

PhD Candidate full name:

Bruno Georgevich Ferreira

Higher Education Institution: University of Porto**Doctoral Programme:** Doctoral Program in Informatic Engineering (FEUP)**Dissertation Title:** Modular and Multi-Stage Semantic Perception System for Robotics**Date:** 27/02/2026 14:00**Location:** Porto (UP) | Engineering (FEUP) | Administration Building | Room 104 - Sala de Atos

Abstract or Public Summary: Modular and Multi-Stage Semantic Perception System for Robotics

The evolution of autonomous robotics benefits largely from the capacity to construct rich, navigable, and semantic representations of the environment, even more so if shared with humans. While the advent of open-vocabulary scene graphs powered by Vision-Language Models (VLMs) has revolutionized perception, these systems face critical hurdles: high rates of hallucinations (False Positives), a lack of topological spatial context, and operational fragility due to heavy reliance on cloud connectivity. This thesis proposes the Hybrid Inference Perception and Mapping System (HIPaMS), framework adaptable to a target system, likely a robotic system that interacts with humans. The HIPaMS is a modular framework designed to bridge the gap between low-level perception and high-level agentic reasoning. A Proof of Concept (PoC) was designed to implement the HIPaMS. This PoC enhances the state-of-the-art ConceptGraphs semantic mapping process and introduces a refined interaction system through four main contributions. First, it introduces the Hybrid Adaptable Resource-Aware Inference Mechanism (HARAIM), which dynamically orchestrates internal models and settings based on runtime resource availability and optimization policies. This mechanism allows any optimization policy to adapt robotic system's operation, possibly allowing zero downtime during network failures, graceful degradation and/or operational efficiency. Second, the semantic mapping pipeline is enhanced with rigorous False Positive filtering protocols, persona-based prompt engineering, and a broad collection of semantic information in an optimized manner during mapping. Third, a Room Semantic Segmentation Routine is proposed to provide topological information to the semantic map during interaction. This transforms unstructured, noisy detections into a hierarchically organized scene graph, anchoring objects within functional topological regions. Fourth, the robotic system now incorporates dynamic knowledge base via the Human-in-the-Loop (HITL) Agentic Retrieval-Augmented Generation (RAG)-based Interaction System (HARBIS). This interface uses short- and long-term memory to understand complex natural language queries. It enables the robot to learn continuously from user interactions, address gaps in perception and

knowledge, maintain temporal consistency, and acknowledge its limitations by proactively asking for clarification. Extensive validation was conducted across 30 diverse environments, involving a total of 3300 interactive requests (depend on semantic map quality). The tested PoC processed 110 user requests per environment, categorized into: direct (30), indirect (30), graceful failure (30), follow-up (10) and time consistency (10). An ablation study was also performed to identify the impact of specific framework and PoC components. The results show that the PoC reduces False Positive detections by $\approx 86\%$, elevating mapping precision from a baseline of ≈ 0.28 to ≈ 0.68 . Although strict filtering reduces raw recall, the integration of HITL learning increased the success rate for complex query resolution to ≈ 0.81 , compared to baseline values of ≈ 0.48 and ≈ 0.55 . Furthermore, the HIPaMS PoC reduced cloud inference costs by up to $\approx 84\%$ in mapping and over $\approx 95\%$ in interaction tasks while ensuring system stability. The presented framework paves the way for increased robotic autonomy and efficiency. The presented PoC demonstrates superior performance, particularly for human-centered scenarios.

Principal Supervisor at INESC TEC: Armando Jorge Miranda de Sousa

Scientific Domain: [Artificial Intelligence] + [Robotics]

Keywords: Semantic Mapping, Open-Vocabulary Perception, Hybrid Inference Architecture, Adaptable Framework, Human-in-the-Loop, Retrieval-Augmented Generation (RAG), Topological Segmentation, Robot@VirtualHome, Vision-Language Models, Agentic AI, Operational Robustness.