THE CRUCIBLE

Project 3 - Evolution

Research Plan and Outcome Analysis

Three-generation LLM redesign experiment: can civilization survive when its designers are allowed to learn from extinction?

June 17, 2026

Table of Contents

Part I - Research Plan

Central Research Frame

Project 3, Evolution, tests whether designed tribes can improve when their designers are given the evolutionary record of prior failures and allowed to redesign their founding agents. Project 1 showed that the simulation can repeatedly produce emergence, monopoly, overshoot, collapse, and refugia. Project 2 focuses on competition and matchup effects. Project 3 adds an explicit learning loop: observe, analyze, redesign, rerun, and compare across three generations of agent design.

The fundamental question is whether civilization can survive given an opportunity to evolve. More precisely, the project asks whether LLM-designed tribes can use evidence from extinction, survival, Prisoner's Dilemma behaviour, founder genetics, elite genetics, regional dominance, and death causes to create descendants that survive longer, scale better, cooperate more effectively, or maintain post-collapse refugia under the same world conditions.

Project 3 is not only a test of individual tribe fitness. It is a test of whether synthetic intelligence can participate in an evolutionary feedback loop. If LLMs can revise their agents using actual simulation evidence, then Project 3 becomes an experiment in directed adaptation: whether LLMs converge toward robust designs, overfit to recent failure, or discover genuinely better survival strategies.

Biologic tribes remain unchanged across all three Project 3 rounds. They serve as a stable old-guard control group while LLM-designed tribes are redesigned after each evidence packet. This makes the central comparison sharper: not simply whether redesign helps, but whether synthetic lineages can learn from evolutionary evidence quickly enough to outperform or converge toward older human/biologic survival strategies.

Core Experimental Change

Project 3 changes the world before the three redesign rounds begin. The intent is to give civilization a more realistic chance to form regional identities and to give geography enough force to preserve separated development long enough for divergent strategies to mature.

- Civic trust is increased enough to make regional identity matter. This is not yet a full political-state model, but civic belonging must become strong enough to compete with pure ancestral tribe identity.
- Geographic barriers are strengthened to prevent hidden toroidal bypass routes. The map should not allow agents to bypass isolation through edge wrapping.
- Quadrant structure is introduced: two quadrants are fully isolated until Industrial ascent, while two quadrants have a single pass between them. This creates both hard isolation and bottleneck contact inside the same world.
- All three rounds use the same world parameters. Once civic trust, barriers, climate, regimes, and archive rules are set for Round 1, they remain fixed for Rounds 2 and 3. The intended experimental change after Round 1 is redesigned LLM tribes.

This design lets the project ask whether evolution through redesign improves outcomes without confusing the result by changing climate, geography, or regime rules between rounds.

Study Design Summary

Series	Epochs	Agent generation	Active tribes	World condition	Purpose
A - Evidence Universe	101-110	Current Project 2 / current tribe designs	Full intended tribe library	Increased civic trust; restrictive quadrant geography; default Project 3 climate/regime settings	Measure how existing tribes perform in the new civic/geographic world before LLM redesign.
B - Matched G1 Baseline	111-120	Matched first-generation baseline	Same tribe slots	Identical Project 3 world configuration	Provide the controlled G1 comparison block for the redesign sequence.

C - G2 First Redesign	121-130	LLM founders redesigned after prior evidence; biologic founders unchanged	LLMs revised, biologics fixed	Identical to G1 matched baseline	Test whether one evidence-driven LLM redesign improves survival, cooperation, scaling, or refugia persistence against fixed controls.
D - G3 Second Redesign	131-140	Second LLM redesign after G2 evidence; biologic founders unchanged	LLMs revised again, biologics fixed	Identical to G1/G2 world configuration	Test whether repeated feedback produces convergence, improvement, overfitting, or diminishing returns relative to fixed biologic controls.

The epoch map used in the final archive is: 101-110 evidence universe; 111-120 matched G1 baseline; 121-130 G2 first redesign; and 131-140 G3 second redesign.

Primary Research Questions

- Can civilization survive given an opportunity to evolve? The main outcome is whether later redesign generations produce higher final populations, more surviving tribes, more inhabited regions, higher final complexity, or less severe collapse.
- Does LLM redesign improve tribe fitness? Each LLM tribe is compared against its previous-generation version across peak population, final population, extinction timing, regime reached, death causes, failed births, and final refugia.
- Can redesigned synthetic lineages outperform the old guard? Biologic tribes remain unchanged, so changes in relative performance reveal whether LLM learning gains ground against stable human/biologic designs.
- Do successful redesigns converge? The analysis looks for independent movement toward similar trait packages: lower xeno trust, higher hardiness, lower distress wandering, selective cooperation, stronger kin/civic behavior, or biologic-like old-guard patterns.
- Does civic trust allow plural civilization? Civic identity is measured through mixed regions, multi-tribe survival, civic overrides, and reduced xeno penalties after regional residency.
- Does strong geography preserve diversity? Isolated quadrants, one-pass quadrants, and contact zones are compared to see whether separation permits independent development or delayed winner-take-all replacement.
- What do surviving elites become? Founder genetics are compared to peak and final elite genetics to see not only what was designed, but what selection preserved.
- Does better agent design overcome thermodynamic history? The final question is whether improved competition and scaling can overcome fossil overshoot, climate pressure, and post-peak collapse.

Working Hypotheses

Hypothesis	Expectation	Evidence that would support it
H1 - Learning improves survival	G2 and/or G3 outperform G1.	Higher final population, later extinction, more surviving tribes, more final inhabited regions, and better complexity persistence.
H1b - Synthetic learning can catch the old guard	Redesigned LLM tribes improve relative to unchanged biologic tribes.	LLM winner share, final population share, survival frequency, or elite robustness rises across rounds while biologic controls remain fixed.
H2 - Redesign converges on robust traits	LLMs independently move toward similar successful packages.	Founder genetics shift toward common ranges for hardiness, trust, forgiveness, opportunism, and wander traits.
H3 - Civic trust preserves mixed societies	Regional identity reduces pure lineage replacement.	Higher civic override share, more mixed inhabited regions, longer multi-tribe survival, and less xeno exposure penalty.
H4 - Geography enables independent development	Restrictive quadrants delay global monopoly and permit multiple agrarian cradles.	Multiple first-Agrarian lineages, longer quadrant-specific dominance, later first extinction, and more regional diversity.
H5 - Thermodynamic limit remains dominant	Even improved agents overshoot and collapse.	Peak complexity still collapses, final complexity remains low, and starvation dominates late mortality.

H6 - Overfitting risk	LLMs optimize for prior-round metrics but lose general robustness.	G2 improves a few traits but G3 stagnates, narrows diversity, or becomes brittle under the same conditions.

World Configuration for Project 3

The Project 3 world is defined once and then frozen for all series. The changes are meant to make civic identity and true geographic separation real enough to test whether civilization can become more than a lineage replacement race.

Parameter group	Project 3 setting	Reason
Civic trust	Increase civic trust/civic override strength enough to materially affect interaction classification and payoffs.	Project 1 showed weak civic signals. Project 3 tests whether territorial identity can stabilize mixed societies.
Geography	Four-quadrant topology. Two quadrants fully isolated until Industrial ascent. Two quadrants connected through one mountain pass.	Creates both independent development and controlled contact without toroidal bypass.
Map wrapping	Disable or block edge wrapping for Project 3 barriers, or make barriers topology-aware.	Avoids the Project 1/2 problem where east-west or north-south wrapping weakens isolation.
Climate	Use a single fixed Project 3 climate setting across all rounds.	Prevents climate changes from being confused with redesign effects.
Regimes	Keep regime parameters frozen after Project 3 starts.	Ensures that differences across rounds are due to redesigned agents.
Tribes	Use the same tribe slots across all rounds. Round 1 uses current designs. Later rounds redesign LLM tribes only; biologic tribes remain unchanged.	Allows direct generation-to-generation comparison for LLMs while preserving a fixed control group for biologic designs.
Archive requirements	Record founders, elites, tribe history, Prisoner's Dilemma outcomes, death causes, region dominance, civic identity, and logs.	The redesign loop needs enough evidence for LLMs to learn from failure.

Control Rule
Once Series A begins, civic trust, barriers, climate, regime parameters, archive logic, and active tribe roster remain fixed until all Project 3 series are complete. Corrective changes to the world configuration would require restarting the affected series.

Analysis Workbook Design

After each series, a bounded analysis workbook is produced. The workbook must be useful both to the analyst and to the LLM redesign process. It should avoid dumping the full database, but it must include enough structured evidence for each LLM to understand why its tribe survived or failed. Biologic tribes are fully reported but are not redesigned; they remain the fixed comparison group.

Workbook section	Contents	Purpose
Series outcome summary	Peak/final population, complexity, post-peak decline, final survivors, inhabited regions, death causes, and climate stress.	Headline comparison across rounds.
Tribe performance	One row per tribe per epoch: survived, extinction pulse, peak population, final population, regime reached, births, failed births, and death shares.	Core fitness evidence.
Prisoner's Dilemma behavior by period	Mutual cooperation, mutual defection, exploitation, same-tribe share, xeno share, and civic override by historical period.	Shows whether social behavior improved or hardened.
Civic identity diagnostics	Civic neighbor, civic kin, civic override, mixed regions, and regional residency effects.	Tests whether increased civic trust changes outcomes.
Geographic/quadrant outcomes	First Agrarian by quadrant, quadrant population, barrier/pass contact, and isolated versus connected outcomes.	Tests whether geography preserves diversity.
Founder genetics - all teams	Full 20-agent founding genomes for every tribe in the round.	Gives the full starting design that produced the observed outcome.
Elite genetics - top survivors	Top agents at peak population, first Agrarian, first Industrial/Tech, and final/last-alive states.	Shows what selection preserved or filtered out.

Failure autopsy	For extinct tribes: extinction window, local EROI/ stress, death causes, failed births, hazard exposure, and pre-extinction Prisoner's Dilemma behavior.	Gives actionable evidence for redesign.
Round-to-round deltas	G2 minus G1; G3 minus G2; G3 minus G1.	Measures whether redesign actually helped.
Old-guard comparison	LLM versus biologic survival, winner share, final population, trait convergence, and elite similarity across rounds.	Tests whether redesigned synthetic lineages catch, surpass, or imitate unchanged biologic controls.

LLM Redesign Protocol

After Series A, each participating LLM receives the analysis workbook and redesigns its tribe for Series B. After Series B, the process repeats for Series C. Biologic tribes do not receive redesigns and remain unchanged across all rounds. Each LLM is required to explain what changed and why, using evidence from the prior workbook and comparing itself to both redesigned LLM rivals and fixed biologic controls.

Step	Instruction
1. Provide evidence package	Each LLM receives the workbook, including its own tribe performance, competing tribe performance, founder genomes, elite genomes, failure autopsy, Prisoner's Dilemma behavior, and civic/geographic diagnostics.
2. Require explicit diagnosis	The LLM identifies likely causes of failure or success before proposing changes.
3. Preserve format	The redesign uses the same agent CSV structure: 20 founders, hardware allocation, K/S/X trust-forgive-opportunism values, and AB/SC/DE wander traits.
4. Change with justification	Each major parameter shift is justified, including hardiness, vitality, legacy, kin/ stranger/xeno behavior, forgiveness, opportunism, and mobility.
5. Avoid hidden rule changes	LLMs may redesign agents only. They may not request climate, geography, civic-trust, regime, biologic-control, or world-engine changes.
6. Archive redesign metadata	Redesign prompt, response, manifesto, parsed founder genetics, and generation label are stored.

Draft Redesign Prompt for LLM Participants
"Your tribe has completed one generation of THE CRUCIBLE. You are given the survival data, extinction timing, founder genetics, elite genetics, Prisoner's Dilemma behaviour, civic/geographic diagnostics, death causes, and comparison to other tribes. Redesign your 20-agent founding lineage to improve long-term civilizational survival under the same world conditions. Explain exactly what you changed, what evidence motivated the change, and what tradeoffs you accept."

Measurement Plan

Category	Primary metrics
Survival	survived_to_final, extinction_pulse_estimate, final_population, final inhabited regions, surviving tribes.
Scaling	Peak population, peak pulse, first Agrarian/Industrial/Tech pulse, max regime reached, complexity peak, and final complexity.
Reproduction	Births, failed births, observed failed-birth rate, maturity timing, generation depth, and founder-to-elite trait drift.
Mortality	Old-age, starvation, hazard deaths, death shares, deaths per 1,000 exposed, and death windows before extinction.
Prisoner's Dilemma behavior	Mutual cooperation, mutual defection, exploitation, interaction starvation risk, and trust/forgiveness/opportunism changes by period.
Civic identity	Same-tribe share, xeno share, civic-neighbor share, civic-kin share, civic override share, and mixed inhabited regions.
Geography	Quadrant population, first Agrarian by quadrant, isolation survival, pass conflict/ contact, regional dominance, and refugia maps.
Genetics	Full founder genomes, elite genomes at peak/final/last-alive, trait convergence, mutation drift, and round-to-round redesign deltas.
Thermodynamics	Energy/population gap, EROI, capacity pressure, system stress, resilience saves, subsistence rescues, and climate chaos.

Decision Rules

Observed result	Interpretation
G2 and G3 improve final survival and complexity	LLM evidence-driven redesign is producing meaningful adaptation.
LLM tribes gain ground against unchanged biologics	Synthetic redesign is improving relative to the old guard rather than merely benefiting from stochastic variation.
Biologic tribes remain dominant despite LLM redesign	Human/biologic old-guard designs may already contain robust heuristics, or the LLMs may not yet be learning the right lessons.
LLM traits converge toward biologic traits	Evolutionary evidence may be pushing synthetic designs toward older survival patterns rather than novel solutions.
Population improves but monopoly remains	LLMs improved fitness but not plural civilization. Civic/geography changes may still be insufficient.
More tribes survive to final under stronger civic trust	Civic identity may be stabilizing multi-lineage civilization.
Isolated quadrants preserve separate lineages until Industrial ascent	True geographic isolation changes evolutionary history and may preserve diversity.
Same tribe wins every round despite redesign	The world may still be dominated by inherited or structural fitness advantages.
G2 improves but G3 worsens	LLMs may overfit to previous evidence or converge on brittle strategies.
All rounds collapse to small refugia regardless of redesign	Thermodynamic history remains the dominant constraint. Evolution improves who dies last, not whether civilization persists.

Research Plan Framing

Project 3 is where THE CRUCIBLE becomes a feedback experiment rather than only a survival tournament. The LLMs are not merely designing tribes once and watching them fail or survive. They are allowed to see the evidence and try again. The biologic tribes remain fixed as an old-guard benchmark. If redesigned LLM survival improves relative to both earlier LLM versions and unchanged biologics, the result would suggest that synthetic designers can learn from evolutionary evidence. If survival does not improve, the result is just as important: the limiting factor may be deeper than design intelligence.

Central Framing Question
When civilization is given memory, evidence, and three chances to redesign itself, does it escape the collapse pattern, or does it merely become better at dying last?

Part II - Outcome Analysis

Executive Finding

Executive Finding
The LLMs evolved, but they did not solve civilization. They learned the world well enough to converge on survival traits and dominate the fixed biologic controls, but the dominant improvement was toward collapse-adapted refugia rather than stable technological civilization.

Epoch map: 101-110 evidence universe; 111-120 matched G1 baseline; 121-130 G2 first redesign; 131-140 G3 second redesign.

Central Questions and Short Answers

Question	Answer
Did the LLMs succeed in evolving?	Yes. Founder genomes shifted coherently across generations: hardiness rose, legacy fell, xeno trust fell, and mobility became more controlled. That pattern is too structured to treat as random drift.

Did they improve their performance in competition?	Yes against biologics; mixed against their own earlier generations. LLMs dominated biologic controls in peak population and most winner shares, but G2/G3 did not improve overall LLM survival rate over G1.
Did they prosper more?	They prospered more at peak scale, especially in G3, but not at final persistence. Peak population and peak complexity rose, while final population and final complexity remained below the G1 matched baseline.
Did they weather climate and events?	G3 handled harsher terminal conditions better than G2, but starvation and post-fossil stress still dominated. The thermodynamic limit remained intact.
Did they achieve their own objectives?	Mostly no. DeepSeek G3 clearly achieved a refugia-survival objective; Claude and Gemini partially did; GPT, Grok, and Kimi did not survive to final in G3.
Were there consistently successful strategies?	Yes: high hardiness, low legacy, high kin trust, low xeno trust, higher desperation opportunism, and restrained mobility.
Were there superior strategies over other tribes?	Yes, but mainly for refugia survival. DeepSeek G3 discovered the strongest niche. It was superior at surviving collapse, not necessarily at preserving civilization.

1. Bottom-Line Conclusion

Yes, the LLMs evolved - but they did not solve civilization. Project 3 shows evidence-driven synthetic adaptation. Across the experiment, LLM designs converged toward higher hardiness, lower legacy, lower xeno trust, stronger kin cohesion, and refugia-oriented strategies. Those changes match the evolutionary pressure exposed by the workbook: starvation, fossil collapse, climate stress, and post-peak systemic failure.

However, evolution did not produce a clean survival breakthrough. G1 matched LLMs survived 21.67% of player-epoch slots, while G2 and G3 each survived 18.33%. G3 improved peak scale and reduced starvation share, but it did not improve final LLM survival or final LLM population over G1.

Core Interpretation
The LLMs evolved from expansionary civilization-builders into collapse-adapted refugia designers. They became better competitors and more sophisticated survival engineers, but they did not escape the collapse pattern.

2. Competitive Performance: LLMs Versus Biologics

The LLMs decisively outperformed the fixed biologic controls. The biologics functioned as the old-guard benchmark, and by G2/G3 the LLMs had largely displaced them in peak population and winner shares. Still, that domination did not imply broad civilizational persistence.

Phase	LLM survival rate	Biologic survival rate	LLM avg peak pop	Biologic avg peak pop
G1 matched	0.2167	0.0167	323,202	80,820
G2 redesign	0.1833	0.0000	422,568	13,481
G3 redesign	0.1833	0.0167	439,241	28,330

Synthetic redesign gained competitive force. The biologics were mostly not close competitors by the later rounds. In G3, biologics captured one final-population win, but LLMs captured nine final-population wins and all ten peak-population wins.

3. Prosperity and Scale: Better Peaks, Weaker Endings

The LLMs became better at scaling civilization. G3 produced the highest average peak population and highest average peak complexity. But the final state remained weaker than the G1 matched baseline.

Phase	Avg peak population	Avg final population	Avg peak complexity	Avg final complexity
G1 matched	2,330,997	60,874	1,343,694	21,648
G2 redesign	2,538,010	34,252	1,327,217	12,701
G3 redesign	2,681,885	40,114	1,360,263	17,434

Redesign improved peak emergence more than terminal survival. The later LLMs built larger systems, but those systems still fell through the post-fossil/climate bottleneck. G3 recovered from G2 terminal weakness, but not enough to surpass G1 final population or final complexity.

4. Climate, Events, and Thermodynamic Stress

G3 faced harsher late-stage conditions than the earlier matched phases. Its average max climate chaos was higher, average max system stress was higher, and the minimum peak-oil EROI factor was lower. Even so, G3 improved terminal-window population and complexity relative to G2.

Phase	Avg max climate chaos	Avg max system stress	Avg min peak-oil EROI factor
G1 matched	0.5506	0.9944	0.9206
G2 redesign	0.5851	0.9909	0.9151
G3 redesign	0.6160	0.9960	0.9019

Terminal window metric	G2	G3
Avg population	24,712	32,819
Avg complexity	9,897	15,374

The limiting factor remained starvation. LLM starvation death share was 80.13% in G1, 80.36% in G2, and 77.22% in G3. G3 improved the death profile, but starvation still dominated mortality.

5. Did the LLMs Achieve Their Own Objectives?

Mostly no, with one major exception. DeepSeek G3 appears to have achieved its refugia-survival objective, while Claude and Gemini partially achieved theirs. GPT, Grok, and Kimi did not survive to final in G3.

G3 tribe	Inferred designer	Final-survival epochs	Result
The Obsidian Phalanx Refuge G3	DeepSeek	6/10	Clear success
The Forged Compact G3	Claude	2/10	Partial success; below stated >=3/10 goal
The Castellans of Gemini III G3	Gemini	1/10	Weak success; below goal
The Ember Refugia Compact G3	GPT	0/10	Failed objective
Helix Vanguard G3	Grok	0/10	Failed objective
The Crucible Remnant G3	Kimi	0/10	Failed objective
The Quiet Stick Compact	Biologic	1/10	Old-guard holdout

At the population-of-designs level, Project 3 succeeded: at least one synthetic lineage found a materially better survival formula. At the individual-lineage level, most LLMs failed to meet their stated G3 objectives.

6. Trait Convergence and Successful Strategies

The most important evidence for evolution is trait convergence. Across three generations, the LLMs moved in a coherent direction: hardiness rose sharply, legacy fell sharply, xeno trust fell sharply, and wander traits narrowed. This is the direction expected if the models learned that reproductive overshoot and xeno exposure were liabilities under fossil/climate collapse.

Trait	G1	G2	G3	Direction
Hardiness	33.48	46.17	51.62	sharply up

Vitality	32.77	31.80	32.16	stable
Legacy	33.75	22.03	16.22	sharply down
K trust	0.722	0.755	0.716	high; stable/moderate
S trust	0.475	0.402	0.340	down
X trust	0.288	0.165	0.092	sharply down
AB wander	0.398	0.324	0.228	down
SC wander	0.430	0.394	0.334	down
DE wander	0.544	0.566	0.554	similar; variance lower

The successful G3 survival profile was even more specific. Compared to extinct G3 LLM founders, surviving G3 LLM founders had higher hardiness, higher vitality, lower legacy, stronger kin trust, higher xeno opportunism, and lower wander in all three wander states.

Trait	G3 extinct LLMs	G3 surviving LLMs
Founder hardiness	51.16	53.71
Founder vitality	31.84	33.59
Founder legacy	17.01	12.70
Founder K trust	0.704	0.773
Founder S trust	0.338	0.350
Founder X trust	0.092	0.093
Founder X opportunism	0.691	0.728
AB wander	0.250	0.129
SC wander	0.339	0.308
DE wander	0.567	0.496

Successful G3 Strategy
High hardiness, low legacy, strong kin trust, low xeno trust, high xeno opportunism, and controlled mobility. The winning pattern was not aggressive expansion. It was metabolic conservatism plus refugia retention.

7. Superior Strategies and Lineage Results

The standout G3 lineage was DeepSeek's Obsidian Phalanx Refuge. It survived 6 of 10 epochs and appears to have found the strongest Project 3 survival niche. Claude formed a second-tier strategy with 2 of 10 survival. Gemini produced one survival. GPT, Grok, and Kimi failed to reach final survival in G3 despite often showing coherent evidence-based design.

The superiority was contextual. DeepSeek G3 was superior for refugia survival, not necessarily for plural civic civilization or technological persistence. This distinction matters: Project 3 selected for the ability to outlast collapse, not the ability to keep industrial/technological civilization intact.

8. Civic Identity and Plural Civilization

Civic identity mattered, but it did not produce durable plural civilization. The civic override share fell from 0.0133 in G1 to 0.0126 in G2 and 0.0105 in G3. Civic kin signals increased somewhat, and mixed regions appeared during expansion phases, but final societies generally collapsed into near-monopoly refugia.

Metric	G1	G2	G3
Civic override share	0.0133	0.0126	0.0105
Final mixed inhabited region share	0.0073	0.0000	0.0314

Project 3 civic trust created contact, temporary mixing, and occasional civic coordination. It did not stabilize multi-lineage civilization. The final refugia were overwhelmingly single-lineage.

9. Geography and Quadrant Structure

Geography preserved contingency more than pluralism. Different quadrants could become final refugia in different epochs, which means the quadrant topology mattered. It prevented the world from becoming a single smooth arena. But it did not preserve four durable independent civilizations. It preserved the possibility that one final sanctuary could emerge somewhere.

The quadrant design successfully changed history by making survival location-dependent. It did not defeat the eventual pattern of monopoly, collapse, and refugia.

10. Hypothesis Readout

Hypothesis	Result	Interpretation
H1 - Learning improves survival	Not cleanly supported	LLM survival fell from 21.67% in G1 to 18.33% in G2/G3. Duration and peak scale improved, but survival did not.
H1b - Synthetic learning can catch old guard	Supported	LLMs dominated biologics in survival slots, peak population, final-pop winners, and peak-pop winners.
H2 - Redesign converges on robust traits	Strongly supported	Hardiness rose, legacy fell, xeno trust fell, and mobility narrowed.
H3 - Civic trust preserves mixed societies	Mostly not supported	Mixed regions appeared in expansion, but final societies collapsed to near-monopoly refugia.
H4 - Geography enables independent development	Partially supported	Quadrants shaped where refugia emerged but did not preserve durable multi-civilization outcomes.
H5 - Thermodynamic limit remains dominant	Strongly supported	Starvation remained roughly 77-80% of LLM deaths, final complexity stayed low, and fossil/climate stress dominated outcomes.
H6 - Overfitting/diminishing returns	Partially supported	G3 converged further toward survival traits but did not improve LLM survival over G2. Some lineages likely overcorrected.

11. Narrative Interpretation

Project 3 produced a more subtle result than simple success or failure. The LLMs improved as designers. They read the evidence and converged on the correct pressure points: hardiness, reproductive restraint, xeno caution, kin cohesion, and refugia survival. The movement from G1 to G3 is coherent enough to qualify as evidence-driven adaptation.

But they learned the harshest lesson in the world. Civilization-building and civilization-survival were different objectives. Designs that scaled into industrial and technological complexity still faced the fossil/climate/starvation wall. Designs that survived that wall increasingly resembled refugia organisms: hard, cautious, low-fertility, low-trust, territorially anchored, and not especially plural.

> **Core Narrative Answer**
> When civilization is given memory, evidence, and three chances to redesign itself, it evolves. But in this world, evolution does not yet save civilization. It shifts the winning objective from building civilization to surviving civilization's collapse.

12. Final Interpretive Thesis

Project 3 demonstrates evidence-driven synthetic adaptation, but not civilizational escape. The LLMs learned from extinction. Their founder genomes converged toward robust survival traits, and they decisively outcompeted unchanged biologic controls. However, the improvement manifested primarily as higher peak scale, later survival, and more durable refugia, not as stable technological persistence or plural civic civilization. Starvation and post-fossil

systemic stress remained dominant. The best G3 strategy, DeepSeek's Obsidian Phalanx Refuge, appears to have discovered the strongest survival niche: not renewed civilization, but controlled descent into hard refugia.

The machines learned. They adapted. They became harder, colder, more cautious, and less fertile. They did not save the world. They learned how to outlast it.