



ISSN: 2231-3656

International Journal of Pharmacy and Industrial Research (IJPIR)

IJPIR | Vol. 16 | Issue 4 | Oct – Dec 2026

www.ijpir.com

DOI: <https://doi.org/10.61096/ijpir.v16.iss4.2026.971-988>

AI Driven Green Analytical Chemistry in Pharmaceutical Technology

Nataraj Palaniyappan^{*1}, Vijayakumar Murugesan², Kannan Arumugam³, Abuthaheer Thajudeen⁴
M. Ravisankar⁵,

^{1*}Scientist, Novitium Pharma LLC, New Jersey, USA

^{2,3.} Research Scholar, School of Pharmacy, JSS Academy of Higher Education and Research, Mauritius

^{4.} Research scholar, PRIST University, Vallam, Thanjavur, Tamil Nadu, India.

^{5.} Department of Pharmaceutical Chemistry, Srinivasan College of Pharmaceutical Science, Trichy.

*Corresponding Author: **Nataraj Palaniyappan**

E-mail: palanatraj2020@gmail.com



Published on:

06.10.2026

Published by:

Futuristic
Publications

2026| All rights
reserved.



Creative Commons
Attribution 4.0
International
License.

Abstract: Pharmaceutical laboratories depend on analytical procedures that are selective, accurate, robust, and suitable for regulated decisions, yet conventional method development can consume substantial solvent, energy, sample, and analyst time. Green analytical chemistry (GAC) seeks to reduce these burdens without weakening the quality of analytical information. Machine learning (ML), chemometrics, and automated optimization can support this goal by predicting analytical responses, selecting informative experiments, and optimizing competing objectives. This critical narrative review examines how artificial intelligence (AI) and ML can be integrated with GAC and analytical quality by design (AQbD) for pharmaceutical method development. It distinguishes established design-of-experiments and chemometric workflows from newer data-driven and closed-loop AI approaches, reviews their use in chromatography and spectroscopy, and evaluates solvent hazard, waste, energy, preparation, and lifecycle metrics. Published examples include neural-network-assisted reversed-phase HPLC, Bayesian optimization of green supercritical fluid chromatography, automated feedback-controlled liquid chromatography, and recent AI-predicted HPLC methods. The evidence indicates that AI can reduce experiment burden and expose useful trade-offs, but it does not by itself ensure a greener or more reliable method. Small and heterogeneous datasets, narrow optimization targets, incomplete environmental inventories, limited external validation, and poor reporting constrain generalization. The review proposes a human-supervised workflow that embeds environmental objectives in the analytical target profile, uses uncertainty-aware experiment selection, validates the final procedure under current ICH guidance, and reports analytical performance and sustainability transparently. Priorities include shared datasets, interpretable models, prospective comparisons, and lifecycle-aware optimization.

Keywords: green analytical chemistry; artificial intelligence; machine learning; pharmaceutical analysis; analytical quality by design

1. Introduction

1.1 Pharmaceutical analysis and its environmental footprint

Analytical procedures underpin decisions throughout pharmaceutical development and manufacture. They are used to establish identity, assay strength, quantify impurities, evaluate degradation, measure dissolution, monitor cleaning, characterize raw materials, and support release and stability testing. The quality of those decisions depends on method performance: results must be accurate and precise enough

for their intended use, and procedures must be sufficiently selective, robust, and transferable. Liquid chromatography remains central to many of these tasks because it can resolve complex mixtures and accommodate a broad range of analytes and matrices. That versatility carries an environmental cost. Mobile phases, sample diluents, extraction solvents, vials, filters, columns, and cleaning solutions contribute to material use and waste. Instrument pumps, ovens, detectors, and autosamplers consume energy, while long development sequences and repeat injections increase resource demand.

The footprint is not confined to the final routine assay. Early method development may involve screening several columns, buffers, organic modifiers, pH values, gradients, and temperatures. When choices are made sequentially through trial and error, each failed experiment adds solvent and instrument time. Sample preparation can add more hazardous materials than the measurement itself. Scale, frequency, and the number of samples matter: a modest reduction per injection can become meaningful when a procedure is used for years or across multiple laboratories. Conversely, a nominally “green” solvent may require more volume, longer run time, elevated temperature, or repeated extraction, shifting rather than eliminating impact. Sustainability therefore needs to be evaluated across the complete analytical workflow, including the decision to measure, sampling, preparation, separation or detection, data processing, cleaning, and waste treatment [1-2].

Green analytical chemistry (GAC) adapts pollution prevention and safer-chemistry principles to analytical practice. Its purpose is not to trade away data quality. A less hazardous technique that cannot quantify a critical impurity at the required level is not fit for that purpose. The relevant design problem is to obtain the necessary information with lower material, energy, hazard, and waste burdens, while retaining suitable accuracy, precision, selectivity, robustness, and practicality. The 12 GAC principles and the SIGNIFICANCE mnemonic provide a useful conceptual basis, while assessment tools such as the Analytical Eco-Scale, GAPI, AGREE, and AGREEprep help communicate selected dimensions of a method’s environmental profile [3-6]. These tools are valuable, but their scores are not interchangeable and do not replace transparent reporting of the underlying conditions and resource use.

1.2 Why AI and machine learning are relevant

Method development is a multivariable problem. Mobile-phase composition, gradient shape, pH, flow rate, temperature, column chemistry, detector settings, extraction conditions, and matrix effects can interact in nonlinear ways. Traditional one-factor-at-a-time experimentation is inefficient when interactions matter. Statistical design of experiments (DoE) improves this by changing multiple factors systematically and estimating their effects. Chemometric and ML models extend the approach by learning complex input–response relationships, predicting performance at untested conditions, and identifying experiments likely to improve an objective. Bayesian optimization is particularly relevant when experiments are costly and the feasible region is constrained, because it can balance exploitation of promising conditions with exploration of uncertain ones.

AI does not automatically make an analytical procedure sustainable. A model trained only to maximize resolution may recommend an environmentally burdensome method. A model that reduces analysis time could increase energy use or require more hazardous additives. Sustainability has to be explicit in the design objective, the experimental data, and the comparison baseline. In recent work, Bayesian optimization has been used in developing a green supercritical fluid chromatographic method for a pharmaceutical ingredient; other studies have examined AI-proposed HPLC conditions against experimentally optimized methods or incorporated automated feedback into chromatographic development [7-10]. These examples are promising, but they are few, diverse in method, and not yet a basis for assuming broad transferability.

1.3 Aim and scope of this review

This review critically examines how AI and ML can support sustainable pharmaceutical method development and optimization. It focuses on the intersection of three fields: GAC, data-driven optimization, and AQbD. It considers small-molecule pharmaceutical analysis as the principal domain, while including spectroscopy, sample preparation, bioanalysis, and process analytical technology (PAT) where these illustrate transferable principles. It distinguishes between models used to predict analytical results, algorithms that recommend experiments, and AI tools that generate candidate conditions from broad knowledge sources. Those categories have different evidence requirements and should not be treated as equivalent.

The central question is whether AI-guided workflows can identify methods that satisfy analytical requirements with a demonstrably smaller environmental burden, and what evidence is needed to support

that claim. The review therefore compares objectives, input data, validation practices, environmental assessment, and practical limitations rather than cataloguing algorithms alone. It also proposes a workflow for integrating sustainability into an analytical target profile, prioritizing experiments, selecting among Pareto-efficient methods, and validating the final procedure.

2. Review approach and scope of evidence

This article is a critical narrative review. A targeted literature search was conducted through October 2026 using PubMed-indexed records, publisher platforms, and official ICH resources. Search combinations included terms for artificial intelligence, machine learning, chemometrics, Bayesian optimization, chromatography, spectroscopy, pharmaceutical analysis, green analytical chemistry, solvent selection, AQbD, and analytical method validation. Foundational studies were retained when they introduced a GAC principle, assessment metric, or regulatory framework needed to interpret newer work. Application studies were prioritized when they reported experimental analytical data, model inputs and outputs, a pharmaceutical analyte or matrix, and at least one relevant performance or sustainability endpoint. Studies in pharmaceutical manufacturing and general analytical chemistry were used as adjacent evidence only when their transferability to method development could be clearly explained.

Because the purpose was critical synthesis rather than pooled effect estimation, no meta-analysis was attempted. The literature is too heterogeneous in analyte, instrument, model, environmental boundary, and outcome definition to support a meaningful combined estimate. The search did not establish a complete screened-study count, and the article should therefore be submitted as a narrative review rather than described as a systematic review. The distinction is important: a journal requiring a systematic or scoping review would require a reproducible database search, duplicate screening, documented exclusions, and a flow diagram based on the authors' final search records. The evidence synthesis below emphasizes the maturity of the claim. Experimental optimization with measured chromatograms is treated as stronger evidence than a computational recommendation without prospective laboratory testing. Sustainability claims are considered strongest when they use a stated boundary, a defined comparator, measured consumption or hazard data, and more than one complementary assessment.

3. Conceptual framework for sustainable analytical method development

3.1 Green analytical chemistry as a design constraint

The GAC principles direct attention to prevention and minimization rather than end-of-pipe mitigation. In practical method development, this can mean asking first whether the required information can be obtained through direct analysis, a smaller sample, a less solvent-intensive detector, a miniaturized separation, or a method with fewer preparation steps. When a separation remains necessary, choices such as solvent identity, flow rate, column dimensions, injection volume, run time, and re-equilibration can influence the material and energy profile. Sample preparation deserves equal attention because extraction and cleanup can dominate solvent consumption and hazardous waste. A GAC strategy should therefore identify high-impact operations before optimization begins and treat their reduction as part of the method's performance specification [1-2].

Greenness is contextual. Ethanol, water, ethyl acetate, and other alternatives are often considered in place of more hazardous solvents, but a solvent's environmental and occupational profile depends on toxicity, flammability, sourcing, recovery, and the full process conditions. A replacement that changes selectivity may need a different column or longer separation. Supercritical fluid chromatography (SFC) can reduce the volume of conventional organic mobile phase in suitable applications, but its practical suitability depends on analyte properties, system availability, and validation requirements. Likewise, replacing a chromatographic method with UV-visible spectrophotometry may reduce solvent and energy use but can compromise specificity in a complex formulation. These are multi-criteria design decisions, not solvent-substitution exercises.

3.2 AQbD and the analytical target profile

AQbD supplies a structured way to connect the purpose of an analytical procedure to its development and control. The analytical target profile (ATP) states what the procedure must measure and the required performance for the intended decision. Critical method parameters and critical analytical attributes can then be identified, studied, and related through experiments and models. The resulting knowledge can support a method operable design region, robustness assessment, system suitability criteria, and lifecycle management. ICH Q14 describes science- and risk-based approaches to analytical procedure development and lifecycle management, while ICH Q2(R2) addresses validation of procedures and includes considerations for multivariate analytical procedures [11-12].

Integrating GAC with AQbD means adding explicit sustainability characteristics alongside the ATP's analytical performance criteria. For example, a stability-indicating impurity procedure may specify resolution of critical pairs, quantitation limits, accuracy, precision, and robustness while also placing limits on hazardous solvent volume per reportable result, total waste, analysis duration, or energy use. These environmental criteria should not override a safety-critical analytical need. Rather, they help discriminate among methods that meet the required analytical profile. If two procedures provide equivalent and sufficiently reliable measurements, the procedure with the lower overall environmental and operational burden can be preferred.

3.3 Sustainability metrics and what they do not measure

The Analytical Eco-Scale uses penalty points for non-ideal inputs and operations; GAPI provides a pictorial assessment across procedural stages; AGREE translates the 12 GAC principles into a normalized visual score; and AGREEp_{rep} focuses attention on sample preparation. BAGI complements green metrics by assessing practical applicability [13]. These tools differ in scope, weighting, required information, and interpretation. One should not compare scores generated by different tools as if they were values on a common scale. A single total score can also conceal trade-offs: a procedure may receive a favorable score despite high solvent use if other criteria compensate, or may appear less green because its sample preparation is reported in greater detail.

For AI-guided optimization, a metric can be used as a screening or reporting layer, but the underlying quantities should also be retained. Useful operational measures include millilitres of each solvent per sample, grams of hazardous waste per result, number of preparation steps, sample mass, number of injections, runtime, energy per batch, and solvent-recovery fraction. Hazard can be characterized using clearly described classification systems and should not be inferred solely from the label “green solvent.” The environmental system boundary should specify whether it covers only the analytical run, sample preparation, instrument operation, consumables, data processing, and waste treatment. Where upstream solvent manufacture or instrument life-cycle impacts are unavailable, those exclusions should be stated.

3.4 A joint design space

The combined design space contains experimental factors, analytical responses, and environmental outcomes. Inputs can include solvent identity and fraction, buffer concentration, pH, gradient, flow, temperature, column dimensions, extraction ratio, sonication time, and detector settings. Analytical responses can include resolution, retention, peak asymmetry, efficiency, sensitivity, recovery, prediction error, and robustness. Sustainability responses can include solvent hazard, volume, total waste, energy, sample consumption, and number of development experiments. The aim is to identify conditions satisfying the ATP while improving sustainability, not to optimize a greenness score in isolation.



Fig 1: Integrated workflow for AI-driven green analytical method development. The schematic links intended use and the analytical target profile to a sustainability inventory, designed experiments,

uncertainty-aware model development, constrained multiobjective optimization, confirmation experiments, validation, and lifecycle monitoring. Human review is included at experiment selection and method acceptance.

4. Artificial intelligence and machine-learning methods

4.1 Terminology and method classes

Artificial intelligence is an umbrella term for computational systems that perform tasks associated with inference, prediction, classification, or decision support. Machine learning refers to algorithms that estimate patterns from data rather than relying only on manually specified rules. Chemometrics overlaps with ML and includes multivariate statistical methods used to extract information from chemical measurements. The labels are sometimes used loosely in pharmaceutical analytical papers. A multiple-regression model, response-surface method, or artificial neural network may all be described as AI, but their data needs, interpretability, and validation requirements differ. A rigorous review should record what the algorithm actually does: predicts a response, classifies a sample, recommends an experimental condition, or automates a sequence of decisions.

Linear and partial least squares (PLS) regression remain useful when relationships are approximately linear or when spectra contain correlated variables. They offer relatively transparent coefficients or latent variables and can be effective for multicomponent spectral analysis. Response-surface models, including polynomial regression based on central composite or Box–Behnken designs, capture curvature and interactions within a defined region. Their main strength is design interpretability; their main limitation is that the selected functional form may not describe highly nonlinear behavior beyond the studied domain. These methods often provide a strong baseline against which more complex models should be evaluated.

Artificial neural networks (ANNs) can approximate nonlinear mappings between chromatographic factors and responses. Adaptive neuro-fuzzy inference systems combine fuzzy rules with learning, potentially representing nonlinearities while retaining rule-like structure. Support-vector regression can perform well in moderate-sized datasets with nonlinear kernels, while random forests and gradient-boosted trees can model interactions and rank features. Gaussian process regression provides both predictions and uncertainty estimates, making it useful in Bayesian optimization. Deep learning may be appropriate for large spectral, image, or chromatographic data, but small method-development datasets seldom justify model complexity without careful regularization and independent validation [14-15].

4.2 Optimization algorithms and adaptive experimentation

Conventional DoE selects a structured set of experiments before observing their outcomes. It is effective for factor screening and response-surface estimation, especially when the number of factors is manageable and the experimental region is known. Sequential optimization updates the next experiment after each result. Bayesian optimization uses a surrogate model to predict objective values and uncertainty, then selects a condition through an acquisition rule such as expected improvement or a lower confidence bound. In chromatography, the inputs may be categorical and continuous: column phase, solvent type, pH, modifier proportion, gradient duration, temperature, or flow rate. Constraints can encode practical limits, equipment ranges, solvent compatibility, and safety requirements.

Active learning is a broader strategy in which the model chooses new samples or experiments expected to improve its knowledge or the decision. Multi-fidelity optimization can combine inexpensive simulations, retention prediction, or small-scale tests with more costly full analyses. Reinforcement learning has conceptual relevance to repeated, feedback-controlled decisions, but its use in pharmaceutical analytical method development remains less established than DoE, regression, and Bayesian optimization. Claims of an autonomous analytical laboratory require more than an algorithm: instrument interfaces, reliable peak processing, sample handling, error detection, data provenance, and human oversight all contribute to the closed loop.

AI-generated method suggestions from large language models or general-purpose assistants constitute a distinct category. Such systems may rapidly propose a starting solvent, column, or pH from learned textual patterns, but a generated recommendation is not equivalent to a calibrated prediction from a defined training set. It may omit an incompatibility, create a plausible but unsupported setting, or conceal uncertainty. A recent ranolazine evaluation illustrates this distinction: the AI platform produced usable starting suggestions, but performance varied by parameter and human refinement remained necessary, especially for chemically specific interactions [10]. LLMs can support ideation or structured extraction, but recommendations must be experimentally checked against the ATP and documented as hypotheses rather than validated outputs.

4.3 Choosing models for purpose

Model selection should be driven by the decision and available evidence. If the objective is to screen a small set of factors and estimate interactions, an appropriately designed factorial or response-surface model may be more informative than a neural network. If the objective is to predict retention for compounds with known descriptors, a mechanistic or hybrid retention model may generalize better than an empirical model trained on one analyte. If experiments are costly and the objective is to discover workable settings with few runs, Bayesian optimization may be suitable. If spectra overlap strongly, supervised multivariate calibration may improve selectivity without a physical separation, but it must be challenged with independent samples, matrix variation, and reference methods.

4.4 Explainability and uncertainty

Interpretability is useful in analytical chemistry because the factors have chemical meaning. A model should ideally show how solvent fraction, pH, flow, or temperature influences the predicted response and whether interactions are plausible. Feature importance and partial-dependence plots can support interpretation, but they do not establish causality. A model may assign high importance to a variable because it is correlated with an unmeasured factor. Experimental confirmation remains necessary when changing a parameter for a regulated method.

Uncertainty estimates are particularly important near acceptance boundaries. A method predicted to achieve minimum resolution with wide uncertainty should not be selected solely because its point estimate is favorable. Bayesian optimization can use uncertainty to seek informative experiments; prediction intervals can be used to define conservative acceptance criteria; and applicability-domain checks can flag conditions unlike the training data. Confidence should be reported separately for model fit, predictive performance, and the final method's analytical validation. A high coefficient of determination on training data is not evidence of reliable external prediction.

5. Data foundations and model-development practice

5.1 Data structure and provenance

The practical value of AI depends on the quality of its training and experimental data. Analytical development records are often fragmented across instrument software, notebooks, spreadsheets, and laboratory information systems. They may record peak areas but omit solvent lot, column age, temperature equilibration, injection sequence, mobile-phase preparation, or analyst. When development outcomes are copied into a model without those conditions, apparently comparable runs may not be comparable. Data curation should preserve the experimental unit, instrument and column metadata, date, sample and matrix, method version, response-processing steps, and reason for failed or excluded runs.

For chromatographic optimization, a data record may include categorical factors such as column chemistry and solvent identity; numeric factors such as pH, modifier proportion, flow, gradient slope, and temperature; and output variables such as retention, resolution, asymmetry, pressure, runtime, and solvent volume. For sample preparation, inputs may include solvent composition, extraction time, temperature, sample-to-solvent ratio, and cleanup material, while outputs include recovery, precision, matrix effect, solvent and material use, and preparation time. Spectral models require instrument, resolution, preprocessing, sample composition, reference method, and measurement conditions. When raw chromatograms or spectra can be retained, future harmonization and secondary analysis become more feasible.

5.2 Sample size, imbalance, and domain shift

Method-development datasets are usually small because each observation represents a physical experiment. Hundreds of rows may still cover only a narrow region for one analyte, instrument, column, or matrix. Randomly splitting replicate injections into training and test sets can produce leakage: near-identical runs appear in both, inflating predictive performance. Stronger tests hold out batches, compounds, days, columns, instruments, or laboratories, depending on the intended generalization. If a model is meant to optimize one analyte on one system, within-system validation may be adequate for an early proof of concept. If it is claimed to support other compounds or sites, validation must challenge those differences directly.

Class imbalance and missingness also matter. Failed separations, no-elution outcomes, detector saturation, or precipitated mobile phases may be less frequently recorded than successful chromatograms. Excluding them can create a biased model that recommends unsafe or impractical conditions. Missing data

should be described and handled with a documented strategy. The model should distinguish “not measured,” “instrument failed,” and “analytical response was below detection,” which have different meanings. Outlier removal should be supported by a documented cause rather than used to improve a desired score.

5.3 Feature engineering and chemical knowledge

Feature representation determines what a model can learn. Direct method variables are usually necessary for optimization, while analyte descriptors can help transfer models across molecules. Descriptors may include ionization constants, logP, molecular weight, hydrogen-bond donors and acceptors, polar surface area, functional-group counts, or calculated solubility. For retention prediction, structure fingerprints or quantum-chemical descriptors may help, but their use requires sufficient chemically diverse data and appropriate representation of stationary-phase and mobile-phase chemistry. Solvent descriptors can encode polarity, hydrogen-bonding capacity, viscosity, toxicity, flammability, and environmental persistence. The descriptors should be selected before seeing outcome labels where possible, and their source, calculation software, and units should be recorded.

Mechanistic knowledge can reduce data needs. Hybrid models can combine chromatographic retention equations or mass-transfer concepts with a learned correction term. Constrained optimization can exclude conditions likely to cause precipitation, exceed pressure limits, or violate detection compatibility. Chemical priors can improve sample efficiency but also encode assumptions that may fail for unusual analytes. The most defensible approach is to state which relationships are learned, which are imposed, and how violations are detected. Training data, code, parameter definitions, and preprocessing should be made available when confidentiality permits; otherwise, a sufficient summary should enable reviewers to assess reproducibility.

5.4 Model assessment

Evaluation should be separated into training, tuning, internal validation, and external testing. For regression, useful statistics may include MAE, RMSE, bias, calibration slope, and prediction-interval coverage. The relevant error depends on the application: a one-minute retention error may be acceptable in early screening but unacceptable when it changes a critical-pair resolution decision. Classification models require sensitivity, specificity, and class-wise performance, especially when false acceptance of a poor method is consequential. Repeated random cross-validation can estimate internal stability but does not substitute for an independent challenge set.

Optimization should be evaluated by prospective laboratory experiments at recommended conditions, including points near predicted boundaries and at least one comparator. The model should report the number of experiments required to achieve an acceptable method, not just a final response. Performance should be compared with an appropriate DoE or expert-led workflow. If a model proposes conditions outside the training domain, uncertainty should be high or the recommendation should be blocked. A useful reporting checklist is provided in Table 4.

6. Applications in pharmaceutical method development

6.1 Liquid chromatography and reversed-phase HPLC

Liquid chromatography is the most visible target for AI-assisted pharmaceutical method development because chromatographic performance depends on multiple interacting settings and each injection provides rich response data. Early data-driven work applied ANN, adaptive neuro-fuzzy inference, and multilinear regression to model HPLC response surfaces for selected analytes, illustrating that nonlinear models can approximate retention behavior under studied conditions [14]. DoE-linked ANN methods have also been used to optimize simultaneous reversed-phase HPLC of cyclosporin A and etodolac across solution, plasma, nanocapsule, and emulsion matrices [15]. These studies support the feasibility of modeling method responses, but they do not establish that the recommended conditions will generalize across analytes, columns, or laboratories.

Automated feedback-controlled LC development moves beyond offline prediction. In this workflow, the instrument performs planned runs, software evaluates the chromatographic outcome, and an optimization algorithm selects subsequent conditions. A 2024 pharmaceutical study demonstrated automated screening and feedback optimization across small molecules and larger modalities, including peptides, proteins, and vaccine-related products [8]. Such systems can reduce repetitive analyst work and standardize experiment selection. Their potential environmental benefit is indirect unless solvent volume,

waste, energy, and number of injections are actually measured. Automation may lower the number of manual repeats, but it may also increase instrument throughput and total solvent use if the system runs continuously without a material-reduction objective.

Bayesian optimization provides a more direct example of coupling adaptive experiments with greener separation. Brunelli and colleagues combined automated multifactorial design and Bayesian optimization to develop an SFC method for a pharmaceutical ingredient [7]. Their comparison reported a 17-30% reduction in a method-specific green-score measure relative to reversed-phase LC, depending on the sequence size. This study is important because it links method optimization to environmental performance and uses a separation mode that can reduce conventional organic-solvent demand. Its findings remain specific to the selected analyte, instrument, and comparison boundary; broader claims require additional prospective studies and consistent definitions of the green score.

Recent studies have begun to compare AI-generated or AI-predicted settings with laboratory-optimized HPLC. A 2025 report evaluated predictions and greenness for simultaneous separation of amlodipine, hydrochlorothiazide, and candesartan [9]. In 2026, a parameter-level evaluation for ranolazine found that a general AI system's accuracy varied across operating and chemical parameters; instrument settings were predicted more reliably than certain chemical-interaction choices, and human optimization improved the final procedure [10]. These studies help expose a practical limitation: language-model outputs can be useful starting points but may fail on reagent necessity, buffer selection, or analyte-specific interactions. They also underline why accuracy should be scored by parameter and verified in the laboratory, rather than inferred from a plausible written protocol.

6.2 Supercritical fluid chromatography and alternative separation modes

SFC uses compressed carbon dioxide, typically with an organic modifier, and can reduce the consumption of conventional organic mobile phase for suitable compounds and applications. In pharmaceutical development it may support achiral and chiral separations, impurity analysis, and screening. The green advantage is not universal: modifier identity and proportion, system energy, pressure, sample preparation, throughput, and laboratory infrastructure all influence the net assessment. A lower volume of liquid solvent is a useful observation but not a full life-cycle assessment.

ML can assist SFC through prediction of retention, solvent-modifier selection, pressure and temperature optimization, or adaptive experiment selection. The 2024 Bayesian optimization study provides a direct pharmaceutical analytical example [7]. A valuable next step is to compare SFC and LC on matched analytical targets and report both methods' performance, solvent volumes, energy, and development experiments. Transferability should be tested across molecules with different polarity and ionization characteristics, since a model optimized for one ingredient does not establish a broadly applicable SFC platform. Decisions should account for method transfer to QC, availability of qualified equipment, and whether the benefits persist at routine sample volumes.

6.3 Solvent selection and sample preparation

Solvent selection is a natural data-driven problem because candidate solvents have measurable physicochemical, safety, and environmental descriptors, while analyte solubility and extraction performance can be experimentally measured or predicted. Machine learning can rank solvents, estimate solubility, and identify mixtures that meet a target while avoiding high-hazard choices. The SolECOs platform illustrates the use of large API-solvent solubility data and predictive models to support more sustainable solvent choice in pharmaceutical manufacturing [16]. This is adjacent to analytical method development rather than direct proof of greener analytical procedures. The transfer opportunity is methodological: an analogous dataset could pair API and impurity descriptors with chromatographic selectivity, solubility, detector compatibility, and solvent hazard for analytical solvents.

Sample preparation can dominate method burden, particularly in bioanalysis and complex dosage forms. AI may help optimize extraction solvent, pH, liquid-to-solid ratio, sonication, protein precipitation, solid-phase extraction, and cleanup. The objectives should include recovery and matrix effect as well as solvent volume, material use, and preparation time. A high-recovery extraction that uses several large-volume solvent washes may be less desirable than a slightly lower-volume method that still meets the predefined accuracy and precision requirements. AGREEprep is particularly relevant for transparent evaluation of this part of the workflow [6].

The use of alternative solvents also requires careful assessment. Ethanol, water, ethyl acetate, and certain bio-based or eutectic systems may be candidates, but their suitability depends on analyte solubility, selectivity, viscosity, volatility, miscibility, detector compatibility, and waste processing. "Natural" or "bio-

based” does not automatically mean benign, and solvent mixtures can have different hazards from their individual components. An ML model should not encode a binary green/non-green flag where graded hazard, sourcing, and operational properties are available. Candidate selection can instead use constraints or multi-objective scoring, followed by laboratory screening and a documented safety review.

6.4 Spectroscopy and chemometric analytical procedures

UV-visible, NIR, Raman, fluorescence, and infrared methods can reduce sample preparation and avoid chromatographic solvent use in appropriate applications. Their main challenge in pharmaceutical mixtures is often overlapping signals, matrix variability, or limited selectivity. Chemometric models such as PLS, principal-component regression, and ANN can use the full spectrum rather than a single wavelength to quantify multiple components. ML can also support spectral preprocessing, wavelength selection, classification, and anomaly detection. The environmental case is strongest when a spectroscopic method directly replaces a solvent-intensive separation without compromising the intended measurement.

Data-driven methods for multicomponent spectral analysis must be validated beyond calibration fit. Calibration samples should span relevant concentration ranges, excipient variation, degradation, instrument variability, and sample preparation conditions. Independent test samples should be analyzed using an appropriate reference procedure. If the model will be used for release or stability testing, its lifecycle and change control require special attention. ICH Q2(R2) explicitly includes model-validation concepts and a multivariate analytical procedure framework, while Q14 places such procedures within analytical development and lifecycle thinking [11-12].

6.5 PAT, dissolution, bioanalysis, and lifecycle applications

PAT and real-time release strategies generate repeated measurements and can benefit from multivariate analysis, anomaly detection, and adaptive sampling. A validated NIR or Raman model may reduce the need for destructive testing or repeated laboratory assays, but only if its prediction uncertainty and operating range are controlled. ML can select informative calibration samples, identify instrument drift, and support maintenance. In dissolution analysis or process monitoring, models can connect sensor signals to product-quality attributes, potentially reducing sample handling and laboratory turnaround. The environmental benefit should be quantified across the complete control strategy, including calibration, cleaning, sensor lifecycle, and confirmatory testing.

Bioanalysis presents additional complexities: matrix effects, low analyte concentrations, extraction recovery, carryover, and sample stability. An ML model that predicts a simpler extraction or direct-injection method must still demonstrate selectivity and sensitivity in relevant biological matrices. Similar caution applies to impurity profiling and stability-indicating procedures, where degradation products may be unknown or present at low levels. A method that reduces solvent use but fails to detect a relevant degradant is not a sustainable improvement in a pharmaceutical control strategy.

6.6 Evidence synthesis and comparative lessons

Table 2 summarizes representative evidence. The examples span different maturity levels, from offline neural-network optimization to experimentally closed-loop LC and recent AI predictions. They should not be ranked by a single performance number because analytes, platforms, targets, and comparator methods differ. The clearest common lesson is that AI can reduce the search burden or improve experiment selection, while method performance and greenness remain empirical outcomes requiring direct measurement.

Table 1. Machine-learning and optimization methods for analytical method development

Method family	Typical analytical role	Strengths	Main limitations and safeguards
Factorial and response-surface DoE	Screen method factors and estimate interactions within a planned region	Efficient, statistically grounded, interpretable factor effects	Requires a defined design region; model form may not extrapolate; include replicates and center points

PLS and other chemometrics	Quantify analytes from correlated spectral variables; compress high-dimensional signals	Works with collinear spectra; established for multivariate calibration	Sensitive to calibration-set coverage and preprocessing; independent validation is essential
ANN and neuro-fuzzy models	Predict nonlinear chromatographic or spectral responses	Flexible approximation of nonlinear relationships	Can overfit small datasets; report architecture, tuning, and external prediction error
Random forest and boosted trees	Predict responses and rank influential factors in tabular data	Handles nonlinearities and mixed predictors; limited scaling requirements	Importance is not causality; extrapolation is weak and must be controlled
Support-vector regression	Predict responses from small or moderate datasets	Can model nonlinear relationships with kernels	Kernel and hyperparameter selection matter; predictions outside training support are unreliable
Gaussian processes and Bayesian optimization	Predict response with uncertainty and recommend the next experiment	Data-efficient sequential search; uncertainty can guide experiments	Computational and kernel choices matter; constraints and stopping rules must be explicit
Active learning and multi-fidelity methods	Select experiments that improve model knowledge or decision quality	Can reduce expensive experiments by using information strategically	Requires reliable acquisition logic and prospective confirmation
Large language models	Generate candidate conditions or summarize method-development evidence	Fast ideation and natural-language interface	Outputs may be ungrounded or chemically incomplete; treat as hypotheses, not validated predictions

Table 2. Representative evidence for AI, ML, and sustainable pharmaceutical analysis

Study or evidence type	Platform and contribution	Sustainability or validation relevance	Interpretation
Abba et al., 2020 [14]	Compared ANN, adaptive neuro-fuzzy inference, and linear modeling for HPLC response-surface prediction	Demonstrated data-driven modeling of chromatographic responses	Early methodological evidence; applicability is bounded by the studied compounds and factors
Rahman et al., 2021 [15]	Combined factorial and central-composite DoE with ANN for simultaneous cyclosporin A and etodolac analysis	Included validation across solution and complex matrices	Shows the value of DoE-linked nonlinear optimization; does not by itself establish a quantified environmental gain
Brunelli et al., 2024 [7]	Automated DoE and Bayesian optimization for green SFC of a pharmaceutical ingredient	Reported a 17-30% reduction in a method-specific green-score measure relative to RPLC, depending on sequence size	Direct example of adaptive optimization linked to an environmental comparison; result is application-specific
Naser Aldine et al., 2024	Feedback-controlled automated LC	Demonstrated a closed-loop workflow and	Strong automation evidence; environmental

[8]	development for diverse pharmaceutical modalities	reduced manual development burden	benefit requires direct measurement of resource use
Lotfy et al., 2025 [9]	Compared AI-predicted and laboratory-optimized HPLC for three pharmaceutical analytes	Explicitly assessed greenness and chromatographic performance	Useful comparative design; broader conclusions require external replication and consistent metrics
SoLECOs, 2025 [16]	Used predictive models and sustainability assessment to support API-solvent selection in manufacturing	Provides an adjacent solvent-selection framework	Transferable data architecture, but not direct evidence for analytical mobile-phase performance
Obaydo et al., 2026 [10]	Evaluated several AI systems for ranolazine HPLC suggestions and scored parameter-level prediction	Reported variable accuracy by parameter and need for human refinement	Illustrates promise and limits of general AI; findings are early and analyte-specific
Green AQbD studies [17-18]	Applied DoE/QbD to sustainable RP-UPLC and HPTLC pharmaceutical procedures	Integrated performance and greenness evaluation without ML-led optimization	Useful baseline for comparing AI-assisted methods against established sustainable development practice

7. Integrating sustainability into multiobjective optimization

7.1 Define the optimization problem before modeling

A sustainable optimization workflow begins by describing the decision. The ATP should state the analyte or quality attribute, matrix, concentration range, intended use, critical impurities or interferences, and minimum analytical performance. The development team can then identify which variables are controllable, which responses are hard constraints, and which environmental outcomes are optimization objectives. This step avoids the common error of choosing an algorithm before deciding what “better” means.

Analytical requirements should normally be represented as feasibility constraints. Examples include minimum resolution for a critical pair, maximum bias, acceptable precision, a quantitation limit below a specification threshold, acceptable recovery, and a robustness criterion. Sustainability objectives may include minimizing hazardous-solvent volume, total waste per reportable result, energy per sample, sample consumption, and the number of development experiments. Practicality objectives may include runtime, cost, availability of consumables, and ease of transfer. These dimensions can conflict, so the method should produce a set of options rather than a single opaque score.

An optimization can be written conceptually as: find method settings x that satisfy analytical constraints $g(x)$ and minimize the vector of environmental and operational outcomes $f(x)$. For example, x may describe a column, solvent composition, gradient, flow, temperature, and extraction protocol. The constraints can require resolution above a threshold, quantitation error below an agreed limit, and pressure within instrument limits. The objectives can then minimize hazardous solvent use and waste while maintaining practical throughput. This formulation makes sustainability visible to the optimizer and permits explicit trade-off analysis.

7.2 Build a consistent environmental inventory

Environmental endpoints need consistent units and boundaries. Solvent use can be reported per injection, sample, or reportable result; waste can include both mobile phase and preparation waste; energy should be measured or estimated with stated equipment assumptions. Development burden can be measured as the number of experiments and total resource consumption before an acceptable method is obtained. For routine use, annual volume depends on sample load, system suitability, blanks, calibration, repeat rate, and cleaning. A comparison based only on nominal runtime can misstate the total burden if one method requires more re-equilibration or repeat analysis.

Hazard classification should be transparent and current. Solvent screening can use a named pharmaceutical or chemical hazard guide, regulatory classification, or life-cycle dataset, with limitations described. Where different solvent guides disagree, the disagreement should be shown rather than collapsed into a single binary category. Biodegradability, renewability, toxicity, flammability, and recovery are related but distinct properties. A solvent's "greenness" should not be asserted from origin alone. If upstream impacts are not included, authors should say that the score is an operational laboratory assessment rather than a full life-cycle assessment.

Although modest ML models may have little computational burden relative to laboratory experiments, compute use can be estimated and reported for resource-intensive pipelines [19].

7.3 Pareto analysis and decision rules

Weighted sums are simple but can hide the value judgments embedded in weights. If environmental score is multiplied by an arbitrary weight, a method with weak performance may appear optimal even when it fails a critical task. A safer approach is to filter first by the ATP, then identify Pareto-efficient methods across environmental and practical outcomes. A solution is Pareto-efficient when no objective can be improved without worsening at least one other objective. The decision-maker can then select a candidate with a documented rationale, such as lower solvent hazard at equivalent performance or shorter analysis time with a modest increase in solvent use.

Sensitivity analysis should examine whether the preferred method changes when weights, thresholds, sample load, solvent recovery, or energy assumptions change. If a method is favored only under a narrow set of assumptions, the conclusion is fragile. If it remains favorable across plausible conditions, the recommendation is more robust. The number of samples analyzed over the procedure's lifecycle can change the choice: a solvent-intensive but fast method might be defensible for a very low sample volume and unfavorable for routine high-throughput QC. Scenario analysis should therefore accompany any total score used to select a method.



Fig 2: Conceptual Pareto front for analytical performance and environmental burden. The schematic illustrates that several non-dominated methods may satisfy the analytical target profile while differing in solvent hazard, waste, energy, or throughput. It is conceptual and contains no experimental data; method selection requires a documented decision rule.

7.4 Sequential optimization under experimental constraints

Bayesian optimization can support multiobjective development by proposing the next condition that is likely to improve one objective or reduce uncertainty near the Pareto front. The acquisition rule can incorporate constraints, such as limiting pressure, avoiding solvent mixtures that precipitate buffer, and prioritizing experiments expected to satisfy critical-pair resolution. A practical sequence begins with

literature and mechanistic priors, uses a small screening DoE to identify influential factors, fits a surrogate model, and then selects informative experiments adaptively. Human review should reject chemically implausible or unsafe recommendations before the instrument is run.

The search should stop when the model has identified a robust region, when additional experiments are unlikely to change the decision, or when a predefined resource budget is reached. A stop rule avoids endless optimization toward marginal improvements in a score. It should be followed by confirmation runs at the selected condition and deliberate perturbations near anticipated operating limits. A final validation sequence must be separate from model tuning; otherwise, performance estimates will be optimistic. Figure 2 illustrates how acceptable methods can occupy a Pareto front rather than a single universal optimum.

7.5 Comparing AI with a conventional baseline

Any claim that AI made method development more sustainable requires an appropriate baseline. The baseline should represent a credible conventional workflow, such as expert-led scouting or a pre-specified DoE, not an intentionally inefficient trial-and-error sequence. Both workflows should start from the same ATP and equipment, and the comparison should record elapsed development time, experiments, solvent and sample use, waste, energy, final analytical performance, and robustness. A paired or prospective design is preferable to comparing separate studies with different analytes and laboratories.

8. Validation, regulatory alignment, and reproducibility

8.1 Validation of the analytical procedure and the model

Two related but distinct objects require evaluation: the analytical procedure and the predictive model used to develop or operate it. The analytical procedure must be fit for its intended purpose. Validation characteristics depend on the use and may include specificity or selectivity, accuracy, precision, reportable range, detection or quantitation limits, robustness, and system suitability. ICH Q2(R2) describes validation of analytical procedures and includes an annex with illustrative approaches for multivariate techniques; ICH Q14 addresses development, knowledge management, robustness, and lifecycle concepts [11-12]. These guidelines do not make an AI output acceptable merely because a model was used. They require the resulting method and, where applicable, analytical model to be supported by appropriate evidence.

For multivariate analytical procedures, model validation should use independent samples with known reference values and performance criteria established in advance. The test set should reflect variation expected in practice. A random split from one narrow batch may not challenge matrix, instrument, site, or analyst variability. Model overfitting can be hidden when preprocessing, variable selection, and hyperparameter tuning are performed before the data are split. A defensible plan separates data used for training and tuning from data used for final independent assessment, and preserves a record of all transformations. If the model is updated after observing validation results, the validation set has become part of development and a new independent test is needed.

For AI-guided method optimization, the model's predictive performance should be evaluated prospectively at conditions selected by the algorithm. Confirmation experiments should include replicates and should test predicted responses at the proposed operating point. Robustness experiments should deliberately vary critical parameters to determine whether the method has an acceptable operating region. The algorithm should not be used to replace system suitability tests, reference standards, forced-degradation studies, or impurity challenge samples where they are needed for the intended claim. In particular, stability-indicating capacity must be demonstrated experimentally using appropriate degradants or stressed samples, not inferred from a high model score.

8.2 Regulatory documentation and lifecycle control

Regulatory readiness depends on traceability. Development records should identify the ATP, risk assessment, experimental design, raw data, model versions, preprocessing code, prediction uncertainty, acceptance criteria, deviations, and reasons for selecting final conditions. Changes to software, instrument firmware, data processing, training data, or model weights may affect the recommendation and should be controlled according to their potential impact. The level of control should be proportionate to the role of the model: exploratory suggestions used by an analyst differ from a model that directly controls release testing or process decisions.

ICH Q14 supports analytical procedure development and lifecycle management through science- and risk-based approaches, while Q2(R2) links validation to intended use and includes model-validation

concepts for multivariate procedures [11-12]. A practical submission can describe AI as a development tool while validating the final method using conventional performance evidence. If a predictive model itself is part of the analytical procedure, the model's calibration and independent validation become part of the procedure's justification. Regulatory expectations for specific AI systems continue to evolve, so authors should verify current jurisdictional requirements at the time of submission and avoid claiming regulatory acceptance in the absence of a relevant precedent or authority decision.

8.3 Reproducibility and minimum reporting

Reproducibility requires more than naming "AI" or identifying a software package. Authors should report the algorithm class, software and version, parameters, feature definitions, preprocessing, training/validation split, hyperparameter search, random seeds where applicable, missing-data handling, and applicability domain. The full set of factors and experimental ranges should be stated, including failed runs. For sequential optimization, authors should explain the initial design, acquisition function, constraints, stop rule, human interventions, and the reason for each experiment. If a language model was used, the model name and version, date of access, prompt or input structure, and experimental verification should be recorded.

9. Limitations, evidence gaps, and implementation barriers

9.1 Limits of the current evidence

The evidence base remains fragmented. Studies use different analytes, instruments, endpoints, and definitions of AI, so quantitative comparison is difficult. Many method-development studies model a response after a designed experiment but do not compare the resource burden with a conventional workflow. Some papers use a greenness metric after optimization rather than incorporating sustainability into the objective. Conversely, method papers that use AI may report chromatographic performance without environmental data. This separation weakens claims that AI itself produced a greener method.

The data are often small and local. A model for one API on one column and instrument may be useful in that laboratory but offer limited transfer. General-purpose AI tools can generate plausible methods from text, but the underlying training data and uncertainty are often unavailable. A benchmark that scores predictions against one analyst-selected final method may conflate model error with the subjectivity of the comparator. Prospective, multi-laboratory studies are needed to estimate reproducibility and determine how often an AI-recommended method actually works without extensive expert correction.

Greenness metrics also have limitations. They simplify multidimensional environmental and occupational impacts into scores or pictograms. Different tools select different factors and weights, and a score may depend on assumptions about hazard, energy, and sample preparation. Two methods can receive similar aggregate scores while having materially different profiles. Scores are useful for screening and communication but should be accompanied by raw quantities and clear boundaries. Full life-cycle assessment is often impractical for a single lab method, yet a minimal inventory of solvents, waste, energy, and consumables can still make comparisons more meaningful.

9.2 Data, access, and organizational barriers

Pharmaceutical development data may be confidential, proprietary, or stored in non-standard formats. Instrument vendors use different data systems, and laboratories may not preserve failed experiments in a machine-readable form. Small academic or resource-limited laboratories may lack automated instruments and computational staff. These realities affect which AI methods are practical. A low-cost model that works with a well-designed small DoE may be more useful than a complex digital platform requiring proprietary software and continuous integration.

9.3 Avoiding sustainability rebound effects

Efficiency can produce rebound effects. A faster method may encourage more testing, increasing total solvent use even as consumption per sample falls. Automation can run unattended overnight and expand the number of experiments. A green-solvent method may require increased flow or temperature, higher energy, or repeated injections. Miniaturization can reduce sample and solvent but increase disposable consumables. A full comparison should show both per-sample impact and expected total workload. Where throughput changes, the analysis should report absolute resource use for the realistic annual sample volume as well as normalized values.

9.4 Evidence needed for stronger claims

Future research should report prospective head-to-head comparisons between AI-guided and conventional development. At minimum, the design should use the same ATP, equipment class, analyte, matrix, and acceptance criteria; measure development experiments and resources; validate the final method independently; and compare at least one complementary GAC metric alongside raw operational quantities. A second laboratory or a held-out analyte would test transferability. Publication of data and code, when possible, would allow independent assessment of leakage, uncertainty, and applicability domain.

The research question should also be specified at the right level. “Can ML help optimize chromatography?” is too broad to guide a study. More answerable questions include whether Bayesian optimization reaches a robust resolution target with fewer solvent-consuming experiments than a pre-specified DoE, whether a model trained across compounds can recommend lower-hazard solvent systems for a new API, or whether direct spectroscopic analysis can replace a separation without increasing false-negative risk for a critical impurity. Explicit hypotheses improve study design and make negative findings valuable.

10. Future directions and implementation roadmap

10.1 Shared and interoperable data

The field would benefit from curated, machine-readable datasets that combine analyte structures, column and mobile-phase metadata, chromatograms or spectra, preparation steps, analytical outcomes, and environmental inventories. A minimum schema should include units, instrument settings, sample matrix, method version, and data-processing rules. Negative outcomes and failed runs should be retained with reason codes. Benchmark datasets should be split by compound, site, or instrument when the intended use involves transfer. Confidentiality can be addressed through controlled access, federated analysis, synthetic test data for software verification, or de-identified descriptors.

10.2 Hybrid, interpretable, and uncertainty-aware models

Hybrid mechanistic-ML models offer a practical path between empirical black boxes and purely mechanistic simulations. Chromatographic retention, solvent-property models, and mass-transfer principles can constrain predictions, while learned components capture residual behavior. Bayesian and ensemble methods can quantify uncertainty and guide experiment selection. Interpretable outputs should include factor-response relationships, feasible ranges, and a warning when recommendations leave the training domain. Model cards or comparable documentation can describe intended use, known limitations, validation datasets, and retraining triggers.

10.3 Autonomous laboratories with human oversight

Automated instruments can close the loop between prediction, execution, and analysis. Future systems could combine sample preparation, LC or SFC, peak processing, green-metric calculation, and safe experimental planning. Human oversight should remain at decisions involving chemical plausibility, safety, impurity significance, and method acceptance. The system should record every recommendation and intervention. Autonomous operation will be most useful when it reduces total development burden and finds robust conditions, not simply when it increases the number of experiments per day.

10.4 Sustainability-aware objectives and lifecycle use

Future optimizers should treat sustainability as a measurable objective throughout development and routine use. Relevant factors include solvent hazards and volumes, sample-preparation materials, water and energy, waste treatment, analysis and equilibration times, model compute, and eventual method transfer. Metrics should be modular so that a laboratory can add more complete life-cycle information over time. Scenario analysis can estimate total impact under different testing volumes, column lifetimes, solvent-recovery rates, and repeat-analysis frequencies. The result should be an interpretable trade-off profile rather than a universal green score.

10.5 Practical roadmap for pharmaceutical laboratories

A laboratory can begin by selecting a procedure with high solvent use, frequent development work, or a clear opportunity for direct analysis. It should define the ATP and baseline method, inventory current resource use, and choose a bounded experimental region. A conventional screening DoE can establish initial data; a simple interpretable model should be tested before a more complex one. AI-guided

experiments can then target uncertain or promising conditions. The final candidate should be selected from methods that satisfy analytical requirements, assessed with complementary greenness and practicality tools, and confirmed through robustness and validation studies.

11. Conclusion

AI and ML can strengthen sustainable pharmaceutical method development by modeling nonlinear responses, guiding informative experiments, comparing solvent and separation choices, and supporting automated feedback. Current examples show value in DoE-linked neural networks, Bayesian optimization for green SFC, closed-loop LC, and AI-assisted HPLC prediction. They also show that results remain context-dependent and that chemically complex choices often need expert correction. The sustainability benefit is established only when the method meets its analytical target and its environmental advantage is measured against a credible baseline.

The most defensible approach integrates GAC and AQbD from the start. The ATP should define analytical and environmental criteria; data should include failed experiments and resource use; models should report uncertainty and be tested prospectively; and final procedures should be validated and managed across their lifecycle. Greenness metrics can aid comparison, but should be reported alongside solvent, waste, energy, sample, and method-performance data. Shared datasets, transparent benchmarks, and multi-laboratory studies will determine whether AI-assisted analytical development can deliver repeatable sustainability gains across pharmaceutical settings.

Table 3. Green and complementary assessment tools relevant to pharmaceutical analysis

Tool	Primary scope	Best use in an AI-enabled workflow	Important limitation
Analytical Eco-Scale [3]	Penalty-based semi-quantitative assessment of reagents and operations	Compare candidate procedures using disclosed penalties	Score depends on penalty assignments and may compress distinct impacts
GAPI [4]	Pictorial evaluation across multiple analytical workflow steps	Reveal which stages contribute to environmental burden	Qualitative color fields are not a life-cycle impact score
AGREE [5]	Weighted assessment aligned with the 12 GAC principles	Summarize method greenness after reporting component measures	Score depends on input assumptions and should not replace raw data
AGREEprep [6]	Environmental profile of sample-preparation operations	Optimize extraction, cleanup, and sample handling	Does not independently characterize the downstream separation
BAGI [13]	Practical applicability and implementation features	Compare feasibility alongside environmental performance	Practicality is not equivalent to environmental impact or analytical validity
Direct inventory and life-cycle indicators	Solvent, waste, energy, consumables, hazard, and defined boundaries	Train explicit objective functions and support transparent comparisons	Requires measurement or defensible assumptions; inventory may be incomplete

Table 4. Minimum reporting checklist for AI-guided green method development

Reporting domain	Minimum information to provide
Analytical purpose	ATP, analyte and matrix, intended decision, critical impurities/interferences, performance thresholds
Experimental design	Factors, ranges, design type, run order, replicates, failed runs, constraints, stopping rule

Dataset	Units, sample and instrument metadata, preprocessing, missingness, exclusions, provenance, data availability
Model	Algorithm and version, features, hyperparameters, tuning process, random seed where applicable, applicability domain
Evaluation	Training/tuning/test separation, cross-validation, external or prospective experiments, comparator, uncertainty, error metrics
Sustainability inventory	Solvents and volumes, hazard source, waste boundary, sample use, energy assumptions, consumables, throughput, metric version
Human oversight	Review points, interventions, reasons for rejecting recommendations, safety checks
Final procedure	Confirmation runs, robustness, validation characteristics, system suitability, lifecycle and change-control approach
Reproducibility	Code or detailed pseudocode, model/data version, instrument method files where shareable, reporting of limitations

References

1. Gałuszka A, Migaszewski ZM, Konieczka P, Namieśnik J. Green chemistry metrics with special reference to green analytical chemistry. *Molecules*. 2015;20(6):10928-10946. doi:10.3390/molecules200610928.
2. Gałuszka A, Migaszewski ZM, Namieśnik J. The 12 principles of green analytical chemistry and the SIGNIFICANCE mnemonic of green analytical practices. *TrAC Trends Anal Chem*. 2013;50:78-84. doi:10.1016/j.trac.2013.04.010.
3. Gałuszka A, Migaszewski ZM, Konieczka P, Namieśnik J. Analytical Eco-Scale for assessing the greenness of analytical procedures. *TrAC Trends Anal Chem*. 2012;37:61-72. doi:10.1016/j.trac.2012.03.013.
4. Płotka-Wasyłka J. A new tool for the evaluation of the analytical procedure: Green Analytical Procedure Index. *Talanta*. 2018;181:204-209. doi:10.1016/j.talanta.2018.01.013.
5. Pena-Pereira F, Wojnowski W, Tobiszewski M. AGREE—Analytical GREENness Metric Approach and Software. *Anal Chem*. 2020;92(14):10076-10082. doi:10.1021/acs.analchem.0c01887.
6. Wojnowski W, Tobiszewski M, Pena-Pereira F, Psillakis E. AGREEprep—Analytical greenness metric for sample preparation. *TrAC Trends Anal Chem*. 2022;149:116553. doi:10.1016/j.trac.2022.116553.
7. Brunelli C, Osborne R, Yule G, Dixon T, Bruce I, Taylor M. Automated multifactorial design of experiment and Bayesian optimisation algorithm approaches to method development for the green analysis by supercritical fluid chromatography of a pharmaceutical ingredient. *J Chromatogr A*. 2024;1732:465214. doi:10.1016/j.chroma.2024.465214.
8. Naser Aldine F, Singh AN, Wang H, Makey DM, Barrientos RC, Wong M, et al. Improved assay development of pharmaceutical modalities using feedback-controlled liquid chromatography optimization. *J Chromatogr A*. 2024;1722:464830. doi:10.1016/j.chroma.2024.464830.
9. Lotfy HM, Erk N, Genc AA, Obaydo RH, Tiris G. Artificial intelligence in chromatography: greenness and performance evaluation of AI-predicted and in-lab optimized HPLC methods for simultaneous separation of amlodipine, hydrochlorothiazide, and candesartan. *Talanta Open*. 2025;11:100473. doi:10.1016/j.talo.2025.100473.
10. Obaydo RH, Keshishian R, Altinawe YS, et al. Critical evaluation of AI-assisted HPLC method development: a case study of ranolazine analysis with systematic assessment of prediction accuracy and human optimization requirements. *BMC Chem*. 2026. doi:10.1186/s13065-026-01915-w.

11. International Council for Harmonisation. ICH Q14: Analytical Procedure Development. Step 4 guideline. Geneva: ICH; 2023.
12. International Council for Harmonisation. ICH Q2(R2): Validation of Analytical Procedures. Step 4 guideline. Geneva: ICH; 2023.
13. Manousi N, Wojnowski W, Płotka-Wasyłka J, Samanidou V. Blue applicability grade index (BAGI) and software: a new tool for the evaluation of method practicality. *Green Chem.* 2023;25:7598-7604. doi:10.1039/D3GC02347H.
14. Abba SI, Usman AG, Işik S. Simulation for response surface in the HPLC optimization method development using artificial intelligence models: a data-driven approach. *Chemom Intell Lab Syst.* 2020;201:104007. doi:10.1016/j.chemolab.2020.104007.
15. Rahman SNR, et al. Application of design of experiments approach-driven artificial intelligence and machine learning for systematic optimization of reverse phase high performance liquid chromatography method to analyze simultaneously two drugs (cyclosporin A and etodolac) in solution, human plasma, nanocapsules, and emulsions. *AAPS PharmSciTech.* 2021;22(4):155. doi:10.1208/s12249-021-02026-6.
16. Ma Y, Gao S, Mehta N, Fu Q, Li W, Benyahia B. SolECOs: a data-driven platform for sustainable and comprehensive solvent selection in pharmaceutical manufacturing. *Green Chem.* 2025;27:12621-12641. doi:10.1039/D5GC04176G.
17. Vanga MG, Bukke SPN, Kusuma PK, Narapureddy BR, Thalluri C. Integrating green analytical chemistry and analytical quality by design: an innovative approach for RP-UPLC method development of ensifentrine in bulk and inhalation formulations. *BMC Chem.* 2025;19:70. doi:10.1186/s13065-025-01448-8.
18. Prajapati P, Shahi A, Acharya A, et al. Green and sustainable analytical chemistry-driven chromatographic method development for stability study of apixaban using Box-Behnken design and principal component analysis. *J Chromatogr Sci.* 2024;62(7):649-659. doi:10.1093/chromsci/bmad033.
19. Lannelongue L, Grealey J, Inouye M. Green Algorithms: quantifying the carbon footprint of computation. *Adv Sci.* 2021;8(12):2100707. doi:10.1002/advs.202100707.