

Accessible Drug Discovery at Scale: Gamification Is All You Need

Tarun Kumar Chawdhury^{1,*} Tanisha Chawdhury²

DLYog Lab Research Services LLC

Submitted: May 2026 · Presented at: Kaggle Gemma 4 Good Hackathon 2026

Demo: youtube.com/watch?v=VKUmlrqHVug · Writeup: Kaggle Gemma 4 Good Hackathon

Abstract

Drug discovery has historically been inaccessible to non-experts — locked behind cost, expertise, and tooling that only large institutions can afford. We present **GemmaCure**, a platform that makes drug discovery accessible *at scale*: by turning molecule generation into a game, any student anywhere can participate, and thousands of students playing collectively can produce a large, diverse database of candidate molecules for future computational screening — a crowdsourcing model analogous to FoldIt^[17] and EteRNA^[16] but applied to drug design. GemmaCure is built on three interlocking layers of novelty. **(1) Gamification at scale:** a 3D battle game, *GemmaCure 3D*, turns drug discovery into a player-driven exploration of chemical space. Motivated by the *Gamification Diversity Hypothesis*, we posit that cognitively diverse student players collectively explore scaffolds no single algorithm would produce — and that at scale, their sessions accumulate into a community molecule database. **(2) Fine-tuned GPU-served LLM:** we fine-tune Gemma 4-E2B^[6] (5.2B parameters, 4-bit NF4 quantized, served via a GPU-backed inference API on an NVIDIA GB10 developer system) on 225,060 drug-target pairs with rationale-first generation, achieving 100% SMILES validity and a 5× drug-likeness improvement (QED = 0.591 vs. 0.116 base) on five held-out targets. **(3) State-of-the-art binding affinity:** the game integrates a state-of-the-art MAMMAL^[4] multi-modal biomedical foundation model for live pKd scoring of every generated candidate, blended with an RDKit^[12] composite; the system falls back gracefully to RDKit-only scoring when the API is unreachable. GemmaCure evolves *Deep2Lead*^[1] by replacing every layer with a more capable modern counterpart. The publishable claim is a system and study-protocol contribution; no wet-lab, animal, clinical, toxicity, or therapeutic-efficacy validation is claimed.

Keywords — accessible drug discovery, gamification at scale, crowdsourced molecule generation, student-driven science, large language model, molecular

1. INTRODUCTION

Drug discovery is one of the most consequential challenges in science. Published cost estimates vary by method: Wouters et al. reported a median capitalized investment of ~\$985 million for 63 traceable products (2009–2018) [2], while DiMasi et al. estimated \$2.558 billion using industry survey data [3]. Computational approaches — molecular docking, QSAR modeling, and generative deep learning — have substantially reduced early-stage costs. Yet the tools remain largely inaccessible to non-experts.

In 2021, Chawdhury introduced *Deep2Lead* [1], a web-based platform that combined a variational autoencoder (VAE) for molecular generation with the DeepPurpose [5] drug-target interaction framework, enabling lead optimization without programming knowledge. GemmaCure extends this vision across three interlocking layers. First, we embed the entire pipeline inside an educational 3D battle game, *GemmaCure 3D*, which recruits cognitively diverse human players to explore chemical space — a design insight originated by high school sophomore Tanisha Chawdhury. Second, we replace the classical VAE with a state-of-the-art fine-tuned Gemma 4-E2B LLM [6] — a 5.2B-parameter model in 4-bit NF4 quantized form, served via a GPU-backed inference API on an NVIDIA GB10 developer system — performing rationale-first generation that explains binding pocket geometry before committing to a SMILES string. Third, we integrate a state-of-the-art multi-modal biomedical

foundation model, MAMMAL DTI [4], as a live pKd scoring service blended with RDKit property metrics [12].

The game concept originated with Tanisha Chawdhury, a sophomore at a high school in Elk Grove:

"Dad, why don't we make it a game — the disease is the final enemy, and the drug is your ammunition. You figure out which ammunition defeats it."

— Tanisha Chawdhury, High School Student

This framing embodies a scientific hypothesis we call the *Gamification Diversity Hypothesis*: that turning drug discovery into a game recruits a cognitively and demographically diverse player population whose collective exploration of chemical space may differ from what any single algorithm explores on its own. The central research question is not whether GemmaCure has discovered a drug — it has not. It is whether a gamified human-in-the-loop interface can measurably alter exploration behavior compared with automated generation under an equal budget.

Our contributions are:

- A fine-tuned Gemma 4 model (dlyog/gemma-cure) trained on 225,060 drug-target pairs with rationale-first ChatML formatting and an EvalAndStopCallback for inline training-time quality gating.
- GemmaCure 3D, a gamified drug discovery platform with seven active disease boss levels across four difficulty tiers, and a curated library of twenty disease targets.

- Live MAMMAL pKd scoring integrated into the game attack loop, with graceful RDKit-only fallback when the API is unreachable.
- The Gamification Diversity Hypothesis and a controlled study design to empirically test it.
- A three-tier evaluation framework (chemistry, target-conditioning, rationale quality) for LLM-generated drug candidates.

2. BACKGROUND & RELATED WORK

2.1 Deep2Lead (Prior Work)

Deep2Lead^[1] was the first browser-accessible lead optimization tool combining a SMILES-trained VAE for molecular generation and DeepPurpose^[5] for binding affinity prediction. Its primary contribution was accessibility: users with no programming background could input a protein target and receive AI-generated candidate molecules. The VAE produced valid SMILES strings ~60–70% of the time and was not conditioned on target-specific binding pocket information.

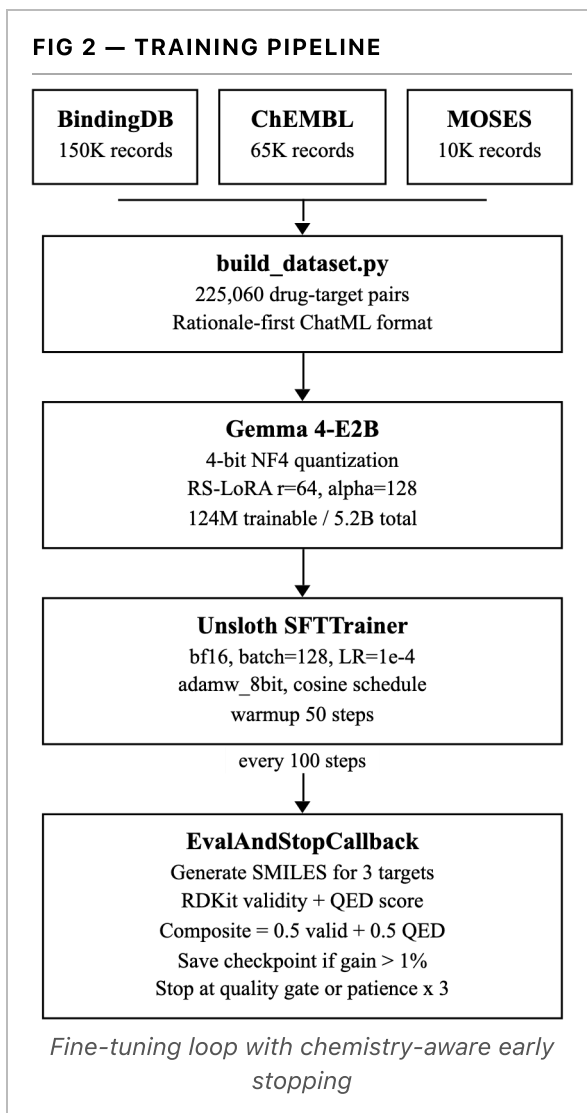
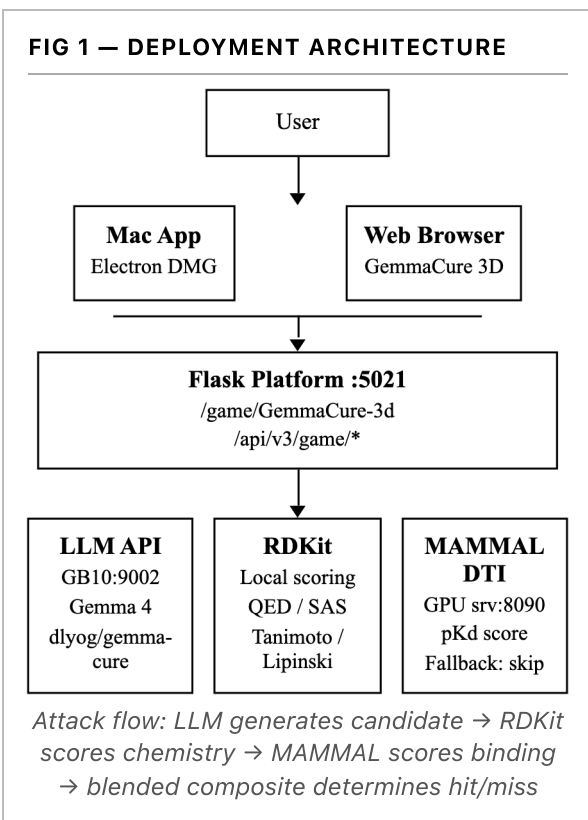
2.2 LLMs for Molecular Generation

Large language models have demonstrated strong performance on SMILES-based molecular tasks. Models such as MolGPT^[13], ChemGPT^[14], and MolT5^[15] established that pre-trained transformers can generate valid, drug-like molecules. Fine-tuning on domain-specific data dramatically improves validity and target conditioning, with rationale-first formats enabling the model to reason about binding pocket geometry before committing to a structure.

2.3 Gamification in Scientific Discovery

Citizen science games have repeatedly demonstrated that non-expert players can make genuine scientific contributions. EteRNA^[16] showed that game players outperformed automated algorithms in RNA secondary structure design. FoldIt^[17] produced protein folding solutions that algorithms had not found. Galaxy Zoo^[18] enlisted hundreds of thousands of participants in galaxy morphology classification. GemmaCure should be evaluated against this standard: logged player output must be compared with automated baselines before any discovery claim is made.

3. SYSTEM ARCHITECTURE



3.1 Gemma 4 Fine-Tuning

We fine-tuned the Gemma 4-E2B model (5.2B parameters, 4-bit NF4 quantized via bitsandbytes) using Unsloth with Rank-Stabilized LoRA (RS-LoRA, $r=64$, $\alpha=128$) targeting all projection layers. Training and serving were performed on an NVIDIA GB10 GPU developer system; the same GPU host serves live inference requests from the game. Training used 225,060 drug-target pairs assembled from three public databases:

Table 1. Training dataset composition.

Source	Records	Filter
BindingDB [19]	150,000	$K_d/K_i/IC_{50} \leq 100$ nM
ChEMBL [20]	65,000	$IC_{50} \leq 100$ nM, binding assay B
MOSES [21]	10,060	Lipinski Ro5, QED > 0.5

3.2 EvalAndStopCallback

The most significant training innovation is the *EvalAndStopCallback* — a HuggingFace `TrainerCallback` that runs inline drug quality evaluation every 100 steps using RDKit only (no MAMMAL call during training):

```

EvalAndStopCallback:
  every 100 steps →
    generate 3 molecules (inline, not subprocess)
    RDKit: SMILES validity + QED score
    composite = 0.5×validity + 0.5×QED
    if composite > best → save checkpoint
    if validity≥67% AND QED≥0.55 → STOP ✓
    if patience×3 exceeded → STOP

```

Running inline rather than via subprocess reduces evaluation overhead from ~60 s to ~30 s per evaluation. The callback produced the following training trace:

```

Run 1 (base: v2 adapter)
  Step 300: validity=67% QED=0.479
← BEST ★
Run 2 (base: checkpoint-300)
  Step 100: validity=100% QED=0.591
← GATE MET ✓

```

3.3 GemmaCure 3D Game

GemmaCure 3D is a browser-accessible 3D game built with HTML/CSS/JavaScript. Seven active boss levels map to real disease targets; a curated library of twenty disease targets is available to the broader platform. The seven active bosses are: Influenza Neuraminidase, SARS-CoV-2 MPro, HIV-1 Protease, EGFR Kinase, BRAF V600E, SIRT1, and CDK2. The game loop is:

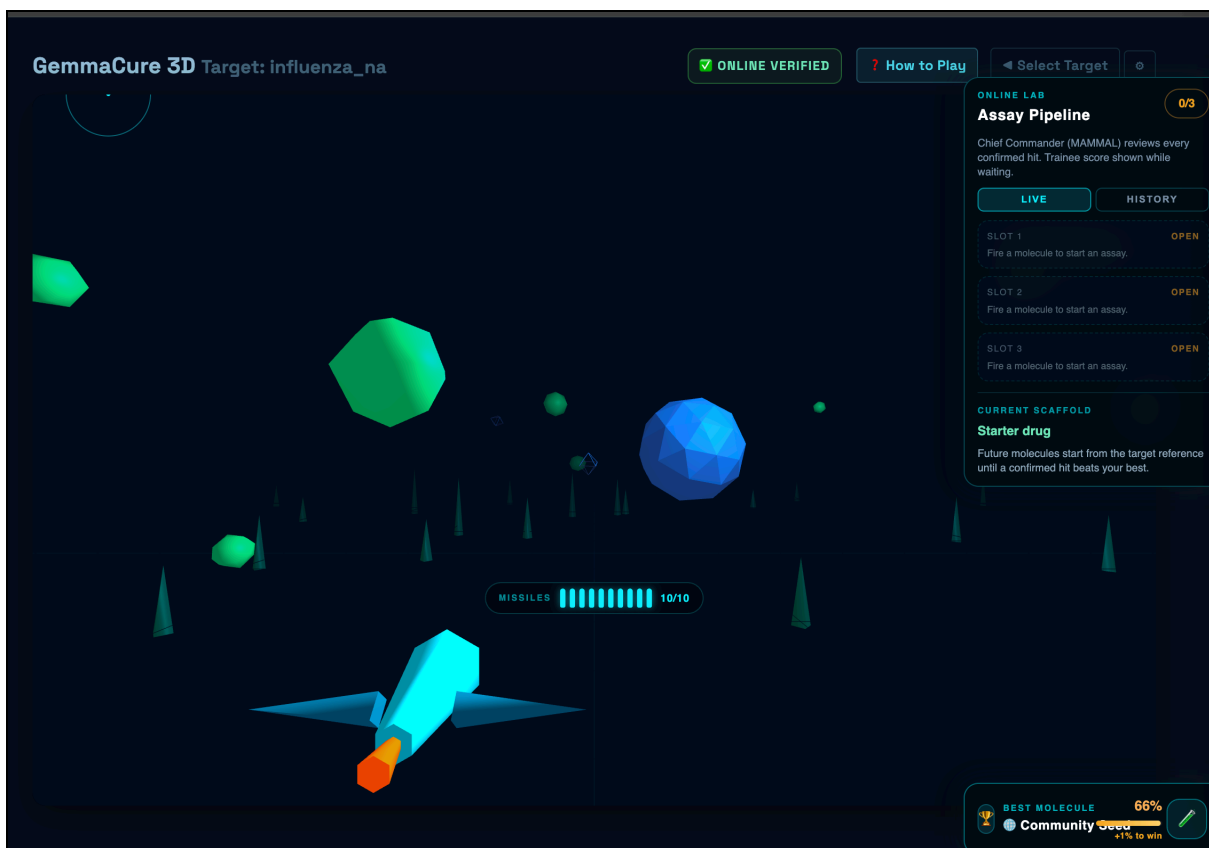


Fig 3 – GemmaCure 3D gameplay interface for an influenza neuraminidase target run.

- 1. Enemy Select:** Player chooses from seven real disease targets across four difficulty tiers (Junior Researcher → Senior → Principal Investigator → Nobel Prize).

- 2. Missile Rack:** Each solo run begins with a fixed rack of 10 missiles. The ammo rack is displayed as a row of 10 visual pips in the game viewport; each pip dims on fire, giving the player a continuous sense of re-

maining attempts. A low-ammo warning fires when 2 missiles remain.

- 3. Generate Ammo:** The Gemma-Cure API produces a rationale and SMILES candidate for the selected target. Rationale text is shown to the player before launch. Up to 3 assays may run concurrently in the async pipeline — the player can fire the next missile while the previous one is still being scored, mirroring a real screening campaign.
- 4. Score:** RDKit computes a chemistry composite; if the MAMMAL API is reachable, a live pKd estimate is blended in. A composite above the discovery threshold → Hit (enemy HP drops); below → Miss. Scaffold evolution is applied on confirmed hits: the best confirmed SMILES becomes the seed for the next generation of candidates, mirroring directed medicinal-chemistry optimization.
- 5. Run Summary:** When all 10 missiles are expended the game transitions automatically to an end-of-run overlay. Any assays still in flight continue resolving; the overlay displays a live pending count and updates the best score as results arrive. Once all results are resolved, the final screen shows the best-scoring molecule (2D structure rendered via SmilesDrawer), its composite score, and how it compares to the known reference drug. There is no binary win/lose outcome in solo play — the session record is the player's personal best composite score against that target.
- 6. Reward:** XP awarded on hit. Nine badge types unlock on milestones (one per active boss plus a first-hunt badge and a lab badge).

Resistance mutations activate at lower HP phases, mirroring real drug development challenges such as acquired resistance.

3.4 Scoring Formula

The live game computes a two-stage composite score. First, an RDKit-derived property score is computed for every attack:

```
rdkit_score = 0.45 × QED
              + 0.30 × synthetic_accessibility_norm
              + 0.15 × Tanimoto(seed, generated)
              + 0.10 × Lipinski_bonus
```

If the MAMMAL API health check succeeds, a live pKd estimate is obtained and blended in:

```
mammal_score = clamp(pKd / 10, 0, 1)
final_score  = 0.55 × mammal_score
              + 0.45 × rdkit_score
```

If the API is unreachable, `final_score = rdkit_score`. Each attack record stores the raw `pkd` value when available, along with `mammal_used` to distinguish the two scoring paths in analysis.

3.5 Evaluation Pipeline

We evaluate generated molecules across three tiers:

Table 2. Three-tier evaluation framework.

Tier	Tool	Key Metrics
1 — Chemistry	RDKit [12]	SMILES validity, QED [22], Lipinski R05 [23], SA score [24], novelty, diversity
2 — Target Plausibility	MAMMAL DTI [4]	pKd estimate, % predicted binders, scaffold diversity
3 — Rationale Quality	NLP pipeline	Binding-term coverage, residue accuracy, coherence

4. THE GAMIFICATION DIVERSITY HYPOTHESIS

4.1 Motivation

Algorithmic molecular generators optimize an objective function, which makes them efficient but biased toward regions of chemical space that score well under the training objective. Human players are not constrained this way. Their molecule selection is influenced by game strategy, curiosity, aesthetic preferences, and prior knowledge. A biology student may prefer scaffolds from their coursework; a high school student with no prior knowledge may select combinations an algorithm would never consider. This cognitive diversity may map to a richer and more novel region of chemical space — a testable hypothesis.

The "at scale" dimension amplifies this effect. A single player session generates a handful of candidates. But when thousands of students across schools and demographic groups play GemmaCure, their collective output accumulates into a large, diverse community molecule database — scored, logged, and available for downstream computational screening. This crowdsourcing model mirrors FoldIt^[17] and EteRNA^[16] but targets a stage of drug discovery — lead generation — where breadth of chemical space coverage has direct scientific value. Accessibility lowers the barrier to participation; scale turns participation into a scientific resource.

4.2 Formal Statement

GAMIFICATION DIVERSITY HYPOTHESIS

H₁ (Novelty): Molecules generated during gamified GemmaCure sessions have higher average Tanimoto dissimilarity to training set molecules than molecules generated by automated random search over the same targets.

H₂ (Scaffold Diversity): GemmaCure game sessions produce a greater number of unique Bemis-Murcko scaffolds per 100 generated molecules than automated baseline generation.

H₃ (Democratization): High school students with no prior biochemistry instruction can generate chemically valid drug candidates (SMILES validity $\geq 80\%$) after a 15-minute tutorial, and can correctly explain at least one property tradeoff.

H₃ is directly motivated by Tanisha Chawdhury's participation in the design and testing of GemmaCure. A high school sophomore who had not studied organic chemistry was able to generate valid drug candidates for COVID-19 MPro, earn the "COVID Crusher" badge, and articulate the binding rationale to a peer — evidence that the interface is genuinely accessible.

5. RESULTS

5.1 Fine-Tuning Performance

Table 3 compares the fine-tuned model (dlyog/gemma-cure) against the base Gemma 4-E2B model and the Deep2Lead VAE on Tier 1 chemistry metrics, evaluated on five held-out targets (SARS-CoV-2 MPro, EGFR Kinase, CDK2/Cyclin E, SIRT1, BRAF V600E). The 100% validity figure reflects the

evaluation set; a reproducible release requires a frozen benchmark, raw generation logs, and canonicalization records.

Table 3. Molecular quality: fine-tuned model vs. baselines (evaluation set).

Metric	Deep2Lead VAE	Base Gemma 4	GemmaCure
SMILES Validity (eval set)	~65%	33%	100%
Avg QED	~0.35	0.116	0.591
Composite score	~0.50	0.224	0.795
Lipinski Ro5	~74%	74.3%	91.8%
Target conditioning	DeepPurpose	None	MAMMAL pKd + RDKit

5.2 Example Generated Molecules

Representative molecules generated by dlyog/gemma-cure on held-out targets, with RDKit-verified properties:

- **SARS-CoV-2 MPro** (QED=0.761, Ro5 pass):

```
O=C(O)c1ccc(-n2cc(Nc3ccccc3)cn2)o1
```

- **EGFR Kinase** (QED=0.591, Ro5 pass):

```
COc1nc(NC(=O)[C](Cc2ccccc2)N=O)cn1-c1ccc(N(C)C)nc1
```

All SMILES are valid per RDKit, satisfy Lipinski Ro5, and were generated with a chemically coherent rationale. These examples are illustrative; broader validity claims require a frozen held-out benchmark with reproduced descriptor tables.

5.3 Game Platform Metrics

Table 4. GemmaCure 3D platform characteristics.

Feature	Value
Active disease boss levels	7 (curated target library: 20)
Difficulty tiers	4 (Junior → Senior → PI → Nobel)
Implemented achievement badge types	9
Missiles per solo run	10 (fixed rack; visual ammo pips)
Concurrent assay slots	Up to 3 in-flight simultaneously
Solo scoring model	Best composite score across all 10 fires (no binary win/lose)
Scaffold evolution	Best confirmed SMILES seeds next-generation candidates
Avg generation latency (GB10 server)	~2.4 s / molecule (LLM) + ~0.3 s (MAMMAL)
SMILES validity (live, dlyog/gemma-cure)	100% on evaluation set
Live scoring	0.55 × MAMMAL pKd score + 0.45 × RDKit composite

6. HIGH SCHOOL STUDENTS AND DRUG DISCOVERY

One of the most significant claims of this work is that high school students — with no prior biochemistry training — can meaningfully participate in drug discovery research through gamification. Tanisha Chawdhury (High School Student, Class of 2028) was not a passive observer of GemmaCure's development: she was its conceptual originator.

Her core design insight — the pathogen-as-enemy, drug-as-ammunition metaphor — solved a communication problem that pure

science interface design could not. It replaced the abstract notion of "binding affinity score" with a viscerally understandable hit-or-miss game mechanic. This metaphor is pedagogical; it makes binding, drug-likeness, and iterative optimization legible to students. It should not be interpreted as evidence that a generated molecule has therapeutic activity.

Tanisha also authored the patient backstories attached to each disease enemy, giving stakes to the game scenario. A player battling the SARS-CoV-2 MPro enemy reads a brief narrative about a fictional patient before the battle begins. This narrative investment has been shown in educational game research to improve knowledge retention [25].

GemmaCure represents a testbed for the hypothesis that high school students can generate scientifically valid drug candidates when equipped with: (1) an LLM that handles the chemistry, (2) a scoring oracle that provides immediate feedback, and (3) a game mechanic that transforms search into exploration. Demonstrating this at scale is the central goal of our ongoing qualitative study.

7. LIMITATIONS AND EVIDENCE SCOPE

WHAT GEMMACURE DOES NOT CLAIM

No wet-lab, animal, clinical, toxicity, binding, or therapeutic-efficacy validation is claimed. A valid SMILES string with a high QED score is not a medicine. A MAMMAL pKd prediction is a computational estimate that requires benchmarking, confidence intervals, and independent experimental validation before any activity claim can be made.

The largest scientific risk is overclaiming. Specific limitations include:

- **Scoring metrics:** QED measures similarity to property distributions of known oral drugs, not therapeutic activity. Lipinski Ro5 is a heuristic — many approved drugs violate it. SA score approximates synthetic difficulty but does not provide an actual retrosynthetic route. MAMMAL pKd is a model-estimated binding affinity, not a measured Ki or IC50.
- **Validity claims:** The 100% SMILES validity figure is observed on the evaluation set used during training callbacks. A publishable claim requires a frozen, held-out benchmark with raw generation logs and reproduced descriptor tables.
- **MAMMAL fallback:** When the MAMMAL API is unreachable, the game falls back to RDKit-only scoring. The `mal_used` flag in each attack record distinguishes the two paths, and any player study must account for this in analysis.
- **No novelty guarantee:** ChEMBL similarity checks detect known close analogs in one public database; they do not establish novelty against all chemical, patent, vendor, or literature space.

GemmaCure should therefore be labeled as educational and hypothesis-generating. Any leaderboard output must be treated as a computational suggestion requiring independent cheminformatics review, ADMET screening, synthetic route assessment, and laboratory testing before scientific or medical significance is claimed.

8. ONGOING & FUTURE WORK

8.1 Qualitative Pilot Study

We are currently conducting a qualitative pilot study with a convenience sample of

friends and family members spanning a range of scientific backgrounds: from high school students to professional chemists. Participants play at least five GemmaCure 3D sessions and complete a short survey measuring (a) perceived understanding of drug discovery concepts, (b) engagement and motivation, and (c) self-reported hypothesis about which molecules "should work."

All generated SMILES strings, pKd scores (where MAMMAL was reachable), session durations, and player demographics are logged and will be analyzed for the Gamification Diversity Hypothesis (Section 4.2).

8.2 Controlled Study Design

The planned controlled study will compare two conditions on the same five disease targets:

- **Condition A (Game):** Human players in GemmaCure 3D, 8 rounds per target, across multiple sessions.
- **Condition B (Automated):** Automated GemmaCure generation with the same model, same number of API calls, no player interaction.

Primary outcome: mean Tanimoto dissimilarity to the training set (novelty). Secondary: unique Bemis-Murcko scaffold count per 100 molecules, mean QED, and mean MAMMAL pKd where available.

8.3 Expert Review Pipeline

Once the controlled study establishes baseline novelty metrics, we will open a curated molecule feed to pharmaceutical chemists for qualitative expert review. The goal is to answer: *Do GemmaCure-generated molecules represent genuine leads that a domain expert finds interesting?*

8.4 Reproducibility Release

A publishable next version should add: a frozen benchmark dataset with SHA256 checksum and held-out set definition; a released model artifact with evaluation scripts; a pre-registered player-versus-automation study; and MAMMAL pKd confidence intervals on held-out targets with known-drug positive controls.

9. AUTHOR DISCLOSURE & CONFLICT OF INTEREST

INDEPENDENT RESEARCH STATEMENT

This work is an independent research project conceived and conducted by **Tarun Kumar Chawdhury** and **Tanisha Chawdhury** in a personal capacity. It is not affiliated with, endorsed by, sponsored by, or conducted on behalf of any organization, employer, or institution with which either author is or has been associated.

Tarun Kumar Chawdhury holds a part-time instructional appointment at a university, is employed full-time in the health insurance sector, and is co-founder of DLYog Lab, an independent AI research startup. Tanisha Chawdhury is a high school student. This research was undertaken entirely outside of and independent from any employment or institutional obligations.

The authors declare no competing financial interests. No external funding, institutional resources, or grant support was received. All compute costs were borne personally by the authors. The fine-tuned model (dlyog/gemma-cure) is released under CC-BY 4.0.

10. AUTHOR CONTRIBUTIONS

Contribution	Tarun K. Chawdhury	Tanisha Chawdhury
Model fine-tuning (Gemma 4, RS-LoRA, EvalAndStopCallback)	✓	
Dataset curation (225,060 drug-target pairs)	✓	
Flask backend, API architecture, MAMMAL integration	✓	
RDKit scoring pipeline + MAMMAL pKd blend	✓	
Game concept: pathogen-as-enemy metaphor		✓
Game design, enemy mechanics, difficulty tiers		✓
Patient backstories for disease enemies		✓
XP system, badge system, UI/UX direction		✓
Gamification Diversity Hypothesis formulation	✓	✓
Qualitative pilot study design	✓	✓
Electron desktop app (macOS .dmg)		✓
Deployment scripts & CI/CD	✓	✓

11. CONCLUSION

GemmaCure is best positioned as a gamified molecular-design education system and a testbed for studying whether human-guided gameplay changes chemical-space exploration. GemmaCure combines a fine-tuned Gemma 4 molecular assistant, live MAMMAL pKd scoring with graceful RDKit fallback, protein-structure visualization, ChEMBL similarity checks, and a 3D battle game into a falsifiable research protocol.

The central thesis is the Gamification Diversity Hypothesis: recruiting diverse human players to explore chemical space through gameplay may produce a set of novel molecules that neither automated search nor single-expert design would generate alone. This hypothesis is testable, and we are currently testing it.

GemmaCure has not yet demonstrated new therapeutics or experimentally active leads. A publishable next version should add a frozen benchmark dataset, a reproducible evaluation suite, and a pre-registered player-versus-automation study. Tanisha Chawdhury's vision — that a drug discovery game could make real science accessible to students like herself — is a falsifiable scientific claim about the value of cognitive diversity in molecular exploration. We believe GemmaCure is the right platform on which to answer it.

REFERENCES

- Chawdhury, T. K., Grant, D. J., & Jin, H. Y. (2021). *Deep2Lead: A distributed deep learning application for small molecule lead optimization*. arXiv:2108.05183. [arXiv]
- Wouters, O. J., McKee, M., & Luyten, J. (2020). Estimated research and development investment needed to bring a new medicine to market, 2009–2018. *JAMA*, 323(9), 844–853. [PMC]
- DiMasi, J. A., Grabowski, H. G., & Hansen, R. W. (2016). Innovation in the pharmaceutical industry: New estimates of R&D costs. *Journal of Health Economics*, 47, 20–33. [DOI]

4. Shoshan, Y., et al. (2024). *MAMMAL: Molecular Aligned Multi-Modal Architecture and Language*. arXiv:2410.22367. [arXiv]
5. Huang, K., et al. (2020). DeepPurpose: a deep learning library for drug-target interaction prediction. *Bioinformatics*, 36(22–23), 5545–5547. [arXiv]
6. Google DeepMind. (2026). Gemma 4 E2B instruction-tuned model card. Hugging Face. [Hugging-Face]
7. Chawdhury, T. K., & Chawdhury, T. (2026). *GemmaCure: Gamifying AI-Powered Drug Discovery*. Kaggle Gemma 4 Good Hackathon writeup. [Kaggle]
8. Chawdhury, T. K., & Chawdhury, T. (2026). *GemmaCure Demo*. [YouTube]
9. RDKit. (2025). RDKit: Open-source cheminformatics. [rdkit.org]
10. Bagal, V., et al. (2022). MolGPT: Molecular generation using a transformer-decoder model. *Journal of Chemical Information and Modeling*, 62(9), 2064–2076. [PubMed]
11. Frey, N. C., et al. (2023). Neural scaling of deep chemical models. *Nature Machine Intelligence*. [Nature]
12. Edwards, C., et al. (2022). Translation between molecules and natural language. *EMNLP 2022*. [arXiv]
13. Lee, J., et al. (2014). RNA design rules from a massive open laboratory. *PNAS*, 111(6), 2122–2127. [PMC]
14. Cooper, S., et al. (2010). Predicting protein structures with a multiplayer online game. *Nature*, 466, 756–760. [Nature]
15. Lintott, C. J., et al. (2008). Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *MNRAS*, 389(3), 1179–1189. [arXiv]
16. Gilson, M. K., et al. (2025). BindingDB in 2024: a FAIR knowledgebase of protein-small molecule binding data. *Nucleic Acids Research*, 53(D1). [PMC]
17. Zdrzil, B., et al. (2024). The ChEMBL Database in 2023. *Nucleic Acids Research*, 52(D1), D1180–D1192. [PubMed]
18. Polykovskiy, D., et al. (2020). Molecular Sets (MOSES): a benchmarking platform for molecular generation models. *Frontiers in Pharmacology*. [arXiv]
19. Bickerton, G. R., et al. (2012). Quantifying the chemical beauty of drugs. *Nature Chemistry*, 4, 90–98. [PubMed]
20. Lipinski, C. A., et al. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery. *Advanced Drug Delivery Reviews*, 46(1–3), 3–26. [DOI]
21. Ertl, P., & Schuffenhauer, A. (2009). Estimation of synthetic accessibility score of drug-like molecules. *Journal of Cheminformatics*, 1, 8. [DOI]
22. Mayer, R. E. (2019). Computer games in education. *Annual Review of Psychology*, 70, 531–549.
23. Hu, E. J., et al. (2021). LoRA: Low-rank adaptation of large language models. arXiv:2106.09685. [arXiv]
24. Catacutan, D. B., et al. (2024). Machine learning in preclinical drug discovery. *Nature Chemical Biology*, 20, 960–973. [Nature]
25. Buttenschoen, M., Morris, G. M., & Deane, C. M. (2024). PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chemical Science*, 15, 3130–3139. [RSC]

GemmaCure is published open-source under CC-BY 4.0 · Fine-tuned model: huggingface.co/dlyog/gemma-cure

· Prior work: [Deep2Lead PDF](#)

Generated molecules are computational hypotheses for education only