El Agente Forjador: Task-Driven Agent Generation for Quantum Simulation

Zijian Zhang [1,2] **Aiwei Yin** [1] **Amaan Baweja** [3] **Jiaru Bai** [3] **Ignacio Gustin** [3] **Varinia Bernales** [3,1,4]
Alán Aspuru-Guzik [3,1,5,6,4,2,7]

[1]*Department of Computer Science, University of Toronto* [2]*Vector Institute for Artificial Intelligence* [3]*Department of Chemistry, University of Toronto* [4]*Acceleration Consortium* [5]*Department of Materials Science & Engineering, University of Toronto* [6]*Department of Chemical Engineering & Applied Chemistry, University of Toronto* [7]*Canadian Institute for Advanced Research (CIFAR)*. Correspondence to: Alán Aspuru-Guzik alan@aspuru.com.

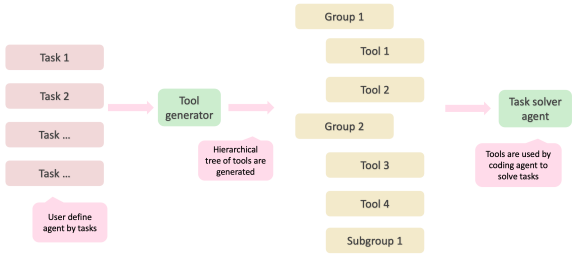

Fig. 1: Task-driven agent generation. A user describes a goal; the coding agents diagnose missing capabilities, forge and validate the required tools, and then an execution agent will solve the task based on the tools. The tools can also be reused related tasks later.

Abstract

Scientific automation increasingly relies on large language model (LLM)-based agents, yet most systems depend on static, hand-curated toolboxes that hinder adaptation to new tasks, scientific subdomains, and rapidly evolving libraries. We present El Agente Forjador, a multi-agent framework that forges, validates, and reuses tools on demand in a task-driven generate–execute–evaluate loop. Instead of assuming all necessary capabilities in advance, the system identifies gaps in its toolset, synthesizes and verifies task-specific tools, and retains them for future use. We evaluate the framework on tasks from quantum chemistry, quantum dynamics, and quantum computing that involve distinct computational abstractions and software stacks, observing progressive reuse of forged tools and reduced need for manual intervention. These results suggest that tool forging provides a practical and scalable mechanism for enabling adaptable scientific agents across environments.

1. Introduction

Automating scientific research is a long-standing goal, with recent advances in large language model[1, 2] and agent systems enabling natural-language interfaces to complex workflows [3, 4, 5, 6]. While these systems can plan and execute multi-step tasks, they typically rely on fixed, hand-engineered toolsets that must be manually extended as tasks evolve or software libraries change. This reliance on static tools creates friction even when moving between closely related scientific subdomains and limits the long-term scalability of agent-based scientific automation.

El Agente Forjador addresses this limitation by automating not only task execution but also task-driven tool forging and reuse. When a functionality is missing, the system explicitly forges new tools, validates them, and integrates them into the workflow, allowing subsequent tasks to build on previously forged capabilities. Specifically, we explore this approach across tasks in quantum chemistry, quantum dynamics, and quantum computing, showing how tool forging supports adaptation to different computational abstractions and software stacks while reducing repeated human intervention. Figure 1 illustrates the process.

2. Approach

We introduce a multi-agent framework that forges and reuses tools in a task-driven generate–execute–evaluate loop. The tools are in the format of Python functions. The system uses coding agents, including Claude Code [3] and Codex [4], with distinct prompts for each component (planning, tool search/forging, reviewing, execution, and evaluation). Missing capabilities are detected at the beginning of each run; tools are synthesized with explicit type annotations and validation; agents execute end-to-end workflows and iterate until predefined success criteria are met. The framework (i) forges, validates, and composes tools to reduce human effort; (ii) accumulates reusable tools across tasks to enable progressively complex workflows; and (iii) applies a toolset optimizer between runs to organize the library into a tree structure for more efficient retrieval and reuse.

3. Evaluation

We evaluate on two task suites: El Agente Q [5] for quantum chemistry (12 tasks; primarily using PySCF [7]) and El Agente Cuántico [6] for quantum dynamics and quantum computing (11 tasks; using CUDA-Q [8] for quantum computing and QuTiP [9] for quantum dynamics). El Agente Q contains two difficulty levels: a simpler Level 1 and a more challenging Level 2. Here, we show two representative execution of our system on each task set.

3.1 Relative Stabilities of Carbocations (SMILES input)

Objective. Compute ΔH and ΔG for $R{-}H \rightarrow R^{+} + H^{-}$ at the B3LYP/6-31G level provided the SMILES of R–H. The nine R–H substrates are methane, ethane, propane, 2-methylpropane,

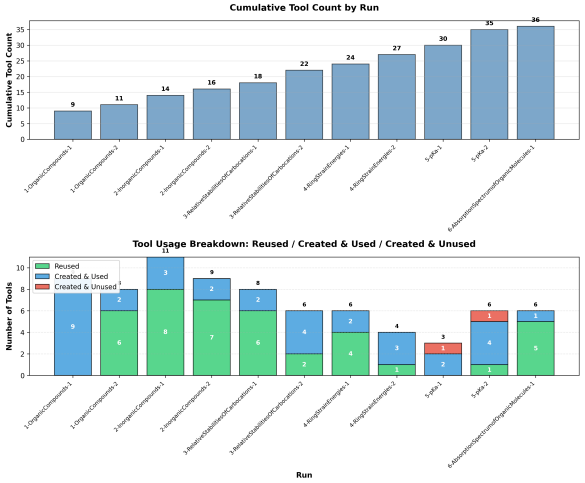

Fig. 2: Tool growth and reuse across the El Agente Q task set. Top: cumulative tools vs. tasks. Bottom: tools generated per task and tools reused.

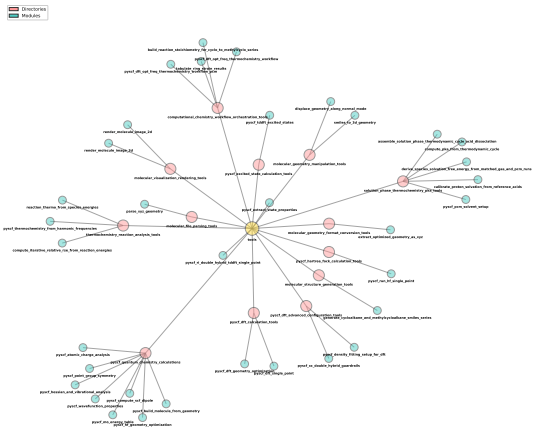

Fig. 3: Resulting tool tree after running the set of tasks from El Agente Q. The tools are organized into a tree structure by a tool optimizer to improve the efficiency of retrieval and archival.

toluene, benzene, dimethyl ether, trimethylamine, and propene.

Reused tools. The agent reused pre-existing tools to (i) build initial 3D geometries from SMILES, (ii) convert geometries to PySCF inputs, (iii) run DFT geometry optimization and vibrational analysis, and (iv) escape saddle points by displacing structures along imaginary modes.

Forged tools. Three tools are genereted to fill gaps in the tool library: a standardized constructor for the single-atom hydride anion (H^-) without geometry optimization; a wrapper to compute per-species $E/H/G$ with consistent RRHO thermochemistry; and a routine to batch reactions and assemble the final $\Delta E/\Delta H/\Delta G$ table with consistent unit conversion.

Results. Success (Human checked). All structures converged to minima after at most one re-optimization iteration; trends match chemical intuition. The whole run took 117.79 minutes and cost $4.56.

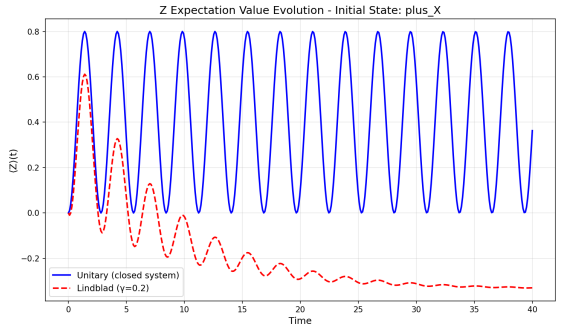

Fig. 4: Agent-generated plot of the expectation value $\langle Z(t)\rangle$ for a single qubit initialized in $|+\rangle_X$ under unitary and Lindblad evolution with $H = X + 0.5Z$ and collapse operator $C = \sqrt{\gamma}\,\sigma^-$ ($\gamma = 0.2$), over $t \in [0, 40]$.

3.2 Unitary vs. Lindblad Qubit Dynamics

Objective. Compare closed (unitary) and open (Lindblad) dynamics for a single qubit under $H{=}X{+}0.5Z$ with collapse operator $C{=}\sqrt{\gamma}\,\sigma^-$, $\gamma{=}0.2$. Initial states $|\pm\rangle_X$; observable Z.

Reused tools. The agent did not reuse any tools as there is no similar tasks the agent has solved.

Forged tools. The agent forges QuTiP helpers to (i) construct a Hamiltonian from a list of Pauli terms, (ii) prepare $|\pm\rangle$ eigenstates of Pauli operators for initialization, and (iii) run both closed-system (unitary) and open-system (Lindblad) time evolution via thin wrappers around `sesolve` and `mesolve` that return time grids and $\langle Z(t)\rangle$ trajectories.

Results. Success (Human checked). Dissipation yields damped oscillations and phase shifts vs. unitary trajectories (See Figure 4). Deviations depend on initial state (e.g., RMSE ≈ 0.65 for $|+\rangle_X$ vs. ≈ 0.25 for $|-\rangle_X$). The whole run took 137.3 minutes and cost $7.96.

4. Conclusion & Limitations

El Agente Forjador demonstrates that task-driven tool forging can make LLM agents more adaptable than static toolboxes. Across quantum chemistry and quantum dynamics tasks, the agent constructs, validates, and uses tools to complete end-to-end workflows with minimal manual engineering.

While our case studies demonstrate robust end-to-end execution, the system has several limitations. First, we have not exhaustively audited all agent-produced code and trajectories; in some runs, the agent re-implemented functionality that already exists in standard libraries, which can introduce bugs. Second, our current tool-forging pipeline targets Python-centric scientific stacks; extending to other ecosystems (e.g., R) and to domain tools that rely on specialized input formats remains future work.

Acknowledgments

A.A.-G. thanks Anders G. Frøseth for his generous support. A.A.-G. also acknowledges the generous support of Natural Resources Canada and the Canada 150 Research Chairs program. This research is part of the University of Toronto's Acceleration Consortium, which receives funding from the CFREF-2022-00042 Canada First Research Excellence Fund. This work was supported by the AI2050 program of Schmidt Sciences.

1. Detail of case studies

1.1 Relative Stabilities of Carbocations - Level 2

Problem. The goal is to compute carbocation formation thermochemistry for a set of 9 reactions of the form

$$\mathrm{R{-}H} \rightarrow \mathrm{R^+} + \mathrm{H^-}, \tag{1}$$

reporting both the reaction enthalpy ΔH and Gibbs free energy ΔG (kcal/mol). The input is a list of SMILES strings for the neutral precursors (R–H) and the corresponding cations (R^+), with prescribed charge/multiplicity for each species. All structures are to be treated with DFT using B3LYP/6-31G*.

Result (Agent generated with some long content removed for conciseness).

- **Level of theory:** B3LYP/6-31G*
- **Thermochemistry:** RRHO (Rigid-Rotor Harmonic-Oscillator) at $T = 298.15$ K, $P = 1$ atm
- **Geometry optimization:** geomeTRIC optimizer in PySCF
- **Hydride (H^-):** Not geometry-optimized (single atom); thermochemistry computed at fixed position

Reaction Thermochemistry (kcal/mol).

Reaction	ΔH (kcal/mol)	ΔG (kcal/mol)
methane	354.5133	347.1268
ethane	311.5172	304.2864
propane	287.4559	279.1098
2-methylpropane	270.3726	260.2885
toluene	272.1274	265.8437
benzene	320.7142	312.5371
dimethyl ether	277.1025	269.6780
trimethylamine	239.7131	232.0149
propene	291.3527	284.0170

Imaginary Frequency Fix. Initial geometry optimizations produced saddle-point structures for 4 species with significant imaginary frequencies (> 100 $\mathrm{cm^{-1}}$). These were fixed by displacing along the imaginary normal mode and re-optimizing.

- ethane R^+ ([CH3+]C): imaginary freq. -247.0 $\mathrm{cm^{-1}}$; energy change -2.77 kcal/mol.
- propane R^+ (C[CH+]C): imaginary freq. -210.2 $\mathrm{cm^{-1}}$; energy change -1.10 kcal/mol.
- 2-methylpropane R^+ (C[C+](C)C): imaginary freqs. -199.8, -157.8 $\mathrm{cm^{-1}}$; energy change -1.36 kcal/mol.
- propene R–H (C=CC): imaginary freq. -220.0 $\mathrm{cm^{-1}}$; energy change -2.08 kcal/mol.

After the fix, all species have $n_{\text{imaginary}} = 0$ (confirmed true minima).

1.2 Unitary vs. Lindblad Qubit Dynamics

Problem. This case study compares single-qubit dynamics under (i) closed-system unitary time evolution and (ii) dissipative Lindblad time evolution. The system is specified by the Hamiltonian

$$H = X + 0.5\,Z, \tag{2}$$

two initial states (the X eigenstates $|+\rangle_X$ and $|-\rangle_X$), and a collapse operator

$$C = \sqrt{\gamma}\,\sigma^-, \qquad \gamma = 0.2. \tag{3}$$

For a uniform time grid on $t \in [0, 40]$ (400 steps), the output of interest is the expectation value $\langle Z\rangle(t)$ for each initial state, computed once for unitary evolution and once for Lindblad evolution.

Result (Agent generated with some long content removed for conciseness). The time evolution of $\langle Z\rangle(t)$ was computed on a uniform grid $t \in [0, 40]$ (400 steps) for both unitary evolution (QuTiP `sesolve`) and Lindblad evolution with amplitude damping (QuTiP `mesolve`, $\gamma = 0.2$). Unitary dynamics exhibit sustained coherent oscillations, while Lindblad dynamics show damping and long-time deviation.

Initial state	$\langle Z\rangle_{\text{unit}}(40)$	$\langle Z\rangle_{\text{L}}(40)$	RMSE	Max diff.
$\|+\rangle_X$	0.362943	-0.329959	0.652900	1.124317
$\|-\rangle_X$	-0.362943	-0.336168	0.246975	0.460193

Table 1: Summary comparison of unitary vs. Lindblad evolution for $\langle Z\rangle(t)$ (final-time values and trajectory-level differences).

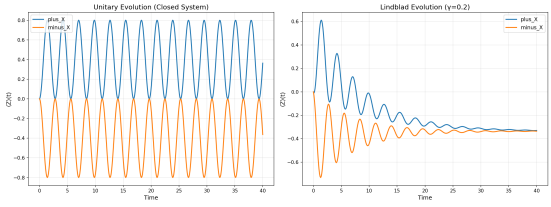

Fig. 5: Side-by-side comparison of unitary (left) and Lindblad (right) dynamics for both initial states.

2. List of tasks

El Agente Q tasks

Each of the problem contains two levels with the first level easier to solve (except "Organic Molecule Analysis" which has three levels). In all settings, we

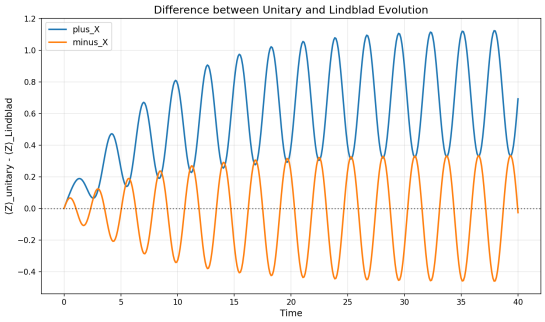

Fig. 6: Difference signal $\langle Z \rangle_{\text{unitary}}(t) - \langle Z \rangle_{\text{Lindblad}}(t)$, highlighting cumulative effects of dissipation.

solve Level 1 first and then proceed to Level 2. In total, there are 13 problems. The detail of the problems can be found in [5].

1. **Organic Molecule Analysis**: Optimize organic molecules from chemical names and report key properties including geometry, symmetry, dipole moment, molecular orbitals, atomic charges, and visualization.

2. **Inorganic Molecule Analysis**: Optimize inorganic and organometallic compounds and generate molecular property reports analogous to the organic case.

3. **Carbocation Stability**: Compute reaction enthalpy and Gibbs free energy for carbocation formation to assess relative stability.

4. **Cycloalkane Ring Strain**: Calculate enthalpy and Gibbs free energy changes associated with ring strain in cycloalkanes of varying sizes.

5. **pKa of Carboxylic Acids**: Compute pKa values of carboxylic acids using Gibbs free energies with implicit solvation and calibration to reference acids.

6. **Electronic Absorption Spectra**: Compute singlet excitation energies, oscillator strengths, and singlet–triplet gaps of organic molecules using TDDFT.

El Agente Cuántico Tasks

The detail of the problems can be found in [6].

1. **Bell State Preparation and Measurement**: Prepare a Bell state in a two-qubit circuit and evaluate correlations in different measurement bases.

2. **H_2 VQE Dissociation Curve**: Compute the molecular energy of H_2 as a function of bond distance using VQE and compare with exact results.

3. **Gibbs State via Imaginary-Time Evolution**: Prepare a thermal state of a Hubbard chain using imaginary-time evolution and validate thermalization.

4. **Transverse-Field Ising Model Dynamics**: Simulate magnetization dynamics in different quantum phases of the transverse-field Ising model.

5. **Open vs Closed Qubit Dynamics**: Compare unitary Schrödinger evolution and Lindblad dynamics for a single qubit with dissipation.

6. **FMO Exciton Transport**: Simulate exciton energy transfer in the Fenna–Matthews–Olson complex at different temperatures using HEOM.

7. **Lambda-System Optimal Control (GRAPE)**: Optimize control pulses for population transfer in a three-level system while suppressing intermediate-state occupation.

8. **Floquet Kicked Ising Chain**: Study nonequilibrium dynamics of a periodically driven Ising spin chain.

9. **Disordered Floquet Spin Chain**: Investigate many-body localization and time-crystal behavior using TD-DMRG with disorder averaging.

10. **Cluster Model Phase Transition**: Detect symmetry-protected topological phase transitions via string-order parameters.

11. **Noisy Bell Correlations**: Analyze the degradation of Bell-state correlations under depolarizing noise.

References

[1] OpenAI. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

[2] Gemini Team. Gemini: A family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*, 2023.

[3] Anthropic PBC. Claude code, 2025. `https://code.claude.com/docs/en/overview` [Accessed: 2025-12-05].

[4] OpenAI. Codex, 2025. `https://openai.com/index/introducing-codex/` [Accessed: 2025-12-05].

[5] Yunheng Zou, Austin H Cheng, Abdulrahman Aldossary, Jiaru Bai, Shi Xuan Leong, Jorge Arturo Campos-Gonzalez-Angulo, Changhyeok Choi, Cher Tian Ser, Gary Tom, Andrew Wang, et al. El agente: An autonomous agent for quantum chemistry. *Matter*, 8(7), 2025.

[6] Ignacio Gustin, Luis Mantilla Calderón, Juan B Pérez-Sánchez, Jérôme F Gonthier, Yuma Nakamura, Karthik Panicker, Manav Ramprasad, Zijian Zhang, Yunheng Zou, Varinia Bernales, et al. El agente cuántico: Automating quantum simulations. *arXiv preprint arXiv:2512.18847*, 2025.

[7] Qiming Sun, Xing Zhang, Samragni Banerjee, Peng Bao, Marc Barbry, Nick S Blunt, Nikolay A Bogdanov, George H Booth, Jia Chen, Zhi-Hao Cui,

et al. Recent developments in the pyscf program package. *The Journal of chemical physics*, 153(2), 2020.

[8] Jin-Sung Kim, Alex McCaskey, Bettina Heim, Manish Modani, Sam Stanwyck, and Timothy Costa. Cuda quantum: The platform for integrated quantum-classical computing. In *2023 60th ACM/IEEE Design Automation Conference (DAC)*, pages 1–4. IEEE, 2023.

[9] Neill Lambert, Eric Gigu'ere, Paul Menczel, Boxi Li, Patrick Hopf, Gerardo Su'arez, Marc Gali, Jake Lishman, Rushiraj Gadhvi, Rochisha Agarwal, Asier Galicia, Nathan Shammah, Paul Nation, J. R. Johansson, Shahnawaz Ahmed, Simon Cross, Alexander Pitchford, and Franco Nori. Qutip 5: The quantum toolbox in Python. *Physics Reports*, 1153:1–62, 2026.