Chapter 3
Human–AI Collaboration:
Balancing Trust, Emotion, and Productivity

Shikha Khullar
https://orcid.org/0009-0000-5598-3145
Poornima University, India

Randhir Singh Baghel
https://orcid.org/0000-0002-5113-3554
Poornima University, India

Priya Swami
Poornima University, India

ABSTRACT

This chapter discusses the role of human-AI partnership within the enterprise metaverse and how AI hallucinations endanger trust, emotional congruency and productivity. The more immersive virtual environments transform organizational work, the more AI agents help in analyses, decision support, and training, but, because of their probabilistic nature, they can give confident and false results. The chapter introduces the TrustEmotion Productivity (TEP) Framework to demonstrate that safe collaboration relies on calibrated trust, emotional intelligence, and design that is qualified by productivity orientation. It also emphasizes the intensification of hallucinations within immersive, real-time, and multimodal environments and analyzes technical and organizational measures like retrieval-augmented generation, confidence calibration, transparency, expert review and governance protocols. It illustrates the operational risk of hallucinations and best practices of constructing credible, ethical, and human-focused AI systems to support metaverse-enabled businesses.

DOI: 10.4018/979-8-3373-7534-2.ch003

1: INTRODUCTION

1.1 Background and Context

The modern business environment is experiencing a radical shift because of the integration of artificial intelligence technologies and virtual reality (Lee et al., 2021). The next wave of digital workplace evolution is the enterprise metaverse a stable and interrelated chain of virtual spaces in which organizations are engaged in business activities, collaborate, and innovate. In this regard, AI systems are more than tools but collaborators, agents of decision support, and autonomous, capable of creating content, opportunities, and suggestions (Brynjolfsson & McAfee, 2017). Nevertheless, the growing power and independence of AI systems presented a serious problem: the so-called AI hallucinations (Ji et al., 2023). They happen when AI models, especially large language models as well as generative systems generate output that seems plausible and coherent but is false, inconsistent with source information, or even fabricated. In conventional computing settings, these errors could be handled with the help of the traditional verification mechanisms. In the immersive, high-stakes, and real-world experience of enterprises operating in the metaverse, AI hallucinations present increased vulnerabilities to the quality of decision making, reputation, and even safety of operations (Slater and Sanchez-Vives, 2016).

1.2 The Significance of AI Hallucination Management

The nature of neural network architectures and training techniques leads to AI hallucinations (Brown et al., 2020). The modern AI systems, and in particular, the transformer-based language models, do not produce outputs based on the deterministic retrieval or logical inference, but instead as the probabilistic prediction. Such approach of generating allows impressive flexibility and creativity at the same time predisposes to the generation of confident-sounding misinformation (Guo et al., 2017). These issues are increased by the enterprise metaverse context due to a number of reasons (McVeigh-Schultz et al., 2022). To begin with, the immersion of virtual space may lead to an increase in psychological presence and a decrease in critical appraisal, which may make virtual users more vulnerable to the acceptance of AI-generated misinformation (Slater et al., 2010). Second, due to the real-time collaborative essence of metaverse interactions, there is often a strong need to make a decision quickly with little time to check it. Third, by incorporating AI in various functional areas, it provides many injections points of hallucination in the work process of the organization.