



MANUSCRIPT ID: BTS0926007

A REVIEW ON PHARMA 4.0: THE INTELLIGENT EVOLUTION OF DRUG DEVELOPMENT

*Thanmai sree Bora, Akhil Majji

Raghu College of Pharmacy, Dakamarri, Visakhapatnam, Andhra Pradesh, India.

Corresponding Author Mail Id: thanmaisreebora@gmail.com**Received:** 15 July 2025 | **Revised & Accepted:** 11 January 2026**Keywords:** AI-artificial intelligence, ML-Machine learning, DoE-Design of Experiments, DL-deep learning, RF-Random Forests, CNNs-Convolutional Neural Networks, SVM-Support Vector Machines, RNNs-Recurrent Neural Networks, IID-Inactive Ingredient Database.**Abstract:**

In the face of rising molecular complexity and shorter medication development schedules, classic trial-and-error approaches to pharmaceutical formulation are increasingly ineffective. This study looks at the rise of Predictive Formulation Science, a transformative discipline that uses Artificial Intelligence (AI) and Machine Learning (ML) to model, simulate, and optimise drug formulations with unprecedented precision and efficiency. We look at the fundamentals of artificial intelligence, such as supervised, unsupervised, and deep learning approaches, and how they can be applied to large-scale pharmaceutical datasets, ranging from physicochemical drug qualities and excipient compatibility to experimental design outcomes. Excipient selection, solubility and bioavailability prediction, formulation optimisation, stability modelling, drug release dynamics, and personalised medicine via AI-guided 3D printing are among the critical applications examined. This article also discusses the developing ecosystem of technologies (e.g., TensorFlow, DeepChem, ChemAxon) and datasets (e.g., PubChem, DrugBank, ExCIPI-ENTdb) that are enabling this prediction revolution. While promising case studies from academia and industry indicate AI's real benefits (shorter development time, better decision-making, and cost savings), significant difficulties remain. Data quality, model interpretability, and regulatory uncertainties must all be resolved before widespread adoption may occur. Looking ahead, we investigate cutting-edge developments such as autonomous formulation labs, digital twins, and federated learning, which have the potential to accelerate, smarten, and adapt drug development. Predictive Formulation Science is poised to transform the future of pharmaceuticals by establishing AI as a co-creator alongside scientists, rather than a replacement.

Introduction:

Earlier drug technologies primarily concentrated on enhancing the fundamental physicochemical characteristics of drugs, including their stability, solubility, and bio-availability. However, biological barriers such as the blood-brain barrier and cell membranes became important obstacles for them. As a result, Drug delivery systems that are especially engineered to get past these biological obstacles were

developed. These developments used strategies including improved penetration, tailored distribution to particular tissues or cells, and defence mechanisms in the body that prevented disintegration [1]. These methods ultimately sought to reduce adverse effects and increase therapeutic efficacy. Despite the advancements made by these drug delivery systems, it became evident that a complete strategy was required



to address both physicochemical and biological limitations simultaneously. This triggered in the development of the another drug delivery technology. These attempted to address both physicochemical and biological problems. They strive to optimize drug characteristics, such as solubility, stability, and release kinetics, as well as techniques to overcome biological barriers and enhance drug delivery to the target site [2]. However Drug discovery is a difficult and costly Endeavour with a high failure rate. On average, developing a new drug costs more than \$2.5 billion and takes more than a decade. Furthermore, only a small percentage of therapeutic candidates who undergo clinical trials gain regulatory approval. Despite their efforts, just 2.01% of drug development initiatives produce a commercially marketable medicine. Along with this some significant obstacles need to be solved in order to successfully surpass the limitations of human physiology. Drug intolerance and pathogenic resistance issues necessitate the rapid development of new ideas for potential novel approaches that could produce superior outcomes. These difficulties show how urgently creative methods are needed to speed up and raise the success rates of drug discovery [3]. Technology development and skill enhancement will be crucial in tackling these issues and closing the knowledge gap between academia and the pharmaceutical sector. In the development of pharmaceutical products, the use of artificial intelligence (AI) and machine learning (ML) has become a potent and enabling force. By facilitating automated decision-making, data-driven optimization, and predictive modeling, these technologies have the potential to completely transform formulation science while also shortening development period, increasing repeatability, and cutting expenses. The pharmaceutical industry is quickly implementing AI, as noted in a

MANUSCRIPT ID: BTS0926007

number of recent reviews, not just in clinical trials and drug development but also more and more in preformulation, excipient compatibility studies, and dosage form design [3].

This development is summed up by the phrase "predictive formulation science." In order to predict formulation behaviour, optimize compositions, and simulate product performance without requiring a lot of laboratory testing, artificial intelligence (AI) algorithms and digital technologies are integrated. Before any physical prototypes are developed, scientists can model in vitro and in vivo drug release kinetics, predict drug–excipient interactions, and simulate a variety of formulation scenarios using historical datasets, simulation tools, and real-time analytic's. Machine learning (ML) is a type of artificial intelligence (AI) that uses data to train computational models for operations. For example, ML may enable one to forecast the stability of a certain medication formulation by analyzing data from numerous past trials on the stability of API formulations. Advancements in machine learning techniques, faster computer hardware, and user-friendly tool-kits have made sophisticated ML models more accessible [4].

AI approaches, such as supervised learning (e.g., regression models, random forests, support vector machines) and deep learning (e.g., artificial neural networks, convolutional neural networks), have been successfully applied to problems such as:

- Predicting the physicochemical properties of active pharmaceutical ingredients (APIs)
- Predicting the model drug release profiles from complicated matrices like hydrogels and nanoparticles [4].
- Predicting tablet performance like hardness, friability and disintegration



- Developing innovative delivery technologies, including lipid nanoparticles.
- AI-assisted Design of Experiments (DoE) improves formulation optimisation [5].

Furthermore, given the increasing number of BCS Class II and IV compounds in the research pipeline, AI has proven very helpful in addressing formulation difficulties associated with poorly soluble drugs. More effectively than conventional techniques, machine learning models trained on formulation and biopharmaceutical datasets can identify appropriate carriers (e.g., SNEDDS, co-crystals, amorphous systems), forecast solubility enhancement strategies, and direct the selection of solubilising chemicals [6].

Another area where AI is being used extensively is excipient selection. Researchers can virtually screen huge excipient libraries using predictive compatibility models trained on known API-excipient interactions and molecular descriptors (e.g., Mol2Vec, fingerprinting), reducing down candidates that are chemically and functionally compatible. This decreases material waste and reduces the danger of formulation instability during development [7].

The increasing accessibility of open-source tools (e.g., Scikit-learn, TensorFlow, KNIME) and pharmaceutical databases (e.g., Pub Chem, Drug Bank, ChEMBL) further supports these developments by enabling researchers to create, test, and implement machine learning models for particular formulation tasks. Cloud-based platforms and enhanced processing power have made AI more accessible, even for smaller R&D teams [8].

However, incorporating AI into formulation development presents a number of obstacles. One significant restriction is data quality and standardization. Many pharmaceutical datasets are limited, private, or

MANUSCRIPT ID: BTS0926007

inconsistently annotated, making it challenging to develop strong and generalisable models. Furthermore, regulatory frameworks governing the use of AI in pharmaceutical development are still emerging, particularly in terms of the transparency and explainability of complicated "black-box" models like deep neural networks. Regulatory agencies, like the FDA and EMA, recognize the necessity of model validation and risk-based techniques in drug development to assure safety and efficacy [9].

In the future, the idea of "self-driving laboratories", in which artificial intelligence (AI) systems use robotics and feedback loops to independently plan and carry out formulation experiments, is starting to take shape [10].

This paper provides a thorough overview of the current state of AI-driven formulation creation, examining its applications in pre-formulation, solid dosage forms, nanomedicine, and controlled release systems [11] [12]. We cover the most often utilized ML techniques, tools, case studies, and datasets, as well as critically assess the merits, limitations, and future directions of predictive formulation science. Finally, integration of AI in pharmaceuticals has the promise of enabling smarter, faster, and more personalized drug development, aligning with the goals of precision medicine and industry [13].

1. PRINCIPLES OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING IN PHARMACEUTICALS

Artificial intelligence (AI) is rapidly changing the landscape of pharmaceutical formulation science. At its heart, AI refers to computing systems that can replicate human cognition by learning from data, finding patterns, and generating predictions. Among its sub-fields, machine learning (ML) and deep learning (DL) have



shown notably useful for addressing the nonlinear, multivariate difficulties associated with pharmaceutical data. In formulation science, these tools are not replacing traditional approaches; rather, they are enhancing existing capabilities by allowing researchers to explore greater datasets, simulate complicated connections, and develop robust prediction models [14].

2.1 AI, ML, and DL: A Foundational Framework

Machine learning is based on algorithms that improve with experience. In pharmaceuticals, machine learning models can be developed to predict drug solubility, optimize excipient ratios, and even forecast in vivo pharmacokinetic characteristics. Deep learning, a more complex branch of ML, leverages layered neural networks to comprehend subtle patterns in large-scale, high-dimensional data. It is especially effective in image recognition and time-series modeling [14].

2.2 Conventional Modelling Architectures in Pharmaceutical Sciences

Several algorithmic approaches have been effectively implemented across pharmaceutical applications:

- Random Forests (RF) are effective in predicting medication release or nanoparticle size due to their ability to handle noisy and multi collinear data.
- Support Vector Machines (SVM) are effective at binary classification tasks such as drug-excipient compatibility.
- Convolutional Neural Networks (CNNs): Initially developed for image analysis, CNNs are currently employed for visual quality control in tablet manufacturing.
- Recurrent Neural Networks (RNNs) are effective for modelling sequential data, such as medication release profiles across time [15].

MANUSCRIPT ID: BTS0926007

- Bayesian Networks are probabilistic models that use prior knowledge and uncertainty to make decisions under restricted data conditions [16].

Although each of these models has unique benefits, the calibre, applicability, and organization of the input data have a significant impact on how well they function.

2.3 Data: The Engine behind Prediction

The data received by AI ultimately limits or empowers its usefulness in pharmaceuticals. Today, formulation scientists have access to an increasing number of data sources, which include:

- Physicochemical drug properties: Platforms such as Pub Chem and Drug Bank provide crucial model inputs such as logP, solubility, pKa, and molecular weight.
- Excipient compatibility databases: Datasets capturing hygroscopicity, compressibility, or chemical reactivity enable AI to forecast formulation concerns before bench testing begins [17].
- DoE outputs: These structured datasets, while historically employed for optimization, can now be used to train models that generalize beyond predetermined experimental conditions [18] [19].

However, issues still exist. The fragmentation, proprietary nature, and lack of standardization of many datasets might hinder the reproducibility and transferability of models. Moreover, models that are trained on sparse or noisy data are still susceptible to over fitting. The answer is found in

- cross-validation,
- data curation, and a growing interest in explainable AI (XAI), which allows for clear interpretation of model predictions [20].

2.4 Along the path to smart pharmaceuticals



MANUSCRIPT ID: BTS0926007

Although it is still in its early phases, the direction of AI integration in formulation science is obvious. We may anticipate that intelligent models will play a crucial role in formulation design as data infrastructure advances and regulatory environments change. The questions that AI is uniquely positioned to solve are why it works and what will work next, rather than just what works [21].

AI is making pharmaceuticals a more proactive, data-driven science by facilitating deeper mechanistic insight, faster iteration, and prediction accuracy. Not only is the future smarter and faster, but it is also interpretable, predictive, and optimized from molecule to market [22].

2. AI FRONTIERS IN THE ART OF FORMULATION

In the field of pharmaceutical sciences, artificial intelligence (AI) is no longer a sci-fi idea; rather, it is a useful and indispensable instrument that is revolutionizing the design, optimization, and personalisation of medication formulations. AI is simplifying formulation procedures, cutting costs, and speeding up development for everything from early preformulation insights to sophisticated delivery platforms and patient-specific treatments. We examine its primary application areas below, arranged by product type and formulation function.

1. Excipient Optimization and Pre-formulation Strategies

Selecting the appropriate excipients is essential to guaranteeing the stability, solubility, and manufacturing viability of drugs. Iterative testing and a lot of lab work were formerly needed for this process. These days, drug–excipient compatibility may be quickly predicted by AI models trained on physicochemical descriptors, chemical fingerprints, and historical formulation data [22].

In identifying stable vs. unstable pairings, machine learning classifiers such as

Support Vector Machines (SVMs) and Random Forests have demonstrated an accuracy of over 90%. Prediction accuracy is further increased by combining AI-enhanced molecular docking tools with Quantitative Structure–Activity Relationship (QSAR) models, particularly for ionisable or reactive APIs [23].

3.2 Predicting Solubility and Bioavailability

SMILES strings and molecular descriptors are currently used by deep learning models to predict aqueous solubility, surpassing the capabilities of conventional regression-based methods [24]. Additionally, AI-assisted BCS classification methods provide a high-throughput method of screening poorly soluble drugs by forecasting bioavailability based on permeability, logP, and molecular size [25].

3.3 Optimization of Formulation

Design of Experiments (DoE), although useful, may overlook non-linear interactions when optimizing API:excipient ratios or processing parameters. Traditional DoE is improved by AI models like Genetic Algorithms, Artificial Neural Networks (ANNs), and XGBoost, which explore larger design spaces and adjust formulation factors in response to predicted feedback [26].

A case study from Insilico Medicine showed how 40% fewer experimental cycles might be achieved by using deep generative models to suggest the best formulation conditions for a targeted cancer treatment. By enabling virtual prototyping, these solutions improve time-to-market and reduce material waste [27].

3.4 Predicting Stability and Shelf-Life

It is both commercially and legally required to forecast a product's stability under different storage conditions. AI-powered models are now able to predict shelf-life, degradation kinetics, and moisture sensitivity depending on environmental factors including

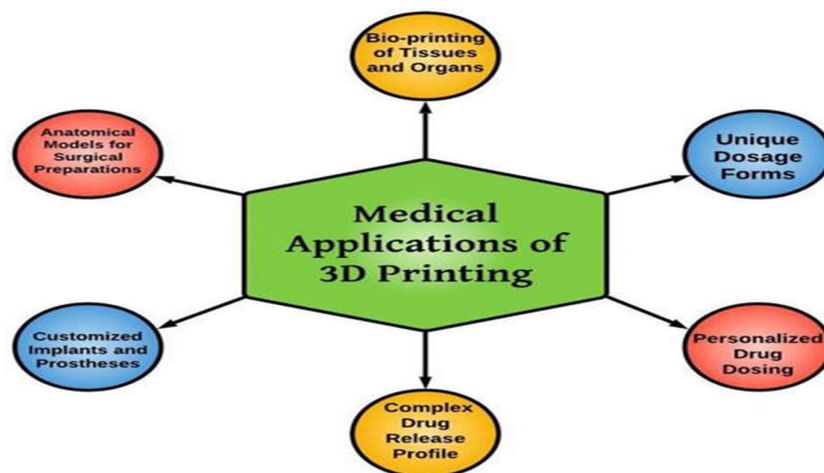
MANUSCRIPT ID: BTS0926007

light exposure, humidity, and temperature [28].

Long-term degradation of small molecules and biologics has been successfully estimated using random forest regressors and

release curves, including burst and plateau phases. These models obviate the need for significant in vitro testing and are currently being verified for QbD submissions [30].

3.6 Personalized Medicine & 3D Printing



Bayesian inference models. Additionally, AI facilitates accelerated stability modelling, which satisfies ICH Q1A(R2) compliance requirements while minimising the need for drawn-out real-time investigations [29].

3.5 Drug Release Modelling

Neural networks trained on formulation and process characteristics correctly predict

AI is also driving 3D-printed drug products, with algorithms optimizing layer-by-layer deposition depending on drug release needs. AI is being used by companies such as FabRx and Exscientia to create personalized dosage forms that combine pharmacogenomics and formulation science [31].

Fig: 1 Medical Applications of 3D Printing

3.7 Practical Uses for Cases

Insilico Medicine decreased development time by using AI to design and optimize the formulation of a lead chemical.

In order to achieve the desired PK profiles with the least amount of iteration, Exscientia used reinforcement learning to suggest formulation changes for a CNS drug candidate.

AI has also been used to predict solid dispersion crystallisation tendencies, liposomal encapsulation efficiency, and nano emulsion behavior [32].

4. AVAILABLE TOOLS, DATABASES, AND FRAMEWORKS

The availability of trustworthy instruments, well selected datasets, and easily accessible computational frameworks are essential for the effective use of artificial intelligence (AI) in pharmaceutical formulation science. With both open-source and proprietary resources available, the environment has grown significantly over the last five years, allowing researchers to use AI at different phases of the formulation lifecycle. Even though these technologies have sped up development, there are still issues that need to be carefully considered in order to integrate them into pharmaceutical workflows.



4.1 Platforms and Computational Frameworks

Numerous AI systems have become well-known due to their adaptability, scalability, and simplicity of integration:

Tensor Flow and PyTorch have emerged as key components for developing custom deep learning models, especially for processing spectral data, predicting solubility, and modelling drug release dynamics. They are widely used in academic research and early-stage industry development due to their open-source nature and active developer communities [33].

Based on Tensor Flow, DeepChem provides a domain-specific library for bioactivity prediction and cheminformatics. Because it has capabilities for docking, QSAR modelling, and molecular featurization, it is ideal for problems involving predictive formulation [34].

There are automated machine learning (AutoML) pipelines offered by Google AutoML, H2O.ai, and IBM Watson Studio. These platforms enable formulation scientists to train, test, and deploy models with little programming experience, reducing the requirement for deep coding expertise. Watson has been used in formulation risk analysis and predictive toxicology [35]. ChemAxon's Marvin and JChem offer commercial cheminformatics solutions, such as drug–excipient interaction analysis and property prediction (e.g., logP, pKa). Access is frequently restricted to enterprise users, despite its robustness [35].

4.2 Formulation and Molecular Properties Databases

The foundation of predictive modelling continues to be high-quality data. AI development in pharmaceuticals is currently supported by a variety of public and proprietary databases:

Target bioactivities, pharmacokinetic information, and chemical structures are publicly available through PubChem,

MANUSCRIPT ID: BTS0926007

ChEMBL, and DrugBank. When creating molecular descriptors for formulation prediction, these resources are frequently utilized [36].

FDA's Inactive Ingredient Database (IID) and Excipient Knowledge Base (ExCIPIENTdb) provide information on safety profiles, regulatory restrictions, and excipient compatibility. Better excipient risk profiling has been made possible by the integration of these databases into ML procedures [37].

Because of their bulk, consistency, and experimental rigour, proprietary industrial datasets, such those kept up to date by Pfizer, Novartis, and AstraZeneca, provide unmatched value. However, unless cooperative agreements are struck, their limited accessibility limits cross-industry innovation [38].

4.3 Actual Case Studies: Adoption in Academic and Industrial Settings

A number of recent case studies demonstrate how AI is influencing formulation in the real world:

Novartis developed a fast-dissolving tablet formulation with 40% fewer tests by combining Bayesian optimisation with AI-enhanced DoE tools [39].

Pfizer greatly shortened development durations by using machine learning to forecast long-term drug product stability from early accelerated deterioration data [40].

Using graph neural networks, researchers at the University of Cambridge combined DeepChem and ExCIPIENTdb to forecast drug loading capacity and nanoparticle stability [41].

These examples demonstrate the versatility of AI tools as well as their capacity to minimise resource usage, shorten time-to-market, and produce formulation techniques that were previously unattainable through conventional means [42].

5. CHALLENGES AND LIMITATIONS

MANUSCRIPT ID: BTS0926007

Artificial intelligence (AI) has made encouraging advances in pharmaceutical formulation, but there are still several obstacles in the way of its broad use. If these issues are not resolved, AI-driven approaches

may not be as scalable, transparent, or accepted by regulators. As the area develops, responsible innovation and long-term integration require a critical assessment of existing constraints.

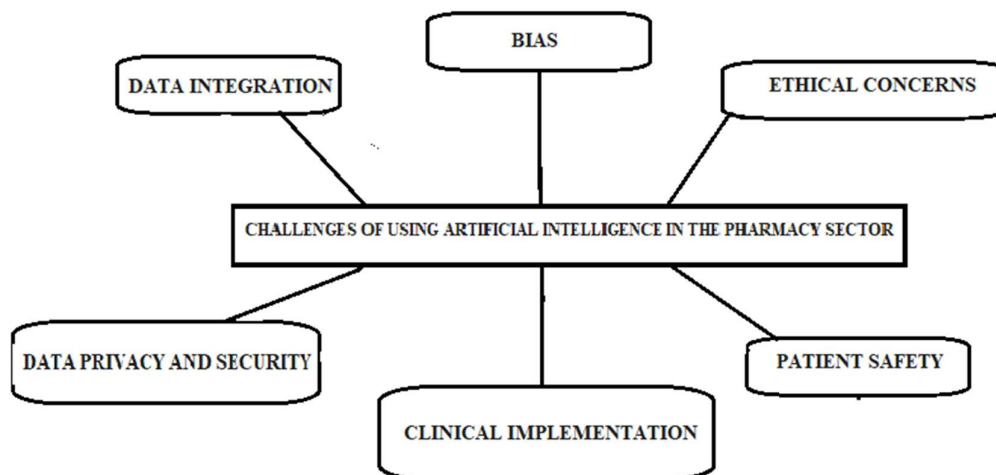


Fig: 2 Challenges of using Artificial Intelligence in the Pharmacy Sector

5.1 Availability, Curation, and Quality of Data

The quality of the data used to train AI models determines their dependability. Data inconsistency and fragmentation continue to be major problems in pharmaceuticals. Many formulation datasets are unsuitable for generalised predictive modeling because they are obtained from small-scale, proprietary studies, lack standardised formats, or are housed in isolated silos [43]. Interoperability of datasets is further complicated by differences in excipient nomenclature, measurement units, and experimental design. The pharmaceutical industry is currently lagging behind industries like finance and e-commerce in terms of using large-scale, high-quality data, despite initiatives like FAIR (Findable, Accessible, Interoperable, Reusable) data principles having increased access and structure [44].

5.2 Interpretability and the Problem of the Black Box

The decision-making processes of AI models get more difficult to understand as they become more complicated, particularly deep learning architectures like neural networks and ensemble approaches. In pharmaceuticals, where comprehending the reasoning behind a model's forecast is frequently just as crucial as the prediction itself, this "black box" characteristic presents a significant obstacle [45]. Explainability is necessary for regulatory agencies like the FDA and EMA to guarantee scientific validity and reproducibility. Although some transparency is provided by tools like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations), few of these have been extensively used in formulation contexts [46]. Therefore, in order to make them more



palatable to regulators and formulators alike, future research must concentrate not only on predicting accuracy but also on developing explainable AI (XAI) frameworks that may provide mechanistic insights.

5.3 Barriers to Regulation and Validation

For a good reason, the pharmaceutical business is subject to strict regulations. Nevertheless, the regulatory framework in place now is insufficiently prepared to assess or authorise AI-based formulation models. There are still issues with assessing risk in accordance with ICH Q8–Q14 guidelines, ensuring model resilience across batches, and validating machine learning predictions [47].

The FDA's 2021 discussion paper on AI/ML in healthcare is one example of an agency that has released preliminary guidance on AI in drug development. However, specific standards for AI in formulation science remain elusive [48]. In the absence of clear regulatory pathways, companies may be reluctant to adopt AI, even when it offers quantifiable benefits.

Model auditing procedures, AI-specific GMP rules, and regulatory sandboxes must be created in order to close this gap. It will be crucial to involve regulators early in the AI design process.

5.4 Integration with Conventional Processes

Integration into current formulation operations is a major operational problem, even when models are accurate and interpretable. The majority of pharmaceutical R&D labs are built around traditional statistical techniques or run on outdated platforms. AI integration necessitates cross-

disciplinary collaboration, staff retraining, and frequently, data infrastructure renovation [49].

5.5 The Path Ahead: From Obstacle to Possibility

The future of AI in formulation science is still very bright, despite these obstacles. Every difficulty listed above is an opportunity to innovate ethically rather than a hindrance. AI is set to evolve from a mystery adjunct to a trusted partner in formulation as explainability tools advance, data ecosystems improve, and regulators get more involved.

Future formulation pipelines will be smarter, safer, and faster thanks to hybrid models, which combine machine intelligence and human knowledge. By foreseeing and tackling these issues today, we can open the door to a time when formulation design is a transparent, data-driven, and predictive science rather than empirical hunch work.

6. FUTURE PERSPECTIVES

Pharmaceutical formulation has already started to change as a result of artificial intelligence (AI), but its full revolutionary potential will be realized when it works in tandem with next-generation technologies. We can imagine a time when formulation creation is quicker, more intelligent, and more flexible than it has ever been as AI develops further. The combination of artificial intelligence (AI), robots, digital twins, and federated learning will revolutionize not only medication formulation but also pharmaceuticals collaboration, regulation, and innovation.

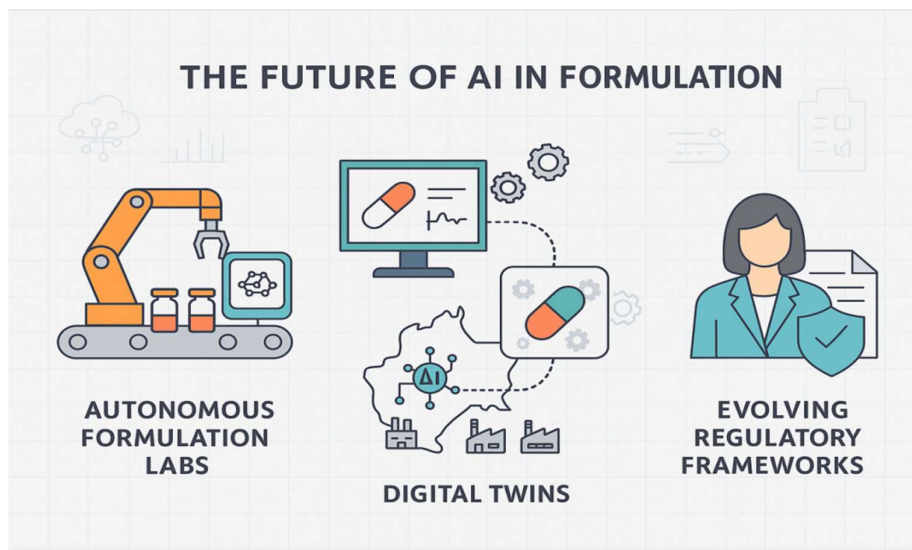


Fig: 3 The future of AI in formulation

6.1 AI and Self-Sustained Formulation Labs

Self-driving formulation laboratories automated systems that can plan, carry out, analyse, and improve experiments without human assistance are becoming possible thanks to the combination of AI and robotics. Based on experimental input, these self-sufficient labs iteratively enhance formulation strategies in real time using reinforcement learning algorithms.

Robotic platforms that autonomously test excipient combinations and optimise release kinetics in less than a week have been proven by businesses like AstraZeneca and university laboratories like the University of Toronto's Acceleration Consortium. These labs maximise reproducibility, minimise human error, and save time and money [50].

6.2 Formulation Science and Digital Twins

In pharmaceutical production, the idea of a digital twin a virtual version of a physical system has gained traction and is currently being applied to formulation research and development. The behaviour of a pharmaceutical product during manufacturing, packaging, storage, and even patient usage can be replicated using a digital twin of the formulation.

Consistent batch quality is ensured by AI-driven digital twins, which allow for real-time optimisation and predictive control of critical quality attributes (CQAs). Additionally, they function as training grounds for novel formulations, allowing variables to be virtually stress-tested before to the start of actual trials [51].

6.3 Federated Education for Safe, Joint Innovation

AI is driven by data, but data silos are still a problem, particularly when confidentiality prevents organisations from sharing secret formulation data. By enabling AI models to be trained across several universities



without sharing raw data, federated learning provides a potent remedy.

Federated architectures that allow collaborative model building while maintaining IP security are being piloted by pharmaceutical consortia and research institutes. Without sacrificing legal or commercial restrictions, this strategy might make it possible to acquire large, varied datasets that support more accurate and broadly applicable formulation models [52].

6.4 Developing Legal Structures for AI-Powered Formulations

The current regulatory framework was not designed to evaluate AI and other dynamic, learning technologies. Because of this, regulatory agencies are currently creating frameworks that support digital tools, adaptive AI models, and data-centric formulation design.

The FDA's "AI/ML-Based Software as a Medical Device" and ICH Q14's focus on model-informed development⁴ are two examples of the guidelines that the U.S. FDA, EMA, and ICH have started investigating to facilitate AI integration.

In order to guarantee trustworthy AI and promote its wider application in formulation design, proactive cooperation between developers and regulators is essential [52].

6.5 Moving Towards Patient-Centric, Predictive, and Adaptive Formulations

The formulation scientist of the future will not be dependent solely on empirical intuition, as AI takes the lead. Instead, they will work together to develop smarter therapies that are faster, safer, and customised for each patient using robotic platforms, digital twins, and algorithms [53].

7. CONCLUSION

Artificial intelligence (AI) has quickly advanced from a theoretical concept to a real and transformational force in pharmaceutical formulation research. By incorporating machine learning algorithms, deep learning architectures, and data-driven modelling

MANUSCRIPT ID: BTS0926007

into the formulation process, researchers and developers can now forecast excipient compatibility, optimise dosage forms, and simulate release kinetics with unprecedented precision and speed.

Over the last five years, tremendous progress has been made in using AI to improve classical methodologies like Design of Experiments (DoE) and Quality by Design. AI systems use high-dimensional data to identify complicated patterns, shorten formulation cycles, and reduce material waste and cost. Case studies from academia and business, including platforms developed by Novartis, AstraZeneca, and university AI labs, show the actual benefits of AI in real-world product development.

This technical advancement is not without difficulties, though. Wider deployment is still hampered by data fragmentation, insufficient model interpretability, and a lack of regulatory frameworks specifically designed for adaptive AI solutions. Furthermore, as regulatory bodies like the FDA and EMA start to create guidelines for AI-driven development, the pharmaceutical industry needs to move from static modelling to dynamic systems that learn, adapt, and explain their predictions.

The future of predictive formulation science is unquestionably bright, notwithstanding these obstacles. Digital twins, federated learning, and robotics combined with AI will allow for autonomous formulation laboratories that can innovate continuously. In addition to expediting development, these technologies are expected to help achieve the long-term objective of tailored and flexible medication distribution. In the future decade, formulators will collaborate with intelligent systems to co-design safer, faster, and more effective medicines, rather than relying exclusively on empirical data. AI will not replace formulation scientists, but rather redefine their roles, allowing them to ask better questions,



interpret deeper findings, and push the limits of pharmaceutical innovation.

REFERENCES

1. Jiao Y, Yang L, Wang R, Song G, Fu J, Wang J, Gao N, Wang H. Drug delivery across the blood-brain barrier: a new strategy for the treatment of neurological diseases. *Pharmaceutics*. 2024 Dec 19;16(12):1611.
2. Tewabe A, Abate A, Tamrie M, Seyfu A, Abdela Siraj E. Targeted drug delivery – from magic bullet to nanomedicine: principles, challenges, and future perspectives. *J Multidiscip Healthc*. 2021 Jul 5;14:1711–24. doi:10.2147/JMDH.S313968.
3. Malheiro A, Vale N. The potential of artificial intelligence in pharmaceutical innovation: from drug discovery to clinical trials. *Pharmaceutics*. 2025;18(6):788.
4. Pandey BS, Durgapal S, Kumar P, et al. Unraveling the artificial intelligence role in drug discovery and pharmaceutical product design: an opportunity and challenges. *Discov Artif Intell*. 2025;5:72. doi:10.1007/s44163-025-00330-9.
5. Bao Z, Bufton J, Hickman RJ, Aspuru-Guzik A, Bannigan P, Allen C. Revolutionizing drug formulation development: the increasing impact of machine learning. *Adv Drug Deliv Rev*. 2023 Nov;202:115108. doi:10.1016/j.addr.2023.115108. Epub 2023 Sep 27. PMID: 37774977.
6. Bannigan P, Aldeghi M, Bao Z, Häse F, Aspuru-Guzik A, Allen C. Machine learning directed drug formulation development. *Adv Drug Deliv Rev*. 2021 Aug;175:113806. doi:10.1016/j.addr.2021.05.016.
7. Dorsey PJ, Lau CL, Chang TC, Dorschuk PC, D'Addio SM. Review of machine learning for lipid nanoparticle formulation and process development. *J Pharm Sci*. 2024 Dec;113(12):3413–33. doi:10.1016/j.xphs.2024.09.015. Epub 2024 Sep 27. PMID: 39341497.
8. Kumar R, et al. Machine learning-aided design of experiments for pharmaceutical formulation optimization. *Int J Pharm*. 2024;635:122795.
9. Liu X, et al. Machine learning approaches to predict drug–excipient compatibility. *Int J Pharm*. 2024;632:122604.
10. Suresh B, et al. Emerging AI technologies in solid dosage forms. *Pharmaceutics*. 2024;16(11):2257.
11. Patel D, et al. Exploring AI and ML in pharmaceutical formulation design. *Int J Nat Rem Pharm Heal*. 2024;4(2):67–75.
12. U.S. Food and Drug Administration (FDA). Artificial intelligence and machine learning in software as a medical device. U.S. Food & Drug Administration; 2024.
13. Carragher N, et al. Digital twins in drug product development: current landscape and future outlook. *Adv Drug Deliv Rev*. 2023;197:114028.
14. Holzinger A, et al. Machine learning and knowledge extraction in digital health. *Nat Rev Drug Discov*. 2021;20(6):353–69.
15. Liao J, et al. Predicting nanoparticle characteristics using machine learning: a formulation dataset approach. *Pharmaceutics*. 2023;15(6):1624.
16. Liu X, et al. Machine learning approaches to predict drug–excipient

MANUSCRIPT ID: BTS0926007

Epub 2021 May 19. PMID: 34019959.



- compatibility using Mol2Vec descriptors. *Int J Pharm.* 2024;632:122604.
17. Sato Y, et al. Deep learning-based tablet defect detection using convolutional neural networks. *J Pharm Sci.* 2024;113(2):135–42.
 18. Zhang T, et al. Deep learning for prediction of in vivo pharmacokinetics from dissolution data. *J Control Release.* 2023;357:483–92.
 19. Mohamed N, et al. Bayesian networks in pharmaceutical development: a case study for uncertainty modeling. *Comput Struct Biotechnol J.* 2022;20:987–95.
 20. Banerjee P, et al. Chemoinformatics tools for drug–excipient interaction modeling. *J Chem Inf Model.* 2023;63(11):3051–62.
 21. Rao V, et al. AI-assisted design of sustained-release tablets using ensemble modeling and genetic algorithms. *Pharmaceutics.* 2023;15(10):2190.
 22. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst.* 2017;30.
 23. Oberoi R, Dogra R. AI-enabled prediction of drug–excipient compatibility: a systematic review. *Int J Pharm.* 2023;635:122857.
 24. Banerjee P, et al. Chemoinformatics tools for drug–excipient interaction modeling. *J Chem Inf Model.* 2023;63(11):3051–62.
 25. Chen B, et al. Machine learning models for solubility prediction using chemical descriptors. *Mol Pharm.* 2023;20(1):112–21.
 26. Zhang Y, et al. Predicting BCS classification with machine learning and in silico descriptors. *AAPS J.* 2022;24(3):57..
 27. Rao V, et al. AI-assisted design of sustained-release tablets using ensemble modeling and genetic algorithms. *Pharmaceutics.* 2023;15(10):2190.
 28. Zhavoronkov A, et al. Artificial intelligence for pharmaceutical discovery and development. *Nat Biotechnol.* 2022;40(10):1343–57.
 29. Mohamed N, et al. AI in accelerated stability studies: modeling degradation kinetics. *Eur J Pharm Sci.* 2022;176:106242.
 30. Deshmukh V, et al. ANN-based modeling of drug release from hydrophilic matrices. *AAPS PharmSciTech.* 2023;24(3):89.
 31. Sandler N, et al. AI and 3D printing in personalized medicine: case studies and future perspectives. *Adv Drug Deliv Rev.* 2021;176:113857.
 32. Liao J, et al. Predicting nanoparticle characteristics using machine learning: a formulation dataset approach. *Pharmaceutics.* 2023;15(6):1624.
 33. Abadi M, et al. TensorFlow: a system for large-scale machine learning. *Proc 12th USENIX Symp Operating Systems Design and Implementation.* 2016;265–83.
 34. Ramsundar B, et al. DeepChem: democratizing deep learning for the life sciences. *J Cheminform.* 2019;11(1):74.
 35. Ghassemi M, et al. IBM Watson in drug development: capabilities and limitations. *AI Health.* 2021;3(2):45–51.
 36. ChemAxon. ChemAxon Suite [Internet]. Available from: <https://chemaxon.com/products>
 37. Kim S, et al. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.* 2021;49(D1):D1388–95.

MANUSCRIPT ID: BTS0926007



- MANUSCRIPT ID: BTS0926007**
38. Kuentz M. Excipients and beyond: a database approach for predictive formulation. *Eur J Pharm Biopharm.* 2022;178:150–62.
 39. Makhija SN, et al. Industry perspectives on integrating machine learning into pharmaceutical R&D. *Pharm Ind.* 2023;85(9):882–92.
 40. Müller C, et al. Accelerating formulation development with AI: a Novartis case study. *Pharmaceutics.* 2023;15(11):2434.
 41. Singh N, et al. Predicting pharmaceutical stability using machine learning. *Int J Pharm.* 2022;624:121865.
 42. Li J, et al. Graph neural networks for nanoparticle formulation prediction. *J Pharm Sci.* 2024;113(3):187–95.
 43. Farooqi MA, et al. Platforms for machine learning in pharmaceutical R&D: an industrial perspective. *Pharm Ind.* 2023;85(12):1156–62.
 44. Wilkinson MD, et al. The FAIR guiding principles for scientific data management and stewardship. *Sci Data.* 2016;3:160018.
 45. Holzinger A, et al. Explainable AI for medicine: a practical guide. *Comput Methods Programs Biomed.* 2021;201:105993.
 46. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst.* 2017;30:4765–74.
 47. International Council for Harmonisation (ICH). ICH Q14: Analytical procedure development. International Council for Harmonisation; 2023.
 48. U.S. Food and Drug Administration (FDA). Artificial intelligence and machine learning in software as a medical device: discussion paper. 2021.
 49. Sarker IH. Machine learning: algorithms, real-world applications and research directions. *SN Comput Sci.* 2021;2:160.
 50. MacLeod BP, et al. Self-driving laboratory for accelerated discovery of thin-film materials. *Sci Adv.* 2020;6(20):eaaz8867.
 51. Rossi S, et al. Digital twins in pharmaceutical manufacturing: from vision to reality. *Int J Pharm.* 2022;627:122254.
 52. Xu Z, et al. Federated learning for healthcare: opportunities and challenges. *IEEE Intell Syst.* 2021;36(2):20–30.
 53. U.S. Food and Drug Administration. Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD): Action Plan. 2021. Available from: <https://www.fda.gov/media/145022/download>