Horizon Europe—The Framework Programme for Research and Innovation

Project funded by the European Commission

Grant Agreement Number 101135707—FORTIS

Multi-Modal and Multi-Aspect Holistic Human-Robot Interaction


# FORTIS Open Call #1

## Annex 1.1 - Technical Specifications

**Version 2.0 -** 17 November 2025

**Disclaimer**

## Document History

| Version | Publication date | Changes |
|---|---|---|
| **1.0** | **01-04-2025** | First version |
| **2.0** | **17-11-2025** | Date of submission stage has been updated:<br>• 17 November 2025 |

# Table of contents

# List of Tables

# 1. Introduction

The aim of the FORTIS Open Call #1, entitled "Boosting Development of the FORTIS Solution" aims at attracting and engaging systems actors and innovators - universities/researchers, mid-caps, SMEs and start-ups, to develop and integrate their technologies within the FORTIS ecosystem. The selected beneficiaries will receive mentoring support and technical guidance to implement the projects.
For Open Call #1, FORTIS will allocate EUR 2M to be distributed to up to 8 successful applications, with up to EUR 250k each.

This document, referred to as **Annex 1.1 FORTIS Technical specification**, offers a comprehensive outline of:

1. The FORTIS toolkit's structure
2. FORTIS topics, specifications and requirements.

## 2. Introduction to FORTIS toolkit's architecture

FORTIS main objectives are:

1. Develop, integrate, and provide a human-centric solution for modelling and analyzing humans.
2. Develop and provide a flexible and agile multi-robotic-centric solution interacting with humans.
3. Integrate and provide the FORTIS solution where a safe and trustworthy Human-Robot interaction (HRI) is guaranteed and provides optimized operations for both humans and robots.
4. Demonstrate the solution in several scenarios, industrial and non-industrial.

The toolkits are defined as sets of components/modules designed to build the HRI at all levels: human, robot and human-robot centric. The FORTIS solution is divided into 3 major activities: human-centric, robot-centric, and human-robot trustworthy interactions. Each one of these activities will result in the delivery of specific toolkits. The toolkits form the Tangible Expected Outcomes (TEOs) of the FORTIS project and will be validated in the FORTIS pilots.

FORTIS OC#1 invites third parties to create modules and components that enhance its HRI solution. The topics open for the proposals are presented in **Chapter 3** and the link with the toolkits is presented in Table 1.

***Table 1.*** *Architecture of the FORTIS toolkits*

| TEO | TEO description | Toolkit | Topics |
|---|---|---|---|
| *Human-centric: TEO 1* | *A set of toolkits that models the human and builds the virtual compliment character which will guarantee best interaction for each human under the human's status.* | **FORTIS AI-based Human Cognition Toolkit** | 1.1 Human Activity Recognition Using Non-wearable<br>1.2 Human Activity Recognition Using Wearable Sensors |
| | | **FORTIS Human-Robot Interaction Context Toolkit** | 2.1 Memory Consolidation and Forgetting Mechanisms<br>2.2 Lifelong and Continual Learning for Personalized Adaptation |
| | | **FORTIS Multi-modal Adaptive Human-Robot Communication Toolkit** | 3.1 Spatialized and Adaptive Audio for Robot Communication<br>3.2 Visual Cues for Enhanced Intelligibility |
| *Robot-centric: TEO 2* | *A set of toolkits and guidelines for building a multi-robotics solution that can interact with humans. This bundle will allow end users to upgrade machinery to an AI-embodied system that is able to interact with the human.* | **FORTIS Robotic Intelligibility Toolkit** | 3.1 Spatialized and Adaptive Audio for Robot Communication<br>3.2 Visual Cues for Enhanced Intelligibility |
| | | **FORTIS Robot Gateway** | 4 Multimodality semantic segmentation/perception |

## 3. Topics for FORTIS OC#1

Applicants must submit their proposals for one of the FORTIS topics.

The objective of the open calls is to allow third parties to contribute to the development of the FORTIS solution, which consists of the development of technology and software, among others. In this way, the participation of third parties will enrich the results of the project and will be a source of synergies, improving the quality of the project solution and deepening the knowledge on the technologies developed. Below a

description of the Fortis topics for OC#1 is provided, with a description of the topic, the technical requirements and key aspects that the proposal should address and an example of application, with suggested real-word scenario(s) where the solution of the topic is applied.

**FORTIS topics for OC#1:**

## Topic 1: Activity Recognition - supported by FORTIS partners: FBK + XLAB + Garcia

In the Open Call, proposals should target the estimation the human activities in a wide variety of working environments, based on the exploitation of 1) sensors mounted on robots, 2) sensors mounted on humans. It is possible to exploit both types of sensor deployments, but the applicant must specify one of the sub-topics as their primary focus.

### Sub-topic 1.1: Human Activity Recognition Using Non-Wearable Sensors

*a) Description:*

High-level Human Activity Recognition (HAR) must capture not only human actions but also the overall environmental context, with the goal of generating a comprehensive textual description that reflects the dynamics of the entire scene. This awareness plays a crucial role in enabling safe and efficient human-robot collaboration (HRC), particularly in dynamic and unstructured environments, such as construction sites, manufacturing facilities, and infrastructure maintenance. Within FORTIS, one of the main objectives falls on robot-centric HAR, meaning that human activities are to be recognised using the perception capabilities embedded in robotic platforms, rather than relying on wearables or external tracking systems. To achieve this, multimodal sensing is leveraged that integrates various perception technologies, for example:

- ***RGB Cameras*** (useful for detecting human presence, identifying body posture, and recognising gestures)
- ***Stereo/depth cameras*** (to provide additional spatial information to enhance tracking and segmentation of human subjects),
- ***3D LiDARs*** (to enable the robust tracking of human movements, even in challenging conditions with occlusions or variable lighting), and,
- ***IMUs*** (indirect via robot movement and inertial data to complement human motion estimation).

The HAR approaches considered in FORTIS may include, depending on the adopted approach:

1. ***Human Tracking*** for the continuous tracking of human workers in the robot's operational environment,
2. ***Body Pose Estimation*** to extract key joint positions and infer actions through movement analysis,
3. ***Sequential Data Classification*** that uses time-series data from joint positions to infer activity states,
4. ***Scene Context Awareness*** that combines human activity recognition with environmental context to anticipate collaborative tasks.

Collaborative perception, where multiple robots share sensor data to improve human activity recognition, is also a key consideration. This is especially relevant in occlusion-prone environments where a single viewpoint may not always provide sufficient information to track and classify human activities.

To validate the HAR solutions developed in the OC projects, real-world datasets will be provided, collected under actual working conditions. These datasets will contain synchronised multimodal data from robots operating in the FORTIS pilots. The open call is looking for innovative solutions that consider the use of zero-shot models, to eliminate the need for the ground truth labelling of activities, and to improve generalisability. Varying levels of action granularity will need to be delivered, and it is essential that direct access to the model is provided, rather than relying on inaccessible vision-language model (VLM) for example.

*b) Technical requirements:*

To ensure compatibility and efficiency within the FORTIS framework, the following technical requirements must be met:

- ***Real-Time Processing:*** The system must be capable of classifying human activities with a minimum update rate of 1Hz to ensure timely response to changes in human behaviour.
- ***Computational Efficiency:*** The implementation should be optimised for real-time deployment on robotic platforms with limited computational resources. GPU usage should be avoided, when possible, prioritising CPU-based inference. Energy efficiency will be considered, with preference for lightweight models.
- ***Multi-Robot Capabilities:*** Solutions that allow for distributed sensing and data fusion across multiple robots to improve human activity recognition are encouraged.

While not mandatory, **ROS2 compatibility** and the use of standard ROS message types are desirable and will be considered positively during the evaluation process.

*c) Real-world scenario(s):*

The proposed solution must be **generic and adaptable**, ensuring its applicability beyond a single use case. To validate the solution, each project must **define and provide its own scenario or pilot** where the technology will be tested.

Potential domains for validation include:

- **Industrial environments** (e.g., manufacturing, logistics, maintenance operations).
- **Critical operations** (e.g., defense, security, emergency response).
- **Human-centric services** (e.g., healthcare, assistive technologies, human-robot collaboration).

Other domains may also be considered, provided they align with the overall objectives of the Open Call.

### Sub-topic 1.2: Human Activity Recognition Using Wearable Sensors

*a) Description:*

Human Activity Recognition using wearable and other human-embedded sensor technologies provides a powerful alternative, or complement, to robot-centric perception by mitigating issues such as occlusions, limited fields of view, and poor lighting conditions. This approach relies on wearable and user-embedded sensing devices - including smartwatches, AR headsets, smartphones, chest bands, and other body-worn sensors - to track human motion, physiological state, and environmental interactions.

This approach enables a more comprehensive assessment of human activities, extending beyond motion tracking to include:

- ***Kinematic Analysis:*** Utilising IMUs, accelerometers, and gyroscopes embedded in wearables to capture fine-grained motion data for detailed movement classification.
- ***Physiological Monitoring:*** Measuring heart rate, stress levels, skin temperature, electrodermal activity, and oxygen saturation to assess worker fatigue, cognitive load, or physical strain.
- ***Motion Analysis:*** Measuring track gaze, gestures, and head movements (for example, using gloves or wrist bands).
- ***Multimodal Sensor Fusion:*** Combining motion data, physiological responses, and environmental context to enable a richer, more context-aware human activity understanding.

Wearable-based HAR is particularly relevant in challenging environments, such as construction sites, manufacturing plants, and underground infrastructure, where frequent occlusions can hinder robot-based human tracking and limited visibility due to dust, lighting variations, or confined spaces can affect standard vision-based solutions. Worker stress and fatigue need to be monitored in real-time to prevent accidents and improve productivity. Wearable HAR solutions should ensure that continuous activity tracking is possible, regardless of their proximity to robots. This personalized data stream will enable enhanced safety, adaptive robot behaviour, and improved worker well-being.

*b) Technical requirements:* To ensure the seamless integration and real-time performance within the FORTIS ecosystem, the following requirements apply:

- ***Edge Processing on Embedded Devices:*** The solution must run locally on embedded devices, preferably Android- or iOS-based platforms (e.g., smartphones, AR headsets, smartwatches). Cloud-based or WiFi-dependent processing should be avoided due to high data throughput and real-time constraints.
- ***Efficient Sensor Utilisation:*** The system should support real-time data processing using IMUs, heart rate monitors, electrodermal activity sensors, and other physiological tracking sensors. Camera-based activity recognition (e.g., from AR headsets) should be optimised for on-device inference to reduce computational load.
- ***Low-Latency Communication and API Integration:*** Processed HAR data should be made accessible over a local network through an efficient backend API (e.g., FlaskAPI or MQTT-based messaging). The system should provide a lightweight and structured output (e.g. JSON-formatted activity labels "activity": "lifting", "stress_level": "moderate"), with real-time updates at ≥1 Hz, and minimal network bandwidth usage to avoid congestion in industrial settings.
- ***Battery and Power Optimisation:*** Solutions must be optimised for low power consumption, ensuring extended battery life for wearable devices used in long working shifts.

*c) Real-world scenario(s):*

The proposed solution must be **generic and adaptable**, ensuring its applicability beyond a single use case. To validate the solution, each project must **define and provide its own scenario or pilot** where the technology will be tested.

Potential domains for validation include:

- **Industrial environments** (e.g., manufacturing, logistics, maintenance operations).
- **Critical operations** (e.g., defense, security, emergency response).
- **Human-centric services** (e.g., healthcare, assistive technologies, human-robot collaboration).

Other domains may also be considered, provided they align with the overall objectives of the Open Call.

## Topic 2: Long-term memory – TAU + XLAB + BEKO

Long-term Memory (LTM) in Human-Robot Interaction (HRI) refers to a robot's ability to store, retrieve, and update information over extended periods of time to improve interactions with humans. It is critical for personalized, adaptive, and context-aware interactions, enabling robots to remember past experiences, recognize individuals, and anticipate future needs. It combines different disciplines like Social Sciences and Humanities (SSH), Artificial Intelligence (AI), Knowledge Representation (KR), and Software Engineering (SE). In FORTIS the Long-term Memory is essential for satisfying the needs to reach holistic and multi-aspect Human-Robot Interaction.

### Sub-topic 2.1: Memory Consolidation and Forgetting Mechanisms

*a) Description*

Memory consolidation in robots refers to the process of stabilizing and integrating learned knowledge over time, ensuring that relevant information is retained while outdated or less useful data is forgotten. This prevents memory overload and optimizes decision-making efficiency. Inspired by human cognition, techniques such as sleep-like memory replay, synaptic plasticity, and adaptive forgetting are crucial for maintaining useful long-term memory in AI-driven systems. This ability is crucial in mobile implementations where shared data should be minimized due to limited computing, network and storage resources. Therefore, this sub-topic is needed in FORTIS project and will enable the consortium to further develop the expected outcomes.

*b) Technical Requirements*

- **Memory Management Algorithms**: Implement memory compression, episodic memory prioritization, and adaptive forgetting mechanisms for optimizing the use of the limited available resources.
- **Real-Time Processing**: The system must be capable of memory management, storage and retrieval in real-time (less than 300 ms) to ensure a timely response to the human.
- **Sharable memory:** The system must implement mechanism to share the memories with other agents for community-like interaction with the human.
- **Efficient Storage and Retrieval**: Implement vector databases (e.g., FAISS) or knowledge graphs for scalable long-term memory for fast information flow.
- **Cognitive Load Optimization**: Use reinforcement learning-based heuristics to determine which information should be retained or discarded.
- **Human Identification:** The system must accurately identify the intended human based on markers, reflectors or sensors (without facial recognition) to ensure the correct memory is updated.

*c) Real-World Scenario*

1. **Human-Machine Interaction**: healthcare and schools represent real world situation where long-term memory can improve the human interaction with the computer systems for daily activities.
2. **AI-Powered Digital Twins**: In smart manufacturing, a digital twin of a production line retains critical historical sensor data but prunes irrelevant short-term fluctuations to optimize system modelling.

### Sub-topic 2.2: Lifelong and Continual Learning for Personalized Adaptation

*a) Description*

Lifelong learning enables robots and AI systems to continuously acquire, refine, and update their knowledge while interacting with new users, tasks, and environments. Unlike traditional AI models that require retraining from scratch, **continual learning** ensures adaptive intelligence without catastrophic forgetting, allowing robots to function efficiently in dynamic real-world conditions. This ability is important in Human-Robot Interaction, like in FORTIS project, where new human may join the interaction or new behavior of the human is observed. Thus, the Robot, or the embodied-AI, must continually learn human trains.

*b) Technical Requirements*

1. **Catastrophic Forgetting Mitigation** – Implement techniques to prevent memory training by over-learning. This includes looking for incremental learning approach to gradually update the AI models without harming the existing knowledge.
2. **Task-Specific Adaptation** – Introduce user-personalized models via federated learning or hybrid AI architectures (symbolic + deep learning).
3. **Efficient Knowledge Transfer** – Implement modular AI architectures that reuse prior knowledge (e.g., transformers fine-tuned for different domains).
4. **Human Identification:** The system must accurately identify the intended human based on markers, reflectors or sensors (without facial recognition) to ensure the correct memory is updated.

*c) Real-World Scenario*

- **Human-Robot Collaboration in Factories:** An AI-powered robot learns different workers' assembly techniques and adapts to individual preferences over time while retaining core manufacturing knowledge.
- **Personalized AI Assistants:** A virtual assistant improves responses over time by learning user habits (e.g., adjusting reminders based on past interactions) while avoiding memory saturation with outdated behaviors.

## Topic 3: Enhancing Robotic Intelligibility Through Directional Audio and Visual Effects – TAU + TEC+ ROBOTNIK + VIAS

**Directional audio and visual effects** are one of the key advances that can contribute to the holistic Human-Robot Interaction domain. Not only does it deliver the message to the human but also provide directional guidance, which enriches the transferred information with no additional message. Hence, enhancing how robots convey information, ensuring clarity, context-awareness, and seamless communication in dynamic environments. FORTIS project foresees this topic to provide an advance in **AI-driven multimodal communication**, enabling robots to adapt their messaging based on user attention, spatial positioning, and contextual needs.

### Sub-topic 3.1: Spatialized and Adaptive Audio for Robot Communication

*a) Description*: Effective robotic communication relies not only on what is said but also how and where it is heard. Spatialized and adaptive audio enables robots to direct sound toward specific users, significantly enhancing speech intelligibility in noisy or multi-user environments. By leveraging beamforming, 3D audio rendering, and directional speakers, robots can ensure that their messages reach intended recipients without overwhelming others. Additionally, advanced AI-driven speech modulation allows robots to dynamically adjust tone, volume, and emphasis based on environmental context and user proximity. This can be particularly useful in industrial settings, public spaces, and assistive robotics, where clear and precise communication is essential. Furthermore, research in psychoacoustics and user perception can help optimize auditory cues to minimize cognitive load and maximize comprehension. By integrating machine learning and real-time audio adaptation, robots can personalize communication strategies, making interactions more natural and effective. Understanding the interaction between spatialized audio and human attention is key to developing intelligent robotic systems that communicate seamlessly in dynamic environments.

*b) Technical requirements:*

- **Directional Audio Technology**: The system must utilize directional audio devices such as beamforming microphones and directional speakers to precisely target sound, improving speech intelligibility in dynamic and fluctuating environments.
- **3D Audio Rendering & Spatialization**: The system should support 3D audio rendering to provide spatialized sound, allowing users to perceive audio as originating from specific directions relative to their position, enhancing the naturalness of communication.
- **Multimodal Sensor Fusion**: Integration of sensors (e.g., cameras, microphones, LIDAR) is essential for real-time tracking of user position, attention, and movements, enabling dynamic adjustments to both audio and visual cues.
- **Real-Time Context-Aware Communication**: The system must feature AI-driven algorithms to assess contextual factors (Scene status, e.g., environment noise or user attention) and adjust audio and visual outputs dynamically to ensure optimal intelligibility in varying settings.

*c) Real-world scenario(s):*

- **Factories**: in a noisy factory environment, a robot needs to communicate with workers without adding to the auditory clutter. Using **directional audio**, the robot can direct instructions or alerts specifically to a worker within its vicinity, while filtering out unnecessary noise.
- **Assistive Robotics for the Elderly**: A robot designed to assist elderly individuals in a home setting uses spatialized audio to provide instructions or reminders in a way that feels natural and intuitive. For example, when the robot needs to remind a person to take their medication, it can direct the message toward the person from a specific location, ensuring clarity.

### Sub-topic 3.2: Visual Cues for Enhanced Intelligibility

*a) Description:* Visual cues are a powerful tool for improving robotic communication, especially when verbal communication may be less effective or disrupted. These cues can include **light signals**, **colour changes and transitions**, **flashing indicators**, and **directional lasers**, all designed to capture attention and convey important information clearly. For example, a robot might use **flashing lights** to signal an alert or draw attention to a specific area, **color transitions** to indicate different states, or **directional lasers** to point to specific objects or guide user actions. By providing clear and visually distinct cues, robots can enhance their ability to communicate their intent or actions, making interactions more intuitive and reducing the potential for misunderstandings in dynamic or noisy environments.

*b) Technical requirements:*

- **Visual Output Devices**: The system must include high-precision light sources, such as directional lasers, LEDs, and RGB lights, capable of generating clear and attention-grabbing visual cues. These devices should be capable of projecting specific patterns (e.g., beams, flashes, or color transitions) to guide users and indicate the robot's state or intent.
- **Control and Signal Processing**: The system must include a real-time control unit that can adjust visual cues based on dynamic inputs. This includes fine-tuning the intensity, color, and movement of light signals or lasers, as well as seamlessly switching between different visual patterns according to the robot's status or user interaction.
- **Environmental Adaptation**: The system must feature environmental sensors (e.g., cameras, proximity sensors, ambient light detectors) to dynamically adjust the visual cues. This ensures that visual cues remain effective under different conditions, such as bright or cluttered spaces, and adapt in real-time to changes in user movement and contextual factors.
- **User Interaction Feedback Mechanism**: The system must support interactive feedback where the robot's visual cues are responsive to user behaviour. This involves detecting when a user's attention has been directed to a cue and adjusting visual output accordingly, such as changing colours or activating additional cues to confirm the user's interaction and maintain effective communication.

*c) Real-world scenario(s):*

- **Robotic Assistance in Noisy Industrial Factories:** In a noisy industrial factory, a robot uses **directional lasers** and **flashing LEDs** to communicate with workers who have their hands full. The robot can project a laser to highlight the location of a tool or machine part, while flashing lights signal critical status updates or alerts. These visual cues allow workers to receive important information without needing to stop their tasks or rely on auditory signals, improving efficiency and safety in a high-noise environment.
- **Autonomous Delivery Robots in Outdoor Environments**: In outdoor environments, autonomous delivery robots use directional lasers, flashing LEDs, and color transitions to communicate with users and pedestrians. The robot projects a laser to indicate the pickup location and uses flashing lights to signal its proximity. If obstacles are detected, the robot changes the light color to signal delays or the need for user assistance.
- **Healthcare Assistive Robots in Hospitals**: In hospitals, assistive robots use color-changing LEDs and directional lasers to guide patients and staff during medical tasks. Flashing lights indicate readiness to deliver medication, while lasers point to specific equipment or areas. These visual cues help ensure clear communication in noisy environments, improving overall safety and efficiency.

## Topic 4: Multimodality semantic segmentation/perception - ING + FBK

*a) Description:* Multimodal semantic segmentation is crucial for effective robot perception, particularly in complex dynamic environments, like those addressed within the FORTIS project. This topic aims to explore and develop innovative methods that seamlessly combine and exploit various robot-embedded sensing modalities, including RGB cameras, stereo cameras, depth cameras, LiDARs, IMUs, microphones, and potentially others. The goal is to create a coherent semantic understanding of the surroundings, identifying and classifying objects, obstacles, humans, and other relevant features within the operational environment. Multimodal data fusion represents a significant research challenge, particularly when dealing with different sensing modalities that provide complementary, but heterogeneous information. RGB cameras offer rich visual data; LiDARs provide accurate geometrical information even in variable lighting conditions; IMUs offer essential motion data of the robot that can be fundamental to identify dynamic obstacles; and microphones add an auditory dimension, useful in noisy or visually occluded environments. Effective fusion strategies - early, intermediate, or late - are encouraged, aiming to leverage the strengths of each modality to enhance segmentation accuracy and robustness.

Proposals should particularly address strategies for multimodal fusion in challenging real-world conditions, such as construction sites and manufacturing settings, where variability in lighting, occlusions, noise, and highly dynamic scenarios are common. Advanced deep-learning techniques, including transformer-based architectures, convolutional neural networks (CNNs), and attention mechanisms, are encouraged for improving segmentation and perception accuracy. Moreover, collaborative perception techniques exploiting multiple robots and their diverse viewpoints and sensor modalities will be positively considered, given their potential to further enhance perception robustness and accuracy.

*b) Technical requirements:* To ensure seamless integration and real-time performance within the FORTIS robotic framework, the proposed solutions must meet the following technical specifications:

- **ROS Compatibility:** The solution must be compatible with ROS1 and should additionally support ROS2, if possible. Implementing ROS2 will be considered a positive factor during the evaluation process.
- **Real-Time Performance:** Semantic segmentation must be computed with a minimum frequency of 1Hz to ensure timely reactions within dynamic scenarios.
- **Computational Efficiency:** Solutions should minimise computational resources and prioritise CPU-based inference, considering limited onboard processing capabilities typical in mobile robotic platforms.
- **Multi-Robot Collaboration:** Methods that support collaborative multimodal perception by sharing and fusing data across multiple robotic units, especially under communication constrained scenarios, are highly recommended.

*c) Real-world scenario(s):* The multimodal semantic segmentation methods will be validated across the diverse FORTIS pilots, each representing distinct operational challenges:

- **Construction Pilot (Pilot #1):** Solutions must demonstrate robust segmentation capabilities in dynamic and cluttered environments typical of construction sites, accurately classifying elements, such as building materials, machinery, tools, and humans.
- **Infrastructure Services Pilot (Pilot #2):** Targeting underground and confined maintenance operations, proposals should focus on reliably recognising infrastructure elements, obstacles, and workers, even in visually challenging environments.

Real-world datasets collected from these scenarios (provided as ROSbags by the FORTIS consortium) will serve as a basis for developing, validating, and demonstrating the efficacy and robustness of the proposed multimodal segmentation solutions.

# 4. Conclusions

This document serves as an informative resource detailing the FORTIS toolkits and the topics to which the applicants should apply. It complements the FORTIS Guidelines for Applicants by providing a comprehensive understanding of the FORTIS solution and topics specification offered by the consortium. The information provided herein should support third parties in elaborating their own proposals.