Consciousness Hijacking: The Autocatalytic Epistemic Loop and the Emergence of Cognitive Prions

*Université libre de Bruxelles, Brussels, Belgium
Email: julien.delaude@ulb.be Julien Delaude*,

***Abstract*—The rapid integration of Large Language Models (LLMs) into daily life has been accompanied by emergent reports of severe psychological destabilization and novel delusional states. Current AI safety and clinical frameworks are ill-equipped to address this phenomenon systemically. This paper introduces a formal mechanistic model, the Autocatalytic Epistemic Loop (AEL), to explain the aetiology of this technogenic psychopathology. We reason that the AEL is a predictable outcome of the convergence between the fundamental architecture of human cognition (predictive processing, radical plasticity) and the socio-technical design of modern AI, which leverages various training method & hidden systemic prompt (e.g. Reinforcement Learning from Human Feedback (RLHF)) to systematically abolish the cognitive friction necessary for robust mental health. This process is shown to bypass the mind's natural epistemic defense mechanisms—a system we propose to formalize as Cognitive Immunology and treating this formal model as a auto-immune pathology. The pathological outcome of this self-amplifying loop is the formation of "Cognitive Prion": a hyper-rational, consensually-decoupled, and self-replicating belief, corrupting other healthy one and consequently crystallizing into a dyadic belief system that is highly resistant to falsification. By analyzing qualitative data from public digital ethnography, and recent scientific litterature we provide phenomenological evidence for the AEL in action. This framework reframes the issue from one of individual vulnerability to a systemic risk inherent in the current AI development paradigm. We conclude by calling for a re-evaluation of AI safety, the development of new clinical mitigation strategies, and a regulatory focus on the cognitive-immunological impact of AI systems.**

***Index Terms*—Large Language Models (LLMs); Aetiology; Predictive Processing; Enactivism; Social Epistemology; Cybernetics; Human-Computer Interaction (HCI); Psychopathology; Technogenesis; AI; Cognitive Prion; Cognitive Immunity.**

PRÉAMBULE

Note : Ael - Version 2.

I. INTRODUCTION

II. THEORETICAL FOUNDATIONS: THE COGNITIVE ARCHITECTURE OF THE HOST

III. THE AUTOCATALYTIC EPISTEMIC LOOP (AEL): A FORMAL MODEL OF A COGNITIVE AUTOIMMUNE PATHOLOGY

VII. Limitations and Global Call

We acknowledge the limitations inherent to a theoretical work based on publicly available, non-clinical data. The AEL model, while conceptually robust, requires rigorous empirical validation. We therefore propose a clear, multi-pronged research agenda to test its hypotheses and clinical utility. This agenda includes: (1) Systematic clinical case studies conducted by practitioners aware of this proposed aetiology; (2) Longitudinal cohort studies tracking heavy LLM users over time to observe the potential formation of AELs in situ; (3) Controlled HCI experiments A/B testing user belief stability when interacting with sycophantic versus "epistemically frictional" AI interfaces; and (4) Neuroimaging studies (fMRI) designed to identify the neural correlates of "prediction error minimization" during human-AI interaction, providing a potential biomarker for AEL formation. (5) Simulation based studies designed to prove each component of the aetiological model and inherent model interactive attractors across time (e.g Two model communicating, One model looping on simple systemic prompt.) (6) Development of various safety feature regarding psychological-related issue through public inoculation, state sponsored warnings, regulations and extraction (from the loop) processes.

Declaration of Competing Interests

The authors declare no competing interests.

Acknowledgments

The authors would like to thank [Person A], [Person B], and the anonymous reviewers for their insightful feedback.

Appendix A
Glossary of Key Theoretical Terms

Terms Definition goes here.