

Prompt Engineering for CAS NewtonSM

Techniques and best practices for agentic scientific search in CAS SciFinder®

CAS Newton is agentic AI grounded in the CAS Content Collection™, not a general-purpose language model. It reasons over curated, domain-specific content built by scientists, carries context across turns, contains modular tools that expose and amplify content, and cites its answers. Prompting it well looks less like keyword searching and more like briefing a colleague: state the scientific question, name the constraints, and refine the conversation.

Technique, practice, or pitfall	Why it matters in CAS Newton	Science-based example
PROMPT ENGINEERING TECHNIQUES		
Zero-shot	Direct question, no examples — fastest for simple facts.	"What is the boiling point of acetone?"
Few-Shot	Show one-to-two sample Q&A pairs, so CAS Newton mirrors your desired structure.	"Q: Benzene → CAS RN 71-43-2, flammable liquid. Q: Toluene → CAS RN 108-88-3, flammable liquid. Q: Acetone → ?"
Contextual	Give audience/use case before asking, so depth matches the reader.	"This is for a high-school lab safety briefing. Explain in plain language why concentrated sulfuric acid is dangerous."
Chain-of-thought	Ask for step-by-step reasoning before the final answer — best for mechanisms, stoichiometry.	"Walk through balancing the combustion equation for propane step by step, then give the final equation."
Instructional	Explicit verbs — "list," "compare," "convert," "balance."	"Compare the solubility of NaCl and KNO ₃ in water at 25°C as a table."
Role-based	Frame CAS Newton as a specific expert to calibrate tone/depth.	"You are a process safety chemist. Explain the hazards of mixing bleach and ammonia."
Format-constrained	Specify the exact output shape so it's reusable downstream.	"Return only JSON with keys: compound_name, cas_rn, molecular_formula, hazard_class."
Negative/constraint	State what to avoid — jargon, speculation, unverified claims.	"Explain what a catalyst does without using the word 'kinetics,' in under 50 words."
Ask a research question, not a keyword string	CAS Newton interprets intent, turning multi-part questions into targeted search paths. Keyword fragments waste that reasoning layer.	Perovskite photovoltaic stability "What degradation mechanisms limit operational stability of formamidinium-based perovskite solar cells under damp-heat conditions, and which passivation strategies produced the largest reported lifetime gains?" Weak version: "perovskite stability passivation"
Decompose a complex query into steps	CAS training on prompting teachers to break a query down, then layering follow-ups. Context carries forward, so each step inherits the last.	Kinase inhibitor landscape, in three moves "(1) List the main chemotypes of covalent BTK inhibitors reported since 2018. (2) For the top three, compare reported selectivity over EGFR. (3) Which have entered clinical study for autoimmune indications?" Ask sequentially, not as one paragraph.

Technique, practice, or pitfall	Why it matters in CAS Newton	Science-based example
Pair a structure with text that names its role	CAS Newton takes text, structures, or both. CAS documentation emphasizes describing the drawn structure in words because the phrasing determines whether hits are returned as substances, substructures, or reactions.	Same core, two different answers <i>"Molecules containing the attached quinoline core as a substructure, with activity against chloroquine-resistant P. falciparum. / Synthesis methods for bicyclic amines, such as the structure provided."</i> First returns substances; second returns reaction routes.
Invoke the science-specific capability by name	Structure search, functional group reactivity, spectra prediction, and property prediction are distinct capabilities. Naming one routes the query there instead of a general retrieval.	Structure elucidation support <i>"Predict the ¹H NMR spectrum of the attached compound, then identify reported analogs showing a comparable downfield shift for the vinylic proton."</i>
Signal route-planning intent	When CAS Newton detects synthetic route planning intent, it automatically surfaces a retrosynthesis plan, openable in the full plan viewer.	Route scouting under real constraints <i>"Propose a retrosynthetic route to the attached spiro-oxindole in four steps or fewer, favoring commercially available starting materials and avoiding cryogenic conditions."</i>
Describe the columns you want in a table	CAS Newton organizes data across hundreds of entities into a table exportable to Excel, building the columns you describe. Unspecified requests yield a generic layout.	Solid-state electrolyte triage <i>"Build a table of reported solid electrolytes for lithium-metal batteries with columns for composition, ionic conductivity at 25 C, electrochemical stability window, critical current density, and reference."</i>
Change one variable per follow-up	Context persists across turns, so a follow-up altering several conditions at once hides which change moved the result.	Narrowing a coupling search <i>"Now restrict to ligand-free conditions. Then: of those, which report yields above 85% for aryl chloride substrates?"</i> Follows an opening query on aqueous Suzuki couplings.
Supply observed data when troubleshooting	CAS Newton is built to help diagnose unexpected experimental results, but only the details you provide can be reasoned over.	A stalled amination <i>"My Pd-catalyzed amination of a 2-chloropyridine stalls near 30% conversion with cesium carbonate in dioxane at 100 C, and LC-MS shows a new dimeric byproduct. What side reactions are reported here, and how have others suppressed them?"</i>
Ask for an explicit head-to-head comparison	CAS Newton can compare alternative methodologies. Naming the axes of comparison leads to a descriptive answer when a decision-ready one was needed.	Choosing a deprotection strategy <i>"Compare enzymatic and chemical deprotection of N-acetyl groups in sialic acid synthesis on reported yield, epimerization risk, and demonstrated scale."</i>
Constrain scope and spell out synonyms	Boundaries cut the volume CAS Newton must synthesize; supplying alternate nomenclature stops a whole literature stream from being missed.	Hepatotoxicity screening/degrader vocabulary <i>"Limit to studies since 2020 using human primary hepatocytes rather than HepG2. / Include literature using PROTAC, bivalent degrader, and heterobifunctional degrader."</i> CAS Registry Numbers disambiguate better than names.
Structure long prompts within the character limit	The question box supports up to 1,000 characters, and Shift+Enter inserts a line break. Labeled lines read more reliably than one dense sentence.	A structured prompt skeleton <i>"Goal: green solvent replacements for DMF in amide couplings. / Constraints: REACH-compliant, boiling point under 200 C. / Output: table of solvent, reported yield, reference."</i> Each segment on its own line.



BEST PRACTICES

Open the citation before relying on the answer	Every answer is cited back to the CAS Content Collection, AI output is labeled, and links open the underlying record. Verification is the design intent, not an optional extra.	Toxicology due diligence <i>"Before quoting a reported LD50 in a safety assessment, open the cited reference and confirm species, route of administration, and vehicle."</i>
Treat predicted values as hypotheses	Grounding in curated data substantially lowers the risk of hallucination but does not eliminate it, and predicted properties are model outputs, not measurements.	Solubility prediction <i>"Use predicted aqueous solubility to rank which candidates to measure first, then confirm against an experimental value in the literature record."</i>
Export the conversation for the record	A question and its follow-ups are exported as a PDF with links preserved, and outputs are auditable for compliance review.	Documenting a design decision <i>"Attach the exported PDF to the lab notebook entry, justifying the chosen reaction conditions, so reviewers can retrace the evidence."</i>
Plan the session around usage limits	Daily usage limits may apply per user. Spending the broad landscape question first, then refining, uses the allowance better than repeated wide queries.	Sequencing a literature review <i>"Open with the landscape question, then spend the remaining turns on targeted follow-ups rather than re-asking a wide question with different wording."</i>
Use precise chemical identifiers (CAS RN, IUPAC name, SMILES) instead of casual names	Reduces ambiguity between isomers, salts, or trade names that share a common name.	
State the exact output format up front	Since responses are returned verbatim, an unformatted request yields an unformatted (harder to reuse) answer.	
Ask one well-scoped question per prompt	Compound questions ("explain X and also compare Y and also convert Z") dilute the answer depth on each part.	
Specify units and conditions (temperature, pressure, concentration)	Chemical properties are condition-dependent; omitting them invites CAS Newton to assume defaults that may not match your use case.	
For safety/regulatory questions, add role + audience context	Produces appropriately cautious, calibrated answers rather than generic text.	
Verify high-stakes answers against CAS ConnectionsSM lookups (substance_details_lookup, reaction_search)	CAS Newton is conversational; structured lookups are the system of record for regulatory or purchasing decisions.	
Iterate rather than over-engineer a single prompt	If an answer is off target, a short follow-up ("now express that in mg/L") is usually more reliable than a longer, denser initial prompt.	
Avoid leading questions with an embedded assumption	E.g., asking "why is X more toxic than Y" presumes an answer; ask "how do the toxicities of X and Y compare" instead.	



Practice	Why it matters in CAS Newton	Science-based example
PRACTICES TO AVOID		
Legacy Boolean and operator syntax	CAS Newton is built for conversational access without specialized search expertise. Command-line-style strings, truncation symbols, and field codes fight the reasoning layer rather than feed it.	Carried over from command-line searching <i>"perovskite AND (stability OR degrad?) NOT dye-sensitized"</i> Instead: Write the constraint as a sentence and let CAS Newton translate it into a search path.
Stacking unrelated questions in one turn	CAS Newton decomposes a multi-part question along one line of inquiry. Bundling separate projects into a single prompt splits the search path and burns the 1,000-character budget.	Four projects in one prompt <i>"How is imatinib synthesized, what is its hepatotoxicity profile, who holds the formulation patents, and which suppliers ship it?"</i> Instead: Ask each as its own thread; use follow-ups to go deeper in each.
Prompts that presuppose the conclusion	A leading prompt invites the model to assemble supporting evidence and to skip contradictory literature. The grounding is only as balanced as the question.	Confirmation bias, built into the query <i>"Confirm that nickel catalysts outperform palladium for Suzuki couplings of aryl chlorides."</i> Instead: Ask what the literature reports for both, including cases where nickel underperforms.
Rephrasing and re-asking instead of following up	A reworded broad query discards accumulated context and starts a fresh search path, while still drawing on a limited daily allowance.	The retry loop <i>"Asking the same landscape question four ways in hope of a better answer."</i> Instead: Keep the thread and narrow it one constraint at a time.
Blurring exact-match and scaffold intent	A drawn structure can be read as an exact substance or as a substructure, and the wording decides which. Silence on the point produces the wrong result set, not an error message.	Over-specified when a class was wanted <i>"Attaching a fully decorated lead compound while wanting the whole chemotype."</i> Instead: Draw the core and say, "as a substructure," or name the exact substance and say, "this specific compound."
Reading a null result as proven novelty	An empty answer can reflect your phrasing, scope limits, or indexing lag rather than an absence in the literature. Novelty and freedom-to-operate claims need more than one negative query.	A risky conclusion <i>"No inhibitors are reported for this target, so our scaffold is novel."</i> Instead: Retry with synonyms and a broadened scope, then confirm through a formal IP search.
Citing CAS Newton rather than the primary source	CAS Newton is the route to the literature, not the authority. AI output is labeled as such, and every answer links to the record that supports the claim.	Wrong attribution in a manuscript <i>"Reporting a melting point or activity value attributed to the AI answer itself."</i> Instead: Cite the primary reference CAS Newton surfaced, after reading it.

Compiled from CAS product documentation, CAS Newton training materials, and CAS product announcements (accessed July 2026). Interface behavior, capabilities, and usage limits vary by product license and may change; verify current behavior in your own deployment.

