INSTITUTO
UNIVERSITÁRIO
DE LISBOA

# An Integrated Framework for the Responsible Adoption of Artificial Intelligence (AI) and New Approach Methodologies (NAMs) in Replacing Animal Testing in the Cosmetic and Pharmaceutical Industries

Joana Anselmo De Magalhães

MSc in Business Administration

Supervisors:
PhD, Renato Jorge Lopes da Costa, Assistant Professor with Habilitation, ISCTE-Instituto Universitário de Lisboa

PhD, Rui Alexandre Henriques Gonçalves, Invited Assistant Professor, ISCTE-Instituto Universitário de Lisboa

October, 2025

BUSINESS
SCHOOL

Department of Marketing, Operations and General Management

# An Integrated Framework for the Responsible Adoption of Artificial Intelligence (AI) and New Approach Methodologies (NAMs) in Replacing Animal Testing in the Cosmetic and Pharmaceutical Industries

Joana Anselmo De Magalhães

MSc in Business Administration

Supervisors:
PhD, Renato Jorge Lopes da Costa, Assistant Professor with Habilitation, ISCTE-Instituto Universitário de Lisboa

PhD, Rui Alexandre Henriques Gonçalves, Invited Assistant Professor, ISCTE-Instituto Universitário de Lisboa

October, 2025

## Acknowledgments

To my supervisor, Professor Renato Jorge Lopes da Costa, for sharing his expertise and passion on the topic, which guided me throughout this dissertation.

To my co-supervisor, Professor Rui Alexandre Henriques Gonçalves, for his invaluable guidance, essential motivation, and inspiring enthusiasm.

To my parents, for their endless support, encouragement, and unwavering belief in me.

To my boyfriend, for his patience, love, and constant encouragement, which gave me the strength to endure the most challenging moments.

To my friends, for the motivation, collaboration, patience, and shared knowledge that made this journey lighter and more enjoyable.

To the interviewees, for generously sharing insights that significantly enriched my research.

Lastly, to everyone who, in one way or another, contributed to this dissertation, no matter how small the contribution.

Thanks to everyone!

**Resumo**

A utilização contínua de testes em animais nos setores de cosméticos e farmacêuticos tem sido cada vez mais contestada, não só por razões éticas, mas também por sua baixa capacidade de prever resultados em humanos. O desenvolvimento da Inteligência Artificial (IA) e das Metodologias de Nova Abordagem (MNAs) oferece alternativas mais éticas e potencialmente mais eficazes, porém ainda há uma falta de investigação integrada sobre as implicações éticas, regulatórias, tecnológicas e financeiras de sua adoção. O objetivo geral desta dissertação é analisar como a IA e as MNAs podem ser implementadas de forma responsável para reduzir ou substituir a experimentação animal. Para isso, foi utilizada uma metodologia qualitativa, baseada em entrevistas semiestruturadas com quinze especialistas de diversas áreas científicas e profissionais. Os resultados apontam vantagens claras em termos de custos, previsibilidade e alinhamento ético, mas também evidenciam barreiras como a aceitação regulatória, a validação tecnológica e a diversidade de políticas globais.

**Palavras-chave:** Inteligência Artificial (IA); Testes em Animais; Metodologias de Nova Abordagem (MNAs); Metodologia Qualitativa

**Classificação *JEL***:

M14 – Cultura Corporativa • Diversidade • Responsabilidade Social

I18 – Política Governamental • Regulação • Saúde Pública

**Abstract**

The ongoing reliance on animal testing in the cosmetics and pharmaceutical industries has been increasingly questioned, not only for ethical reasons but also due to its limited ability to predict human outcomes. Advances in Artificial Intelligence (AI) and New Approach Methodologies (NAMs) offer more ethical and potentially more effective alternatives, yet there is still a lack of integrated research exploring their ethical, regulatory, technological, and financial implications. The main goal of this dissertation is to examine how AI and NAMs can be ethically and responsibly implemented to reduce or replace animal testing. A qualitative approach was employed, involving semi-structured interviews with fifteen experts from diverse scientific and professional backgrounds. The results emphasize obvious benefits such as cost savings, better predictive accuracy, and ethical improvements, but also reveal substantial challenges, including regulatory approval, technological validation, and the diversity of global policy frameworks. This study helps shed light on the future of safety testing in these fields.

**Keywords:** Artificial Intelligence (AI); Animal Testing; New Approach Methodologies (NAMs); Qualitative Research

**JEL Classification**:

M14 – Corporate Culture • Diversity • Social Responsibility

I18 – Government Policy • Regulation • Public Health

**General Index**

**Index of Figures**

**Index of Tables**

**Abbreviations List**

(q)AOPs – (Quantitative) Adverse Outcome Pathways

3Rs – Replacement, Reduction, Refinement

ADME – Absorption, Distribution, Metabolism, and Excretion

AI – Artificial Intelligence

ANNs – Artificial Neural Networks

AOPs – Adverse Outcome Pathways

AR – Augmented Reality

ASPIS – Animal-free Safety Assessment of Chemicals

CEFIC – European Chemical Industry Council

DART – Developmental and Reproductive Toxicity

DDTs – Drug Development Tools

DFG – German Research Foundation

DL – Deep Learning

DNT – Developmental Neurotoxicity

DOAJ – Directory of Open Access Journals

DPRA – Direct Peptide Reactivity Assay

DTs – Decision Trees

EBA – Exposure-Based Adaptation

ECHA – European Chemicals Agency

EFSA – European Food Safety Authority

EMA – European Medicines Agency

ESA – Environmental Safety Assessment

EU – European Union

FAIR – Findable, Accessible, Interoperable, Reusable

FDA – Food and Drug Administration

GPMT – Guinea Pig Maximization Test

h-CLAT – Human Cell Line Activation Test

HSI – Humane Society International

HT – High-Throughput

HTS – High-Throughput Screening

IATA – Integrated Approaches to Testing and Assessment

IATAs – Integrated Approaches to Testing and Assessment (plural usage)

ICCVAM – Interagency Coordinating Committee on the Validation of Alternative Methods

ICE – Integrated Chemical Environment

ISTAND – Innovative Science and Technology Approaches for New Drugs

KY – K-Step Yard Sampling

LLNA – Local Lymph Node Assay

LM – Language Model

LR – Logistic Regression

ML – Machine Learning

MPS – Microphysiological Systems

multi-OoC – Multiorgan-on-Chip

MUSST – Myeloid U937 Skin Sensitization Test

NAMs – New Approach Methodologies

NGO – Non-Governmental Organization

NGRA – Next Generation Risk Assessment

NGS – Next-Generation Sequencing

NIST – United States National Institute of Standards and Technology

NRC – National Research Council

OECD – Organization for Economic Co-operation and Development

OECD TG – OECD Test Guideline

OIDs – Orally Inhaled Drugs

OoC – Organ-on-Chip

PBK – Physiologically Based Kinetic Modelling

PBPK – Physiologically Based Pharmacokinetic

PBTK – Physiologically Based Toxicokinetic

PK/PD – Pharmacokinetic/Pharmacodynamic

QIVIVE – Quantitative In Vitro to In Vivo Extrapolation

QMRF – QSAR Model Reporting Format (or Model Reporting Format)

QSARs – Quantitative Structure-Activity Relationships

R&D – Research and Development

RAAF – Read-Across Assessment Framework

REACH – Registration, Evaluation, Authorization and Restriction of Chemicals

SARs – Structure-Activity Relationships (Databases)

SVMs – Support Vector Machines

TILSQ – Iterative Least Squares Linear Discriminant

TK – Toxicokinetic

TTC – Threshold of Toxicological Concern

U.S. – United States

WHO – World Health Organization

## Chapter I – Introduction

In recent years, ethical concerns and scientific limitations surrounding traditional animal testing have gained significant attention, driving a global search for innovative and reliable alternatives. This dissertation focuses on replacing animal testing with Artificial Intelligence (AI) and other New Approach Methodologies (NAMs) in the cosmetics and pharmaceutical industries. It rethinks the core of preclinical testing: moving away from long-standing animal experimentation practices and instead utilizing technologies that can more accurately, efficiently, and ethically predict human responses to cosmetics and pharmaceuticals. As confirmed by the United States (U.S.) Food and Drug Administration (FDA, 2025) roadmap, 90% of drugs tested on animals ultimately fail in human trials, mainly due to safety and/or efficacy concerns.

Additionally, NAMs also have enormous cost-saving potential. This encompasses not only a scientific shift but also a broader transformation that touches on corporate decision-making, regulatory compliance, financial investment, consumer trust, and ethical responsibility.

The importance of this topic stems from the convergence of social, regulatory, scientific, and financial pressures. Ethically, the principle of the "3Rs" - Replacement, Reduction, and Refinement of animal use - has been a foundational concept for decades. Regulatory agencies and the scientific community increasingly see replacement as the most preferred goal. As one guideline states, "In accordance with Directive 2010/63/EU, the principle of the 3Rs needs to be considered when selecting testing approaches to be used for regulatory testing of human and veterinary medicinal products." (EMA, 2016, p.3). In practice, this ethical principle has been legally mandated in multiple regions.

The European Union (EU), for example, banned animal testing for cosmetic products and ingredients under the Cosmetics Regulation 1223/2009 and strengthened its commitment through the REACH (Registration, Evaluation, Authorization, and Restriction of Chemicals) legislation, which requires industry to minimize or avoid animal testing whenever scientifically possible. (Silva & Tamburic, 2022). These measures are not isolated; other countries and regulatory bodies are increasingly adopting similar restrictions.

From a scientific perspective, the limitations of animal models have become impossible to ignore. Evidence suggests that animal testing is not always predictive of human

responses, leading to failures in drug development pipelines, misinterpretation of safety signals, and wasted investment. The pharmaceutical industry, in particular, has faced high attrition rates, with eight out of nine drug candidates entering the clinical testing phase failing, primarily due to issues with safety and efficacy, despite having passed preclinical testing in animals (Van Berlo et al., 2021). This raises questions not only of ethics but also of scientific validity. AI, in silico modeling, organ-on-chip (OoC) systems, and other NAMs present an opportunity to model human biology more directly, generating data that is both more relevant, more predictive, and less expensive (Bailey & Balls, 2019).

Financial considerations further strengthen the argument for adopting AI alternatives. Traditional animal testing is slow and expensive, with some carcinogenicity bioassays costing "over $1 million" and lasting several years (Taylor & Alvarez, 2020, p. 2). By contrast, AI-based in silico models, machine learning (ML) algorithms, and high-throughput (HT) in vitro systems offer the possibility of producing results at a fraction of the cost and within significantly shorter timeframes. This makes the shift toward AI not only ethically and scientifically appealing but also strategically sound for companies facing competitive global markets, increasing consumer demand for cruelty-free products, and pressure to reduce development timelines (Winarto & Wisesa, 2024; Sihaloho et al., 2023).

The objectives of this thesis are therefore threefold. First, to review and synthesize the existing literature across multiple disciplines, including toxicology, regulatory science, ethical branding, financial modeling, and consumer behavior, to provide a comprehensive understanding of the current landscape of alternatives to animal testing. Second, to critically examine the motivations, benefits, and limitations of implementing AI and related technologies in the replacement of animal testing within the cosmetics and pharmaceutical sectors. Third, to investigate how companies can responsibly transition toward AI-driven testing methods while addressing the ethical, regulatory, technological, and financial considerations that inevitably shape this process.

The main aim of this dissertation, entitled *An Integrated Framework for the Responsible Adoption of AI and NAMs in Replacing Animal Testing in the Cosmetic and Pharmaceutical Industries*, is to explore how AI can be implemented to reduce and ultimately replace animal testing in these sectors. At its core, the central research problem is how cosmetic and pharmaceutical companies can adopt these innovations in ways that safeguard safety and efficacy, comply with regulatory requirements, manage financial and operational realities, and maintain public trust (Westmoreland et al., 2022).

Importantly, the shift from animal testing to alternative testing methods is not uniform. It is an ongoing process that varies across industries and regions (Amador, 2021). Some companies have already integrated AI and other NAMs into their pipelines, while others remain hesitant. As one source notes, "Replacement of in vivo by in vitro testing falls into two distinct scenarios: Replacement of in vivo testing by in vitro for well-established products that have been on the market for a long time", and "Avoiding the use of in vivo testing and instead encouraging the development of in vitro tests whenever possible for new products. The finances needed to establish in vitro tests should not block the research efforts, and investments in standardized reagents, reference preparations, and scientific methods by the industrial, regulatory, and scientific communities should be encouraged." (Romberg et al., 2012, p.104). This indicates that adoption is often easier in cases where products are already well understood, while innovation-heavy sectors may face greater challenges.

This dissertation is structured into seven chapters, each designed to address a specific aspect of the research in a logical and cumulative progression.

*Chapter I (Introduction)* contextualizes the research topic, defines the problem, presents the study's primary aim, and outlines themes to be addressed in Chapter V (Results and Discussion). It emphasizes the relevance of transitioning from animal testing to NAMs, with a focus on AI-driven alternatives in the cosmetic and pharmaceutical sectors.

*Chapter II (Literature Review)* examines the replacement of animal testing with AI and NAMs across various dimensions, including ethical, regulatory, technological, and financial considerations. It is organized around four themes: (1) the current role of AI and NAMs in preclinical testing and safety assessment; (2) potential benefits for the industries; (3) costs and challenges for the industries, and (4) implementation pathways, from hybrid approaches to full replacements, critically assessing the conditions under which alternative methods could fully replace animal testing. This dissertation synthesizes the fragmented literature into a coherent foundation for the empirical component, providing companies with guidance on how to navigate the transition responsibly.

*Chapter III (Theoretical Approach)* connects the literature review with the seven research questions, underscoring the interdisciplinary nature of the transition and aiming to bridge ethical imperatives with corporate realities.

*Chapter IV (Methodology)* details the qualitative research design, including the interview guide, participant criteria, and analytical framework.

*Chapter V (Results and Discussion)* presents insights from expert interviews, linking them to the perspectives of the literature review, drawing comparisons, and identifying trends related to the adoption of AI and NAM.

*Chapter VI (Conclusion)* synthesizes the findings and directly addresses the central aim: whether AI and NAMs can reduce or replace animal testing in the cosmetic and pharmaceutical industries. It also discusses study limitations, provides recommendations for future research, and offers personal reflections, offering both a final conclusion to the research aim and a critical perspective on the field's current state and future directions.

## Chapter II – Literature Review: Regulatory, Financial, Technological, and Ethical Considerations

### 2.1 Current Role of AI and NAMs in Testing

Artificial intelligence (AI) is reshaping toxicology, pharmaceutical development, and cosmetic safety testing as part of the broader shift toward NAMs, which are grounded in the 3Rs. (Madden et al., 2020; Alves et al., 2021; Van Berlo et al., 2021). European centers and platforms emphasize the research, validation, and uptake of 2-D/3-D human cell assays across neurotoxicity, embryotoxicity, endocrine disruption, and adjacent domains; these efforts build Adverse Outcome Pathways (AOPs) and support regulatory toxicology, with education programs expanding in vitro and in silico literacy (Neuhaus et al., 2022).

Nonetheless, essential gaps persist in developmental neurotoxicity (DNT), developmental and reproductive toxicity (DART), and chronic toxicity, so reduction and refinement remain essential interim aims. For orally inhaled drugs (OIDs), high clinical failure rates reflect weak animal-to-human predictivity; in vitro NAMs can better anticipate efficacy and safety and eliminate weak candidates earlier (Movia & Prina-Mello, 2020). As early as 1997, regulators were already considering many in vitro testing methods (EMA, 1997).

Beyond complex in vitro and computational in silico categories, multiple innovations strengthen a non-animal ecosystem: ex vivo human tissues can be exposed to a drug to assess localized toxic effects or immune cell infiltration; high-throughput/high-content screens of human cells survey off-target effects across diverse genotypes; microdosing with PET imaging yields early human pharmacokinetic (PK) and distribution data; and refined in vivo strategies reduce animal numbers during transition. In practice, a human organ-chip for toxicity, plus physiologically based PK and physiologically based toxicokinetic (PBPK/PBTK) for PK, and an AI immunogenicity predictor can reproduce the informational value of a traditional whole-animal study with greater human relevance (US FDA, 2025).

As Madden et al. (2020, pp. 146–149) emphasize, information from a combination of techniques is required: “In silico modelling, in vitro assays, high-throughput screening” (HTS), “organ-on-a-chip technology, omics, and mathematical biology, to provide complementary information to develop a complete picture of the potential response of an

organism to a chemical stressor". In silico methods encompass a range of technologies, including "Databases, Structure-activity relationships (SARs); structural alerts, Quantitative structure-activity relationships (QSARs), (Quantitative) adverse outcome pathways ((q)AOPs) and Physiologically based kinetic (PBK) modelling".

Since the EU cosmetics ban, firstly in 2004 and then in 2013, AI-driven in silico and computational solutions have become essential (Kimura et al., 2025, p. 1), with skin-sensitization models spanning diverse ML classifiers (Grech et al., 2024). Under "The Interagency Coordinating Committee on the Validation of Alternative Methods" (ICCVAM), Logistic Regression (LR) and Support Vector Machines (SVM) integrating Direct Peptide Reactivity Assay (DPRA), Human Cell Line Activation Test (h-CLAT), and read-across, achieved up to "92% accuracy", and some models integrating log P and KeratinoSens. Complementary work notes simulation can outperform animal experiments by leveraging human data, supported by stem cells, biochips, and 3D images (Kabene & Baadel, 2019, p. 8).

Current in vitro tests include h-CLAT, DPRA, LuSens, Myeloid U937 Skin Sensitization Test (MUSST), and KeratinoSens, which offer strong performance but typically interrogate a single AOP. Therefore, they are often coupled with in silico models and historical epidemiology to extend their scope (Grech et al., 2024). On the experimental side, the classification and prediction of chemical toxicity using automated phenotypic profiling of Caenorhabditis elegans demonstrates how NAMs connect phenotype to hazard, while, as a microscopic model organism, the costs associated with its development, maintenance, and housing are significantly lower compared to traditional rabbit or murine models, making research more accessible with limited budgets (Samah H. O. Zarroug, 2025).

Additionally, Transgenic C. elegans strains expressing human disease-associated proteins, such as those associated with Parkinson's and Alzheimer's diseases, are engineered to aid in understanding neurodegeneration, neuron-coordinated behavioral changes, and cognitive defects (Gutleb et al., 2025). "Additionally, the genetic similarities between certain lower-rank animals and humans supports their translational relevance to humans" (Manful et al., 2023, p.9, as cited in Pitchakarn et al., 2021).

The Next Generation Risk Assessment (NGRA) has enabled "cruelty-free" skin sensitization via tiered frameworks, with similar approaches currently being developed for other endpoints. Consumer-facing AI screens ingredients for allergens and irritants, supporting personalized routines (Hash et al., 2025, p. 3). However, privacy and security concerns necessitate strong protective measures and careful data governance (Grech et al., 2024; Hash et al., 2025).

Regulatory complexity persists. According to Knight et al. (2021, p. 667), the EU cosmetic regulations conflict due to Regulation EC 1223/2009, which bans in vivo testing, and the Registration, Evaluation, Authorization, and Restriction of Chemicals (REACH), which can require it. Their analysis of REACH dossiers revealed "3,206 chemicals used solely in cosmetics", with "419 classified as cosmetic-only". Among these, "63 had in vivo tests conducted after the ban". Many registrants "used alternative methods, but some still performed in vivo tests to meet REACH requirements". The European Chemicals Agency (ECHA), the agency responsible for evaluating REACH dossiers, has requested additional in vivo tests, which may necessitate further testing in the future. Current tracking methods "fail to distinguish REACH tests on cosmetics from those on industrial chemicals". The most recent "report, from 2020, states that no testing was reported for the Cosmetic Regulation"; however, "REACH records make it clear that testing of cosmetic ingredients continues under REACH".

This situation presents challenges for ingredient manufacturers, who may face legal requirements for testing that could harm market acceptance, and for brands that risk consumer backlash if they are found to be using ingredients that have been tested. Consequently, consumer trust in cruelty-free cosmetics is at risk.

AI is driving momentum in drug discovery, making processes more efficient, accurate, and cost-effective (Malik et al., 2024, p. 4; Liu et al., 2021, p. 2604). At the experimental interface, microfluidics, multi-OoC models, and digital twins reduce sample needs, improve predictive power, and simulate physiological responses pre-clinically, while iPSCs, organoids, and GANs capture human heterogeneity for in silico clinical trials (Kimura et al., 2025; Van Berlo et al., 2021; Gangwal & Lavecchia, 2025; Zushin et al., 2023).

Predictive toxicology now fuses big data with AI/ML/DL and bioinformatics, moving beyond QSAR limits by integrating diverse datasets (Kleinstreuer & Hartung, 2024, as cited in Hartung & Hoffmann, 2009; Gangwal & Lavecchia, 2025). Exemplars such as DeepTox, DeepDILI, and AnimalGAN, together with advanced (Q)SAR methods supported by DNN, SVM, k-nearest neighbor, and random forest, extend predictive accuracy and cut late-stage failures (Gangwal & Lavecchia, 2025; Silva, 2023; Kumar, 2021, as cited in Silva, 2023). Regulators, however, stress the need for mechanistic transparency over "black-box" outputs, with the European Food Safety Authority (EFSA)

targeting broad NAM adoption by 2027[1] while still requesting new in vivo studies in some cases (Schmeisser et al., pp. 4-1, 2023; Hartung, 2023; Wood et al., 2025). Finally, concrete use cases in genetic toxicology and manufacturing show how AI improves sensitivity, specificity, and quality control through automated imaging, next-generation sequencing, and classifier systems (Alnasser et al., 2025; Bhattamisra et al., 2023). Certain areas, such as "skin and eye irritation, skin sensitization and genotoxicity", are already reliably addressed by NAMs and should be required and accepted where validated (Eskes et al., 2007; Alépée et al., 2019; Strickland et al., 2022).

Ecotoxicology shows a similar trajectory. ML algorithms enhance prediction accuracy and lessen the need for experimental validation by analyzing large datasets. Regulatory summaries link this to reduced animal use as models become more accurate and reliable. This aligns with agencies such as the Organization for Economic Co-operation and Development (OECD), which are integrating alternative methods into risk management rather than "relying heavily on animal data to establish safety standards and make decisions about the approval or restriction of various products" (Pastorino et al., 2024).

The clinical and consumer domains reflect this shift. In dermatology, AI supports personalized regimens, outcome prediction, and data-driven treatment selection, improving patient experience (Kania et al., 2024). Ingredient-level AI screening for allergens/irritants enables personalization with safety (Hash et al., 2025)

Ethically, AI should augment human judgment within the longstanding traditions of applied ethics and risk assessment (Belenguer, 2022), aligning with the broader push for alternatives (Madden et al., 2020). Adoption gaps persist where usability and communication with medicinal chemistry are weak (Kapustina et al., 2024). Some argue that animal use remains justifiable in medicine when no alternatives exist and the benefits are substantial (less so in cosmetics), a distinction that shapes policy and practice (Kabene & Baadel, 2019).

Overall, AI now connects human-relevant models, probabilistic assessment, and real-world feedback across testing. It does not solve every endpoint, but it reduces animal reliance and accelerates evidence generation, especially where regulation, validated genotoxicity and ecotoxicology models, and end-to-end innovations open the door (Kimura et al., 2025; Alnasser et al., 2025; Pastorino et al., 2024; Shiammala et al., 2023).

---

[1] EFSA, 2021b. EFSA strategy 2027. https://www.efsa.europa.eu/sites/default/files/202 1-07/efsa-strategy-2027.pdf.

## 2.2 Benefits of the Transition

AI facilitates the shift to ethical and consumer-aligned testing, particularly evident among younger buyers. The "Cruelty-Free" label enhances brand image, attracts millennials and Gen Z consumers concerned with ethical consumption, and increases market share and loyalty (Mendivelso et al., 2024, p. 14). At the same time, AI promotes transparency in raw material procurement and ethical production procedures, with sustainability initiatives significantly influencing preferences and "promoting transformative change" (Winarto & Wisesa, 2024, pp. 20–21). This ethical pressure is reinforced by the EU's 2013 ban on animal testing for cosmetics, which has made non-animal approaches a regulatory and reputational necessity (Silva & Tamburic, 2022; Kimura, 2025).

Scientifically, human-relevant approaches give stronger signals. According to Movia and Prina-Mello (2020, p. 8), "Human-relevant approaches offer crucial advantages of speed and much more robust and exacting data than any animal experiment could deliver," and the toolbox of non-animal testing models is growing, improving disease understanding and discovery. At the same time, alternative methods and invertebrates can reduce costs and time, increase efficiency and scale of testing, making evaluation more systematic (Pastorino et al., 2024). Public support is not insignificant - "Three-quarters of Europeans are in favour of phasing out animal testing" - and this social license matters for adoption (Marshall et al., 2022, p. 2). Together, these trends strengthen the argument that NAMs, combined with AI, improve predictivity while respecting welfare (Pastorino et al., 2024).

In the realm of cosmetics safety, tangible benefits are evident. As mentioned before, AI-enabled in silico tools are less expensive, quicker, and not exclusive to laboratories compared to in vitro models (Grech et al., 2024). "In silico methods also offer significant advantages in the toxicology of cosmetic ingredients via safety assessments, enabling high-throughput screening of extensive chemical libraries without requiring physical experiments. (Grech et al., 2024, p. 8). From a human-relevance angle, data obtained through individual animals are "more akin to technical replicates", which limits reflection of the "diversity of human genetics", and in any case, "over 110 million animals, including mice, rats, frogs, dogs, cats, rabbits, hamsters, guinea pigs, monkeys, fish, and birds, are sacrificed per year, in the US alone," so reduction is meaningful (Zushin et al., 2023, p. 2; Pai et al., 2024, p. 10330, citing People for the Ethical Treatment of Animals (PETA)).

Overall, AI offers substantial benefits heralding a new era in cosmetic and pharmaceutical safety assessment (Grech et al., 2024). "The different natures of tests conducted on animals in the fields of medicine and cosmetics tend to have clear negative implications". However, the "benefits in the field of medicine outweigh those in cosmetics" (Kabene & Baadel, 2019, pp. 8, 9).

Drug discovery and development have also sped up. Authors report that AI can provide a solution through predictive modeling, cutting time and cost to optimize candidates (Patel & Patel, 2024). The field of genetic toxicology "has evolved from traditional methods to advanced high-throughput technologies, leading to improvements in chemical risk inference and associated disease understanding. The field is crucial for predicting long-term effects, assessing drug hazards, and designing strategies for hazard prevention. Non-animal testing approaches and HTS are vital for maximizing the potential of genetic toxicology." (Alnasser et al., 2025, p. 16).

"In vitro and in silico methods, along with the use of invertebrates as alternative models, are poised to play pivotal roles, offering efficient, ethical, and cost-effective alternatives. As technology, data science, and validation protocols progress, researchers can anticipate a future where ecotoxicological assessments are not only more predictive but also more aligned with the principles of humane and responsible scientific practices" (Pastorino et al., 2025, p. 8).

Romberg et al. (2012, p. 105) emphasize the "consistency approach" in vaccine testing, where in vitro analysis ensures batch quality during production rather than relying on endpoint animal tests. This stepwise replacement, such as shifting challenges to serology, introducing in-process in vitro tests, and phasing out animal use after sufficient parallel validation, reduces variability and ethical burden. As they note, regulators and assessors should provide the replacement of in vivo by in vitro testing. Importantly, animal testing does not always guarantee human safety, since interspecies differences can trigger unexpected reactions, making alternative methods a stronger predictor (Singer, 1975a, as cited in Mendivelso et al., 2024).

At the interface between wet lab and computation, MPS and digital methods add power. AI has augmented the pharmaceutical industry with novel technologies, allowing scientists to study molecules more comprehensively. Advancements have led to an increase in promising alternatives to animal testing, including multi-OoC platforms (Malik et al., 2024; Kimura et al., 2025). "AI-powered LMs offer an unprecedented opportunity for medical officers and others who ensure that drugs are safe and effective in supporting RWE (real-world evidence) generation for regulatory decision making and better patient outcomes" (Liu et al., 2021, p. 2602). "Organ-on-chip models possess the potential to predict human-specific side effects more

accurately. They are also invaluable in scenarios where patient populations are tiny, offering a new dimension of personalized medicine". (Yamazaki and Ishida, 2025, p.3)

Biosimulations and digital twins bring scale. Reviews highlight the ethical advantages of true replacement (OoC, DTs), as well as the stronger predictive power achieved by integrating multi-omics and clinical data (Gangwal & Lavecchia, 2025). Cost and throughput benefits are concrete: "VeriSIM Life's AI-enabled biosimulations allow evaluation of 1000 times more candidates in the same timeframe," which changes triage and reduces late failures due to species gaps (Gangwal & Lavecchia, 2025, pp. 7, 13). Personalized modeling is also possible when organoids are linked to AI, giving patient-specific responses in preclinical research (Gangwal & Lavecchia, 2025; Kimura et al., 2025).

Regulatory momentum is building to make these benefits usable. U.S. policy windows and pilots are opening. Authors note "AI-driven technologies have the potential to increase the reliability and efficacy of medical products but have stressed the need to safeguard patients," so governance must keep pace (Shiammala et al., 2023, p. 2). Programs and guidance around OoC, NAMs, and patient-evaluation systems are expanding in both the U.S. and EU (Schmeisser et al., 2023; Pistollato et al., 2025).

Legislative changes, such as the FDA Modernization Act 2.0, also enable the use of NAMs and AI as alternatives, signaling that the pharmaceutical industry recognized a transformative shift in the regulatory landscape for early-stage drug development. This legislation officially authorizes the use of advanced technologies, including cell-based assays, predictive computer models, and OoC, as alternatives to traditional non-human animal models for non-clinical drug testing. (FDA, 2025; Yamazaki and Ishida, 2025, p. 3; Connor et al., 2022).

From an industrial strategy perspective, the use of AI-powered systems enables companies to scale innovation, reduce infrastructure needs (e.g., animal housing and care), and lower overall operational costs (Samah H. O. Zarroug, 2025; Fontana et al., 2021). Madden et al. (2020, p. 149) highlight that "*In silico* (computational), *in vitro*, omics, organ-on-a-chip technology, high-throughput screening (HTS) and mathematical biology can all play a role in providing complementary information" associated with safety assessments, while still maintaining regulatory confidence. These tools also make safety testing more feasible for small and medium-sized enterprises (SMEs), democratizing access to high-quality assessment methods. AI methodologies promise significant cost and time savings.

Taylor and Alvarez (2020, p.2) quantify that “replacement methods, particularly in silico techniques for the more long-term animal tests, can be significantly cheaper than the corresponding animal test.”. Van Berlo et al. (2021, p. 8) state that traditional drug development is enormously costly and inefficient, with “12 years” or more required to bring a single drug to market, and costs reaching up to “$2.6 billion” per successful drug. AI-driven NAMs offer an opportunity to fail early and cheaply, refining candidates before entering costly animal and human trials. Pharmaceutical Research and Development (R&D) can achieve “up to 70% cost reductions” through AI integration; similarly, the cosmetics industry benefits economically from bans on animal testing, driving innovation (Rudroff, 2024; Silva & Tamburic, 2022).

Practical access is improving too. Computational platforms and data resources lower the barrier for labs and regulators, and there is a push toward open, shared tools, such as the Integrated Chemical Environment (ICE), Sysrev, and DistillerSR, which facilitate data curation and mining for chemical safety assessment. Freely accessible computational tools, such as OPERA, VEGA, T.E.S.T., and the QSAR Toolbox, democratize toxicological research and support regulatory submissions (Ram et al., 2022). In biomedicine and cosmetics alike, AI is becoming increasingly common in screening and design; however, some authors still consider adoption uneven, highlighting the need for training and improved interfaces (Kapustina et al., 2024; Poh & Stanslas, 2024).

Improved access to NAM data is essential for validation and regulatory assessment. Enriching databases, such as ECHA’s chemical inventory, can facilitate a deeper understanding and greater acceptance. The concept of *safe spaces* was proposed to allow free exchange of methodologies and data between industry and regulators without immediate regulatory consequences. This could involve the simultaneous submission of NAM data alongside traditional data, requiring sufficient expertise and resources from all parties. (Cronin et al., 2025). These infrastructural pieces make the transition stick, not just a short trend (Schmeisser et al., 2023).

Van Norman (2019, pp. 851, 845) stresses the scientific limits of animal models, arguing that “because animals and humans are classic examples of incompletely understood complex systems” and that “it may simply be scientifically invalid to assume that toxicity in any one species can reliably predict toxicity in any other.” He also emphasizes that alternatives are “significantly faster and less expensive.” Despite bans in 2004, 2009, and 2013, new in vivo tests still occur under REACH, since registrants often begin with in vitro methods but ultimately conduct animal tests to comply. This highlights the need for stronger policy follow-through (Knight et al., 2021; Fentem et al., 2021).

Global regulatory trends reinforce these benefits. Sreedhar et al. (2020) document the cascading bans on cosmetic animal testing in countries such as India, New Zealand, Taiwan, and Turkey, driven by both ethical advocacy and scientific readiness. These bans signal a shift in market expectations, making non-animal methods a strategic necessity. "The EU is a wide marketplace for all cosmetics companies" and "this policy has forced cosmetics companies to develop alternatives" (Sreedhar et al., 2020, p. 114).

Many argue that NAMs can provide complementary information to develop a complete picture, reminding us that tests in animals are not always predictive of human responses; therefore, method choice should follow the decision context, not tradition (Taylor & Alvarez, 2020; Madden et al., 2020). When this logic is applied to MPS with AI analytics, teams increase the reliability of predicting both efficacy and safety and reduce costly late-stage failures (Van Berlo et al., 2021; Malik et al., 2024).

Rinaldi et al. (2020, p. 1) describe a pivotal legal ruling in Italy that places the burden of proof on researchers to justify the use of animals, stating, "the burden of proof as to the usefulness of the use of animals for laboratory tests is borne by the experimenter." This reflects broader shifts in ethical and legal perspectives toward prioritizing alternative methods. López and Franceschini (2025) note that the Treaty of Lisbon codifies animal sentience in EU law, leading to harmonized ethical governance across member states.

In summary, the benefits of transitioning to AI-supported, non-animal testing are broad, spanning higher predictive accuracy, cost efficiency, regulatory compliance, ethical legitimacy, transparency, market access, and scalability. As Bhattamisra et al. (2023, p. 11) observe, "Researchers are fascinated by the recent developments in AI, especially its application in healthcare and pharmaceutical research and service. Smart hospitals and healthcare facilities equipped with AI, ML, and Big Data will shape the future of the healthcare sector. Pharmaceutical industries are in constant advancement with their technologies, and AI will be an opportunity for minimizing the cost and time for drug development." Overall, the literature suggests these methods are becoming the gold standard in testing and product development.

### 2.3 Costs and Challenges of the Transition

AI and ML systems operate faster than humans but can propagate errors; therefore, the U.S. FDA and other U.S. agencies impose strict controls during the drug development process.

Accountability rests with developers and users, who must validate models before submitting to regulators. Ethical issues, such as privacy, bias, and fairness, primarily stem from data quality and insufficient validation, not from algorithms alone. Historical evidence suggests that fears of mass job loss are overstated; therefore, policy should prioritize education and upskilling over stifling innovation. Misuse and misleading content highlight safety and security risks, and extrapolating beyond test conditions raises concerns about generalizability and ethics.

Additionally, opacity and proprietary limits can restrict explainability. The FDA is developing trustworthy AI guidance that covers explainability, reliability, privacy, safety, security, and bias reduction (Niazi, 2023).

Public opinion and practice diverge. Kabene and Baadel (2019) note that animal testing can benefit both humans and animals, for instance, in the context of shared diseases. According to Linzey A[2]., from Kabene and Baadel (2019, pp. 2,3), "about 50–100 million animals" are used for experiments annually, and "1.37 million" for U.S. drug research in 2010. A published report[3] stated that acceptance splits by use: for medical research, 65% are "acceptable," 17% are "mostly unacceptable" if no alternative is available, and 17% are "not acceptable"; for cosmetics, 80% are "unacceptable." On necessity: 38% "completely necessary," 23% "somewhat necessary," 20% "not very necessary," 16% "completely unnecessary".

In the cosmetics industry, labels and messaging often confuse buyers. One study states that the absence of clear and universal regulation encourages misleading marketing practices, such as "greenwashing", and calls for a consistent, transparent, and clearly established regulatory or license system, because, if public policies promote the adoption of sustainable practices and discourage "greenwashing", it could improve consumer confidence and incentivize companies to invest in greener technologies and production processes (Timpanaro & Cascone, 2025, p. 10).

Overall, policy, perception, and practice must align before costs go down. Authors note that requirements persist in some contexts in Mainland China, creating practical tension for global brands (Silva & Tamburic, 2022). The luxury sector reports that major houses still conduct animal testing and sell in Mainland China, which demonstrates how compliance pressures can override cruelty-free goals (Amador, 2021). More broadly, "the conflict between

---

[2] Linzey A. The Global Guide to Animal Protection, 1st ed. USA: University of Illinois Press; 2010.

[3] Anonymous. Statistics of scientific procedures on living animals, Great Britain 2016. [cited 2019 October]; Available from: https://www.gov.uk/government/statistics/statistics-of- scientific-procedures-on-living-animals-great-britain-2016

REACH and the Cosmetic Regulation poses a serious dilemma for all segments of the cosmetic industry: for ingredient manufacturers, as they can be legally required under REACH to conduct in vivo tests on their ingredients, but the cosmetic market may reject ingredients with such tests; for cosmetic brands, as they cannot easily identify REACH testing of ingredients in their supply chain, but if such testing is identified, a brand risks backlash from consumers if it continues to use the ingredient, but finding an alternative can be difficult and costly". This demonstrates why companies must comply with different rules in different markets and why mutual acceptance is necessary (Knight et al., 2021, p. 668; Fentem et al., 2021).

Legislation is advancing, but not at a sufficient pace. "The European Directive 2010/63 on the Protection of Animals for Scientific Purposes" articulates the need for animal welfare at its heart and throughout, stating "the final goal of full replacement of procedures on live animals for scientific and educational purposes as soon as it is scientifically possible to do so", "However, it is apparent when studying the available data on animal use for scientific purposes that effective implementation of the Directive alone is not sufficient to drive meaningful reduction, let alone full replacement."[4] Global harmonization of testing strategies and standardization of the tools will be necessary to improve confidence in NAMs data (Marshall et al., 2022, pp. 1-12).

International inconsistency persists: "over 30 countries" ban animal testing for cosmetics, but enforcement, exemptions, and legal clarity vary, and bans often do not cover imported cosmetics (Sreedhar et al., 2020, p. 114).

Technical and scientific gaps persist as well. OoC and organoid platforms represent transformative tools; however, their widespread adoption remains hindered by challenges such as "scalability, system complexity, and reproducibility". Kimura et al. (2025, pp.1-14) explain that while these systems simulate physiological responses more accurately than traditional in vitro models, technical difficulties remain in "Reproducibility, Immaturity and Limited Complexity, Vascularization, Access and Culture, Ethical and Regulatory, High Cost and Limited Accessibility, and Incomplete Recapitulation of In Vivo Functions". Gaps remain for complex endpoints, as DART, respiratory sensitization, and immunotoxicity (Westmoreland et al., 2022). Transitioning to new testing methods requires

[4] Annex B

significant research and development investments, infrastructure changes, and training personnel.

Introducing AI into existing businesses rarely yields immediate productivity gains. To reach its market potential, firms must make both tangible investments (such as equipment, software, and infrastructure) and intangible investments (including business and technology development, organizational restructuring, and worker training). Studies show industries often face output declines when adopting new technologies, due to adjustment costs and learning delays. In the case of AI, companies need extensive research and time to identify opportunities in business, operations, technology development, and production processes. Adoption also requires structural changes before core or peripheral functions can be transformed. Implementation typically involves ongoing human-machine interaction, since many tasks are complex, and algorithms are not always readily available.

This iterative learning process demands significant effort and resources before the productivity benefits of AI can be fully realized. (Brynjolfsson et al., 2019; Atkeson & Kehoe, 1993; Hornstein & Krusell, 1996; Jovanovic & Nyarko, 1994; Greenwood & Yorukoglu, 1997; Jovanovic & Rousseau, 2005; McKinsey, 2017; Ransbotham et al., 2020; Raisch & Krakowski, 2021, as cited in Lee et al., 2022).

Data, bias, and privacy are persistent worries. Biased and homogeneous datasets skew outputs, imaging variability, limited time, and inadequate user training hinder performance (Gholizadeh et al., 2024). In vitro tests often probe a single Adverse Outcome Pathway (AOP), and technical limitations remain. Handling biometric and consumer data requires strong protective measures (Grech et al., 2024; Hash et al., 2025). Although interspecies differences limit the applicability of animal data, alternatives must still be validated to ensure human safety and efficacy (Singer, 1975b, as cited in Mendivelso et al., 2024). Authors caution that “deep learning approaches” are “black-box models” and are “considered unethical since stakeholders may doubt the accuracy and impartiality of AI suggestions”, and that “bias is a significant issue (Data bias, algorithm bias, and Implementation bias)” when data reflect historical inequalities, “leading to incorrect predictions and recommended treatment” (Patel & Patel, 2024, pp. 683, 688).

In practice, stakeholders still worry about generalization and fairness in evidence synthesis and trial modeling (Niazi, 2023). These are fixable problems, but they take time and shared standards.

Data access and costs remain high. Labeled data are needed for LM fine-tuning, but data sharing is constrained: “AI faces challenges in reproducibility because researchers have

difficulty reproducing many vital results, hindering their real-world applications." (Liu et al., 2021, p. 2605). Moreover, organizations face high initial costs and skills gaps (Malik et al., 2024; Gholizadeh et al., 2024). Even advanced digital twins and OoCs only approximate physiology, and model limitations and computational complexity complicate scale-up. Regulators demand broad validation due to *black-box* behavior complicating review, and mechanistic clarity is preferable (Gangwal & Lavecchia, 2025; Pastorino et al., 2024; Wood et al., 2025). Despite upfront costs and time, "the transition to in vitro would pay back after four years" (Romberg et al., 2012, p. 103).

Legal conflicts and legacy rules raise costs. Under REACH, ECHA has requested new in vivo tests for cosmetic-only ingredients, which clash with the EU cosmetics ban (Knight et al., 2021). Harmonization remains uneven. The requirements of Mainland China demonstrate how compliance pressures can override cruelty-free goals (Silva & Tamburic, 2022; Akkermans et al., 2020; Amador, 2021). Frameworks from the OECD and the EU Reference Laboratory for Alternatives to Animal Testing (EURL-ECVAM) exist, but their adoption is slow, and *industry lobbying* and entrenched interests have sometimes slowed or diluted advances (Romberg et al., 2012; López & Franceschini, 2025).

Non-Governmental Organization (NGO) activism and consumer scrutiny increase reputational risks, so clear and unambiguous guidance should be provided with case examples that show potential scenarios when animal testing is not necessary. Such guidance could have prevented recent cases where animal testing was performed unnecessarily, helping regulators accept non-animal data. Over time, this can help reduce cross-border conflicts (Wood et al., 2025, p. 25; Amador, 2021). L'Oréal has led the use of reconstructed human skin models ("pruebas In vitro") since around 1979, using synthetic skins (Episkin) in centers located in Lyon, France, and Shanghai, China. These models allow for evaluating product safety and effectiveness without the need for animal testing (Belén, 2020a; L'Oréal, n.d., p. 13, as cited in Mendivelso et al., 2024).

Standards and reproducibility are recurring needs. Authors call for "fitness-for-purpose criteria for NAMs", noting in silico predictions are rarely used alone for regulatory decisions (Madden et al., 2020, pp. 149, 168). Multi-OoC advocates caution that an Absorption, Distribution, Metabolism, and Excretion (ADME)-chip[5] is still immature for regulatory testing (Van Berlo et al., 2021, p. 11). Accurate in silico work depends on

[5] "absorption, distribution, metabolism, and excretion" (Van Berlo et al., 2021, p.8)

curated chemical and biological data. “KNIME”[6] (the first tool to integrate all key steps of chemical and biological data curation into a single framework) helps integrate curation steps and flags mixtures and conflicting replicates. Still, manual review remains necessary, especially in small datasets, and in particular for neglected diseases, where every data point matters (Alves et al., 2021, pp. 75, 77).

Worldwide regulatory acceptance is essential but complex, and data curation and overload can cause reproducibility issues, so scientist groups stress validation to ensure model reliability (Ram et al., 2022; Poh & Stanslas, 2024). These practical hurdles slow down scaling even when the science is ready.

Single NAMs are rarely sufficient. Regulators require integrated frameworks and weight-of-evidence (WoE) approaches. An Environmental Safety Assessment (ESA) utilizing Integrated Approaches to Testing and Assessment (IATAs) is envisioned, combining toxicity, bioaccumulation, and toxicokinetics (TK), with coordination with the Water Framework Directive being essential (Cronin et al., 2025). As emphasized by Fentem et al. (2021), a paradigm shift to NGRA across diverse datasets is necessary; however, implementing NGRA still requires the development of coordinated scientific, regulatory, and economic infrastructures that are ongoing.

As mentioned earlier in the FDA (2025, p.1) roadmap, “90% of drugs tested on animals ultimately fail in human trials, predominantly due to safety and/or efficacy issues.” On the other hand, some authors counter that statement - “animal rights activist Peter Egan argued that “90 percent of all tests on animals as models for human diseases fail.” This statistic is meaningless given that the 10% successes greatly help to enhance human health” (Hunter, 2023, p.1).

Overall, while AI-supported non-animal testing offers significant advantages, its adoption is limited by intertwined scientific, regulatory, financial, ethical, and cultural challenges. Progress will require technical innovation, regulatory reform, international harmonization, public investment, and cultural change within the scientific community.

### 2.4 Implementation Pathways

[6] Neves B, Moreira-Filho J, Silva A, et al. Automated framework for developing predictive machine learning models for data-driven drug discovery. J Braz Chem Soc 2021; 32: 110-132.

The shift to AI and NAMs rests on coordinated strategies, including structured validation, regulatory pilot programs, policy development, institutional coordination, and technology transfer (EMA, 2020; Alves et al., 2021; Van Berlo et al., 2021; Fentem et al., 2021; Romberg et al., 2012).

One of the clearest entry points is through regulatory guidance and annexes. The European Medicines Agency's (EMA) 2020 Strategic Reflection highlights progress in the 3Rs and recommends annex-based guidance to prioritize NAMs for meeting testing requirements (EMA, 2020). This annex-based model allows targeted implementation without requiring a system-wide overhaul, and it is supported by training, education, and transparent roadmaps tied to relevant guidelines, promoting knowledge sharing among regulators, researchers, and industry (EMA, 2020). Historically, CPMP/SWP/728/95 (EMA, 1997) demonstrates the slow pace of regulatory change and the ongoing need to revise assumptions that default to animal methods tests.

Meanwhile, road mapping and collaboration efforts support implementation. From the Brussels workshop on The Roadmap Towards Phasing Out Animal Testing for Chemical Safety Assessments (Dec 11–12, 2023), Cronin et al. (2025) highlight endpoint-by-endpoint analysis to identify alternatives, fill gaps, and determine development needs. Their recommendations include involving a wide range of stakeholders through workshops and communication, improving cooperation among EU agencies and expert committees, and creating advisory scientific committees focused on developing and implementing non-animal methods.

They emphasize the importance of mutual acceptance of data across jurisdictions, engagement with non-EU partners and global bodies such as the OECD and the World Health Organization (WHO), as well as regulatory measures to enhance chemical data accessibility and foster deeper engagement with the scientific community.

To support these collaborative efforts, scientific priorities and performance criteria are also central. NAMs should be benchmarked not only against legacy animal data but against human exposure and disease outcomes, with clear criteria for robustness, reliability, reproducibility, sensitivity, and specificity. Achieving this requires curated, standardized datasets with both qualitative and quantitative human data, as well as expanded applicability domains for complex substances (e.g., nanomaterials, polymers).

Additionally, it requires stronger in vitro ADME and quantitative in vitro to in vivo extrapolation (QIVIVE) models. Moreover, NAMs should generate meaningful Points of Departure (PoDs) for risk and hazard classification, including the repurposing of available

data, and prioritize human-based approaches across relevant dose ranges rather than predicting NOAELs/LOAELs[7] from animal studies. Progress also relies on advances in exposure science, including the integration of life cycle and accidental exposures, systematic hazard–exposure linking, and the broader sharing of practical experiences across different substances and scenarios, especially for complex endpoints such as DART (Westmoreland et al., 2022).

Validation strategies remain essential. Alves et al. (2021) call for OECD best practices in cheminformatics and AI model validation, while Westmoreland et al. (2022) highlight tools such as QSAR, QMRF, IATA, and RAAF to ensure the interpretability and reproducibility of models. Exposure assessment is equally essential: expanding the Threshold of Toxicological Concern (TTC), applying internal TTC, and adapting mature pharmaceutical and cosmetic exposure practices can reduce reliance on animals. Cronin et al. (2025) add that stronger TK models of internal exposure in humans and environmental species would reduce uncertainty and strengthen regulatory confidence.

Institutional coordination is essential for progress. Fentem et al. (2021) emphasize the European Partnership for Alternatives to Animal Testing, which aligns policy, funding, and industry practices. Current initiatives showcase this momentum, including ECHA workshops and challenge papers, EFSA's 2022 roadmap on NAMs, and EMA's 3Rs Working Party. Industry and NGOs also play a role; CEFIC promotes global standards and ongoing verification, while HSI advocates for transparent roadmaps, milestones, and safe data-sharing environments. EU-funded projects like PARC, the ASPIS Cluster, and EPAA's "NAM Designathon 2023" further advance tiered strategies and classification systems (Cronin et al., 2025, p.5).

The concept of scientific legitimacy is evolving. Fentem et al. (2021) contend that the EU's commitment to non-animal methods under REACH requires a paradigm shift in chemical safety assessment, incorporating NGRA. While modern non-animal methods already influence EU assessments for cosmetics and food, chemical regulation remains behind. Under the Chemicals Strategy for Sustainability (CSS) and the Green Deal, experts advocate for a flexible, science-based framework that integrates multiple data sources, making animal testing a last resort.

---

[7] No-Observed-Adverse- Effect Levels (NOAELs), Lowest-Observed-Adverse-Effect Levels (LOAELs) - Agency for Toxic Substances and Disease Registry. (1995). *Appendix A: User's guide*. In *Toxicological profile for tetryl (2,4,6-trinitrophenyl-N-methylnitramine)* (pp. A-1-A-7). U.S. Department of Health and Human Services, Public Health Service. https://www.atsdr.cdc.gov/toxprofiles/tp80-a.pdf

Reaching this tipping point demands collaboration among policymakers, scientists, regulators, and lawyers, with harmonization of acceptance criteria across sectors and regions as a top priority. Internationally, the U.S. ICCVAM roadmap and the European Partnership for the Assessment of Risks from Chemicals' (PARC) NGRAroute initiative offer guidance, with NGRAroute aiming to establish NGRA as the long-term default risk assessment approach across EU chemicals legislation (Cronin et al., 2025).

Policy frameworks establish evidence standards. The FDA Predictive Toxicology Roadmap details how emerging methods undergo regulatory review: EMA, EFSA, the United States National Institute of Standards and Technology (NIST), the U.S. FDA, and Congress are strengthening guidance on safely implementing AI as software tools and medical devices in the drug approval process. (Connor et al., 2022; Yamazaki & Ishida, 2025).

From this point, regulatory design and legislative texts become essential. Updating guidance and refining tools such as REACH annexes can explicitly enable tiered schemes that combine exposure assessments and NAMs without resorting to animal studies. Regulators are encouraged to find a balance between flexibility and legal certainty, allowing the widespread use of validated NAMs across different sectors, and aligning their efforts with the EU CSS and the "One Substance - One Assessment" principle. Additionally, expanded channels for pre-submission scientific advice can help clarify when and how NAMs might fulfill regulatory information requirements (Westmoreland et al., 2022, p.2).

The paradigm shift itself requires broad, multidisciplinary expertise in toxicology, in vitro methods, QSAR, (PBPK)/QIVIVE, exposure science, statistics, and regulatory science. Training needs to span from university curricula to ongoing professional development. Key actions include raising awareness, promoting more NAM-based regulatory submissions, and developing a common language among industry, authorities, NGOs, and society. Forums like the EPAA are helpful by offering case studies, opportunities for alignment, and connections to broader EU initiatives such as the Green Deal (Westmoreland et al., 2022). Concrete examples also demonstrate how implementation occurs in practice. Romberg et al. (2012) discuss vaccine consistency testing carried out through collaborative studies involving regulators, manufacturers, and reference laboratories, where harmonized standard operating procedures (SOPs) were crucial for regulatory approval.

Similarly, Liu et al. (2021, p. 2597) recommend "establishing a voluntary-based biomedical labeling consortium to facilitate biomedical data annotation. Second, a reorganization of publicly available biomedical corpus would be useful for addressing specific BioNLP tasks." "Third, labeling tools could be a solution to facilitate the manual data curation and annotation process", so that AI models have diverse and reliable training sets to depend on.

In Germany, the German Research Foundation (DFG) emphasizes planning data management, enabling access, and ensuring long-term archiving. Supporting this, the 3R-Center Rhine-Neckar provides educational resources on selecting journals (via the Directory of Open Access Journals [DOAJ]), reimbursing publication costs, converting animal data according to the FAIR principles ("Findable, Accessible, Interoperable, Reusable"), and proper data storage (Neuhaus et al., 2022, p. 398).

Programs like Innovative Science and Technology Approaches for New Drugs "(ISTAND)" and the "SaMD Action Plan" expand Drug Development Tools (DDTs) and accelerate the adoption of AI/ML in regulatory science. European and international initiatives focus on human-centric NAMs to improve translatability, emphasizing cross-sectoral dialogue to harmonize strategies and enable evidence-based policy (Pistollato et al., 2025, p. 1). Achieving the National Research Council (NRC) "Option IV" vision (HS human cell testing and computation) requires ongoing investment in developing, validating, recognizing, and gaining regulatory acceptance for in silico and big-data methods, as well as intelligently integrating weight-of-evidence approaches (Ram et al., 2022, pp. 1–7).

Technological innovations further expand pathways for implementation. Kimura et al. (2025) demonstrate that microfluidic systems replicate in vivo-like conditions with modularity, scalability, and high spatial control through manipulating microstructures and flow, enabling precise control of cell culture environments. In consumer-facing sectors, Elder et al. (2024) show how AI-enabled remote diagnostics, augmented reality (AR), and 3D modeling support personalized formulations, allowing patients to safely operate home-use laser devices and receive tailored regimens via virtual platforms. Across all these domains, dialogue and interdisciplinary literacy remain crucial. Madden et al. (2020) emphasize the importance of communication among biology, data science, and regulatory science, while legal precedents can further promote adoption. For example, a ruling by the Italian Council of State shifted the burden of proof to researchers, leading to widespread use of in vitro and in silico protocols and transforming laboratory practices (Rinaldi et al., 2020). Overall progress is achieved through a combination of strategies: regulatory annexes and pilots, validation frameworks, institutional partnerships, industrial use cases, legal rulings, digital infrastructure, education, and global

cooperation. No single approach is enough, but together, they unlock the scientific, ethical, and economic potential of NAMs. As Malik et al. (2024) conclude, realization requires collaboration among academia, industry, policymakers, technologists, and environmentalists to ensure sustainable, energy-efficient, economically viable, and safe AI integration.

## Chapter III – Theoretical Approach

The transition from animal testing to AI-based methods and NAMs in cosmetics and pharmaceuticals offers substantial ethical benefits, potential financial savings, and technological advancements. However, regulatory acceptance, validation, upfront costs, and global harmonization remain significant challenges. Coordinated efforts among regulatory bodies, academia, industry, and consumers are crucial to harness the full potential of AI-supported NAMs responsibly and effectively (Fentem et al., 2021; EMA, 2020; Malik et al., 2024).

Animal models often fail to accurately predict human outcomes; for example, a significant proportion of drug candidates that pass animal tests ultimately fail in human trials (FDA, 2025). However, some authors advise caution when interpreting headline failure statistics, reminding us to consider the broader context. (Hunter, 2023). Animal testing raises significant ethical issues and faces strong societal and consumer opposition, particularly in the cosmetic industry (Singer, 1975, as cited in Mendivelso et al., 2024; Marshall et al., 2022). Animal studies are expensive, time-consuming, and resource-intensive, with carcinogenicity bioassays costing over $1 million and lasting several years (Taylor & Alvarez, 2020). Conflicting regulations (e.g., EU Cosmetics Regulation vs. REACH) create compliance difficulties and can mandate animal testing despite bans (Knight et al., 2021; Silva & Tamburic, 2022). AI integrates in vitro, in silico, and ex vivo data, providing models that are more directly relevant to human biology and thereby improving prediction accuracy (Movia & Prina-Mello, 2020; Grech et al., 2024; Kleinstreuer & Hartung, 2024). AI-supported NAMs align with the 3Rs principle by replacing or reducing animal use, addressing ethical concerns, and fulfilling consumer demand for cruelty-free products (EMA, 2016; Pistollato et al., 2025).

AI methods significantly accelerate testing and reduce costs, enabling faster candidate screening and development (Van Berlo et al., 2021; Gangwal & Lavecchia, 2025). Emerging technologies, such as OoC systems, digital twins, and machine-learning models, enable complex biological simulations and HTS (Kimura et al., 2025; Gangwal & Lavecchia, 2025). AI also allows personalized safety assessments and formulations, supporting consumer-facing applications, particularly in dermatology (Elder et al., 2024; Hash et al., 2025).

AI models require extensive validation to gain regulatory trust and confidence. Current frameworks and harmonization are evolving but incomplete (Alves et al., 2021; EMA, 2020; Fentem et al., 2021; Cronin et al., 2025). AI systems face issues such as "black-box" opacity, computational demands, reproducibility concerns, and the need for large, high-quality, and unbiased datasets (Patel & Patel, 2024; Liu et al., 2021; Shiammala et al., 2023). Despite advances, AI and NAMs may not yet fully replicate complex systemic interactions, metabolic processes, or immune responses captured in whole-animal studies (Van Berlo et al., 2021; Kimura et al., 2025).

Many NAMs are already usable/accepted for specific endpoints, but gaps persist for DART, respiratory sensitization, and immunotoxicity, areas that require deeper mechanistic coverage and integrated strategies (Westmoreland et al., 2022; Neuhaus et al., 2022). Significant upfront costs for infrastructure, skilled personnel, and validation can limit adoption, especially for smaller companies (Malik et al., 2024; Gholizadeh et al., 2024; Lee et al., 2022). Conflicting or incomplete international regulations slow adoption and create compliance uncertainty (Sreedhar et al., 2020; Knight et al., 2021; Amador, 2021; Akkermans et al., 2020).

With the pros and cons of AI and animal testing still being debated, the first research question emerges: *(RQ1) What are the main advantages and limitations of animal testing compared to AI-supported NAMs in the cosmetics and pharmaceutical sectors?*

This debate is framed by evidence of limited animal-to-human predictivity and high costs (FDA, 2025; Van Norman, 2019; Taylor & Alvarez, 2020), the ethical imperative to reduce animal suffering (Singer, 1975 as cited in Mendivelso et al., 2024; Marshall et al., 2022), and demonstrations that AI/NAMs can provide faster, human-relevant evidence while still facing challenges of scope, reproducibility, and explainability (Movia & Prina-Mello, 2020; Kimura et al., 2025; Patel & Patel, 2024; Liu et al., 2021; Westmoreland et al., 2022; Kleinstreuer & Hartung, 2024).

Given ethical bans and regulatory acceptance (e.g., EU Cosmetics Regulation), AI-supported in silico models and in vitro tests can fully replace animal testing for many cosmetics safety assessments (Grech et al., 2024; EMA, 2016), and validated non-animal strategies have already supplanted LLNA/GPMT in defined contexts (Eskes et al., 2007; Alépée et al., 2019; Strickland et al., 2022). In pharmaceuticals, AI complements traditional methods by enhancing early-stage discovery, prioritizing candidates, and facilitating mechanistic understanding, thereby reducing, but not yet entirely eliminating, the need for animal use (Madden et al., 2020; Van Berlo et al., 2021; Pai et al., 2024). In vaccine and biologics production, in vitro

"consistency" replaces animal batch release tests, improving reproducibility and ethics (Romberg et al., 2012).

AI-enabled data integration (e.g., PBK/QIVIVE, read-across, model ensembles) strengthens weight-of-evidence approaches (Silva et al., 2023; Blümmel et al., 2024; C. Westmoreland et al., 2022). The gradual integration of AI with existing non-animal methods can facilitate validation, regulatory acceptance, and scaling, using AI to fill gaps while ensuring safety (EMA, 2020; Connor et al., 2022; Fentem et al., 2021). This debate raises a second research question: *(RQ2) In which situations can AI replace or complement animal testing?*

As demonstrated throughout the literature review, the ethical imperative to replace, reduce, and refine animal testing (the 3Rs) is a driving force behind adopting AI and NAMs. Replacement is increasingly viewed as the most desirable goal (EMA, 2016; Silva & Tamburic, 2022), and public support for phasing out animal testing is strong (Marshall et al., 2022). Cruelty-free and sustainability signals shape demand and loyalty (Mendivelso et al., 2024; Winarto & Wisesa, 2024), while inconsistent labels invite "greenwashing" and call for more transparent labeling and standards (Timpanaro & Cascone, 2025).

Ethical governance must also encompass data ethics, including privacy, bias, and explainability, so that AI's benefits are realized without undermining trust (Hash et al., 2025; Niazi, 2023; Patel & Patel, 2024). Some argue that the limited use of animals in medical research remains more defensible than in cosmetics when no substitutes exist (Kabene & Baadel, 2019). Based on the presented review of the literature, the third research question emerges: *(RQ3) What ethical principles and guidelines should guide this transition?*

As stated before, agencies such as the FDA and EMA are developing guidelines to incorporate AI and NAMs into regulatory decision-making (EMA, 2020; Connor et al., 2022; Yamazaki & Ishida, 2025). The EU banned animal testing for cosmetics (Regulation 1223/2009), but REACH requirements sometimes still mandate in vivo studies, creating tensions and compliance challenges (Silva & Tamburic, 2022; Knight et al., 2021). Acceptance hinges on rigorous validation aligned with OECD principles and reproducibility standards. Documentation tools such as QMRF, IATA, and RAAF support uptake, alongside defined approaches and NGRA frameworks (Alves et al., 2021; Madden et al., 2020; Fentem et al., 2021). Concretely, OECD TGs and performance benchmarks (e.g., ICCVAM-style integrated LR/SVM models for sensitization) are becoming practical currencies of trust (Grech et al., 2024; Liu et al., 2023).

Global rules remain divided: more than 30 countries restrict cosmetic animal testing, but enforcement and exemptions vary, with Mainland China pressures complicating practice (Sreedhar et al., 2020; Amador, 2021). European foresight anticipates broader integration of NAM-based assessments by approximately 2027, with AI-assisted weight-of-evidence methods converging across sectors (Wood et al., 2025; Pastorino et al., 2024). So, the fourth research question becomes clear: *(RQ4) Which regulatory frameworks should companies adhere to during this transition?*

AI-driven in silico models, machine-learning classifiers (SVM, ANN, decision trees, Bayesian networks), OoC, MPS, and digital twins simulate human biology with rising fidelity and are increasingly integrated with omics and imaging for mechanistic insight (Madden et al., 2020; Kimura et al., 2025). Recent toolsets include deep-learning toxicity models (e.g., DeepTox, DeepDILI) and GAN-based generators (e.g., AnimalGAN), which can reproduce clinical-pathology patterns without animal data (Silva et al., 2023; Gangwal & Lavecchia, 2025). In vitro batteries (e.g., DPRA, h-CLAT, KeratinoSens, MUSST, LuSens) within integrated approaches have achieved regulatory-relevant performance and, in specific contexts, have replaced legacy LLNA/GPMT (Grech et al., 2024).

Ongoing challenges include "black-box" opacity, computational complexity, reproducibility, data bias/privacy, and single-AOP scope limits, issues for which standards (QMRF/RAAF), interpretable modeling, robust curation workflows, and benchmarking are designed to mitigate (Patel & Patel, 2024; Liu et al., 2021; Alves et al., 2021; Neuhaus et al., 2022; Madden et al., 2020). OoC/digital-twin platforms need continued validation for regulatory acceptance (Van Berlo et al., 2021; Kimura et al., 2025), while deployment depends on data-sharing consortia and usable interfaces (Liu et al., 2021; Kapustina et al., 2024; Ram et al., 2022). Outside the lab, dermatology diagnostics, personalized skincare, and virtual try-ons extend AI's role post-market (Elder et al., 2024; Hash et al., 2025). Ecotoxicology is following a similar trajectory toward ML-refined, mechanistic predictions with reduced reliance on animals (Pastorino et al., 2024).

At the same time, animal models still provide integrated whole-organism responses that can be difficult to fully emulate today, which helps explain the continued requests for animal data in some regimes and the different treatment of medical versus cosmetic contexts (Kabene & Baadel, 2019; Knight et al., 2021). With that, the fifth research question is: *(RQ5) Which technological implementations are most essential to adopt, and what are the main technological challenges?*

AI and NAMs offer significant cost and time savings compared to traditional animal testing. Carcinogenicity bioassays can cost over one million dollars and last for years, whereas in silico and AI methods accelerate early triage (Taylor & Alvarez, 2020; Van Berlo et al., 2021). Beyond direct test costs, biosimulations change portfolio economics by boosting throughput, enabling far more candidates in the same timeframe and reducing late-stage failures (Gangwal & Lavecchia, 2025). AI-enabled platforms reduce infrastructure needs (e.g., animal housing) and broaden access for SMEs (Fontana et al., 2021; L'Oréal, n.d.; Belén, 2020a, as cited in Mendivelso et al., 2024).

However, upfront investments in infrastructure, validation, and training can be heavy, particularly for smaller companies (Malik et al., 2024; Gholizadeh et al., 2024; Romberg et al., 2012), and global compliance tensions (e.g., Mainland China) can add costs (Amador, 2021). On the revenue side, an ethical/transparency positioning, along with faster iteration, can expand market share (Winarto & Wisesa, 2024; Mendivelso et al., 2024). Accordingly, the light is cast on a sixth research question: *(RQ6) What are the financial implications of this transition in these sectors?*

While animal testing remains entrenched due to historical, regulatory, and biological complexity factors, AI-supported NAMs provide ethically superior, cost-effective, and technologically advanced alternatives that are already replacing animal use in cosmetics and specific pharmaceutical contexts. The future lies in strategic integration, complete replacement where validation and regulation permit, and structured complementarity elsewhere, supported by international harmonization, investment in validation, *safe spaces* for data exchange, open tools, and interdisciplinary training (Fentem et al., 2021; EMA, 2020; Cronin et al., 2025; Ram et al., 2022). Practical enablers include coordinated multi-stakeholder initiatives such as the European Partnership for Alternatives to Animal Testing, which fosters alignment between regulators, researchers, and industry (Fentem et al., 2021), and demonstration projects like the vaccine "consistency approach" that successfully replaced animal batch tests through harmonized protocols and regulatory collaboration (Romberg et al., 2012). Broader access to curated data infrastructures, biomedical labeling tools, and interoperable platforms also facilitates implementation by ensuring reproducibility and regulatory confidence (Liu et al., 2021; Blümmel et al., 2024).

At the same time, several barriers persist. The reproducibility and maturity of OoC and microfluidic systems remain limiting factors, alongside variability in exposure models and the absence of robust datasets for complex endpoints such as developmental and reproductive toxicity (Kimura et al., 2025; Westmoreland et al., 2022). Financial hurdles,

including high upfront investments in infrastructure and training, pose particular challenges for small and medium enterprises (Malik et al., 2024; Gholizadeh et al., 2024).

Moreover, uneven regulatory frameworks, conflicting requirements (e.g., REACH vs. Cosmetics Regulation), and slow policy uptake contribute to compliance uncertainty (Knight et al., 2021; Schmeisser et al., 2023). Cultural and institutional inertia, coupled with hesitancy to abandon established animal-based practices, further slows adoption. However, legal precedents, such as rulings shifting the burden of proof onto animal users, can act as accelerators (Rinaldi et al., 2020). All of that considered, the final and seventh question unfolds: *(RQ7) Which practical enablers and barriers most influence implementation in a real-world setting?*

## Chapter IV – Methodology

This chapter presents the research methodology adopted in this dissertation. It explains the choices that guided the design, data collection, and analysis procedures used to address the central aim: to understand how AI and NAMs can be responsibly implemented to reduce and, where appropriate, replace animal testing in the cosmetic and pharmaceutical industries, considering ethical, regulatory, technological, and financial dimensions.

### 4.1 Research Model

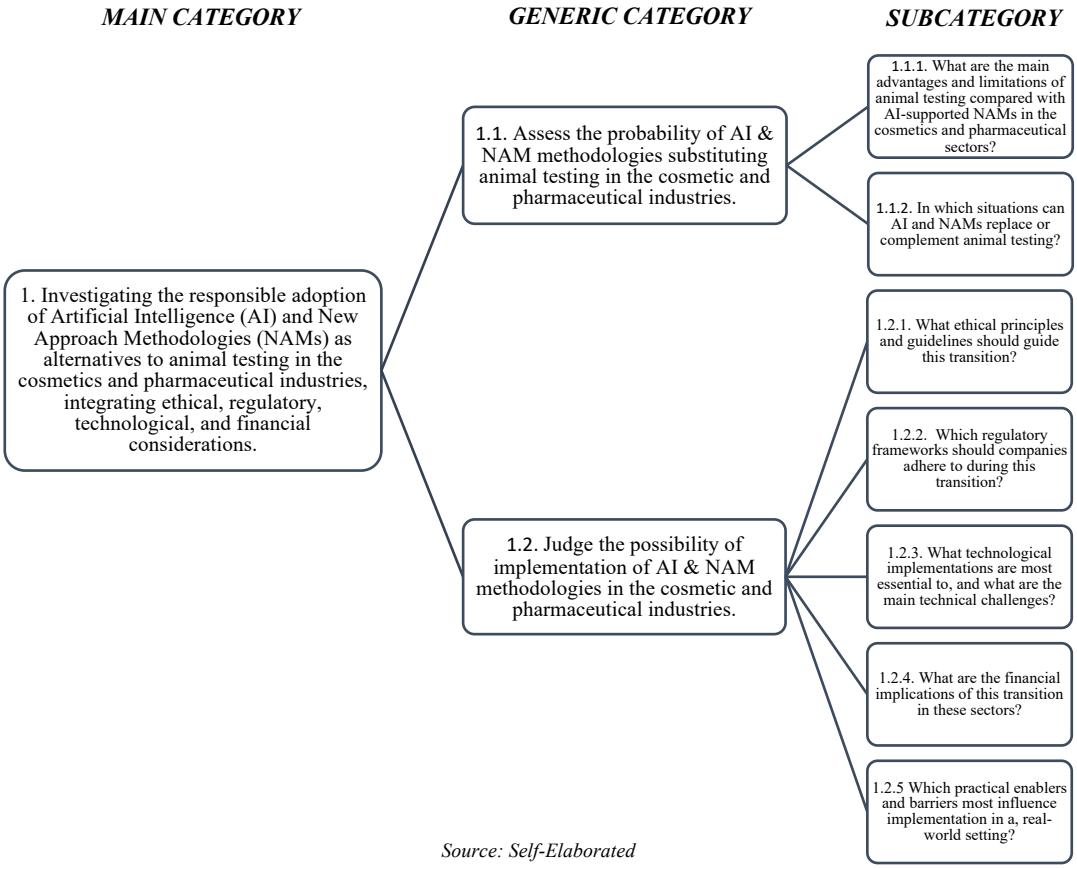
This investigation was conducted in four sequential phases. First, a comprehensive review of existing literature was performed to consolidate current knowledge and identify gaps related to the study's focus.

Second, key theoretical concepts were mapped onto observable realities, aligning the conceptual framework with the examined context. In the third phase, fieldwork involved semi-structured interviews with fifteen multidisciplinary experts, including: a pharmacist and product manager; a medical doctor with a PhD in pharmaceutical biotechnology; a professor of predictive and computational toxicology and chemist; a philosophy teacher specializing in animal ethics and author of *Causeries Animales and De l'or dans la tête*, who also contributes to Marianne magazine (France); a Laboratory Animal Scientist with expertise in animal welfare, experimental design, education, automation, and thermobiology; an economist and former EVP of Finance & Operations and CFO of the BIAL Group, now serving as a non-executive board member; a chartered biologist and European Registered Toxicologist (CBiol, CSci, ERT, FRSB); a biotechnology process manager at Ethicell Solutions LTDA (2025); a professor of evidence-based toxicology at Johns Hopkins and Director of the Center for Alternatives to Animal Testing (CAAT), specializing in NAMs, MPS/OoC, in silico/AI, validation, and regulatory science; an emeritus professor of toxicology and dermato-cosmetology; a Communications Leader at Johns Hopkins CAAT; a veterinary doctor specialized in animal reproduction, laboratory animals, and welfare; a master's graduate in biotechnology and current PhD candidate in translational biomedicine; a veterinary professional at DGAV responsible for the evaluation and authorization of scientific projects involving animals, licensing of

personnel, and inspection of establishments; and the Director of the 3Rs Centre Utrecht, trained as a cellular (neuro)immunologist.

Finally, the data collected from these interviews underwent qualitative analysis to identify patterns, insights, and thematic intersections based on expert experience.

Figure 4.1 - Research Model Diagram




*Source: Self-Elaborated*

## 4.2 Data Collection Method

Data were collected via semi-structured interviews using purposive convenience sampling. Participants were selected based on their relevance to AI/NAMs, non-animal testing, or roles in the pharmaceutical and cosmetics industries, with a focus on those who possess expertise or decision-making authority from academic, scientific, or policy backgrounds. The goal was to gain insights from professionals with operational, strategic, or epistemic knowledge on AI-enabled methods and NAMs challenges and opportunities. Participants represented diverse roles in regulatory science, ethics, industry innovation, and academia.

According to Vilelas (2020), qualitative studies usually require 15 to 20 interviews to reach data saturation. This study conducted fifteen interviews because the specialized focus led to a

quick convergence of ideas. Data collection ended when key themes recurred, indicating saturation. Although the sample size is small, it prioritized depth over generalizability, aligning with the study's objectives. The semi-structured format balances consistency and flexibility, allowing for rich responses without rigid questioning. Interviews aimed to explore patterns and reasoning rather than test hypotheses. Following Bardin's (2018) content analysis model, the analytic process involved three stages: first, pre-analysis, which included organizing the material and familiarizing with the transcripts; second, exploration of the data through coding and categorization; and third, treatment and interpretation of results, focused on thematic synthesis and connection to the research questions.

MAXQDA 24.4 was used to manage transcripts, apply codes systematically, and maintain an audit trail. It supported the deductive structuring of top-level themes based on the four analytical dimensions (Ethics, Regulation, Technology, Finance) and the inductive identification of subthemes from participants' real-world experiences.

### 4.3 Interview Procedure

A critical stage of the research involved planning the interviews and designing the interview guide. The guide was structured to align with the study's overarching objective: to explore expert perspectives on the feasibility, limitations, and strategic implications of adopting AI-enabled NAMs as alternatives to animal testing, particularly in the pharmaceutical and cosmetics sectors.

The interview guide was divided into two main parts. The first contained a single descriptive question intended to characterize the participant's role, sector, and experience. The second section focused on the four analytical dimensions of the study - Ethics, Regulation, Technology, and Finance - through open-ended core questions accompanied by optional probes. These prompts encouraged participants to provide concrete examples, real-world cases, and reflective insights based on their professional expertise.

Participants were contacted via professional networks, LinkedIn, academic referrals, or direct email outreach. Each participant was provided with a brief explanation of the research purpose, which emphasized the investigation of the ethical, financial, regulatory, and technological challenges associated with integrating AI-driven NAMs into existing safety and efficacy frameworks. All interviews were conducted remotely via secure video conferencing platforms (primarily Zoom or Microsoft Teams) or phone calls between

August 27, 2025, and September 19, 2025. Each interview lasted between 10 and 35 minutes, depending on the participant's availability and the depth of the discussion. With informed consent, all sessions were audio-recorded for transcription and analysis.

Confidentiality and data handling protocols were explained at the beginning of each session, and participants were reminded that their responses could be anonymized in all outputs. To ensure a shared understanding of key terms, the interviewer provided a concise contextual explanation of NAMs and their intersection with AI, focusing particularly on relevance to the interviewee's domain of expertise. All fifteen interviews were transcribed verbatim. Where necessary, content in Portuguese was translated into English to ensure consistency during analysis.

### 4.4 Sample Characterization

The sample for this study consisted of fifteen participants from France, Brazil, United States, United Kingdom, Spain, and Portugal, selected based on their professional expertise, practical involvement in NAMs and AI-driven methods, and relevance to the pharmaceutical and/or cosmetics sectors. The selection aimed to capture a range of expert perspectives - scientific, ethical, regulatory, and operational - on the feasibility and implications of transitioning toward AI-enabled non-animal testing approaches.

To characterize the sample, the following parameters were considered: primary area of professional activity, sectoral experience, years of experience, and disciplinary diversity.

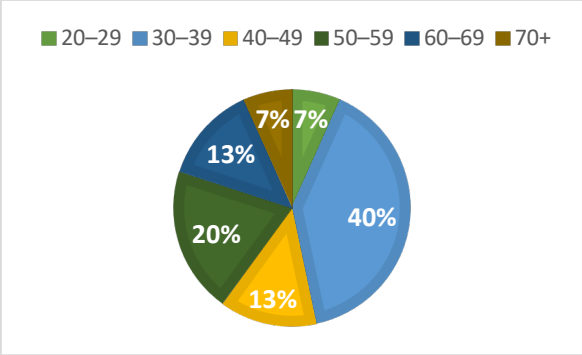
Variables such as gender and age were not used as analytic categories; however, they are included below for descriptive completeness, as all participants consented to being identified in the context of their professional roles.

Regarding Gender and Age, although not used as analytical categories, data on these factors were collected for descriptive completeness and are represented in Figures 4.2 and 4.3.

Participant ages ranged from the late 20s to the early 70s, with the most significant proportion (40%) falling within the 30-39 age range, as shown in Figure 4.2.

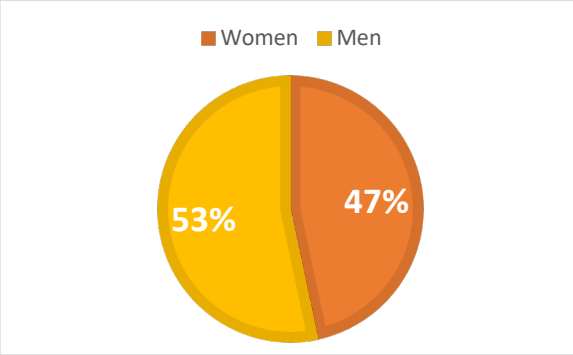
The sample consisted of seven women (46.7%) and eight men (53.3%), reflecting a relatively balanced gender distribution, as shown in Figure 4.3.

*Figure 4.2 - Age Groups of Participants*




*Source: Self-Elaborated*

*Figure 4.3 - Gender of Participants*




*Source: Self-Elaborated*

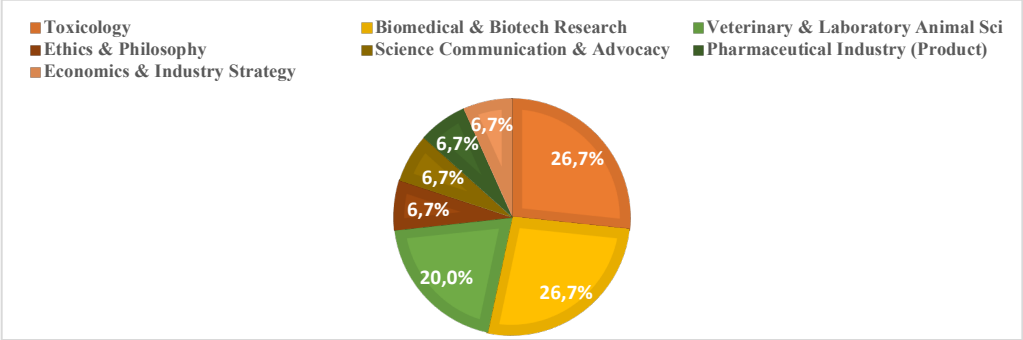
Regarding years of experience, although the precise years varied, the majority of participants held mid-to-senior-level positions with extensive exposure to the scientific, regulatory, or ethical aspects of NAMs and AI. Many had over a decade of experience in their respective fields, with several holding emeritus or leadership roles in academia, research centers, or industry.

### 4.5 Professional Background

Participants represented a diverse but coherent range of professional domains relevant to the development, assessment, regulation, and communication of NAMs and AI. The largest groups (26.7%, 4 out of 15 each) worked primarily in toxicology, covering computational, regulatory, dermato-cosmetology, and evidence-based specializations, as well as in biomedical and biotechnology research, including pharmaceutical biotechnology, bioprocess development, translational biomedicine, and cellular immunology. Veterinary and laboratory animal science accounted for 20.0% of the sample, with expertise spanning animal welfare, reproduction, laboratory animal science, and regulatory oversight of animal use for scientific purposes.

The remaining participants contributed perspectives from ethics and philosophy (6.7%), science communication and advocacy (6.7%), the pharmaceutical industry in product management (6.7%), and economics and strategic operations in the pharmaceutical sector (6.7%). This multidisciplinary composition highlights the complexity of transitioning to animal-free science, which necessitates the integration of scientific innovation, regulatory frameworks, ethical reflection, economic considerations, and effective communication.

*Figure 4.4 - Participants' Area of Expertise*

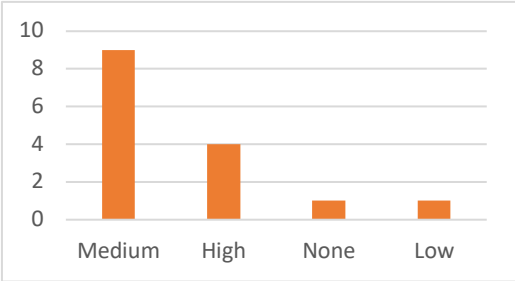



*Source: Self-Elaborated*

When looking at Figure 4.4, participants spanned academia, regulatory science, industry (pharmaceutical and biotechnology), scientific communication, and ethics. This interdisciplinary composition reflects the complex nature of NAMs adoption, which intersects regulatory, scientific, economic, and sociotechnical concerns.

When asked about their familiarity with NAMs and AI methods in the cosmetic and pharmaceutical industries, in Figure 4.5, participants' responses reflected a heterogeneous but generally moderate level of knowledge. Most participants (9 out of 15) indicated a medium familiarity, suggesting that while they are aware of and engage with these methodologies, there is still room for more profound technical expertise or practical application.

*Figure 4.5 - Familiarity with NAMs and AI in Cosmetics and Pharmaceuticals*




*Source: Self-Elaborated*

A smaller but significant proportion (4 out of 15) reported high familiarity, reflecting participants who are either actively involved in the development, application, or regulatory assessment of these methods. Only one participant described their familiarity as low, and another admitted to having no familiarity at all (none). This distribution highlights two crucial insights. First, NAMs and AI are no longer marginal concepts but are increasingly embedded in the professional awareness of experts across academia, regulation, and industry. Second, the predominance of "medium" familiarity points to a transitional phase: stakeholders recognize the relevance and potential of NAMs/AI. However, challenges are still being navigated.

## Chapter V – Results and Discussion

In this chapter, the data collected from the interviews are analyzed and discussed in relation to the literature reviewed in Chapter II. The focus is on the transition from traditional animal testing to AI-supported methods and NAMs in the cosmetics and pharmaceutical sectors. Particular attention is given to the ethical, regulatory, financial, and technological aspects of this transition, as well as to the practical barriers and enablers identified by both experts and the literature. By linking the participants' perspectives with academic and regulatory insights, this chapter aims to critically examine the opportunities and limitations of AI and NAMs, while assessing how these methods can gradually replace or complement animal testing in real-world settings.

Ultimately, the aim is to integrate empirical and theoretical findings to provide a comprehensive understanding of the challenges and pathways that define this paradigm shift, by focusing on the two specific objectives/generic categories of this dissertation: Firstly, 1.1. Assess the probability of AI & NAM methodologies substituting animal testing in the cosmetic and pharmaceutical industries, with two subcategories/research questions (1.1.1; 1.1.2). Secondly, 1.2. Judge the possibility of implementation of AI & NAM methodologies in the cosmetic and pharmaceutical industries, with five subcategories/ research questions (1.2.1; 1.2.2; 1.2.3; 1.2.4; 1.2.5).

### 5.1 Assess the probability of AI and NAM methodologies substituting animal testing in the cosmetic and pharmaceutical industries.

For this investigation, the first generic category, "1.1 Assess the probability of AI and NAM methodologies substituting animal testing in the cosmetic and pharmaceutical industries," encompasses two subcategories. The first subcategory (1.1.1) focused on the main advantages and limitations of animal testing compared to AI-supported NAMs in the cosmetics and pharmaceutical sectors. This involved ethical, regulatory, financial, and technological implications.

*Table 5.1 - Main advantages and limitations of animal testing compared with AI-supported methods and NAMs in the cosmetics and pharmaceutical sectors.*

| TEXT | GENERIC CATEGORY | SUBCATEGORY | TIMES MENTIONED | PARTICIPANTS |
|---|---|---|---|---|
| Ethical advantage of NAMs, namely the reduction or elimination of animal suffering. | 1.1 | 1.1.1 | 4 | P2, P4, P11, P14 |
| Poor translation of animal data to clinical outcomes | 1.1 | 1.1.1 | 6 | P2, P3, P4, P10, P13, P14 |
| Speed, scalability, and capacity to integrate large datasets across multiple modalities | 1.1 | 1.1.1 | 5 | P2, P5, P8, P10, P13 |
| NAMs still lack the systemic integration that in vivo models provide | 1.1 | 1.1.1 | 6 | P2, P3, P8, P13, P14, P15 |
| Validation and regulatory acceptance, data quality and bias in AI models, and high initial costs | 1.1 | 1.1.1 | 8 | P2, P3, P14, P5, P9, P10, P13, P7 |
| Legacy strengths of animal testing, such as familiarity, systemic exposure, and extensive historical datasets | 1.1 | 1.1.1 | 3 | P6, P12, P14 |

*Source: Self-Elaborated*

In Table 5.1, it is possible to identify the perspectives of the fifteen participants, highlighting a complex but converging view: while animal testing has historically provided a framework for safety evaluation, its ethical, translational, and financial limitations are increasingly evident, and AI-supported NAMs are progressively recognized as superior alternatives in many contexts, 4 out of 15 participants emphasized the ethical advantage of NAMs, namely the reduction or elimination of animal suffering. This aligns with Silva and Tamburic (2022), Neuhaus et al. (2022) and Fentem et al. (2021), who note that the EU regulatory framework and the principle of the 3Rs make replacement not only desirable but legally mandated wherever possible.

One participant explicitly linked AI-supported NAMs to advancing science “with a more ethical approach, perfectly aligned with the 3Rs,” echoing Cronin et al. (2025), who argue for international harmonization to accelerate non-animal testing. Consumer-oriented perspectives (Mendivelso et al., 2024; Winarto & Wisesa, 2024; Marshall et al., 2022) further reinforce this view, demonstrating that markets, NGOs, and the public are increasingly demanding ethical testing methods.

From a human relevance perspective, 6 out of 15 participants stressed the poor translation of animal data to clinical outcomes. As two participants underlined, animal models often lack predictive power, resulting in high attrition rates in the development of new drugs. This is

consistent with the FDA's (2025) reports, which indicate that over 90% of drugs that pass animal tests fail in humans. Movia and Prina-Mello (2020) and Van Berlo et al. (2021) similarly point to NAMs as more human-relevant, with OoC and in silico models offering enhanced predictive validity. Grech et al. (2024), Malik et al. (2024), and Pastorino et al. (2024, 2025) highlight that these methods not only improve accuracy but also accelerate early-stage candidate screening, while L'Oréal's industrial work (Belén, 2020a, as cited in Mendivelso et al., 2024) demonstrates their feasibility at scale.

The technological advantages of AI-supported NAMs were widely recognized. One-third of the participants highlighted the speed, scalability, and capacity to integrate large data sets of multiple alternative modalities. This resonates with Hartung (2023), Grech et al. (2024), and Kleinstreuer and Hartung (2024), who emphasize AI's ability to integrate big data and AI/ML, which can surpass classic QSAR limits by fusing diverse evidence streams. While 7% of the participants emphasized the ability of AI to interpolate results across methodologies, another 7% highlighted the acceleration of information processing: "It has the advantage of accelerating information processing and validating tests and procedures". The literature supports this, with QSARs, biosimulations, and DL toxicology models (Gangwal & Lavecchia, 2025; Fontana et al., 2021; Taylor & Alvarez, 2020; Bhattamisra et al., 2023) consistently cited as both cost-saving and more predictive.

However, participants also raised limitations of NAMs and AI. A recurring theme amongst 40% of the participants was that NAMs still lack the systemic integration that in vivo models provide. As one participant put it, "none of the current NAMs can mimic the interplay between immune, neurological, or endocrine systems," reflecting Van Berlo et al. (2021) and Kimura et al. (2025), who note that OoC technologies, though promising, remain limited in scale, reproducibility, and complexity. Similarly, 7% emphasized that critical areas such as repeated dose toxicity and reproductive toxicology lack validated non-animal equivalents, echoing Westmoreland et al. (2022). As one participant put it, "I don't see AI being able to replace animal in vivo tests in the short to medium term, but it is already an important part of the discovery program, and this will increase".

Other limitations mentioned by 8 out of the 15 participants (53%) included validation and regulatory acceptance, data quality, and bias in AI models, and high initial costs: "Limitations include a lack of AI prepared for scientific functions, bias, and hallucinations". These findings align with those of Romberg et al. (2012), who describe the slow pace of validation and regulatory endorsement, and with Patel and Patel (2024), who

caution against “black box” AI models lacking explainability. Hartung (2023) similarly stresses that a mechanistic understanding is needed for regulatory trust.

Interestingly, some participants (3/15) acknowledged the legacy strengths of animal testing, such as familiarity, systemic exposure, and extensive historical datasets: “Animal testing has a long legacy of regulatory familiarity and can capture whole-organism systemic exposure… also, obviously has extensive historical datasets”. This aligns with Knight et al. (2021) and Amador (2021), who explain why regulatory conservatism and legal conflicts (e.g., REACH vs. Cosmetics Directive) perpetuate the entrenchment of animal testing in certain cases. Hunter (2023) adds nuance by rejecting the idea that animal testing is a “90% failure,” stressing instead that the 10% of drugs that succeed are lifesaving, while Kabene and Baadel (2019) argue that animal testing retains defensibility in medicine but not in cosmetics.

In summary, the participants’ responses reinforce the consensus in the literature: AI-supported NAMs represent the future of toxicology and preclinical testing due to their ethical legitimacy, human relevance, speed, and scalability (Mendivelso et al., 2024, Winarto & Wisesa, 2024, Silva & Tamburic, 2022, Kimura et al., 2025, Kabene & Baadel, 2019, Movia & Prina-Mello, 2020, Pastorino et al., 2024; Alnasser et al., 2025 Romberg et al., 2012; Singer, 1975 as cited in Mendivelso et al., 2024; Malik et al., 2024; Yamazaki and Ishida, 2025; Gangwal & Lavecchia, 2025; Shiammala et al., 2023; Schmeisser et al., 2023; Pistollato et al., 2025; Samah H. O. Zarroug, 2025; Fontana et al., 2021; Taylor & Alvarez, 2020; Bhattamisra et al., 2023).

However, their adoption is constrained by validation gaps, systemic complexity, regulatory inertia, and data challenges. As both experts and authors argue, the path forward is not immediate replacement but structured complementarity: deploying NAMs where validated, while gradually reducing reliance on animal models until systemic barriers are overcome (Kimura et al., 2025; Westmoreland et al., 2022; Alves et al., 2021; Neuhaus et al., 2022; Knight et al., 2021; Kapustina et al., 2024).

The second subcategory (1.1.2) focused on identifying cases where AI could replace or complement animal testing in the cosmetics and pharmaceutical sectors. The responses from the participants illustrate both optimism and caution regarding the replacement or complementarity of animal testing with AI-supported NAMs. Their perspectives align closely with the literature, which highlights validated niches of replacement, promising areas of complementarity, and persistent gaps that necessitate a stepwise transition.

*Table 5.2 - Cases where AI could replace or complement animal testing in the cosmetics and pharmaceutical sectors.*

| TEXT | GENERIC CATEGORY | SUBCATEGORY | TIMES MENTIONED | PARTICIPANTS |
|---|---|---|---|---|
| AI's potential to substitute animal testing in early-stage ingredient screening and toxicity prediction, particularly relevant for cosmetics | 1.1 | 1.1.2 | 2 | P2, P10 |
| Validated use cases such as skin sensitization hazard classification | 1.1 | 1.1.2 | 1 | P14 |
| NAMs cannot yet mimic systemic complexities such as immune or endocrine interplay | 1.1 | 1.1.2 | 2 | P13, P15 |
| AI models must undergo qualification and validation before regulators will accept them as standalone tests | 1.1 | 1.1.2 | 1 | P3 |
| AI's ability to guide and target animal studies, improve pharmacokinetic and tolerance predictions, and accelerate results | 1.1 | 1.1.2 | 3 | P7, P8, P11 |
| Robust substitution is possible when AI is integrated with in vitro, in silico and in chemico methods, forming multi-layered testing batteries. | 1.1 | 1.1.2 | 2 | P5, P12 |
| Issue of data sharing | 1.1 | 1.1.2 | 1 | P1 |
| While animal models provide systemic data, their physiology is not human-equivalent, so AI may ultimately offer more reliable predictions | 1.1 | 1.1.2 | 1 | P4 |

*Source: Self-Elaborated*

Looking at Table 5.2, a high percentage of participants, 40%, emphasized that replacement is already feasible in certain contexts. For example, 2 out of the 15 highlighted AI's potential to replace animal testing in early-stage ingredient screening and toxicity prediction, particularly relevant for cosmetics, where EU law prohibits the use of animals in testing. One provided the most detailed account, identifying existing, validated use cases: "Today, in practice, NAMs can replace skin sensitization hazard classification using Defined Approaches combining in chemico, for example DPRA, in vitro, like KeratinoSens, h-CLAT and in silico under OECD Test Guideline (OECD TG) 497.

In the pharmaceutical industry, genotoxic impurity assessment is conducted via ICH M7(R2) with two complementary QSAR models to classify mutagenicity and guide

controls. Oh, and the eye and skin irritation/corrosion using validated in vitro test methods, like the OECD TG 492/RhCE and the TG 431/439". The literature strongly corroborates and adds to these perspectives: Van Berlo et al. (2021), Alépée et al. (2019), Eskes et al. (2007), and Strickland et al. (2022) all confirm that NAMs have regulatory traction in these domains, with accuracy rates matching or surpassing animal assays. Similarly, Romberg et al. (2012) demonstrated in vaccines that consistency approaches based on in vitro methods can replace animal batch release tests, validating gradual but concrete substitution.

Yet, both participants and authors acknowledged that full replacement is not yet possible across all toxicological endpoints. As one participant added to Alves et al. (2021) from the literature review, "It can complement short to mid-term in some areas, like the prioritization and waiving strategies under REACH, (read-across, weight-of-evidence) using AI-enhanced similarity or read-across frameworks (RAAF).

Additionally, quantitative risk assessment is conducted by linking NAM effect concentrations to human exposure using PBK/IVIVE, as outlined in OECD guidance document No. 331. In pharmaceuticals, it can complement carcinogenicity; the ICH S1B(R1) weight-of-evidence approach can sometimes replace a traditional two-year rat study when scientifically justified. And, of course, there's the use of MPS, such as OoC models, which can elicit mechanism-specific human responses. These are already advancing towards defined regulatory contexts of use through pathways like the FDA ISTAND program".

13% noted that NAMs cannot yet mimic systemic complexities such as immune or endocrine interplay, a limitation also highlighted by Kimura et al. (2025) and Westmoreland et al. (2022), who warn that domains like DART, immunotoxicity, and respiratory sensitization remain unresolved. One emphasized that AI models must undergo qualification and validation before regulators will accept them as standalone tests, a concern echoed by Romberg et al. (2012) and Patel and Patel (2024), who stress that reproducibility and explainability are prerequisites for regulatory trust. Likewise, Mendivelso et al. (2024) and Knight et al. (2021) emphasize that while replacement is ethically urgent, regulators still mandate in vivo testing in certain contexts, particularly in the pharmaceutical industry.

Where participants were most unified was in identifying AI's capacity to complement animal testing and enhance efficiency. 20% described AI's ability to guide and target animal studies, improve PK and tolerance predictions, and accelerate results - roles that parallel Fentem et al. (2021) and Madden et al. (2020), who envision AI-enhanced NGRA frameworks. Another 13% stressed that robust substitution is possible when AI is integrated with in vitro and in chemico methods, forming multi-layered testing batteries. This aligns with Madden et al.

(2020), Silva and Tamburic (2022), and Kleinstreuer and Hartung (2024), who advocate for tiered approaches where batteries of complementary NAMs collectively approximate systemic insight.

The participants also pointed to situations where complementarity is already delivering value. One raised the issue of data sharing, noting that AI could replace duplicative studies if negative results were openly available: a barrier explicitly flagged by Cronin et al. (2025), who call for *safe spaces* for toxicology data exchange. Another argued that while animal models provide systemic data, their physiology is not human-equivalent, so AI may ultimately offer more reliable predictions, a position consistent with Van Norman (2019) and Movia and Prina-Mello (2020), who stress human relevance as a decisive strength of NAMs.

In synthesis, both participants and the literature converge on a tiered, endpoint-specific transition strategy. Replacement is already validated in certain domains (e.g., skin, eye, sensitization, genotoxicity). Complementarity is effective in early-stage discovery, prioritization, and PK, but persistent gaps remain for systemic and long-term toxicities. The literature emphasizes, and participants confirm, that AI-supported NAMs are not yet an absolute substitute. Still, when deployed in batteries, integrated with mechanistic NAMs, and supported by robust data infrastructures, they can meaningfully reduce and refine animal use while progressively building regulatory and scientific confidence.

## 5.2 Judge the possibility of implementation of AI and NAM methodologies in the cosmetic and pharmaceutical industries

For this investigation, the second generic category, "1.2 Judge the possibility of implementation of AI and NAM methodologies in the cosmetic and pharmaceutical industries," encompasses five subcategories. The first subcategory (1.2.1) of the second generic category focused on each participant's knowledge of the ethical principles and guidelines that should guide the transition from animal testing to alternative methods.

*Table 5.3 - Ethical principles and guidelines that should guide the transition from animal testing to alternative methods.*

| TEXT | GENERIC CATEGORY | SUBCATEGORY | TIMES MENTIONED | PARTICIPANTS |
|---|---|---|---|---|
| 3Rs principle | 1.2 | 1.2.1 | 5 | P2, P6, P10, P14, P15 |

| | | | | |
|---|---|---|---|---|
| Transparency | 1.2 | 1.2.1 | 10 | P1, P2, P4, P5, P6, P7, P10, P11, P14, P15 |
| Accountability | 1.2 | 1.2.1 | 1 | P6 |
| Adherence to international standards | 1.2 | 1.2.1 | 1 | P14 |
| Ethical debates are often unproductive | 1.2 | 1.2.1 | 1 | P3 |
| The ethical goal is not simply to replace animal tests, but to generate scientifically valid data ensuring human safety | 1.2 | 1.2.1 | 2 | P12, P13 |

*Source: Self-Elaborated*

When looking at Table 5.3, the responses converge strongly around a set of recurring principles mentioned throughout the literature review, by Neuhaus et al. (2022), EMA (2016), Grech et al. (2024), Rinaldi et al. (2020), Connor et al. (2022), EMA (2020), Cronin et al. (2025) and Silva and Tamburic (2022), the 3Rs, transparency, accountability, and the need for an ethically defensible scientific framework to guide the transition away from animal testing. Participants linked these directly to societal expectations, regulatory frameworks, and the evolving responsibilities of researchers and industry actors.

So, 5 out of the 15 participants explicitly invoked the 3Rs principle, enshrined in Directive 2010/63/EU, as the ethical baseline. 7% stressed that the principle remains "the skeleton of animal protection" in Europe, ensuring that any project involving animals is authorized only if it adheres to 3Rs standards. This mirrors Silva and Tamburic (2022) and Fentem et al. (2021), who highlight that the 3Rs underpin EU law and remain central to regulatory decisions. The emphasis on refinement and reduction echoes Mendivelso et al. (2024) foundational critique of animal suffering and subsequent calls to minimize harm wherever possible. Pistollato et al. (2025) and Espinosa and Castañeda (2020, cited in Mendivelso et al., 2024) further underscore the moral obligation to replace animal models whenever viable human-relevant alternatives exist.

Transparency was repeatedly emphasized. 67% participants argued for openness not only in animal testing but also in AI-supported NAMs, where the training data, algorithms, and decision processes must be explainable, as one participant noted, "the transition from animal testing to AI-supported NAMs should be guided by clear ethical principles that ensure the protection of animals, people, and the environment while preserving scientific integrity. The 3Rs remain the ethical cornerstone, as enshrined in Directive 2010/63/EU, which requires all approved animal research in the EU to comply with these principles. Likewise, transparency must also apply to AI methods, such as training data, algorithms, and decision-making

processes, which should be disclosed and open to evaluation, fostering trust in AI-enabled NAMs".

Another participant linked transparency to public education, suggesting that open, visible practices could build trust in biomedical transitions and even influence perceptions of AI adoption beyond this field: "Adoption must be consistent and incremental. The entire biomedical enterprise is still very much structured around animal models, so change has to be at the institutional level". This view parallels Marshall et al. (2022), who document strong public support for phasing out animal testing, and Winarto and Wisesa (2024), who argue that consumer trust in cruelty-free branding hinges on transparency. Hash et al. (2025) note that transparency in AI models is essential not only for maintaining consumer confidence but also for addressing issues of bias and ensuring explainability.

Accountability also emerged as a central theme. 7% emphasized that responsibility ultimately rests with principal investigators, who should be held accountable for research paths and decisions involving animal use. This resonates with Cronin et al. (2025), who propose governance frameworks and incentives to align research practices with ethical commitments. Similarly, another 7% emphasized adherence to international standards (e.g., GIVIMP, GCCP 2.0) and systematic review practices to ensure reproducibility and minimize bias. These align with Hartung (2023) and Kleinstreuer and Hartung (2024), who argue that transparent, evidence-based practices are crucial for the ethical integration of AI and NAM.

A minority of participants highlighted different angles. One suggested that ethical debates are often unproductive, proposing that progress is best made by focusing on scientific validity rather than moral arguments. This pragmatic stance mirrors Romberg et al. (2012), who emphasize scientific consistency approaches (e.g., vaccine testing) as drivers of replacement. It also reflects Hunter (2023), who pushes back against the "90% failure" narrative of animal testing, reminding that despite high attrition rates, animal models have underpinned lifesaving breakthroughs - an argument for grounding the debate in evidence rather than passion.

Likewise, Kabene and Baadel (2019) accept that animal testing is ethically indefensible for cosmetics but defend limited use in medicine when human benefits are substantial, underscoring the ethical trade-off in different sectors.

Ultimately, participants stressed that ethical principles cannot be detached from societal expectations. 13% tied transparency and innovation to maintaining public trust, while another 13% argued for clarity of purpose: "the ethical goal is not simply to replace

animal tests, but to generate scientifically valid data ensuring human safety". This resonates with Mendivelso et al. (2024), who demonstrate that consumer demand for cruelty-free and ethical products is driving industry transitions, and with L'Oréal (Belén, 2020a, as cited in Mendivelso et al., 2024), which shows how the transparent adoption of NAMs can align corporate ethics with social legitimacy.

In synthesis, participants and literature align on a common set of guiding principles: 3Rs remain the ethical baseline, legally codified and socially reinforced; transparency and accountability are essential for both public trust and scientific reproducibility; scientific rigor must complement ethical imperatives, ensuring NAMs are validated and reliable; ethical arguments are context-dependent, with cosmetics facing stronger imperatives for replacement than pharmaceuticals, where human benefit may justify trade-offs. Together, these principles suggest that the ethical framework for transition is not static but evolving it must balance compassion for animals, scientific credibility, societal expectations, and sector-specific realities.

The second subcategory (1.2.2) of this generic category focused on each participant's knowledge of which regulatory frameworks companies should adhere to during the transition from animal testing to alternative methods.

*Table 5.4 - Regulatory frameworks companies should adhere to during the transition from animal testing to alternative methods.*

| TEXT | GENERIC CATEGORY | SUBCATEGORY | TIMES MENTIONED | PARTICIPANTS |
|---|---|---|---|---|
| REACH | 1.2 | 1.2.2 | 9 | P1, P2, P4, P5, P10, P11, P12, P14, P15 |
| EU Cosmetics Regulation (EC No. 1223/2009) and the validation procedures of EURL ECVAM | 1.2 | 1.2.2 | 1 | P2 |
| OECD Test Guidelines (TGs) and related standards (e.g., PBK Guidance No. 331, GIVIMP No. 286) | 1.2 | 1.2.2 | 4 | P2, P7, P10, P14 |
| FDA | 1.2 | 1.2.2 | 7 | P2, P5, P6, P8, P9, P11, P14 |
| MAPA, CONCEA | 1.2 | 1.2.2 | 2 | P7, P10 |
| MoCRA (2022) for U.S. cosmetics | 1.2 | 1.2.2 | 1 | P14 |
| Anvisa, Brazil | 1.2 | 1.2.2 | 1 | P5 |
| Global co-ordination and between sectors, streamlining the EU Green Deal's "One Substance, One Assessment". | 1.2 | 1.2.2 | 1 | P12 |
| Regulation 1223/2009 under DG GROW and SCCS oversight for cosmetics | 1.2 | 1.2.2 | 2 | P14, P15 |

| | | | | |
|---|---|---|---|---|
| (ICH M7(R2), ICH S1B(R1)) for pharmaceuticals | 1.2 | 1.2.2 | 2 | P14, P15 |
| International law harmonization | 1.2 | 1.2.2 | 2 | P3, P12 |

*Source: Self-Elaborated*

Looking at Table 5.4, it is evident that the participants' responses to this question reveal a strong awareness of the fragmented yet evolving regulatory environment governing the transition from animal testing to AI-supported NAMs. While their emphasis varies depending on geographical and sectoral contexts, a consistent theme emerges: adherence to frameworks such as REACH, EMA, FDA, and OECD TG is viewed as essential, though participants also stress the need for harmonization, validation, and coordination across jurisdictions.

13% of the participants highlighted REACH as a cornerstone of European regulation. This echoes Knight et al. (2021), who argue that REACH, while initially intended to reduce animal use, still creates contradictions with the EU Cosmetics Regulation by requiring animal data in some cases. Similarly, Fentem et al. (2021) emphasize that REACH needs to be progressively aligned with NGRA to resolve these conflicts.

One participant provided a remarkably detailed overview, citing not only REACH and EMA guidance, but also the EU Cosmetics Regulation, ECVAM validation procedures, FDA guidelines, and OECD TG. This reflects the comprehensive perspective of EMA (1997, 2016, 2020) and OECD initiatives, which promote mutual acceptance of data, as well as Pistollato et al. (2025), who call for coordinated international frameworks to accelerate NAM adoption. 20 % pointed to frameworks outside of Europe, including ANVISA, MAPA, and CONCEA in Brazil, as well as U.S. regulators such as the FDA. This mirrors the analysis of Amador (2021) and Akkermans et al. (2020), who emphasize that global inconsistencies, particularly China's continued mandate of animal testing for cosmetics, remain a barrier to complete harmonization.

Some participants, 27%, placed particular importance on EMA and FDA. Their responses reflect the leading role of these agencies in setting international standards, primarily through initiatives such as the FDA Modernization Act 2.0 (2022), which permits non-animal methods in regulatory submissions, and the EMA (2020) Reflection Paper, which explicitly supports NGRA approaches. Connor et al. (2022), Cronin et al. (2025), and Yamazaki and Ishida (2025) similarly emphasize that the FDA and EMA play a pivotal role in legitimizing AI and NAMs for risk assessment.

Three participants provided a more expansive view. One argued for adherence to all frameworks but underscored the inefficiency of current regulatory fragmentation, pointing to the EU's "One Substance, One Assessment" initiative as a way forward. This aligns directly with Fentem et al. (2021) and Westmoreland et al. (2022), who argue that streamlined and unified assessments are crucial for reducing duplication and enhancing efficiency.

The other two offered the most comprehensive regulatory mapping, citing frameworks spanning cosmetics (EU Regulation 1223/2009, MoCRA in the U.S.), chemicals (REACH, OECD TG), and pharmaceuticals, as one added: "the International Council for Harmonization (ICH) has established key guidance, including M7(R2) on QSAR for mutagenic impurities and S1B(R1) enabling a weight-of-evidence carcinogenicity approach, both adopted by the EMA and the FDA. For drugs, the FDA Modernization Act 2.0 (2022) authorizes nonclinical tests, such as in vitro, in silico, and MPS, for Investigational New Drug submissions and approvals. In parallel, the FDA ISTAND program offers a pathway to qualify novel tools, including MPS and AI-enabled methods, for defined regulatory uses".

Their perspective closely mirrors the literature consensus: while validated NAMs already exist for endpoints such as skin sensitization, irritation, and genotoxicity (Van Berlo et al., 2021; Alépée et al., 2019; Strickland et al., 2022), broader regulatory adoption requires integration of tools such as PBK/QIVIVE and consistent global standards (Cronin et al., 2025; Pistollato et al., 2025; Westmoreland et al., 2022).

Finally, one participant specifically pointed to the EU Cosmetics Regulation 1223/2009, noting: "In Europe, cosmetics are regulated under Regulation 1223/2009, which is currently being reviewed for simplification and is overseen by DG GROW. In the coming years, responsibility for the Scientific Committee on Consumer Safety (SCCS) will shift to ECHA, although the SCCS will remain an independent committee". This is consistent with Silva and Tamburic (2022) and Knight et al. (2021), who highlight that the cosmetics sector is unique in having an outright ban on animal testing, but still struggles with regulatory tensions, particularly under REACH.

Taken together, the responses indicate that while participants name different regulators, they converge on the same structural issues emphasized in the literature: fragmentation, slow validation, regulatory contradictions, and the urgent need for international harmonization (Cronin et al., 2025; Schmeisser et al., 2023; Wood et al., 2025). As such, companies should not only comply with sector- and region-specific frameworks but also anticipate the global trajectory toward NGRA, harmonized OECD standards, and flexible qualification pathways (e.g., FDA ISTAND, EMA pilots). The evidence suggests that success in this transition will

depend on combining regulatory adherence with proactive engagement in validation, harmonization, and transparent reporting.

The third subcategory (1.2.3) focused on each participant's knowledge of the technological implementations that are most essential to adopt and what the main technological challenges are in transitioning from animal testing to alternative methods.

*Table 5.5 - Technological implementations that are most essential to adopt, and what the main technological challenges are in transitioning from animal testing to alternative methods.*

| TEXT | GENERIC CATEGORY | SUBCATEGORY | TIMES MENTIONED | PARTICIPANTS |
|---|---|---|---|---|
| Data and infrastructure | 1.2 | 1.2.3 | 5 | P1, P5, P9, P10, P13 |
| Validation and regulatory integration | 1.2 | 1.2.3 | 5 | P2, P3, P4, P12, P14 |
| Innovation in methods and systems approaches | 1.2 | 1.2.3 | 5 | P6, P7, P8, P11, P15 |
| Transcriptomics, 3D cultures, HTS, AI, and PBPK modeling | 1.2 | 1.2.3 | 1 | P15 |
| FAIR data assets, robust QSARs, mechanistic NAMs (TG 497, PBK/IVIVE, MPS), and adherence to quality frameworks like GIVIMP and GCCP 2.0 | 1.2 | 1.2.3 | 1 | P14 |
| Batch effects, model drift, traceability, and integration into risk assessments | 1.2 | 1.2.3 | 1 | P14 |
| Linking validation to economic benefits | 1.2 | 1.2.3 | 1 | P4 |

*Source: Self-Elaborated*

Examining Table 5.5, the responses indicate a strong consensus among participants that technological progress in AI-supported NAMs depends not only on innovation but also on infrastructure, validation, and data governance. Their answers mirror recurring themes in the literature, showing that while the scientific toolbox is already diverse, systemic and regulatory barriers continue to slow its widespread adoption.

Considering that, 5 out of the 15 participants placed data infrastructures and quality at the center of the debate. They stressed that without curated, interoperable datasets, AI models cannot be trained to generate reliable predictions. This aligns with Neuhaus et al. (2022), who argue that AI-driven NAMs require FAIR (Findable, Accessible, Interoperable, and Reusable) datasets, harmonized ontologies, and systematic benchmarking to address concerns about reproducibility.

One of the participants' references to France's relatively late launch of a 3R center and uncertainty about the technology present in France further illustrates Schmeisser et al. (2023)'s point that uneven national capacities undermine EU-wide progress. The emphasis

on feeding AI with high-quality data from 7% of the participants resonates with Patel and Patel (2024) and Niazi (2023), who warn that "black-box" opacity and bias threaten regulatory trust when training datasets are incomplete or flawed. 20% identified validation as the most pressing challenge, aligning with Van Berlo et al. (2021) and Alépée et al. (2019), who note that while NAMs already show strong performance in specific domains (skin sensitization, genotoxicity), regulatory endorsement requires costly and cumbersome validation.

One pointed directly to the European Commission's roadmap (Cronin et al., 2025), emphasizing that validation today is inefficient, slow, and underfunded, while another highlighted the importance of linking validation to economic benefits, a position reinforced by Fontana et al. (2021), who argue that validated NAMs lower long-term R&D costs. Romberg et al. (2012) demonstrated this pathway in vaccine testing, where gradual substitution through validated consistency tests led to global regulatory acceptance.

Additionally, 7% offered a distinct perspective, claiming that funding, not technology, is the real bottleneck. This observation aligns with Schmeisser et al. (2023) and Wood et al. (2025), who note that while new methods proliferate, validation frameworks often lack sufficient financial backing. Malik et al. (2024) similarly warn that without sustained funding for validation and training, even promising NAMs risk remaining in research silos rather than becoming regulatory defaults.

Three participants broadened the debate, pointing to methodological diversity, infrastructure, and cost. One emphasized that not all progress comes from making models more complex, but also from tiered testing strategies, breaking down difficult toxicological questions into manageable steps. This aligns with Madden et al. (2020) and Pai et al. (2024), who advocate for combining in vitro, in silico, and AI-driven approaches into a battery of methods for weight-of-evidence decision-making. At the same time, these participants echoed Westmoreland et al. (2022) and Kimura et al. (2025), who warn that NAMs remain underdeveloped for systemic endpoints such as DART, respiratory sensitization, and immunotoxicity, areas where animal models remain entrenched.

Only 7% observed that AI is already widely available in drug discovery and R&D, highlighting its role in accelerating early-phase research. This aligns with Gangwal and Lavecchia (2025) and Alnasser et al. (2025), who describe how ML and generative models accelerate candidate screening, optimize compound selection, and reduce reliance on late-stage animal testing.

One provided the most detailed and technically grounded response, calling for FAIR data assets, robust QSARs, mechanistic NAMs (TG 497, PBK/IVIVE, MPS), and adherence to

quality frameworks such as GIVIMP and GCCP 2.0. They also highlighted challenges such as batch effects, model drift, traceability, and integrating these into risk assessments. These concerns align with Hartung (2023), Neuhaus et al. (2022), and Kleinstreuer and Hartung (2024), who emphasize the importance of explainability and reproducibility in tracking AI applications across various regulatory contexts. Their call for rigorous frameworks reflects the emerging push toward NGRA (EMA, 2020; Pistollato et al., 2025).

Another 7% pointed to advanced molecular and systems technologies, including transcriptomics, 3D cultures, HTS, AI, and PBPK modeling. These approaches are precisely those described by Madden et al. (2020), Kimura et al. (2025), Gangwal and Lavecchia (2025), US FDA (2025), Westmoreland et al. (2022), Elder et al. (2024), Kabene and Baadel (2019) and Silva et al. (2023), who emphasize their ability to create personalized synthetic trials and capture human-specific biology. L'Oréal's Episkin platform (Belén, 2020a, as cited in Mendivelso et al., 2024) further demonstrates that these technologies are not only theoretical but already scalable in industrial practice.

In essence, participants and authors alike agree that the science is advancing rapidly, however, adoption hinges on addressing structural barriers, including funding, validation, harmonization, and trust in AI transparency. Until these issues are addressed, NAMs and AI will continue to complement, rather than fully replace, animal models in complex systemic endpoints.

The fourth subcategory (1.2.4) centered on understanding each participant's knowledge of the financial implications of this transition in these sectors. The financial dimension of transitioning from animal testing to AI-supported NAMs emerged as a recurring concern among participants, with perspectives oscillating between acknowledging high upfront costs and expecting significant long-term savings. The literature strongly corroborates this duality, framing the transition as a high-investment, high-return paradigm.

*Table 5.6 - Financial implications of the transition from animal testing to alternative methods in cosmetics and pharmaceuticals.*

| TEXT | GENERIC CATEGORY | SUBCATEGORY | TIMES MENTIONED | PARTICIPANTS |
|---|---|---|---|---|
| Substantial initial investments | 1.2 | 1.2.4 | 5 | P2, P4, P7, P12, P14 |
| Economic benefits of new methods hinge on their predictive validity | 1.2 | 1.2.4 | 4 | P3, P6, P9, P13 |
| Cost of infrastructure and data requirements | 1.2 | 1.2.4 | 4 | P5, P10, P11, P13 |
| Public and private funding | 1.2 | 1.2.4 | 2 | P10, P11 |
| Long-term return on investment (ROI) | 1.2 | 1.2.4 | 3 | P4, P8, P14 |

| | | | | |
|---|---|---|---|---|
| Uncertainty or skepticism | 1.2 | 1.2.4 | 2 | P1, P15 |

*Source: Self-Elaborated*

Table 5.6 shows that 33% of the participants emphasized the substantial initial investments required to implement NAMs, including the purchase of AI tools, laboratory infrastructure, and training of specialized staff. One participant, for instance, noted that although upfront costs are "high", the long-term benefits justify the investment through reduced reliance on animals and diminished housing and care expenses. This mirrors Taylor and Alvarez (2020) and Romberg et al. (2012), who demonstrated that while NAM validation is resource-intensive, the break-even point can be reached in as little as four years. Similarly, Fontana et al. (2021) and Malik et al. (2024) stress that NAMs ultimately reduce R&D costs and increase efficiency across portfolios, making the financial case favorable over time.

Furthermore, 4 of the 15 participants expressed cautionary perspectives, emphasizing that the economic benefits of new methods hinge on their predictive validity. As one participant highlighted, if technologies "prove to be non-predictive," not only will investment be wasted, but broader trust in new technologies may be eroded. This is consistent with Westmoreland et al. (2022) and Kimura et al. (2025), who warn that gaps in developmental and reproductive toxicology may delay widespread regulatory acceptance, keeping financial risks high. Lee et al. (2022) further argue that productivity dips often accompany early AI adoption due to adjustment costs, suggesting that short-term financial instability should be expected.

Participants also stressed the cost of infrastructure and data requirements (27%). As one observed, AI adoption itself carries hidden costs, including the expense of data servers, GPUs, and high energy consumption, which raises sustainability concerns. This resonates with Gholizadeh et al. (2024), who identify equipment, training, and data management as practical cost barriers, and with Niazi (2023), who highlights the environmental impact of computationally intensive AI models. For 2 participants, the call for public and private funding highlights the literature's focus on shared financial contributions responsibility: Pistollato et al. (2025) and Schmeisser et al. (2023) stress the importance of coordinated funding initiatives, such as the EU Green Deal, to spread costs and accelerate adoption.

On the other hand, 20% of the participants pointed to the long-term return on investment (ROI), emphasizing cost savings from faster drug development and reduced late-stage clinical failures, as one noted: "For many endpoints, validated in vitro and defined approaches are faster and less expensive than in vivo studies, creating a strategic ROI that outweighs initial expenditures". They added that NAMs may be costly initially but could prevent expensive attrition in clinical phases. This aligns with Gangwal and Lavecchia (2025) and Bhattamisra et

al. (2023), who demonstrate how AI-enabled triaging can reduce late-stage drug failures: a key driver of R&D costs, often exceeding billions of dollars per failed compound (Rudroff, 2024; Silva & Tamburic, 2022; Fontana et al., 2021).

Another one explicitly tied financial benefits to accelerated drug development timelines, a view supported by Liu et al. (2021) and Patel and Patel (2024), who highlight AI's role in compressing development cycles.

Finally, 13% did not provide concrete views, reflecting either uncertainty or skepticism, as one stated: "I do not have the answer, but am interested in finding out". This echoes Hunter (2023), who questions narratives of cost reduction by arguing that animal testing, despite its inefficiencies, has historically produced irreplaceable, life-saving breakthroughs. Such positions highlight that, while the financial promise of NAMs is compelling, skepticism persists in parts of the scientific and regulatory community.

In synthesis, participants' perspectives align with the literature in describing the financial implications of the transition as a short-term burden but a long-term benefit. High costs for infrastructure, validation, and training create immediate obstacles, especially for smaller actors. However, NAMs offer reduced attrition, faster timelines, and better alignment with consumer and regulatory demands in the medium to long term. As Romberg et al. (2012) show, the transition can be cost-effective within a few years if funding mechanisms and validation pathways are strategically coordinated.

The fifth and final subcategory (1.2.5) of this investigation highlighted the practical enablers and barriers that most influence the implementation of alternative methods in a real-world setting.

*Table 5.7 - Practical enablers and barriers that most influence the implementation of alternative methods in a real-world setting.*

| TEXT | GENERIC CATEGORY | SUBCATEGORY | TIMES MENTIONED | PARTICIPANTS |
|---|---|---|---|---|
| Expertise, training, and professional culture as facilitators | 1.2 | 1.2.5 | 10 | P1, P2, P5, P6, P7, P10, P11, P12, P13, P14 |
| Multi-stakeholder partnerships, combining regulators, scientists, and industry actors to validate and harmonize new approaches as facilitators | 1.2 | 1.2.5 | 4 | P2, P14, P5, P11 |
| Societal and consumer forces as facilitators | 1.2 | 1.2.5 | 1 | P11 |
| Entrenched reliance on animal tests within institutions and regulatory culture as barriers | 1.2 | 1.2.5 | 6 | P4, P6, P7, P11, P12, P15 |
| Technical and infrastructural limitations as barriers | 1.2 | 1.2.5 | 8 | P6, P7, P9, P10, P11, P12, P13, P14 |
| Initial financial investment as barrier | 1.2 | 1.2.5 | 1 | P11 |

*Source: Self-Elaborated*

According to Table 5.7, the responses of participants emphasize that the transition from animal testing to AI-supported NAMs is not merely a technical or financial challenge but is deeply embedded in cultural, institutional, and regulatory contexts. Consistent with the literature, the participants identified both facilitators, such as training, stakeholder collaboration, ethical and consumer pressure, and regulatory engagement, as well as barriers, including entrenched institutional reliance on animal models, regulatory fragmentation, cultural inertia, and resource constraints.

A recurrent theme among over half of the participants, 10 out of 15, was the centrality of expertise, training, and professional culture. As two of them noted, the "training of professionals" and investing in younger scientists are essential for cultivating an openness to innovation. At the same time, one emphasized that retraining must begin as early as secondary education to foster generational change. This view aligns directly with Cronin et al. (2025) and EMA (2020), who emphasize that structured training and institutional "safe spaces" for testing new methodologies are crucial to building regulatory confidence and accelerating NAM adoption. Similarly, Romberg et al. (2012) demonstrated in the vaccine consistency framework that systematic training and validation pipelines can serve as scalable models for transitioning away from animals.

Collaboration was also underlined as a vital enabler. 13% emphasized the importance of multi-stakeholder partnerships, which involve combining regulators, scientists, and industry actors to validate and harmonize new approaches. This aligns with Fentem et al. (2021) and the FDA (2025), who emphasize the importance of consortia, cross-sector partnerships, and international harmonization as foundational to an effective transition.

Two participants further highlighted the role of industry partnerships, echoing industrial pioneers such as L'Oréal (Belén, 2020a, as cited in Mendivelso et al. 2024), which have already demonstrated the large-scale feasibility of non-animal methods in the cosmetics sector. Beyond institutional actors, societal and consumer forces were also identified as facilitators by 7% of the participants. This resonates with Marshall et al. (2022), Mendivelso et al. (2024), and Winarto and Wisesa (2024), who demonstrate that consumer demand for cruelty-free and transparent products continuously exerts pressure on companies to move away from animal testing, particularly in the cosmetics industry.

As one participant added to the literature review (Connor et al., 2022; Yamazaki and Ishida, 2025; Romberg et al., 2012; Cronin et al., 2025), "key enablers include strong regulatory and executive support, with clearly defined contexts of use and decision rules, supported by early

scientific advice mechanisms such as EMA's ITF or FDA's ISTAND. Additionally, establishing a culture of quality through the mandatory adoption of frameworks like GIVIMP and GCCP 2.0, along with the use of reproducibility checklists, further enhances trust in the results. Collaborative consortia also play a vital role by enabling the sharing of reference data, benchmarks, and validation studies across institutions".

Conversely, substantial barriers were noted. A significant challenge is the entrenched reliance on animal tests within institutions and regulatory culture, identified by 6 out of 15 participants. One explicitly noted "institutional reliance" as the biggest obstacle, while another underscored the persistence of the "in vivo veritas" mindset among older regulators and researchers. This aligns with the findings of Knight et al. (2021), who argue that legacy legal frameworks, such as REACH, often maintain animal testing as the default in many cases, even when validated alternatives exist. Amador (2021) and Akkermans et al. (2020) similarly emphasize that global divergence, particularly China's continued requirements for specific animal tests, creates compliance contradictions that undermine progress.

Technical and infrastructural limitations were also cited by 53%. For instance, one identified a lack of expertise as a bottleneck, while another pointed to challenges in data governance, reproducibility, and traceability. These concerns are echoed in the literature: Westmoreland et al. (2022) and Kimura et al. (2025) note that many OoC and MPS remain technically immature and difficult to reproduce across laboratories; Patel and Patel (2024) and Niazi (2023) warn that AI's "black-box" opacity and bias undermine regulatory confidence; and Gholizadeh et al. (2024) stress that high infrastructure and training costs hinder adoption, particularly for SMEs. 7% also underlined the barrier of initial financial investment, a challenge that Lee et al. (2022) link to productivity dips during the adoption of AI-driven processes, which can slow institutional change despite long-term efficiency gains.

Cultural and political factors were equally noted by 5 out of 15 participants. One described political "arcaicismo" (backwardness) as a barrier. At the same time, another 13% stressed national culture and misinformation as obstacles to scientific progress, as one added, "Differences in legislation in different parts of the world, internet nonsense and fake news". These insights align with Schmeisser et al. (2023), who argue that while regulatory pilots exist, uptake remains slow due to inertia and lack of political will.

Taken together, the participants' responses confirm that enablers and barriers to NAMs adoption operate across several dimensions: regulatory (frameworks and harmonization), institutional (legacy reliance and organizational inertia), technical (data, reproducibility,

validation), financial (infrastructure and training costs), and socio-cultural (professional mindsets, public perception, and consumer demand). The literature agrees on the same point: successful transition requires both structural and cultural change. Regulatory innovation, international harmonization, and technical validation must be supported by training, stakeholder collaboration, and alignment with societal ethics and values. The challenge, therefore, is not only to develop technological alternatives but also to integrate them into a system that has traditionally relied on the use of animals in testing.

# Chapter VI – Conclusion

## 6.1 Final Considerations

Taking into consideration the previous analysis of data gathered from both participants and authors regarding the topics of AI and NAMs, ethics, regulation, technology, and finances, this final chapter aims to summarize the pivotal findings of this investigation. The subject of animal testing has been a topic of debate for decades. Still, in recent years, its scientific and ethical limitations have become increasingly apparent, while AI-driven non-animal approaches have gained significant traction and attention. Equally, regulatory science and toxicology are evolving to integrate these methods into decision-making frameworks. Additionally, ethical and financial dimensions have been attracting interest, as this transition carries profound societal, economic, and moral implications. As one participant stated, "3Rs as the ethical baseline to move from animal by default to human-relevant by design", corroborating the findings of Silva & Tamburic (2022), Neuhaus et al. (2022), Fentem et al. (2021).

Primarily, the first research question of this investigation intended to identify the main advantages and limitations of animal testing compared with AI-supported NAMs in the cosmetics and pharmaceutical sectors. To address this, interviewees were asked to reflect on predictive value and ethical implications. It was possible to infer that while animal testing continues to provide systemic insights for endpoints such as repeated-dose or developmental toxicity, its lack of human relevance undermines its credibility.

Conversely, AI-supported NAMs were consistently described as ethically preferable, more mechanistically anchored, and more human-relevant. However, challenges remain in reproducibility, standardization, and validation for complex systemic endpoints. As one participant stated, "Although animal testing has a legacy of regulatory familiarity and extensive historical datasets, animal testing also has limitations, like human predictivity often being limited, and their high clinical failure rates highlight translational gaps from animals to humans. They are costly, slow, and ethically problematic, and not scalable for modern chemical and drug portfolios." For NAMs, their limitations fall under "coverage gaps, for example, some reproductive and developmental toxicities, and their data quality, curation, and applicability domain limits for ML and QSAR", adding to Van Berlo et al. (2021), Kimura et al. (2025), Westmoreland et al. (2022), and Patel and Patel (2024) perspectives.

The goal of the second research question was to determine in which situations AI can replace or complement animal testing. From the interviews and literature reviewed, it became clear that in cosmetics, replacement is already a reality for endpoints such as skin and eye irritation, sensitization, and specific genotoxicity tests, where validated in vitro and in silico models are internationally accepted. In pharmaceuticals, replacement is more limited, with NAMs currently functioning as complementary tools in integrated assessment strategies. These help reduce reliance on animal data and improve predictive capacity, but have not yet eliminated the use of animal models in complex domains, such as developmental or reproductive toxicology.

As a participant put it, "NAMs can complement animal testing by enabling better analysis and assessment of results, identifying toxicity more rapidly, and comparing large numbers of animal test results to identify patterns. They can be replaced by combining large amounts of information from multiple NAMs for hazard and exposure to gain a response that can be used in chemical safety assessment", as Movia and Prina-Mello (2020), Van Berlo et al. (2021), Gangwal and Lavecchia (2025), Taylor and Alvarez (2020), Bhattamisra et al. (2023), and Van Norman (2019) stated throughout the literature review.

The third research question aimed to identify the ethical principles and guidelines that should guide this transition. Both the literature and interviewees highlighted the enduring importance of the 3Rs principle as well as EU Directive 2010/63/EU and broader international guidelines. In the cosmetics industry, the ethical imperative is particularly strong, as cruelty-free innovation aligns with both consumer demand and legal requirements. In contrast, the ethical debate in pharmaceuticals is complicated by the life-saving imperatives. Nonetheless, there was consensus that cruelty-free science is not only a moral obligation but also a societal expectation, corroborating with EMA (1997, 2016, 2020) principles and Westmoreland et al. (2022) and Cronin et al. (2025) perspectives.

One participant summarized the ethical basis clearly: "Ethical principles encompass reduction of animal use, a reduction in the use of resources. Guidelines should follow a clear understanding of what we are trying to do, for instance, ensure a chemical is safe for use… the purpose should not be to replace the animal test, but to provide information that can be used to assure safety."

The fourth research question examined which regulatory frameworks should guide companies during this transition. The findings confirm that the OECD TG, ICH frameworks, and legislative initiatives such as the US FDA Modernization Act 2.0 are essential in advancing NAMs. At the same time, regulatory fragmentation remains a significant barrier. Conflicting

requirements, for example, between REACH obligations and the EU Cosmetics Regulation, or between EU bans and mandatory animal tests in Asian markets, create uncertainty and duplication.

Experts emphasized the need for harmonization, cross-sector consistency, and more explicit regulatory guidance to accelerate the adoption of these measures. One participant noted that "all regulatory frameworks should be adhered to," but highlighted the complexity of "over 40 pieces of EU chemical legislation, plus diverging requirements in the US, Canada, Japan, China, and India." They argued this "calls for global coordination and cross-sector alignment," pointing to the EU Green Deal's "One Substance, One Assessment" as a strong example, as Cronin et al. (2025) and Pistollato et al. (2025) defended.

The fifth research question was designed to identify the most essential technological implementations and technical challenges. The findings highlight OoC systems, MPS platforms, PBK modeling, and AI integration with in vitro and in chemico assays as key enablers, with models like DeepTox and AnimalGAN showcasing their potential. However, challenges such as data bias, reproducibility, algorithmic opacity, and the absence of standardized FAIR-compliant datasets remain significant. As participants emphasized, "data should be one of the most important," with success depending on "robust database and infrastructure" and ensuring that "for AI, the main technological challenge is what data to feed, and to make sure it is of high quality."

Participants highlighted techniques such as "transcriptomics, 3D cultures, high-throughput screening, AI, and PBPK modeling" as crucial. However, concerns remain over "data scarcity, batch effects, generalization & drift," as well as the need for "traceability" to support regulatory review. They also emphasized that the main barriers are not purely technical, but rather linked to validation and funding.

The sixth research question focused on the financial implications of the transition. Although significant upfront costs in infrastructure, training, and data management are required, both literature and interviewees highlighted long-term benefits. These include faster decision-making, earlier failure detection, reductions in R&D expenditure of up to 70%, and improved portfolio efficiency. In addition, companies benefit from reputational and branding advantages by embracing cruelty-free innovation, which has become an increasingly important driver of differentiation, particularly in the cosmetics sector.

As one participant stated, "there is a large investment needed at the outset to develop NAMs, also to take them beyond the initial stages. Academic labs, as well as small

businesses, will never be able to afford this. The move to NAMs should make chemical safety testing cheaper and more efficient in the long term", which goes hand in hand with Lee et al. (2022), Niazi (2023) and Gholizadeh et al. (2024) ideals.

The seventh question identified key enablers and barriers. Success depends on access to high-quality, auditable data, robust validation processes, and an organizational capacity for effective regulation and data management. Barriers include institutional inertia, high start-up costs, inconsistent regulatory frameworks, and limited expertise in smaller firms. Facilitators include international partnerships, pre-competitive collaborations, demonstration projects, and clear roadmaps that foster confidence and accelerate progress.

Another participant added that "current regulations mostly require animal test data, so those considerations need to be rethought, and regulations rewritten and redesigned to move to the 21st century." Moreover, "there needs to be many levels of retraining and education. The new mindset should be introduced in secondary schools and reinforced in undergraduate and postgraduate education, so that the new generation of scientists will understand the issues and benefits of NAMs. Naturally, education is also required for the current professionals", such as Niazi (2023), Marshall et al. (2022), Neuhaus et al. (2022), and EMA (2020) advised.

Overall, the evidence suggests that the shift away from animal testing is progressing, with certain endpoints already being replaced in cosmetics and AI supporting gradual changes in pharmaceuticals. Ethical concerns, technological progress, economic incentives, and evolving regulations are converging to create a safer, more humane, and scientifically credible system. Although incomplete and uneven, this transformation marks a significant step toward future innovation where safety assessment is both human-relevant and ethically responsible.

### 6.2 Theoretical and Practical Implications

This study offers an integrated framework of complementarity: AI-supported methods and NAMs should replace animal testing when validated, serving as complements until evidence, reproducibility, and regulations support complete replacement. This combines ethical, regulatory, technological, and financial factors into a pragmatic guide for responsible use, moving beyond polarized debates. Practically, it highlights that companies need to integrate NAMs and AI by connecting endpoints to validated solutions, creating clear decision rules, and establishing early triage to optimize resources.

Capabilities in data management, validation, AI governance, and regulatory knowledge are crucial for credibility. Transparent communication and stakeholder engagement are vital for maintaining social license and trust. On policy, international harmonization should expand OECD mutual acceptance, clarify legal issues, and develop endpoint-specific regulations. Collaborations, data sharing, and projects like L'Oréal's reconstructed skin models or vaccine testing demonstrate paths for broader adoption.

### 6.3 Limitations

Despite rich findings, several limitations exist. The qualitative research, comprising 15 interviews, provided detailed insights but limited generalizability. Cultural and language biases may affect perspectives. As AI and regulation evolve rapidly, conclusions may need to be revised. The study relied on literature and expert views to discuss cost-effectiveness and accuracy, but lacked quantitative validation. It also didn't thoroughly examine issues like data privacy, bias, and model opacity. Systemic endpoints remain underexplored due to the limitations of current development methods. Organizational aspects, such as capacity building and workforce training, need further study, as the research only captures snapshots rather than ongoing processes.

### 6.4 Directions for Future Research

Future research should focus on empirical comparisons of AI/NAMs and animal testing across various endpoints, product categories, and scales to inform policy. Developing validation frameworks that incorporate reproducibility, uncertainty, and mechanistic explanations can enhance regulatory confidence and acceptance. Studying regulatory and cultural differences globally will help clarify adoption challenges, particularly in terms of legal conflicts and societal attitudes. Organizational research can track a company's adaptation and structural changes over time. Examining AI governance issues, such as data quality, transparency, bias, and privacy, is crucial for establishing trust. Incorporating consumer, advocacy, and investor perspectives can address societal drivers and prevent greenwashing. Additionally, investigating frontier technologies such as digital twins, generative adversarial networks, and multi-OoC platforms for scalability, readiness, and regulatory acceptance will guide future adoption.

## References


Akkermans, A., Chapsal, J. M., Coccia, E. M., Depraetere, H., Dierick, J. F., Duangkhae, P., Goel, S., Halder, M., Hendriksen, C., Levis, R., Pinyosukhee, K., Pullirsch, D., Sanyal, G., Shi, L., Sitrin, R., Smith, D., Stickings, P., Terao, E., Uhlrich, S., Viviani, L., & Webster, J. (2020). Animal testing for vaccines: Implementing replacement, reduction and refinement—Challenges and priorities. Biologicals, 68, 92–107. https://doi.org/10.1016/j.biologicals.2020.07.010

Alépée, N., Adriaens, E., Abo, T., Bagley, D., Desprez, B., Hibatallah, J., Mewes, K., Pfannenbecker, U., Sala, À., Van Rompay, A. R., Verstraelen, S., & McNamee, P. (2019). Development of a defined approach for eye irritation or serious eye damage for liquids, neat and in dilution, based on Cosmetics Europe analysis of in vitro STE and BCOP test methods. Toxicology in Vitro, 57, 154–163. https://doi.org/10.1016/j.tiv.2019.02.019

Alnasser, S. M. (2025). Revisiting the approaches to DNA damage detection in genetic toxicology: Insights and regulatory implications. BioData Mining, 18, 33. https://doi.org/10.1186/s13040-025-00447-8

Alves, V. M., Auerbach, S. S., Kleinstreuer, N., Rooney, J. P., Muratov, E. N., Rusyn, I., Tropsha, A., & Schmitt, C. (2021). Curated Data In - Trustworthy In Silico Models Out: The Impact of Data Quality on the Reliability of Artificial Intelligence Models as Alternatives to Animal Testing. Alternatives to laboratory animals: ATLA, 49(3), 73–82. https://doi.org/10.1177/02611929211029635

Amador, V. (2021, January 26). Luxury brands and the pressures to end their use of animal skin and animal testing (Master's thesis, Universidade Católica Portuguesa). Ciência-UCP. http://hdl.handle.net/10400.14/34883

Bailey, J., & Balls, M. (2019). Recent efforts to elucidate the scientific validity of animal-based drug tests by the pharmaceutical industry, pro-testing lobby groups, and animal welfare organisations. BMC Medical Ethics, 20(1), 16. https://doi.org/10.1186/s12910-019-0352-3

Bardin, L. (2018). Análise de Conteúdo (E. 70, Ed.).

Belenguer L. (2022). AI bias: exploring discriminatory algorithmic decision-making models and the application of possible machine-centric solutions adapted from the pharmaceutical industry. AI and ethics, 2(4), 771–787. https://doi.org/10.1007/s43681-022-00138-8

Bhattamisra, S. K., Banerjee, P., Gupta, P., Mayuren, J., Patra, S., & Candasamy, M. (2023). Artificial Intelligence in Pharmaceutical and Healthcare Research. Big Data and Cognitive Computing, 7(1), 10. https://doi.org/10.3390/bdcc7010010

Connor, S., Li, T., Roberts, R., Thakkar, S., Liu, Z., & Tong, W. (2022). Adaptability of AI for safety evaluation in regulatory science: A case study of drug-induced liver injury. Frontiers in artificial intelligence, 5, 1034631. https://doi.org/10.3389/frai.2022.1034631

Contreras López, C., & Montes Franceschini, M. (2025). Animal welfare and the law: History, definition, and ethics. Metode Science Studies Journal, 15(5), e29421. https://doi.org/10.7203/metode.16.29421

Cronin, M. T. D., Berggren, E., Camorani, S., Desaintes, C., Fabbri, M., Fabrega, J., Herzler, M., Ingram, J. D. E., Lacasse, K., Louhimies, S., Maxwell, G., Schutte, K., Sobanski, T., Streck, G., Terron, A., & Worth, A. P. (2025). Report of the European Commission workshop on "The roadmap towards phasing out animal testing for chemical safety assessments", Brussels, 11–12 December 2023. Regulatory Toxicology and Pharmacology, 161, 105818. https://doi.org/10.1016/j.yrtph.2025.105818

Elder, A., Cappelli, M. O., Ring, C., & Saedi, N. (2024). Artificial intelligence in cosmetic dermatology: An update on current trends. Clinics in dermatology, 42(3), 216–220. https://doi.org/10.1016/j.clindermatol.2023.12.015

Eskes, C., Cole, T., Hoffmann, S., Worth, A., Cockshott, A., Gerner, I., & Zuang, V. (2007). The ECVAM international validation study on in vitro tests for acute skin irritation: selection of test chemicals. Alternatives to laboratory animals: ATLA, 35(6), 603–619. https://doi.org/10.1177/026119290703500608

European Medicines Agency. (1997). Replacement of animal studies by in vitro models. Retrieved from https://www.ema.europa.eu/en/documents/scientific-guideline/replacement-animal-studies-vitro-models_en.pdf

European Medicines Agency. (2016, December 15). Guideline on the principles of regulatory acceptance of 3Rs (replacement, reduction, refinement) testing approaches (EMA/CHMP/CVMP/JEG-3Rs/450091/2012). https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-principles-regulatory-acceptance-3rs-replacement-reduction-refinement-testing-approaches_en.pdf

European Medicines Agency. (2020). EMA regulatory science to 2025: Strategic reflection. https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/ema-regulatory-science-2025-strategic-reflection_en.pdf

Fentem, J., Malcomber, I., Maxwell, G., & Westmoreland, C. (2021). Upholding the EU's Commitment to 'Animal Testing as a Last Resort' Under REACH Requires a Paradigm Shift in How We Assess Chemical Safety to Close the Gap Between Regulatory Testing and Modern Safety Science. Alternatives to laboratory animals: ATLA, 49(4), 122–132. https://doi.org/10.1177/02611929211040824

Gangwal, A., & Lavecchia, A. (2025). Artificial intelligence in preclinical research: enhancing digital twins and organ-on-chip to reduce animal testing. Drug discovery today, 30(5), 104360. https://doi.org/10.1016/j.drudis.2025.104360

Gholizadeh, N., Rokni, G. R., & Babaei, M. (2024). Advantages and disadvantages of using AI in dermatology. Dermatological Reviews, 5, e248. https://doi.org/10.1002/der2.248

Gómez Mendivelso, J, Fandiño Polo, M y Ardila Duran, Y. (2024). Análisis de estudio del caso L'oreal en la incidencia en las empresas del sector de la industria cosmética en Colombia. . Corporación Unificada Nacional de Educación Superior - CUN. Disponible en: https://repositorio.cun.edu.co/handle/cun/10334

Grech, V. S., Kefala, V., & Rallis, E. (2024). Cosmetology in the Era of Artificial Intelligence. Cosmetics, 11(4), 135. https://doi.org/10.3390/cosmetics11040135

Gutleb, A. C., Murugadoss, S., Stępnik, M., SenGupta, T., El Yamani, N., Longhin, E. M., Olsen, A.-K. H., Wyrzykowska, E., Jagiello, K., Judzinska, B., Cambier, S., Honza, T., McFadden, E., Shaposhnikov, S., Puzyn, T., Serchi, T., Weber, P., Arnesdotter, E., Skakalova, V., Jirsova, K., Grudzinski, I. P., Collins, A., Rundén-Pran, E., Dusinska, M. (2025). New approach methods (NAMs) for genotoxicity assessment of nano- and advanced materials; Advantages and challenges. Mutation Research - Genetic Toxicology and Environmental Mutagenesis, 904, 503867. https://doi.org/10.1016/j.mrgentox.2025.503867

Hash, M. G., Forsyth, A., Coleman, B. A., Li, V., Vinagolu-Baur, J., & Frasier, K. M. (2025). Artificial Intelligence in the Evolution of Customized Skincare Regimens. Cureus, 17(4), e82510. https://doi.org/10.7759/cureus.82510

Kabene, S., & Baadel, S. (2019). Bioethics: a look at animal testing in medicine and cosmetics in the UK. Journal of medical ethics and history of medicine, 12, 15. https://doi.org/10.18502/jmehm.v12i15.1875

Kabene, S., & Baadel, S. (2019). Bioethics: a look at animal testing in medicine and cosmetics in the UK. Journal of medical ethics and history of medicine, 12, 15. https://doi.org/10.18502/jmehm.v12i15.1875

Kapustina, O., Burmakina, P., Gubina, N., Serov, N., & Vinogradov, V. (2024). User-friendly and industry-integrated AI for medicinal chemists and pharmaceuticals. Artificial Intelligence Chemistry, 2(2), 100072. https://doi.org/10.1016/j.aichem.2024.100072

Kimura, H., Nishikawa, M., Kutsuzawa, N., Tokito, F., Kobayashi, T., Kurniawan, D. A., Shioda, H., Cao, W., Shinha, K., Nakamura, H., Doi, K., & Sakai, Y. (2025). Advancements in microphysiological systems: Exploring organoids and organ-on-a-chip technologies in drug development—Focus on pharmacokinetics related organs. Drug Metabolism and Pharmacokinetics, 60, 101046. https://doi.org/10.1016/j.dmpk.2024.101046

Kleinstreuer, N., & Hartung, T. (2024). Artificial intelligence (AI)-it's the end of the tox as we know it (and I feel fine). Archives of toxicology, 98(3), 735–754. https://doi.org/10.1007/s00204-023-03666-2

Knight, J., Rovida, C., Kreiling, R., Zhu, C., Knudsen, M., & Hartung, T. (2021). Continuing animal tests on cosmetic ingredients for REACH in the EU. ALTEX, 38(4), 653–668. https://doi.org/10.14573/altex.2104221

Lee, Y. S., Kim, T., Choi, S., & Kim, W. (2022). When does AI pay off? AI-adoption intensity, complementary investments, and R&D strategy. Technovation, 118, 102590. https://doi.org/10.1016/j.technovation.2022.102590

Liu, Y., Cai, Z., Li, N., & Alépée, N. (2023). How to facilitate the implementation of 3D models in China by applying good in vitro method practice for regulatory use. Frontiers in toxicology, 5, 1080528. https://doi.org/10.3389/ftox.2023.1080528

Liu, Z., Roberts, R. A., Lal-Nag, M., Chen, X., Huang, R., & Tong, W. (2021). AI-based language models powering drug discovery and development. Drug discovery today, 26(11), 2593–2607. https://doi.org/10.1016/j.drudis.2021.06.009

Madden, J. C., Enoch, S. J., Paini, A., & Cronin, M. T. D. (2020). A Review of In Silico Tools as Alternatives to Animal Testing: Principles, Resources and Applications. Alternatives to laboratory animals: ATLA, 48(4), 146–172. https://doi.org/10.1177/0261192920965977

Malik, S., Muhammad, K., & Waheed, Y. (2024). Artificial intelligence and industrial applications — A revolution in modern industries. Ain Shams Engineering Journal, 15(9), 102886. https://doi.org/10.1016/j.asej.2024.102886

Manful, M. E., Ahmed, L., & Barry-Ryan, C. (2023). New approach methodologies (NAMs) for safety testing of complex food matrices: A review of status, considerations, and regulatory adoption. Trends in Food Science & Technology, 142, 104191. https://doi.org/10.1016/j.tifs.2023.104191

Marshall, L. J., Constantino, H., & Seidle, T. (2022). Phase-In to Phase-Out-Targeted, Inclusive Strategies Are Needed to Enable Full Replacement of Animal Use in the European Union. Animals: an open access journal from MDPI, 12(7), 863. https://doi.org/10.3390/ani12070863

Movia, D., & Prina-Mello, A. (2020). Preclinical development of orally inhaled drugs (OIDs)—Are animal models predictive or shall we move towards in vitro non-animal models? Animals, 10(8), 1259. https://doi.org/10.3390/ani10081259

Neuhaus, W., Reininger-Gutmann, B., Rinner, B., Plasenzotti, R., Wilflingseder, D., De Kock, J., Vanhaecke, T., Rogiers, V., Jírová, D., Kejlová, K., Knudsen, L. E., Nielsen, R. N., Kleuser, B., Kral, V., Thöne-Reineke, C., Hartung, T., Pallocca, G., Rovida, C., Leist, M., Hippenstiel, S., … Spielmann, H. (2022). The Current Status and Work of Three Rs Centres and Platforms in Europe. Alternatives to laboratory animals: ATLA, 50(6), 381–413. https://doi.org/10.1177/02611929221140909 Niazi S. K. (2023). The Coming of Age of AI/ML in Drug Discovery, Development, Clinical Testing, and Manufacturing: The FDA Perspectives. Drug design, development and therapy, 17, 2691–2725. https://doi.org/10.2147/DDDT.S424991

Pai, S., Binu, A., Lavanya, G. S., Harikumar, M., Herga, S. K., Citartan, M., & Mani, N. K. (2025). Advancements of paper-based microfluidics and organ-on-a-chip models in cosmetics hazards. RSC Advances, 15, 10319–10335. https://doi.org/10.1039/d4ra07336c

Pai, S., Binu, A., Lavanya, G. S., Harikumar, M., Kedlaya Herga, S., Citartan, M., & Mani, N. K. (2025). Advancements of paper-based microfluidics and organ-on-a-chip models in cosmetics hazards. RSC advances, 15(13), 10319–10335. https://doi.org/10.1039/d4ra07336c

Pastorino, P., Prearo, M., & Barceló, D. (2024). Ethical principles and scientific advancements: In vitro, in silico, and non-vertebrate animal approaches for a green ecotoxicology. Green Analytical Chemistry, 8, 100096. https://doi.org/10.1016/j.greeac.2024.100096

Pereira da Silva, V., de Carvalho Brito, L., Mesquita Marques, A., da Cunha Camillo, F., & Raquel Figueiredo, M. (2023). Bioactive limonoids from Carapa guianensis seeds oil and the sustainable use of its by-products. Current research in toxicology, 4, 100104. https://doi.org/10.1016/j.crtox.2023.100104

Pistollato, F., Furtmann, F., Abitbol, S., Agostinho, M., Andreu, T., Baluta, B., Beets-Tan, R. G. H., Beken, S., Ciotti, S., Coecke, S., Correia, T., Desaintes, C., Destrebecq, F., Detilleux, P., Di Virgilio, S., Effraimidou, A., Ferreira Bastos, L., Gastaldello, A., Habermann, J. K., Hanes, I., Jakovljevic, M., Laverty, H., Maisonneuve, B., Marshall, L. J., Movia, D., Pallocca, G., Pappalardo, F., Pastorino, R., Radvilaite, V., Reid, K., Riegger, S., Russo, G., Salvatori, D., Speirs, V., Straccia, M., Tagle, D. A., Turner, J., van den Eijnden-van Raaij, J., Viviani, L., & Constantino, H. (2025). Leveraging innovative research tools to meet public health challenges: A BioMed21 workshop report. NAM Journal, 1, 100023. https://doi.org/10.1016/j.namjnl.2025.100023

Poh, W. T., & Stanslas, J. (2024). The new paradigm in animal testing - "3Rs alternatives". Regulatory toxicology and pharmacology : RTP, 153, 105705. https://doi.org/10.1016/j.yrtph.2024.105705

Ram, R. N., Gadaleta, D., & Allen, T. E. H. (2022). The role of 'big data' and 'in silico' New Approach Methodologies (NAMs) in ending animal use – A commentary on progress. Computational Toxicology, 23, 100232. https://doi.org/10.1016/j.comtox.2022.100232

Rinaldi, R., Cioffi, A., & Ciallella, C. (2020). Reflections concerning the legitimacy of animal testing. Ethics, Medicine and Public Health, 15, 100569. https://doi.org/10.1016/j.jemep.2020.100569

Rohankumar Patel, & Ankur Patel. (2024). Revolutionizing Drug Development: AI-Driven Predictive Modeling for Accelerated Small Molecule and Biologic Therapeutics. Well Testing Journal, 33(S2), 668–691. Retrieved from https://welltestingjournal.com/index.php/WT/article/view/148

Romberg, J., Lang, S., Balks, E., Kamphuis, E., Duchow, K., Loos, D., Rau, H., Motitschke, A., & Jungbäck, C. (2012). Potency testing of veterinary vaccines: the way from in vivo to in vitro. Biologicals:journal of the International Association of Biological Standardization, 40(1), 100–106. https://doi.org/10.1016/j.biologicals.2011.10.004

Schmeisser, S., Miccoli, A., von Bergen, M., Berggren, E., Braeuning, A., Busch, W., Desaintes, C., Gourmelon, A., Grafström, R., Harrill, J., Hartung, T., Herzler, M., Kass, G. E. N., Kleinstreuer, N., Leist, M., Luijten, M., Marx-Stoelting, P., Poetz, O., van Ravenzwaay, B., Roggeband, R., … Tralau, T. (2023). New approach methodologies in human regulatory toxicology - Not if, but how and when!. Environment international, 178, 108082. https://doi.org/10.1016/j.envint.2023.108082

Shiammala, P. N., Duraimutharasan, N. K. B., Vaseeharan, B., Alothaim, A. S., Al-Malki, E. S., Snekaa, B., Safi, S. Z., Singh, S. K., Velmurugan, D., & Selvaraj, C. (2023). Exploring the artificial intelligence and machine learning models in the context of drug design difficulties and future potential for the pharmaceutical sectors. Methods (San Diego, Calif.), 219, 82–94. https://doi.org/10.1016/j.ymeth.2023.09.010

Sihaloho, M. P., Kurniawati, & Masnita, Y. (2023). Antecedents and consequences of using artificial intelligence (AI) color cosmetics. International Journal of Social Research, 2(8), 2496–2505. https://ijsr.internationaljournallabs.com/index.php/ijsr/article/view/1293

Silva, R. J., & Tamburic, S. (2022). A State-of-the-Art Review on the Alternatives to Animal Testing for the Safety Assessment of Cosmetics. Cosmetics, 9(5), 90. https://doi.org/10.3390/cosmetics9050090

Sreedhar, D., Manjula, N., Pise, A., Pise, S., & Ligade, V. S. (2020). Ban of cosmetic testing on animals: A brief overview. International Journal of Current Research and Review, 12(14), 113–116. https://doi.org/10.31782/IJCRR.2020.121424

Strickland, J., Truax, J., Corvaro, M., Settivari, R., Henriquez, J., McFadden, J., Gulledge, T., Johnson, V., Gehen, S., Germolec, D., Allen, D. G., & Kleinstreuer, N. (2022). Application of Defined Approaches for Skin Sensitization to Agrochemical Products. Frontiers in toxicology, 4, 852856. https://doi.org/10.3389/ftox.2022.852856

Taylor, K., & Rego Alvarez, L. (2020). Regulatory drivers in the last 20 years towards the use of in silico techniques. Computational Toxicology, 17, 100115. https://doi.org/10.1016/j.comtox.2020.100115

Timpanaro, G., & Cascone, G. (2025). Consumer behavior and decision-making process to buy green cosmetics: A systematic literature review. Heliyon, 11, e42233. https://doi.org/10.1016/j.heliyon.2025.e42233

Timpanaro, G., & Cascone, G. (2025). Consumer behavior and sustainability: Exploring Italy's green cosmetics market with prickly pear seed oil. Heliyon, 11(3), e42233. https://doi.org/10.1016/j.heliyon.2025.e42233

U.S. Food and Drug Administration. (2025). Roadmap to reducing animal testing in preclinical safety studies. https://www.fda.gov/files/newsroom/published/roadmap_to_reducing_animal_testing_in_preclinical_safety_studies.pdf

van Berlo, D., van de Steeg, E., Amirabadi, H. E., & Masereeuw, R. (2021). The potential of multi-organ-on-chip models for assessment of drug disposition as alternative to animal testing. Current Opinion in Toxicology, 27, 8–17. https://doi.org/10.1016/j.cotox.2021.05.001

Van Norman G. A. (2019). Limitations of Animal Studies for Predicting Toxicity in Clinical Trials: Is it Time to Rethink Our Current Approach?. JACC. Basic to translational science, 4(7), 845–854. https://doi.org/10.1016/j.jacbts.2019.10.008

Vilelas, J. (2020). Investigação o Processo de Construção do Conhecimento (3rd ed.). Sílabo

Westmoreland, C., Bender, H. J., Doe, J. E., Jacobs, M. N., Kass, G. E. N., Madia, F., Mahony, C., Manou, I., Maxwell, G., Prieto, P., Roggeband, R., Sobanski, T., Schütte, K., Worth, A. P., Zvonar, Z., & Cronin, M. T. D. (2022). Use of New Approach Methodologies (NAMs) in regulatory decisions for chemical safety: Report from an EPAA Deep Dive Workshop. Regulatory Toxicology and Pharmacology, 135, 105261. https://doi.org/10.1016/j.yrtph.2022.105261

Winarto, L., & Wisesa, A. (2024). Analyzing the impact of artificial intelligence and sustainability on Gen Z consumer purchase intentions: A case study of L'Oréal cosmetics Indonesia. European Journal of Business and Management Research, 9(5). https://doi.org/10.24018/ejbmr.2024.9.5.2241

Wood, A., Atienzar, F., Basili, D., Coulet, M., Fernandez, R., Galano, M., Marin-Kuan, M., Montoya, G., Piechota, P., Punt, A., Reale, E., Wang, S., & Hepburn, P. (2025). Countdown to 2027 - maximising use of NAMs in food safety assessment: closing the gap for regulatory assessments in Europe. Regulatory toxicology and pharmacology: RTP, 162, 105863. https://doi.org/10.1016/j.yrtph.2025.105863

Yamazaki, D., & Ishida, S. (2025). Global expansion of microphysiological systems (MPS) and Japan's initiatives: Innovation in pharmaceutical development and path to regulatory acceptance. Drug metabolism and pharmacokinetics, 60, 101047. https://doi.org/10.1016/j.dmpk.2024.101047

Zarroug, S. H. O. (2025). Caenorhabditis elegans as in vivo model for the screening of natural plants-derived novel anti-aging compounds: A short introduction. Journal of Asian Natural Products Research, 27(4), 577–590. https://doi.org/10.1080/10286020.2024.2414189

Zushin, P. H., Mukherjee, S., & Wu, J. C. (2023). FDA Modernization Act 2.0: transitioning beyond animal models with human cells, organoids, and AI/ML-based approaches. The Journal of clinical investigation, 133(21), e175824. https://doi.org/10.1172/JCI175824

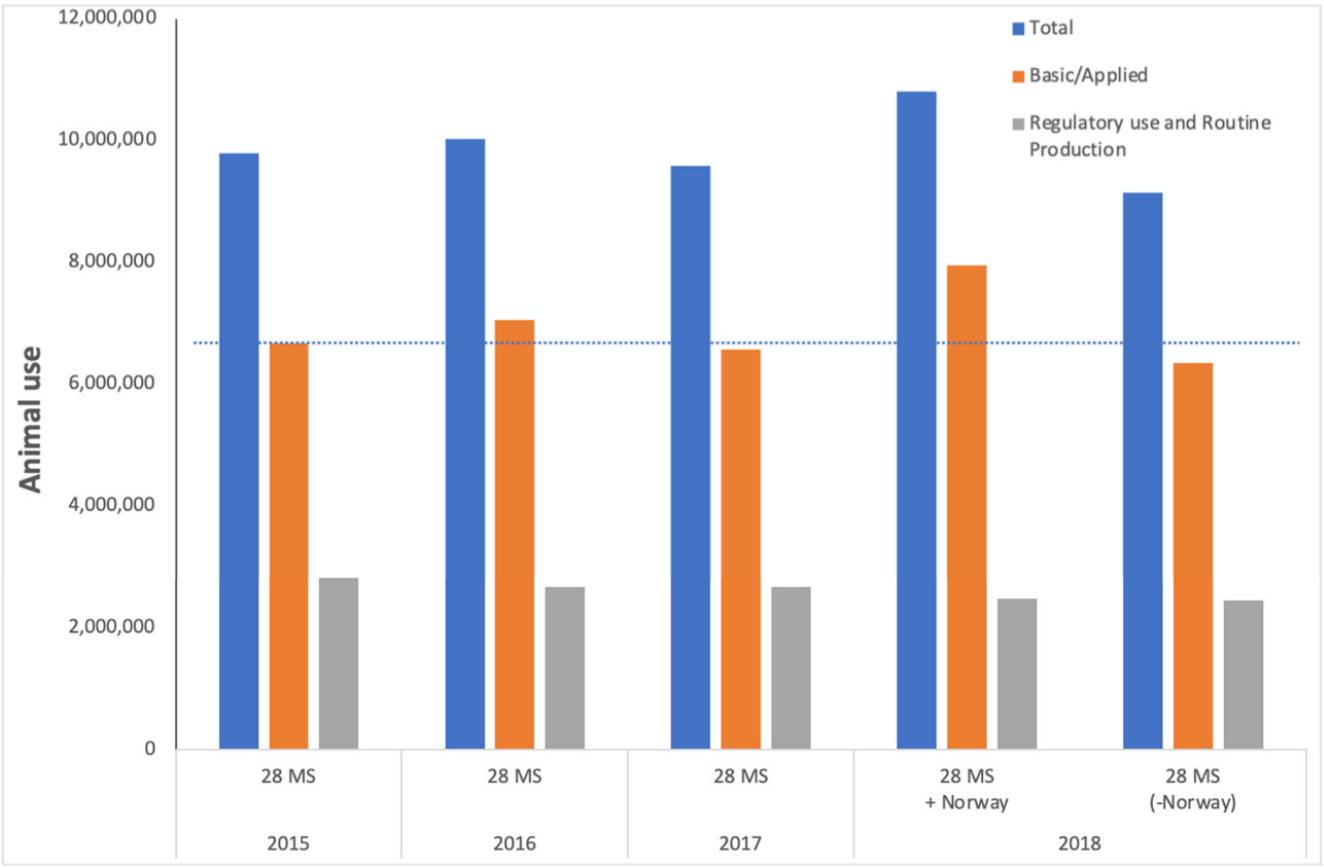
## Annexes

### Annex A – Interview Script

1. Could you briefly describe your profession or area of expertise?
2. What is your level of familiarity with current animal-testing practices and AI technologies in these industries? High, Medium, Low, or None?
3. In your experience, what are the main advantages and limitations of animal testing compared with AI-supported NAMs in the cosmetics and pharmaceutical sectors?
4. In which situations do you think AI can replace or complement animal testing?
5. What ethical principles and guidelines should guide this transition?
6. Which regulatory frameworks should companies adhere to during this transition?
7. What technological implementations are most essential to adopt, and what are the main technological challenges?
8. What are the financial implications of this transition in these sectors?
9. In your opinion, which practical enablers and barriers most influence implementation in a real-world setting?

**Annex B – ALURES database**

"Animal use across the European Union is not undergoing any sustained decline. The blue bars indicate total animal use recorded for each year for all purposes, and according to the counting requirements defined in Directive 2012/63/EU and recorded in ALURES. The orange bars are combined animal use for Basic research and Translational and applied research and the grey bars represent the number of uses of animals for Regulatory use. Note that data from Norway are included in the ALURES database for the first time in 2018, creating an artificial increase in animal use. This is addressed with the data presented as the bars on the far right, which are from the 28 Member States (in direct comparison with the data from 2015, 2016, and 2017). The dotted line indicates the level of use for basic research and translational and applied research in 2015 and tracking this across to the bars on the far right illustrates the absence of any significant decrease in animal use for these purposes."




*Source:* Marshall et al. (2022)