Natural Resources for Human Health
eISSN: 2583-1194

Review Article

Received 08 July 2026
Revised 07 Aug 2026
Accepted 20 Aug 2026

Natural Resources for Human Health 2026, Vol. 6 (18s): 1158–1169

KEYWORDS:
artificial intelligence; machine learning; pharmaceutical formulation; drug delivery; quality by design.

AI-Assisted Formulation Development: From Trial-and-Error to Predictive Drug Delivery

Md. Khalid Anwer[*1], and Ramadan Alshdefat[2]

[1]Department of Pharmaceutics, College of Pharmacy, Prince Sattam Bin Abdulaziz University, 11942, Al-kharj, Saud Arabia.
[2]Department of Pharmaceutical Sciences, Faculty of Pharmacy, Jadara University-733, Irbid 2111, Jordan

ABSTRACT Conventional development depends on experiment design, mechanistic comprehension, and repeated experimentation, but as formulation complexity rises, it becomes more challenging to explore the search space. To forecast material properties, drug–excipient compatibility, dissolution, stability, particle attributes, drug loading, release kinetics, pharmacokinetics, and process behavior, artificial intelligence (AI), machine learning (ML), Bayesian optimization, deep learning, generative models, and hybrid mechanistic–data-driven systems are being introduced. The stronger assertion that AI has typically replaced conventional development or consistently produced transferable superiority is not yet supported by the literature, but it does support a significant change from unstructured trial-and-error toward model-guided experimentation. Narrative reviews, retrospective models, tiny private datasets, and application-specific demonstrations continue to make up most of the published evidence. In order to assess whether AI-assisted formulation becomes predictive rather than merely descriptive, this review looks at the computational techniques, formulation domains, data needs, validation standards, regulatory consequences, and translational impediments. A workable implementation architecture is put out in which AI enhances rather than replaces quality by design (QbD), design of experiments (DoE), process analytical technology (PAT), mechanistic modeling, and human formulation knowledge.

1. INTRODUCTION

Through the process of formulation development, an active pharmaceutical ingredient (API) and a group of excipients are transformed into a medicinal product with clinically relevant performance and repeatable quality [1]. Molecular, material, process, product, and patient dimensions are all included in the design variables. API solid form, particle-size distribution, excipient identity and concentration, solvent system, pH, mixing and granulation conditions, compression force, coating parameters, particle engineering, device shape, and storage environment are a few examples. Solubility, dissolution, release kinetics, dose uniformity, tablet strength, disintegration, stability, bioavailability, encapsulation efficiency, particle size, surface charge, and pharmacokinetic exposure are all multivariate outputs [2]. These variables are described as a connected design space rather than a series of discrete optimization problems in reviews of AI in pharmaceutical technology. The principal input–output relationships and their associated validation expectations are summarized in Table 1.

Table 1. AI methods and their most defensible formulation roles

Method family	Typical inputs	Typical outputs	Best-supported role	Main validation risk
Tree ensembles and boosting	Material, formulation, and process descriptors	CQAs, compatibility, stability, particle attributes	Nonlinear prediction in small-to-medium tabular datasets [6]	Leakage, confounding by project or supplier, poor extrapolation
Neural and deep learning	Molecular graphs, spectra, images, temporal profiles, high-dimensional descriptors	Solubility, release, PK, particle properties	Complex representations and large compiled datasets [3]	Overfitting and weak robustness on small formulation datasets [4]
Gaussian processes and Bayesian optimization	Experimental design variables plus measured responses	Predicted response, uncertainty, next experiment	Sequential optimization when experiments are expensive [5]	Invalid uncertainty outside the design domain
Generative models	Molecular, formulation, or process representations plus constraints	Candidate molecules, compositions, or structures	Candidate generation and inverse design [2, 7]	Unrealizable or noncompliant candidates
Mechanistic–ML hybrids	Mechanistic states plus formulation and process data	Corrected or constrained CQAs and trajectories	Better extrapolation and scientific interpretability	Incorrect mechanistic assumptions or poorly calibrated residual models

The attraction of AI is therefore not simply computational speed. It is the possibility of learning nonlinear relationships from historical and experimental data, ranking the next experiments by expected information gain, and navigating formulation spaces too large for one-factor-at-a-time experimentation. Across the literature, AI and ML have been proposed for preformulation, drug–excipient compatibility, solubility prediction, stability and shelf-life estimation, oral solid dosage forms, nanoparticles, lipid systems, 3D-printed medicines, controlled release, and pharmacokinetic modeling [2, 3].

But the data is frequently overshadowed by the terminology used in the field. Without offering a prospective comparison, a truly external validation set, or a precise definition of "traditional," many papers claim that AI is speeding up development, cutting expenses, or exceeding conventional techniques. Four claims must be distinguished in a scientific evaluation: (1) AI can fit formulation data; (2) AI can predict held-out data; (3) AI can direct experiments more effectively than a reliable comparator; and (4) AI-guided products maintain their robustness when transferred across APIs, scales, sites, suppliers, and manufacturing conditions. The first two have the best literature, whereas the last two have far less. [2, 3]

This article assesses the translational discipline of AI-assisted formulation development. It questions where predictive value is demonstrated, where it remains conditional, and what evidence is needed for routine usage in pharmaceutical research. It focuses on the chain from data production to model deployment. [2, 3]

2. SCOPE AND METHODS

2.1 Search approach

Using queries covering AI and ML in pharmaceutical formulation, nanoparticle and drug-delivery optimization, QbD and predictive modeling, generative AI, digital twins, and pharmaceutical development standards, a focused in-session literature search was carried out across academic and reputable web sources. Candidate articles were evaluated for source authority, development-method relevance, and direct applicability to formulation or drug delivery development. When accessible, full

texts were retrieved, and pertinent sources were arranged according to formulation domain, computational method, input/output structure, validation approach, key discovery, and limitation. [2, 3]

Alongside primary application research, the resulting evidence base includes a significant number of contemporary narrative reviews. As a result, the review gives primary research that describe specific datasets, model performance, and experimental validation more weight while treating broad claims from narrative reviews as theories about the discipline. The search should not be regarded as comprehensive because it is not a registered systematic review. [2, 3]

2.2 Eligibility logic

At least one of the following topics was covered by the included sources: formulation or preformulation prediction; drug-excipient compatibility; dosage-form or drug-delivery optimization; dissolution, release, or stability prediction; nanoparticle or lipid-system engineering; process or scale-up modeling; or regulatory and QbD implications of AI. Translational modeling principles were only informed by sources that were exclusively focused on target identification, pharmaceutical chemistry, or clinical AI without a formulation or delivery component. [2, 3]

3. FORMULATION DEVELOPMENT AS A PREDICTIVE-DESIGN PROBLEM

3.1 From empirical screening to structured design spaces

Traditional formulation development is not "trial and error" by nature. QbD, DoE, risk assessment, mechanistic investigations, and process understanding are already used in contemporary pharmaceutical development. The key contrast is between unstructured iteration and organized experimentation. AI can learn from previous experiments, recognize nonlinear relationships, assess uncertainty, and suggest experiments within a predetermined design space. It is most useful when it expands an organized development program. When used in place of material characterization or experimental validation, it is least reliable. [2, 3]

The formulation design problem can be expressed as a mapping from inputs x to outputs y. The method families, typical inputs and outputs, and validation risks discussed below are summarized in Table 1 [2, 3]:

$$y = f(x) + \epsilon$$

where y may include one or more critical quality attributes (CQAs) or performance outcomes, x may include API descriptors, excipient identity and concentration, process parameters, environmental conditions, and dosage-form architecture, and ϵ stands for measurement noise, batch variability, and unobserved causes. Developers actually require more than a point prediction. They require proof that the mapping is true when the process or product changes, as well as uncertainty and sensitivity to controlled variables. [2, 3]

Formulation composition, material qualities, process conditions, and environmental factors are frequently identified as model inputs in the literature. Solubility, permeability, stability, degradation rate, shelf life, dissolution, drug release, particle size, surface charge, drug loading, encapsulation effectiveness, bioavailability, tablet hardness, friability, and disintegration are among the reported outcomes [2, 3].

3.2 The role of AI relative to DoE and QbD

AI shouldn't be positioned as DoE's replacement because DoE offers a methodical approach to producing instructive trials and estimating effects within a given area. By sorting potential formulations, learning intricate nonlinearities, or updating a model in response to additional tests, machine learning (ML) can enhance DoE. Because it combines a surrogate model with an acquisition function that strikes a balance between exploring doubtful regions and exploiting promising formulations, Bayesian optimization is especially well-suited to this context.[2, 3]

Reviews discuss how ML is integrated with DoE, PAT, QbD, Bayesian optimization, and computational fluid dynamics for oral solid dose forms [5]. Therefore, the position that can be supported by science is hybrid: QbD is used to establish the quality target product profile and risk structure, DoE or active learning is used to generate informative data, ML is used to model complicated relationships, and experiments are used to test predictions and maintain process understanding.

4. COMPUTATIONAL METHODS USED IN FORMULATION DEVELOPMENT

4.1 Classical supervised learning

Because many development datasets are arranged as paired observations of inputs and measured outcomes, supervised learning continues to be the most popular method in formulation applications. While classification models forecast compatibility classes, stability results, or failure risks, regression models forecast continuous attributes like dissolution time, particle size, hardness, or drug loading. [2, 3]

The literature frequently uses random forests, gradient boosting, support vector machines, k-nearest neighbors, Gaussian-process regression, and artificial neural networks. These techniques are described in conjunction with deep learning, unsupervised learning, and hybrid mechanistic models in reviews of oral solid dose forms [5]. Because they may depict nonlinear interactions without the data requirements of massive neural networks, tree-based models are frequently appealing in small and heterogeneous datasets. Although its processing complexity and kernel assumptions become restrictive as datasets grow, Gaussian processes offer a helpful uncertainty representation for Bayesian optimization. Their most defensible formulation roles and principal validation risks are compared in Table 1.

4.2 Deep learning and representation learning

For molecular, particle, formulation, and pharmacokinetic problems, deep neural networks, convolutional networks, recurrent networks, graph convolutional networks, and neural ordinary differential equations have been employed or suggested. Molecular structure can be directly represented by graph-based models; temporal release or concentration profiles can be represented by recurrent or neural-ODE systems; pictures, spectra, or spatial representations can be processed by convolutional systems. [2, 3]

Not everyone benefits from deep learning's methodological advantages. Large datasets, complicated input representations, and the availability of transfer learning all make it more likely. Deep models may be less reliable than simpler models in limited formulation datasets, particularly when there are more formulation variables than independent batches. According to an examination of drug-delivery applications, LightGBM demonstrated potential while neural networks did badly in one long-acting injectable example due to the tiny dataset. [4].

4.3 Ensemble learning and hybrid models

To lower variance and increase robustness, ensembles mix several learners or model classes. An ensemble achieved an external Solubility Challenge 2019 RMSE of 0.865 in preformulation solubility prediction, compared to an average RMSE of 1.62 for 37 other models, according to a review. The best individual modeling results included an XGBoost-like approach with MAE 0.458, RMSE 0.613, and R^2 0.918 on a held-out split [6]. This is a helpful example because it contains an external benchmark, but it is still a property-prediction task rather than evidence that an AI-guided formulation enhanced a manufacturing or therapeutic outcome.

Statistical learning and mechanical limitations are combined in hybrid models. ML models based on mass-transfer relationships, PBPK, process models, dissolution physics, or material balances are a few examples. The reasoning is compelling: ML can identify residual nonlinearities that are challenging to define analytically, while mechanistic structure can limit extrapolation and enhance interpretability. Whether the hybrid model performs better under distribution shift—rather than just whether it matches the training data better—is the crucial validation question. [2, 3]

4.4 Bayesian optimization and active learning

Formulation development is treated as sequential decision-making in Bayesian optimization. An acquisition function chooses the subsequent experiment, while a surrogate model forecasts the response and uncertainty among potential formulations. When each experiment is costly, time-consuming, or material-limited, this can cut down on the number of tests. By choosing the subsequent observations that are anticipated to enhance the model or eliminate uncertainty, active learning adheres to a similar approach. [2, 3]

A restricted design area with a well-defined goal, trustworthy measurements, and a feedback loop from experiment to model is the best use case. When the model is requested to extrapolate well beyond the observable domain, the target is not well defined, or the data are skewed or noisy, the approach becomes less dependable. A candidate may require acceptable solubility, stability,

manufacturability, content homogeneity, and patient acceptance all at once, hence the acquisition function for pharmaceutical development should frequently be multi-objective. [2, 3]

4.5 Generative AI and inverse formulation design

Instead of just rating current choices, generative models try to suggest new compounds, excipient combinations, particle designs, or process parameters. Formulation reviews include diffusion techniques, generative adversarial networks, variational autoencoders, and structure-synthesis systems [2, 3]. Particle engineering and in silico formulation optimization utilizing a generative structure-synthesis approach are described in a generative AI study. [7].

Validity, utility, and realizability are the three main requirements for generative formulation design. A created candidate must meet physical and chemical requirements, accomplish the desired product performance, and be able to be produced with the tools and resources at hand. Constraint checking, uncertainty estimation, experimental design, and process feasibility filters should all be integrated with generative models. A design ideation tool, not a predictive development system, is a model that produces feasible compositions but is unable to take into consideration excipient availability, regulatory status, scale-up, or stability. [2, 3]

4.6 Explainable AI

Because formulation scientists must comprehend why a model suggests a composition or process condition, interpretability is important. Although they can be useful, feature importance, partial dependency, SHAP, LIME, monotonicity requirements, sensitivity analysis, and counterfactual explanations are not the same as causal knowledge. When a feature is associated with an unmeasured variable, a supplier, a batch, or a measurement convention unique to a publication, it may be predictive. [2, 3]

Regulatory confidence is hampered by black-box behavior, restricted interpretability, data provenance, and reproducibility, according to reviews [2, 8]. Therefore, explainability should be combined with experimental perturbation: if a model states that polymer concentration drives release, then altering polymer concentration while keeping other variables constant should result in the anticipated directional effect. Compared to a feature-importance plot alone, this combination of statistical explanation and experimental investigation is more convincing.

5. APPLICATION DOMAINS

5.1 Preformulation and molecular-property prediction

Because many outputs are quantifiable, repeatable across projects, and connected to chemical or material descriptions, preformulation is an ideal starting point for AI. Aqueous solubility, partition coefficient, polymorphism, crystallization, thermal behavior, hygroscopicity, and drug-excipient compatibility are all pertinent responsibilities. Molecular graphs, SMILES, electrostatic-potential maps, molecular volume, surface area, and hundreds of molecular descriptors are reported as inputs; solubility, logP/logKow, polymorph or solvate formation, and compatibility scores are reported as outputs. [6].

When a property prediction can focus experiments on a smaller area of the space or reject obviously inappropriate candidates early, the translational value is significant. However, formulation prediction and molecular-property prediction are not the same. A projected polymorph does not ensure manufacturability; compatibility in a binary mixture does not provide stability in a multicomponent product; and solubility in a specified medium does not determine dissolution from a tablet. Therefore, unless they are linked to dosage-form performance data, models should be used as upstream filters and prioritizing tools. [2, 3]

5.2 Oral solid dosage forms

One of the most developed application areas is oral solid dosage forms. Formulation composition, granulation, compression, coating, dissolution, stability, and continuous manufacturing are all areas where machine learning has been used. Particle-size distribution, moisture, compression force, mixing time, impeller speed, binder addition rate, drying temperature, spray rate, atomization pressure, and pan speed are among the inputs. Hardness, friability, disintegration, dissolution, flow, granule-size distribution, blend homogeneity, coating thickness, and stability are among the outputs. [5].

Compressibility, compactibility, and manufacturability factors were used in a main study to predict wet-granulation scale-up [9]. Another study predicted the dissolution of bicalutamide from powder systems using data-driven modeling [10]. ML prediction

of drug-release patterns and kinetic parameters based on tablet formulation is reported in a subsequent work [11]. When taken as a whole, these studies show how to go from forecasting a single CQA to modeling temporal performance.

Scale transfer is the primary obstacle. Relationships that vary with equipment geometry, shear history, residence duration, humidity, raw-material variability, or operator practice may be captured by a model trained on laboratory batches. Therefore, scale-up validation has to test additional batches, new equipment, and ideally new campaigns in addition to random held-out samples from the same distribution. The application-specific evidence and its principal translational bottlenecks are summarized in Table 2. [2, 3]

Table 2. Application domains and evidence maturity

Domain	Representative outputs	Evidence maturity in the reviewed set	Translational bottleneck
Preformulation	Solubility, logP, polymorph, compatibility	Strongest for retrospective and external property prediction [6]	Property prediction does not guarantee dosage-form performance
Oral solid dosage forms	Hardness, disintegration, dissolution, granulation, coating	Multiple application studies, but limited cross-site and scale validation [5, 9]	Material and equipment variability during scale-up
Controlled release	Full release profiles and kinetic parameters	Emerging prospective evidence [11-12]	Profile extrapolation and clinically relevant conditions
Nanoparticles and lipid systems	Size, charge, loading, encapsulation, release, uptake	Broad activity with mostly project-specific datasets [13]	Process-context dependence and biological translation
3D printing/personalized dosage	Geometry, dose, disintegration, release	Compiled datasets and proof-of-concept studies [17-19]	Individual-patient validation and software/hardware control
Stability and shelf life	Degradation, shelf life, stability-indicating attributes	Mechanistically plausible but transfer evidence limited [20]	Accelerated-to-real-time extrapolation
Process modeling and digital twins	Process state, PAT outputs, nanosuspension attributes	Early-stage and infrastructure-dependent [21]	Sensor quality, change control, and lifecycle maintenance

5.3 Modified release and controlled delivery

Because the output is a curve rather than a single number, controlled-release systems provide an especially appealing but challenging target for AI. Release rate, lag time, burst release, kinetic parameters, or the complete dissolution profile can all be predicted by models. Polymer type and concentration, drug loading, particle size, matrix architecture, coating thickness, pore-former concentration, and process conditions are examples of inputs. [2, 3]

AI applications to drug-release kinetics and controlled-release formulations are described in reviews [3, 12]. Modeling the entire profile has the benefit of revealing trade-offs between early release, sustained release, and dose dumping danger while identifying formulations that match a target temporal behavior. Overfitting poses a risk because a model may be able to replicate the recorded timepoints for a limited range of formulations while failing at therapeutically relevant circumstances, agitation regimes, pH transitions, or storage states.

Clinically relevant situations and profile-level measures should be used in validation. One correlation coefficient is not enough. Absolute profile error, prediction intervals, bootstrap stability, and performance at independent formulations purposefully chosen close to the boundaries of the design space should all be examined by developers. [2, 3]

5.4 Nanoparticles and nanocarriers

Lipid or polymer composition, solvent and mixing conditions, surfactants, ligand density, drug loading, and process scale are all factors in the high-dimensional field of nanoparticle formulation. Particle size distribution, polydispersity, zeta potential, loading, morphology, release, cellular uptake, biodistribution, and effectiveness are among the outputs. Reviews show ML uses in precision nanomedicine, targeted nanocarriers, polymeric nanoparticles, and lipid nanoparticles [13].

The main examples are AI-driven design and optimization of nanoparticle systems, ML-assisted optimization of lipid-based ceftriaxone administration, and ML techniques for PLGA nanoparticles in antiviral delivery [14–16]. Larger gathered datasets can facilitate model-based analysis of formulation and manufacturing variables, as demonstrated by a study of over 900 3D-printed medication delivery systems. [17, 18].

Nanoparticle models have a unique generalizability issue: mixing regime, equipment, scale, raw-material grade, analytical technique, and storage history can cause ostensibly comparable formulations to behave differently. Additionally, several methods with varying biases may be used to measure the same particle-size output. Not only must a model encode the nominal composition but also the measurement technique and process environment. [2, 3]

5.5 Lipid systems and biologic delivery

Sensitivity to lipid identity, molar ratios, ionization, pH, mixing, encapsulation conditions, and biological environment is increased by lipid-based delivery methods. Biological endpoints are further downstream than physicochemical CQAs, but AI can help prioritize lipid combinations, forecast particle characteristics, and simulate release or uptake. Even with great encapsulation efficiency, instability, immunogenicity, tissue distribution, or insufficient intracellular release might lead a formulation to fail. [2, 3]

Thus, a tiered architecture—predicting physicochemical characteristics first, testing biological performance, and then assessing pharmacokinetics and efficacy—is supported by the findings. Although it may seem appealing to combine these into a single end-to-end model, doing so calls for considerably larger, harmonized datasets and rigorous causal interpretation. [2, 3]

5.6 3D-printed and personalized dosage forms

Digital design characteristics can be linked to geometry, medication loading, disintegration, release, and patient-specific dose requirements using AI-assisted 3D printing. Reviews detail web-based and fused-deposition modeling techniques that generate dosage forms using formulation stocks and printing parameters [19]. An example of utilizing a sizable gathered dataset to investigate correlations between printing and product attributes is provided by machine learning applied to more than 900 printed medicine delivery devices. [17].

Personalization is the value proposition, but it raises issues with validation. For a patient's dose, physiology, adherence, or comorbidities, a model may be accurate at the population level but unreliable. Validated hardware, material traceability, software control, dose verification, and a clear release strategy are all necessary for deployment. [2, 3]

5.7 Stability and shelf life

Because data accumulates across temperatures, humidity, light exposure, packaging, formulations, and time, stability modeling is a good fit for machine learning. Environmental factors, degradation profiles, excipient interactions, packing materials, and process variables are examples of inputs; degradation rates, routes, stability-indicating metrics, and shelf-life estimates are examples of outputs. [20].

Extrapolation is the crucial problem. Predictions of shelf life frequently rely on kinetic assumptions and accelerated conditions. The scientific foundation for extrapolation to labeled storage settings should not be obscured by machine learning (ML), even though it can detect nonlinearities or interactions. When the underlying process is understood, a hybrid approach that combines deterioration kinetics with ML residual correction is more convincing than a strictly black-box model. [2, 3]

5.8 Process development and digital twins

The goal of AI-assisted process modeling is to forecast how equipment and material circumstances result in product qualities. By keeping a computer representation of a process up to date with real-time data, digital twins expand on this idea. The

application of this idea to pharmaceutical systems is demonstrated by a study on artificially intelligent digital twins for solidified nanosuspensions. [21].

The process model needs to be linked to sensors, data historians, PAT, batch records, and a verified control plan in order for a digital twin to be beneficial. It's not only an algorithmic problem. Sensor calibration, missing data, synchronization, cybersecurity, change control, and preserving model performance when materials and equipment change are all included. [2, 3]

6. DATA FOUNDATIONS

6.1 Data quality is the limiting reagent

Across reviews, data quality, scarcity, heterogeneity, and fragmented ownership are the most consistent barriers [3, 6, 20]. Formulation datasets are often small because each experiment consumes API and labor, and because proprietary development data are rarely published in machine-readable form. Published datasets may also overrepresent successful formulations, omit failed experiments, use inconsistent analytical methods, or mix batches and formulations without preserving lineage.

A model cannot distinguish a real formulation relationship from a laboratory artifact unless the dataset preserves experimental context. Minimum metadata should include API identity and solid form, excipient grade and supplier, batch identifier, equipment, process history, analytical method, units, environmental conditions, and whether the result is a replicate or a new formulation. [2, 3]

6.2 Independent observations and data leakage

Randomly splitting rows is often inappropriate. Replicate measurements from the same batch, formulations from the same DoE run, or near-identical compositions across a single project can leak information from training to test data. More realistic validation schemes hold out entire formulation families, API compounds, suppliers, campaigns, sites, or time periods. If the deployment question is "will this model work for a new API?" then the test set must contain new APIs. If the question is "will it work at commercial scale?" then the test set must contain scale or site shifts. [2, 3]

6.3 Data standardization and ontology

The field needs common representations for excipients, dosage forms, process variables, analytical methods, CQAs, and failure modes. Without standardization, two studies may appear to model the same endpoint while measuring different constructs. A shared ontology should distinguish formulation composition from process conditions, nominal from measured attributes, and batch-level from unit-level observations. [2, 3]

6.4 Uncertainty and out-of-domain detection

A formulation recommendation without uncertainty is incomplete. Prediction intervals, conformal prediction, ensemble dispersion, Gaussian-process variance, and applicability-domain methods can estimate when a model is uncertain. Developers also need out-of-domain detection to identify candidate formulations unlike the training data. An uncertainty estimate should trigger either an additional experiment or a human review—not be treated as a cosmetic confidence score. [2, 3]

7. VALIDATION: FROM FIT TO DEPLOYMENT

7.1 A hierarchy of evidence

AI formulation evidence can be organized into five levels:

1. **Retrospective fit:** the model reproduces historical data.
2. **Internal prediction:** the model predicts held-out observations from the same project or distribution.
3. **External prediction:** the model predicts data from a different dataset, laboratory, or source.
4. **Prospective experimental guidance:** the model selects new formulations that are tested experimentally.
5. **Operational or clinical translation:** the AI-guided product improves development efficiency, robustness, manufacturing performance, or patient-relevant outcomes relative to a credible comparator. [2, 3]

Much of the literature reaches levels 1–2. Some property-prediction examples reach level 3; the external solubility benchmark is one such example [6]. Far fewer studies demonstrate level 4, and level 5 evidence is sparse in the reviewed material. The hierarchy and minimum evidence package for each intended use are presented in **Table 3.**

Table 3. Minimum evidence package by intended use

Intended use	Minimum evidence package	Decision authority
Exploratory prioritization	Internal validation, applicability domain, expert review, confirmatory experiments	Formulation team
CQA prediction	Grouped holdout testing, uncertainty calibration, predefined error threshold, independent batch confirmation	Development and analytical science
Next-experiment selection	Prospective sequential experiment, comparator strategy, experiment count and cost, stopping rule	Formulation lead and project governance
Scale-up support	New equipment, scale, supplier, and campaign testing; transfer acceptance criteria	Process development and manufacturing science
Process control or release support	Validated data pipeline, audit trail, change control, monitoring, human override, regulatory documentation	Quality and regulatory functions

7.2 Performance metrics

Metrics must match the task. Regression should report MAE or RMSE alongside calibration and prediction intervals; classification should report sensitivity, specificity, precision, calibration, and class imbalance; release profiles require profile-level error and time-dependent assessment; optimization requires the number and cost of experiments needed to reach a predefined target. Reporting only R^2 can conceal systematic bias, poor performance at the design-space boundary, or clinically meaningful absolute error. [2, 3]

7.3 Prospective validation and comparator choice

The relevant comparator is not "no intelligence." It may be a standard DoE, a formulation scientist's predefined strategy, a mechanistic model, or a conventional statistical model. A credible prospective study should predefine the target, candidate space, stopping rule, experimental budget, comparator, and success criterion. Otherwise, an AI-guided workflow can appear successful simply because the objective or candidate set was changed after observing results. [2, 3]

7.4 Scale, site, and supplier transfer

A model can be accurate in one laboratory and fail after transfer. Validation should explicitly test changes in API lot, excipient supplier, equipment, scale, operator, analytical method, and storage condition. The more the model is used for control rather than prioritization, the more stringent this testing must be. [2, 3]

8. REGULATORY AND QUALITY CONSIDERATIONS

8.1 AI does not remove pharmaceutical quality obligations

An AI model does not become acceptable because it is accurate on a benchmark. It must be embedded in a pharmaceutical quality system with traceable data, controlled software, documented intended use, validated interfaces, change control, audit trails, and defined human oversight. Reviews consistently identify regulatory acceptance, interpretability, reproducibility, privacy, and bias as unresolved barriers [2, 3].

8.2 Intended use and model lifecycle

Every model should have an intended-use statement specifying the input domain, output, decision supported, acceptable uncertainty, and consequences of error. The lifecycle should include data qualification, training, validation, deployment, monitoring, retraining, and retirement. Retraining must be controlled because an apparently improved model may alter the design space or decision threshold. [2, 3]

8.3 Human oversight

Human oversight should be proportional to risk. A model that ranks excipients for exploratory screening can operate with lighter controls than a model that sets a process control limit or releases a batch. In both cases, the formulation scientist remains responsible for assessing whether the recommendation is chemically, physically, clinically, and operationally plausible. [2, 3]

8.4 Explainability and documentation

Regulatory reviewers need to understand not only the algorithm but the evidence supporting its use. Documentation should include the dataset lineage, exclusions, missing-data handling, feature definitions, model architecture, hyperparameters, validation splits, external tests, uncertainty, failure modes, and change history. Explainable AI can support this documentation but cannot substitute for validation. [2, 3]

9. COMMON FAILURE MODES

9.1 Overfitting and optimistic validation

Small formulation datasets make overfitting easy. High-dimensional descriptors, repeated measurements, and flexible models can produce impressive internal metrics that collapse on external data. Nested cross-validation, grouped splits, and pre-specified external testing are essential. [2, 3]

9.2 Confounding by project, supplier, or method

A model may learn that one supplier or analytical method is associated with a particular outcome rather than learning the formulation mechanism. This problem is especially likely when data are pooled from different projects without harmonization. [2, 3]

9.3 Treating correlation as intervention evidence

A predictive feature is not necessarily a controllable causal lever. If moisture predicts dissolution, changing moisture may not reproduce the same effect if moisture is merely a marker of a different granulation history. Experimental perturbation is needed before using a feature for optimization.[2, 3]

9.4 Extrapolation beyond the design space

AI is often strongest near the data it has seen. Generative systems may propose chemically or operationally implausible formulations, while black-box predictors may produce confident but invalid outputs outside the applicability domain. Constraint handling and uncertainty-aware candidate selection are mandatory. [2, 3]

9.5 Optimizing one CQA at the expense of the product

A formulation with excellent dissolution may have poor stability, unacceptable hardness, difficult scale-up, or inadequate bioavailability. Multi-objective optimization should use a clinically and manufacturably meaningful utility function, not a single laboratory endpoint.[2, 3]

10. A PRACTICAL IMPLEMENTATION FRAMEWORK

Step 1: Define the decision

Specify whether the model will screen materials, predict a CQA, select the next experiment, optimize a formulation, support scale-up, or control a process. Different decisions require different validation levels. [2, 3]

Step 2: Define the quality target and constraints

Translate the quality target product profile into CQAs, acceptable ranges, hard constraints, and soft objectives. Include manufacturability, stability, analytical uncertainty, and patient-use requirements. [2, 3]

Step 3: Audit the data

Assess completeness, batch independence, measurement consistency, failed experiments, missingness, outliers, duplicate records, and data lineage. Do not begin with algorithm selection; begin with whether the data can answer the decision question. [2, 3]

Step 4: Establish a credible baseline

Compare AI models with simple statistical models, DoE, mechanistic predictions, and expert-defined rules where appropriate. A more complex model should earn its complexity through improved external or prospective performance. [2, 3]

Step 5: Use uncertainty-aware modeling

Report prediction intervals and define the applicability domain. Use uncertainty to prioritize experiments and to prevent automated decisions in poorly supported regions.

Step 6: Close the experimental loop

Use active learning or Bayesian optimization only when the model's recommendations are experimentally tested and returned to the dataset. Record not only successful formulations but failures and reasons for exclusion. [2, 3]

Step 7: Validate transfer

Test new batches, materials, equipment, scale, and sites. If transfer is not part of the intended use, state that explicitly rather than implying generalization.

Step 8: Deploy with quality controls

Implement version control, audit trails, model monitoring, change control, retraining rules, and human review. Treat the model as a controlled component of the pharmaceutical development system.

11. CONCLUSIONS

AI-assisted formulation development is best understood as a progression from retrospective prediction toward closed-loop, uncertainty-aware experimentation. The literature demonstrates broad feasibility across preformulation, oral solid dosage forms, controlled release, nanoparticles, lipid systems, 3D printing, stability, and process modeling. It also shows that performance can be strong for narrowly specified tasks with well-defined data and external benchmarks, including solubility prediction and selected dosage-form or delivery attributes.

The evidence does not yet justify a general claim that AI routinely outperforms conventional formulation development or can replace formulation scientists. The decisive limitations are small and heterogeneous datasets, data leakage, inconsistent analytical methods, weak external validation, uncertain extrapolation, limited prospective evidence, and regulatory requirements for traceability and control. The most defensible development model is hybrid: QbD and DoE define the problem; mechanistic knowledge constrains it; AI learns complex relationships and prioritizes experiments; and laboratory, pilot, and manufacturing data verify whether the predicted design remains valid. AI becomes predictive drug delivery only when the model is validated across the transitions that matter—new formulations, new materials, new scales, new sites, and ultimately patient-relevant performance.

Funding: "The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU- 2024/03/29792)".

Acknowledgments: "The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU- 2024/03/29792)".

Ethics Declaration: Not Applicable.

Data Availability Statement: All the data generated in this study have been included in the manuscript.

Conflicts of Interest: "The authors declare no conflict of interest.".

REFERENCES

[1] Bhelkar N. *Role of Artificial Intelligence in Formulation Design*. 2026.[12]
[2] Thakur P, et al. *Smart Pharmaceutical Formulation and Drug Delivery Using AI*. 2026.[2]
[3] Dey H, et al. *Exploring the Role of Artificial Intelligence and Machine Learning in Pharmaceutical Formulation Design*. 2024.[3]
[4] Dangeti A, et al. *Revolutionizing Drug Formulation: Harnessing Artificial Intelligence and Machine Learning for Enhanced Stability, Formulation Optimization, and Accelerated Development*. 2023.[20]
[5] Harika Harika N, et al. *Prediction of Preformulation Using AI and Machine Learning Approaches*. 2026.[6]
[6] Parvathaneni M, et al. *Application of Artificial Intelligence and Machine Learning in Drug Discovery and Development*. 2023.[19]
[7] Patel A. *Artificial Intelligence in Formulation Design: A New Era in Pharmaceutics*. 2015.[8]
[8] Vora LK, et al. *Artificial Intelligence in Pharmaceutical Technology and Drug Delivery Design*. 2023.[4]
[9] Padhiar JS, et al. *Evolution of Artificial Intelligence in Drug Delivery System Design: Applications, Challenges and Future Perspectives*. 2026.[13]
[10] Alla MR. *A Review of Artificial Intelligence-Driven Predictive Strategies for the Formulation Development of Oral Solid Dosage Forms*. 2026.[5]
[11] Suksaeree J. *A Review of Artificial Intelligence-Driven Smart and Sustainable Drug Delivery Systems*. 2025.[47]
[12] Yar MAE, et al. *Artificial Intelligence in Pharmaceutics: Current Applications, Methodological Challenges, and Translational Potential—A Review*. 2026.[48]
[13] Singh A, et al. *A Review Analysis of Drug Delivery System Using Artificial Intelligence*. 2025.[49]
[14] Muñiz Castro B, et al. *Machine Learning Applied to Over 900 3D Printed Drug Delivery Systems*. 2021.[17]
[15] Haider R, et al. *Smart Pharmaceutics: Integrating Machine Learning into Controlled Drug Release Systems*. 2025.[50]
[16] Kadibagil V, et al. *Machine Learning Assisted Design and Simulation of Drug Delivery Mechanisms*. 2026.[51]
[17] Kumar KV, et al. *Optimization of Nanoparticulate Drug Delivery Systems Using Hybrid Machine Learning Architectures*. 2026.[52]
[18] Wang Q, et al. *Machine Learning-Enhanced Nanoparticle Design for Precision Cancer Drug Delivery*. 2025.[53]
[19] Kapoor DU, et al. *AI-Driven Design and Optimization of Nanoparticle-Based Drug Delivery Systems*. 2024.[16]
[20] Salah M. *Machine Learning for Nanoparticle Drug Delivery from 2017 to 2026: A Systematic Review*. 2026.[54]
[21] Saravanan G, et al. *Machine Learning-Assisted Optimization of Polymeric Nanoparticles for Targeted Therapy*. 2026.[55]
[22] Loureiro DRP, et al. *Optimization of Lipid-Based Ceftriaxone Delivery System via Machine Learning*. 2022.[14]