"You can never cross the ocean until you have the courage to lose sight of the shore."
Christopher Columbus (1451-1506)

"Progress is made by lazy men looking for easier ways to do things."
Robert A. Heinlein, American science fiction author (1907-1988)

Food for Thought ...

E-validation – Unleashing AI for Validation

Thomas Hartung[1,2], Alexandra Maertens[1] and Thomas Luechtefeld[1,3]

[1]Center for Alternatives to Animal Testing (CAAT), Doerenkamp-Zbinden-Chair for Evidence-based Toxicology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA; [2]CAAT-Europe, University of Konstanz, Germany; [3]ToxTrack LLC, Bethesda, MD, USA

Received September 21, 2024;

Correspondence:
Thomas Hartung, MD, PhD,
Center for Alternatives to Animal Testing (CAAT),
Johns Hopkins University,
615 N Wolfe St., Baltimore, MD, 21205, USA
(THartun1@jhu.edu)

ALTEX 41(4), 567-587.
doi:10.14573/altex.2409211

Abstract

The validation of new approach methods (NAMs) in toxicology faces significant challenges, including the integration of diverse data, selection of appropriate reference chemicals, and lengthy, resource-intensive consensus processes. This article proposes an artificial intelligence (AI)-based approach, termed e-validation, to optimize and accelerate the NAM validation process. E-validation employs advanced machine learning and simulation techniques to systematically design validation studies, select informative reference chemicals, integrate existing data, and provide tailored training. The approach aims to shorten current decade-long validation timelines, using fewer resources while enhancing rigor. Key components include the smart selection of reference chemicals using clustering algorithms, simulation of validation studies, mechanistic validation powered by AI, and AI-enhanced training for NAM education and implementation. A centralized dashboard interface could integrate these components, streamlining workflows and providing real-time decision support. The potential impacts of e-validation are extensive, promising to accelerate biomedical research, enhance chemical safety assessment, reduce animal testing, and drive regulatory and commercial innovation. While the integration of AI and machine learning offers significant advantages, challenges related to data quality, complexity of implementation, scalability, and ethical considerations must be addressed. Real-world validation and pilot studies are crucial to demonstrate the practical benefits and feasibility of e-validation. This transformative approach has the potential to revolutionize toxicological science and regulatory practices, ushering in a new era of predictive, personalized, and preventive health sciences.

Plain language summary

Validating new methods to replace traditional animal testing for chemicals can be slow and costly, often taking up to ten years. This article introduces e-validation, an artificial intelligence (AI)-powered approach designed to speed up and improve this process. By using advanced computer techniques, e-validation selects the best chemicals for testing, designs efficient studies, and integrates existing data. This approach would cut validation time and use fewer resources. E-validation includes a smart system for choosing test chemicals, virtual simulations to predict study outcomes, and AI tools to understand the biological effects of chemicals. It also provides training in these new methods. E-validation could accelerate medical research, improve chemical safety, reduce the need for animal testing, and help create safer products faster. While promising, this new approach will need real-world testing to prove its benefits and address potential challenges.

Glossary of terms and abbreviations

3Rs (replacement, reduction, refinement): Ethical principles aimed at minimizing the use of animals in research.

AI (artificial intelligence): The simulation of human intelligence processes by machines, especially computer systems.

BERT (Bidirectional Encoder Representations from Transformers): A powerful NLP model developed by Google that uses transformers to achieve high performance on a variety of language tasks.

Clustering algorithms: Machine learning techniques used to group similar data points into clusters for analysis.

Density-based spatial clustering of applications with noise (DBSCAN): An unsupervised machine learning algorithm used to identify clusters of varying shapes and densities in a dataset.

E-validation: An AI-based approach designed to optimize and accelerate the validation of new approach methods (NAMs) in toxicology, newly introduced in this article.

Extrapolation: The process of estimating unknown values by extending or projecting from known data.

Green toxicology: An approach that integrates principles of green chemistry and toxicology to design safer chemicals and processes that reduce environmental impact.

K-means clustering: A popular partitioning method used in machine learning to divide a dataset into k distinct, non-overlapping subsets (clusters).

LLM (large language models): Advanced AI models trained on vast amounts of text data, capable of understanding and generating human-like language.

ML (machine learning): A type of AI that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so.

Mechanistic validation: An approach to validation that focuses on understanding the biological pathways and molecular events that lead to toxic effects (introduced in Hartung et al., 2013b).

NAMs (new approach methods): Innovative methods for testing chemical safety that aim to replace or reduce animal testing.

NLP (natural language processing): A branch of AI that helps computers understand, interpret, and respond to human language.

PBPK (physiologically based pharmacokinetic) models: Computational models that simulate the absorption, distribution, metabolism, and excretion of chemicals in the body.

qAOP (quantitative adverse outcome pathway): A framework that quantitatively links molecular-level events to adverse health outcomes, aiding in risk assessment and regulatory decision-making.

QSAR (quantitative structure-activity relationship): Computational models that predict the effects of chemical structure on biological activity.

Simulation studies: Virtual experiments using computational models to predict the outcomes of real-world studies.

t-SNE (t-distributed stochastic neighbor embedding): A machine learning algorithm for dimensionality reduction, often used for visualizing high-dimensional data.

UMAP (uniform manifold approximation and projection): A dimensionality reduction technique for visualizing complex data sets.

1 Introduction

Validation of new approach methods (NAMs) to replace animal testing faces numerous challenges that hinder their adoption and use in regulatory contexts. These include difficulty finding relevant reference data, lack of appropriate reference chemicals, complex integration of existing validation data, insufficient or inappropriate chemical testing in studies, and resource-intensive, lengthy validation processes. These obstacles create a "valley of death" between NAM development and regulatory acceptance, stifling innovation in human-relevant approaches.

To address these challenges, we propose an artificial intelligence (AI)-based approach called "e-validation" to optimize and accelerate NAM validation. This transformative improvement in speed, cost, and rigor could finally unlock the long-dormant potential of human-relevant NAMs and integrated testing strategies.

2 Key components of e-validation

E-validation integrates five main AI/ML components to facilitate validation (Fig. 1):

I. Smart selection of reference chemicals
II. Simulation of validation studies
III. Mechanistic validation: AI-powered assessment of biological relevance
IV. AI-enhanced training: Revolutionizing NAM education and implementation
V. A centralized dashboard interface guides toxicologists through streamlined workflows to coordinate validation studies by integrating these components

2.1 Smart selection of reference chemicals

Stephen Hawking famously stated, "*The greatest enemy of knowledge is not ignorance, it is the illusion of knowledge*." Too often, we base validation studies on what we believe to know, e.g., which substances are the true positives or negatives for a given hazard. A critical challenge in validating NAMs is the selection of such reference chemicals (Hoffmann et al., 2008; Petersen et al., 2021), the benchmarks against which the perfor-

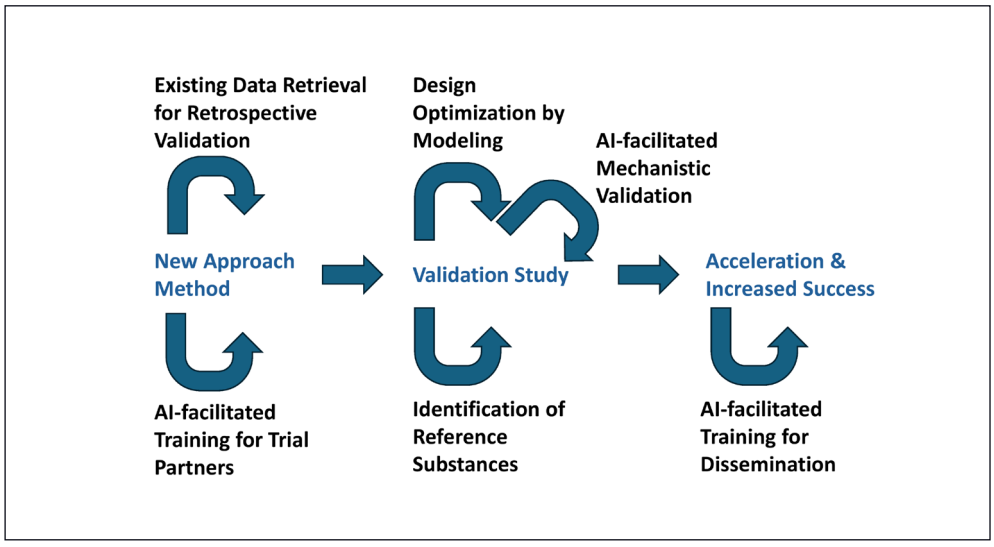

Fig. 1: The components of e-validation

Tab. 1: Advisable properties of the chemicals/substances used in the different validation modules
Modified from Hoffmann et al. (2008), with permission.

Validation module	Chemical/substance type (used in the respective module)	Advisable properties of the chemicals/substances used in the respective module[a]
1) Test definition	Controls	**Stability**; **continuous availability**; selective toxicity; safe and practical handling; **homogeneity**; **purity**; **adequate potency within response range of test**; cost; **scientific soundness of choice**
2) Within-laboratory reproducibility	Test chemicals/substances with existing reference results	**Stability**; availability; **homogeneity**; characterization (identity(/(im)purity/ physico-chemical properties/use categories); **adequate potency within response range of test**; cost; scientific soundness of choice; safety information
3) Transferability	Test chemicals/substances with existing reference results	**Stability**; availability; **homogeneity**; characterization (identity/(im)purity/ physico-chemical properties/use categories); **adequate potency within response range of test**; cost; scientific soundness of choice; safety information
4) Between-laboratory reproducibility	Test chemicals/substances with existing reference results	**Stability**; availability; **homogeneity**; characterization (identity/(im)purity/ physico-chemical properties/use categories); **adequate potency within response range of test**; cost; scientific soundness of choice; safety information
5) Relevance	Test chemicals/substances with existing reference results	**Stability**; availability; **homogeneity**; **characterization (identity/(im)purity/ physico-chemical properties/use categories)**; adequate potency within response range of test; cost; **scientific soundness of choice**; **traceability**; **safety information; availability of (high-)quality reference results**
7) Performance standards	Reference chemicals/ substances	**Certification**, **historical result range**

[a] Highest priority properties for a given module are shown in bold. Module 6, the applicability domain, is not included as usually no dedicated testing is performed for this module.

mance of a NAM is evaluated. They should fulfill several criteria, which depend on where in the validation process the reference substance is being used (Tab. 1).

Traditional approaches to selecting reference chemicals often rely on convenience in sampling or expert opinion, which can lead to biased or incomplete evaluations. Especially when the same reference compounds are used over and over, they can lead to bias in model development. E-validation can address this challenge through the implementation of advanced clustering algorithms and machine learning (ML) techniques to suggest representative reference substances that provide comprehensive coverage of diverse toxicological mechanisms. AI is uniquely suited to retrieve information on these various properties and optimize selection.

2.1.1 The importance of diverse reference chemicals

The selection of reference chemicals must ensure testing across the NAM's full applicability domain. This is essential for several reasons:

a) *Comprehensive evaluation:* A diverse set of chemicals allows for the assessment of the NAM's performance across a wide range of toxicological mechanisms and chemical properties. Retrieving a larger pool of possible candidate substances from the scientific literature or databases with large language models (LLM) broadens choices for the test set.
b) *Avoiding bias:* Overrepresentation of certain chemical classes or mechanisms can lead to skewed validation results that do not reflect the NAM's true performance.
c) *Defining the applicability domain:* By testing a diverse set of chemicals, the boundaries of the NAM's applicability can be more accurately defined.
d) *Regulatory acceptance:* Demonstrating the NAM's performance across a broad chemical space increases confidence in its reliability and relevance for regulatory purposes.

We have demonstrated how AI can facilitate this process based on the example of identifying demyelinating reference compounds through an extensive systematic literature review (Chesnut et al., 2021). This review involved retrieving 5,223 articles from PubMed and using the SWIFT-Review (Sciome Workbench for Interactive computer-Facilitated Text-mining) to prioritize and categorize these articles based on predefined search filters. The AI-driven prioritization algorithm sorted the studies, allowing us to identify and review 143 relevant studies in detail. This process helped in the selection of nine potential test chemicals, of which four were chosen for further study in a brain organoid model to assess their impact on myelination.

2.1.2 Clustering algorithms for chemical selection

E-validation employs advanced unsupervised learning approaches, particularly clustering algorithms, to intelligently select reference chemicals. These methods include:

a) *Dimensionality reduction techniques:* Methods such as principal component analysis (PCA) or t-distributed stochastic neighbor embedding (t-SNE) are used to reduce the high-dimensional chemical descriptor space to a manageable number of dimensions while preserving important variability.
b) *Density-based spatial clustering:* Algorithms like DBSCAN (density-based spatial clustering of applications with noise) are applied to identify clusters of chemicals with similar properties in the reduced dimensional space.
c) *K-means clustering:* This method partitions the chemical space into k clusters, where k is chosen to balance between diversity and manageability of the reference set.
d) *Hierarchical clustering:* This approach creates a tree-like structure of chemical clusters, allowing for flexibility in selecting chemicals at different levels of similarity.

These clustering methods consider multiple descriptors simultaneously, including:

a) Structural features (e.g., molecular weight, number of rings, functional groups)
b) Physicochemical properties (e.g., logP, solubility, pK_a)
c) Biological activity profiles (e.g., *in vitro* assay results, known mechanisms of action)

By considering this multidimensional space, the clustering algorithms can identify chemicals that are truly representative of the diverse landscape of toxicological mechanisms.

2.1.3 Integration with public databases

The smart selection process in e-validation goes beyond mere clustering by integrating with public databases to enhance the quality and utility of the selected reference chemicals. This integration involves:

a) *Data availability checks:* The system queries public databases to ensure that sufficient toxicological and chemical property data are available for each potential reference chemical.
b) *Quality assessment:* Data quality metrics are applied to prioritize chemicals with robust, high-quality data.
c) *Mechanistic information:* Where available, known mechanisms of action are incorporated to ensure coverage of diverse toxicological pathways.
d) *Regulatory relevance:* Information on regulatory classifications and decisions is considered to align the reference set with regulatory needs.

This integration allows for the selection of chemicals that are not only structurally and mechanistically diverse but also well-characterized and relevant to regulatory contexts.

2.1.4 Balancing diversity and feasibility

While maximizing diversity is crucial, e-validation will have to also consider practical constraints in the selection process:

a) *Availability:* The system prioritizes chemicals that are commercially available in sufficient quantities for testing.
b) *Cost:* Extremely expensive chemicals may be deprioritized if suitable alternatives exist.
c) *Handling requirements:* Chemicals with extreme storage or handling requirements are considered carefully.
d) *Ethical considerations:* For *in vivo* validation components, e.g., refinement methods, chemicals with existing animal data are prioritized to minimize new animal testing.

2.1.5 Iterative refinement

The smart selection process is not a one-time event but an iterative procedure that can be refined as new data become available:

a) *Feedback loop:* As validation studies progress, performance data can be fed back into the selection algorithm to identify areas where additional reference chemicals may be needed.
b) *Gap analysis:* The system continuously assesses the coverage of the chemical space and can suggest additional chemicals to fill identified gaps.
c) *Incorporation of new data:* As new chemicals are characterized or new public data become available, the reference set can be updated to maintain its relevance and comprehensiveness. This is particularly relevant for retrospective validations (Hartung et al., 2004; Corvi et al., 2008) and hybrids of retro- and prospective studies.

2.1.6 Transparency and reproducibility

To ensure transparency and reproducibility, e-validation should provide detailed documentation of the selection process:

a) *Selection criteria:* All parameters and thresholds used in the clustering and selection processes are clearly documented.
b) *Data sources:* The specific databases and versions used are recorded.
c) *Algorithmic details:* The exact algorithms and their implementations are specified.
d) *Rationale for inclusion:* For each selected reference chemical, a clear rationale for its inclusion is provided.

This documentation not only supports regulatory acceptance but also allows for critical evaluation and improvement of the selection process over time.

2.1.7 Conclusion

The smart selection of reference chemicals is a cornerstone of the e-validation approach to accelerating and improving NAM validation. By leveraging advanced clustering algorithms, integrating diverse data sources, and balancing theoretical diversity with practical constraints, this component ensures that NAMs are evaluated against a truly representative set of chemicals. This comprehensive and unbiased evaluation is crucial for building confidence in NAMs and ultimately facilitating their adoption in regulatory and research contexts. As the field of computational toxicology continues to advance, this smart selection process will evolve, incorporating new data types and algorithmic approaches to further refine the identification of optimal reference chemical sets.

2.2 Simulation of validation studies

Simulation is designed to create virtual validation studies by modeling expected outcomes across a range of variables including chemicals, doses, timepoints, and other experimental parameters. The idea is to model different study designs similarly to the modeling of clinical trials. This powerful tool allows for rapid iteration of proposed study designs, enabling researchers to identify optimal combinations that balance ethical considerations (Hartung, 2024b), practical constraints, scientific rigor, and probability of success, e.g., statistical power.

This will require retrieval and data integration for a given NAM from literature and databases to reduce duplication of efforts. AI lends itself to such retrieval of publicly available information. Natural language processing (NLP) mines textual data to extract details on past uses of the NAM.

2.2.1 The need for simulation in validation

Traditional validation studies are often resource-intensive, time-consuming, and ethically challenging, particularly when animal testing is involved. The simulation approach addresses these issues by:

a) *Reducing animal use:* By simulating outcomes, fewer animals are needed for actual experiments in case of refinement alternatives.
b) *Saving time and resources:* Virtual studies can be conducted much faster and at a fraction of the cost of wet-lab experiments.
c) *Exploring a wider parameter space:* Simulations allow for testing of scenarios that might be impractical or impossible in real-world settings.
d) *Optimizing study design:* Iterative simulations can identify the most informative experimental conditions before any actual testing begins.

2.2.2 Core components of the simulation

The simulation approach for validation studies will have to integrate several sophisticated modeling approaches:

a) *Physiologically based pharmacokinetic (PBPK) models and* in vitro *biokinetics*
   PBPK models simulate the absorption, distribution, metabolism, and excretion (ADME) of chemicals in the body. Originally designed to predict actual organ concentrations of drugs given to a patient, they have since been adapted to toxicants and their environmental exposures, and in reverse for *in-vitro*-to-*in-vivo* extrapolation (IVIVE) (Bouvier d'Yvoire et al., 2007; Tsaioun et al., 2016; Hartung, 2017). Furthermore, the concept of PBPK has been applied to the fate of a chemical in an *in vitro* setting as *in vitro* biokinetics (Blaauboer, 2010; Hamon et al., 2015; Proença et al., 2019). These models:
   – Predict tissue or *in vitro* concentrations over time
   – Account for species differences in metabolism and physiology
   – Help translate *in vitro* concentrations to *in vivo* doses
b) *Quantitative structure-activity relationship (QSAR) models*
   QSAR models relate chemical structure to biological activity. In the simulation they:
   – Predict potential toxicity endpoints based on chemical structure
   – Estimate physicochemical properties relevant to toxicity
   – Help fill data gaps for chemicals with limited experimental data
c) *Systems biology models*
   These models simulate complex biological networks and pathways. They are used to:
   – Predict perturbations in biological systems caused by chemical exposure
   – Model dose-response relationships at the molecular and cellular levels
   – Integrate data from multiple biological scales (molecular, cellular, tissue, organ)
d) *Statistical and ML models*
   Advanced statistical and ML techniques are employed to:
   – Predict variability in responses across different experimental conditions
   – Identify potential confounding factors in study designs
   – Estimate the predictive performance of NAMs under various scenarios

2.2.3 Creating virtual validation studies

The simulation should use these components to create comprehensive virtual validation studies:

a) *Exploration of the chemical space*

- Simulates outcomes for a wide range of chemicals, including those selected by the smart selection of reference chemicals
- Explores how chemical diversity affects NAM performance

b) *Dose-response modeling*
- Models responses across a range of doses, from very low to very high
- Identifies optimal dose ranges for detecting effects
- Predicts potential non-monotonic responses

c) *Temporal dynamics*
- Simulates outcomes at multiple timepoints
- Helps determine optimal sampling times for detecting effects
- Models potential delayed or long-term effects

d) *Experimental design optimization*
- Varies parameters such as sample size, number of replicates, and experimental conditions
- Identifies designs that maximize statistical power while minimizing resource use

e) *Variability and uncertainty analysis*
- Incorporates Monte Carlo simulations to account for biological variability and measurement uncertainty
- Helps determine robust sample sizes and replication needs

2.2.4 Balancing constraints

A key promise of validation simulation is its ability to balance various constraints:

a) *Ethical considerations*
- Minimizes the resources needed by identifying the most informative experiments, with the lowest number of laboratories, test substances and replicates
- Prioritizes designs that use the fewest animals or least severe procedures when *in vivo* studies are necessary for refinement alternatives

b) *Practical constraints*
- Considers resource limitations (e.g., budget, time, available equipment)
- Accounts for feasibility of proposed experimental designs in real-world laboratory settings

c) *Scientific rigor*
- Ensures sufficient statistical power to detect meaningful effects
- Maintains broad applicability domain and chemical diversity
- Addresses potential confounding factors and sources of bias

2.2.5 Iterative refinement process

The simulation module enables an iterative approach to study design:

I. Initial design: Based on input parameters and constraints
II. Simulation: Run virtual studies using the initial design
III. Analysis: Evaluate simulated outcomes against validation goals
IV. Refinement: Adjust design parameters based on analysis
V. Re-simulation: Run updated virtual studies
VI. Optimization: Repeat steps III-V until an optimal design is achieved

This iterative process allows for rapid exploration of many possible designs, leading to optimized protocols before any wet-lab work begins.

2.2.6 Integration with other e-validation approaches

The simulation is tightly integrated with other e-validation components. It incorporates reference chemicals identified by the smart selection of reference chemicals and feeds optimized designs into the training for protocol development. Additionally, the simulation provides simulated data to the data integration system for comparison with real-world results and informs the mechanistic validation process by simulating pathway-level effects. Predictive toxicology tools, especially based on AI, can predict the outcome especially of *in vitro* tests because of their simpler make-up and often larger homogenous test sets, for example from robotized testing. While these *in silico* tools are not yet accepted to substitute for the *in vitro* test, they can already be used to model the likely outcome in a validation study.

2.2.7 Output and interpretation

The simulation promises rich, interpretable outputs. These include visualizations of predicted outcomes across various conditions, sensitivity analyses showing which parameters most strongly influence results, and power calculations for different study designs. The simulation also provides recommendations for optimal experimental protocols and estimates of resource requirements for different design options.

2.2.8 Limitations and future directions

While powerful, the simulation approach has limitations that are important to acknowledge. These include dependence on the quality of underlying models and data, inability to capture all possible biological complexities, and potential for overlooking unexpected effects or interactions.

Future developments will have to address these limitations through continuous update and refinement of underlying models, integration of new data types (e.g., multi-omics data) as they become available, and development of more sophisticated AI approaches to capture complex biological interactions.

2.2.9 Conclusion

Simulation represents a paradigm shift in validation study design. By creating virtual validation studies, it allows for the exploration of a vast experimental space, optimization of study parameters, and balancing of ethical, practical, and scientific constraints. This not only accelerates the validation process but also enhances its rigor and relevance. As computational models continue to improve and integrate more diverse data types, simulation will become an increasingly powerful tool in the validation of NAMs, ultimately contributing to more efficient, ethical, and scientifically robust toxicological assessments.

2.3 Mechanistic validation: AI-powered assessment of biological relevance

Mechanistic validation represents a paradigm shift in the evaluation of NAMs (Hartung et al., 2013b), moving beyond simple

correlation to establish a deeper understanding of biological relevance. The integration of AI, particularly LLMs and ML techniques, into this process marks a revolutionary advance in our ability to assess and establish the mechanistic basis of toxicological effects.

2.3.1 The imperative for mechanistic understanding

Traditional validation approaches often rely heavily on correlative evidence, comparing NAM results to those from conventional animal studies or human data. While useful, this approach has limitations:

- It may perpetuate biases and errors inherent in traditional methods.
- It provides little insight into the biological mechanisms underlying toxicity.
- It struggles to account for species differences in toxicological responses.

Mechanistic validation addresses these issues by focusing on the biological pathways and molecular events that lead to adverse outcomes (Leist et al., 2017). This approach aligns with the adverse outcome pathway (AOP) framework and supports the development of human-relevant toxicology.

2.3.2 AI-enhanced evidence-based approaches

E-validation leverages AI to supercharge evidence-based approaches to mechanistic validation. This involves:

a) *Systematic review automation*
   AI-powered tools can rapidly scan vast amounts of scientific literature, extracting relevant information on mechanisms of toxicity. This process, which might take human researchers months or years, can be accomplished in days or even hours.
   - NLP algorithms identify and extract key information from papers, including experimental methods, results, and conclusions.
   - ML classifiers categorize studies based on quality and relevance.
   - Network analysis tools map relationships between different mechanistic components.

b) *Bradford-Hill criteria assessment*
   The Bradford-Hill criteria provide a framework for assessing causal relationships in biological systems. AI enhances the application of these criteria:
   - *Strength of association:* ML algorithms quantify the strength of associations between molecular events and adverse outcomes across multiple studies.
   - *Consistency:* AI tools identify patterns of consistency (or inconsistency) across diverse data sets and experimental conditions.
   - *Specificity:* NLP and ML techniques help determine the specificity of molecular interactions and their outcomes.
   - *Temporality:* AI-powered time series analysis establishes the temporal relationships between events in toxicological pathways.
   - *Biological gradient:* ML models can detect and characterize complex dose-response relationships, including non-monotonic responses.
   - *Plausibility:* LLMs can assess the biological plausibility of proposed mechanisms based on current scientific understanding.
   - *Coherence:* AI systems integrate data from multiple sources to evaluate the coherence of mechanistic explanations with broader biological knowledge.
   - *Experiment:* AI can design and simulate experiments to test mechanistic hypotheses, prioritizing the most informative real-world studies.
   - *Analogy:* ML algorithms can identify analogous mechanisms across different chemicals or biological systems, enhancing our understanding of shared toxicological pathways.

c) *Knowledge synthesis*
   The application of LLMs to mechanistic validation represents a possible quantum leap in our ability to synthesize and interpret toxicological knowledge (Tetko et al., 2022; Lin and Chou, 2022; Rodríguez-Belenguer et al., 2023; Hartung, 2023a,b; Kleinstreuer and Hartung, 2024). LLMs, trained on vast scientific corpora, can:
   - Summarize the current state of knowledge on specific toxicological mechanisms
   - Identify gaps in mechanistic understanding
   - Generate hypotheses about potential mechanisms based on available data

d) *Cross-domain integration*
   LLMs excel at making connections across diverse scientific domains. In the context of mechanistic validation, they can:
   - Link molecular biology insights with toxicological outcomes
   - Integrate knowledge from pharmacology, biochemistry, and systems biology to inform toxicological mechanisms
   - Identify potential off-target effects or unexpected pathway interactions

e) *Temporal analysis of scientific progress*
   By analyzing the progression of scientific understanding over time, LLMs can:
   - Trace the evolution of mechanistic concepts in toxicology
   - Identify emerging trends and shifting paradigms in mechanistic understanding
   - Predict future directions in mechanistic toxicology research

f) *Natural language querying of complex datasets*
   LLMs enable researchers to interact with complex toxicological datasets using natural language queries. This democratizes access to mechanistic insights, allowing even non-specialists to explore toxicological mechanisms in depth.

2.3.3 ML for pathway mapping and prediction

Advanced ML techniques contribute to mechanistic validation in several ways:

a) *Pathway reconstruction*
   - Graph neural networks map complex interaction networks between genes, proteins, and metabolites involved in toxicological responses (Shah et al., 2021; Baranwal et al., 2020).
   - Unsupervised learning algorithms identify novel patterns and clusters in high-dimensional omics data, potentially revealing new mechanistic insights.

b) *Predictive modeling of pathway perturbations* (Costello and Martin, 2018)
   – Deep learning models trained on large toxicogenomic datasets can predict pathway-level effects of new chemicals (Tonoyan and Siraki, 2024).
   – Ensemble methods combine predictions from multiple models to improve robustness and account for uncertainty in mechanistic predictions.
c) *Transfer learning for mechanism generalization*
   – Transfer learning techniques allow models trained on data-rich toxicological domains to be applied to less-studied areas, accelerating mechanistic understanding across diverse chemical spaces (Luechtefeld and Hartung, 2017).

2.3.4 AI-driven integration of *in silico*, *in vitro*, and *in vivo* data to quantitative AOPs (qAOPs)

E-validation's mechanistic validation uses AI to seamlessly integrate data from diverse sources:
– *In silico* predictions of molecular interactions and pathway activations
– *In vitro* high-throughput screening data on molecular and cellular responses
– *In vivo* toxicological outcomes from animal studies and human data

AI algorithms weigh and synthesize these multi-scale data to build comprehensive mechanistic models that span from molecular events to organism-level outcomes. AI significantly enhances the development of qAOPs as pursued in the ongoing ONTOX project (Vinken et al., 2021; Corradi et al., 2022; van Ertfelde et al., 2023):
– ML models quantify relationships between key events in AOPs.
– Bayesian networks capture uncertainty and variability in AOP components.
– Reinforcement learning algorithms optimize the structure of AOPs based on available data.

2.3.5 Challenges and future directions

While AI brings unprecedented capabilities to mechanistic validation, challenges remain:
a) *Interpretability:* Ensuring that AI-derived mechanistic insights are interpretable and scientifically meaningful.
b) *Bias mitigation:* Addressing potential biases in training data or algorithm design that could skew mechanistic understanding. In particular, literature extracted information is prone to publication biases and the prejudice of existing knowledge.
c) *Uncertainty quantification:* Developing robust methods to quantify uncertainty in AI-derived mechanistic predictions. This calls for probabilistic risk assessment as discussed earlier (Maertens et al., 2022, 2024a).
d) *Regulatory acceptance:* Building confidence in AI-powered mechanistic validation among regulatory bodies and the broader scientific community. This is being addressed in the Implementation Moonshot Project for Alternative Chemical Testing (IMPACT) (Sillé et al., 2024) and requires multi-stakeholder communication (von Aulock et al., 2022).

Future developments need to address these challenges through:
– Development of explainable AI techniques tailored to toxicological mechanisms
– Creation of curated, high-quality datasets specifically for training mechanistic models
– Integration of expert knowledge with AI systems through advanced human-AI collaboration interfaces
– Standardization of AI-powered mechanistic validation protocols to enhance reproducibility and regulatory acceptance

2.3.6 Conclusion

The integration of AI, particularly LLM and advanced ML techniques, into mechanistic validation represents a transformative advance in toxicology. By rapidly synthesizing vast amounts of scientific knowledge, uncovering hidden patterns in complex datasets, and generating novel mechanistic hypotheses, AI empowers researchers to develop deeper, more nuanced understandings of toxicological mechanisms. This AI-enhanced approach to mechanistic validation not only accelerates the development and acceptance of NAMs but also paves the way for a more predictive, human-relevant toxicology. As these technologies continue to evolve, they promise to revolutionize our ability to assess chemical safety, ultimately leading to more effective protection of human health and the environment.

2.4 AI-enhanced training: Revolutionizing NAM education and implementation

AI can strongly impact education (Box 1), both the training of participating laboratories and their staff in validation studies and the dissemination of NAMs after successful validation. This dual impact on laboratory training and method dissemination is a critical aspect of the validation process for NAMs. During validation studies, participating laboratories and their staff undergo intensive training to ensure consistent and accurate implementation of the new methods in ring trials. This process not only enhances the reliability of the validation results but also creates a cadre of skilled professionals who become proficient in these innovative techniques. Consequently, these trained individuals serve as valuable resources for the broader scientific community, facilitating the transfer of knowledge and expertise. Following successful validation, the dissemination of NAMs benefits significantly from this network of trained professionals. They act as ambassadors for the new methods, providing practical insights and troubleshooting advice to other laboratories adopting these techniques. This hands-on experience proves invaluable in overcoming initial implementation challenges and accelerating the widespread adoption of NAMs across different research and regulatory settings. Thus, the validation process serves not only to verify the scientific validity of new methods but also as a crucial mechanism for building capacity and fostering the diffusion of innovative approaches in toxicology and safety assessment.

Box 1: AI's contributions to education

1. *Personalized learning:* AI enables adaptive learning systems that can tailor content and pacing to individual student needs.

2. *Intelligent tutoring:* AI-powered tutoring systems like Khanmigo[1] can provide personalized, interactive learning experiences using inquiry-based approaches.
3. *Writing assistance:* AI tools can serve as writing assistants, editors, and thought partners to enhance students' writing processes.
4. *Accessibility:* AI has potential to assist students with disabilities, support universal design principles, and enhance curriculum accessibility.
5. *Assessment:* AI can facilitate new forms of assessment, enabling more continuous evaluation of student learning.
6. *Administrative support:* AI can help automate administrative tasks, freeing up teacher time for instruction.
7. *Data-driven insights:* AI analytics can provide educators with deeper insights into student performance and learning patterns.
8. *Preparation for future:* Exposure to AI tools prepares students for a workforce increasingly shaped by these technologies.
9. *Expanded educational resources:* AI enables creation of new learning materials and experiences, like interactive simulations.

The training approach of e-validation represents a paradigm shift in how researchers and regulators are educated about NAMs and prepared to generate high-quality validation data. By leveraging cutting-edge AI technologies, this transcends traditional training approaches and can offer personalized, adaptive, and immersive learning experiences that are crucial for the successful implementation of NAMs in toxicology and safety assessment.

2.4.1 The imperative for advanced NAM training

As toxicology transitions from traditional animal-based methods to innovative NAMs, there is an urgent need for comprehensive and effective training. This can be evidenced by the enormous uptake of our COURSERA classes on *Toxicology for the 21st Century*[2] *and Evidence-based Toxicology*[3] (von Aulock et al., 2022), for which more than 20,000 active learners have enrolled over the last six years. The complexity and diversity of NAMs, coupled with the rapid pace of technological advancement, create unique educational challenges:

a) *Diverse learner backgrounds:* Trainees come from varied scientific disciplines, each requiring tailored educational approaches.
b) *Rapidly evolving methodologies:* NAMs are continually refined and updated, necessitating agile training solutions.
c) *Complex data interpretation:* Many NAMs generate high-dimensional data that require sophisticated analytical skills.
d) *Regulatory considerations:* Training must address the nuanced regulatory landscape surrounding NAM adoption.

2.4.2 AI-powered personalized learning paths

E-validation training should employ advanced AI algorithms to create personalized learning experiences. Adaptive learning algorithms use ML models to analyze learner performance and behavior, dynamically adjusting content difficulty and pacing. Reinforcement learning algorithms optimize the sequence of training modules for each individual, maximizing knowledge retention and skill development. NLP for content customization can analyze learner queries and responses to tailor explanations and examples to individual comprehension levels. Sentiment analysis of learner feedback continuously refines and improves course content. AI-driven skill gap analysis uses predictive models to assess learner profiles against required competencies for specific NAMs, identifying personalized skill gaps. Automated recommendation systems suggest targeted learning resources to address individual knowledge deficits.

2.4.3 Immersive and interactive learning environments

AI technologies enable the creation of highly engaging and effective learning environments. Virtual and augmented reality (VR/AR) simulations can use AI-powered VR environments to simulate complex laboratory procedures, allowing risk-free practice of NAM techniques. AR overlays provide real-time guidance and information during hands-on training sessions. AI-enhanced interactive case studies can use generative AI to create diverse, realistic case studies that adapt based on learner decisions. ML algorithms can analyze learner approaches to problem-solving, providing personalized feedback and suggestions. Intelligent tutoring systems can provide 24/7 support, answering questions and guiding learners through complex concepts. Natural language understanding allows these systems to engage in nuanced, context-aware dialogues with learners. While this is still visionary at this point, AI-facilitated learning is rapidly progressing, and this overall progress can be leveraged for this specific purpose by using the increasingly available foundational models and platforms.

2.4.4 Continuous learning and performance support

The training opportunities extend beyond initial education to provide ongoing support. AI-powered knowledge repositories can use NLP-driven search engines to allow researchers to quickly find relevant information from vast databases of NAM literature and protocols. ML algorithms continuously update these repositories with the latest research findings and best practices. Predictive maintenance of skills uses AI models to track individual skill decay over time, proactively recommending refresher training. Personalized microlearning modules deliver targeted skill reinforcement based on predicted knowledge gaps.

[1] https://kodexolabs.com/ai-in-education/
[2] https://www.coursera.org/learn/toxicology-21
[3] https://www.coursera.org/learn/evidence-based-toxicology

Real-time performance support utilizes AR systems to provide *in situ* guidance during actual NAM implementation, reducing errors and enhancing data quality. AI-powered chatbots offer immediate troubleshooting support for common issues encountered during NAM execution.

2.4.5 Collaborative learning and community building

AI facilitates and enhances collaboration among NAM practitioners. For example, intelligent discussion forums can employ NLP algorithms to moderate and categorize forum discussions, ensuring relevant information is easily accessible. Recommendation systems connect learners with similar interests or complementary expertise. AI-facilitated virtual workshops can use ML algorithms to optimize participant groupings for virtual breakout sessions. Real-time speech analysis provides facilitators with insights into participant engagement and understanding. Crowd-sourced knowledge validation involves AI systems aggregating and analyzing community-contributed content, validating its accuracy and relevance. Blockchain technology can ensure the integrity and traceability of crowd-sourced NAM knowledge.

2.4.6 Assessment and certification

The training can employ AI to ensure rigorous and fair assessment of NAM competencies. Adaptive testing uses AI-powered tests that adjust question difficulty based on learner responses, providing more accurate assessments of competency. ML models analyze response patterns to detect potential misconceptions or knowledge gaps. Performance-based assessment utilizes computer vision and ML to evaluate learner performance in virtual lab simulations, assessing practical skills alongside theoretical knowledge. NLP analyzes written reports and oral presentations for comprehensive evaluation of communication skills crucial for NAM implementation. Continuous competency monitoring involves AI models tracking on-the-job performance metrics, providing ongoing assessment of NAM proficiency. Predictive analytics flag potential skill degradation, triggering targeted interventions to maintain competency.

2.4.7 Future directions

There might be at some point ethical considerations and challenges, which include ensuring data privacy while leveraging learner data for personalized training, mitigating potential algorithmic biases that could disadvantage certain groups of learners, maintaining transparency about how AI is used in training and assessment, and balancing AI-driven automation with human expertise in training delivery and evaluation.

As AI technologies continue to evolve, the training aspects will expand its capabilities. This includes incorporating "emotion AI" to create more empathetic and responsive learning experiences, leveraging "quantum ML" to handle increasingly complex NAM datasets and simulations, exploring brain-computer interfaces for enhanced learning and skill acquisition, and developing advanced AI-human hybrid teaching systems that seamlessly blend AI capabilities with human expertise.

2.4.8 Conclusion

AI-enhanced training for e-validation represents a quantum leap in preparing researchers and regulators for the era of NAMs in toxicology. By harnessing the power of AI, this can deliver personalized, immersive, and continuously adaptive learning experiences that are essential for mastering the complexities of NAMs at scale. Benjamin Franklin put it nicely: "*Tell me and I forget. Teach me and I remember. Involve me and I learn*." The integration of AI will not only accelerate the learning process but also ensure that training remains current with the rapidly evolving field of NAMs. As these technologies continue to advance, the training will play a pivotal role in building a skilled workforce capable of fully leveraging NAMs to revolutionize toxicology and safety assessment. This AI-driven educational paradigm is not just about transferring knowledge but about cultivating a new generation of innovative, adaptable, and critically thinking scientists who will drive the future of human-relevant toxicology.

2.5 A centralized dashboard interface guides toxicologists through streamlined workflows to coordinate validation studies by integrating the components

The e-validation approach should be anchored by a centralized dashboard interface that serves as the command center for coordinating and executing validation studies. This intuitive, AI-powered interface can seamlessly integrate all the contributions of e-validation, providing toxicologists with a streamlined, end-to-end workflow management system. Ideally, this is combined with centralized institutions for the strategic development of safety sciences (Busquet and Hartung, 2017) or platforms like the Integrated Chemical Environment (ICE)[4] (Bell et al., 2017, 2020). The dashboard offers real-time visualization of study progress, automated data quality checks, and intelligent decision support based on ML analysis of ongoing results. It allows researchers to effortlessly navigate between different aspects of the validation process, from selecting reference chemicals and designing experiments to analyzing data and generating reports. The interface adapts to user roles and preferences, presenting the most relevant information and tools for each stage of the validation process. Advanced analytics and predictive modeling capabilities are embedded within the dashboard, enabling researchers to forecast study outcomes, identify potential bottlenecks, and make data-driven decisions to optimize the validation process. Furthermore, the dashboard facilitates collaboration by providing shared workspaces, version control for protocols and data, and integrated communication tools. Security features ensure data integrity and compliance with regulatory requirements. By centralizing control and providing a holistic view of the entire validation process, this intelligent dashboard significantly re-

[4] https://ntp.niehs.nih.gov/whatwestudy/niceatm/comptox/ct-ice/ice

duces the cognitive load on researchers, minimizes errors, and accelerates the overall timeline of validation studies, ultimately fast-tracking the adoption of NAMs in toxicology.

3 Scientific rationale

The scientific foundation for e-validation is built upon a robust body of evidence demonstrating the power of AI and advanced modeling approaches to revolutionize scientific work[5,6], which is applied here to the validation process for NAMs in toxicology. This rationale is supported by several key pillars of modern computational science and data analytics.

3.1 Predictive simulations and complex systems modeling

Predictive simulations of the complex dynamics between chemical exposures, biological mechanisms, and toxicological outcomes represent powerful tools for understanding whether NAMs correctly predict human toxicity. These simulations leverage advanced computational techniques such as physiologically based PBPK modeling, systems biology approaches, and ML algorithms to capture the intricacies of biological responses to chemical perturbations. By integrating diverse data types – from molecular interactions to organ-level effects – these models can provide a holistic view of toxicological processes that is difficult to achieve with traditional experimental methods alone.

Recent advancements in deep learning and artificial neural networks have significantly enhanced our ability to model complex biological systems. For instance, graph neural networks have shown remarkable success in predicting protein-protein interactions and drug-target binding, which are crucial for understanding toxicological mechanisms. Moreover, the integration of multi-omics data into these models allows for a more comprehensive representation of the biological state, enabling more accurate predictions of toxicity across diverse chemical spaces.

3.2 Advanced clustering and chemical space exploration

The application of advanced clustering algorithms to the challenge of selecting representative chemical subsets has demonstrated significant potential in enhancing the efficiency and effectiveness of validation studies. Techniques such as k-means clustering, hierarchical clustering, and DBSCAN have been successfully applied to large chemical databases to identify diverse yet representative sets of compounds. These methods consider multiple dimensions of chemical properties simultaneously, including structural features, physicochemical attributes, and known biological activities.

Recent innovations in dimensionality reduction techniques, such as t-distributed stochastic neighbor embedding (t-SNE) and uniform manifold approximation and projection (UMAP), have further enhanced our ability to visualize and analyze high-dimensional chemical spaces. This allows for more intuitive and effective selection of reference chemicals that span the relevant chemical and biological diversity needed for comprehensive validation of NAMs.

3.3 Natural language processing and knowledge extraction

NLP techniques[7] have emerged as powerful tools for unlocking valuable information from the vast scientific literature and unstructured data sources. NLP is enabling computers to understand and manipulate human language, allowing understanding document contents, extracting information, categorizing and organizing documents. They are rule-based, probabilistic (including statistical and neural network-based) ML methods. Advanced NLP models, such as BERT (bidirectional encoder representations from transformers)[8] and its derivatives, have demonstrated remarkable capabilities in understanding context and extracting relevant information from complex scientific texts (Kang et al., 2022). BERT is a language model leveraging transformer architecture for superior performance in language understanding introduced in 2018 by Google, which sparked the rise of LLMs. These models can be fine-tuned for specific toxicological domains, enabling the automated extraction of experimental protocols, results, and mechanistic insights from published studies.

The ability of NLP to parse and synthesize information from diverse sources addresses a critical challenge in toxicology: the integration of fragmented knowledge across multiple studies and data types. By automating the process of literature review and data extraction (Tonoyan and Siraki, 2024), NLP promise to significantly accelerate the compilation of existing evidence relevant to NAM validation, reducing the risk of overlooking crucial information and enabling more comprehensive assessments of NAM performance.

3.4 Iterative modeling and optimization

The application of iterative modeling techniques to optimize resource allocation in validation studies represents a significant advancement over traditional approaches. Techniques such as Bayesian optimization, reinforcement learning, and evolutionary algorithms enable the efficient exploration of vast experimental design spaces to identify optimal configurations that balance scientific rigor with practical constraints.

These methods can dynamically adjust experimental parameters based on accumulated data, allowing for adaptive study designs that maximize information gain while minimizing resource expenditure. For example, active learning algorithms can intelligently select the most informative experiments to perform next, based on the current state of knowledge and uncertainty. This approach is particularly valuable in the context of NAM validation, where the space of possible experiments is often too large to explore exhaustively.

[5] Fink, F., Hartung, T., Lee, S. Y. et al. (2024). AI for scientific discovery pioneering new frontiers in knowledge. In World Economic Forum, Top 10 Emerging Technologies of 2024, Flagship Report. https://www.weforum.org/publications/top-10-emerging-technologies-2024/in-full/1-ai-for-scientific-discovery/

[6] http://wef.ch/aiforscience

[7] https://www.ibm.com/topics/natural-language-processing

[8] https://towardsdatascience.com/bert-explained-state-of-the-art-language-model-for-nlp-f8b21a9b6270

3.5 E-validation's integration of AI pillars

A seamless integration of the AI pillars within e-validation would create a synergistic approach that is uniquely positioned to transform the validation process for NAMs. The combination of these technologies addresses multiple challenges simultaneously:

3.5.1 Comprehensive chemical space coverage

By leveraging advanced clustering algorithms for smart chemical selection, e-validation ensures that validation studies cover a meaningful diversity of chemical structures and properties. This approach not only enhances the robustness of validation outcomes but also helps define the applicability domain of NAMs more precisely. The pruning of redundant chemicals through intelligent selection algorithms optimizes resource utilization without compromising the breadth of validation.

3.5.2 Balancing rigor and feasibility

The virtual simulation capabilities of e-validation allow for the rapid exploration of numerous experimental scenarios, identifying those that strike an optimal balance between scientific rigor and practical feasibility. This approach enables researchers to anticipate potential challenges, optimize study designs, and make informed decisions about resource allocation before initiating costly and time-consuming wet-lab experiments.

3.5.3 Knowledge integration

The text mining and NLP components of e-validation mitigate duplicate efforts by efficiently extracting and synthesizing relevant information from existing literature and databases. This not only saves time and resources but also ensures that validation studies extend the current state of knowledge rather than inadvertently repeating work that has already been done.

3.5.4 Quality assurance in data generation

The training modules and guided workflows provided by e-validation establish an optimal infrastructure for generating high-quality validation data. By standardizing protocols, providing interactive training, and offering real-time guidance, these components help ensure consistency and reliability in data generation across different laboratories and studies.

3.5.5 Advanced data interpretation

E-validation's AI-powered analytics tools facilitate sophisticated interpretation of complex validation data. ML algorithms can identify subtle patterns and relationships that might be missed by traditional statistical approaches, leading to more nuanced and comprehensive assessments of NAM performance.

3.5.6 Adaptive and iterative validation strategies

The platform's ability to continuously learn and adapt based on incoming data enables the implementation of iterative validation strategies. This dynamic approach allows for real-time optimization of study designs and resource allocation, maximizing the efficiency and effectiveness of the validation process.

3.5.7 Conclusion

In conclusion, the scientific rationale underlying e-validation is grounded on the proven capabilities of AI and advanced modeling techniques to address the complex scientific challenges translatable to NAM validation. By integrating these powerful computational approaches, e-validation promises a transformative solution to accelerate the development, validation, and adoption of innovative, human-relevant methods in toxicology. Such a platform could not only enhance the efficiency and rigor of the validation process but also paves the way for a more predictive and mechanistically informed approach to chemical safety assessment.

4 Novelty

The suggested e-validation approach represents a paradigm shift in the approach to validating NAMs in toxicology, marking a significant departure from traditional practices. While AI and ML techniques have been increasingly incorporated into the development of NAMs themselves, the application of these advanced computational approaches to strategize and optimize the validation process is a novel concept. This innovation addresses a critical gap in the current toxicological landscape, where the validation of new methods often remains a bottleneck in their adoption and regulatory acceptance.

4.1 Current paradigm and its limitations

The existing paradigm for designing and executing validation studies largely relies on trial-and-error approaches, expert opinion, and historical precedent (Hartung and Leist, 2008; Leist et al., 2008). This traditional methodology suffers from several key limitations:

a) *Inefficiency:* The trial-and-error approach often leads to suboptimal use of resources, with multiple iterations required to refine study designs.
b) *Limited scope:* Human experts, while knowledgeable, are constrained in their ability to consider vast amounts of data and complex interrelationships simultaneously.
c) *Bias:* Historical precedents may perpetuate biases and outdated practices, potentially limiting innovation in study design. Especially the bias toward animal data as point of reference represents a major limitation.
d) *Lack of adaptability:* Traditional approaches struggle to keep pace with the rapid evolution of NAMs and the increasing complexity of toxicological data.

4.1.1 E-validation's novel approach

In contrast, e-validation suggests a comprehensive approach that harnesses the power of simulation, optimization, and ML to chart efficient validation trajectories. This approach is complemented by integrated data retrieval and customized learning modules, creating a comprehensive ecosystem for NAM validation. The novelty of e-validation lies not just in its use of advanced technologies, but in its holistic, data-driven approach to reimagining the entire validation process. Specifically novel aspects include:

a) *Prospective simulation of validation outcomes*
The use of design of experiments (DoE) and systems modeling to prospectively simulate validation outcomes represents a groundbreaking approach in toxicology, similar to the modeling of clinical studies (Abbas et al., 2006; Malikova, 2016; Sverdlov et al., 2019)[9]. There are some interesting parallels between modeling clinical trials and modeling validation studies for NAMs (Box 2). The key differences are that NAM validation typically does not involve human subjects and aims to predict animal or human toxicology results rather than clinical outcomes. However, many of the statistical and modeling approaches could be adapted from the clinical trial domain to enhance NAM validation studies.

Box 2: Opportunities to translate clinical trial optimization to the optimization of NAM validation

1. *Study design optimization*
   - For clinical trials, optimal designs aim to maximize statistical power and efficiency while minimizing sample size and costs.
   - For NAM validation, optimal designs could similarly aim to maximize the information gained about a method's performance while minimizing resources required.
2. *Adaptive designs*
   - Adaptive clinical trial designs allow modifications based on interim data.
   - NAM validation could potentially use adaptive approaches to refine protocols or add/remove test compounds as data is collected.
3. *Simulation studies*
   - Clinical trial simulations are used to evaluate different design options and statistical analysis plans.
   - Simulations could be valuable for NAM validation to explore different validation strategies and predict outcomes.
4. *Dose-response modeling*
   - Modeling dose-response relationships is crucial in many clinical trials.
   - For NAMs, modeling concentration-response relationships is often important and could leverage similar statistical approaches.
5. *Population PK/PD modeling*
   - Population approaches model variability between subjects in clinical trials.
   - For NAMs, modeling variability between experimental runs or laboratories could be analogous.
6. *Multi-arm multi-stage designs*
   - Allow efficient comparison of multiple treatments in clinical trials.
   - Could be adapted to compare multiple NAM variants or evaluate NAMs across multiple endpoints.
7. *Bayesian methods*
   - Increasingly used in clinical trials to incorporate prior information.
   - Could be valuable for NAM validation to leverage existing knowledge about similar assays or chemical classes.
8. *Biomarker validation*
   - Clinical biomarker validation has established frameworks that could inform NAM validation approaches (Hartung et al., 2024).
9. *Benefit-risk assessment*
   - Balancing efficacy and safety is key in clinical trials.
   - For NAMs, balancing sensitivity, specificity, and other performance metrics may be analogous.

This aspect of e-validation allows researchers to explore a vast space of possible study designs and outcomes before committing resources to actual experiments. The novelty lies in the integration of multi-scale models, where e-validation combines PBPK models, systems biology networks, and ML predictors to create comprehensive simulations of toxicological processes. Unlike static simulation approaches, e-validation would employ reinforcement learning algorithms for dynamic optimization, continuously refining the validation strategy by adjusting experimental parameters based on simulated outcomes. Additionally, advanced Bayesian methods are used for uncertainty quantification, providing a more nuanced understanding of potential risks and benefits associated with different validation approaches.

b) *AI-driven chemical selection*
The selection of representative chemicals using clustering algorithms that balance diversity and feasibility is a novel application of AI in validation planning. This approach goes beyond simple structural or property-based clustering by employing multi-objective optimization. E-validation could use evolutionary algorithms to simultaneously optimize for chemical diversity, data availability, cost, and experimental feasibility. The platform could incorporate adaptive sampling, using active learning techniques to iteratively refine the selection of chemicals based on accumulated data and model performance, ensuring optimal coverage of the chemical space relevant to the NAM being validated. Unlike traditional approaches that focus solely on chemical properties, e-validation incorporates toxicological mode of action information into the clustering process, ensuring that selected chemicals are mechanistically relevant to the NAM under evaluation. This is similar to the information economics we discussed earlier in this series (Maertens et al., 2022), which means that AI can suggest what to test and what not based on the informative value and whether this really affects the overall assessment.

[9] https://trialkey.ai/blog/optimizing-clinical-trial-design-with-ai/

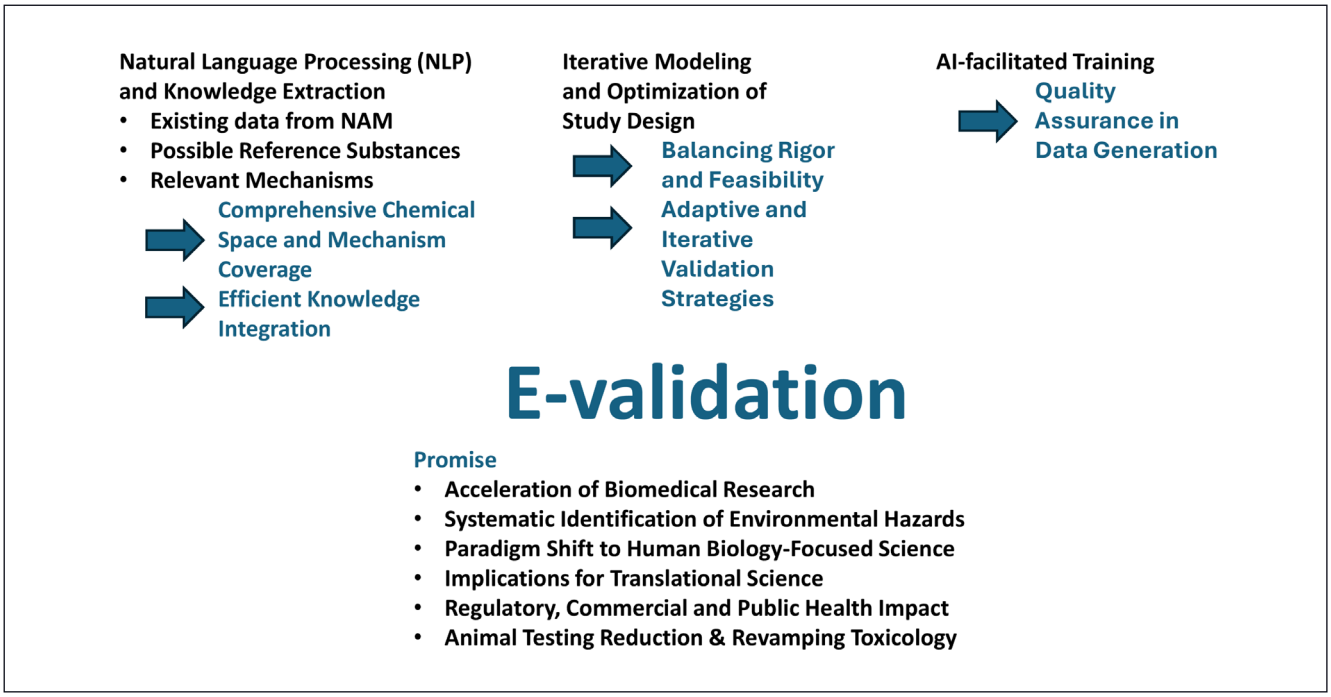

Fig. 2: The potential impact of e-validation

c) *Transparent data integration*
E-validation's approach to referencing extensively documented public data resources for transparency is novel in its scale and implementation. The platform should employ advanced NLP techniques for automated data curation, automatically extracting, validating, and integrating data from diverse public sources to ensure up-to-date and comprehensive coverage. E-validation might implement a novel blockchain-based system for data provenance, tracking the origin and transformations of all data used in validation studies, ensuring unprecedented levels of transparency and reproducibility. To address data privacy concerns, e-validation pioneers the use of federated learning techniques in toxicology, allowing models to be trained on distributed datasets without centralizing sensitive data.

d) *Intelligent literature mining*
The platform's approach to datamining published literature goes beyond simple keyword searches or citation analysis. E-validation needs to employ advanced NLP models capable of semantic understanding, extracting relevant information even when it is not explicitly stated by understanding context and nuance in scientific texts. By analyzing patterns across large corpora of toxicological literature, e-validation can generate novel hypotheses about potential validation strategies or unexplored mechanisms of toxicity. The platform could also automatically identify and codify experimental designs from published studies, building a knowledge base of validation strategies that can inform future studies.

e) *Personalized e-learning solutions*
E-validation's provision of bespoke e-learning solutions represents a novel approach to addressing skill deficiencies in the toxicology community. The vision is that AI creates adaptive learning paths, personalizing learning trajectories for each user by dynamically adjusting content difficulty and focus based on individual performance and learning styles. For this purpose, e-validation would need to incorporate cutting-edge virtual reality (VR) technology to provide immersive, hands-on training experiences for complex NAM protocols, a first in toxicology education. Unlike traditional training approaches, e-validation could employ continuous competency assessment through ongoing, AI-driven evaluations to ensure sustained competency in NAM implementation and data interpretation.

4.1.2 Paradigm shift in validation

E-validation heralds a new era where validation transitions from being a reluctantly tolerated Achilles heel to a strategically embraced core competency. This shift is characterized by:

a) *Proactive design:* Moving from reactive troubleshooting to proactive, data-driven study design.
b) *Continuous optimization:* Replacing fixed protocols with dynamically optimized validation strategies.
c) *Integrated knowledge management:* Shifting from siloed data to a holistic, interconnected knowledge ecosystem.
d) *Predictive power:* Transitioning from descriptive to predictive approaches in assessing NAM performance.

In conclusion, e-validation represents a revolutionary approach to NAM validation, introducing novel technologies and methodologies that promise to transform the field of toxicology. By reimagining validation as a data-driven, AI-powered process, e-validation has the potential to accelerate the adoption of innova-

tive, human-relevant methods in toxicology, ultimately leading to more efficient and effective protection of human health and the environment.

5 Potential impact

The potential impact of e-validation extends far beyond the realm of toxicology, promising to revolutionize biomedical research, environmental health, and public safety (Fig. 2). By radically condensing validation timelines and reducing associated costs, this innovative approach has the potential to catalyze transformative changes across multiple sectors and disciplines.

5.1 Acceleration of biomedical research

E-validation's promise to expedite the validation of NAMs could have a profound impact on biomedical research across a wide spectrum of diseases and health conditions. In the field of drug discovery and development, validated NAMs could enable rapid, high-throughput screening of potential drug candidates, significantly reducing the time and cost of early-stage drug discovery and frontloading safety assessments. By providing more human-relevant data, validated NAMs could increase the success rate of drug candidates in clinical trials, addressing the high attrition rates that currently plague the pharmaceutical industry. Additionally, NAMs based on human cells and tissues could facilitate the development of personalized therapies, taking into account individual genetic and physiological variations.

In terms of disease modeling, validated NAMs, particularly those involving 3D organoids and microphysiological systems, could provide unprecedented insights into complex diseases such as cancer, neurodegenerative disorders, and autoimmune conditions (Marx et al., 2016, 2020; Roth et al., 2019). For rare diseases where patient numbers are limited, validated NAMs could provide crucial tools for understanding disease mechanisms and testing potential therapies.

In the field of regenerative medicine, accelerated validation of NAMs could fast-track the development and safety assessment of stem cell-based therapies for a range of conditions, from spinal cord injuries to heart disease. Validated *in vitro* models could accelerate the development of engineered tissues and organs for transplantation, addressing critical shortages in donor organs. These advancements could potentially revolutionize the field of regenerative medicine, offering new hope for patients with previously untreatable conditions.

5.2 Systematic identification of environmental hazards

E-validation's impact on environmental health and safety could be transformative:

a) *Chemical safety assessment*
Rapid validation of NAMs could enable the systematic screening of thousands of chemicals currently in use but lacking comprehensive safety data. Validated NAMs could also provide new tools for assessing the combined effects of chemical mixtures, addressing a major gap in current risk assessment practices.

b) *Emerging contaminants*
Fast-tracked validation of NAMs could enable quicker assessment of emerging environmental contaminants, such as micro- and nanoplastics (Maertens and Hartung, 2019), allowing for timelier regulatory and remediation responses. Additionally, validated *in silico* models (Hartung, 2023a; Kleinstreuer and Hartung, 2024) could help predict the potential hazards of new chemicals before they enter the environment, supporting green chemistry initiatives (Crawford et al., 2017; Maertens and Hartung, 2018; Maertens et al., 2021, 2024b).

c) *Climate change and health*
As climate change alters environmental conditions, rapidly validated NAMs could provide agile tools for assessing new and evolving health risks, from altered pathogen distributions to changing air pollution profiles. The many connections between climate change and health effects of chemicals are often overlooked (Box 3). In summary, climate change is likely to exacerbate many existing chemical-related health risks while also creating new exposure scenarios and vulnerabilities that will require proactive assessment and management strategies[10,11,12] (Boxall et al., 2009; Balbus et al., 2013). The complex interactions between climate factors and chemical pollutants pose significant challenges for protecting public health in a changing environment.

Box 3: Connections between climate change and the health impacts of chemicals

1. *Altered chemical behavior and exposure*
   Climate change affects the fate and transport of chemicals in the environment through changes in temperature, precipitation, humidity, wind conditions, and erosion. This can lead to increased release and mobility of chemicals, potentially increasing human exposure.
2. *Increased chemical use*
   Climate change may lead to increased use of pesticides and fertilizers due to changes in agriculture and pest populations.
   There may be greater use of certain chemicals for adaptation purposes (e.g., new pest control methods).
3. *New exposure pathways*
   Changes in human behavior due to climate change (e.g., dietary changes) can affect how people come into contact with contaminated air, water, and food.
4. *Enhanced toxicity*
   Climate-related stressors like heat stress and air pollution

10 https://assets.publishing.service.gov.uk/media/65705ea1739135000db03bc1/HECC-report-2023-chapter-12-chemicals.pdf
11 https://www.ilo.org/sites/default/files/wcmsp5/groups/public/%40ed_dialogue/%40lab_admin/documents/publication/wcms_887111.pdf
12 https://minamataconvention.org/climatechange-report/

may increase human vulnerability to the toxic effects of chemicals.

5. *Indirect health effects*
   Climate change can increase chemical contamination of food and water supplies, indirectly impacting human health.
6. *Industrial accidents*
   Extreme weather events linked to climate change may increase the risk of chemical releases from industrial sites and waste facilities.
7. *Impacts on vulnerable populations*
   Workers in certain industries (e.g., agriculture, construction) and people in low/middle-income countries may face disproportionate risks from the combined effects of climate change and chemical exposures.
8. *Challenges for risk assessment*
   Current chemical risk assessment and management practices may need to be updated to account for how climate change alters exposure patterns and human vulnerabilities.

d) *Support of occupational health*

Validated NAMs could significantly enhance the assessment of occupational exposures, leading to improved safety standards and practices across various industries. These methods could provide more rapid, cost-effective, and ethically sound ways to evaluate potential health risks associated with workplace chemicals, particulates, and other hazards. By offering high-throughput screening capabilities, NAMs could enable the assessment of a wider range of substances and exposure scenarios than traditional animal testing allows. This could lead to more comprehensive and nuanced understanding of occupational hazards, including the effects of low-dose, long-term exposures that are often challenging to study in animal models. Furthermore, NAMs based on human cells or tissues could provide more relevant data for human health risk assessment, potentially improving the accuracy of exposure limits and safety guidelines. The implementation of validated NAMs in occupational health contexts could also facilitate more frequent reassessment of existing standards as new data becomes available, ensuring that worker protection measures remain up-to-date with the latest scientific evidence. Additionally, these methods could be particularly valuable in emerging industries or for novel materials where traditional toxicology data may be limited. Overall, the integration of validated NAMs into occupational health practices has the potential to enhance worker safety, reduce the incidence of occupational diseases, and support the development of safer industrial processes and materials.

5.3 Paradigm shift toward human biology-focused science

The increasing integration of quality-verified NAMs facilitated by e-validation contributes to a broader paradigm shift in biomedical science and toxicology. This shift is characterized by a move towards a systems biology approach (Hartung et al., 2012, 2017), moving away from reductionist animal models to support a more integrative, systems-level understanding of human biology and pathology, i.e., towards human-centric models based on human cells and tissues. This transition provides a more directly relevant context for studying human health and disease, potentially leading to more accurate and translatable research outcomes.

In the realm of systems toxicology (Sauer et al., 2015), NAMs enable the study of toxicological effects as perturbations of complex biological networks rather than isolated endpoints. This holistic approach allows for a more comprehensive understanding of how substances interact with biological systems, capturing the intricate interplay between different physiological processes and pathways.

Furthermore, the field of computational biology is significantly enhanced by validated NAMs, which provide rich datasets for developing and refining computational models of human biology and toxicology. These data-driven models can simulate complex biological processes, predict toxicological outcomes, and generate hypotheses for further research, all while reducing the need for animal testing. They offer the potential for more relevant, efficient, and ethical research practices that could accelerate scientific discovery and improve human health outcomes.

This transition could catalyze the development of network medicine approaches, where diseases are understood as perturbations of complex biological networks rather than isolated molecular events. Ultimately, this enables the creation of digital twins (Sillé et al., 2024). NAMs based on human embryonic tissues and stem cells could revolutionize our understanding of human development and developmental toxicity, areas where animal models have significant limitations.

In the realm of precision health, human-derived NAMs could capture the genetic and physiological diversity of human populations more effectively than animal models, supporting the development of precision health strategies. Especially, induced pluripotent stem cell (iPSC)-based models allow the comparison of individual patients' genetic backgrounds toward chemical exposure (Suciu et al., 2023; Butera et al., 2023). Some NAMs, such as organs-on-chips, could enable long-term studies of an individual's cellular and tissue responses, providing insights into chronic diseases and aging processes (Smirnova et al., 2018).

Furthermore, validated NAMs could provide crucial tools for exposome research, allowing for the assessment of the totality of environmental exposures (the exposome) and their health impacts throughout an individual's lifetime. This is particularly important in the context of a possible Human Exposome Project (Sillé et al., 2020, 2024; Hartung, 2023a; Sillé and Hartung, 2024). These advancements collectively represent a significant paradigm shift towards human biology-focused science, promising more relevant and accurate insights into human health and disease.

5.4 Implications for translational science

For translational scientists, validated NAMs offer several significant advantages. By facilitating and accelerating validation, e-validation lends itself also to test methods that do not formally require validation, but where important product development decisions profit from tool validation. In terms of improved com-

pound prioritization, NAMs that more accurately reflect human biology could provide greater confidence in compound selection for further development (Hartung, 2024a). Validated NAMs could enable selection of compounds based on specific mechanistic actions, rather than relying on phenotypic outcomes in animal models. This approach could lead to more efficient and targeted drug development processes.

In bridging the preclinical-clinical gap, NAMs could help identify and validate biomarkers that are more translatable from preclinical to clinical stages (Hartung et al., 2024). PBPK models integrated with validated NAMs could improve the prediction of safe and efficacious doses for first-in-human trials. This integration could significantly reduce the risk and increase the efficiency of early-stage clinical trials.

Furthermore, in the realm of failure analysis, when clinical trials fail, validated NAMs could provide platforms for in-depth mechanistic analysis, informing future drug design and development strategies. Similarly, when drug candidates fail in concomitant animal safety studies, they support the derisking of these substances in investigative toxicology (Beilmann et al., 2019). This capability could not only help in understanding why certain compounds fail but also guide the development of more successful drug candidates in the future, potentially reducing the high attrition rates in drug development.

5.5 Regulatory impact

E-validation's promise to provide robust, scientifically sound validation of NAMs could have far-reaching regulatory implications. In the realm of risk assessment, validated NAMs could provide more detailed, mechanism-based information for risk-benefit assessments, enabling more nuanced regulatory decisions. This enhanced level of detail could allow regulators to make more informed and precise judgments about the safety and efficacy of new products. Additionally, by providing human-relevant data, NAMs could reduce uncertainties associated with extrapolating from animal data to human outcomes, potentially leading to more accurate and reliable risk assessments.

The impact on policy evolution could be significant (von Aulock et al., 2022). The availability of validated NAMs could drive the evolution of regulatory policies towards more flexible, science-based approaches. This shift could allow for more adaptive and responsive regulatory frameworks that can keep pace with rapid scientific advancements. Furthermore, globally accepted validation processes could facilitate international harmonization of regulatory requirements, streamlining product development and approval processes across different countries and regions. This harmonization could significantly reduce the time and cost associated with bringing new products to market globally.

In terms of post-market surveillance, validated NAMs could provide valuable tools for quickly assessing newly identified safety concerns for marketed products. This rapid response capability could enhance the ability of regulatory agencies to address emerging safety issues promptly, potentially preventing adverse outcomes and improving public health protection. Overall, the regulatory impact of e-validation and validated NAMs could lead to more efficient, scientifically robust, and harmonized regulatory processes, benefiting both industry and public health.

5.6 Commercial impact

For commercial enterprises, the adoption of validated NAMs could drive significant improvements across various aspects of their operations (Meigs et al., 2018). In terms of R&D productivity, rapid and reliable screening using NAMs could accelerate product development timelines across pharmaceuticals, chemicals, cosmetics, and consumer products. This acceleration could lead to faster time-to-market for new products, providing companies with a competitive edge. Moreover, by providing early prediction of potential toxicities, NAMs could reduce late-stage failures, significantly cutting R&D costs and improving overall efficiency in the development process.

In the realm of innovation, validated NAMs could enable the development of products that are challenging to assess using traditional methods, such as complex combination therapies or advanced nanomaterials. This capability could open new avenues for product development and innovation, allowing companies to explore and bring to market novel products that were previously difficult or impossible to evaluate for safety and efficacy.

Furthermore, NAMs could contribute significantly to sustainability efforts (Maertens et al., 2024b). By supporting the development of safer, more sustainable chemicals and materials, NAMs align with the growing market demand for environmentally friendly products. This alignment could not only improve a company's environmental footprint but also enhance its market position and brand image among increasingly environmentally conscious consumers. Overall, the adoption of validated NAMs presents a compelling opportunity for commercial enterprises to improve their productivity, innovate more effectively, and align with sustainability goals, potentially leading to significant competitive advantages in their respective markets.

5.7 Public health impact

The integration of validated NAMs into safety assessment and biomedical research could have profound implications for public health. In the realm of predictive toxicology, NAMs could provide early indicators of potential health risks from environmental exposures or new products, enabling proactive public health interventions. This early warning system could allow for more timely and effective measures to protect public health. Furthermore, human cell-based NAMs could better account for variability in susceptibility across different population groups, including children, the elderly, and individuals with pre-existing conditions. This enhanced understanding of population-specific vulnerabilities could lead to more tailored and effective public health strategies.

Regarding combined exposures, validated NAMs could provide new tools for assessing the health impacts of complex environmental mixtures, addressing a major gap in current risk assessment practices. This capability is particularly important given the reality that humans are typically exposed to multiple substances simultaneously, rather than single chemicals in isola-

tion. Additionally, NAMs could help elucidate how genetic factors interact with environmental exposures to influence disease risk, supporting more targeted public health strategies. This gene-environment interaction insight could pave the way for more personalized approaches to public health interventions.

In the context of emerging health threats, the ability to quickly validate new testing approaches could significantly enhance preparedness for novel challenges, from new pathogens to new classes of environmental contaminants. This rapid validation capability could be crucial in responding effectively to unexpected public health crises, allowing for faster development of diagnostic tools, treatments, and preventive measures. Overall, the integration of validated NAMs into public health practices promises to enhance our ability to predict, prevent, and respond to a wide range of health threats, potentially leading to significant improvements in population health outcomes.

5.8 Animal testing reduction

The expedited verification of NAMs through e-validation has significant implications for reducing animal testing. From an ethical standpoint, validated NAMs provide a scientifically sound basis for reducing and ultimately replacing many forms of animal testing, addressing long-standing ethical concerns about the use of animals in research. This shift towards NAMs aligns with growing public and scientific sentiment favoring more humane research practices.

In terms of improved predictivity, NAMs that focus on human biology can often provide more relevant and predictive information than animal models, especially for human-specific biological processes. This increased relevance to human physiology not only enhances the quality of scientific research but also potentially improves the translation of findings from laboratory to clinical applications.

The regulatory landscape is also likely to evolve as more NAMs become validated. Regulatory agencies are expected to increasingly accept and even prefer these methods over traditional animal tests. This shift in regulatory acceptance could significantly accelerate the adoption of NAMs across various industries and research fields.

Finally, e-validation provides a pragmatic approach to implementing the 3R principle (replacement, reduction, refinement) in toxicology and biomedical research (Balls et al., 2024). By offering a systematic and efficient method for validating alternatives to animal testing, e-validation supports the broader goal of minimizing animal use in scientific research while maintaining or improving the quality of scientific outcomes. This approach not only addresses ethical concerns but also promotes the development of more sophisticated, human-relevant research methodologies.

5.9 Revamping of toxicology

In its grandest incarnation, e-validation will contribute to the revision of the toolbox of toxicology, representing a fundamental reimagining of toxicology and safety science (Leist et al., 2008; Hartung and Leist, 2008). By accelerating the building of trust through improved validation, the modernization of toxicology is boosted, allowing more methods to be validated with a higher success rate. This transformation is characterized by several key developments. First, data integration is becoming increasingly sophisticated, incorporating information from genomics, proteomics, metabolomics, and other high-dimensional biological measurements to provide a comprehensive view of toxicological responses. This multi-omics approach allows for a more holistic understanding of how substances interact with biological systems.

AI plays a crucial role in this revamping, with ML and LLM being leveraged to develop predictive models that can anticipate toxicological outcomes for novel chemicals and mixtures. These advanced computational tools are enhancing our ability to extrapolate from existing data to new scenarios, potentially reducing the need for extensive testing of every new substance.

The field is also moving towards personalized toxicology, accounting for genetic, epigenetic, and physiological variations in toxicological assessments. This approach recognizes that individuals may respond differently to the same exposure based on their unique biological makeup, leading to more nuanced and accurate risk assessments.

Furthermore, there is a shift towards dynamic assessments, moving from fixed testing batteries to adaptive approaches that evolve based on accumulated data. This flexibility allows for more efficient and targeted testing strategies that can be adjusted as new information becomes available.

By replacing dated animal models with validated human-focused approaches, this movement promises to enhance our understanding of the complex interplay between environmental factors, genetic predispositions, and lifestyle choices in determining health outcomes. This shift could lead to more effective strategies for disease prevention, more targeted therapeutic interventions, and ultimately, improved public health outcomes. The revamping of toxicology through e-validation and related approaches represents a significant step forward in our ability to assess and mitigate risks associated with chemical exposures, potentially revolutionizing how we approach safety science and public health protection.

In conclusion, the potential impact of e-validation extends far beyond the realm of toxicology, promising to catalyze transformative changes across biomedical research, environmental health, regulatory science, and public health. By accelerating the validation and adoption of human-relevant NAMs, e-validation could play a pivotal role in ushering in a new era of predictive, personalized, and preventive health sciences.

6 Conclusions

E-validation aims to transform rate-limiting validation barriers into catalysts for progress by providing the missing AI ingredients needed to realize next-generation safety sciences. Similar to technological leaps in other fields, NAM translation likely necessitates disruptive innovation versus incremental change. E-validation provides the vehicle to determine valid replacements in toxicology's modernization to protect public health with 21st century tools.

Several additional challenges to the validation process could potentially be addressed or mitigated using AI: AI could help

tackle the challenge of integrating diverse data types from multiple sources in validation studies. ML algorithms could be employed to analyze and synthesize data from *in vitro* assays, *in silico* models, historical animal studies, and human clinical/epidemiological sources to provide a more comprehensive assessment of a new method's performance. AI could also assist in addressing the complexity of validating integrated testing strategies (Hartung et al., 2013a; Rovida et al., 2015) or batteries of tests, where traditional one-to-one comparisons with animal tests are not feasible. NLP could be used to more efficiently extract relevant information from the scientific literature to inform validation study designs and interpretations. Additionally, AI simulation models could help predict a wider range of possible outcomes and scenarios in validation studies, potentially identifying edge cases or unexpected results that human experts might overlook. This could enhance the robustness of validation protocols and improve the assessment of a method's applicability domain. Finally, AI-powered decision support systems could aid regulators in navigating the complex landscape of validation data, helping to balance multiple factors and stakeholder perspectives in a more systematic and transparent manner.

This is not to contend e-validation alone determines the fate of NAM validation, nor that NAMs warrant endorsement without due diligence. But enhancing assay credibility to justify incorporation has proved pivotal towards changed practices. E-validation aims to tip the balance from isolated successes proving concepts to integrated transformations ushering impact at scale. The initiative channels cultural currents toward improved safety science that protects populations. This vision has considerable challenges:

1. *Dependence on AI and data quality*
   While the integration of AI and ML is a strength, it also poses challenges. The effectiveness of e-validation heavily depends on the quality and comprehensiveness of the data used to train AI models. Any biases or gaps in the data could undermine the reliability of the validation process.
2. *Complexity and implementation*
   The implementation of e-validation involves multiple sophisticated AI and ML components, which could be complex and resource-intensive. We will need a more detailed discussion on the practical aspects of implementing e-validation, including the required infrastructure, expertise, and potential barriers.
3. *Scalability and generalizability*
   While this article provides a comprehensive framework, the scalability and generalizability of e-validation across different types of NAMs and toxicological studies remain uncertain. More evidence and case studies demonstrating successful applications across various contexts would strengthen the proposal.
4. *Ethical and regulatory challenges*
   The adoption of AI-driven approaches in regulatory contexts can face ethical and regulatory hurdles. To address these challenges more explicitly, it will be necessary to discuss strategies to ensure ethical AI use, mitigate biases, and gain regulatory acceptance.
5. *Need for real-world validation*
   The theoretical advantages of e-validation are compelling, but real-world validation and pilot studies are necessary to demonstrate its practical benefits and feasibility. We must develop a roadmap for conducting such validation studies to build confidence in the approach.

In summary, e-validation proposes an ambitious and innovative framework for transforming the validation of NAMs using AI and ML. The approach has the potential to address significant challenges in traditional validation methods, offering a more efficient, transparent, and adaptive process. However, the success of e-validation will depend on addressing practical implementation challenges, ensuring high-quality data, and navigating ethical and regulatory landscapes. Real-world validation and case studies will be crucial to demonstrate the feasibility and impact of this approach in advancing toxicological science and regulatory practices. The vision of e-validation laid out here presents a timely adaptation of the revolutionary changes AI is bringing to science. It will have to prove whether it can hold these promises – failure is an option, not trying it is not!

References

Abbas, I., Rovira, J., Cobo, E. et al. (2006). Simulation models for optimizing the design of clinical trials. *Qual Reliab Enging Int 22*, 683-691. doi:10.1002/qre.799

Balbus, J. M., Boxall, A. B., Fenske, R. A. et al. (2013). Implications of global climate change for the assessment and management of human health risks of chemicals in the natural environment. *Environ Toxicol Chem 32*, 62-78. doi:10.1002/etc.2046

Balls, M., Bass, R., Curren, R. et al. (2024). 60 Years of the 3Rs symposium: Lessons learned and the road ahead. *ALTEX 41*, 179-201. doi:10.14573/altex.2403061

Baranwal, M., Magner, A., Elvati, P. et al. (2020). A deep learning architecture for metabolic pathway prediction. *Bioinformatics 36*, 2547-2553. doi:10.1093/bioinformatics/btz954

Beilmann, M., Boonen, H., Czich, A. et al. (2019). Optimizing drug discovery by investigative toxicology: Current and future trends. *ALTEX 36*, 3-17. doi:10.14573/altex.1808181

Bell, S. M., Phillips, J., Sedykh, A. et al. (2017). An integrated chemical environment to support 21st-century toxicology. *Environ Health Perspect 125*, 054501. doi:10.1289/ehp1759

Bell, S., Abedini, J., Ceger, P. et al. (2020). An integrated chemical environment with tools for chemical safety testing. *Toxicol In Vitro 67*, 104916. doi:10.1016/j.tiv.2020.104916

Blaauboer, B. J. (2010). Biokinetic modeling and in vitro-in vivo extrapolations. *J Toxicol Environ Health B 13*, 242-252. doi:10.1080/10937404.2010.483940

Bouvier d'Yvoire, M., Prieto, P., Blaauboer, B. J. et al. (2007). Physiologically-based kinetic modelling (PBK modelling): Meeting the 3Rs agenda. The report and recommendations of ECVAM workshop 63. *Altern Lab Anim 35*, 661-671. doi:10.1177/026119290703500606

Boxall, A. B., Hardy, A., Beulke, S. et al. (2009). Impacts of climate change on indirect human exposure to pathogens and chemicals from agriculture. *Environ Health Perspect 117*, 508-514. doi:10.1289/ehp.0800084

Busquet, F. and Hartung, T. (2017). The need for strategic de-

velopment of safety sciences. *ALTEX 34*, 3-21. doi:10.14573/altex.1701031

Butera, A., Smirnova, L., Ferrando-May, E. et al. (2023). Deconvoluting gene and environment interactions to develop "epigenetic score meter" of disease. *EMBO Mol Med 15*, e18208. doi:10.15252/emmm.202318208

Chesnut, M., Paschoud, H., Repond, C. et al. (2021). Human 3D iPSC-derived brain model to study chemical-induced myelin disruption. *Int J Mol Sci 22*, 9473. doi:10.3390/ijms22179473

Corradi, M. P. F., de Haan, A. M., Staumont, B. et al. (2022). Natural language processing in toxicology: Delineating adverse outcome pathways and guiding the application of new approach methodologies. *Biomater Biosys 7*, 100061. doi:10.1016/j.bbiosy.2022.100061

Corvi, R., Albertini, S., Hartung, T. et al. (2008). ECVAM retrospective validation of in vitro micronucleus test (MNT). *Mutagenesis 23*, 271-283. doi:10.1093/mutage/gen010

Costello, Z. and Martin, H. G. (2018). A machine learning approach to predict metabolic pathway dynamics from time-series multiomics data. *NPJ Syst Biol Appl 4*, 19. doi:10.1038/s41540-018-0054-3

Crawford, S. E., Hartung, T., Hollert, H. et al. (2017). Green toxicology: A strategy for sustainable chemical and material development. *Environ Sci Eur 29*, 16. doi:10.1186/s12302-017-0115-z

Hamon, J., Renner, M., Jamei, M. et al. (2015). Quantitative in vitro to in vivo extrapolation of tissues toxicity. *Toxicol In Vitro 30*, 203-216. doi:10.1016/j.tiv.2015.01.011

Hartung, T., Bremer, S., Casati, S. et al. (2004). A modular approach to the ECVAM principles on test validity. *Altern Lab Anim 32*, 467-472. doi:10.1177/026119290403200503

Hartung, T. and Leist, M. (2008). Food for thought … on the evolution of toxicology and phasing out of animal testing. *ALTEX 25*, 91-96. doi:10.14573/altex.2008.2.91

Hartung, T., van Vliet, E., Jaworska, J. et al. (2012). Systems toxicology. *ALTEX 29*, 119-128. doi:10.14573/altex.2012.2.119

Hartung, T., Luechtefeld, T., Maertens, A. et al. (2013a). Integrated testing strategies for safety assessments. *ALTEX 30*, 3-18. doi:10.14573/altex.2013.1.003

Hartung, T., Stephens, M. and Hoffmann, S. (2013b). Mechanistic validation. *ALTEX 30*, 119-130. doi:10.14573/altex.2013.2.119

Hartung, T. (2017). A comprehensive overview of the current status and application of predictive ADMET. In S. Chackalamannil, D. Rotella and S. E. Ward (eds), *Comprehensive Medicinal Chemistry III* (Chapter 4.08, 150-155). doi:10.1016/B978-0-12-409547-2.12378-9

Hartung, T., FitzGerald, R., Jennings, P. et al. (2017). Systems toxicology – Real world applications and opportunities. *Chem Res Toxicol 30*, 870-882. doi:10.1021/acs.chemrestox.7b00003

Hartung, T. (2023a). ToxAIcology – The evolving role of artificial intelligence in advancing toxicology and modernizing regulatory science. *ALTEX 40*, 559-570. doi:10.14573/altex.2309191

Hartung, T. (2023b). Artificial intelligence as the new frontier in chemical risk assessment. *Front Artif Intell 6*, 1269932. doi:10.3389/frai.2023.1269932

Hartung, T. (2024a). The (misleading) role of animal models in drug development. *Front Drug Discovery 4*, 1355044. doi:10.3389/fddsv.2024.1355044

Hartung, T. (2024b). The validation of regulatory test methods – Conceptual, ethical, and philosophical foundations. *ALTEX 41*, 525-544. doi:10.14573/altex.2409271

Hartung, T., King, N. P. M., Kleinstreuer, N. et al. (2024). Leveraging biomarkers and translational medicine for preclinical safety - Lessons for advancing the validation of alternatives to animal testing. *ALTEX 41*, 545-566. doi:10.14573/altex.2410011

Hoffmann, S., Edler, L., Gardner, I. et al. (2008). Points of reference in validation – The report and recommendations of ECVAM Workshop. *Altern Lab Anim 36*, 343-352. doi:10.177/026119290803600311

Kang, H., Goo, S., Lee, H. et al. (2022). Fine-tuning of BERT model to accurately predict drug-target interactions. *Pharmaceutics 14*, 1710. doi:10.3390/pharmaceutics14081710

Kleinstreuer, N. and Hartung, T. (2024). Artificial intelligence (AI) – It's the end of the tox as we know it (and I feel fine) – AI for predictive toxicology. *Arch Toxicol 98*, 735-754. doi:10.1007/s00204-023-03666-2

Leist, M., Hartung, T. and Nicotera, P. (2008). The dawning of a new age of toxicology. *ALTEX 25*, 103-114. doi:10.14573/altex.2008.2.103

Leist, M., Ghallab, A., Graepel, R. et al. (2017). Adverse outcome pathways: Opportunities, limitations and open questions. *Arch Toxicol 31*, 221-229. doi:10.1007/s00204-017-2045-3

Lin, Z. and Chou, W-C. (2022). Machine learning and artificial intelligence in toxicological sciences. *Toxicol Sci 189*, 7-19. doi:10.1093/toxsci/kfac075

Luechtefeld, T. and Hartung, T. (2017). Computational approaches to chemical hazard assessment. *ALTEX 34*, 459-478. doi:10.14573/altex.1710141

Maertens, A. and Hartung, T. (2018). Green toxicology – Know early about and avoid toxic product liabilities. *Toxicol Sci 161*, 285-289. doi:10.1093/toxsci/kfx243

Maertens, A. and Hartung, T. (2019). Green toxicology meets nanotoxicology: The process of sustainable nanomaterial development and use. A. Gajewicz and T. Puzyn (eds), *Computational Nanotoxicology: Challenges, Pitfalls, and Perspectives* (495-506). Jenny Stanford Publishing. doi:10.1201/9780429341373-11

Maertens, A., Golden, E. and Hartung, T. (2021). Avoiding regrettable substitutions: Green toxicology for sustainable chemistry. *ACS Sustain Chem Eng 9*, 7749-7758. doi:10.1021/acssuschemeng.0c09435

Maertens, A., Golden, E., Luechtefeld, T. H. et al. (2022). Probabilistic risk assessment – The keystone for the future of toxicology. *ALTEX 39*, 3-29. doi:10.14573/altex.2201081

Maertens, A., Antignac, E., Benfenati, E. et al. (2024a). The probable future of toxicology – Probabilistic risk assessment. *ALTEX 41*, 273-281. doi:10.14573/altex.2310301

Maertens, A., Luechtefeld, T. and Hartung, T. (2024b). Alternative methods go green! Green toxicology as a sustainable approach for assessing chemical safety and designing safer chemicals. *ALTEX 41*, 3-19. doi:10.14573/altex.2312291

Malikova, M. A. (2016). Optimization of protocol design: A path

to efficient, lower cost clinical trial execution. *Future Sci OA 2*, FSO89. doi:10.4155/fso.15.89

Marx, U., Andersson, T. B., Bahinski, A. et al. (2016). Biology-inspired microphysiological system approaches to solve the prediction dilemma of substance testing using animals. *ALTEX 33*, 272-321. doi:10.14573/altex.1603161

Marx, U., Akabane, T., Andersson, T. B. et al. (2020). Biology-inspired microphysiological systems to advance medicines for patient benefit and animal welfare. *ALTEX 37*, 364-394. doi:10.14573/altex.2001241

Meigs, L., Smirnova, L., Rovida, C. et al. (2018). Animal testing and its alternatives – The most important omics is economics. *ALTEX 35*, 275-305. doi:10.14573/altex.1807041

Petersen, E. J., Nguyen, A., Brown, J. et al. (2021). Characteristics to consider when selecting a positive control material for an in vitro assay. *ALTEX 38*, 365-376. doi:10.14573/altex.2102111

Proença, S., Paini, A., Joossens, E. et al. (2019). Insights into in vitro biokinetics using virtual cell based assay simulations. *ALTEX 36*, 447-461. doi:10.14573/altex.1812101

Rodríguez-Belenguer, P., March-Vila, E., Pastor, M. et al. (2023). Usage of model combination in computational toxicology. *Toxicol Lett 389*, 34-44. doi:10.1016/j.toxlet.2023.10.013

Roth, A. and MPS-WS Berlin 2019 [Marx, U., Vilén, L., Ewart, L. et al.] (2019). Human microphysiological systems for drug development. *Science 373*, 1304-1306. doi:10.1126/science.abc3734

Rovida, C., Alépée, N., Api, A. M. et al. (2015). Integrated testing strategies (ITS) for safety assessment. *ALTEX 32*, 171-181. doi:10.14573/altex.1506201

Sauer, J. M., Hartung, T., Leist, M. et al. (2015). Systems toxicology: The future of risk assessment. *Int J Toxicol 34*, 346-348. doi:10.1177/1091581815576551

Shah, H. A., Liu, J., Yang, Z. et al. (2021). Review of machine learning methods for the prediction and reconstruction of metabolic pathways. *Front Mol Biosci 8*, 634141. doi:10.3389/fmolb.2021.634141

Sillé, F. C. M., Karakitsios, S., Kleensang, A. et al. (2020). The exposome – A new approach for risk assessment. *ALTEX 37*, 3-23. doi:10.14573/altex.2001051

Sillé, F. and Hartung, T. (2024). Metabolomics in preclinical drug safety assessment: Current status and future trends. *Metabolites 14*, 98. doi:10.3390/metabo14020098

Sillé, F. C. M., Busquet, F., Fitzpatrick, S. et al. (2024). The implementation moonshot project for alternative chemical testing (IMPACT) toward a human exposome project. *ALTEX 41*, 344-362. doi:10.14573/altex.2407081

Smirnova, L., Kleinstreuer, N., Corvi, R. et al. (2018). 3S – Systematic, systemic, and systems biology and toxicology. *ALTEX 35*, 139-162. doi:10.14573/altex.1804051

Suciu, I., Pamies, D., Peruzzo, R. et al. (2023). GxE interactions as a basis for toxicological uncertainty. *Arch Toxicol 97*, 2035-2049. doi:10.1007/s00204-023-03500-9

Sverdlov, O., Ryeznik, Y. and Wong, W. K. (2019). On optimal designs for clinical trials: An updated review. *J Stat Theory Pract 14*. 10. doi:10.1007/s42519-019-0073-4

Tetko, I. V., Klambauer, G., Clevert, D. A. et al. (2022). Artificial intelligence meets toxicology. *Chem Res Toxicol 35*, 1289-1290. doi:10.1021/acs.chemrestox.2c00196

Tonoyan, L. and Siraki, A. G. (2024). Machine learning in toxicological sciences: Opportunities for assessing drug toxicity. *Front Drug Discov 4*, 1336025. doi:10.3389/fddsv.2024.1336025

Tsaioun, K., Blaauboer, B. J. and Hartung, T. (2016). Evidence-based absorption, distribution, metabolism, excretion and toxicity (ADMET) and the role of alternative methods. *ALTEX 33*, 343-358. doi:10.14573/altex.1610101

van Ertvelde, J., Verhoeven, A., Maerten, A. et al. (2023). Optimization of an adverse outcome pathway network on chemical-induced cholestasis using an artificial intelligence-assisted data collection and confidence level quantification approach. *J Biomed Inform 145*, 104465-104465. doi:10.1016/j.jbi.2023.104465

Vinken, M., Benfenati, E., Busquet, F. et al. (2021). Safer chemicals using less animals: Kick-off of the European ONTOX project. *Toxicology 458*, 152846. doi:10.1016/j.tox.2021.152846

von Aulock, S., Busquet, F., Locke, P. et al. (2022). Engagement of scientists with the public and policymakers to promote alternative methods. *ALTEX 39*, 543-559. doi:10.14573/altex.2209261

Conflict of interest

Thomas Luechtefeld is founder and owner of Tox-Track Inc. and Insilica LLC. Thomas Hartung holds stock options in and consults Tox-Track LLC Inc. and Insilica LLC. Both are consultants for computational toxicology for Underwriters Laboratories (UL) and receive shares of their respective sales. Thomas Hartung is a member of Apple's Green Chemistry Advisory Board.

Data availability

No datasets were generated in this study.

Acknowledgements

The concept of e-validation was awarded a Complement-ARIE award by NIH[13]. This work has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 963845 (ONTOX).

[13] https://commonfund.nih.gov/complementarie/challengewinnersummaries