science Ready for AI?". The workshop report concluded that AI adoption requires ongoing dialogue, adaptability, and AI-trained personnel to harness its potential for regulatory responsibilities in the evolving 21st-century landscape (Hartung et al., 2025). The FDA commissioner outlined the ten key principles for regulating AI (**Figure 16**) whilst fulfilling the FDA's dual role as a regulatory and public health agency in using AI to enhance patient care and public health (Warraich et al., 2025).

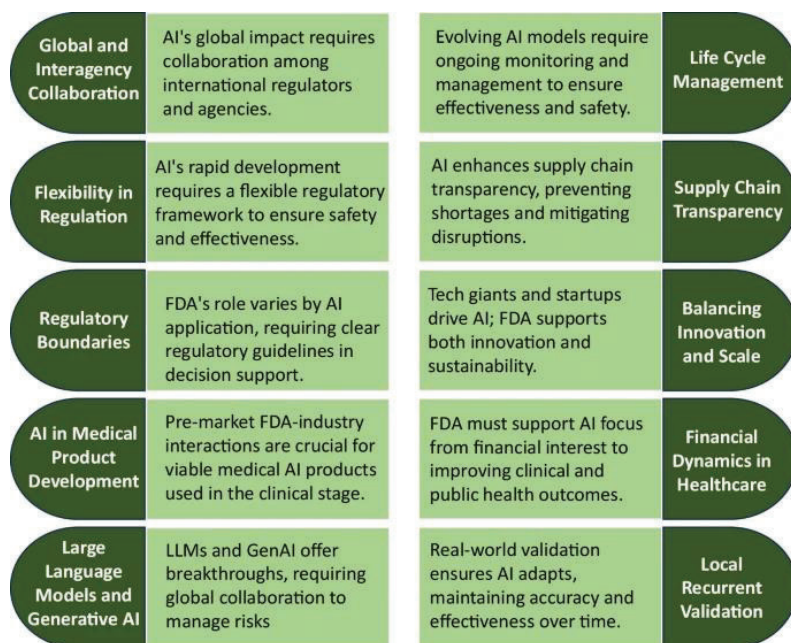


Figure 16. Ten key principles guide AI integration in regulatory science, underscoring the FDA's dual role in regulation and public health and emphasizing the need for adaptable policies to keep pace with rapid AI advancements. (Figure taken from Hartung et al., 2025)

290. From the [New Approach Methodologies \(NAMs\) to Support Regulatory Decisions for Chemical Safety | Published in FSA Research and Evidence](#) (2024) report it highlighted that no assessment of the regulatory readiness of these technologies or regulatory guidelines was found in the public domain.

291. From a data perspective, to integrate these AI technologies regulatory agencies will have to take into account a series of fundamentals with a flexible approach. Some of the fundamentals are the following (**Figure 17**):

Scalable: Leverage scalable infrastructure that can handle growing datasets and complex model deployments within the department's architecture.

Real Time Updates: Utilize a modern data architecture capable of handling the continuous flow of data from different data sources.

Cross Validation: Design validation methods, like cross-validation, to be modular and consistently applied to prevent data leakage and bias.

Storage: Data lifecycle management (DLM): Implement strategies like data tiering, archiving, and deletion to manage storage costs efficiently. This ensures that data is stored on the most cost-effective platforms based on its value and access frequency.

Computing Power: Keep up to date specialised hardware and servers to process vast data.

Collaborative Platforms: Access to data across different sources.

Data Governance: Create a framework that has policies and standards to maintain data secure, available and maintain integrity.

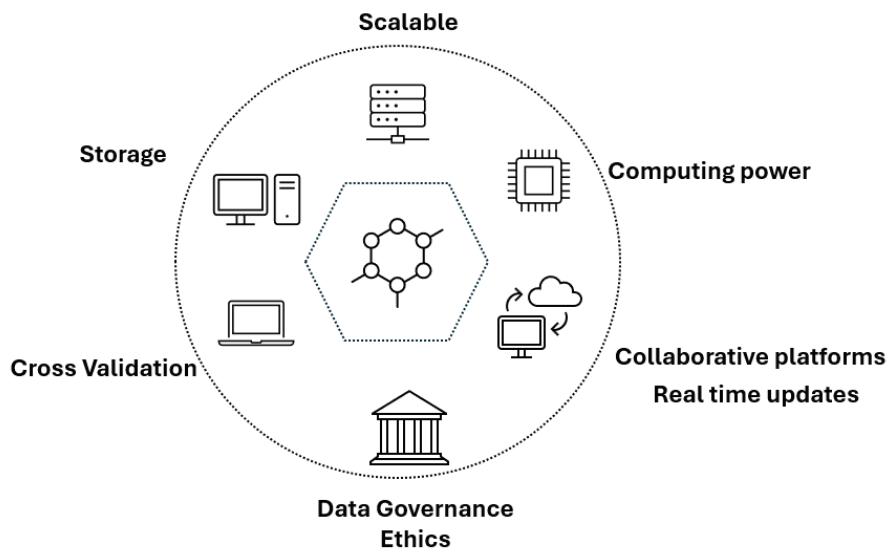


Figure 17. The data ecosystem fundamentals in AI integration.

292. The below are some principles proposed to integrate AI in risk assessment:

Principles that are needed to integrate AI in risk assessment

Flexibility and Agility

Ability and agility to adapt its AI strategies, models, and infrastructure to evolving technological advancements.

Accountability

Establish clear lines of responsibility for AI systems and their decisions, with frameworks for addressing errors or negative impacts.

Balancing Innovation

Regulatory Sandboxes can help innovators and regulators understand how new technologies interact with regulation.

Industry Engagement

Engage with stakeholders early on.

Training and development

Internal upskilling and next generation training.

Monitoring drift and bias

Monitoring drift and bias is crucial in AI to maintain model accuracy, ensure fairness, and build trust.

Trust

The GCRSR large language models (LLMs) Taskforce was established in early 2023 to develop a roadmap for the use of LLMs in regulatory application with participations from 13 countries. The team realized that trust is one of the most important aspects of the uptake of LLMs in the regulatory setting (Tong and Renaudin, 2024). For that,

they proposed the TREAT principle, which consists of Trustworthiness, Reproducibility, Explainability, Applicability, and Transparency (**Figure 18**).



Figure 18. TREAT outlines five key characteristics for AI integration in regulatory science, though their importance varies by application. For instance, can AI models with different features but similar predictions be trusted? If scientific theories must be testable, how critical is explainability? Should a rigid context of use be balanced with adaptability and transparency?

It was suggested that we may need rethink traditional definitions of these concepts, particularly in guiding the regulatory applications of the rapid advancements in modern AI. The rationale behind the TREAT principle is to treat AI like humans, where trust is earned through reproducible and consistent performance.

Finally, it was suggested that regulatory bodies such as the FDA may play a role to bring various stakeholders together with mechanisms for secure data-sharing across industry, academia, and regulatory agencies. Some advanced technologies of sharing such as federated learning or blockchain could also facilitate development of AI models in a privacy-preserving fashion.

Transparent and Explainable

Transparent reporting and validation

Communication/Engagement Plan

Data management and Governance

Ensure the data used to train AI models is accurate, complete, and free from significant bias to prevent flawed or suboptimal decision.

Create robust governance structures with clear policies and oversight to manage AI systems throughout their lifecycle.

Human oversight and interpretability

Human oversight in the development and application of AI models to provide context and domain expertise that can identify potential biases.

Regulatory and policy recommendations.

Regulatory agencies should establish guidelines and frameworks for the validation and deployment of AI models. Strategies in place to review this fast-moving field.

The New Frontier-Era of the new 4Rs: Readiness Resources Risk Responsibility

293. The COT and FSA have developed a live UK roadmap towards acceptance and integration of these new approach methodologies including predictive toxicology methods using computer modelling into safety and risk assessments for regulatory decision making.

294. In the roadmap they proposed that this will not only require the historic 3Rs approach (i.e. replacement, reduction and refinement of animal experiments) and the expansion to the 6R principle: (to also include) reproducibility, relevance, and regulatory acceptance.

295. There is a proposal to add new 4Rs in relation to AI (**Figure 19**):

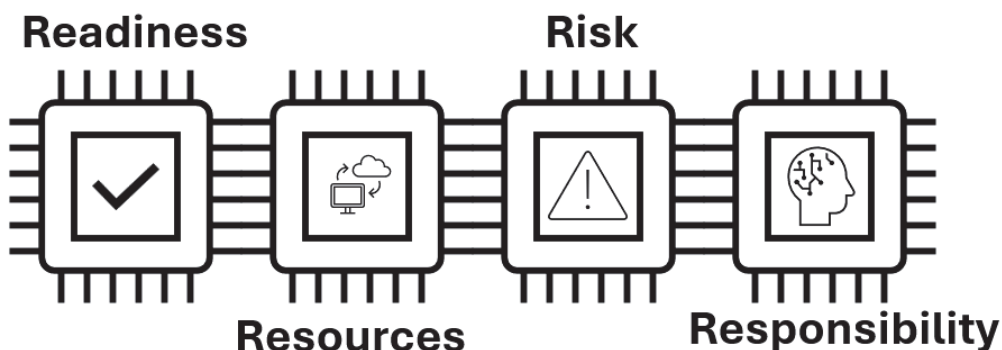
Resources Readiness Risk Responsibility

Figure 19. The 4 pillar chips of AI Adoption: Readiness, Resources, Risk and Responsibility.

- **Readiness:** Ready to be agile and flexible. Regulatory Readiness by governance: Establishing clear rules and standards for AI use in chemical risk assessment, ensuring that AI tools are used responsibly and ethically. This includes maintaining human oversight for AI-assisted decisions.
- **Resources:** Right computational infrastructure (including data capabilities, storage, hardware) and Right talent (People). Cloud platforms to combine disparate datasets from different sources and modalities. Subject matter experts: Involving toxicologists and chemists in all stages of the AI workflow, from data annotation to model validation, is critical for model relevance and accuracy.
- **Risk:** Know the risks of AI including bias and hallucinations. Establishing rigorous validation processes for AI-based methods is crucial to ensure their reliability, trustworthiness, and regulatory acceptance. New methods for validating and monitoring AI-based safety testing are needed to ensure their reliability and trustworthiness.
- **Responsibility:** Responsible for development of next generation tools. Accountability for transparency, reproducibility, and ethical considerations will

guide the responsible application of AI in the field. Establishing clear protocols and ongoing monitoring strategies to validate AI-driven predictions is essential for their reliable integration into risk assessment.

The Algorithm for Adoption - Concluding thoughts

296. At the Global Summit on Regulatory Science (GSRS24), regulators agreed that successful AI adoption requires ongoing dialogue, adaptability, and AI-trained personnel to harness its potential for regulatory responsibilities in the evolving 21st-century landscape.

297. These efforts will ultimately support better-informed regulatory decisions, improved public health outcomes, and greater trust in the scientific process.

298. As the field continues to evolve, we need to keep up to date ensuring scientific rigor and transparency.

299. The AI technological landscape is constantly changing and rapidly evolving, requiring regulatory agencies to adapt their AI strategies and infrastructure to monitor and review continuously.

300. By collaborating and considering the 10 principles we can develop new methodologies by using a phased approach to AI adoption which should prioritise and enhance environmental and human health as well as societal benefits.

Secretariat

October 2025

*“We can only see a short distance ahead,
but we can see plenty there that needs to be done.”*

- Alan Turing (1950)

References

Alpha Fold Website: [AlphaFold Protein Structure Database](#)

AI UK Government White paper: [A pro-innovation approach to AI regulation: government response - GOV.UK](#)

AI in UK government departments: [AI in UK government departments - House of Commons Library](#)

AI Playbook for the UK Government: [AI Playbook for the UK Government - GOV.UK](#)

AI UK HMG Opportunities Action Plan: [CP 1241 – AI Opportunities Action Plan – January 2025](#)

AI Safety Summit (2023) [AI Safety Summit 2023 - GOV.UK](#)

A pro-innovation approach to AI regulation UK Government (2023) [A pro-innovation approach to AI regulation - GOV.UK](#)

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F.L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S. and Avila, R., 2023. Gpt-4 technical report. *arXiv preprint* [GPT-4 Technical Report](#)

Anstine, D.M. and Isayev, O., 2023. Generative models as an emerging paradigm in the chemical sciences. *Journal of the American Chemical Society*, 145(16), pp.8736-8750.

Agarwal, C., Queen, O., Lakkaraju, H. and Zitnik, M., 2023. Evaluating explainability for graph neural networks. *Scientific Data*, 10(1), p.144.

The Association for the Advancement of Artificial Intelligence Website: [The Association for the Advancement of Artificial Intelligence](#)

Banh, L. and Strobel, G., 2023. Generative artificial intelligence. *Electronic Markets*, 33(1), p.63.

Barber, C., Fowkes, A., Hanser, T. and Heghes, C., 2024. From model performance to decision support–The rise of computational toxicology in chemical safety assessments. *Computational Toxicology*, 31, p.100303.

Belfield, S.J., Cronin, M.T., Enoch, S.J. and Firman, J.W., 2023. Guidance for good practice in the application of machine learning in development of toxicological quantitative structure-activity relationships (QSARs). *PloS one*, 18(5), p.e0282924.

Blümmel, T., Rehn, J., Mereu, C., Graf, F., Bazing, F., Kneuer, C., Sonnenburg, A., Wittkowski, P., Padberg, F., Bech, K. and Eleftheriadou, D., 2024. Exploring the use of artificial intelligence (AI) for extracting and integrating data obtained through new approach methodologies (NAMs) for chemical risk assessment. *EFSA Supporting Publications*, 21(1), p.8567E.

Bran, A.M., Cox, S., Schilter, O., Baldassari, C., White, A.D. and Schwaller, P., 2023. Chemcrow: Augmenting large-language models with chemistry tools. *arXiv preprint arXiv:2304.05376*. [\[2304.05376\] ChemCrow: Augmenting large-language models with chemistry tools](#)

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A. and Agarwal, S., 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33, pp.1877-1901.

Brundage, M., 2015. Taking superintelligence seriously: Superintelligence: Paths, dangers, strategies by nick bostrom (Oxford university press, 2014). *Futures*, 72, pp.32-35.

Bruna, J., Zaremba, W., Szlam, A. and LeCun, Y., 2013. Spectral networks and locally connected networks on graphs. *arXiv preprint [1312.6203] Spectral Networks and Locally Connected Networks on Graphs*

Bhatia, S., Schultz, T., Roberts, D., Shen, J., Kromidas, L. and Api, A.M., 2015. Comparison of cramer classification between toxtree, the OECD QSAR Toolbox and expert judgment. *Regulatory Toxicology and Pharmacology*, 71(1), pp.52-62.

- Born, J., Markert, G., Janakarajan, N., Kimber, T.B., Volkamer, A., Martínez, M.R. and Manica, M., 2023. Chemical representation learning for toxicity prediction. *Digital Discovery*, 2(3), pp.674-691.
- Bohde, M., Manjrekar, M., Wang, R., Ji, S. and Coley, C.W., 2025. Diffms: Diffusion generation of molecules conditioned on mass spectra. *arXiv preprint [2502.09571]* [DiffMS: Diffusion Generation of Molecules Conditioned on Mass Spectra](#)
- Bottini, A.A., Alepee, N., Phillips, B., Gribaldo, L., De Silva, O., Hartung, T., Hendriksen, C., Kuil, J., Pazos, P., Rhein, C. and Schiffelers, M.J., 2008. Optimisation of the post-validation process: The report and recommendations of ECVAM Workshop 67. *Alternatives to Laboratory Animals*, 36(3), pp.353-366.
- Canada, A., Capella-Gutierrez, S., Rabal, O., Oyarzabal, J., Valencia, A. and Krallinger, M., 2017. LimTox: a web tool for applied text mining of adverse event and toxicity associations of compounds, drugs and genes. *Nucleic acids research*, 45(W1), pp.W484-W489.
- Ciallella, H.L., Russo, D.P., Aleksunes, L.M., Grimm, F.A. and Zhu, H., 2021. Revealing adverse outcome pathways from public high-throughput screening data to evaluate new toxicants by a knowledge-based deep neural network approach. *Environmental Science & Technology*, 55(15), pp.10875-10887.
- Cheng, W. and Ng, C.A., 2019. Using machine learning to classify bioactivity for 3486 per-and polyfluoroalkyl substances (PFASs) from the OECD list. *Environmental Science & Technology*, 53(23), pp.13970-13980.
- Chenthamarakshan, V., Hoffman, S.C., Owen, C.D., Lukacik, P., Strain-Damerell, C., Fearon, D., Malla, T.R., Tumber, A., Schofield, C.J., Duyvesteyn, H.M. and Dejnirattisai, W., 2023. Accelerating drug target inhibitor discovery with a deep generative foundation model. *Science advances*, 9(25), p.eadg7865.
- Chen, D., Lin, Y., Li, W., Li, P., Zhou, J. and Sun, X., 2020, April. Measuring and relieving the over-smoothing problem for graph neural networks from the topological view. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 34, No. 04, pp. 3438-3445).

Chiang, W.L., Liu, X., Si, S., Li, Y., Bengio, S. and Hsieh, C.J., 2019, July. Cluster-gcn: An efficient algorithm for training deep and large graph convolutional networks. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 257-266).

Chopard, D., Treder, M.S., Corcoran, P., Ahmed, N., Johnson, C., Busse, M. and Spasic, I., 2021. Text mining of adverse events in clinical trials: deep learning approach. *JMIR Medical Informatics*, 9(12), p.e28632.

Chou, W.C. and Lin, Z., 2023. Machine learning and artificial intelligence in physiologically based pharmacokinetic modeling. *Toxicological Sciences*, 191(1), pp.1-14.

Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., Barham, P., Chung, H.W., Sutton, C., Gehrmann, S. and Schuh, P., 2023. Palm: Scaling language modeling with pathways. *Journal of Machine Learning Research*, 24(240), pp.1-113.

Cooper, M., Ji, Z. and Krishnan, R.G., 2023. Machine learning in computational histopathology: Challenges and opportunities. *Genes, Chromosomes and Cancer*, 62(9), pp.540-556.

Committee on Toxicity Horizon Scanning 2025: [Scanning Potential future discussion items – horizon scanning](#)

COT (2003a) *COT agenda and papers: 21 October 2003*, *The National Archives*. Available at: [\[ARCHIVED CONTENT\] COT: COT agenda and papers: 21 October 2003](#) (Accessed: 7 July 2025).

COT (2003b) *COT statement on physiologically based pharmacokinetic modelling*, *The National Archives*. Available at: (Accessed: 7 July 2025).

COT (2007) *COT statement on the COT workshop on evolving approaches to chemical risk assessment*, *The National Archives*. Available at: [\[ARCHIVED CONTENT\] UK Government Web Archive - The National Archives](#) (Accessed: 7 July 2025).

COT (2009) *COT statement on the COT workshop on 21st century toxicology*, *The National Archives*. Available at: [\[ARCHIVED CONTENT\] COT: COT statement on the COT workshop on 21st century toxicology](#) (Accessed: 7 July 2025).

COT (2021a) *COT FSA PBPK for Regulators Workshop Report*, *Committee on Toxicity*. Available at: [TOX-2021-06 PBPK for Regulators Report v.02](#) (Accessed: 7 July 2025).

COT (2021b) *Introduction to physiologically based pharmacokinetic (PBPK) modelling - Handbook 2021 Workshop*, *Committee on Toxicity*. Available at: [COT FSA PBPK for Regulators Handbook](#) (Accessed: 7 July 2025).

Corvi, J., Díaz-Roussel, N., Fernández, J.M., Ronzano, F., Centeno, E., Accuosto, P., Ibrahim, C., Asakura, S., Bringezu, F., Fröhlicher, M. and Kreuchwig, A., 2025. PretoxTM: a text mining system for extracting treatment-related findings from preclinical toxicology reports. *Journal of cheminformatics*, 17(1), p.15.

Digital and Technologies Sector Plan UK Government (2025) [The UK's Modern Industrial Strategy - DIGITAL AND TECHNOLOGIES - Sector Plan](#)

Devlin, J., Chang, M.W., Lee, K. and Toutanova, K., 2019, June. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)* (pp. 4171-4186).

Devadas, R.M. and Sowmya, T., 2025. Quantum machine learning: A comprehensive review of integrating AI with quantum computing for computational advancements. *MethodsX*, p.103318.

“Data Page: Computational capacity of the fastest supercomputers”. Our World in Data (2025). Data adapted from Dongarra et al. 2024 Retrieved from [Computational capacity of the fastest supercomputers](#) [online resource] (archived on June 24, 2025).

Ebers, M., 2020. Regulating explainable AI in the European Union. An overview of the current legal framework (s). *An Overview of the Current Legal Framework*

(s)(August 9, 2021). *Liane Colonna/Stanley Greenstein (eds.), Nordic Yearbook of Law and Informatics*.

ECETOC Integrating AI into chemical safety assessment – Flash Report:

[ECETOC-Flash-Report-Integrating-AI-into-chemical-safety-assessment.pdf](#)

EFSA Roadmap for actions on artificial intelligence for evidence management in risk assessment

[Roadmap for actions on artificial intelligence for evidence management in risk assessment | EFSA](#)

Egeghy, P.P., Judson, R., Gangwal, S., Mosher, S., Smith, D., Vail, J. and Hubal, E.A.C., 2012. The exposure data landscape for manufactured chemicals. *Science of the Total Environment*, 414, pp.159-166.

European Medicines Agency, 2023. Reflection paper on the use of artificial intelligence (AI) in the medicinal product lifecycle. [Reflection paper on the use of Artificial Intelligence \(AI\) in the medicinal product lifecycle 240903](#)

Epstein, Z., Hertzmann, A., Investigators of Human Creativity, Akten, M., Farid, H., Fjeld, J., Frank, M.R., Groh, M., Herman, L., Leach, N. and Mahari, R., 2023. Art and the science of generative AI. *Science*, 380(6650), pp.1110-1111.

Gao, W. and Coley, C.W., 2020. The synthesizability of molecules proposed by generative models. *Journal of chemical information and modeling*, 60(12), pp.5714-5723.

Geels, F.W., 2002. Technological transitions as evolutionary reconfiguration processes: a multi-level perspective and a case-study. *Research policy*, 31(8-9), pp.1257-1274.

Guidance Good Machine Learning Practice for Medical Device Development:

Guiding Principles: [Good Machine Learning Practice for Medical Device Development: Guiding Principles - GOV.UK](#)

Gilmer, J., Schoenholz, S.S., Riley, P.F., Vinyals, O. and Dahl, G.E., 2017, July. Neural message passing for quantum chemistry. In *International conference on machine learning* (pp. 1263-1272). Pmlr.

Gelernter, H., Rose, J.R. and Chen, C., 1990. Building and refining a knowledge base for synthetic organic chemistry via the methodology of inductive and deductive machine learning. *Journal of chemical information and computer sciences*, 30(4), pp.492-504.

Goodfellow, I., Bengio, Y., Courville, A. and Bengio, Y., 2016. Deep learning (Vol. 1, No. 2). Cambridge: MIT press.

Google Lens [Google Lens - Search What You See](#)

Griffiths, R.R. and Hernández-Lobato, J.M., 2020. Constrained Bayesian optimization for automatic chemical design using variational autoencoders. *Chemical science*, 11(2), pp.577-586.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F. and Pedreschi, D., 2018. A survey of methods for explaining black box models. *ACM computing surveys (CSUR)*, 51(5), pp.1-42.

Guo, W., Liu, J., Dong, F., Song, M., Li, Z., Khan, M.K.H., Patterson, T.A. and Hong, H., 2023. Review of machine learning and deep learning models for toxicity prediction. *Experimental Biology and Medicine*, 248(21), pp.1952-1973.

Guo, T., Nan, B., Liang, Z., Guo, Z., Chawla, N., Wiest, O. and Zhang, X., 2023. What can large language models do in chemistry? a comprehensive benchmark on eight tasks. *Advances in Neural Information Processing Systems*, 36, pp.59662-59688.

Global Summit on Regulatory Science (GSRS24) Theme: Digital Transformation in Regulatory Science [home-GSRS24](#)

Jaganathan, K., Tayara, H. and Chong, K.T., 2022. An explainable supervised machine learning model for predicting respiratory toxicity of chemicals using optimal molecular descriptors. *Pharmaceutics*, 14(4), p.832.

Jia, X., Wang, T. and Zhu, H., 2023. Advancing computational toxicology by interpretable machine learning. *Environmental Science & Technology*, 57(46), pp.17690-17706.

Jiménez-Luna, J., Grisoni, F. and Schneider, G., 2020. Drug discovery with explainable artificial intelligence. *Nature Machine Intelligence*, 2(10), pp.573-584.

- Jordan, M.I. and Mitchell, T.M., 2015. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), pp.255-260.
- Jo, E., Lee, Y., Lee, Y., Baek, J. and Kim, J.G., 2023. Rapid identification of counterfeited beef using deep learning-aided spectroscopy: Detecting colourant and curing agent adulteration. *Food and Chemical Toxicology*, 181, p.114088.
- Han, H. and Liu, X., 2022. The challenges of explainable AI in biomedical data science. *BMC bioinformatics*, 22(Suppl 12), p.443.
- Hamilton, W., Ying, Z. and Leskovec, J., 2017. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30.
- Hatakeyama-Sato, K., Yamane, N., Igarashi, Y., Nabae, Y. and Hayakawa, T., 2023. Prompt engineering of GPT-4 for chemical research: what can/cannot be done?. *Science and Technology of Advanced Materials: Methods*, 3(1), p.2260300.
- Hartung, T. and Kleinstreuer, N., 2025. Challenges and opportunities for validation of AI-based new approach methods. *ALTEX-Alternatives to animal experimentation*, 42(1), pp.3-21.
- Hartung, T., Whelan, M., Tong, W. and Califf, R.M., 2025. Is regulatory science ready for artificial intelligence?. *npj Digital Medicine*, 8(1), p.200.
- Hartung, T., Maertens, A. and Luechtefeld, T., 2024. E-validation—Unleashing AI for validation. *ALTEX-Alternatives to animal experimentation*, 41(4), pp.567-587.
- Hartung, T., Hoffmann, S. and Whaley, P., 2025. Assessing risk of bias in toxicological studies in the era of artificial intelligence. *Archives of Toxicology*, pp.1-26.
- HESI AI Health Protection Meeting 2025
[HESI AI Health Protection Meeting 2025 - HESI - Health and Environmental Sciences Institute](#)
- HESI AI Health Protection Meeting Report 2025
[Building a Roadmap for AI-Enabled Human and Environmental Health Protection](#)
- Hennessy, J.L. and Patterson, D.A., 2011. *Computer architecture: a quantitative approach*. Elsevier.

Huang, L., Duan, Q., Liu, Y., Wu, Y., Li, Z., Guo, Z., Liu, M., Lu, X., Wang, P., Liu, F. and Ren, F., 2025. Artificial intelligence: A key fulcrum for addressing complex environmental health issues. *Environment International*, p.109389.

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B. and Liu, T., 2025. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, 43(2), pp.1-55.

Huang, Z., Lou, S., Wang, H., Li, W., Liu, G. and Tang, Y., 2024. Attentiveskin: To predict skin corrosion/irritation potentials of chemicals via explainable machine learning methods. *Chemical Research in Toxicology*, 37(2), pp.361-373.

Industrial Strategy UK Government (2025) [The UK's Modern Industrial Strategy](#)

Kakkad, J., Jannu, J., Sharma, K., Aggarwal, C. and Medya, S., 2023. A survey on explainability of graph neural networks. *arXiv preprint* [\[2306.01958\] A Survey on Explainability of Graph Neural Networks](#)

Kalian, A.D., Benfenati, E., Osborne, O.J., Gott, D., Potter, C., Dorne, J.L.C., Guo, M. and Hogstrand, C., 2023. Exploring dimensionality reduction techniques for deep learning driven QSAR models of mutagenicity. *Toxics*, 11(7), p.572.

Kalian, A.D., Benfenati, E., Gott, D., Potter, C., Dorne, J.L., Osborne, O.J., Guo, M. and Hogstrand, C., 2024. P05-37 Graph attention networks using knowledge graphs, for predicting novel points of departure for brominated flame retardants. *Toxicology Letters*, 399, pp.S146-S147.

Kalian, A.D., Otte, L., Lee, J., Benfenati, E., Dorne, J.L., Potter, C., Osborne, O.J., Guo, M. and Hogstrand, C., 2025. Comparison of Optimised Geometric Deep Learning Architectures, over Varying Toxicological Assay Data Environments. *arXiv preprint arXiv:2507.17775*.

Kalian, A.D., Silva, A.C., Lee, J., Dorne, J.L.C., Potter, C., Benfenati, E., Osborne, O.J., Guo, M. and Hogstrand, C., 2025. Predicting Organ-Specific Toxicity of Selective Androgen Receptor Modulators, using Transfer Learning on Graph Convolutional Networks. *bioRxiv*, pp.2025-08. [Predicting Organ-Specific Toxicity of](#)

[Selective Androgen Receptor Modulators, using Transfer Learning on Graph Convolutional Networks | bioRxiv](#)

Kamiya, Y., Handa, K., Miura, T., Yanagi, M., Shigeta, K., Hina, S., Shimizu, M., Kitajima, M., Shono, F., Funatsu, K. and Yamazaki, H., 2021. In silico prediction of input parameters for simplified physiologically based pharmacokinetic models for estimating plasma, liver, and kidney exposures in rats after oral doses of 246 disparate chemicals. *Chemical research in toxicology*, 34(2), pp.507-513.

Karim, A., Singh, J., Mishra, A., Dehzangi, A., Newton, M.H. and Sattar, A., 2019a, August. Toxicity prediction by multimodal deep learning. In *Pacific Rim Knowledge Acquisition Workshop* (pp. 142-152). Cham: Springer International Publishing.

Karim, A., Mishra, A., Newton, M.H. and Sattar, A., 2019b. Efficient toxicity prediction via simple features using shallow neural networks and decision trees. *Acs Omega*, 4(1), pp.1874-1888.

Karim, M.R., Islam, T., Shajalal, M., Beyan, O., Lange, C., Cochez, M., Rebholz-Schuhmann, D. and Decker, S., 2023. Explainable AI for bioinformatics: methods, tools and applications. *Briefings in bioinformatics*, 24(5), p.bbaf236.

Kipf, T.N., 2016. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv preprint [1609.02907] Semi-Supervised Classification with Graph Convolutional Networks*

Khatun, A. and Brown, D.G., 2023. Reliability Check: An Analysis of GPT-3's Response to Sensitive Topics and Prompt Wording. *arXiv preprint [2306.06199] Reliability Check: An Analysis of GPT-3's Response to Sensitive Topics and Prompt Wording*

Kleinstreuer, N. and Hartung, T., 2024. Artificial intelligence (AI)—it's the end of the tox as we know it (and I feel fine). *Archives of toxicology*, 98(3), pp.735-754.

Kadurin, A., Nikolenko, S., Khrabrov, K., Aliper, A. and Zhavoronkov, A., 2017. druGAN: an advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico. *Molecular pharmaceutics*, 14(9), pp.3098-3104.

Kuan, J., Radaeva, M., Avenido, A., Cherkasov, A. and Gentile, F., 2023. Keeping pace with the explosive growth of chemical libraries with structure-based virtual screening. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 13(6), p.e1678.

Lai, H., Liu, J., Bai, C., Liu, H., Pan, B., Luo, X., Hou, L., Zhao, W., Xia, D., Tian, J. and Chen, Y., 2025. Language models for data extraction and risk of bias assessment in complementary medicine. *npj Digital Medicine*, 8(1), p.74.

Lee, S.H., Choi, E., Park, J., Yoon, S., Song, M.H., Lee, J.Y., Seo, J., Shin, S.K., Lee, S.H. and Oh, H.B., 2025. Prediction of reproductive and developmental toxicity using an attention and gate augmented graph convolutional network. *Scientific Reports*, 15(1), p.18186.

Life Sciences Sector Plan UK Government (2025) [life Sciences Sector Plan](#)

Lie, M. and Sørensen, K.H., 1996. Making technology our own? domesticating technology into everyday life.

Li, T., Tong, W., Roberts, R., Liu, Z. and Thakkar, S., 2021. DeepCarc: Deep learning-powered carcinogenicity prediction using model-level representation. *Frontiers in artificial intelligence*, 4, p.757780.

Lim, J., Ryu, S., Kim, J.W. and Kim, W.Y., 2018. Molecular generative model based on conditional variational autoencoder for de novo molecular design. *Journal of cheminformatics*, 10(1), p.31.

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H. and Neubig, G., 2023. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM computing surveys*, 55(9), pp.1-35.

Li, S., Zhou, J., Xu, T., Huang, L., Wang, F., Xiong, H., Huang, W., Dou, D. and Xiong, H., 2021, August. Structure-aware interactive graph neural networks for the prediction of protein-ligand binding affinity. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining* (pp. 975-985).

Lin, Z., Basili, D. and Chou, W.C., 2024. Preface to the special issue of food and chemical toxicology on “new approach methodologies and machine learning in food safety and chemical risk assessment: development of reproducible, open-source,

and user-friendly tools for exposure, toxicokinetic, and toxicity assessments in the 21st century". *Food and Chemical Toxicology*, 190, p.114809.

Lin, Z. and Chou, W.C., 2022. Machine learning and artificial intelligence in toxicological sciences. *Toxicological Sciences*, 189(1), pp.7-19.

Lin, Z., Chou, W.C., Cheng, Y.H., He, C., Monteiro-Riviere, N.A. and Riviere, J.E., 2022. Predicting nanoparticle delivery to tumors using machine learning and artificial intelligence approaches. *International journal of nanomedicine*, pp.1365-1379.

Liang, B., Gong, H., Lu, L. and Xu, J., 2022. Risk stratification and pathway analysis based on graph neural network and interpretable algorithm. *BMC bioinformatics*, 23(1), p.394.

Luo, R., Sun, L., Xia, Y., Qin, T., Zhang, S., Poon, H. and Liu, T.Y., 2022. BioGPT: generative pre-trained transformer for biomedical text generation and mining. *Briefings in bioinformatics*, 23(6), p.bbac409.

Luo, D., Cheng, W., Xu, D., Yu, W., Zong, B., Chen, H. and Zhang, X., 2020. Parameterized explainer for graph neural network. *Advances in neural information processing systems*, 33, pp.19620-19631.

Lundberg, S.M. and Lee, S.I., 2017. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.

Hatakeyama-Sato, K., Yamane, N., Igarashi, Y., Nabae, Y. and Hayakawa, T., 2023. Prompt engineering of GPT-4 for chemical research: what can/cannot be done?. *Science and Technology of Advanced Materials: Methods*, 3(1), p.2260300.

HESI Building a Roadmap for AI-Enabled Human and Environmental Health Protection Workshop Report: [Building a Roadmap for AI-Enabled Human and Environmental Health Protection](#)

Hong, J. and Kwon, H., 2025. Multimodal deep learning for chemical toxicity prediction and management. *Scientific Reports*, 15(1), p.19491.

LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. *nature*, 521(7553), pp.436-444.

- Liu, X., Zhao, J., Zhang, G., Wei, S., Song, X. and Yin, D., 2025. Deep learning-based screening approach for priority pollutants: a case study on retired power battery recycling. *Environmental Pollution*, p.126849.
- Loeffler, H.H., He, J., Tibo, A., Janet, J.P., Voronov, A., Mervin, L.H. and Engkvist, O., 2024. Reinvent 4: modern AI-driven generative molecule design. *Journal of Cheminformatics*, 16(1), p.20.
- Luechtefeld, T., Rowlands, C. and Hartung, T., 2018. Big-data and machine learning to revamp computational toxicology and its use in risk assessment. *Toxicology research*, 7(5), pp.732-744.
- Mayr, A., Klambauer, G., Unterthiner, T. and Hochreiter, S., 2016. DeepTox: toxicity prediction using deep learning. *Frontiers in Environmental Science*, 3, p.80.
- National Audit Office-Use of artificial intelligence in government value for money report: [Use of artificial intelligence in government](#)
- Nascimento, C.M.C., Moura, P.G. and Pimentel, A.S., 2023. Generating structural alerts from toxicology datasets using the local interpretable model-agnostic explanations method. *Digital Discovery*, 2(5), pp.1311-1325.
- Mastropietro, A., Pasculli, G. and Bajorath, J., 2023. Learning characteristics of graph neural networks predicting protein–ligand affinities. *Nature Machine Intelligence*, 5(12), pp.1427-1436.
- Medsker, L.R. and Jain, L., 2001. Design and applications. *Recurrent Neural Networks*, 5(64-67), p.2.
- Méndez-Lucio, O., Nicolaou, C.A. and Earnshaw, B., 2024. MolE: a foundation model for molecular graphs using disentangled attention. *Nature Communications*, 15(1), p.9431.
- Meskó, B. and Topol, E.J., 2023. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *NPJ digital medicine*, 6(1), p.120.
- MHRA Guidance-Predetermined change control plans for machine learning-enabled medical devices: [Predetermined change control plans for machine learning-enabled medical devices - GOV.UK](#)

Mittal, A., Mohanty, S.K., Gautam, V., Arora, S., Sapru, S., Gupta, R., Sivakumar, R., Garg, P., Aggarwal, A., Raghavachary, P. and Dixit, N.K., 2022. Artificial intelligence uncovers carcinogenic human metabolites. *Nature Chemical Biology*, 18(11), pp.1204-1213.

New Approach Methodologies (NAMs) to Support Regulatory Decisions for Chemical Safety: [New Approach Methodologies \(NAMs\) to Support Regulatory Decisions for Chemical Safety | Published in FSA Research and Evidence](#)

Osborne, O.J., Boobis, A., Botham, P., Price, S.C., Kuhnle, G.C., Mulholland, C., Potter, C. and Gott, D., 2024. The new normal chemical landscape: the future of risk assessment toward optimum consumer safety. *Toxicology Research*, 13(2), p.tfae016.

Osborne, O.J., Botham, P., Mulholland, C., Potter, C., Gott, D., Adkin, A. and Boobis, A., 2024. Food for thought—Paving the way for a UK roadmap towards optimum consumer safety: Development, Endorsement and Regulatory acceptance of New Approach Methodologies (NAMs) in Chemical Risk Assessment and Beyond. *Regulatory Toxicology and Pharmacology*, 153, p.105701.

OpenAI, 2018. AI and compute. [AI and compute | OpenAI](#)

O'shea, K. and Nash, R., 2015. An introduction to convolutional neural networks. *arXiv preprint* : [\[1511.08458\] An Introduction to Convolutional Neural Networks](#)

Ong, J.C.L., Chang, S.Y.H., William, W., Butte, A.J., Shah, N.H., Chew, L.S.T., Liu, N., Doshi-Velez, F., Lu, W., Savulescu, J. and Ting, D.S.W., 2024. Ethical and regulatory challenges of large language models in medicine. *The Lancet Digital Health*, 6(6), pp.e428-e432.

Pandey, D., Niwaria, K. and Chourasia, B., 2019. Machine learning algorithms: a review. *Mach. Learn*, 6(2).

Paini, A., Campia, I., Cronin, M.T., Asturiol, D., Ceriani, L., Exner, T.E., Gao, W., Gomes, C., Kruisselbrink, J., Martens, M. and Meek, M.B., 2022. Towards a qAOP framework for predictive toxicology-linking data to decisions. *Computational Toxicology*, 21, p.100195.

Pawar, G., Madden, J.C., Ebbrell, D., Firman, J.W. and Cronin, M.T., 2019. In silico toxicology data resources to support read-across and (Q) SAR. *Frontiers in pharmacology*, 10, p.561.

Pérez Santín, E., Rodríguez Solana, R., González García, M., García Suárez, M.D.M., Blanco Díaz, G.D., Cima Cabal, M.D., Moreno Rojas, J.M. and López Sánchez, J.I., 2021. Toxicity prediction based on artificial intelligence: A multidisciplinary overview. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 11(5), p.e1516.

Popescu, M.C., Balas, V.E., Perescu-Popescu, L. and Mastorakis, N., 2009. Multilayer perceptron and neural networks. *WSEAS Transactions on Circuits and Systems*, 8(7), pp.579-588.

Polykovskiy, D., Zhebrak, A., Sanchez-Lengeling, B., Golovanov, S., Tatanov, O., Belyaev, S., Kurbanov, R., Artamonov, A., Aladinskiy, V., Veselov, M. and Kadurin, A., 2020. Molecular sets (MOSES): a benchmarking platform for molecular generation models. *Frontiers in pharmacology*, 11, p.565644.

PROAST [PROAST | RIVM](#)

UK National AI Strategy [National AI Strategy](#)

Rae, J.W., Borgeaud, S., Cai, T., Millican, K., Hoffmann, J., Song, F., Aslanides, J., Henderson, S., Ring, R., Young, S. and Rutherford, E., 2021. Scaling language models: Methods, analysis & insights from training gopher. *arXiv preprint arXiv:2112.11446*.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W. and Liu, P.J., 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, 21(140), pp.1-67.

Rattan, A.K., 2018. Data integrity: history, issues, and remediation of issues. *PDA Journal of Pharmaceutical Science and Technology*, 72(2), pp.105-116.

Rydning, D.R.J.G.J., Reinsel, J. and Gantz, J., 2018. The digitization of the world from edge to core. *Framingham: International Data Corporation*, 16, pp.1-28.

Russell, S.J. and Norvig, P., 2021. Artificial Intelligence: A Modern Approach, Global Edition 4e.

- Rusch, T.K., Bronstein, M.M. and Mishra, S., 2023. A survey on oversmoothing in graph neural networks. *arXiv preprint [2303.10993] A Survey on Oversmoothing in Graph Neural Networks*
- Sanchez-Lengeling, B. and Aspuru-Guzik, A., 2018. Inverse molecular design using machine learning: Generative models for matter engineering. *Science*, 361(6400), pp.360-365.
- Sandhu, R.S., 1993, September. On Five Definitions of Data Integrity. In *DBSec* (pp. 257-267).
- Seal, S., Mahale, M., García-Ortegón, M., Joshi, C.K., Hosseini-Gerami, L., Beatson, A., Greenig, M., Shekhar, M., Patra, A., Weis, C. and Mehrjou, A., 2025. Machine Learning for Toxicity Prediction Using Chemical Structures: Pillars for Success in the Real World. *Chemical research in toxicology*, 38(5), pp.759-807.
- Seal, A., 2025. AOP Net: An AI-Enhanced Software Platform for the Visualization and Analysis of Complex Toxicological Pathways. *bioRxiv*, pp.2025-08. [AOP Net: An AI-Enhanced Software Platform for the Visualization and Analysis of Complex Toxicological Pathways | bioRxiv](#)
- Scarselli, F., Gori, M., Tsoi, A.C., Hagenbuchner, M. and Monfardini, G., 2008. The graph neural network model. *IEEE transactions on neural networks*, 20(1), pp.61-80.
- Schultz, T.W., Diderich, R., Kuseva, C.D. and Mekenyan, O.G., 2018. The OECD QSAR toolbox starts its second decade. In *Computational toxicology: methods and protocols* (pp. 55-77). New York, NY: Springer New York.
- Schoene, A.M., Basinas, I., Van Tongeren, M. and Ananiadou, S., 2022. A narrative literature review of natural language processing applied to the occupational exposome. *International journal of environmental research and public health*, 19(14), p.8544.
- Sinha, K., Ghosh, N. and Sil, P.C., 2023. A review on the recent applications of deep learning in predictive drug toxicological studies. *Chemical Research in Toxicology*, 36(8), pp.1174-1205.
- Silver D, Huang A, Maddison CJ, Guez A, Sifre L, Van Den Driessche G, Schrittwieser J, Antonoglou I, Panneershelvam V, Lanctot M, Dieleman S. Mastering

the game of Go with deep neural networks and tree search. *Nature* **529**, 484–489 (2016).

Simões, R.S., Maltarollo, V.G., Oliveira, P.R. and Honorio, K.M., 2018. Transfer and multi-task learning in QSAR modeling: advances and challenges. *Frontiers in pharmacology*, *9*, p.74.

Sivathanu, G., Wright, C.P. and Zadok, E., 2005, November. Ensuring data integrity in storage: Techniques and applications. In *Proceedings of the 2005 ACM workshop on Storage security and survivability* (pp. 26-36).

Snijders, C.C.P., Matzat, U. and Reips, U.D., 2012. " Big Data": big gaps of knowledge in the field of internet science. *International journal of internet science*, *7*(1), pp.1-5.

Sonnenburg, A., van der Lugt, B., Rehn, J., Wittkowski, P., Bech, K., Padberg, F., Eleftheriadou, D., Dobrikov, T., Bouwmeester, H., Mereu, C. and Graf, F., 2024. Artificial intelligence-based data extraction for next generation risk assessment: Is fine-tuning of a large language model worth the effort?. *Toxicology*, *508*, p.153933.

Sundararajan, M., Taly, A. and Yan, Q., 2017, July. Axiomatic attribution for deep networks. In *International conference on machine learning* (pp. 3319-3328). PMLR.

Sypetkowski, M., Wenkel, F., Poursafaei, F., Dickson, N., Suri, K., Fradkin, P. and Beaini, D., 2024. On the scalability of gnns for molecular graphs. *Advances in Neural Information Processing Systems*, *37*, pp.19870-19906.

Swanson, K., Walther, P., Leitz, J., Mukherjee, S., Wu, J.C., Shivnaraine, R.V. and Zou, J., 2024. ADMET-AI: a machine learning ADMET platform for evaluation of large-scale chemical libraries. *Bioinformatics*, *40*(7), p.btae416.

Tan, Z., Lin, K., Zhao, Y. and Zhou, T., 2025. Generative discovery of safer chemical alternatives using diffusion modeling: A case study in green solvent design for cyclohexane/benzene extractive distillation. *Journal of Environmental Sciences*, *154*, pp.390-401.

Thompson, L.A., Evans, J.G. and Matthews, S.T., 2025. AmesFormer: State-of-the-Art Mutagenicity Prediction with Graph Transformers. *Chemical Research in Toxicology*, *38*(7), pp.1167-1182.

- Tran, T.T.V., Tayara, H. and Chong, K.T., 2023. Artificial intelligence in drug metabolism and excretion prediction: recent advances, challenges, and future perspectives. *Pharmaceutics*, 15(4), p.1260.
- Tian, S., Jin, Q., Yeganova, L., Lai, P.T., Zhu, Q., Chen, X., Yang, Y., Chen, Q., Kim, W., Comeau, D.C. and Islamaj, R., 2023. Opportunities and challenges for ChatGPT and large language models in biomedicine and health. *Briefings in Bioinformatics*, 25(1).
- Tong, W. and Renaudin, M., 2024. Context is Everything in Regulatory Application of Large Language Models (LLMs). *Drug Discovery Today*, pp.103916-103916.
- Tropsha, A., 2010. Best practices for QSAR model development, validation, and exploitation. *Molecular informatics*, 29(6-7), pp.476-488.
- Turpin, M., Michael, J., Perez, E. and Bowman, S., 2023. Language models don't always say what they think: Unfaithful explanations in chain-of-thought prompting. *Advances in Neural Information Processing Systems*, 36, pp.74952-74965.
- Warraich, H.J., Tazbaz, T. and Califf, R.M., 2025. FDA perspective on the regulation of artificial intelligence in health care and biomedicine. *Jama*, 333(3), pp.241-247.
- Wittwehr, C., Blomstedt, P., Gosling, J.P., Peltola, T., Raffael, B., Richarz, A.N., Sienkiewicz, M., Whaley, P., Worth, A. and Whelan, M., 2020. Artificial Intelligence for chemical risk assessment. *Computational Toxicology*, 13, p.100114.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A. and Bridgland, A., 2021. Highly accurate protein structure prediction with AlphaFold. *nature*, 596(7873), pp.583-589.
- Litsa, E.E., Das, P. and Kavraki, L.E., 2020. Prediction of drug metabolites using neural machine translation. *Chemical science*, 11(47), pp.12777-12788.
- Lighthill, J., 1973. Artificial intelligence: a paper symposium. Science Research Council, London.

Moor, J. (2006). The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years. *AI Magazine*, 27(4), 87. [The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years | AI Magazine](#)

Kleinstreuer, N. and Hartung, T., 2024. Artificial intelligence (AI)—it's the end of the tox as we know it (and I feel fine). *Archives of toxicology*, 98(3), pp.735-754.

Samuel, A.L., 1959. Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, 3(3), pp.210-229.

Singh, A.V., Bhardwaj, P., Laux, P., Pradeep, P., Busse, M., Luch, A., Hirose, A., Osgood, C.J. and Stacey, M.W., 2024. AI and ML-based risk assessment of chemicals: predicting carcinogenic risk from chemical-induced genomic instability. *Frontiers in toxicology*, 6, p.1461587.

Tong, W. and Renaudin, M., 2024. Context is Everything in Regulatory Application of Large Language Models (LLMs). *Drug Discovery Today*, pp.103916-103916.

USA FDA Machine Learning in the Development of Drug and Biological Products (2023) [Using Artificial Intelligence & Machine Learning in the Development of Drug and Biological Products](#)

US EPA Benchmark Dose Tools [Benchmark Dose Tools | US EPA](#)

Vall, A., Sabnis, Y., Shi, J., Class, R., Hochreiter, S. and Klambauer, G., 2021. The Promise of AI for DILI Prediction: Andreu Vall, Yogesh Sabnis, Jiye Shi, Reiner Class, Sepp Hochreiter, Günter Klambauer. *Frontiers in Artificial Intelligence*, 4(638410).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P. and Bengio, Y., 2017. Graph attention networks. *stat*, 1050(20), pp.10-48550.

Veale, M. and Zuiderveen Borgesius, F., 2021. Demystifying the Draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, 22(4), pp.97-112.

VEGA HUB: [in Silico Models – VEGA HUB](#)

Wang, M.W., Goodman, J.M. and Allen, T.E., 2020. Machine learning in predictive toxicology: recent applications and future directions for classification models. *Chemical research in toxicology*, 34(2), pp.217-239.

Wang, Z., Walker, G.W., Muir, D.C. and Nagatani-Yoshida, K., 2020. Toward a global understanding of chemical pollution: a first comprehensive analysis of national and regional chemical inventories. *Environmental science & technology*, 54(5), pp.2575-2584.

Walters, W.P. and Murcko, M., 2020. Assessing the impact of generative AI on medicinal chemistry. *Nature biotechnology*, 38(2), pp.143-145.

Walter, M., Webb, S.J. and Gillet, V.J., 2024. Interpreting neural network models for toxicity prediction by extracting learned chemical features. *Journal of Chemical Information and Modeling*, 64(9), pp.3670-3688.

Wassenaar, P.N., Rorije, E., Vijver, M.G. and Peijnenburg, W.J., 2021. Evaluating chemical similarity as a measure to identify potential substances of very high concern. *Regulatory Toxicology and Pharmacology*, 119, p.104834.

Wu, Z., Wang, J., Du, H., Jiang, D., Kang, Y., Li, D., Pan, P., Deng, Y., Cao, D., Hsieh, C.Y. and Hou, T., 2023. Chemistry-intuitive explanation of graph neural networks for molecular property prediction with substructure masking. *Nature Communications*, 14(1), p.2585.

Wellawatte, G.P., Gandhi, H.A., Seshadri, A. and White, A.D., 2023. A perspective on explanations of molecular prediction models. *Journal of Chemical Theory and Computation*, 19(8), pp.2149-2160.

Williams, A.J. and Richard, A.M., 2025. Three pillars for ensuring public access and integrity of chemical databases powering cheminformatics. *Journal of Cheminformatics*, 17(1), pp.1-5.

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C. and Yu, P.S., 2020. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, 32(1), pp.4-24.

- Wu, Y., Liu, Z., Wu, L., Chen, M. and Tong, W., 2021. BERT-based natural language processing of drug labeling documents: a case study for classifying drug-induced liver injury risk. *Frontiers in Artificial Intelligence*, 4, p.729834.
- Wu, P.Y., Chou, W.C., Wu, X., Kamineni, V.N., Kuchimanchi, Y., Tell, L.A., Maunsell, F.P. and Lin, Z., 2025. Development of machine learning-based quantitative structure–activity relationship models for predicting plasma half-lives of drugs in six common food animal species. *Toxicological Sciences*, 203(1), pp.52-66.
- Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C. and Yu, P.S., 2020. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, 32(1), pp.4-24.
- Wu, W., Qian, J., Liang, C., Yang, J., Ge, G., Zhou, Q. and Guan, X., 2023. Geodili: A robust and interpretable model for drug-induced liver injury prediction using graph neural network-based molecular geometric representation. *Chemical research in toxicology*, 36(11), pp.1717-1730.
- Xu, K., Hu, W., Leskovec, J. and Jegelka, S., 2018. How powerful are graph neural networks?. *arXiv preprint [\[1810.00826\] How Powerful are Graph Neural Networks?](https://arxiv.org/abs/1810.00826)*
- Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Zhang, W., Cui, B. and Yang, M.H., 2023. Diffusion models: A comprehensive survey of methods and applications. *ACM computing surveys*, 56(4), pp.1-39.
- Yenduri, G., Ramalingam, M., Selvi, G.C., Supriya, Y., Srivastava, G., Maddikunta, P.K.R., Raj, G.D., Jhaveri, R.H., Prabadevi, B., Wang, W. and Vasilakos, A.V., 2024. Gpt (generative pre-trained transformer)—A comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions. *IEEE access*, 12, pp.54608-54649.
- Yi, J.C., Yang, Z.Y., Zhao, W.T., Yang, Z.J., Zhang, X.C., Wu, C.K., Lu, A.P. and Cao, D.S., 2024. ChemMORT: an automatic ADMET optimization platform using deep learning and multi-objective particle swarm optimization. *Briefings in Bioinformatics*, 25(2), p.bbae008.

- Ying, Z., Bourgeois, D., You, J., Zitnik, M. and Leskovec, J., 2019. Gnnexplainer: Generating explanations for graph neural networks. *Advances in neural information processing systems*, 32.
- Yuan, K. and Nault, R., 2025. Application of a metabolic network-based graph neural network for the identification of toxicant-induced perturbations. *Toxicological Sciences*, 206(1), pp.19-29.
- Yun, S., Jeong, M., Kim, R., Kang, J. and Kim, H.J., 2019. Graph transformer networks. *Advances in neural information processing systems*, 32.
- Zeng, X., Wang, F., Luo, Y., Kang, S.G., Tang, J., Lightstone, F.C., Fang, E.F., Cornell, W., Nussinov, R. and Cheng, F., 2022. Deep generative molecular design reshapes drug discovery. *Cell Reports Medicine*, 3(12).
- Zeng, H., Zhou, H., Srivastava, A., Kannan, R. and Prasanna, V., 2019. Graphsaint: Graph sampling based inductive learning method. *arXiv preprint [\[1907.04931\]](https://arxiv.org/abs/1907.04931)*
[GraphSAINT: Graph Sampling Based Inductive Learning Method](https://arxiv.org/abs/1907.04931)
- Zhang, J., Wang, Q., Wen, H., Gerbaud, V., Jin, S. and Shen, W., 2024. Multi-objective optimization strategy for green solvent design via a deep generative model learned from pre-set molecule pairs. *Green Chemistry*, 26(1), pp.412-427.
- Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C. and Sun, M., 2020. Graph neural networks: A review of methods and applications. *AI open*, 1, pp.57-81.
- Zhou, Y., Ai, J., Cai, L., Yan, Y., Wang, B., Ma, H., Yu, Q., Zhou, J. and Huo, X., 2023. Deep learning enabled miniature mass spectrometer for rapid qualitative and quantitative analysis of pesticides on vegetable surfaces. *Food and Chemical Toxicology*, 180, p.114000.
- Zhu, K., Huang, M., Wang, Y., Gu, Y., Li, W., Liu, G. and Tang, Y., 2024. MetaPredictor: in silico prediction of drug metabolites based on deep language models with prompt engineering. *Briefings in Bioinformatics*, 25(5).

Abbreviations

AI	Artificial Intelligence
ADME	Absorption, Distribution, Metabolism, and Excretion/Elimination
ASCII	American Standard Code for Information Interchange
AOP	Adverse outcome pathway
qAOP	quantitative AOP
AD	Applicability Domain
AUC	Area under the curve
BERT	Bidirectional Encoder Representations from Transformers
BFRs	Brominated Flame Retardants
BMD	Benchmark Dose
COU	Context of use
COT	Committee on Toxicity of Chemicals in Food, Consumer Products and the Environment
CRA	Chemical Risk Assessment
DL	Deep Learning
DICT	Drug-Induced Cardiotoxicity
DILI	Drug-Induced Liver Injury
DIRI	Drug-Induced Renal Injury
DNN	Deep neural network
DSIT	Department for Science, Innovation and Technology
EFSA	European Food Safety Authority
EPA	Environment Protection Agency
FARAD	Food Animal Residue Avoidance Databank
FAIR	Findability, accessibility, interoperability, and reusability
FLOPS	Floating point operations per second
FDA	Food and Drug Administration
XAI	Explainable AI
GATs	Graph Attention Networks
GenAI	Generative AI
GNNs	Graph Neural Networks
GCNs	Graph Convolutional Networks
GPT	Generative Pre-trained Transformer
HTS	High-throughput screening
HBGVs	Health-Based Guidance Values
IDC	International Data Corporation
IVIVE	<i>in vitro</i> to <i>in vivo</i> extrapolation
LIME	Local Interpretable Model-agnostic Explanations
LLMs	Large Language Models

ML	Machine Learning
MIT	Massachusetts Institute of Technology
MHRA	Medicines and Healthcare products Regulatory Agency
MIE	Molecular initiating event
NLP	Natural Language Processing
NAMs	New Approach Methodologies
NLP	Natural Language Processing
NGRA	Next-generation risk assessment
OECD	Organisation for Economic Co-operation and Development
OPPS	Office for Product Safety and Standards
PCA	Principal component analysis
PFAS	per- and polyfluorinated alkyl substances
PBPK	Physiologically Based Pharmacokinetic
RP	Reference point
RPFs	Relative potency factors and
RNNs	Recurrent Neural Networks
SARMs	Selective Androgen Receptor Modulators
SHAP	SHapley Additive exPlanations
SMILES	Simplified Molecular Input Line-Entry System
SME	Subject-Matter Expert
SPIDO	Science Studies and Project Identification & Development Office
SVMs	Support vector machines
SNS	Synthetic Nervous System
T-NLG	Turing Natural Language Generation
TEFs	Toxic equivalency factors
VAEs	Variational Autoencoders

Technical Terms

Applicability Domain	In chemistry and machine learning, the applicability domain of a quantitative structure-activity relationship model defines the boundaries within which the model's predictions are considered reliable. It represents the chemical, structural, or biological space covered by the training data used to build the model.
Bayesian approaches	Bayesian approaches is a method of statistical inference in which Bayes' theorem is used to calculate a probability of a hypothesis, given prior evidence, and update it as more information becomes available.
Central processing unit	A central processing unit, also called a central processor, main processor, or just processor, is the primary processor in a given computer.
Chat bot	A chatbot is a software application or web interface designed to have textual or spoken conversations.
Convex hull	A convex hull is the smallest convex shape containing a given set of points or a given geometric object.
Cover's theorem	Cover's theorem is a statement in computational learning theory and is one of the primary theoretical motivations for the use of non-linear kernel methods in machine learning applications. It is so termed after the information theorist Thomas M. Cover who stated it in 1965, referring to it as counting function theorem.
Data ecosystem	A data ecosystem is the complex environment of co-dependent networks and actors that contribute to data collection, transfer and use.
<i>In silico</i>	In biology and other experimental sciences, an in silico experiment is one performed on a computer or via computer simulation software.
Girvan–Newman algorithm	The Girvan–Newman algorithm (named after Michelle Girvan and Mark Newman) is a hierarchical method used to detect communities in complex system
Hyperparameter	a hyperparameter is a parameter that can be set in order to define any configurable part of a model's learning process.
Lipinski's Rule of 5	Lipinski's Rule of 5 is a set of guidelines used in drug discovery to predict if a chemical compound has the properties needed for good oral bioavailability in humans.
Natural language processing	Natural language processing is the processing of natural language information by a computer.
Neural Network	In machine learning, a neural network (also artificial neural network or neural net) is a computational model inspired by the structure and functions of biological neural networks.
Rogerian argument	Rogerian argument (or Rogerian rhetoric) is a rhetorical and conflict resolution strategy based on empathizing with others, seeking common ground and mutual understanding and

	learning, while avoiding the negative effects of extreme attitude polarization
Synthetic Nervous System	Synthetic Nervous System is a computational neuroscience model that may be developed with the Functional Subnetwork Approach to create biologically plausible models of circuits in a nervous system.
Transformer Architecture	transformer is an architecture based on the multi-head attention mechanism, in which text is converted to numerical representations called tokens, and each token is converted into a vector via lookup from a word embedding table
Transformative Technology	Transformative technology refers to advanced, innovative, and science-based hardware or software that produces fundamental, positive changes in human experience, society, or existing systems by disrupting established norms and creating new capabilities..

**COMMITTEE ON CARCINOGENICITY OF CHEMICALS IN FOOD, CONSUMER
PRODUCTS AND THE ENVIRONMENT**

Horizon scanning update – November 2025

COMEAP SETE evaluations on nano- and microplastic air pollution and traffic-related air pollution.

These papers are attached for Members; they are not reproduced here for copyright reasons, but can be obtained from: [COMEAP: Statement on airborne nano- and microplastic particles and fibres - GOV.UK](#)

**Secretariat
November 2025**



Working paper for COMEAP ‘Statement on airborne nano- and microplastic particles and fibres’

Interim assessment for the Synthesis and Integration of Epidemiological and Toxicological Evidence (SETE) for the population health effects from the inhalation of environmental airborne nano- and microplastic particles and fibres (NMPs).

Approach

1. This paper presents a provisional assessment of the strength of evidence for a risk to human health from inhalation exposure to current environmental levels of nano-microplastics (NMPs).
2. This assessment has used the framework described in a report of the Joint COT and COC Synthesis and Integration of Epidemiological and Toxicological Evidence subgroup (SETE), which reviewed approaches for synthesising and integrating epidemiological and toxicological evidence¹. COMEAP discussed the application of this framework to its work at the May and November 2022 COMEAP meetings, details of which can be found in the meeting minutes². Discussion points included that it may be more difficult to apply the approach to a complex mixture, such as particulate air pollution, than a well-defined chemical entity. Additionally, COMEAP’s approach to integrating epidemiological and toxicological evidence may be different to that used in other chemical risk assessment settings: it was suggested that COMEAP interpreted the axis “epidemiological evidence for causation” as the strength of epidemiological evidence for a risk to health and the axis “experimental evidence for causation” as the strength of experimental evidence for a risk to health. Following discussion at the COMEAP meeting held in March 2025, the labelling of the axes has been amended to make it clear that the evaluation is of a risk to population health from current environmental exposures.

¹ [SETE | Committee on Toxicity \(food.gov.uk\)](https://www.food.gov.uk/committees/committee-on-toxicity)

² Minutes of COMEAP meetings are available at: [Committee on the Medical Effects of Air Pollutants](https://www.food.gov.uk/committees/committee-on-the-medical-effects-of-air-pollutants).

3. The SETE approach requires that the integration of evidence, and visualisation, reflect the considered views of all of those evaluating the evidence, as discussed at each stage of the review process. This SETE assessment has been developed following an evaluation of the evidence as discussed in the Statement. The epidemiological and mechanistic evidence reviewed is not comprehensive and, therefore, the assessment of health risk should be considered provisional.

4. The diagram shown provides a means of visually indicating the consensus view of the Committee on the overall strength of the epidemiological and experimental (mechanistic) evidence that the inhalation of current levels of environmental NMPs poses a risk to human health. The diagram is not intended to reflect a probabilistic or numerical approach but, rather, it provides a representation of how the different lines of evidence assessed in the statement influence the strength of the overall conclusion on risk. To provide context, the assessment and diagram could be compared to other SETE assessments. For example, assessments of the health risks from current environmental, inhalation exposure to traffic related air pollution (TRAP).

Lines of evidence

5. Detecting and quantifying airborne NMPs is difficult due to limitations in current analytical methods. There is, therefore, limited data on the concentrations, and characteristics, of NMPs in the size fractions that are relevant for inhalation exposure and deposition in the lung. As a result, there is a lack of epidemiological studies on the effects of short- and long-term inhalation of environmental levels of NMPs on human health.

6. There is some evidence from occupational studies that exposure to high concentrations of NMPs, much greater than levels experienced by the general population, can increase the risk of restrictive (fibrotic) lung disease.

7. Currently, there is a lack of good quality toxicological studies in the literature using well characterised NMP particles, validated reproducible methods, and using other particles with similar physicochemical properties for comparison. Most toxicity studies have been performed using pristine particles, mainly polystyrene, which do not represent plastic particles in the environment. These pristine polystyrene spheres may not represent suitable model particles and would be unsuitable for assessing the health risks associated with exposure to polystyrene NMPs in the environment. Most studies used inappropriately high exposure concentrations of NMPs with inadequate characterisation, meaning that their relevance to real-world exposures is limited, preventing a meaningful consideration of relevant toxicological pathways and the potential human health impact. In addition, most studies do not report the

toxicological effects of NMPs in comparison with other PM components, at an equivalent dose.

8. There is a lack of research on the uptake, distribution, persistence and elimination of NMPs and their dosimetry within the human body. There are a limited number of studies reporting the presence of NMPs in human lung tissue. Studies reporting large NMPs in tissues and biological samples of a size significantly greater than 1 µm are contrary to the current understanding of how particles are transported within the body. In addition, for the majority of these studies, it is unclear whether translocation of the NMPs detected would have occurred in the lung following inhalation or the gastrointestinal tract following ingestion, which is more likely.

9. Overall, there is currently insufficient epidemiological and toxicological evidence to provide an informative assessment of the risk to health from inhalation exposure to NMPs in the environment. Further research is needed to understand exposure and the potential health effects associated with inhaled NMPs to better inform risk assessment.

Table 1: Summary of the strengths and weaknesses of the data examined for health effects from the inhalation of airborne nano- and microplastics (NMPs) and the influence of the lines of evidence on the overall conclusion.

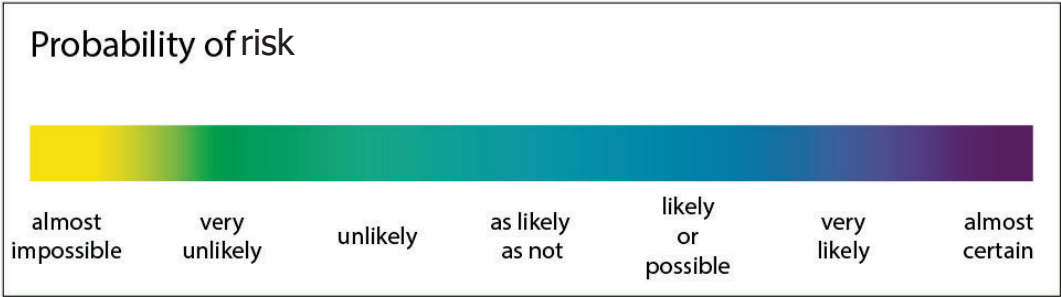
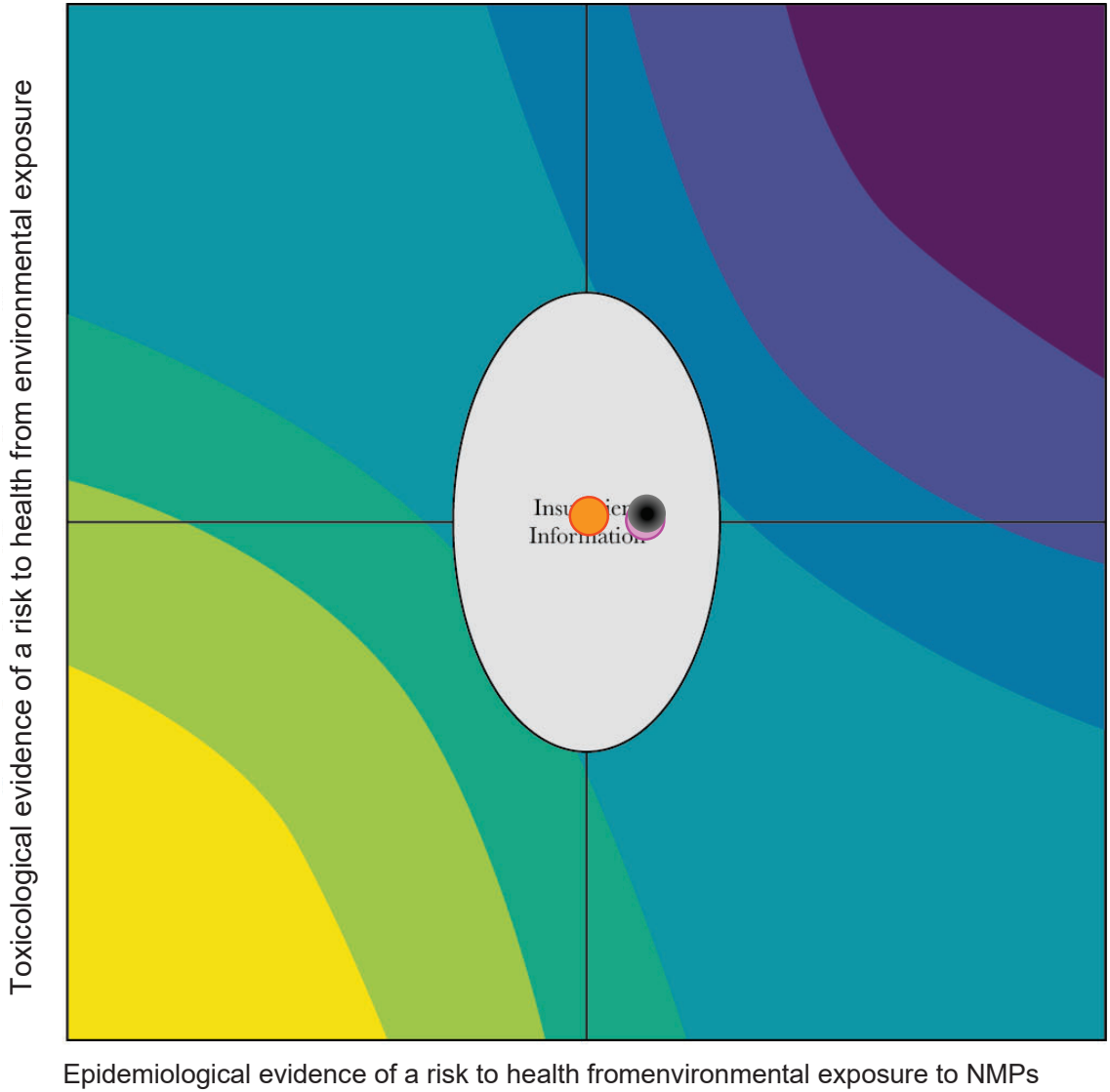
Lines of evidence and their main strengths (S) and weaknesses (W)	Influence on Conclusion
<p><u>Epidemiological data</u></p> <p>S – There are numerous epidemiological studies of workers exposed to NMPs in the plastics and textile industries. There is some evidence of reduced pulmonary function and specific lung pathology, such as interstitial lung disease, from these occupational studies.</p> <p>W – Studies of occupational exposure are based on workers exposed to extremely high concentrations, much greater than ambient levels to which the general population might be exposed.</p> <p>W – There is currently a lack of data on the concentrations and characteristics of airborne NMPs, in the size fractions that are relevant for inhalation exposure, to accurately measure and assess exposure. Without better exposure data epidemiological studies describing the association between the inhalation of airborne</p>	<p>There is evidence of hazard following prolonged high inhalation exposure to certain types of microplastics. However, there is currently a lack of data on the concentrations and characteristics of airborne NMPs, in the size fractions that are relevant for inhalation exposure, to accurately measure and assess exposure. Without this data, meaningful epidemiological studies of the relationship between environmental exposure and potential health effects are not possible.</p>

<p>NMPs and health effects are not possible to interpret.</p>	
<p><u>Mechanistic data</u></p> <p>W – There is a lack of data on the effects of inhaled microplastics in mammalian species and their retention in the lung is unclear.</p> <p>W - Most studies investigating NMP toxicity to date have used pristine, polystyrene spheres. However, there are significant physiochemical differences between polystyrene nano and microspheres made for biological and analytical studies and environmentally generated NMPs. Therefore, these may not represent suitable model particles and would be unsuitable for assessing the health risks associated with inhalation exposure to NMPs in the environment.</p> <p>W/S - Some limited evidence from toxicity studies to suggest that NMPs deposited in the lung induce oxidative stress, inflammation and cytotoxicity. These toxic effects resemble those induced by other solid and insoluble particles, however, none of the reported studies directly compared NMP particles with known particulate matter pollutants.</p> <p>W – Most studies used inappropriately high exposure concentrations of NMPs with inadequate characterisation of both the particles and of the adverse outcome pathway, meaning that their relevance to real-world inhalation exposures is limited</p> <p>W – There are a limited number of studies reporting NMP particles and fibres in human lung tissue. Studies reporting large MPs in tissues and biological samples of a size significantly greater than 1 µm are contrary to the current understanding of how particles are transported within the body. In addition, for the majority of these studies it is unclear whether translocation of the MPs detected would have occurred in the lung following inhalation or the gastrointestinal tract following ingestion, which is more likely.</p> <p>W/S – it may be possible to read across toxicological effects of NMPs based on</p>	<p>Due to limitations in current analytical methods, there are limited data on the concentrations, and characteristics, of NMPs in the size fractions that are relevant for inhalation exposure and deposition in the lung.</p> <p>Most studies investigating the inhalation toxicity of NMPs use pristine, polystyrene spheres, and many use inappropriately high exposure concentrations with inadequate characterisation. More data is needed on the effects of size, shape, chemical composition and other factors from exposure to real-world NMPs, at environmentally relevant concentrations, and in comparison, with other types of particles and fibres with similar properties.</p> <p>There is insufficient data on the fate of inhaled NMPs within the human body including their potential to accumulate in organs and tissues.</p>

knowledge of other airborne particles with similar physicochemical properties, such as, ultrafine particles (UFP), diesel exhaust particles (DEPs), silica and asbestos. However, there is a lack of clarity and understanding of the actual physical (and chemical) characteristics of NMPs to which humans could become exposed.	
<u>Conclusions on risk to health</u>	<p>Currently, there is a lack of evidence for the level of association between inhalation exposure to NMPs in the environment and the risk of adverse health effects.</p> <p>There is insufficient data quantifying and characterising NMP exposure in air to carry out meaningful environmental epidemiological studies.</p> <p>There is a lack of good quality toxicological studies in the literature using well characterised, representative NMPs, validated reproducible methods, and using other particles with similar physicochemical properties for comparison.</p>

10. The diagram is a visual representation of the consensus view of the Committee on the overall strength of the epidemiological and experimental (mechanistic) evidence that inhalation of current levels of environmental NMPs pose a risk to human health. The axes do not portray probabilistic or numerical estimates but, rather, reflect the views of the Committee on how the different lines of evidence assessed in the statement influence the overall conclusion on risk. To provide context, the assessment and diagram could be compared to other SETE assessments. For example, assessments of the health risks from current environmental, inhalation exposure to traffic related air pollution (TRAP)

Figure 1: Interim assessment and visualisation of the risk to population health from environmental, inhalation exposure to NMPs.



The pink circle is representative of all of the epidemiological evidence assessed (both the environmental and occupational evidence); the orange circle of all of the toxicological evidence assessed. The black circle represents the conclusion of the risk to health from

integrating the evidence. The circles are in the grey area in the centre of the figure indicating that there is currently insufficient information on the epidemiology and toxicological mechanisms to inform a conclusion. However, the position of the circles will change as more evidence becomes available. For comparison, the assessment and diagram should be compared to an assessment of a risk to health from environmental, inhalation exposure to traffic related air pollution (TRAP).

**COMEAP Airborne nano- and microplastics drafting group
September 2025**



Working paper for COMEAP ‘Statement on airborne nano- and microplastic particles and fibres’

Interim assessment for the Synthesis and Integration of Epidemiological and Toxicological Evidence (SETE) for the population health effects from the inhalation of traffic-related air pollution (TRAP).

Approach

1. This paper presents an assessment of the strength of evidence for health risks in humans due to inhalation exposure to current environmental levels of traffic-related air pollution (TRAP). The assessment has used the COT/COC framework for the Synthesis and Integration of Epidemiological and Toxicological Evidence (SETE), and was undertaken to provide context for the SETE assessment of the health effects from inhalation exposure to current environmental levels of nano-microplastics (NMPs). Unlike NMPs, there is ample evidence available on the risk from exposure to TRAP.
2. This SETE assessment draws on a number of previous COMEAP papers: the ‘Statement on the differential toxicity of particulate matter according to source or constituents: 2022’¹; ‘The evidence for health effects associated with exposure to non-exhaust particulate matter from road transport’²; and the ‘Statement on the evidence for the effects of nitrogen dioxide on health’³. Full details of the evidence used in this assessment can be found in these reports.
3. The SETE framework is described in a report by the Joint COT and COC Synthesis and Integration of Epidemiological and Toxicological Evidence subgroup (SETE), which reviewed approaches for synthesising and integrating epidemiological

¹ [Statement on the differential toxicity of particulate matter according to source or constituents: 2022](#)

² [COMEAP statement on the evidence for health effects associated with exposure to non-exhaust particulate matter from road transport](#)

³ [Statement on the evidence for the effects of nitrogen dioxide on health](#)

and toxicological evidence⁴. COMEAP discussed the application of this framework to its work at the May and November 2022 COMEAP meetings, details of which can be found in the meeting minutes⁵. Discussion points included that it may be more difficult to apply the approach to a complex mixture, such as particulate air pollution, than to a well-defined chemical entity. Additionally, COMEAP's approach to integrating epidemiological and toxicological evidence may be different from that used in other chemical risk assessment settings: it was suggested that COMEAP interpreted the axis "epidemiological evidence for causation" as the strength of epidemiological evidence for a risk to health and the axis "experimental evidence for causation" as the strength of experimental evidence for a risk to health. Following discussion at the COMEAP meeting held in March 2025, the labelling of the axes has been amended to make it clear that the evaluation is of a risk to health from current environmental exposures.

4. The SETE approach requires that the integration of evidence, and visualisation, reflect the considered views of all of those evaluating the evidence, as discussed at each stage of the review process. This SETE assessment has been developed following an evaluation of the evidence (as described in paragraph 2 above). The epidemiological and mechanistic evidence reviewed is not comprehensive and, therefore, the assessment of health risk should be considered provisional.

5. The diagram shown provides a means of visually indicating the consensus view of the Committee on the overall strength of the epidemiological and experimental (mechanistic) evidence that the inhalation of current levels of environmental TRAP poses a risk to human health. The diagram is not intended to reflect a probabilistic or numerical approach but, rather, it provides a representation of how the different lines of evidence assessed influence the overall conclusions on risk.

Lines of evidence

6. It is not practical or feasible to measure all of the components of the traffic pollutant mix, therefore, surrogates of traffic-related pollution have been used for assessing the contribution of traffic emissions to ambient air pollution and for estimating traffic exposure. The most commonly used traffic-pollutant surrogates include CO, NO₂, elemental carbon (EC) (or black carbon (BC), or black smoke [BS]), particulate matter (PM), benzene, and ultrafine particles (UFP). However, none of these pollutants is unique to emissions from motor vehicles.

⁴ [SETE | Committee on Toxicity \(food.gov.uk\)](https://www.food.gov.uk/committee-on-toxicity)

⁵ Minutes of COMEAP meetings are available at: [Committee on the Medical Effects of Air Pollutants](https://www.food.gov.uk/committee-on-the-medical-effects-of-air-pollutants).

7. The BC/EC-content of particulate matter has been the subject of extensive research and is often used as an indicator of PM from combustion, for example, from traffic exhaust (or in some countries from coal burning).

8. Specific sources of PM such as diesel exhaust are rich in ultrafine particles. For example, in controlled conditions, more than 85% of the particle numbers in diesel exhaust were ultrafine. Sources such as those from vehicle exhaust, may contain a number of harmful constituents (for example organic carbon compounds, constituents with high oxidative capacity) that could increase their potential toxicity. UFPs also have a greater surface area per unit mass, which some studies have suggested as important in driving toxicity, alongside UFP surface reactivity and number.

Table 1: Summary of the strengths and weaknesses of the data examined for health effects from the inhalation of traffic-related air pollution (TRAP) and the influence of the lines of evidence on the overall conclusion.

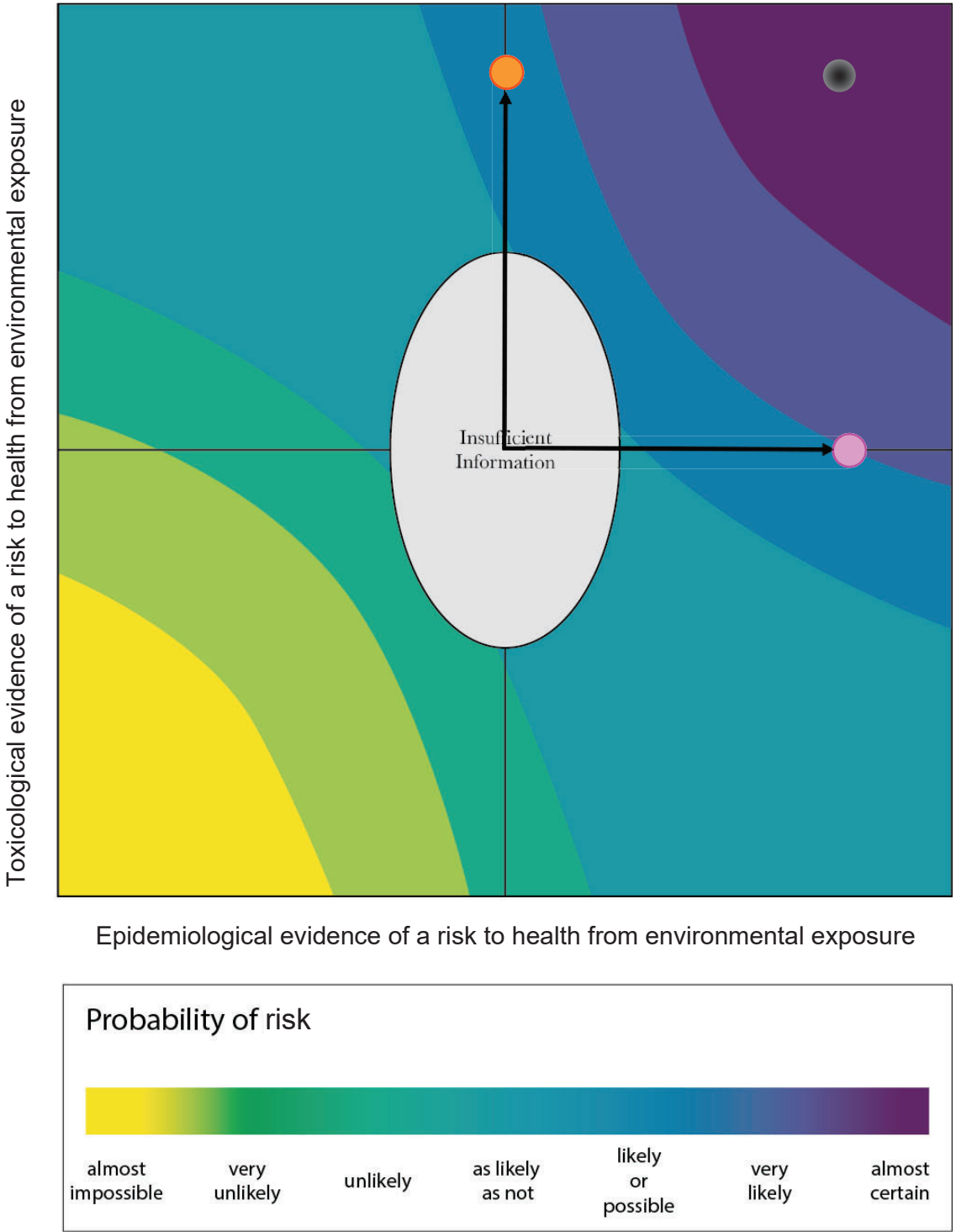
Lines of evidence and their main strengths (S) and weaknesses (W)	Influence on Conclusion
<p><u>Epidemiological data</u></p> <p>S – There is evidence linking traffic pollutants to adverse health parameters or outcomes. The highest levels of evidence in humans are for short-term exposure to traffic-related BC (mortality, respiratory and cardiovascular) and road dust (respiratory).</p> <p>S – There is evidence to support an association between exposure to TRAP and respiratory disease. For example, there are a number of studies that show a consistent positive association with exacerbation of childhood asthma. There is some evidence to suggest an association between exposure to TRAP and onset of childhood asthma, non-asthma respiratory symptoms, and impaired lung function.</p> <p>S – There is consistent evidence for short-term exposure to NO₂ having direct effects on respiratory morbidity. In addition, there is some suggestive evidence of a positive (adverse) association between short-term exposure to NO₂ and hospital admissions for cardiovascular disease and all-cause mortality. These associations have been shown to be robust to adjustments for other pollutants.</p>	<p>There is strong evidence linking TRAP to adverse health effects, particularly short-term exposure to black carbon (BC) and road dust affecting respiratory and cardiovascular health. TRAP is consistently associated with childhood asthma exacerbation and may contribute to its onset, non-asthma symptoms, and reduced lung function. The evidence is more limited linking TRAP to cardiovascular issues like heart rate variability and atherosclerosis. There is consistent, strong evidence of an association of both short and long-term exposure to NO₂ and a range of health effects. Research on non-exhaust particles and health effects remains sparse and inconclusive.</p>

<p>S – Studies of long-term exposure to NO₂ have reported associations with all-cause respiratory and cardiovascular mortality, children's respiratory symptoms and lung function.</p> <p>S/W – There is some but limited evidence that there is an association between exposure to TRAP and cardiovascular morbidity, for example, heart rate variability and atherosclerosis.</p> <p>W – A number of epidemiological studies have sought to measure associations between non-exhaust particle concentrations from road transport and adverse health outcomes. Taken as a whole, the current body of published work is small and does not provide a coherent and convincing narrative of adverse health effects of exposure to non-exhaust particles.</p>	
<p><u>Mechanistic data</u></p> <p>S/W - A modest amount of evidence exists from animal studies, confirming the capacity of vehicle exhaust to have detrimental biological effects.</p> <p>S/W - It has already been established that diesel exhaust can induce lung inflammation and adverse effects on the cardiovascular system in human controlled exposure studies. However, this approach has not been used to ascertain the biological effects of lower exposure regimes or directly compare different sources, constituents or sizes of PM.</p> <p>S – There is some evidence from experimental studies of adults and animals for airway hyperresponsiveness associated with long-term exposure to NO₂ and an allergic response from repeated long-term and short-term exposure.</p> <p>W – There is a general paucity of toxicological studies considering the potential health effects of the non-exhaust PM from road transport.</p>	<p>Evidence from human exposure studies have established that exposure to diesel exhaust can cause adverse respiratory and cardiovascular effects. Evidence from animal and cellular studies show that vehicle exhaust PM can cause inflammation and oxidative stress. There is experimental evidence for biological plausibility between NO₂ exposure and respiratory effects, particularly asthma exacerbation. There is limited data on low exposure regimes and non-exhaust PM.</p>
<p><u>Conclusions on risk to health</u></p>	<p>Many aspects of the epidemiological and toxicological evidence relating adverse human health effects to exposure to primary traffic-generated air pollution remain incomplete.</p>

	<p>Nonetheless, evidence from epidemiological and toxicological studies gives a high level of confidence that exposure to TRAP as a whole leads to adverse health effects. The highest levels of evidence of an adverse effect in humans are obtained for road traffic BC (respiratory health, cardiovascular health, which is less convincing, and all-cause mortality); and road dust (respiratory health). This is consistent with evidence for individual components, such as, black carbon, inorganic carbon and some metals. There is extensive evidence of the harmful effects of diesel engine exhaust, diesel PM and gasoline engine exhaust. There is also sufficient evidence of exposure to NO₂ resulting in respiratory effects, particularly exacerbation of asthma.</p>
--	---

9. The diagram is a visual representation of the consensus view of the Committee on the overall strength of the epidemiological and experimental (mechanistic) evidence that inhalation of current levels of environmental TRAP poses a risk to human health. The axes do not portray probabilistic or numerical estimates but, rather, reflect the views of the Committee on how the different lines of evidence assessed influence the overall conclusion on risk.

Figure 1: Interim assessment and visualisation of the risk to health from inhalation exposure to environmental levels of TRAP.



The pink circle is representative of all epidemiological evidence assessed; the orange circle of all toxicological evidence assessed. The black circle represents the conclusion of the risk to health of the integrated evidence.

COMEAP Airborne nano- and microplastics drafting group
September 2025