



InSIS

Indian Structural Integrity Society

Preventing Hallucinated IS/IRC Code Citations in AI-Generated Structural Health Monitoring Reports: A Retrieval-Augmented Validation Architecture

Srinivas Venkataramanappa

B.E., M.Tech (IIT Madras) · M.S. (WPI, USA)

CivilMind, Mysore – 570 009, Karnataka, India

sринi@civil-mind.com · civil-mind.com

Contact No (optional): +918197218126

Abstract

Large Language Models (LLMs) generate plausible but factually incorrect content at a documented rate of 50–82% in unconstrained settings. In structural health monitoring (SHM) for civil infrastructure, a hallucinated Indian Standard (IS) or Indian Roads Congress (IRC) clause reference in an anomaly report constitutes a potential legal liability that could invalidate maintenance decisions or insurance claims. This technical note describes an IS/IRC-grounded LLM narrative generation pipeline and post-generation citation validation architecture developed for CivilMind, an artificial intelligence (AI)-powered SHM platform targeting Indian civil infrastructure. The pipeline retrieves authoritative IS/IRC clauses from a vector store using semantic search and cross-encoder reranking, grounds LLM generation in retrieved context via Retrieval-Augmented Generation (RAG), and applies a citation validator that rejects any clause number absent from the vector store before report persistence. On a 50-scenario evaluation set, the validator correctly rejected 100% of hallucinated clause references ($n = 23$) while retaining 97.8% of valid citations. The architecture is, to the author's knowledge, the first application of RAG to the IS/IRC knowledge base for real-time structural health monitoring.

Keywords: Structural health monitoring · IS/IRC code retrieval · Retrieval-Augmented Generation · LLM hallucination · Citation validation · Indian Standards