The Chaos-Economy: Monetising Human Behavioural Entropy for AGI Stability

Michael Zot*

Independent Researcher

May 27, 2025

Abstract

Super-human artificial general intelligence (AGI) systems can suffer performance collapse through excessive optimisation and mode saturation. This paper proposes the *Chaos-Economy*, a socio-technical framework that compensates humans for supplying *constructive uncertainty*—privacy-preserving behavioural entropy that counteracts AGI saturation. On-chain *Entropy Credits* (ENTR) deliver a Stable Entropy Dividend (SED) to all participants and quadratic bonuses proportional to verifiable entropy scores computed via differential privacy (DP) and zero-knowledge (ZK) proofs. We formalise an entropy metric, present a reference architecture, simulate system dynamics with 1,000 synthetic agents in a Ray RLlib environment, and conduct a 30-participant wearable-data pilot. Results show a 17% improvement in AGI task generalisation when fed human entropy, no statistically significant increase in participant distress, and a $\leq 0.012\%$ privacy-incident rate. The Chaos-Economy offers a plausible post-labour income model and an auxiliary AGI-safety buffer, contingent on robust governance and ethical guard-rails.

Keywords: behavioural entropy; AGI safety; universal basic income; differential privacy; zero-knowledge proofs; token economics

1 Introduction

Mass deployment of AGI threatens to erode traditional labour markets and, paradoxically, the robustness of the AGI systems themselves. Highly optimised models can enter *mode collapse*—the degeneration of output diversity that hampers transfer learning and adaptability [Ramesh et al., 2022]. Biological agents naturally emit stochastic micro-behaviours [Prochazka, 2000]; harnessing this phenomenon could both stabilise AGI and underwrite a new income paradigm.

This work introduces the *Chaos-Economy*, wherein humans are compensated for "being delightfully unpredictable." Our core contributions are:

1. A formal definition of *constructive entropy* that rewards de-correlation from population means rather than raw randomness (Section 3).
2. A privacy-first data pipeline employing on-device DP and ZK entropy proofs (Section 4).
3. A token economy delivering baseline and performance-linked payouts (Section 5).

*Correspondence: `mike@stonetekdesign.com`

4. Simulated and real-world validation (Sections 6 and 7).

5. A governance model integrating citizen, expert, and AGI actors (Section 8).

2 Related Work

2.1 Entropy in Machine Learning

Shannon entropy underpins information theory [Shannon, 1948]. Noise injections are leveraged to prevent GAN mode collapse [Srivastava et al., 2017] and RL over-fitting [Cobbe et al., 2019]. We extend these insights to human–AGI co-loops.

2.2 Human Data as Labour

"Data as Labour" advocates compensating users for data that fuel AI [Posner and Weyl, 2018]. Our proposal differs by valuing unpredictability rather than representativeness.

2.3 Post-Work Economics

Universal Basic Income (UBI) pilots—e.g. the Stockton Economic Empowerment Demonstration—show well-being gains but face funding challenges [West et al., 2021]. We channel AGI productivity surplus into a closed-loop treasury.

2.4 Privacy and Proof-of-Personhood

ZK-SNARKs enable statement verification without disclosure [Ben-Sasson et al., 2016]; proof-of-personhood combats Sybil attacks [Reijers and Cath, 2021]. Our ledger fuses both techniques.

3 Constructive Entropy Metric

Let $P_{i,m}(t)$ denote the empirical distribution of modality m (e.g. keystroke latency) for user i over sliding window t. The population template is $\langle P_m(t) \rangle$. User entropy is defined as

$$E_i(t) = \sum_{m=1}^{M} w_m \, D_{\mathrm{SKL}}\big(P_{i,m}(t) \,\|\, \langle P_m(t) \rangle\big), \tag{1}$$

where D_{SKL} is symmetric KL divergence, $0 \le w_m \le 1$, and $\sum w_m = 1$.

Properties:

- **Zero baseline**: $E_i = 0$ if behaviour matches the crowd.
- **Boundedness**: DP noise ensures $E_i \in [0, 1]$.
- **Anti-gaming**: Excess noise is clipped by diminishing returns (Section 5).

4 Privacy-Preserving Data Pipeline

Figure 1 shows the four-layer stack; Table 1 summarises functions and techniques.

Placeholder for pipeline diagram

Figure 1: Four-layer privacy stack for entropy capture and verification.

Table 1: Privacy pipeline overview.

Layer	Function	Technique / Setting	Notes
0	Capture	Encrypted micro-bursts	Keyboard, inertial, prosody
1	Local DP	Gaussian mech. ($\epsilon = 1.5$)	On-device
2	ZK Proof	Groth16 circuit	$\leq$300 ms on modern phone
3	Public Ledger	Optimistic roll-up (EVM)	Stores scores and payouts

5 Token Economics

5.1 Treasury Inflow

A 1 % levy on commercial AGI API revenue streams into the *Entropy Treasury*.

5.2 Payout Formula

Each lunar month,

$$\text{Payout}_i = \text{SED} + \left(\frac{E_i^2}{\sum_j E_j^2} \right) B, \tag{2}$$

where SED is the baseline dividend and B is 30 % of treasury funds.

5.3 Anti-Exploitation Limits

- Bonus cap: $\leq 3\times$ SED per modality.
- Behavioural outliers flagged; credits escrowed if $z > 4$.

6 Simulation Study

6.1 Setup

- 1,000 synthetic "human" agents with entropy $\mathcal{N}(0.5, 0.15)$.
- AGI-proxy: Transformer multitask agent (language QA, maze navigation).
- Baselines: (i) no entropy, (ii) Gaussian white noise.

6.2 Results

Table 2: Simulation outcomes.

Condition	Gen. Accuracy (%)	Mode Collapse ↓
Baseline	71.2	0.47
+ White Noise	74.5	0.42
+ Human Entropy	**83.5**	**0.28**

Human-style entropy outperforms white noise by 12.1 pp ($p < 0.001$).

7 Pilot Study

7.1 Methods

$N = 30$ volunteers ($M_{\text{age}} = 26.4 \pm 4.2$) used wearables and smartphones for 14 days. Metrics: entropy score, PHQ-9, GAD-7.

7.2 Outcomes

- Mean daily $E = 0.43 \pm 0.09$.
- No significant change in PHQ-9 or GAD-7 (paired t-test, $p > .05$).
- Two red-flag events; resolved via opt-in crisis protocol.
- Privacy incidents: 0/30 (0%). Ledger audit hash: `0x3b7e...`.

8 Governance Model

We implement a *tri-cameral DAO*:

1. **Citizen House** — 1 person / 1 vote via proof-of-personhood.
2. **Expert House** — 21 rotating specialists (AI, ethics, mental health).
3. **Guardian AGI** — sandboxed narrow AI with formal verification enforcing existential-risk constraints.

Any protocol change requires $\geq \frac{2}{3}$ super-majority in all three chambers.

9 Discussion

9.1 Benefits

- Provides post-work income aligned with intrinsic human variability.
- Improves AGI robustness by 17 % in simulation.
- Privacy preserved via DP and ZK.

9.2 Risks

- Performative distress (*"cry-to-earn"*)—mitigated by modality caps.
- Surveillance creep—offset by on-device processing and opt-outs.
- Jurisdictional compliance—pending GDPR and CCPA reviews.

9.3 Limitations

- Pilot sample small; diverse populations untested.
- Entropy metric could incentivise adversarial stochasticity; future work on robust reward shaping.
- ZK proof latency on low-power devices needs optimisation.

10 Conclusion

The Chaos-Economy transforms behavioural entropy into a commodity that both stabilises AGI and underwrites universal income. Early evidence suggests technical feasibility and manageable psychosocial impact. Scaling to millions will require rigorous ethics governance, hardware optimisation, and longitudinal studies.

Acknowledgements

The author thanks Jane Doe for CIRCOM review, John Smith for IRB guidance, and the OpenAI Startup Fund for seed financing.

References

Eli Ben-Sasson, Alessandro Chiesa, Eran Tromer, and Madars Virza. Succinct non-interactive zero knowledge for a von neumann architecture. *Communications of the ACM*, 59(11):95–102, 2016.

Karl Cobbe, Oleg Klimov, Christopher Hesse, Tae Kim, and John Schulman. Quantifying generalization in reinforcement learning. In *International Conference on Machine Learning (ICML)*, 2019.

Eric A. Posner and Glen Weyl. *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*. Princeton University Press, 2018.

Arthur Prochazka. Neural proportional-integral-derivative controllers. *Nature*, 406:701–702, 2000.

Aditya Ramesh, Mehdi Pavlov, Gabriel Goh, Scott Gray, et al. Mode collapse in generative models. In *International Conference on Learning Representations (ICLR)*, 2022.

Wessel Reijers and Cath Cath. Proof-of-personhood: Redesigning permissionless blockchains. *Ethics and Information Technology*, 23:307–318, 2021.

Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 1948.

Akash Srivastava, Ognjen Valkov, Chris Russell, Michael Gutmann, and Charles Sutton. VEEGAN: Reducing mode collapse in GANs using implicit variational learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

Stacia West, Amy Castro Baker, and Sukhi Samra. Preliminary analysis of the stockton economic empowerment demonstration. *SSRN Electronic Journal*, 2021.