

1

Introduction to AI in Chemistry: Landscape, Motivation, and Scientific Workflow

Lin Tan

Guangzhou National Laboratory, Guangzhou, 510005, Guangdong, China

1.1 Foundations of Artificial Intelligence: From Ideas to Infrastructure

1.1.1 The Emergence of Artificial Intelligence and Its Scientific Significance

Artificial intelligence (AI) has evolved from a set of intellectual ambitions into a general-purpose scientific and technological infrastructure. Across medicine, engineering, finance, and the natural sciences, AI systems now shape how data is interpreted, how hypotheses are prioritized, and how decisions are made at scale. At a practical level, AI refers to computational methods that enable machines to perform tasks associated with intelligence—learning from experience, recognizing patterns, making inferences, and selecting actions under uncertainty. At a deeper level, AI represents a new way to translate information into decisions.

The scientific significance of AI stems from its capacity to extract structure from complex data. Many scientific domains have become limited not by the ability to generate data, but by the ability to interpret it and act on it efficiently. AI fills this gap by providing tools for (i) discovering patterns that exceed human perceptual bandwidth, (ii) building predictive models that generalize beyond observed cases, and (iii) enabling iterative decision-making in high-dimensional spaces. Importantly, AI does not replace scientific thinking; it augments it by scaling pattern discovery and decision support.

In chemistry, these capabilities are unusually impactful because chemical systems are combinatorial, nonlinear, and context-dependent. Chemical discovery is rarely limited to a single variable; outcomes depend simultaneously on molecular structure, reaction environment, and experimental conditions. AI offers a pathway to navigate this complexity systematically, shifting parts of chemistry from artisanal intuition toward scalable, data-driven workflows (Figure 1.1)

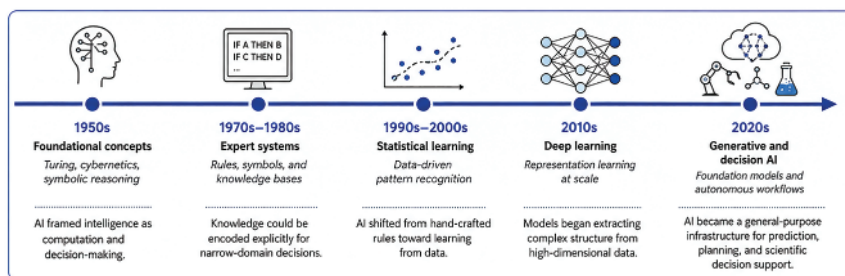


Figure 1.1 Evolution of artificial intelligence. A timeline illustrating key milestones in AI, including the Turing Test, expert systems, statistical learning, and deep learning [1, 3, 5, 14].

1.1.2 Defining Artificial Intelligence: Scope, Objectives, and Core Concepts

AI is best understood as a family of methods and systems designed to transform information into intelligent behavior. The field spans multiple subareas—machine learning, natural language processing, computer vision, planning, and robotics—but these areas are unified by a common objective: building systems that can perceive, learn, reason, and act in complex environments.

For the purposes of chemistry, four concepts provide a useful working decomposition of AI:

- **Learning:** improving performance by extracting regularities from data rather than relying on explicit programming.
- **Reasoning:** drawing conclusions or selecting actions based on data, constraints, and knowledge (explicit or implicit).
- **Problem-Solving and Planning:** finding effective solutions in large spaces of possibilities under constraints.
- **Perception:** interpreting complex signals (e.g. spectra, images, chromatograms) into structured information.

This decomposition is not meant to be philosophical; it is practical. It clarifies why AI is useful for chemistry: chemistry involves learning from experimental outcomes, reasoning under uncertainty, planning synthetic routes, and perceiving chemical signals from instruments (Figure 1.2).

1.1.3 Narrow AI vs. General AI; Weak AI vs. Strong AI

Most AI systems used today are narrow AI: they achieve strong performance within well-defined tasks, such as recognizing patterns in spectra, predicting molecular properties, or proposing retrosynthetic disconnections. These systems are often called weak AI because they do not possess general intelligence; they do not understand the world broadly, and they do not transfer knowledge flexibly across unrelated domains without adaptation.

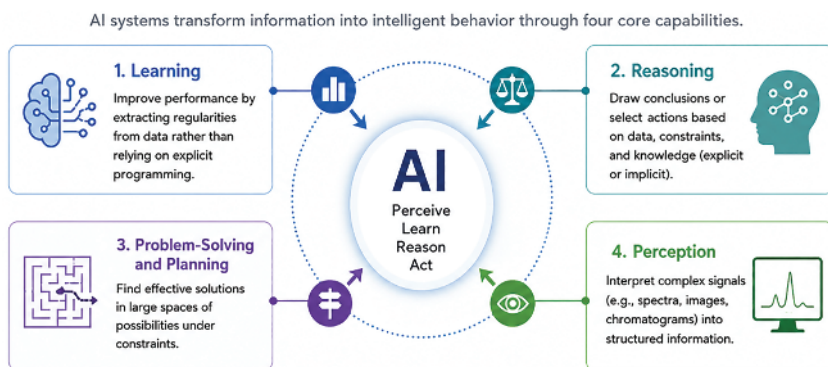


Figure 1.2 Core concepts of AI. A diagram illustrating learning, reasoning, problem-solving/planning, and perception and how they connect to scientific decision-making [3, 5].

General AI (often aligned with the term strong AI) describes a hypothetical class of systems with human-level flexibility—capable of learning and reasoning across domains, forming robust abstractions, and adapting in open-ended environments. Whether general AI will emerge, and what form it might take, remains uncertain and is not required for chemistry to benefit profoundly from AI today.

For chemists, the key message is pragmatic: modern chemistry can be transformed by narrow AI systems integrated into robust workflows without invoking grand claims of human-equivalent intelligence. The value comes from systematic scaling of prediction, prioritization, and experimentation (Figure 1.3).

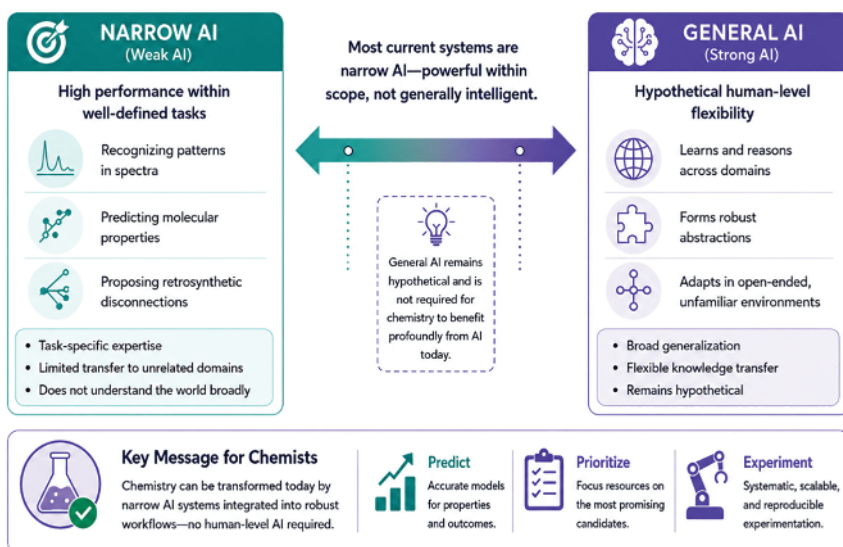


Figure 1.3 Narrow AI vs. general AI. A visual comparison showing task-specific competence vs. broad transfer and adaptation [3].

1.1.4 Early Intellectual Roots: Logic, Cybernetics, and the Turing Test

The foundations of AI draw from logic, philosophy, and early theories of computation. Alan Turing's 1950 proposal—later known as the Turing Test—framed intelligence in operational terms: if a machine's responses are indistinguishable from a human's in conversation, it can be considered intelligent. While the test is not a scientific definition of intelligence, it influenced decades of thinking about what machines might accomplish.

In parallel, cybernetics emphasized feedback, control, and adaptation—ideas that resonate strongly with modern autonomous experimentation. The central insight was that intelligent behavior can emerge from systems that sense, decide, and act in loops. This feedback viewpoint foreshadowed closed-loop chemical discovery, where experiments generate data, models update, and the next experiment is chosen adaptively (Figure 1.4).

1.1.5 Symbolic AI and Expert Systems

Early AI research was dominated by symbolic AI, which attempted to encode knowledge explicitly as rules and symbols. Expert systems emerged as practical implementations: knowledge bases plus inference engines that could make decisions within narrow domains. In chemistry, early expert systems influenced retrosynthetic thinking and knowledge representation, offering structured ways to store and apply human expertise.

Symbolic approaches achieved successes in constrained settings, but they struggled with uncertainty, ambiguity, and scale. Chemical reality is messy: experimental noise, incomplete reporting, and context-dependent outcomes make it difficult

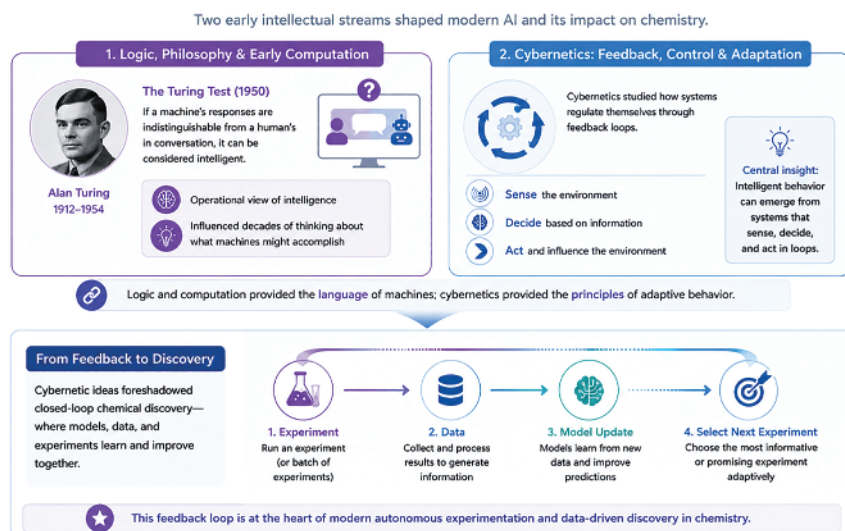


Figure 1.4 Turing Test and cybernetics. A schematic connecting early ideas of computation and feedback control to modern AI workflows [1, 2, 14].

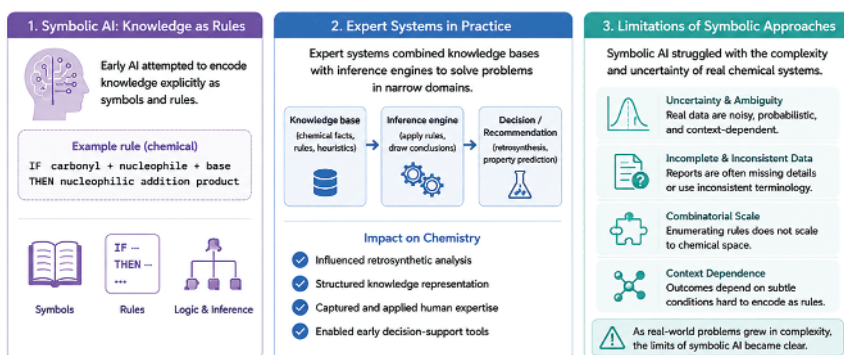


Figure 1.5 Symbolic AI and expert systems. A flowchart showing a knowledge base, inference rules, and decision outputs [3, 4].

to enumerate rules exhaustively. These limitations became increasingly visible as researchers attempted to apply symbolic logic to real-world scientific problems (Figure 1.5).

1.1.6 The First AI Winter and Lessons for Scientific AI

Periods of intense optimism in AI were followed by disappointment when systems failed to generalize beyond toy problems. Funding declined, expectations were reset, and the field entered an “AI winter.” For chemistry, the historical lesson is direct and enduring: scientific AI must be evaluated in realistic conditions—with noise, bias, incomplete information, and shifting distributions.

Two lessons from this era still shape AI in chemistry today:

- Robustness matters more than demos. A model that works on curated examples but fails on real laboratory variation is not scientifically useful.
- Data and evaluation are as important as algorithms. Many failures arose not from a lack of cleverness but from insufficient data quality, narrow problem framing, and unrealistic assumptions.

These lessons reappear throughout the book as we move from landscape to practice (Figure 1.6).

1.1.7 Machine Learning and Representation Learning: Why Modern AI Works

Modern AI is shaped by a fundamental shift: learning from data rather than specifying knowledge entirely by hand. Machine learning systems infer patterns from examples and can generalize to new cases under appropriate conditions. This is scientifically aligned with chemistry, where empirical observations often precede mechanistic certainty and where outcomes are probabilistic rather than deterministic.

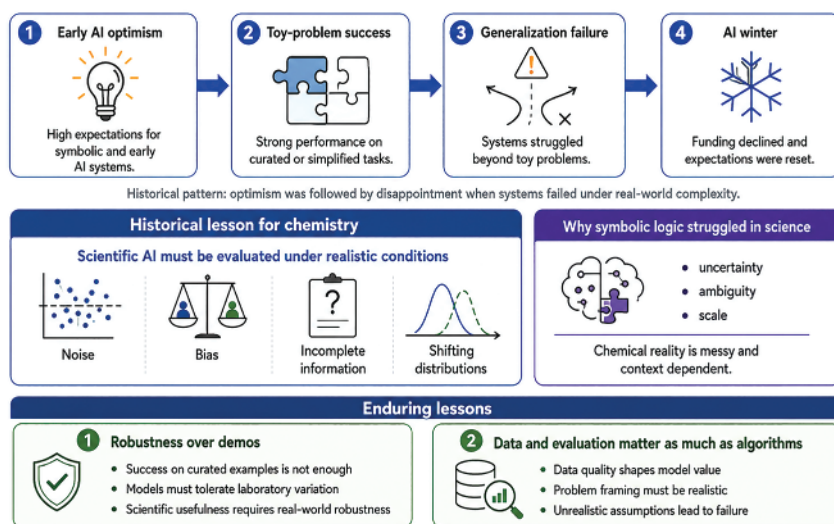


Figure 1.6 The first AI winter and its causes. A diagram linking limited compute, brittle rule systems, and unrealistic expectations [3–5, 17, 18].

A second shift—representation learning—further increased AI’s impact. Instead of depending exclusively on handcrafted descriptors, modern models can learn task-relevant representations from structured chemical inputs such as molecular graphs, reaction strings, or instrument signals. This approach has enabled broader applicability across chemical domains and reduced barriers for nonspecialists to adopt predictive tools.

At the same time, these advances introduced new challenges—interpretability, bias, and reliability—that are especially important in chemistry, where mechanistic understanding and risk-aware decision-making are central (Figure 1.7).

1.1.8 Data, Compute, and Algorithms: The Convergence That Enabled AI at Scale

The practical success of AI is inseparable from the convergence of three forces:

- **Data:** large, digitized corpora of molecules, reactions, spectra, and material properties; increasingly, standardized data from automated experimentation.
- **Compute:** GPUs, cloud infrastructure, and scalable training and inference pipelines that make complex models feasible.
- **Algorithms:** advances that stabilize learning, improve generalization, and enable models to learn from partially labeled or noisy data.

For chemistry, this convergence is particularly powerful because it aligns with the rapid digitization of chemical knowledge: literature mining, public databases, electronic lab notebooks, and high-throughput platforms. The result is a feedback loop—more data enables better models, better models guide better experiments, and better experiments generate better data (Figure 1.8).

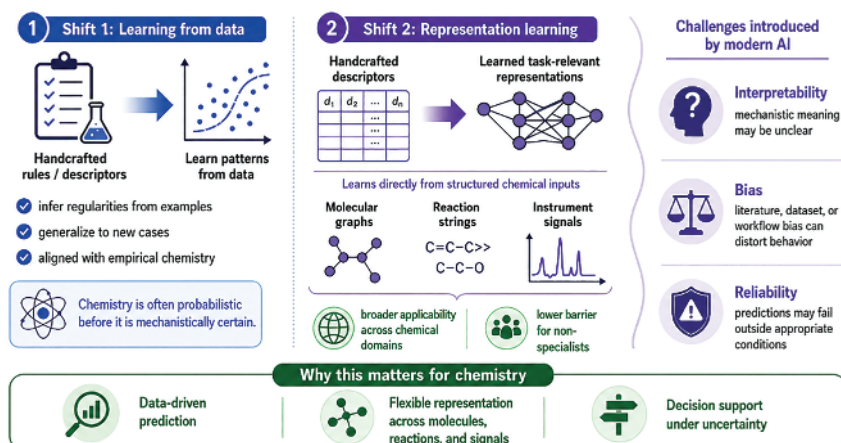


Figure 1.7 From rules to learning. A conceptual comparison between rule-based reasoning and data-driven inference [5, 10, 11].

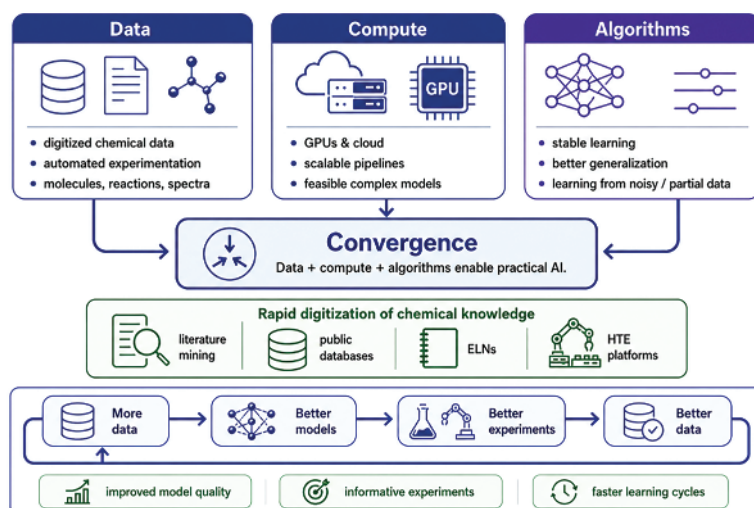


Figure 1.8 The AI triad. A diagram illustrating how data, compute, and algorithms jointly enable modern AI capabilities [5, 7, 14].

1.1.9 From Academic Curiosity to Scientific Infrastructure

AI is increasingly embedded in the infrastructure of science. In chemistry, AI is no longer only a post hoc analytics tool; it is becoming part of how experiments are designed, how hypotheses are prioritized, and how knowledge is encoded and reused. Techniques migrate across domains: methods developed for language modeling inspire reaction prediction; planning algorithms inform synthesis design; reinforcement-style optimization influences adaptive experimentation.

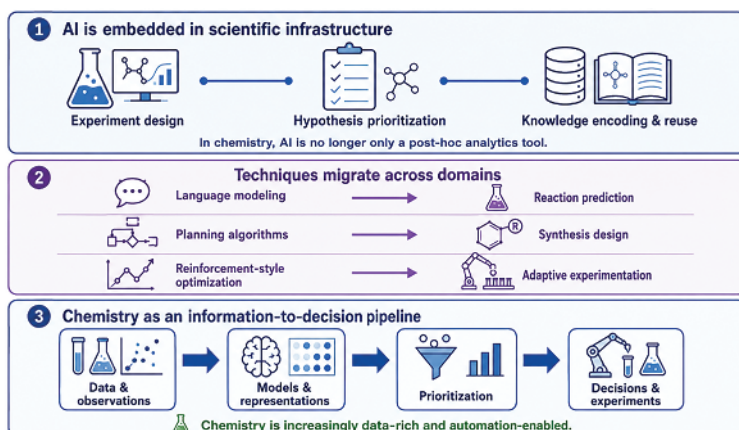


Figure 1.9 AI as scientific infrastructure. A schematic showing AI moving from analysis to planning and decision support across the chemical research cycle [11, 14, 15].

This shift reframes chemistry as an information-to-decision pipeline. Understanding AI’s foundations is therefore not optional for modern chemical research—it is becoming part of the basic literacy required to operate effectively in a field that is increasingly data-rich and automation-enabled (Figure 1.9).

1.2 Why Chemistry Needs AI: Scale, Data, and Decisions

1.2.1 The Scale of Chemical Space and the Limits of Human Throughput

Chemistry is an experimental science operating in a combinatorial universe. Even restricting attention to small organic molecules, the number of synthetically accessible structures is astronomically large, far exceeding what can be explored by human-guided experimentation. Similarly, reaction outcomes depend on many interacting variables—substrates, catalysts, solvents, additives, temperature, concentration, time, and mixing/irradiation profiles—creating condition spaces that cannot be exhaustively searched.

Historically, chemists navigated this space through a combination of mechanistic reasoning, literature knowledge, and intuition. This paradigm remains essential but is increasingly constrained by scale and complexity. As chemistry targets harder problems—selective transformations, complex functional group tolerance, sustainable processes, and rapid iteration—decision-making under uncertainty becomes the norm rather than the exception.

AI offers a complementary approach: it can prioritize experiments, predict outcomes, and identify patterns across datasets that no single researcher can hold in mind. The goal is not to replace chemical intelligence but to amplify it—moving from trial and error toward systematic exploration guided by data (Figure 1.10).

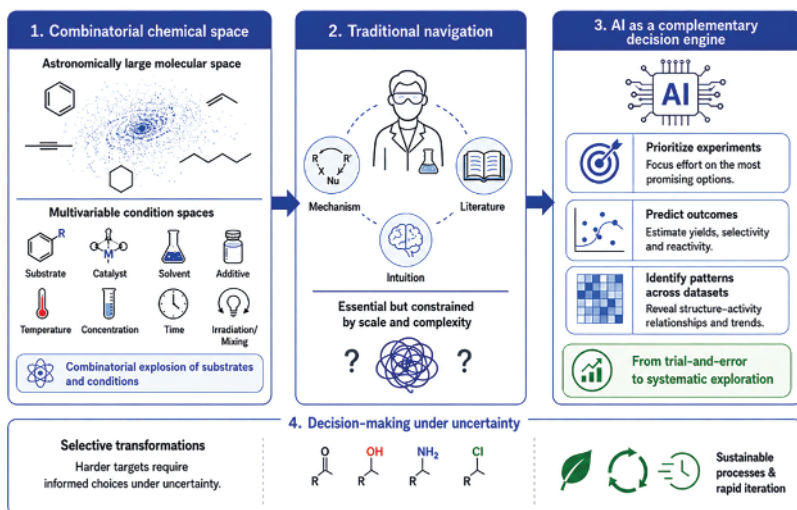


Figure 1.10 Why chemistry needs AI. A diagram linking chemical space size, experimental constraints, and AI-enabled prioritization [7, 11].

1.2.2 Chemical Data: Structure, Heterogeneity, and Noise

AI in chemistry is ultimately limited—and empowered—by chemical data. Chemical datasets are heterogeneous: molecular structures, reaction records, spectra, chromatograms, computational outputs, and materials descriptors have different formats, uncertainties, and missingness patterns. Much data are high-dimensional and relational: a reaction is not a single object but a transformation involving multiple components and a context.

Chemical data are also noisy and biased. Instrument drift, batch effects, operator variability, and incomplete reporting introduce systematic errors. Negative results are underreported, and literature data overrepresent popular transformations. These realities shape what AI can learn and how models should be interpreted.

The implication is central: methods imported from other domains must be adapted to chemistry's data realities. Reliable chemical AI requires careful attention to data provenance, standardization, and the relationship between measured outcomes and underlying chemical mechanisms (Figure 1.11).

1.2.3 Representations of Molecules and Reactions: Encoding Chemistry for Machines

How chemical information is represented determines what AI models can learn. Molecules and reactions are not naturally expressed as fixed-length arrays; they are relational objects defined by connectivity, geometry, and transformation. Practical AI, therefore, depends on representations that preserve chemical meaning while remaining computationally tractable.

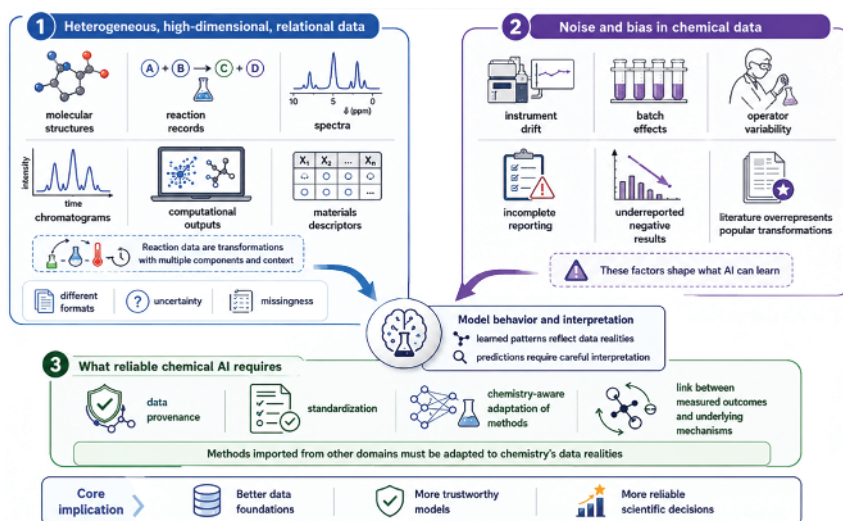


Figure 1.11 The complexity of chemical data. An infographic showing key chemical data types and their associated challenges (noise, bias, missingness, context dependence) [17, 18].

Three broad representation families are widely used:

- Descriptor-based representations (e.g. fingerprints and physicochemical features), which provide a compact, engineered summary of structure.
- Graph-based representations, where atoms and bonds form a graph aligned with chemical intuition.
- Sequence-based representations (e.g. SMILES-like strings) that treat chemistry as a language and enable powerful pattern learning.

Each representation imposes assumptions about what information matters. In that sense, representations are not mere technical details; they are scientific hypotheses about what aspects of chemical systems are relevant for a task. Throughout this book, we emphasize representation choices as a core bridge between chemical thinking and machine learning (Figure 1.12).

1.3 The AI–Chemistry Landscape: What AI Enables Across Chemical Subdisciplines

This section surveys where AI is already changing chemistry. The intent is not to teach each domain in depth (later chapters do that) but to provide a coherent map of tasks, decisions, and workflows.

1.3.1 AI in Drug Discovery: Screening, Design, and Multi-Objective Optimization

Drug discovery is a natural arena for AI because it requires decisions across enormous chemical spaces under multiple constraints. AI models support virtual

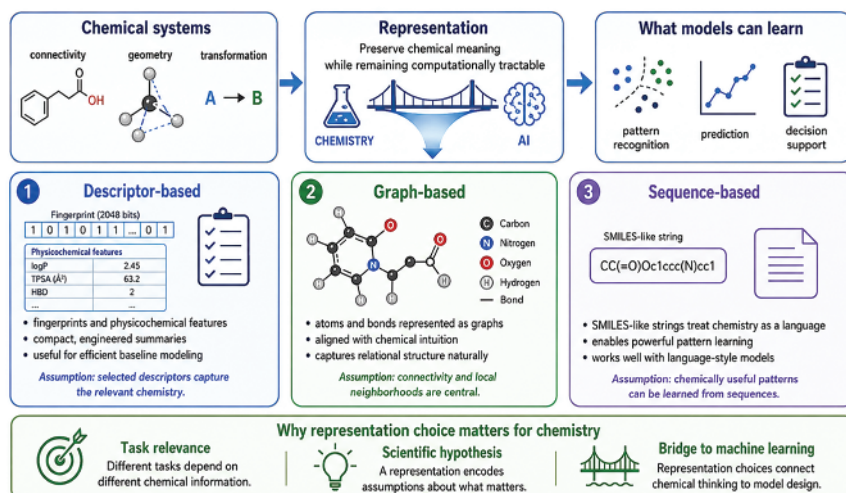


Figure 1.12 Molecular and reaction representations. A comparison of descriptor vectors, molecular graphs, and reaction sequences [8, 10, 11].

screening by prioritizing candidates likely to satisfy biological and physicochemical criteria. Beyond screening, generative approaches can propose novel structures optimized for multiple objectives, accelerating ideation and iteration.

Critically, AI changes the division of labor: models propose and prioritize; chemists define objectives, enforce constraints (synthetic accessibility, safety, novelty), and interpret results. The most effective systems embed AI inside an experimental validation loop rather than treating predictions as endpoints (Figure 1.13).

1.3.2 AI for Reaction Prediction and Retrosynthesis: Scaling Chemical Knowledge

Reaction prediction and retrosynthesis sit at the core of synthetic chemistry. AI systems trained on curated reaction corpora can forecast likely outcomes and suggest disconnections, providing rapid access to patterns distilled from decades of literature. In practice, these tools function as decision support: they broaden the set of options, accelerate route ideation, and help organize chemical knowledge into reusable computational form.

The value is not only speed. By externalizing chemical knowledge into models, AI enables more systematic comparison of alternative routes and supports reproducible planning—especially when paired with standardized reaction representations and robust data curation (Figure 1.14).

1.3.3 AI-Guided Reaction Optimization: From Local Tuning to Systematic Search

Reaction optimization has traditionally been guided by iterative manual tuning. AI enables a different approach: structured exploration of condition spaces guided

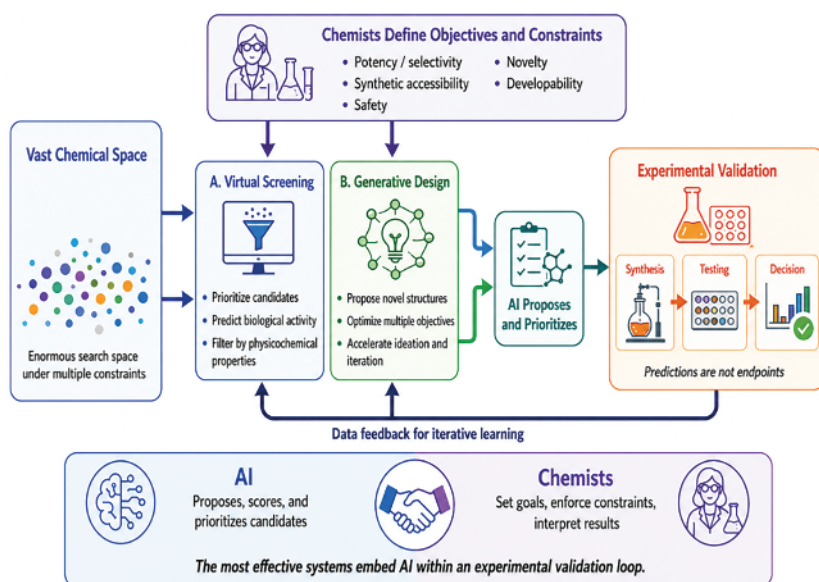


Figure 1.13 AI in drug discovery. A workflow from target definition to screening/design, synthesis, and validation with AI embedded in decision points [6, 8].

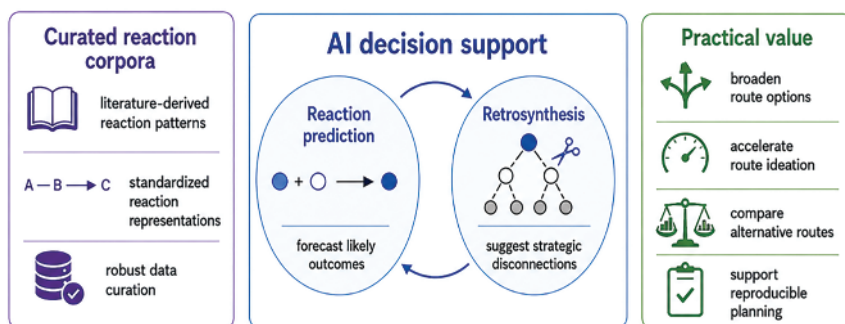


Figure 1.14 Reaction prediction and retrosynthesis. A schematic showing outcome prediction and backward planning from target to precursors [11, 14].

by data. The main benefit is not that AI “knows chemistry,” but that it can efficiently navigate combinatorial spaces and identify nonobvious interactions among variables.

When integrated with high-throughput experimentation (HTE), AI-guided optimization can reduce redundancy, prioritize informative experiments, and accelerate convergence toward robust conditions. This is one of the clearest examples of AI acting as a decision engine rather than a prediction engine (Figure 1.15).

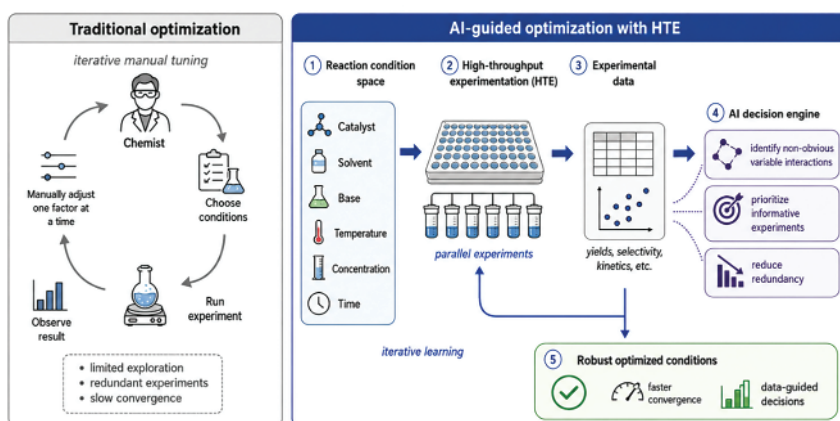


Figure 1.15 AI-guided optimization. A conceptual loop: experiment → data → model update → next experiment recommendation [9, 11, 13].

1.3.4 AI in Materials Discovery: Property Prediction and Inverse Design

Materials discovery often demands navigating vast composition–structure–property spaces. AI models can approximate property landscapes rapidly, prioritizing candidates for synthesis and characterization. Increasingly, inverse design frameworks propose candidate materials that satisfy predefined targets, transforming discovery from trial-and-error to model-guided exploration.

The most powerful paradigm emerges when AI and experimentation coevolve: experiments refine the model; the model guides the next experiments; the loop accelerates discovery and improves reproducibility (Figure 1.16).

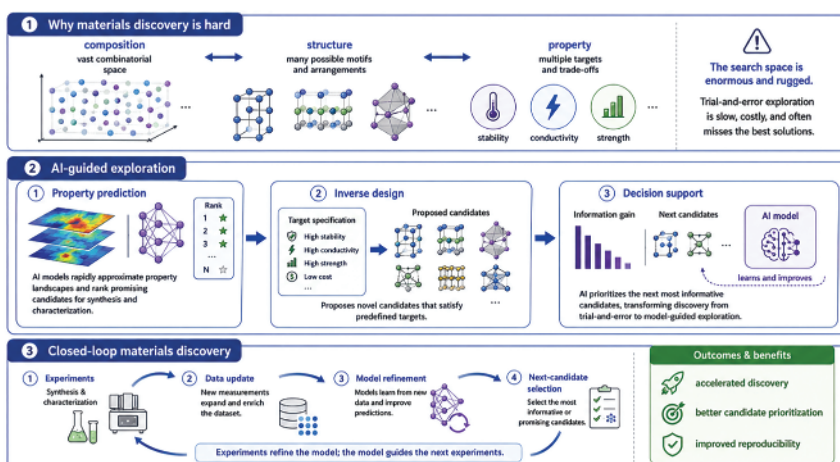


Figure 1.16 AI in materials discovery. A pipeline linking candidate generation, property prediction, synthesis, and validation [7, 14].

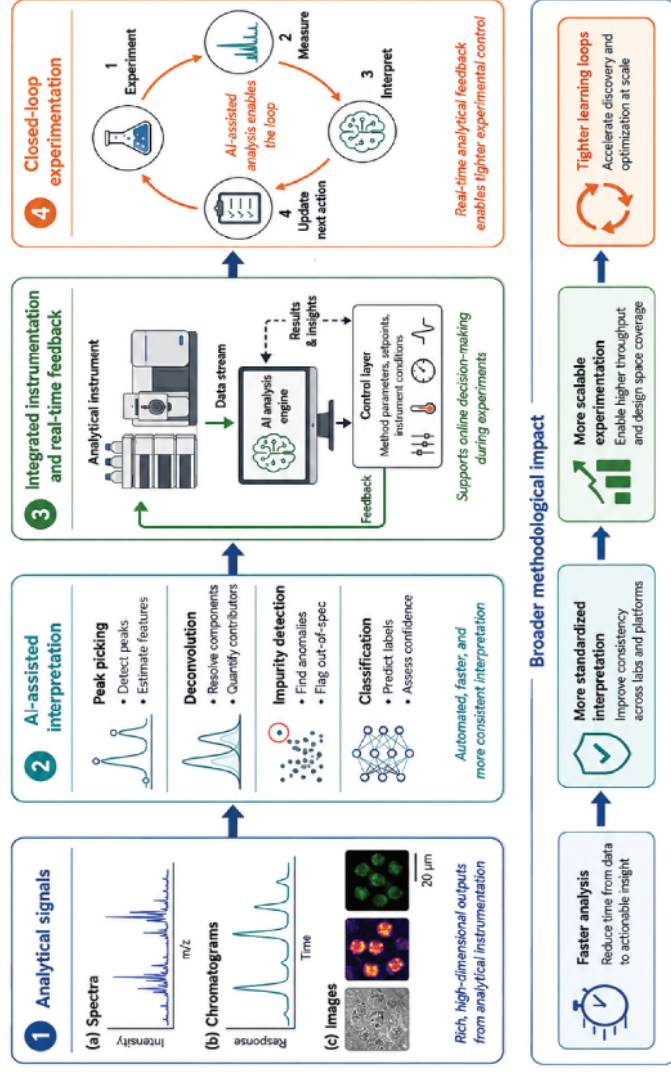


Figure 1.17 AI for spectroscopy and analytics. A diagram showing how AI converts raw signals into structured chemical information and decisions [19, 20].

1.3.5 AI for Spectroscopy and Analytical Chemistry: From Signals to Structures

Analytical chemistry produces rich signals—spectra, chromatograms, images—whose interpretation can be labor-intensive and expert-dependent. AI can automate peak picking, deconvolution, impurity detection, and classification, enabling faster and more consistent interpretation. When integrated into instrumentation, AI supports real-time feedback—an essential capability for closed-loop experimentation.

The broader implication is methodological: as analysis becomes faster and more standardized, experimentation becomes more scalable, and learning loops become tighter (Figure 1.17).

1.4 From Prediction to Action: Closed-Loop Chemistry and Human–AI Collaboration

1.4.1 Self-Driving Laboratories and Closed-Loop Discovery

The most transformative vision of AI in chemistry is closed-loop discovery, where models and experiments form a continuous cycle. Self-driving laboratories combine automation (to execute experiments reliably) with AI (to decide what to do next). In such systems, AI does not merely analyze; it plans. The laboratory becomes an adaptive engine that explores chemical space systematically, improving with experience.

Closed-loop workflows matter because they address chemistry's bottleneck: the gap between what is possible to imagine and what is feasible to test. By coupling algorithmic prioritization with automated execution, discovery becomes scalable and reproducible (Figure 1.18).

1.4.2 Integrating AI with High-Throughput Experimentation: The Data Engine of Modern Chemistry

HTE provides the experimental capacity to explore multidimensional spaces. AI provides the interpretive and decision-making capacity to make that exploration efficient. Together, they form a new research primitive: standardized data generation coupled to adaptive prioritization.

In this paradigm, the objective is not simply “more experiments,” but more informative experiments. The tight coupling of HTE and AI enables structured datasets, better comparability across campaigns, and faster translation from observations to actionable insights (Figure 1.19).

1.4.3 Limitations, Bias, and Interpretability: Why Chemical AI Must Be Trustworthy

Despite rapid progress, chemical AI has fundamental limitations. Models learn from data distributions; when deployed beyond those distributions, they can

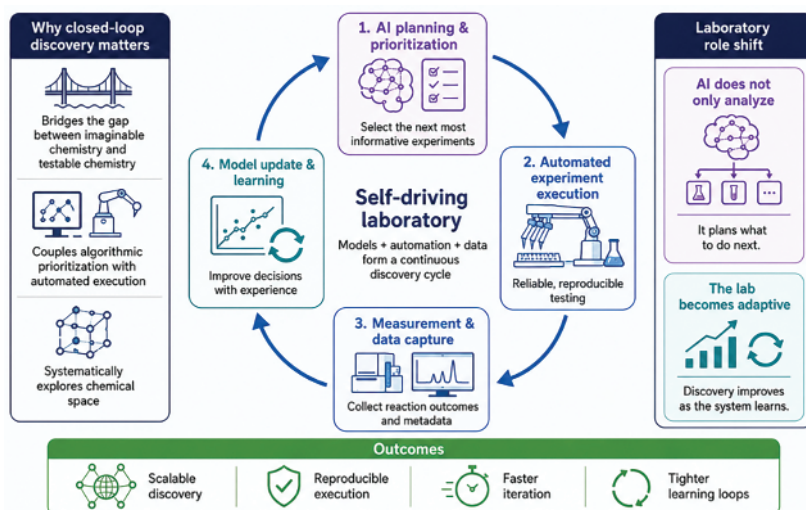


Figure 1.18 Self-driving laboratory. A closed-loop schematic linking robotics, experimentation, analytics, modeling, and decision-making [2, 14, 15].

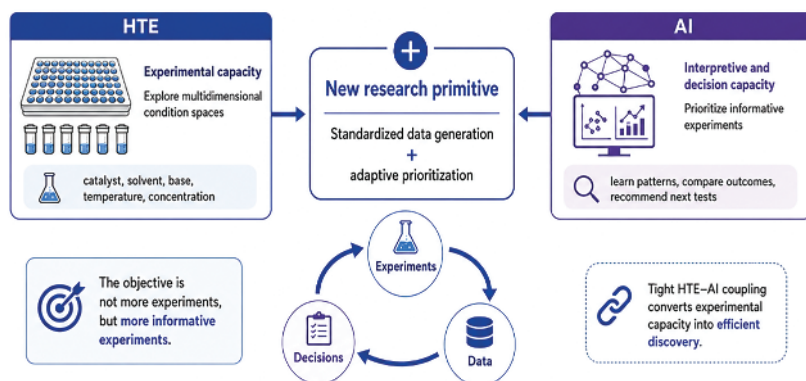


Figure 1.19 AI + HTE integration. A workflow showing parallel experimentation feeding models that recommend subsequent experiments [9, 13, 14].

fail silently. Biases in literature reporting, proprietary datasets, or experimental conventions can distort model behavior. Deep models may be difficult to interpret mechanistically, creating tension between predictive utility and scientific understanding.

For chemistry, the key principle is not blind trust but disciplined use. AI outputs should be treated as hypotheses—actionable but provisional—requiring validation, error analysis, and context-aware interpretation. Trustworthy AI in chemistry depends on data quality, transparent workflows, and careful alignment between the question asked and the evidence available (Figure 1.20).

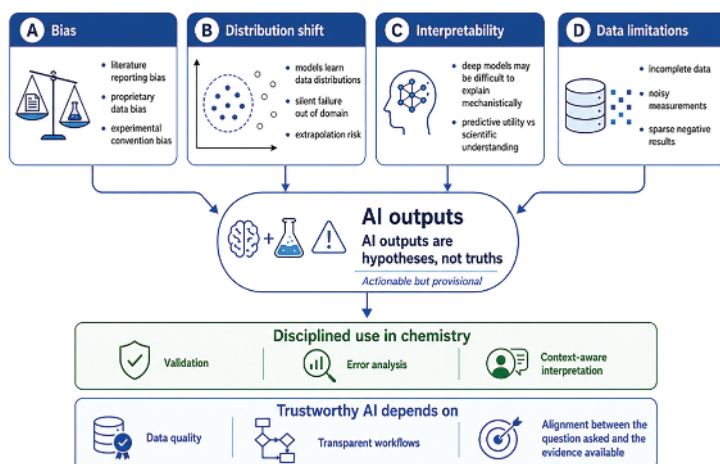


Figure 1.20 Challenges in chemical AI. An infographic highlighting bias, distribution shift, interpretability, and data limitations [16–18].

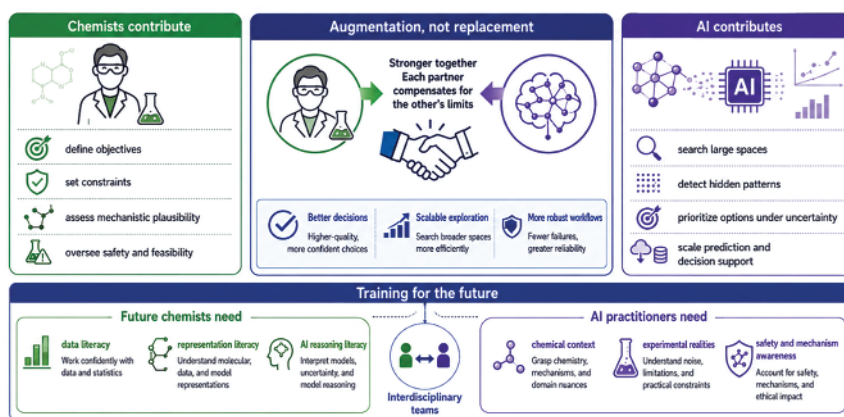


Figure 1.21 Human–AI collaboration. A diagram showing complementary roles: human judgment and AI-scale decision support [14–16].

1.4.4 Human–AI Collaboration: The Chemist's Role in the Age of AI

AI changes chemistry most effectively when it augments, rather than replaces, human expertise. Chemists provide objectives, constraints, mechanistic plausibility, and safety oversight. AI provides scale: it searches large spaces, detects patterns, and prioritizes decisions under uncertainty. The synergy is powerful because each partner compensates for the other's limitations.

This collaboration also changes training needs. Future chemists will benefit from literacy in data, representations, and AI reasoning, while AI practitioners must learn chemical context and experimental realities. The future belongs to interdisciplinary teams—and to chemists who can operate fluently at the interface of molecules, data, and machines (Figure 1.21).

1.5 Trends, Frontiers, and Outlook: Toward Intelligent Chemical Science

1.5.1 Current Trends: From Point Solutions to End-to-End Workflows

A clear trend in AI-driven chemistry is the transition from isolated predictive models to integrated workflows that connect data generation, modeling, and experimentation. AI is increasingly embedded in the research process as a planning and decision component, enabling adaptive experimentation, automated analysis, and rapid iteration. At the same time, emphasis is shifting toward robustness—models that work reliably under real laboratory variability.

1.5.2 Generative Models and Chemical Creativity

Generative AI has expanded the scope of what models can do: from evaluating candidates to proposing them. By learning distributions over chemical structures and transformations, generative models can suggest molecules, materials, and routes that satisfy predefined objectives. In practice, this is best viewed as “creativity at scale”: expanding the candidate pool far beyond what humans would enumerate, while relying on chemists to enforce feasibility and relevance.

1.5.3 AI for Sustainable and Green Chemistry

Sustainability is becoming a defining constraint in chemical innovation. AI supports green chemistry by enabling multi-objective decision-making—balancing yield and selectivity with solvent choice, energy input, waste reduction, and safety. When paired with automated experimentation and standardized metrics, AI enables systematic exploration of greener processes rather than incremental trial and error.

1.5.4 Precision Chemistry: From Molecules to Personalized Interventions

The convergence of chemical design with biological and clinical data is enabling more precise interventions—therapeutics, diagnostics, and delivery systems tailored to specific contexts. AI provides the integrative layer that links chemical structure to biological complexity, enabling faster iteration and improved targeting. The challenge—and opportunity—lies in connecting chemical models to real-world constraints: manufacturability, safety, and regulatory expectations.

1.5.5 Automation, Robotics, and Digital Laboratories

Digital laboratories—where experimental execution, tracking, and analytics are integrated—will increasingly define how data are produced and reused. AI becomes the coordinating intelligence, turning high-throughput capability into high-throughput learning. The near-term frontier is not fully autonomous chemistry but robust semiautonomous systems that reliably accelerate discovery while preserving human oversight.

1.5.6 Open Science, Standards, and Global Platforms

AI benefits disproportionately from shared datasets, benchmarks, and standardized reporting. Open science accelerates method development and comparability, while standards improve reproducibility and data reuse. At the same time, chemistry faces tensions between openness and proprietary value. Navigating these tensions responsibly—through governance, quality control, and equitable access—will shape the pace and direction of progress.

1.5.7 Ethical, Educational, and Societal Implications

As AI influences decisions with real-world impact—pharmaceutical development, environmental remediation, and industrial manufacturing—ethical concerns become unavoidable. Bias, transparency, accountability, and misuse risk must be addressed through both technical methods and institutional practices. Education must evolve accordingly: AI literacy will become part of core chemical training, just as spectroscopy and kinetics are today. (A detailed treatment of ethics and best practices is provided in Chapter 15.)

1.5.8 Outlook: Toward an Intelligent Chemical Science

AI is redefining chemistry as an information-to-decision science. The most important shift is not that models can predict but that workflows can learn: experiments generate standardized data, models improve, and the next experiments are chosen more intelligently. This feedback loop transforms the pace, scale, and reproducibility of chemical discovery.

Together, these developments define the central thesis of this book: AI in chemistry is not merely a collection of predictive tools, but an emerging scientific workflow in which data, models, automation, and human judgment are integrated to support reliable chemical decision-making.

The remainder of this book moves from landscape to mechanics and practice. Chapter 2 introduces the essential concepts of machine learning required for chemists to engage critically with AI tools. Subsequent chapters develop data science foundations, chemical representations, automation integration, and domain-specific applications—culminating in practical exercises and best practices for responsible deployment.

References

- 1 Turing, A.M. (1950). Computing machinery and intelligence. *Mind* 59: 433–460.
- 2 Wiener, N. (1948). *Cybernetics: or Control and Communication in the Animal and the Machine*. Cambridge, MA: MIT Press.
- 3 Russell, S. and Norvig, P. (2021). *Artificial Intelligence: A Modern Approach*, 4th ed. Pearson: Hoboken, NJ.

- 4 Lindsay, R.K., Buchanan, B.G., Feigenbaum, E.A., and Lederberg, J. (1993). DEN-DRAL: a case study of the first expert system for scientific hypothesis formation. *Artificial Intelligence* 61: 209–261.
- 5 LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521: 436–444.
- 6 Vamathevan, J., Clark, D., Czodrowski, P. et al. (2019). Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery* 18: 463–477.
- 7 Butler, K.T., Davies, D.W., Cartwright, H. et al. (2018). Machine learning for molecular and materials science. *Nature* 559: 547–555.
- 8 David, L., Thakkar, A., Mercado, R., and Engkvist, O. (2020). Molecular representations in AI-driven drug discovery: a review and practical guide. *Journal of Cheminformatics* 12: 56.
- 9 Shields, B.J., Stevens, J., Li, J. et al. (2021). Bayesian reaction optimization as a tool for chemical synthesis. *Nature* 590: 89–96.
- 10 Wigh, D.S., Goodman, J.M., and Lapkin, A.A. (2022). A review of molecular representation in the age of machine learning. *WIREs Computational Molecular Science Journal* 12: e1603.
- 11 Schwaller, P., Vaucher, A.C., Laplaza, R. et al. (2022). Machine intelligence for chemical reaction space. *WIREs Computational Molecular Science Journal* 12: e1604.
- 12 Segler, M.H.S., Preuss, M., and Waller, M.P. (2018). Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* 555: 604–610.
- 13 Pomberger, A., Mo, Y., Nandiwale, K.Y. et al. (2022). The effect of chemical representation on active machine learning towards closed-loop optimization. *Reaction Chemistry and Engineering* 7: 1368–1379.
- 14 Tom, G., Schmid, S.P., Baird, S.G. et al. (2024). Self-driving laboratories for chemistry and materials science. *Chemical Reviews* 124: 9633–9732.
- 15 Lo, S., Baird, S.G., Schrier, J. et al. (2024). Review of low-cost self-driving laboratories in chemistry and materials science: the “Frugal Twin” concept. *Digital Discovery* 3: 842–868.
- 16 Gallegos, M. et al. (2024). Explainable chemical artificial intelligence from accurate machine learning of real-space chemical descriptors. *Nature Communications* 15: 4345.
- 17 Crusius, D., Cipcigan, F., and Biggin, P.C. (2025). Are we fitting data or noise? Analysing the predictive power of commonly used datasets in drug-, materials-, and molecular-discovery. *Faraday Discuss.* 256: 304–321.
- 18 Jiang, J., Zhang, C., Ke, L. et al. (2025). A review of machine learning methods for imbalanced data challenges in chemistry. *Chemical Science* 16: 7637–7658.
- 19 Westermayr, J. and Marquetand, P. (2025). Machine learning spectroscopy to advance computation and analysis. *Chemical Science* 16: 21660–21676.
- 20 Flanagan, A.R., Dalal, D., and Glavin, F.G. (2025). Exploring generative artificial intelligence and data augmentation techniques for spectroscopy analysis. *Chemical Science* 125: 6130–6155.