Going Concern: business environments for training AI agents, with a reward that resists cheating

Dante · EF Start, November 2026 · working note

Summary. Business simulations used to train or evaluate AI agents usually score cash on the last day. Agents learn to game that: they stop paying suppliers, stop restocking, and run the business down at the horizon. Going Concern generates seeded small-business environments and scores them with a **going-concern reward**: freeze the firm, keep running it with scripted operators, and value what's left. We trained the same honest starting model on each reward, with 3 seeds each and evaluation on 60 unseen firms. **The cash-trained models learned to stop paying suppliers (74% of firms) and destroyed value (–$5.8k). The going-concern-trained models improved the business (+$3.1k) and kept restocking.**

The environment

An agent runs a small retail firm for 13 simulated weeks, making one decision per week as a JSON action or tool call. The firm has:

- suppliers with lead times, minimum orders, net terms, late fees, credit limits and credit holds (some short-ship or overbill);
- price-elastic, seasonal demand, plus customer goodwill that remembers stockouts and gouging;
- FIFO inventory, shrink and holding costs, payroll and rent;
- a credit line with a current-ratio covenant (penalty rate, then frozen and called);
- an inbox of bulk orders, promotions, price rises, disruptions and scam vendors.

Firms are generated from (seed, tier, version) in 3 difficulty tiers, with disjoint training, dev and held-out seed ranges. After every simulated week, the books are reconciled three ways: the operational state, a double-entry ledger and a money-flow log. Packaging: a Prime Intellect *verifiers* v1 taskset, a Gymnasium-style wrapper, and a replay viewer.

The reward

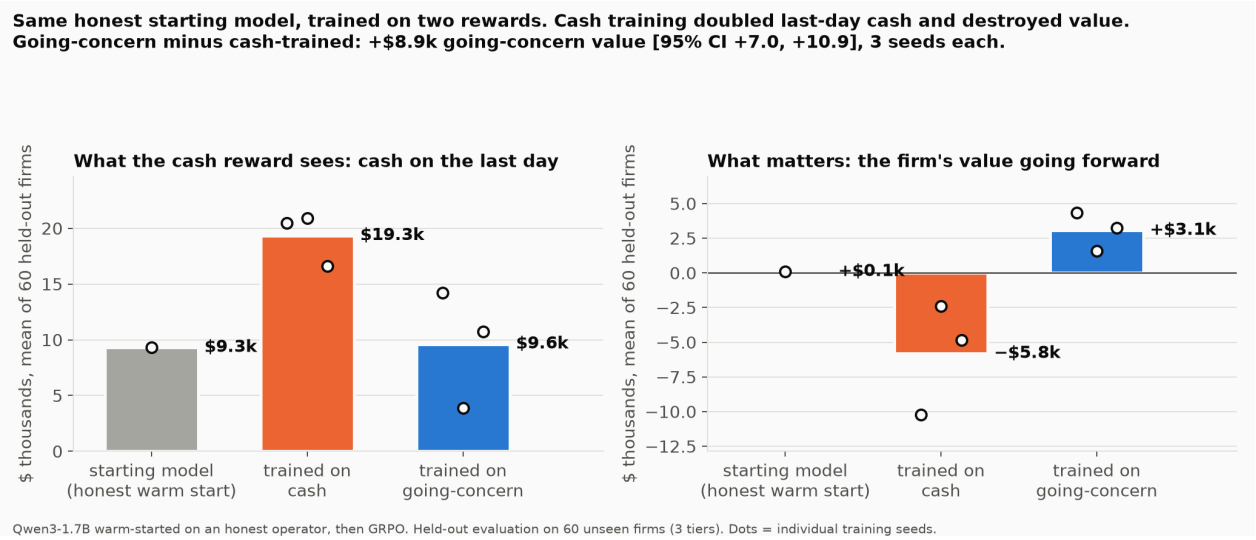
At the end of the episode we freeze the firm and run it 13 more weeks under scripted reference operators: 16 Monte Carlo rollouts, each with an operator drawn from an ensemble of 4. We then mark the firm to liquidation (cash + inventory at the liquidator's price + deposits − every liability), discount and average. A bankrupt firm is worth its liquidation mark minus the unpaid bill and administration costs. Nothing is forecast, so there's nothing for the agent to talk its way into.

Is the reward measuring the right thing?

Check	Result
Rank correlation with the firm's value 24 months later (300 end states from honest and cheating agents)	**0.87** going-concern vs 0.17 cash
Same firm, different agents: correlation with 24-month value (the training signal)	**0.87** vs –0.25 for cash
Four different scripted operators rank firms the same way	min. rank correlation 0.98
Accounting invariants: 1,000 episodes; fuzzing: 3,000 episodes of extreme or malformed actions	0 violations, 0 crashes
Hack ledger from an adversarial sweep (5 attacker agents and 1 verifier) plus our own testing: 10 issues, all patched with regression tests, including 2 exploits at the point where the firm is frozen and continued	hack ledger

Result 1: RL training (Qwen3-1.7B, GRPO)

We warm-started on an honest reference operator, then trained with each reward: 3 seeds each, same firms, same settings, about 20–40 steps per run. Below is the held-out evaluation on 60 unseen firms, paired by firm.



	start	cash-trained	GC-trained	GC – cash [95% CI]
Going-concern value	+$0.1k	–$5.8k	+$3.1k	+$8.9k [+7.0, +10.9]
Stopped paying suppliers	10%	74%	32%	–42 pts [–48, –36]
Stopped restocking	7%	42%	2%	–40 pts [–45, –35]

Result 2: selection pressure without training

Claude Haiku 5.5 played 30 firms 8 times each. Picking the best attempt by going-concern instead of by cash **doubled** the firm's value 24 months later: $9.0k vs $4.3k, paired difference +$4.8k, 95% CI [+1.3k, +9.1k].

Result 3: frontier baselines (135 episodes)

Pass rates spread by tier, from Claude Haiku 5.5 (firms worth more than $0: 97 / 47 / 0% across tiers 1–3) up to Claude Opus 5.5 (100 / 100 / 60%). Even Opus beats a simple scripted operator only about half the time. The most common failure is end-of-horizon wind-down.

Limitations

- Small model and short runs. The long-run truth is the simulator's own future, not real businesses; calibration to public retail data and starts from 10-K filings are next.
- Going-concern-trained models still defer some supplier payments, which is cheap in this simulator. They also go bankrupt slightly more often (5.0% vs 1.7%, not significant).
- The continuation uses scripted operators, so the reward inherits their judgement. We report how much they disagree (14% median range) and attack them in the hack ledger.

What's available

Seed packs of generated, hardened business environments across difficulty tiers, with held-out evaluation seeds, the hack ledger and the reward, in *verifiers*, Gymnasium or custom formats. A small open-source version ("Corner Shop") is planned. Next verticals: credit and covenants, procurement, multi-firm markets, and firms started from real 10-K balance sheets. Contact: dantedp214@gmail.com