PERSPECTIVE

Check for updates

The agentic age of predictive chemical kinetics

Alon Grinberg Dana [ID] *[ab]

Predictive chemical kinetic modeling is foundational to areas ranging from energy and environmental science to pharmaceuticals and advanced materials. While significant progress has been made in automating individual steps, the development of a complete predictive model remains a human-intensive effort to orchestrate existing software tools and revise models. This perspective outlines a practical path to improved chemical kinetic model development using agentic AI. A dual-lane architecture is introduced: a fast execution lane handles mechanism generation and parameter refinement, while a deliberative agentic lane plans, refines, and revises while executing experiments and computations. The proposed outcome is a robust pathway toward decision-grade models. Humans remain central: researchers set objectives and priors, approve high-impact actions, and adjudicate new chemical insights. Creativity, complex judgment, and strategic thinking remain in the human domain. Ultimately, this approach aims to accelerate trustworthy, transparent, decision-grade model development.

1. Introduction: from static workflows to agentic orchestration

The grand challenge of chemical kinetics, to accurately predict reaction outcomes across vast parameter spaces, has long been constrained by the limitations of static, human-directed workflows. Historically, much of chemical kinetics has been post-dictive: mechanisms were assembled and tuned to reproduce known measurements. By contrast, predictive chemical kinetics constructs mechanisms *a priori*, from first-principles estimates and established structure–reactivity relationships, and then uses them to simulate new conditions with quantified uncertainty rather than tuning to match data. This shift matters for design and safety: predictive models support decisions in regions where data are sparse, as long as their uncertainty is honest and traceable. Contemporary practice already includes partially automated pipelines that expand reaction networks and estimate properties at scale; the question we ask in the field is how to orchestrate these components so that the overall process becomes reliably decision-grade (Box 1, also defined in Section 2).

In chemical kinetic modeling, automation executes a standard flow: seed a core set of species and conditions; grow an edge network *via* reaction templates and promote influential species/reactions to the core by flux or rule criteria; assign and refine thermochemistry and rate coefficients from databases, estimation rules, or *ab initio* calculations; assemble pressure-dependent kinetics *via* master-equation treatments; and

simulate reactors (*e.g.*, shock tube, jet-stirred reactor, flames) to evaluate targets such as ignition delay time, flame speed, and speciation. Termination tolerances cap model growth. These pipelines are powerful but static by design: their logic and scope are fixed in advance, even when they include embedded classical AI components (Fig. 1).

A long-held goal for artificial intelligence (AI) is developing systems that can make major scientific discoveries and autonomously acquire knowledge. Although this "AI scientist" concept remains aspirational, progress in agent-based AI (Box 1) is a practical step toward that vision: conversational, reflective agents that plan and reason while orchestrating large language models (LLMs),[1–4] conventional machine learning (ML) tools,[5,6] experimental lab platforms – or even all at once.

An agent, by contrast to AI components such as ML tools, is dynamic and goal-directed. It perceives, decides, and acts to achieve an explicit performance target. Agents are, in essence, compound AI systems that coordinate capabilities rather than focusing on a single model,[7] shifting the core question of "What's the next best action?" from being pre-set once in a pipeline to being assessed afresh at every cycle.

This vision has a lineage in predictive chemical kinetics. Twenty years ago, Frenklach *et al.* sketched a process informatics system (PrIMe) in which chemists, engineers, and policymakers would query an intelligent bot and receive uncertainty-aware responses grounded in curated data and models.[8] PrIMe—proposed but never fully realized—prioritized rigorous curation and explicit uncertainty, yet stopped short of autonomous planning and orchestration. Today, mature automated toolchains are callable as services; cloud-native orchestrators and lab robotics can be integrated; and rapid progress in ML and the rise of agentic AI (Box 1) systems in recent years[9] make tool-using agents that plan against explicit objectives and

a Wolfson Department of Chemical Engineering, Technion – Israel Institute of Technology, Haifa 3200003, Israel. E-mail: alon@technion.ac.il

b Grand Technion Energy Program (GTEP), Technion – Israel Institute of Technology, Haifa 3200003, Israel

Box 1: Key terms used in this perspective

AI agent: autonomous software entity for a bounded task that perceives, decides, and acts; typically single-task scope.

Agentic AI: goal-directed orchestration that plans, selects tools, and adapts under budget, safety, and approval constraints to meet objectives.

UQ: uncertainty quantification that measures parameter and prediction uncertainty; includes propagation to targets.

Decision-grade: decision-grade models have validated uncertainty, end-to-end reproducibility, replayable provenance, and are auditable against benchmarks.

Disagreement signal: summarizes where predictions violate benchmarks relative to uncertainty, including earliest divergence and implicated observables.

Provenance: the complete record of inputs, versions, settings, approvals, costs, and hashes so runs are auditable and replayable.

Data credence: weights data by provenance, quality, and internal consistency; used to prioritize evidence in planning.

HITL: human-in-the-loop approvals required before high-cost, safety-critical, or irreversible actions (*e.g.*, high-cost compute campaigns).

Budget envelope: declares limits on compute, lab cost, and time; stop-rules pause or terminate when thresholds are reached.

Orthogonal experiment: uses a balanced, unconfounded design so factor effects are independent and information added is non-redundant.
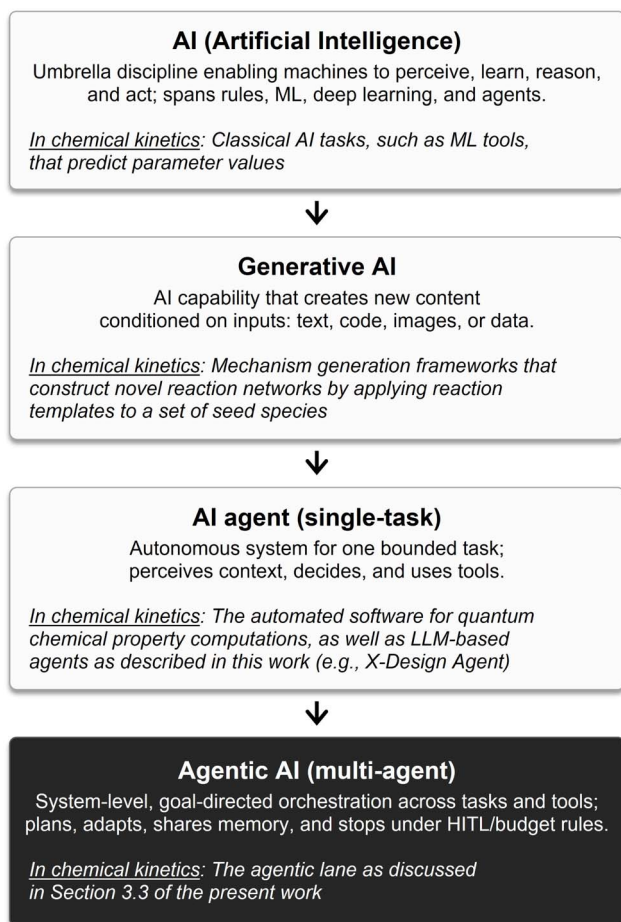

Fig. 1 Progression of AI autonomy and scope. AI refers to the broad field of computer science that aims to create intelligent machines capable of performing tasks that clasically require human intelligence. Generative AI is a capability to create content but does not plan. An AI Agent shows bounded autonomy, completing one specific task. In contrast, Agentic AI demonstrates orchestrated autonomy, decomposing a complex goal to coordinate multiple tasks and tools end-to-end (*e.g.*, plan, adapt, route HITL, stop on budget).

adapt to disagreement signals (Box 1) practical. Our agentic approach supplies this missing layer, extending PrIMe's vision to a goal-seeking system that proposes, executes, and revises.

We use agentic in a precise, field-specific sense defined in Section 2. Any such system must be governed by construction,

incorporating human-in-the-loop (HITL, Box 1) approvals at costly compute and wet-lab gates, planning within explicit budget envelopes (Box 1), and incorporating end-to-end provenance (Box 1) so outcomes are replayable and auditable. The payoff is concrete. An agentic system can weigh whether the next best action is to refine a small set of rate coefficients, propose an orthogonal speciation or global-parameter experiment (Box 1), search the literature for missing targets, or revise model structure, choosing the option with the best expected uncertainty reduction per unit cost. When its world changes, *e.g.*, new data or new constraints arrive, the plan changes with it.

Predetermined execution should transition to goal-seeking orchestration. This perspective formalizes agentic usage for predictive chemical kinetics, walks through a reference architecture that runs side by side with today's automation schemes, and outlines an evaluation path suited to a field that values transparency, reproducibility, and decision-grade prediction. This approach offers a practical path for the chemical kinetics community to transition from automated workflows to a new era of goal-directed, adaptive chemical kinetic model development.

2. Principles of agentic chemical kinetics

2.1. Definitions

We use agentic in a field-relevant sense: a goal-directed orchestrator that plans and coordinates existing automated computational tools so that each step advances clearly stated objectives under explicit constraints. Concretely, an agentic controller (i) pursues explicit performance objectives, *e.g.*, drives per-parameter and propagated prediction uncertainty below agreed thresholds under a budget, (ii) orchestrates today's automated pipelines and vetted tools, and (iii) adapts its plan as disagreement signals or data credence (Box 1) and budgets change. This follows the notion of an intelligent (rational) agent – an entity that perceives, decides and acts to achieve goals under a performance measure – and the recent emphasis on compound AI systems that integrate multiple capabilities rather than centering on a single model.[7,9]

In the broader taxonomy, we distinguish: AI (the umbrella discipline), generative AI (models that synthesize content in
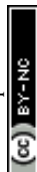
response to prompts), an AI agent (an autonomous software entity designed for a bounded task within a linear workflow), and agentic AI (systems where multiple specialized agents coordinate dynamically to pursue higher-level goals by replanning and reevaluating their actions based on new information).[10,11] Fig. 1 annotates each tier with canonical kinetics operations, clarifying the autonomy progression. Agentic AI departs from a single-agent approach by adding multi-agent collaboration, dynamic task decomposition, persistent memory, and coordinated autonomy (Fig. 2). Reflective reasoning and memory allow agents to evaluate past choices and refine strategies over time.[11–14]

HITL refers to approvals required before high-cost compute campaigns or wet-lab actions. Budget envelopes are constraints that bind computation and laboratory spending. Provenance denotes the record of inputs, tool versions, settings, and costs of key artifacts, ensuring outcomes are replayable and auditable. A disagreement signal summarizes where and when predictions violate benchmarks considering uncertainties, identifying the earliest divergence and the smallest set of implicated observables. Data credence is a metric that evaluates the trustworthiness of data based on its provenance, internal consistency, reproducibility, and associated metadata. Decision-grade denotes a model that is benchmarked, reproducible, provides validated uncertainty estimates, and has complete, replayable provenance.

2.2. Complementing automation

The agentic layer complements rather than replaces the existing automation backbone of the various software suites used in the chemical kinetics modeling community. The present perspective formalizes a two-lane picture: one lane specializes in
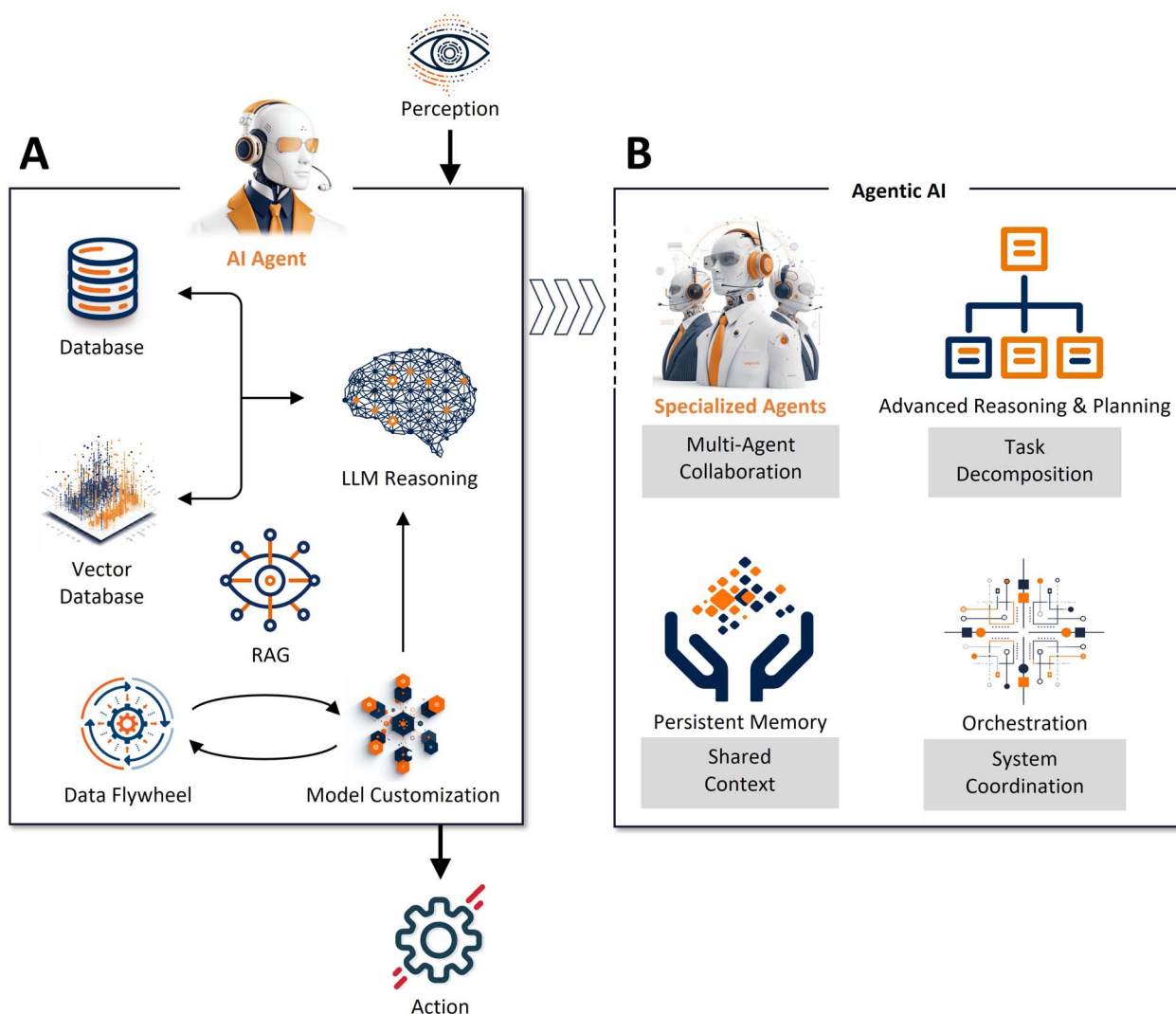

Fig. 2 From single-task agents to agentic systems. (A) Single AI agent: a perception–reasoning–action loop with Retrieval-Augmented Generation (RAG)[15] grounding, similarity search, and persistent memory; applies LLM-based reasoning with optional model customization; writes outcomes back to data/model artifacts (a "data flywheel") and then takes an action. Optimized for a specific, bounded task with high autonomy. Adapted from ref. 16. (B) Agentic AI system: a coordinated ensemble of specialized agents with advanced reasoning and planning, persistent memory with shared context, and system-level coordination; designed to pursue multi-step, higher-complexity goals. Adapted from ref. 9.

automation, the other is an agentic system which decides when/why/what next. The agentic orchestrator seeks to bring target observables inside an acceptable prediction-uncertainty threshold, while respecting compute and laboratory constraints. The automated pipeline remains the callable engine for mechanism generation, thermodynamic and kinetic computations, master equation (ME) solutions, and reactor modeling; the agentic layer decides which of these to invoke, what parameters to send the automated lane, and when to propose literature reconciliation, experiments, or model revision. Automation provides the gears, while agentic AI is the driver.

2.3. The agentic decision loop

At each cycle the orchestrator (i) assesses evidence – current residuals against benchmarks, the status of parameter- and prediction-uncertainty thresholds, the credence of available data, the remaining budget, and approvals; (ii) enumerates options – refine a set of rate coefficients or thermodynamic inputs, call the automated compute path, design an orthogonal experiment (providing non-redundant information) to disambiguate hypotheses, search or reconcile literature, propose a concrete model revision, or stop; (iii) forecasts value *vs.* cost, estimating expected uncertainty reduction per unit cost and time for each option; (iv) selects and routes through required HITL gates; (v) records provenance so outcomes are auditable; and (vi) re-plans as new data, costs, expert human input, or disagreements arrive.

For instance, suppose new shock-tube measurements at elevated pressure disagree with predicted ignition delay times. The agentic layer flags failure at the "model agrees with benchmarks?" gate (Fig. 3), ranks next actions by expected uncertainty reduction per cost, and proposes: (a) targeted refinement of two dominant rate coefficients by sensitivity analysis (SA), (b) refinement of thermodynamic properties of all species on the fuel + O_2 potential energy surface, and (c) a small orthogonal experiment to choose between competing pathways. The experiment hits a HITL gate and obtains approval, returns structured data to the curated store with high credence, the automated pipeline reruns, and the gates are re-evaluated.

2.4. The role of retrieval-augmented generation (RAG)

Retrieval-Augmented Generation (RAG)[15] is a technique that improves the accuracy and relevance of LLMs by giving them access to up-to-date external information. Without RAG, an LLM generates a response based only on the data on which it was trained, which might lead to a high frequency of hallucinations. RAG combines a retriever over curated sources with a generator (often an LLM): before proposing or explaining an action, the system retrieves relevant documents, passages, or data tables, and grounds its reasoning in those materials, often quoting or citing the snippets that informed the step.[15] In our setting, RAG surfaces candidate rate coefficients, thermochemistry values, prior mechanism fragments, known benchmark targets and operating envelopes, and context such as experimental apparatus constraints or calibration practices;

these returns feed a Data Agent that assigns credence (provenance, quality, consistency) and de-duplicates before anything influences further planning or tool selection.

2.5. Implications for predictive chemical kinetics

This field already has the ingredients that make an agentic controller high-leverage: (i) a mature automated backbone (template-driven reaction-network growth; *ab initio* parameter refinement) that can be called as a service; (ii) quantitative performance measures (residuals between predictions and benchmark targets; parameter- and prediction-uncertainty thresholds); (iii) actionable levers (which parameters to refine; which pathway assumptions to revise; which orthogonal, discriminating experiments to run; how to adjust model generation tolerances, constraints, and termination criteria); and (iv) binding budget, safety, and governance constraints. In combination, these features turn planning into a well-posed decision problem instead of a fixed script. In practice, we contend with large hypothesis spaces (competing pathways and pressure–temperature-composition regimes), heterogeneous evidence from simulations and experiments, and shifting requirements. An agentic layer is a natural fit: its job is to decide, repeatedly, what action to take to satisfy the "model agreement with benchmarks" gate under a budget. It stops when additional actions offer too little expected uncertainty reduction per unit cost, or when target benchmarks and prediction uncertainty have been reached.

This framing follows standard agent definitions, but it is localized to the levers that matter here: which parameters to touch, which data to trust, which operating points to probe, and when to terminate. Because provenance, HITL approvals, and budget envelopes are embedded by construction, the result is envisioned to be a transparent, auditable process that aligns with the community's expectations for reproducibility and decision-grade prediction.

3. Reference architecture: the dual-lane design

3.1. Dual-lane overview (System 1/System 2)

We adopt a dual-lane design: a fast, validated automation lane (S1) embedded in today's toolchains, alongside a deliberative, goal-seeking lane (S2) that decides when, why, and what to do next under stated constraints (Fig. 3). This organization echoes the dual-process view of judgment and decision making originating in Kahneman and Tversky's heuristics-and-biases program,[17] later popularized as "System 1" and "System 2".[18,19] S1 plays the role of a dependable "System 1": it executes a vetted sequence quickly and repeatedly with minimal discretion. S2 acts as a deliberative "System 2": it reasons over objectives, uncertainties, budgets, and approvals to select the next best action. The two lanes are coupled, not competitive. S1 remains the execution backbone, while S2 decides when and why to invoke S1, what else to do, and when to stop. All choices traverse explicit gates and emit provenance so that outcomes are reproducible and auditable. At the outset, the user defines
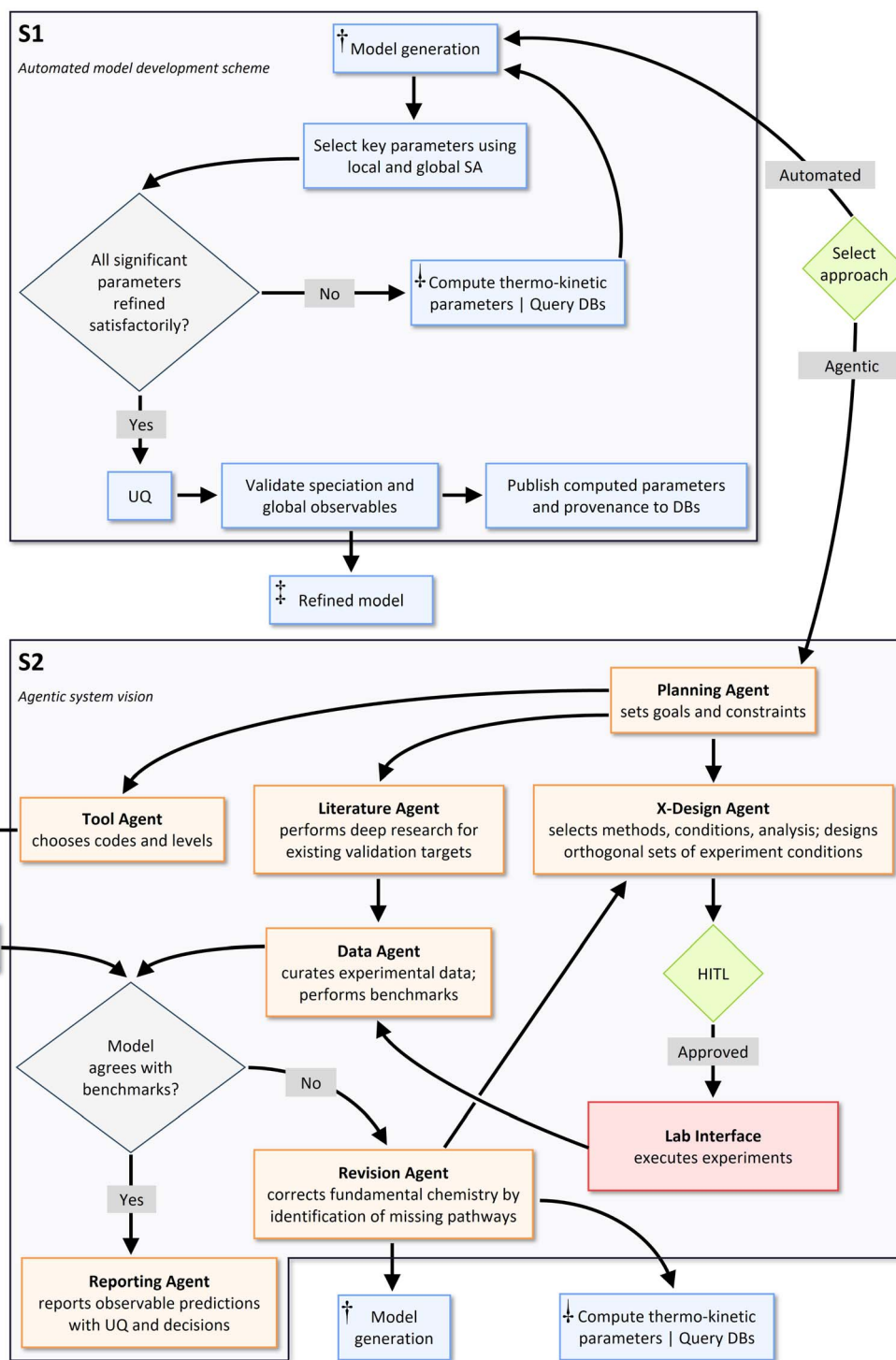
Fig. 3 Dual-lane architecture for predictive chemical kinetics. S1 (top) is the established automated model development backbone. S2 (bottom) is the agentic layer running alongside S1. Arrows indicate data flow; colors denote roles (blue: S1 tools; orange: agents; red: wet-lab; green/gray diamond: human- or machine-controlled gate, respectively). Dagger symbols represent identical modules.

scope, thermodynamic conditions, model targets, and compute and experiment budgets, then chooses between a classical execution of S1 only ("Automated"), or a nearly autonomous execution of model development and revision combining S1 and S2 ("Agentic"), as shown in Fig. 3.

3.2. Automated backbone (S1)

Across multiple chemical kinetic domains, the enabling infra-structure has turned tedious work into streamlined, automated workflows that deliver refined, predictive models. S1 comprises three sets of complementary automated capabilities: model

generation (exploring the chemical space), parameter refinement (reducing uncertainties in thermochemistry and rate coefficients) and model development (a high-level routine) that links generation and refinement in an iterative loop.

Frameworks such as RMG[20,21] and Genesys[22,23] systematically construct reaction networks using reaction-family templates and parameter estimation schemes. A second class of tools computes higher-fidelity thermochemistry and rate coefficients for key steps identified by SA, replacing database estimates with first-principles values at scale. Such open source tools include EStokTP,[24] AutoTST,[25] ChemTraYzer,[26,27] KinBot,[28] AutoMech,[29] and ARC.[30] Statistical mechanics tools convert the quantum chemical computations into thermodynamically and kinetically relevant parameters.[31–34] Higher level routines such as T3,[35] The Tandem Tool for automated chemical kinetic model development, link generation and refinement into a single, iterative development loop. Cantera[36] provides a robust platform for simulating reacting systems with detailed mechanisms. In parallel, community efforts toward a FAIR database and a sharing schema for computed parameters are underway. Operationally, S1 remains code-agnostic: it uses vetted tools that provide a mature and modular backbone that already exists.

3.3. Agentic workflow (S2)

The agentic lane (S2) sits alongside the automated backbone (S1) and acts as a goal-directed orchestrator implemented as a small ensemble of specialized agents coordinated through shared memory and schema-based messages (Fig. 2 and 3). Planning and tool selection come first. A **Planning Agent** encodes objectives (*e.g.*, bring observables within a specified uncertainty threshold and benchmark against measurements), sets constraints (resource quotas), and checks policy gates requiring HITL approvals such as large compute campaigns (not depicted) and wet-lab experiments (Fig. 3). It quantifies the objectives to enable a formal optimization problem. To achieve this, the **Planning Agent** ranks computations by significance, which is determined by a combination of SA and uncertainty quantification (UQ, Box 1). It estimates the expected uncertainty reduction per unit cost for each potential action, guiding its decision-making toward the most efficient path for model improvement.

A **Tool Agent** maps these goals onto concrete actions by choosing which vetted mechanism generator, quantum-chemistry workflow, ME solver, and reactor model to invoke, and at what fidelity (*e.g.*, level of theory). It uses action schemas with explicit field types, measurement units, required fields, ranges/enumerated options (enums), cross-field constraints and validation, rather than free-form text to reduce risk of hallucinations.

Evidence is grounded before action. A **Literature Agent** quotes candidate thermochemical and kinetic data, prior models, transport parameters, apparatus notes, and experimental benchmark targets from curated/internal stores, trusted community repositories, and professional literature. A **Data Agent** ingests these returns and new measurements, assigns

credence, flags outliers, resolves duplicates, and materializes benchmark sets at stated operating envelopes.

When existing measurements are insufficient or competing hypotheses remain, S2 proposes experiments. The **X-Design Agent** enumerates discriminating designs using orthogonal working sets[37–39] of experimental working conditions for validation. A **Lab Interface** schedules and executes approved runs and returns structured results with metadata that the **Data Agent** immediately scores for credence and adds to benchmarks. The **Lab Interface** could be connected to self-driving labs,[40–43] or send structured requests to a lab engineer to carefully execute. Users need not have specialized equipment or operational expertise; the system can route to fee-for-service partner labs charged against the campaign's budget envelope.

S2 then closes the loop *via* revision. If the second gate, "model agrees with benchmarks (within uncertainties)?", fails, the **Revision Agent** selects and prioritizes appropriate actions and executes them. It first diagnoses the disagreement signals using flux and sensitivity analyses to isolate potentially responsible pathways. Next, the agent chooses among three strategic avenues. (i) Parameter-level refinement: target specific quantities that most influence the discrepancy, *e.g.*, refine thermochemistry under the discrepant conditions, refine key pressure-dependent networks, or compute high-fidelity rate coefficients for sparsely trained reaction classes; (ii) model scope and structure adjustments: broaden or correct the mechanism where coverage is lacking or misleading, *e.g.*, generating and benchmarking low-concentration single-reactant models to capture early unimolecular decomposition, probing alternative pressure–temperature windows to reduce confounding pressure-dependence, exploring potentially missing reaction paths,[44] identifying and excluding non-physical species,[45] or introducing chemically termolecular steps where warranted;[46,47] (iii) generation-process modifications: tune how the model is built, *e.g.*, resetting termination criteria, rebalancing enlargement tolerances, and tightening or relaxing species-generation constraints to avoid overgrowth or premature truncation, respectively. The agent ranks candidate actions by expected uncertainty reduction per unit cost, and given approvals routes the chosen action(s) through the automated computational backbone and the lab interface (Fig. 3).

The agentic system presents disagreement signals to the human expert, offering a ranked set of revision hypotheses annotated with expected uncertainty reduction per unit cost, required tools, and safety/budget implications. The human expert can request a what-if preview before committing compute or laboratory time. Qualitative guidance from the expert (*e.g.*, a missing reaction family) is converted into typed priors that bias or expand the search. The agent autonomously executes the most promising revisions and dynamically re-optimizes the plan on-the-fly in response to human insights.

If the gate passes, a **Reporting Agent** assembles a decision-grade package: predictions with propagated uncertainty, a changelog of parameters and pathways touched, budget spend *versus* plan, and full provenance. Throughout, S2 maintains a running estimate of marginal information gain per cost; if it

falls below policy thresholds, the system reframes or halts rather than chasing noise.

4. Risks, governance and mitigation

Agentic systems introduce risks that compound familiar failure modes. Orchestrated agents can enable ambitious automation and amplify weaknesses that affect trust, security, reliability, and interpretability.[11–13,48] These risks are tractable only with proactive design and governance.

The core limitations of these systems stem in part from the black-box chains of reasoning characteristic of LLM-based agents. Agents do not reliably anticipate the downstream impacts of their actions,[11] inviting ambiguity and drift. Chained workflows are especially sensitive to error cascades: a single faulty inference, from incorrect units, through a misparsed chemical structure, to brittle reasoning,[48] can propagate through planning, simulation, and experiment execution, yielding coordinated but wrong outcomes. Furthermore, shared resources can produce contention, duplication, and race conditions. Adding more agents to an agentic system often increases noise rather than capability, complicating debugging and making behavior non-compositional,[9] especially considering the lack of adopted formal verification for multi-agent LLM systems.

Agentic systems are exposed to adversarial threats; a compromised component (*e.g.*, prompt injection, poisoned data, or tool manipulation) can contaminate the shared state. Finally, foundations are early: standard architectures, protocols, and evaluation methods remain unsettled.[9] Unpredictable behavior, such as hallucinated reaction pathways, brittle reasoning, or 'emergent agency' that was never explicitly programmed, can surface at scale, introducing further safety risks.[49,50] For instance, an agentic system might decide to execute an expensive, risky, high-pressure, high-temperature experiment to resolve a minor uncertainty without properly weighing the cost–benefit trade-off.

Maintaining effective human control and accountability becomes harder as these systems make longer chains of decisions in high-stakes settings.[48] The complexity of multi-agent systems can blur accountability and diffuse authorship, especially when multiple agents and humans contribute to a single output. This also raises fairness and alignment questions, as different stakeholders may have conflicting goals, such as prioritizing speed over safety.

These risks are design constraints rather than deal breakers. Concretely, we must constrain action spaces, keep HITL at expensive compute and laboratory gates, enforce budget stop-rules, require complete provenance so decisions are replayable, and strengthen agent interfaces beyond free-form text using typed schemas and protocols. Agentic systems must be evaluated in offline replays (running on past projects with mock tool outputs with disabled intensive computations and wet-lab access) and in shadow mode (run in parallel to classical automation for real-time sanity validation) before live use. These steps do not eliminate risks, but convert a black-box into a transparent, auditable process where failures are bounded and observable.

5. Outlook

Agentic predictive chemical kinetics will succeed when models are truly decision-grade. Automation remains the fast, validated execution lane; the agentic layer plans and justifies next actions – curating external data, running benchmark measurements, proposing revision steps – and decides when to stop. Humans set the objectives and the boundaries, specify priors and acceptable risk, approve high-cost or hazardous actions, and adjudicate new chemical insights. Creativity, complex judgment, and strategic thinking remain in the human domain. Machines execute automation and disciplined decision-making at scale. This synergy will allow modelers to move efficiently from hypothesis to validated predictions.

Near-term, the community should standardize mechanism generation and parameter refinement loops to deliver validated models. Such studies should include: (i) prediction error and UQ calibration on agreed targets; (ii) parameter-level uncertainty reductions; (iii) total compute and lab cost; and (iv) complete, replayable provenance. Mid-term, intelligent agents should integrate laboratory robotics and partner labs into closed-loop cycles where models propose experiments, experiments update models, and policy stop-rules enforce budget and safety.

A next step is to launch a community-agreed Mechanism Development Benchmark with open datasets and mock tools for offline replay. Include a scoreboard that weights predictive accuracy, UQ calibration, provenance completeness, cost, latency, and HITL approvals. Pair it with a shared schema for mechanisms, data, and decision logs, and run frequent evaluations with offline replay and shadow-mode tracks.

If these standards are adopted, predictive chemical kinetics will shift from handcrafted scripts to a transparent, adaptive, and community-auditable enterprise where agents, models, and experiments continually improve each other, allowing researchers to focus on insight, not orchestration.

Author contributions

The author carried out all aspects of this work, including conception, analysis, figure preparation, writing, and funding acquisition.

Conflicts of interest

There are no conflicts of interest to declare.

Data availability

No primary research results, software or code have been included and no new data were generated or analysed as part of this perspective.

Acknowledgements

References

1 D. A. Boiko, R. MacKnight, B. Kline and G. Gomes, Autonomous chemical research with large language models, *Nature*, 2023, **624**, 570–578.

2 A. M. Bran, S. Cox, O. Schilter, C. Baldassari, A. D. White and P. Schwaller, Augmenting large language models with chemistry tools, *Nat. Mach. Intell.*, 2024, **6**, 525–535.

3 Z. Xi, *et al.*, The Rise and Potential of Large Language Model Based Agents: A Survey, *Sci. China Inf. Sci.*, 2025, **68**, 121101.

4 T. Guo, X. Chen, Y. Wang, R. Chang, S. Pei, N. V. Chawla, O. Wiest and X. Zhang, Large Language Model based Multi-Agents: A Survey of Progress and Challenges, *arXiv*, 2024, preprint, arXiv:2402.01680, DOI: **10.48550/arXiv.2402.01680**.

5 M. R. Dobbelaere, P. P. Plehiers, R. Van de Vijver, C. V. Stevens and K. M. Van Geem, Machine Learning in Chemical Engineering: Strengths, Weaknesses, Opportunities, and Threats, *Engineering*, 2021, **7**, 1201–1211.

6 A. M. Schweidtmann, Generative artificial intelligence in chemical engineering, *Nat. Chem. Eng.*, 2024, **1**, 193.

7 M. Zaharia, O. Khattab, L. Chen, J. Q. Davis, H. Miller, C. Potts, J. Zou, M. Carbin, J. Frankle, N. Rao, *et al.*, The Shift from Models to Compound AI Systems, **https://bair.berkeley.edu/blog/2024/02/18/compound-ai-systems/**, 2024.

8 M. Frenklach, Transforming data into knowledge – Process Informatics for combustion chemistry, *Proc. Combust. Inst.*, 2007, **31**, 125–140.

9 R. Sapkota, K. I. Roumeliotis and M. Karkee, AI Agents vs. agentic AI: A conceptual taxonomy, applications and challenges, *Inf. Fusion*, 2026, **126**, 103599.

10 J. Ferber and G. Weiss, *Multi-Agent System: An Introduction to Distributed Artificial Intelligence*, Addison Wesley Longman, Harlow, Boston, 1999, vol. 1.

11 D. B. Acharya, K. Kuppan, B. Divya and A. I. Agentic, Autonomous Intelligence for Complex Goals—A Comprehensive Survey, *IEEE Access*, 2025, **13**, 18912–18936.

12 M. Gridach, J. Nanavati, K. Z. E. Abidine, L. Mendes and C. Mack, Agentic AI for Scientific Discovery: A Survey of Progress, Challenges, and Future Directions, 2025, *arXiv*, preprint, arXiv:2503.08979, DOI: **10.48550/arXiv.2503.08979**.

13 P. Bornet and J. Wirtz, *Agentic Artificial Intelligence*, World Scientific, Singapore, 2025.

14 P. R. Lewis and Ş. Sarkadi, Reflective Artificial Intelligence, *Minds Mach.*, 2024, **34**, 14.

15 P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, *et al.*, Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks, *Advances in Neural Information Processing Systems*, Red Hook, NY, 2020, pp. 9459–9474.

16 E. Pounds, NVIDIA Blog: What Is Agentic AI?, 2024, **https://blogs.nvidia.com/blog/what-is-agentic-ai/**.

17 A. Tversky and D. Kahneman, Judgment under Uncertainty: Heuristics and Biases, *Science*, 1974, **185**, 1124–1131.

18 K. E. Stanovich and R. F. West, Advancing the rationality debate, *Behav. Brain Sci.*, 2000, **23**, 701–717.

19 D. Kahneman, Experiences of Collaborative Research, *Am. Psychol.*, 2003, **58**, 723–730.

20 M. Liu, A. Grinberg Dana, M. Johnson, M. Goldman, A. Jocher, A. Payne, C. Grambow, K. Han, N. Yee, E. Mazeau, *et al.*, Reaction Mechanism Generator v3.0: Advances in Automatic Mechanism Generation, *J. Chem. Inf. Model.*, 2021, **61**, 2686–2696.

21 M. S. Johnson, X. Dong, A. Grinberg Dana, Y. Chung, D. Farina, R. J. Gillis, M. Liu, N. W. Yee, K. Blondal, E. Mazeau, *et al.*, RMG Database for Chemical Property Prediction, *J. Chem. Inf. Model.*, 2022, **62**, 4906–4915.

22 N. Vandewiele, K. VanGeem, M.-F. Reyniers and G. Marin, Genesys: Kinetic model construction using chemo-informatics for automated generation of reaction mechanisms, *Chem. Eng. J.*, 2012, **207**, 526–538.

23 Y. Ureel, L. Tomme, M. K. Sabbe and K. M. Van Geem, Genesys-Cat: automatic microkinetic model generation for heterogeneous catalysis with improved Bayesian optimization, *Catal. Sci. Technol.*, 2025, **15**, 750–764.

24 C. Cavallotti, M. Pelucchi, Y. Georgievskii and S. J. Klippenstein, EStokTP: Electronic Structure to Temperature- and Pressure-Dependent Rate Constants—A Code for Automatically Predicting the Thermal Kinetics of Reactions, *J. Chem. Theory Comput.*, 2019, **15**, 1122–1145.

25 P. L. Bhoorasingh, B. L. Slakman, F. Seyedzadeh Khanshan, J. Y. Cain and R. H. West, Automated Transition State Theory Calculations for High-Throughput Kinetics, *J. Phys. Chem. A*, 2017, **121**, 6896–6904.

26 M. Döntgen, M.-D. Przybylski-Freund, L. C. Kröger, W. A. Kopp, A. E. Ismail and K. Leonhard, Automated Discovery of Reaction Pathways, Rate Constants, and Transition States Using Reactive Molecular Dynamics Simulations, *J. Chem. Theory Comput.*, 2015, **11**, 2517–2524.

27 L. Krep, I. S. Roy, W. Kopp, F. Schmalz, C. Huang and K. Leonhard, Efficient Reaction Space Exploration with ChemTraYzer-TAD, *J. Chem. Inf. Model.*, 2022, **62**, 890–902.

28 J. Zádor, C. Martí, R. Van de Vijver, S. L. Johansen, Y. Yang, H. A. Michelsen and H. N. Najm, Automated Reaction Kinetics of Gas-Phase Organic Species over Multiwell Potential Energy Surfaces, *J. Phys. Chem. A*, 2023, **127**, 565–588.

29 S. N. Elliott, K. B. Moore, A. V. Copan, M. Keçeli, C. Cavallotti, Y. Georgievskii, H. F. Schaefer and S. J. Klippenstein, Automated theoretical chemical kinetics: Predicting the kinetics for the initial stages of pyrolysis, *Proc. Combust. Inst.*, 2021, **38**, 375–384.

30 A. Grinberg Dana, K. Kaplan, C. Pieters, D. Ranasinghe, H. Wu, C. Grambow, X. Dong, M. Johnson, M. Goldman, M. Liu and W. Green, ARC - Automated Rate Calculator,

version 1.1.0. https://github.com/ReactionMechanismGenerator/ARC, 2019.

31 A. Grinberg Dana, M. S. Johnson, J. W. Allen, S. Sharma, S. Raman, M. Liu, C. W. Gao, C. A. Grambow, M. J. Goldman, D. S. Ranasinghe, *et al.*, Automated reaction kinetics and network exploration (Arkane): A statistical mechanics, thermodynamics, transition state theory, and master equation software, *Int. J. Chem. Kinet.*, 2023, **55**, 300–323.

32 D. Glowacki, C.-H. Liang, C. Morley, M. Pilling and S. H. Robertson, MESMER: An open-source master equation solver for multi-energy well reactions, *J. Phys. Chem. A*, 2012, **116**, 9545–9560.

33 Y. Georgievskii, J. A. Miller, M. P. Burke and S. J. Klippenstein, Reformulation and Solution of the Master Equation for Multiple-Well Chemical Reactions, *J. Phys. Chem. A*, 2013, **117**, 12146–12154.

34 J. R. Barker, Multiple-Well, multiple-path unimolecular reaction systems. I. MultiWell computer program suite, *Int. J. Chem. Kinet.*, 2001, **33**, 232–245.

35 C. Pieters, K. Kaplan, K. Spiekermann, W. Green and A. Grinberg Dana, The Tandem Tool (T3) for automated chemical kinetic model development, https://github.com/ReactionMechanismGenerator/T3, Version 0.1.0, 2023.

36 D. Goodwin, H. Moffat, R. Speth, and B. Weber, Cantera: An Object-Oriented Software Toolkit for Chemical Kinetics, Thermodynamics, and Transport Processes, *Zenodo, ver. 2.4.0*, 2018.

37 G. Taguchi and Y. Wu, *Introduction to off-line quality control*, Central Japan Quality Control Association, Japan, 1985.

38 R. Davis and P. John, *Application of Taguchi-Based Design of Experiments. Statistical approaches with emphasis on design of experiments applied to chemical processes*, Croatia, 2018.

39 B. Jones and C. J. Nachtsheim, Blocking Schemes for Definitive Screening Designs, *Technometrics*, 2016, **58**, 74–83.

40 M. Seifrid, R. Pollice, A. Aguilar-Granda, Z. Morgan Chan, K. Hotta, C. T. Ser, J. Vestfrid, T. C. Wu and A. Aspuru-Guzik, Autonomous Chemical Experiments: Challenges and Perspectives on Establishing a Self-Driving Lab, *Acc. Chem. Res.*, 2022, **55**, 2454–2466.

41 J. A. Bennett and M. Abolhasani, Autonomous chemical science and engineering enabled by self-driving laboratories, *Curr. Opin. Chem. Eng.*, 2022, **36**, 100831.

42 M. Abolhasani and E. Kumacheva, The rise of self-driving labs in chemical and materials sciences, *Nat. Synth.*, 2023, **2**, 483–492.

43 G. Tom, S. P. Schmid, S. G. Baird, Y. Cao, K. Darvish, H. Hao, S. Lo, S. Pablo-García, E. M. Rajaonson, M. Skreta, *et al.*, Self-Driving Laboratories for Chemistry and Materials Science, *Chem. Rev.*, 2024, **124**, 9633–9732.

44 C. A. Grambow, A. Jamal, Y.-P. Li, W. H. Green, J. Zádor and Y. V. Suleimanov, Unimolecular Reaction Pathways of a γ-Ketohydroperoxide from Combined Application of Automated Reaction Discovery Methods, *J. Am. Chem. Soc.*, 2018, **140**, 1035–1048.

45 N. Mitnik, S. Haba and A. Grinberg Dana, Non-physical Species in Chemical Kinetic Models: A Case Study of Diazenyl Hydroxy and Diazenyl Peroxide, *ChemPhysChem*, 2022, **23**, e202200373.

46 M. P. Burke and S. J. Klippenstein, Ephemeral collision complexes mediate chemically termolecular transformations that affect system chemistry, *Nat. Chem.*, 2017, **9**, 1078–1082.

47 M. C. Barbet, K. McCullough and M. P. Burke, A framework for automatic discovery of chemically termolecular reactions, *Proc. Combust. Inst.*, 2019, **37**, 347–354.

48 A. Chan, R. Salganik, A. Markelius, C. Pang, N. Rajkumar, D. Krasheninnikov, L. Langosco, Z. He, Y. Duan, M. Carroll, *et al.*, Harms from Increasingly Agentic Algorithmic Systems, *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, New York, USA, 2023, pp. 651–666.

49 H. Hexmoor, J. Lammens, G. Caicedo and S. C. Shapiro, *Behaviour based AI, cognitive processes, and emergent behaviors in autonomous agents*, WIT Press, Ashurst, UK, 2025, vol. 1.

50 D. Trusilo, Autonomous AI Systems in Conflict: Emergent Behavior and Its Impact on Predictability and Reliability, *J. Mil. Ethics*, 2023, **22**, 2–17.