Agentic Markov Blankets for Trustworthy Governance in Hybrid Reality

Massimiliano, Pirani PhD
Pegaso University, Italy
Department of Information Science and Technologies
Palazzo Zapata Piazza Trieste e Trento, 48 - 80132 – Napoli, Italy
massimiliano.pirani@unipegaso.it

Alessandro, Carbonari PhD
Università Politecnica delle Marche, Italy
Department of Construction, Civil Engineering and Architecture (DICEA)
Via Brecce Bianche, 12, 60131 Ancona AN, Italy
alessandro.carbonari@univpm.it

Alessandro, Cucchiarelli
Università Politecnica delle Marche, Italy
Department of Information Engineering (DII)
Via Brecce Bianche, 12, 60131 Ancona AN, Italy
a.cucchiarelli@univpm.it

Andrea, Generosi PhD
Pegaso University, Naples, Italy
Department of Information Science and Technologies
Palazzo Zapata Piazza Trieste e Trento, 48 - 80132 – Napoli, Italy
andrea.generosi@unipegaso.it

Tariq, Naeem
Università Politecnica delle Marche, Italy
Department of Information Engineering (DII)
Via Brecce Bianche, 12, 60131 Ancona AN, Italy
t.naeem@pm.univpm.it

Luca, Spalazzi PhD
Università Politecnica delle Marche, Italy
Department of Information Engineering (DII)
Via Brecce Bianche, 12, 60131 Ancona AN, Italy
l.spalazzi@univpm.it

Claudio, Tomazzoli PhD
Pegaso University, Naples, Italy
Department of Information Science and Technologies
Palazzo Zapata Piazza Trieste e Trento, 48 - 80132 – Napoli, Italy
claudio.tomazzoli@unipegaso.it

Abstract

Far beyond current artificial intelligence (AI) tools such as Large Language Models (LLMs) and Generative AI, the emerging agentic paradigm of AI is progressively assuming a leading role in governance, management, innovation, and sustainability-oriented decision-making processes. Agentic AI systems are expected to reshape socio-technical environments in a disruptive way by introducing increasingly autonomous entities capable of initiative, adaptation, and proactive interaction. As Karl Friston observes, truly agentic AI is characterized by curiosity and by the capability of prompting humans rather than merely responding to them. This transition implies that societies and governance systems will soon need to coexist with forms of artificial agency that cannot be treated as passive tools anymore. This paper argues that current socio-technical and governance structures are not yet prepared to cope with this transformation. In response, we propose a conceptual and technological framework aimed at establishing a sustainable equilibrium between human society and artificial agencies. The proposed approach combines two complementary technological directions with a causality-oriented perspective grounded in a revisited Bayesian Networks framework. In particular, we introduce the concept of Agentic Markov Blankets (AMBs) as agentic structures capable of acting as trustworthy and human-centered interposition layers between humans and increasingly complex AI ecosystems. Beyond its conceptual contribution, the framework can already build upon existing technological foundations, including constructivist trust models based on blockchain technologies and recent advances in Artificial General Intelligence architectures. The overall objective is to support the development of trustworthy, transparent, and sustainability-oriented agentic AI systems within future Hybrid Reality environments.

Keywords: Agentic AI, Hybrid Reality, Markov Blankets, Trustworthy AI, Cybernetics, Socio-Technical Systems