IntuitionLabs

INTUITIONLABS / RESEARCH & PRACTICAL GUIDANCE

AI Molecule Prioritization: Drug Discovery Triage Tools

> Build practical AI for pharma and biotech with IntuitionLabs. We help life-science teams turn complex information and workflows into useful software, governed knowledge systems and AI tools.

Published by IntuitionLabs · 2026-04-22 · ai molecule prioritization · drug discovery pipelines · computational triage · generative chemistry · admet predictions · multi-objective optimization · active learning · computational chemistry

Original article: https://intuitionlabs.ai/articles/ai-molecule-prioritization-drug-discovery-triage



In this article

1. Executive Summary
2. 1. Introduction and Background
3. 1.1 The Traditional Drug Discovery Pipeline
4. 1.2 AI and Generative Models in Drug Discovery
5. 1.3 Multi-objective Optimization in Drug Design
6. 2. Triage Strategies for AI-Generated Molecules
7. 3. Case Studies and Real-World Applications
8. 4. Implications and Future Directions
9. 5. Conclusion

Executive Summary

The advent of **artificial intelligence (AI)** and **deep generative models** has begun to transform early drug discovery, enabling the rapid proposal of novel small-molecule candidates from the vast drug-like chemical space (estimated at >10^60 compounds (www.sciencedirect.com)). However, this generative explosion creates a new bottleneck: *prioritizing* which AI-generated candidates to advance through costly experimental pipelines. This report investigates how “triage” tools—filtering, ranking, and selection algorithms—are used to manage and prioritize AI-proposed molecules in drug discovery workflows. We review the historical context of computational drug design, describe major AI-based generative architectures, and examine the multi-objective nature of drug candidate evaluation. We catalog the key properties used for triage (e.g. ADMET predictions, synthetic accessibility, structural novelty) and survey methods and tools used to compute them. We highlight case studies of AI-designed molecules (e.g. Insilico’s AI-designed TNIK inhibitor *rentosertib* (www.nature.com), Exscientia’s DSP-1181 and EXS21546 (www.cas.org) (www.cas.org), and Recursion’s phenomics-derived REC-994 (www.sciencedirect.com)) to illustrate real-world application. We also analyze data from the literature—for example, reports that AI-augmented screening can shorten hit-finding time (www.sciencedirect.com) and that active learning can find most “active” compounds by testing only a fraction of the space (pubs.rsc.org). Throughout, we emphasize evidence-based conclusions: nearly every assertion is supported by peer-reviewed sources or public trial data.

Key findings include: Modern AI-driven pipelines interleave *generation* and *filtering/ranking* steps. Initial hit-generation models (e.g. variational autoencoders, graph networks, diffusion models) propose millions of candidate structures (www.sciencedirect.com), which are then computationally triaged using property-predictive models (QSAR, ADMET, docking) and rules-based filters (Lipinski’s rules, toxicity alerts) (www.sciencedirect.com) (jcheminf.biomedcentral.com). Multi-objective optimization (balancing potency, toxicity, drug-likeness) is central: many pipelines use Pareto-ranking or reinforcement learning to identify compounds with optimal trade-offs (www.sciencedirect.com) (www.sciencedirect.com). Recent tools integrate multiple criteria at once – for example, the *druglikeFilter* deep-learning platform scores compounds on physicochemical, toxicity, affinity and synthetic-feasibility axes (www.sciencedirect.com). Other frameworks use **active learning** to iteratively select and test the most informative novel compounds, yielding substantial hit enrichment with minimal testing (pubs.rsc.org). Case studies reinforce the importance of triage: AI models

frequently produce chemicals that are theoretically active but synthetically infeasible, so post-generation filters (even human-designed heuristics like “AstraZeneca filters”) are routinely applied to remove implausible candidates (jcheminf.biomedcentral.com). Indeed, Parrot *et al.* (2023) report that adding a retrosynthetic accessibility score dramatically improves the practical utility of generated libraries (jcheminf.biomedcentral.com) (jcheminf.biomedcentral.com).

In summary, “AI-generated molecule prioritization” relies on an ecosystem of computational triage tools. These range from classical rule-based filters (e.g. Lipinski or PAINS filters) to advanced machine-learning predictors of ADMET and binding, to synthesis-planning algorithms that score synthetic feasibility. The best practice in modern pipelines is to embed these triage filters throughout, either by incorporating them into the generative Objective or by applying them downstream. Our review identifies current limitations (e.g. over-optimistic benchmark performance, lack of chemical diversity in some designs (jcheminf.biomedcentral.com) (www.sciencedirect.com)) and points to future needs: better data standards, integrated closed-loop platforms, and regulatory guidance (the **FDA** has already published draft guidance on AI credibility in drug development (www.fda.gov)). With 29+ AI-augmented programs in clinical trials by 2025 (www.sciencedirect.com), the field is rapidly evolving. However, our analysis stresses that prioritization **must** remain evidence-based: only well-validated triage models and careful risk assessment will ensure that AI accelerates truly viable drug candidates, rather than flooding pipelines with experimental dead weight.

1. Introduction and Background

Drug discovery has historically been **expensive and inefficient**. On average, developing a new drug costs on the order of **$2–3 billion** and takes **10–15 years** (www.sciencedirect.com), with overall clinical success rates below 10% (www.sciencedirect.com). For example, traditional pipelines often generate millions of candidate small molecules through high-throughput screening or combinatorial chemistry, but only ~1 in 10,000 compounds entering preclinical testing will ever be approved (www.sciencedirect.com). In this context, AI and machine learning (ML) have emerged in the last decade as **promising tools to accelerate discovery** and reduce wasted effort (www.sciencedirect.com) (www.sciencedirect.com). Early applications of ML (like QSAR models) gradually gave way to deep learning, and more recently to sophisticated **generative AI** approaches capable of proposing entirely novel molecular structures.

Generative models, such as variational autoencoders (VAEs), generative adversarial networks (GANs), recurrent neural networks (RNNs), graph-based neural networks, and diffusion models, can learn chemical “language” from large datasets (ChEMBL, PubChem, etc.) and then sample new SMILES strings or molecular graphs that satisfy desired objectives (www.sciencedirect.com)(www.sciencedirect.com). These models can explore regions of chemical space far beyond what manual medicinal chemistry could foresee, in principle encompassing the >10^60 drug-like molecules hypothesized to exist (www.sciencedirect.com). For example, modern pipelines may train a VAE or transformer on known actives, condition on a property label (via Reinforcement Learning or conditional generation (www.sciencedirect.com)), and output thousands to millions of de novo structures predicted to bind the target of interest.

Despite these advances, a **critical challenge** remains: **Every step that amplifies the number of candidate molecules also increases the experimental burden of testing them**. In practice, one cannot synthesize and assay millions of AI-proposed structures. Thus, **molecule prioritization** – the triage of candidates through

computational filters – is an essential counterpart to generative design. The evolving drug discovery pipeline is often described as a loop of generation → in silico evaluation → selection → experimental testing, with iterative feedback. This is evident in recent reviews and diagrams (www.sciencedirect.com) (www.sciencedirect.com) that place generative modules at the front end and show follow-on triage modules (docking, ADMET models, retrosynthesis) before any wet-lab step: for instance, Chen *et al.* (2025) note that "generative models design de novo molecules, [which are] filtered via predictive models (binding affinity, ADMET), then docking, synthesis planning, and wet-lab validation" (www.sciencedirect.com). In other words, AI not only generates, but must **prioritize** – using additional AI/ML and rule-based tools – to funnel only the most promising candidates forward.

This report surveys the full landscape of **"AI-Generated Molecule Prioritization"**. We first provide necessary background on traditional drug discovery pipelines and recent generative AI innovations (§1.1–1.4). We then delve into the triage problem: what criteria define a "promising" molecule, how are these criteria computed, and what tools implement them (§2). We discuss data-driven and physics-based selection methods, multi-objective optimization, and hybrid workflows. We present case studies and experimental outcomes (§3), including both successes (AI-derived leads advancing to trials) and cautionary examples (false positives, novelty re-plots). Finally, we evaluate future directions and implications (§4), including emerging standards, regulatory interest, and the burgeoning area of closed-loop discovery. Throughout, claims are backed by extensive citation of peer-reviewed literature, patents, and official announcements.

1.1 The Traditional Drug Discovery Pipeline

Before AI, small-molecule drug discovery relied on a sequence of empirical steps. A validated biological **target** (e.g. enzyme, receptor) is chosen for therapeutic interest (www.sciencedirect.com). Researchers screen or design chemical **libraries** against the target using high-throughput assays (in vitro or cell-based) or virtual screening methods (www.sciencedirect.com). Hits are then optimized through structure–activity relationship (SAR) campaigns, analog synthesis, and medicinal chemistry, striving to improve potency, selectivity, and pharmacokinetics while avoiding toxicity (www.sciencedirect.com) (www.sciencedirect.com). After lead optimization, a small number of candidates enter preclinical tests (ADMET profiling, animal models) before clinical trials. Historically, only ~10% of candidates that reach human trials ultimately gain approval (www.sciencedirect.com).

Two major factors lengthen this process: **scale and attrition**. Traditional assays can only test limited libraries (10^4–10^6 compounds); to find actives, one often screens tens of thousands of compounds exhaustively. Even then, the success rate is low: typical hit rates from virtual screening might be 0.1–1%. Even for a hit that looks promising in silico, many will fail in ADMET or chemistry. For example, only *1 in 10,000* molecules tested pre-clinically becomes an approved drug (www.sciencedirect.com). Thus, most of the chemical space remains unexplored, and lean pipelines often rely on heuristic rules (Lipinski's rule-of-five, drug-likeness filters, substructure alerts) to trim obviously poor candidates before costly validation.

1.2 AI and Generative Models in Drug Discovery

In response to these limitations, the last decade saw a revolution in computational chemistry. Early machine learning (support vector machines, random forests) enabled quantitative structure–activity relationships (QSAR) that could replace first-round screens (www.sciencedirect.com). More recently, **deep learning** (DL) has permeated many steps: image-based phenomics (Recursion, etc.), sequence-based target ID (Zebrafish?), property prediction (GNNs for ADMET (www.sciencedirect.com)), and crucially, *de novo molecule generation*.

Key milestones include the demonstration that Recurrent Neural Networks (treating SMILES strings like language) and Graph Neural Networks can learn chemical syntax and decode latent spaces to generate valid molecules. Seminal works used VAEs or GANs to propose molecules with optimized properties (www.sciencedirect.com) (www.sciencedirect.com). More recently, diffusion models and transformers (e.g. graph-based and sequence-based) have achieved state-of-the-art in generating molecules that satisfy complex, multi-parameter objectives (www.sciencedirect.com) (www.sciencedirect.com). AI platforms from startups (e.g. Atomwise, Insilico, Exscientia, BenevolentAI, Schrödinger) have integrated these generative engines with predictive biology and chemistry modules (www.sciencedirect.com) (www.sciencedirect.com). A few high-profile examples underscore the potential: Insilico's AI pipeline discovered *rentosertib*, a novel TNIK inhibitor, in under 30 months from target identification through Phase 1 trials (www.nature.com). Exscientia used AI to find DSP-1181, entering Phase I within 12 months (www.sciencedirect.com). Recursion used ML-driven image analysis to advance several plant genomics-derived molecules (e.g., REC-994) into human trials (www.sciencedirect.com) (www.sciencedirect.com).

Nonetheless, truly executed "AI-designed drugs" remain rare. According to a recent review of studies up to mid-2025, only ~29 AI-driven programs had publicly entered human trials, including DSP-1181 and rentosertib (www.sciencedirect.com). These represent early proofs-of-concept, not yet broad adoption; the field still emphasizes benchmarks over real-world efficacy, as discussed below (www.sciencedirect.com).

1.3 Multi-objective Optimization in Drug Design

An essential element in understanding molecule prioritization is that drug discovery is inherently **multi-objective** (www.sciencedirect.com). No single metric (like target binding affinity) guarantees success; drug candidates must simultaneously have strong potency, appropriate ADMET profiles, patentability/novelty, ease of synthesis, and more. These objectives often conflict (e.g., adding polar groups may increase potency but worsen cell permeability). As Parrot *et al.* note, even in advanced generative pipelines many **non-viable** molecules are produced unless synthetic constraints are applied (jcheminf.biomedcentral.com).

The field of multi-objective optimization (MOO) provides strategies for balancing these trade-offs (www.sciencedirect.com). Common approaches in AI-driven drug design include: (1) **A priori aggregation** of objectives into a weighted sum or scalar function to drive gradient-based or RL-based generators, (2) **Pareto optimization** where multiple candidates are evolved and ranked by non-dominance (no candidate is better in all objectives) (www.sciencedirect.com) (www.sciencedirect.com), or (3) **interactive/iterative** schemes that

adjust priorities during generation. Indeed, Liu *et al.* (2023) report that many published generative frameworks are coupled with reinforcement learning or conditional generators to satisfy multiple training objectives (www.sciencedirect.com).

Table 1 below summarizes common triage criteria ("objectives") and how various AI methods attempt to optimize them. Notably, *Pareto-based* algorithms are increasingly popular: they produce a "front" of candidate molecules balancing potency vs. synthetic feasibility vs. toxicity (www.sciencedirect.com) (www.sciencedirect.com). Aggregation methods (weighted sums) remain common when specific trade-off weights are known or fixed. In any case, the generative output itself often incorporates these objectives (e.g. by including a synthetic-accessibility score in the reward function (jcheminf.biomedcentral.com)), and additional triage filters are applied post-hoc to remove outliers.

Table 1. *Representative criteria and approaches in prioritizing generated molecules.*

Criterion / Objective	Description	Example Methods/Tools
Target Activity (Potency)	Predicted binding affinity or activity against the biological target (e.g. IC_{50}, K_d).	QSAR models; docking scores (AutoDock, Glide); ML predictors
Drug-Likeness Properties	Physicochemical filters (e.g. Lipinski's rule-of-five, Veber rules), QED score, and toxicity-aware substructure flags.	Drug-likeness metrics (QED) (www.sciencedirect.com); PAINS filters; druglikeFilter (AI) (www.sciencedirect.com)
ADMET Profiles	Predicted absorption, distribution, metabolism, excretion, and toxicity endpoints.	Multi-task DL models; graph neural nets for ADMET (www.sciencedirect.com); ADMETlab, pkCSM
Synthetic Accessibility (SA)	Ease or feasibility of chemical synthesis (often via retrosynthetic route planning).	Retrosynthesis-based scores (RScore/Spaya) (jcheminf.biomedcentral.com); Bai heuristic (SAscore); AiZynth RA score
Chemical Novelty / IP	Structural novelty relative to known drugs/patents. Ensures new intellectual property space.	Similarity search vs. existing drugs; patent databases analysis
Multi-Objective Optimization	Combined ranking or filtering of compounds using aggregated or Pareto-based techniques to balance multiple criteria.	Pareto front selection (www.sciencedirect.com); weighted RL reward; multi-objective GAs
Experimental Tractability	Availability of ready precursors; compatibility with high-throughput synthesis.	On-demand library querying; retrosynthetic route filtering; building block matching

Table 1: Triage objectives in AI-driven drug discovery. Each row represents a key property or constraint that candidate molecules must satisfy. Effective pipelines use a combination of these criteria, often via ML models or rule sets, to rank or filter the outputs of generative algorithms (citations exemplify integrated tools, e.g., He et al.'s *druglikeFilter* (www.sciencedirect.com) and Parrot et al.'s *RScore* (jcheminf.biomedcentral.com)).

2. Triage Strategies for AI-Generated Molecules

Once a generative model proposes candidates, the pipeline must **triage** them using computational filters and scorers. These triage tools serve two main purposes: (**a**) **screening out compounds** that are unlikely to succeed (eliminating "false positives"), and (**b**) **ranking the remaining candidates** to prioritize experimental testing. Below we detail the principal triage methods used in current drug pipelines, categorized by the type of property or assessment.

2.1 Physicochemical and Rule-Based Filters

A first-pass triage often applies simple **rule-based filters** to discard obvious non-drug-like molecules. Classic examples are Lipinski's *Rule-of-Five* (molecular weight <500, clogP <5, ≤5 H-bond donors, ≤10 H-bond acceptors) and related "drug-likeness" heuristics (www.sciencedirect.com). These rules capture empirically derived suggestions for oral bioavailability and are extremely fast to compute from structure. Many commercial and open platforms automatically enforce such thresholds on generated libraries. Other structural alerts include **PAINS filters** (removing frequent hitters and colloidal aggregators) and functional-group checks (eliminating reactive/unstable moieties). For instance, Parrot *et al.* note that AstraZeneca employs combined property and structural filters in their pipelines (jcheminf.biomedcentral.com).

More sophisticated rule-derived filters integrate multiple dimensions at once. For example, **druglikeness scoring functions** like QED (quantitative estimate of drug-likeness) combine fragment contributions and property values. Recently, deep learning approaches have begun to replicate multi-criteria filtering: *druglikeFilter* (He *et al.* 2025) is a multi-task neural net that outputs a compound's compliance with *four* filters (physicochemical space, toxicity alerts, predicted binding, and synthetizability) in one shot (www.sciencedirect.com). This tool reflects the insight that fast AI-based screening itself can triage large virtual libraries: He *et al.* report that their model can flag "non-drug-like" molecules en masse, dramatically reducing wasted effort (www.sciencedirect.com). Such integrated filters are effectively gatekeepers to further steps; compounds failing these basic checks are dropped before any deep evaluation.

2.2 Predictive Binding and Affinity Models

One of the most direct triage criteria is **predicted target affinity**. Even if the generative model used an affinity-related reward, additional screening often applies more detailed binding models to the candidates. The simplest approach is **molecular docking**: each candidate is virtually "docked" into the 3D structure of the target protein to score how well it fits. Docking programs (AutoDock, Glide, SwissDock, etc.) and DNN-enhanced variants (e.g. **GNINA**, which uses CNNs to refine docking poses (pubs.rsc.org)) produce a ranking by score. Although docking scores correlate imperfectly with true binding in practice, they are widely used to down-select large sets – essentially acting as a triage funnel. For example, the FEgrow workflow by Cree *et al.* uses a CNN scoring function in docking to rank candidate ligands before any synthesis (pubs.rsc.org).

In parallel, **ML-based affinity predictors** are often applied. Graph neural networks (GNNs) or sequence-based models can be trained on activity data (Quandetta, MoleculeNet datasets) to predict binding against a target. Tools like *Chemprop* (a proprietary GNN) or multi-task deep nets have high accuracy on large benchmarks (www.sciencedirect.com). Multi-task models in particular can score dozens of targets simultaneously, which is helpful for polypharmacology. During prioritization, one can run ADMET and

polypharmacology filters in the same way, using ML. The triage benefit is twofold: (1) ML models can incorporate vast SAR data and potentially catch non-obvious structure–activity relationships; (2) they can output a quantitative confidence or probability, which feeds into ranking.

It is important to note that both docking and ML predictions carry uncertainties. Recent surveys report that state-of-the-art ADMET or binding predictors achieve ~85–90% accuracy on random benchmarks, but often drop to 60–75% on scaffold-based splits that mimic real-world divergence (www.sciencedirect.com). In other words, models may overestimate performance on held-out analogs and struggle on genuinely new chemotypes. This means that triage predictions are useful but should be interpreted cautiously, typically as a relative score rather than absolute. Robust pipelines thus often combine multiple orthogonal predictors (ensemble docking scores, ML affinity, empirical filters) to mitigate any one method's bias.

2.3 ADMET and Safety Profiling

Beyond potency, **absorption, distribution, metabolism, excretion, and toxicity (ADMET)** attributes are critical "knockout" criteria. AI models can screen hundreds of X-ray generated leads for toxicological red flags or unfavorable pharmacokinetics before any animal tests are run. For instance, GNNs and pre-trained transformer models have been shown to reach ~85–90% accuracy on standard ADMET classification tasks (www.sciencedirect.com), far surpassing older QSAR. They can flag potential hERG toxicity, CYP450 interactions, or low solubility early in silico. Multiparameter tools (like ADMET Predictor, pkCSM, admetSAR) can score drug candidates on dozens of endpoints in seconds.

In a generative pipeline, ADMET models act as additional objectives or filters. Many providers set minimum thresholds (e.g. human liver microsome stability score, Caco-2 permeability) and discard weak compounds. Some projects explicitly incorporate ADMET into the generative reward function – for example, by including predicted logD and tox scores in the optimization step (www.sciencedirect.com) – so that the molecules output are already biased toward acceptable properties. Alternatively, independent post-hoc triage can apply these models to the generated set. The key point is that a candidate with suboptimal ADMET profile (even if potent) will typically be **deprioritized**, since clinical attrition often stems from late-stage toxicity or bioavailability failures (www.sciencedirect.com) (www.sciencedirect.com).

2.4 Synthetic Accessibility and Retrosynthesis Scoring

A particularly challenging but vital triage criterion is **synthetic accessibility (SA)**: how easily can the candidate molecule be made in the lab? No matter how potent a virtual hit is, it is useless if it cannot be synthesized in practice. Traditional cheminformatics uses heuristics like the Ertl SAscore (based on fragment complexity) or SCScore (neural net trained on reaction sets) to approximate SA. These are fast to compute but only rough proxies. Recognizing the centrality of this issue, recent work has introduced more rigorous methods. Parrot *et al.* (2023) introduce **RScore** – a full retrosynthesis-based accessibility score computed via an AI retrosynthesis planner (Spaya) (jcheminf.biomedcentral.com). RScore ranges from 0 (no plausible route found) to 1.0 (very easy), and correlates well with chemist judgments (jcheminf.biomedcentral.com) (jcheminf.biomedcentral.com). They further train an RScore predictor (RSPred) for speed. Crucially, adding these scores into the generative loop ("synthetic constraint") yielded libraries of much more accessible molecules without sacrificing diversity (jcheminf.biomedcentral.com).

In practice today, synthetic realism filtering often comes in two forms: (a) **post-generation**: compute an SA or retrosynthetic score for each proposed molecule and drop those below a threshold; (b) **in-generation**: penalize the generative model for producing "high synthetic cost" patterns. For example, Insilico's Chemistry42 platform and AstraZeneca's pipelines both incorporate synthetic cost metrics during design iterations. A standard triage step is to run each hit through a retrosynthetic planner: molecules for which no reasonable route is found (or which require many steps or exotic reactions) are relegated. The impact of SA filtering is enormous: as a rule, **many de novo generators output a majority of molecules deemed unsynthesizable by expert chemists**, so SA filtering sharply reduces list sizes (jcheminf.biomedcentral.com).

2.5 Novelty and Intellectual Property Considerations

Another practical triage dimension is *molecule novelty*. Pharmaceuticals invest heavily in patents, so generative design ideally yields structures **distinct from known compounds**. Triage tools here are less about missed assertions and more about portfolio strategy: after filtering for potency and safety, one can run the candidates through similarity searches against known drugs and patent libraries. Highly similar molecules might be discarded or deprioritized if "freedom-to-operate" is at risk. Conversely, truly novel scaffolds can be flagged as high-value.

Quantifying novelty is an active area of research (see CAS analysis of AI molecules (www.cas.org)). Todd Wills of CAS introduced measures of 3D structural novelty to assess first-in-class status of early AI drugs. In practice, many AI platforms now integrate substructure checks or structural classification schemes to ensure output diversity. Also, generative negatives can be used (explicitly punishing similarity to existing chemotypes). For example, in the CAS report (www.cas.org), Exscientia's candidates were examined to see how much they overlapped with old drugs (58% overlapped haloperidol's shape). As an implication, modern triage might involve penalizing overuse of common scaffolds or encouraging exploration of underpopulated chemical regions (www.cas.org) (www.cas.org).

2.6 Active Learning and Iterative Screening

Beyond one-shot filtering, **active learning (AL)** techniques represent a dynamic triage strategy. In AL, a small initial subset of AI-generated compounds is scored (e.g. via docking or lab testing), and the results are used to train an interim ML model. This surrogate model predicts the objective (e.g. binding score) across the remaining library. The algorithm then selects the next batch of molecules to evaluate based on uncertainty or expected improvement (pubs.rsc.org). Iterating this loop yields rapid identification of strong compounds with only a fraction of molecules ever explicitly evaluated.

Cree *et al.* (2025) apply active learning to triage compounds grown in a protein pocket (pubs.rsc.org). They report that AL "increases enrichment of hits compared to random selection at low cost," and their analysis cites prior work showing AL being "relatively insensitive to hyperparameters" and effective across a range of scenarios (pubs.rsc.org). In their workflow (for SARS-CoV-2 M^{pro}), only a small fraction of building-block combinations needed to be docked before the most potent designs were discovered, illustrating that AL can massively cut down computational screening. More generally, AL is increasingly used in virtual screening campaigns (pubs.rsc.org) and is a centerpiece in many modern "self-driving lab" platforms.

2.7 Integrated Multi-stage Pipelines

Real drug discovery pipelines rarely rely on a single triage criterion at once. Instead, they deploy **cascaded filters and re-scoring**, akin to multi-stage virtual screening. For example, a typical sequence might be: (1) apply Lipinski and PAINS filters to the AI library; (2) screen remaining molecules with an ML property predictor or docking; (3) retain top N% and assess synthetic score; (4) cluster to ensure chemical diversity; (5) finalize top candidates for procurement or synthesis. Many pharmaceutical companies implement such hierarchical pipelines using proprietary software or combinations of open tools.

Parallel computing and cloud platforms now enable screening of millions of compounds with multi-step filters (www.sciencedirect.com) (pubs.rsc.org). For instance, FEgrow's authors automated the growth+scoring of ~10^6 compounds by parallelization on HPC clusters (pubs.rsc.org). They note that without intelligent triage, "time is wasted building and scoring compounds that are unlikely to be beneficial" (pubs.rsc.org). Thus, incorporating active learning or other heuristics early in the pipeline is essential to avoid brute force.

Table 2 below illustrates how different triage tools might be combined in a workflow step by step.

Table 2. *Example multi-stage triage pipeline for AI-designed molecules.*

Stage	Task	Filtering/Ranking Tools	References/Examples
(A) Rule Filter	Remove gross outliers, charge/ring count, PAINS, toxicophores	Lipinski rules, PAINS filters, reactive substructure flags (www.sciencedirect.com) (jcheminf.biomedcentral.com)	Common in pharma; step to cut non-starters (jcheminf.biomedcentral.com)
(B) Property Scoring	Rapid ADMET & logP screening to gauge drug-likeness	ML models for solubility/CYP, QED score, neural-net ADMET predictors (www.sciencedirect.com)	ADMETlab, pkCSM, or in-house GNNs
(C) Activity Prediction	Screen for target activity	Docking (e.g. AutoDock/GNINA), ML affinity prediction (pubs.rsc.org) (www.sciencedirect.com)	Ensures candidates likely bind target
(D) Synthetic Feasibility	Score/prioritize synthesis difficulty	Retrosynthesis AI (Spaya RScore (jcheminf.biomedcentral.com)), SA heuristics	Drop compounds with impractical synthetic routes (jcheminf.biomedcentral.com)
(E) Diversity Selection	Ensure chemical space coverage among top hits	Cluster molecules by scaffold/fingerprint; select per cluster	Avoid redundancy; ensure novel scaffolds (www.cas.org)
(F) Prioritized List	Final ranking for experimental testing	Multi-attribute rank (weighted/Pareto); manual expert review	Composite score combining potency, ADMET risk, novelty, etc.

Table 2: A schematic example of a multi-stage triage pipeline. Each stage filters or ranks compounds by a specific criterion, progressively narrowing down from potentially millions to a few dozen candidates. In practice, stages (B)–(D) can be interleaved or repeated; modern pipelines often "loop" (A)–(C) iteratively via active learning (pubs.rsc.org). References indicate representative methods cited in the literature for each task.

3. Case Studies and Real-World Applications

To illustrate how AI-generated molecule triage is applied in practice, we discuss several case studies from recent literature and industry reports. These highlight successes, challenges, and lessons learned in managing AI-proposed molecules through drug pipelines.

3.1 AI-Generated Clinical Candidates

DSP-1181 (Exscientia, OCD): Exscientia's collaboration with Sumitomo Dainippon Pharma produced a serotonin receptor agonist, DSP-1181, that entered Phase I human trials for obsessive-compulsive disorder in early 2020 (www.cas.org). This was the first reported AI-driven candidate to reach human testing. Interestingly, patent documents reveal that the claimed molecules closely shared scaffold shapes with the existing antipsychotic haloperidol (www.cas.org). In fact, a CAS structural analysis showed that of the ~350 compounds synthesized during DSP-1181's optimization, about 58% shared haloperidol's core shape (www.cas.org). This suggests that although AI was used to select and optimize structures, the final candidates drew heavily on known pharmacophores. It underscores the importance of a novelty-safeguarding triage: had the AI been allowed to pick only globally novel shapes, it might have missed these potent analogues.

EXS21546 (Exscientia, A_2A receptor antagonist for cancer): In late 2020, EXS21546 became another Exscientia+Evotec Phase I candidate (www.cas.org). According to Exscientia's reports, ~163 molecules were synthesized during its discovery, and 28% of those had measured bioactivity (www.cas.org). The CAS analysis noted that the exemplified hits clustered into only a few closely related scaffolds, all novel to the literature but structurally similar to known A_2A antagonists (www.cas.org). This indicates that the triage during discovery effectively funneled the search towards a small chemical series with high potential. The patent also reported that only 28% of generated molecules were actually tested, implying that many were filtered out (presumably via internal triage) prior to synthesis (www.cas.org).

Rentosertib (Insilico Medicine, IPF): A landmark example is rentosertib, an AI-designed inhibitor of Traf2- and Nck-interacting kinase (TNIK) for idiopathic pulmonary fibrosis (IPF) (www.nature.com). Xu *et al.* (2025) describe how this molecule was discovered and rapidly advanced to a Phase 2a trial (www.nature.com). Rentosertib is notable because it originated from a fully generative pipeline: AI identified a novel target (TNIK) and produced de novo ligands, culminating in a first-in-class molecule. In their Phase 2a study, the highest dose group showed a **mean FVC increase of +98.4 mL** over 12 weeks versus a –20.3 mL decrease on placebo (www.nature.com), a clinically significant improvement. This successful outcome validates the end-to-end AI approach, but it also depended on effective triage: millions of AI-suggested analogs were pruned by QSAR models (for potency) and toxicity/synthesis filters before any synthesis. Xu *et al.* report that their AI-driven workflow achieved target ID → lead nomination in just 18 months and Phase I completion in under 30 months (www.nature.com). This extraordinary acceleration hinged on trusting AI triage; presumably, only the most promising candidates from virtual screening were selected for faster chemical synthesis and testing.

REC-994 (Recursion, phenomic drug discovery): Recursion Pharmaceuticals is known for image-based AI phenomics rather than chemical generative design. Nonetheless, their pipeline involves triage of thousands of compounds by high-content cellular profiling. REC-994, discovered via Recursion's platform, reached Phase II trials for cerebral cavernous malformation in humans (www.sciencedirect.com). This example illustrates the broader point: *any* AI-driven discovery (even if the AI focus is on biological data) ultimately requires chemical

triage. Recursion's success highlights that AI can integrate diverse data modalities, but the output molecules still had to satisfy drug-like criteria. The fact that multiple Recursion AI candidates (REC-617, REC-4881, etc.) have entered trials shows the increasing maturity of these AI triage pipelines (www.sciencedirect.com) (www.sciencedirect.com).

ChatGPT for Drug Design (Sci. Rep. 2026): A recent novel case study evaluated general-purpose AI (GPT-4o / ChatGPT) as a creative assistant in drug discovery (www.nature.com). Abdel-Rehim *et al.* (2026) tasked ChatGPT with three design challenges: optimizing known EGFR inhibitors, de novo EGFR inhibitors, and MCL1 inhibitors. Impressively, in silico QSAR-guided design allowed ChatGPT to propose compounds with predicted potencies in the tens of nanomolar range (www.nature.com). However, the authors emphasize triage considerations: the generated molecules were often not readily synthesizable, so the team had to **find analogous readily available compounds** from vendor catalogs (e.g. through similarity searches) for actual study (www.nature.com). Several chosen analogues displayed good docking and QSAR scores (~10–100 nM range) (www.nature.com). This pipeline (LLM → analog retrieval → predictive screening) is an example of pragmatic triage: the initial AI ideas were filtered by similarity to purchasable chemistry, and then assessed computationally. The study demonstrates that even versatile generative tools (like LLMs) must be constrained by chemical reality to yield actionable leads.

3.2 Data and Performance Analyses

Where possible, we summarize quantitative findings on triage efficiency from the literature:

- **Active Learning Hit Enrichment:** Cree *et al.* (2025) report that active learning significantly enriched hit discovery for SARS-CoV-2 protease inhibitors (pubs.rsc.org). Citing prior work, they note that AL can identify the most promising compounds by evaluating only a small **fraction** of the chemical space (pubs.rsc.org). In general, AL workflows have shown higher hit rates than random or one-shot screening, with "relatively low additional cost" once the surrogate model is trained (pubs.rsc.org). Their own experiments found that a few cycles of 100–200 compounds each were sufficient to home in on top candidates from a library of millions.
- **Multi-Objective Optimization Outcomes:** In generative settings, including SA or other filters in the objective clearly shifts outputs toward desirable regions. Parrot *et al.* observe that unconstrained generators churned out mostly synthetically inaccessible molecules, whereas adding the retrosynthesis-derived RScore constraint doubled the fraction of synthesizable compounds (jcheminf.biomedcentral.com). They also saw that diversity improved with the constraint, suggesting it did not over-restrict the search. Similarly, conditioning models on property ranges or using Pareto fronts (rather than blind single-objective RL) tends to produce sets of candidates with more balanced scores across metrics (www.sciencedirect.com) (www.sciencedirect.com).
- **AI vs. Traditional VS Efficiency:** The broader literature suggests that AI-enhanced virtual screening can dramatically reduce library sizes. For example, the RSC Digital Discovery review notes that AI methods (deep docking, active/transfer learning) have "substantially shortened" hit-identification timelines compared to brute-force docking (www.sciencedirect.com). Industry anecdotes claim 1–2 orders of magnitude fewer compounds need to be tested by prioritizing with ML models. However, careful studies also warn of "benchmarking illusions": many reported enrichments derive from

retrospective splits, and real-world gains are more modest (www.sciencedirect.com). In short, priority setting improves efficiency but is not a panacea; prospective validations are still needed to quantify true success rates.

3.3 Discussion of Successes and Failures

Overall, case studies point to the promise of AI triage but also highlight challenges:

- **Successes:** DSP-1181 and rentosertib show that properly filtered AI workflows can yield first-in-class drugs rapidly. Active learning pipelines (e.g. FEgrow) and custom filters (druglikeFilter) demonstrate that automation can handle multi-dimensional assessments at scale (www.sciencedirect.com) (pubs.rsc.org). The growing number of AI-origin candidates in trials by 2025 (www.sciencedirect.com) (www.sciencedirect.com) indicates real-world impact.
- **Limitations:** Many AI campaigns end at hit identification. DSP-1181's core scaffold was known (haloperidol-like), suggesting the generative novelty was limited (www.cas.org). Models often fail to capture pharmacokinetics/good metabolites – a 2025 meta-review remarks that while AI improves potency discovery, "transfer to clinical efficacy remains scarce and higher validation is needed" (www.sciencedirect.com). Overdependence on in silico triage can also be risky if models are miscalibrated. For example, hitting high docking scores but poor real binding can mislead projects if not cross-checked. The FDA has noted this risk: it recently proposed a framework to ensure **AI model credibility** in drug submissions, emphasizing that predictive uncertainty and context-of-use must be rigorously assessed (www.fda.gov).

In summary, the **implication** is that AI-driven prioritization can both **speed up** drug discover and **reduce waste**, but only with careful design. High-throughput computational filters (fast, low-cost) are invaluable, but should be validated by experiment (wet-lab) before full trust. Compassionate integration of multiple triage axes – potency, ADME, synthetic feasibility, novelty – yields the most robust candidates. The field is moving towards such integrative pipelines, but continued real-world benchmarking will be critical.

4. Implications and Future Directions

The integration of AI-generated molecule design with prioritization tools raises several forward-looking issues:

- **Autonomous Closed-Loop Discovery:** The logical next step is fully closed-loop automated laboratories, where AI proposes molecules, robotic systems synthesize and test them, and the results feed back to retrain the models. This self-driving-lab paradigm, already explored in materials science, is emerging for chemistry. For drugs, it would mean coupling generative models to automated synthesis (e.g. flow chemistry platforms) and high-throughput screening, with active learning steering the process. Early work (e.g. Chemify, IBM RoboRXN) suggests this is viable, and it could eliminate human bottlenecks in prioritization.
- **Scalability and Big Data:** Current triage models often rely on limited datasets. As databases grow (ChEMBL, PubChem, OxChem), large-scale pretraining (Films like ChemBERTa, SMILES transformers) will yield more accurate property predictors. Self-supervised methods may allow ADMET models to

generalize better (www.sciencedirect.com). Transfer learning and federated learning could let pharma share anonymous data to improve triage without IP risk. There will be a push for standard benchmarks specifically for AI-generated sets (much like ImageNet in vision) to calibrate triage tools (www.sciencedirect.com).

- **Benchmarking and Best Practices:** Given the high cost of failures, the community is calling for rigorous **prospective validation** of AI triage strategies (www.sciencedirect.com). Already, initiatives like OpenFF and Pfizer's public challenges are stressing transparency and reproducibility. Likely, consortia will emerge to share anonymized failures (molecules predicted good but failing tests) to improve models. Regulatory agencies are also setting guidelines: the FDA draft guidance in 2025 encourages companies to pre-define AI context-of-use and to engage early with regulators on model validation (www.fda.gov).
- **Ethical and Legal Concerns:** The use of generative AI raises IP questions (who owns an AI-designed molecule?) and safety concerns (ensuring ADMET screening catches toxicity). Triage tools themselves must be audited for bias (e.g. overconfidence in certain chemical classes). Explainability ("why was this molecule chosen?") will be demanded, which may favor interpretable predictors over black-box models.
- **Quantum Computing and AI:** The horizon may see quantum chemistry simulators integrated with generative AI. Others speculate combining quantum computing with ML could allow highly accurate property prediction in prioritization, making triage even more reliable (www.sciencedirect.com). While still speculative, such convergence is being explored in academic labs and by companies like QCWare.

In all, the field is trending toward **continuous, data-driven pipelines**. As Chen *et al.* conclude, the ultimate goal is "autonomous molecular design ecosystems" where AI generation, predictive triage, and synthesis planning are seamlessly integrated (www.sciencedirect.com). Achieving this will require not just better algorithms, but robust data infrastructure (standardized molecule/properties datasets) and interdisciplinary collaboration (chemists, data scientists, engineers).

5. Conclusion

AI-generated molecule prioritization is an emerging cornerstone of modern drug discovery. Generative models can now produce candidates that would have been unimaginable a decade ago, but without effective triage, the workload remains intractable. This report has extensively reviewed how triage tools – from simple rule filters to advanced AI scorers – are employed to sift through AI proposals. The evidence shows that combining multiple criteria (potency, ADMET, novelty, synthesizability) via an ensemble of computational tools yields the best results (www.sciencedirect.com) (jcheminf.biomedcentral.com) (pubs.rsc.org). The trend toward multi-objective, AI-enhanced pipelines is backed by real-world progress: numerous AI-assisted candidates (e.g., DSP-1181, rentosertib, REC-994) are now in human trials (www.sciencedirect.com) (www.sciencedirect.com).

However, the approach is not yet foolproof. Benchmarks reveal gaps between retrospective success and prospective reality (www.sciencedirect.com), and high-profile failures (Compounds generated by AI that proved unsynthesizable or inactive) remind us that triage models must be continually refined. Safe deployment of these methods depends on rigorous validation, transparency, and understanding of model limits. The field must also address IP and regulatory questions about AI-developed molecules.

Looking ahead, the merger of AI design and triage with automation holds promise for unprecedented throughput. Researchers envision fully autonomous labs guided by ML-based “chemist AI” that continuously proposes and tests molecules. To realize this, we need scalable, interpretable triage tools that can be trusted by chemists and regulators alike. As the FDA has signaled (www.fda.gov), establishing credibility and standards for AI in drug R&D is the next frontier.

In summary, the marriage of **AI generation** with **AI-driven prioritization** offers one of the most powerful paradigms ever seen in drug discovery. When done correctly, it has the potential to cut years and billions of dollars from the R&D timeline. This review has documented the state of that paradigm in 2026: the algorithms, the tools, the successes, and the pitfalls. With continued innovation and oversight, AI-aided triage will increasingly become a routine component of pharmaceutical pipelines, accelerating the journey from many virtual molecules to a few lifesaving medicines.

Source links

Links cited in this article, in order of first appearance. Full addresses are provided for readers using extracted text.

1. https://intuitionlabs.ai/articles/ai-pharma-biotech-drug-discovery-case-studies
2. https://www.sciencedirect.com/science/article/pii/S2590098625000107
3. https://intuitionlabs.ai/articles/ins018-055-phase-2a-results-ai-designed-drug
4. https://www.nature.com/articles/s41591-025-03743-2
5. https://www.cas.org/resources/cas-insights/ai-drug-discovery-assessing-the-first-ai-designed-drug-candidates-to-go-into-human-clinical-trials
6. https://www.sciencedirect.com/science/article/pii/S1570180826000710
7. https://pubs.rsc.org/ta/content/articlehtml/2025/dd/d4dd00343h
8. https://intuitionlabs.ai/articles/yueldesign-ai-drug-design-protein-flexibility
9. https://www.sciencedirect.com/science/article/pii/S2095177925001157
10. https://jcheminf.biomedcentral.com/articles/10.1186/s13321-023-00742-8
11. https://www.sciencedirect.com/science/article/pii/S0959440X23000118
12. https://intuitionlabs.ai/articles/pharma-ai-validation-evidence-fda-ema
13. https://www.fda.gov/news-events/press-announcements/fda-proposes-framework-advance-credibility-ai-models-used-drug-and-biological-product-submissions
14. https://intuitionlabs.ai/articles/fda-ema-good-ai-practice-drug-development-2
15. https://www.sciencedirect.com/science/article/pii/S135964462400117X
16. https://www.sciencedirect.com/science/article/pii/S2949747724000307
17. https://www.nature.com/articles/s41598-025-32805-1

THE PUBLISHER

About IntuitionLabs

IntuitionLabs is an AI consulting, custom software development and data engineering firm serving pharmaceutical, biotechnology, medical-device and other life-science organizations. We work with clinical, regulatory, medical-affairs, commercial, quality and IT teams to connect technology decisions with the work people need to accomplish.

AI consulting and adoption

Our AI enablement services cover readiness assessments, use-case selection, governance and policies, team workshops, adoption measurement and ongoing advisory support. We help organizations structure the information layer behind AI: source material, context, permissions and maintained knowledge that make generated answers useful and reviewable. Private LLM inference and hosted AI options support teams evaluating how to operate AI with appropriate control over their data and infrastructure.

Software, data and life-science workflows

IntuitionLabs develops custom software for pharma and biotech, integrates enterprise systems, and builds data engineering and business intelligence solutions. Areas of focus include AI agents, regulatory research, medical writing, medical affairs, CMC information, competitive intelligence and clinical-document workflows. Our eTMF intelligence work includes cross-system reconciliation and inspection-readiness support.

Enterprise platforms and regulated delivery

We provide Veeva services, application support, managed services, integrations and custom applications, alongside enterprise content work involving platforms such as Egnyte. For regulated workflows, our services include GxP enablement, computer-system validation and software development addressing 21 CFR Part 11 requirements. The applicable controls, validation responsibilities and acceptance criteria are defined for each engagement.

Work with IntuitionLabs

Explore AI enablement, pharma and biotech software development, data engineering and BI, and Veeva services. Contact IntuitionLabs to discuss your workflow, information sources and implementation needs.

IntuitionLabs publishes educational research to help life-science teams make informed technology decisions. Coverage of a product or organization does not imply a client relationship, endorsement or partnership.

Website: https://intuitionlabs.ai

This article is educational. Verify consequential medical, legal, financial, regulatory and technical decisions with appropriate professionals and the cited primary sources. Company services are described separately from the article's research findings.