AI-BOOST

Delivering the next level of European
AI Open competitions

GENAI-BASED NATURAL LANGUAGE MISSION GENERATOR FOR AUTONOMOUS ROBOTS IN AGRICULTURE

CHALLENGE DESCRIPTION

1. ANNEX 2: CHALLENGE DESCRIPTION

1 CHALLENGE DESCRIPTION

1.1 Title

GenAI-Based Natural Language Mission Generator for Autonomous Robots

1.2 Organisation Description

Organisation name: Consorzio Intellimech - Joiintlab
Website: https://www.intellimech.it/en/ - https://www.joiintlab.com/
Sector / Industry: Manufacturing/Agriculture
Country: Italy

Consorzio Intellimech is a private no-profit research consortium founded in Bergamo in 2007 to fill the gap between the research and the industrial sector, promoting the collaboration of companies of different sizes and from various industrial domains. The consortium currently involves more than 70 high-tech enterprises, making it one of the most important Italian private initiatives in this field. Intellimech's research activity is primarily focused on Data Analytics, AI, ICT, Augmented/Virtual Reality, Robotics and human-robot interaction. Intellimech manages applied R&D and interdisciplinary experimental activities into pre-competitive technological platforms and prototypes production for innovative infra-sectorial applications to serve the partners. Moreover, Intellimech has experience implementing products and services relying on the outcomes from shared research activities and concretized through vertical projects tailored for specific partners.

For this specific challenge, Intellimech relies on JOiiNT LAB, established in 2020 in partnership with Istituto Italiano di Tecnologia (IIT). This laboratory is focused on advanced robotics technologies and aims at strengthening the technology transfer mission, bridging research activities and industrial needs, training high-level professional figures with advanced technical-scientific skills and enhancing the technological excellence of the area. This joint laboratory is unique in terms of size and strategic importance, making it the ideal base for IIT's projection towards supporting the Lombard and national industry.

1.3 Challenge Description

Programming robots still requires specialized knowledge of robotic frameworks (e.g., ROS), scripting, and mission configuration. This creates a strong dependency on skilled programmers and limits the accessibility and scalability of robotic systems, particularly for SMEs and non-technical operators.
This challenge aims to develop a Generative AI-based mission generator capable of translating natural language instructions into executable robotic commands. The system should convert spoken language into text, interpret the intent using Large Language Models, and automatically translate it into structured ROS-compatible commands that can be executed by the robot.
The objective is to move from robot programming (traditional coding) to natural and multimodal interaction. At the core of this innovation is the deployment of Vision-Language-Action (VLA) models designed to process several kinds of inputs—specifically speech commands, text instructions, visual data, and spatial maps. By leveraging these VLA models for high-level reasoning and translating human intent into ROS-compatible commands via robotic middleware, we are creating an accessible, AI-driven mission programming layer that bridges the gap between natural language and standardized robotic execution.

OBJECTIVES

1. Develop a pipeline converting spoken instructions into structured textual commands.
2. Implement LLM-based intent recognition and task decomposition.
3. Translate natural language tasks into executable ROS-compatible mission plans.
4. Integrate visual/map inputs for context-aware mission generation.
5. Test and validate the VLA pipeline output on a robotic platform.

Funded by the European Union

EXPECTED OUTCOMES AND TRL LEVEL

1. The expected result is a functional Proof of Concept capable of autonomously generating and executing robot missions from natural language instructions (voice to ROS commands) in a controlled environment. TRL 5

1.4 Expected Impacts and KPIs

The specific demonstrative use-case will enhance productivity in agriculture, but the approach should be modular and extendable to other domains such as robotic inspection, industrial maintenance, and autonomous field operations. By democratizing robotic control through AI, the challenge contributes to accelerating the digital and intelligent automation transition across European industry. Its expected to achieve the following set of key performance indicators:

KPI1. Intent Recognition Accuracy: measure the system's ability to correctly interpret user intent and extract relevant parameters from natural language instructions (voice or text).
Metric: percentage of correctly identified intents and associated parameters (e.g., task type, target objects, location, constraints) compared to ground truth annotations.
Evaluation Method: Comparison against the environment predefined set of tasks and actions.
Target: ≥ 90% accuracy
KPI2. Task Planning and Execution Success Rate: evaluate the system's capability to generate correct ROS-compatible task sequences and successfully execute them to achieve the intended goal.
Metric: percentage of missions in which the generated task sequence is correct and leads to successful execution in simulation or controlled environment.
Evaluation Method: compare generated task sequences against reference plans and validate successful execution in ROS-based environment.
Target: ≥ 85% successful mission completion
KPI3. Robustness to Natural Language Variability: assess the system's ability to handle different linguistic expressions of the same task while maintaining consistent performance.
Metric: task success rate across multiple variations of the same instruction.
Evaluation Method: for each mission, test at least 3–5 semantically equivalent language variations and measure consistency and success rate across variations.
Target: ≥ 85% success rate across all tested variations
Optional KPI4. End-to-End Processing Latency: measure the responsiveness of the system from user input to executable mission generation.
Metric: total time from speech input to ROS-compatible command generation.
Evaluation Method: average latency measured across multiple test scenarios.
Target: ≤ 5–10 seconds

1.5 Data Framework

DATASETS PROVIDED

Intellimech will provide a labeled dataset of grape clusters in vineyards manually annotated by domain experts. In addition, some datasets are derived from state-of-the-art publicly available sources, from which further metadata and annotations may be leveraged. The estimated error rate is low (not formally quantified but assumed minimal due to expert validation).

Training data for perception tasks (e.g., labelled images for grape detection and classification), robot capability descriptions (structured list of available robot actions and constraints) will be made available. These datasets will support training, validation, and experimentation phases.

For final evaluation real-world, real-time data coming from the outdoor vineyard environment will be used to demonstrate the system in realistic operating conditions. This ensures evaluation on unseen and dynamic data, reflecting real deployment conditions. This setup minimizes risks of overfitting and data leakage, ensuring a fair and robust evaluation.

DATA RIGHTS (LEGAL & ETHICAL)

Funded by the European Union

The datasets are primarily open-source and/or shared under appropriate licenses (Wgisd - Creative Commons Attribution-NonCommercial 4.0 International Public License; AI4Agriculture-grape-detection MIT License3; wGrapeUNIPD-DL - Creative Commons Attribution 4.0 International) that allow their use within the scope of the challenge. Participants will be granted access in compliance with the applicable licenses. Any third-party datasets will remain subject to their original licensing terms.

The datasets do not contain personal or sensitive data, therefore no anonymization or pseudonymization procedures are required.

Open datasets are used with proper permissions. Data can be transferred and used across secure environments (e.g., local servers, cloud platforms, HPC systems).

Class imbalance (e.g. stronger representation of specific grape types such as red grapes) or environmental bias (data may be collected under specific lighting or seasonal conditions) could be present in the datasets. These factors may affect generalization and should be considered during model development.

Besides the specificities mentioned above, no specific legal, ethical, or contractual restrictions are foreseen beyond compliance with dataset licensing terms.

1.6 Evaluation Metrics and Protocol

The proposed solutions will be compared against a traditional robot programming approach based on Graphical User Interfaces (GUI) and manual configuration, predefined and structured task pipelines and execution based on rule-based and conventional programming techniques. This baseline reflects the current state of practice, where mission definition requires technical expertise and lacks flexibility in handling natural language inputs.
A random split strategy will be applied for training and validation datasets. More concretely, partitioning of the global dataset into 70% for training, 15% for validation, and the remaining 15% as the test set.
The evaluation will prioritize performance and efficiency, specifically the task Execution Success Rate (percentage of missions successfully completed according to the user's intent) and the Mission Generation Time (the time required to convert a natural language command into a ROS-compatible executable plan. Secondary evaluation will focus on system reliability (degree to which the robot behaves as expected, avoiding incorrect or unintended actions), usability (ease of integration with existing robotic systems and middleware), and transparency (ability of the system to provide interpretable descriptions of the generated task plan).

1.7 Infrastructure

The solution will be ideally tested in a real-world outdoor environment, specifically within a vineyard row. A dedicated experimental setup will be available at Joiintlab premises (Italy), where a small vineyard row will be used to validate the system under realistic operating conditions. This setup enables testing of perception, navigation, and task execution in a controlled but representative environment. Participants will have access to a robotic platform equipped for field operations, an environmental maps and contextual data and a controlled testing environment for experimentation and validation.

The proposed solution is expected to integrate with existing robotic systems and infrastructure. For the robotic platform integration, compatibility with the provided robot hardware and software stack should be ensured. Integration with ROS1-based architectures for communication, task planning and execution is foreseen.

1.8 Responsible AI

Some operational factors that could affect the solution are the variations in environmental settings (e.g., lighting, weather conditions) and the differences in grape types and vineyard configurations. Potential gaps may arise from limited variability in these conditions, which could affect generalization.

Proposed solutions should ensure a sufficient level of explainability and traceability, particularly in the interpretation of natural language commands and the generation of task sequences and mission plans. The system should provide interpretable outputs, such as structured representations of planned actions.

Funded by the European Union

Participants must address potential risks related to system safety and reliability, including incorrect command interpretation, which may lead to unintended robot actions, unsafe robot behaviour, particularly in navigation and task execution and mechanical or operational risks affecting the mobile robotic platform. Solutions should include safeguards such as the validation of generated commands before execution and fail-safe mechanisms and emergency stop capabilities. Human oversight is required, especially during mission validation before execution and monitoring of robot behavior in real-world environments.

No specific sectoral regulatory constraints are imposed within the scope of this challenge. However, participants are expected to follow general AI safety and trustworthiness principles and ensure compliance with applicable standards for robotic systems and data usage.

1.9 Additional Support Offered by the Challenge Owner

Consorzio Intellimech will participate in co-creation and mentoring during the competition phases.

Funded by the European Union