Funded by the European Union

Marie Skłodowska-Curie Actions

Green Field Data

Exceptional
36-months Double Degree PhD Scholarship
Position-L
Phd in Computer Science from CNRS and
PhD in Agriculture, Environment and Bioenergy from UniMI

Conversational AI for Sustainable Agriculture: Natural Language Interfaces over Robotic and Analytical Farming Systems

EU Recruiting institutions	University Claude Bernard Lyon 1, Doctoral School Informatics and Mathematics, Lyon, France (18 Months), Supervisor: G. Vargas-Solar University of Milan, Dept. of Agricultural and Environmental Sciences - Production, Landscape, Agroenergy, Italy (18 Months), Supervisor: R. Oberti
Keywords	Agri-robots, IoT, data streams, conversational analytics, large language models

Exceptional benefits at a glance

- ***International PhD training excellence (here)***
- ***Renowned supervisors & top-tier labs***
- ***Interdisciplinary & multi sectoral research***
- ***Competitive MSCA salary & allowances***
- ***Global academic & industrial network***
- ***Non-academic secondments***

Salary	Living Allowance	Mobility Allowance*	Family Allowance**
Gross amount	EUR 4736	EUR 710	EUR 660

Long Term leave allowance (if applicable)

Special needs allowance (if applicable)

*private mobility-related costs (e.g. travel and accommodation costs), not their professional costs under the action

**doctoral candidate has or acquires family obligations during the action duration, i.e. persons linked to him/her by (i) marriage, or (ii) a relationship with equivalent status to a marriage recognised by the legislation of the country or region where this relationship was formalised; or (iii) dependent children who are actually being maintained by the researcher, the family allowance must be paid to him/her as well

GreenFieldData Project at glance

GreenFieldData: "IoRT Data management and analysis for Sustainable Agriculture" is a project funded under the action HORIZON Marie Sklodowska-Curie Action (MSCA) Joint Doctoral Network. **GreenFieldData** will train a new generation of researchers able to tackle digital and green transition challenges using a human-centric approach to ensure the robustness and relevance of the solutions responding to the specific needs of the EU market in a context of climate change and increasing socio-economic constraints. At a policy level, **GreenFieldData** outcomes will feed in directly to the aims of the HE

Strategic Plan 2025-2027, EU Partnership Agriculture of Data and Digital EU Program. **GreenFieldData** proposes a high-level interdisciplinary, inter-sectoral and international (triple ‘i’) research project and training network on new IoRT (Internet of Robotic Things) based solutions for sustainable agriculture. **GreenFieldData** will mobilize 14 Doctoral Candidates (DCs) enrolled in Double Degree Doctorate programmes with 12 academic main beneficiary partners, across 7 EU countries. Moreover, 21 non-academic associated partners, and 3 academic associated partners will provide support to the DCs. The partners form a high quality network, where Academic partners have previous research collaborations as outlined in a common vision paper. The ambitious project will provide the DCs with a unique toolbox of cutting-edge knowledge, tools and strategies which will boost their employability and benefit the next generation operational workforce (researchers, Digital Technologies (DTs) and agricultural stakeholders). The project results will also benefit EU innovation as the human-centric IoT devices & robotics, and data-based solutions tailored to EU context will enable the agricultural sector to assess and mitigate the impacts of climate change, and define new sustainable low input practices, thus increasing resilience and competitiveness.

PhD Position L – “Conversational AI for Sustainable Agriculture: Natural Language Interfaces over Robotic and Analytical Farming Systems”

Context: Agricultural robots are autonomous machines deployed to improve the efficiency of crop production and to reduce the amount of human labor, holding a strong potential as well for enhancing the quality of produce and positively influencing the sustainability of agricultural processes. The market of agricultural robots grew at a rate of 19.3% from 2021 until 2022. The market is predicted to grow from 5.9 billion USD in 2022 to 11.9 billion USD by 2026. Such a fast growth implies the necessity of a highly educated staff from both fields of agriculture and IT, to develop techniques for planning the work of Agri robots, scheduling their tasks, deploying IoT devices in fields as well as for collecting, integrating, storing, analysing, building prediction models, and visualizing immense volumes of various types of data, generated by agri-robots and IoT devices.

The most popular classes of autonomous agricultural devices include: weeding robots, seeding robots, harvesting robots, milking robots, autonomous tractors, robotic carriers, agricultural drones. Ground and airborne platforms can cooperate in the so-called multi-modal systems [Agra] to perform their tasks efficiently. The control of multiple robots through a centralized software platform can bring a large-scale farm automation. However, agricultural environments often feature unpredictable terrain, unknown obstacles, in season changing working scenarios, and a range of weather conditions that can impair autonomous navigation and operation and limit reliability. Additionally, agricultural regions are often in highly rural areas, where connectivity and access to repair and maintenance services can be limited. Coordinating Agri robots and running smart farming needs assistance from technologies such as IoT, big data, analytics, computer vision, cloud computing, and artificial intelligence (AI). IoT devices help in data collection. Sensors plugged in tractors and trucks as well as in fields, soil, and plants aid in the collection of real-time data directly from the ground. Based on these historical data predictive machine learning (ML) algorithms that can alert even before a problem occurs. Combining field-level data with other data available in the cloud, such as weather data and food pricing models allow to further optimize smart farming.

Among the most impactful operations for sustainable crop management are crop protection treatments aimed at preventing or mitigating the effects of biotic stressors, such as pests and diseases, that impair crop establishment and development. While these treatments ensure the production of abundant, healthy, and affordable food, they predominantly rely on pesticide applications. Growing societal concerns regarding the environmental and human health risks associated with pesticide use, together with the increasing incidence of resistance cases in pests and pathogens, have prompted global policy initiatives to reduce pesticide dependence.

Integrating robotic sensing and IoT with precision actuation technologies can significantly enhance decision-making in crop protection, enabling substantial reductions in pesticide use through optimized timing, dosage, and spatial targeting of treatments. At the same time, such selective and precise interventions can dramatically improve the efficacy of naturally derived compounds or physical control methods, such as irradiation with UV, which greatly benefit from selective and accurate application to the specific plant tissues or organs requiring protection.

Nonetheless, while proliferation and advances in IoT services and robots enable tremendous automation and scalability opportunities, new productivity, usability and maintainability challenges have also emerged. Typically experts create data analysis pipelines and robot control processes as well as providing interfaces (e.g., Web, mobile app, desktop) and dashboards to support interactions with robots. Thus, current generation of technologies in agri-robots and analytics primarily target data scientists. In fact, much of these available technologies presuppose technical expertise comparable to that of professional programmers, including skill fully employing different low-level APIs to access various data sources, sensors, together with procedural data flow constructs to create and maintain complex data curation scripts (e.g., using Python).

In smart robotic agriculture, **data** acts as the central nervous system of the entire ecosystem. From soil sensors and aerial drones to autonomous tractors or equipment, vast and heterogeneous streams of data are continuously generated. These include structured data (e.g., GPS coordinates, temperature, humidity), semi-structured inputs (e.g., XML/JSON sensors feed), and unstructured content (e.g., images, videos, audio, human input via voice or text). Managing and interpreting this deluge of multimodal data requires robust **data ingestion**, **stream processing**, and **real-time analytics** infrastructures. In this context, **stream processing frameworks** like Apache Spark Structured Streaming or Flink are essential for enabling near real-time decision-making—such as adapting crop treatments strategies during sudden weather shifts or rerouting harvesting robots in case of terrain changes—by processing incoming sensor and IoT data on-the-fly.

Natural Language Processing (**NLP**) and Large Language Models (**LLMs**) bring a complementary human-centric dimension to this technologically complex environment. NLP techniques allow stakeholders (e.g., farmers, agronomists, maintenance technicians) to interact with systems using natural language, simplifying command issuing, querying data, and receiving feedback. For example, a farmer could ask, "How much pesticide was sprayed on the northern field yesterday, and under what conditions?"—a query that an NLP interface backed by structured data and metadata could resolve and respond to. LLMs, particularly in fine-tuned or domain-adapted versions (e.g., AgriGPT-style models), further enhance this by enabling conversation-based interfaces that support task planning, explanation of analytical results, or even co-piloting robotic actions. These capabilities democratize access to complex data-driven insights and empower non-programmers, including future farmers, to leverage the full potential of precision agriculture.

Moreover, LLMs can assist in **semantic interpretation** of diverse agricultural data sources. For instance, they can analyze farmer notes, maintenance logs, scientific articles, or weather advisories to extract relevant knowledge and align it with real-time robotic operations. This is particularly relevant for sustainability tasks—like interpreting historical documentation on optimal pesticide use or recommending data-efficient practices based on literature analysis. When paired with **streaming data pipelines**, LLMs can process and summarize evolving datasets (e.g., daily yield metrics or satellite observations) to support timely interventions and policy decisions.

Overall, the synergy between real-time data streams, scalable analytics engines, and LLM-powered natural language interfaces transforms smart agriculture into an inclusive, responsive, and data-sovereign ecosystem. It enables not only operational efficiency but also greater transparency, accountability, and resilience across the agricultural lifecycle.

Objectives: Generative AI foundation models are increasingly shaping the landscape of digital systems, from decision-making processes to autonomous control and analytics. This PhD project aims to investigate and design novel abstractions, models, and algorithms that enable the superimposition of human-in-the-loop, context-aware natural language interactions over smart agricultural robotics and IoT infrastructures. Building on state-of-the-art advances in large language models (LLMs) such as BERT, BLOOM, GPT-4, Codex, and multimodal systems like DALL·E and Whisper, we will create a unified conversational layer that allows end-users—regardless of technical background—to command, monitor, and adapt agricultural robots and streaming analytics workflows using natural text or voice-based inputs.

The goal is to provide a seamless entry point where users can express high-level tasks (e.g., “harvest ripe crops in field B before rain”), perform situational queries (e.g., “show plots with soil moisture below threshold”), and receive explanatory feedback in conversational form. These requests will be dynamically mapped to orchestrated sequences of API calls, data retrieval, and robotic actions across a heterogeneous and distributed ecosystem of IoT devices, drone services, and farm robots.

Technically, the project will focus on the integration of:

- **Conversational orchestration over streaming dataflows** (e.g., Spark Structured Streaming).
- **Semantic grounding and intent recognition** for task disambiguation and adaptation to physical constraints (terrain, weather, crop health).
- **Multimodal LLMs** for interpreting textual, visual, and spatial data representations.
- **Feedback loops** for user supervision and trust calibration.
- **Latency-aware deployment strategies** in edge-cloud environments to ensure robust and timely operations.
- LLM-driven generation of workflow and configuration scripts for robots

The outcomes will contribute to sustainable, inclusive, and interpretable smart agriculture, allowing non-expert farmers, technicians, and researchers to benefit from cutting-edge AI through accessible and ethical interfaces.

Methodology

To achieve the objectives of this PhD project, we will develop a multi-layered methodology that combines advances in robotics, natural language processing (NLP), and knowledge-driven automation.

- Algorithms and a methodology - along with supporting tools – for generating robot control processes over situated agriculture environments, robot capabilities, and tasks, given natural language utterances. These workflows will consider the specificities of situated agricultural environments—such as terrain, weather, and crop types—as well as the capabilities and constraints of various agri-robots.
- Novel parameter configuration techniques that leverage LLM capabilities to automatically extract appropriate parameter tuning values from possible large number of evolving and unstructured knowledge articles and configuration manuals (e.g., text documents describing spraying manuals, or best practice in crop protection operations) to effectively guide robotic spraying. This will transform robotic spraying from ad-hoc, possibly harmful and wasteful process to a process that is knowledge-driven, safe, efficient and to a large extend, automated. By embedding this contextual knowledge into robotic processes, particularly for sensitive tasks like pesticide spraying, we aim to move from reactive or generic deployments to proactive, precise, and environmentally conscious operations.
- Novel techniques through which: (i) stakeholders are empowered to interactively and collaboratively systematically analyze sensors data, identify robot breakdown events and obstacles, (ii) relevant recovery and repair strategies can be constructed systematically and incrementally, (iii) natural language utterances are used to interactively specify corrections and feedbacks aligning model outcomes with expert knowledge and (iv) human-AI models continuously synthesize robot control policies and adapt themselves by reasoning and acting upon corrections and feedback by stakeholders. This will be achieved through a dialogue-based interface where natural language is used not only to diagnose and validate observations, but also to iteratively update control models. The system will support feedback loops where human expertise is incrementally encoded into decision policies, enabling human-AI co-evolution of robot behaviour.

- Lastly, we will develop methods for dynamically synthesizing and adapting control policies in real-time, based on human feedback, streaming observations, and evolving task goals. These adaptive models will reason about environmental variability, user corrections, and system performance to refine both analytic and operational components of the agricultural platform.

Together, these contributions will constitute a robust, conversational, and adaptive architecture for natural language–driven smart agriculture robotics, contributing both foundational knowledge and applied tools for sustainable AI in agri-tech ecosystems.

References

1. Chen, Z., Jiang, C., Xie, L., & Liu, Y. (2021). **Deep learning for autonomous vehicle perception and control: A review and case studies in agriculture**. *Computers and Electronics in Agriculture*, *187*, 106278. https://doi.org/10.1016/j.compag.2021.106278

2. Kamilaris, A., Kartakoullis, A., & Prenafeta-Boldú, F. X. (2017). **A review on the practice of big data analysis in agriculture**. *Computers and Electronics in Agriculture*, *143*, 23–37. https://doi.org/10.1016/j.compag.2017.09.037

3. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). **Attention is all you need**. *Advances in Neural Information*

4. Wagner, N., Kraus, M., Minker, W., Griol, D., & Callejas, Z. (2025). *A Survey on Multi-User Conversational Interfaces*. *Applied Sciences*, *15*(13), 7267. https://doi.org/10.3390/app15137267 mdpi.com+7mdpi.com+7researchgate.net+7researchgate.net+1researchgate.net+1

5. Rozanski, P., & Murphy, R. (2022). *Programming cobots by voice: A pragmatic, web-based approach*. *Industrial Robot: An International Journal*, 49(5), 660–672. https://doi.org/10.1080/0951192X.2022.2148754 tandfonline.com

6. Symeonaki, E., Arvanitis, K., Piromalis, D., & Papoutsidakis, M. (2020). Conversational user interface integration in controlling iot devices applied to smart agriculture: analysis of a chatbot system design. In *Intelligent Systems and Applications: Proceedings of the 2019 Intelligent Systems Conference (IntelliSys) Volume 1* (pp. 1071-1088). Springer International Publishing.

7. Brabra, H., Báez, M., Benatallah, B., Gaaloul, W., Bouguelia, S., & Zamanirad, S. (2021). Dialogue management in conversational systems: a review of approaches, challenges, and opportunities. *IEEE Transactions on Cognitive and Developmental Systems*, *14*(3), 783-798.

8. Gaci, Y., Benatallah, B., Casati, F., & Benabdeslem, K. (2024, December). Bias Exposed: The BiaXposer Framework for NLP Fairness. In *International Conference on Service-Oriented Computing* (pp. 312-326). Singapore: Springer Nature Singapore.

9. Vargas-Solar, G., Cerquitelli, T., Espinosa-Oviedo, J. A., Cheval, F., Buchaille, A., & Polgar, L. (2023). Conversational Data Exploration: A Game-Changer for Designing Data Science Pipelines. *arXiv preprint arXiv:2311.06695*.

10. Vargas-Solar, G. (2024). Processing the Narrative: Innovative Graph Models and Queries for Textual Content Knowledge Extraction†. *Electronics (2079-9292)*, *13*(18).

11. Oberti R., Schmilovitch Z. (2021). Robotic Spraying for Precision Crop Protection. In: *Innovation in Agricultural Robotics for Precision Agriculture*. p. 117-150, Springer International Publishing, doi: 10.1007/978-3-030-77036-5_6

12. Oberti R., Caffi T., Calcante A (2025). Developments in precision crop protection systems. In: P*recision agriculture for sustainability (2nd Ed)*. p. 501-540. Burleigh Dodds Science Publishing

PRACTICAL INFORMATION

Recruiting and host institutions	• University Claude Bernard Lyon 1, Doctoral School Informatics and Mathematics, Lyon, France (18 Months) (Recruiting institution) • University of Milan, Dept. of Agricultural and Environmental Sciences - Production, Landscape, Agroenergy, Italy (18 Months),
Doctoral schools	• InfoMaths @ University Claude Bernard Lyon 1, Aarhus, Denmark • AAB @ University of Milan, Italy
Supervisors	• Dr. Genoveva Vargas-Solar (CNRS, University Claude Bernard Lyon 1, Lyon, France) • Prof. Roberto Oberti (University of Milan, Italy)
Non-academic mentors	• Mohamad Othman Abdallah (ProBayes, France) • Marco Penco (Info Solution SpA, Italy) • B. Benatallah (DCU, France)
Secondments **(1 to 6 hosting months)**	• ProBayes, Grenoble, France (Mohamad Othman Abdallah) months 12-18, conversational solutions and validation tests of experimental results related to developed prototypes. Months 12-18 (maximum 6 months) for designing conversational solutions and 18th month -six months for the assessment of conversational solutions after having worked on LLM's. • Info Solution, Vimodrone, Italy (Marco Penco), month 24th up to 4 months, working with autonomous navigation system with the aim of simplifying the interaction between farmers (or non-expert users) and agri-robots for programming field missions.
Contact information	• genoveva.vargas-solar@cnrs.fr • roberto.oberti@unimi.it

RECRUITMENT CRITERIA

General criteria

- MSCA Mobility Rule: researchers must not have resided or carried out their main activity (work, studies, etc.) in **France** for more than 12 months in the 36 months immediately before their date of recruitment
- All researchers recruited in a DN must be doctoral candidates (i.e. not already in possession of a doctoral degree at the date of the recruitment)
- An applicant must have received the equivalent of 300 ECTS with a major in computer science, data science/engineering, or artificial intelligence, databases, from which at least 60 ECTS corresponds to

a master's degree. The master's degree must be granted by a university recognised by the International Association of Universities.

- Scientific excellence to fit the PhD project
- Fluent (oral and written) English skills, as the project operates in the English language
- Knowledge of the language of the host country may be considered a merit
- Team-mindedness

Required skills

- MSc (or equivalent) in Computer Science / Data Science / AI with strong results
- Proficiency in Python and experience with ML / deep learning libraries
- Knowledge of data management, databases and/or big-data / streaming frameworks
- Interest in NLP and large language models, and motivation to work at the interface of AI, data engineering and sustainable agriculture
- Ability to work in an international, interdisciplinary team

Desirable / nice-to-have skills

- Experience with IoT, robotics or cyber-physical systems, or at least strong motivation to learn about Agri-robots and sensor infrastructures.
- Basic knowledge or strong interest in agriculture, agri-food systems or environmental sustainability (the domain knowledge can be acquired during the PhD, but curiosity here is important).
- Prior experience with conversational interfaces or NLP applications (chatbots, dialogue systems, question answering, text mining, etc.).
- Familiarity with version control (Git), and reproducible experiments.
- Previous exposure to interdisciplinary or collaborative projects (e.g. working with non-CS partners, applied projects in another domain).

APPLICATION

How to apply?

- **All information are provided here**

Deadline: **15th April 2026**

Other information

The **National Centre for Scientific Research (CNRS**), better known by its acronym CNRS, is the largest public scientific research organisation in France. It operates in all fields of knowledge. It is a public institution of a scientific and technological nature (EPST) under the administrative supervision of the Ministry of Higher Education, Research, and Innovation. CNRS - Computer Science has coordinated research conducted at CNRS in the areas of computer science, automation, signal and image processing, robotics, and chip design. The development of computer science is at the heart of contemporary digital transformations and impacts research methodologies in other disciplines. The mission of CNRS Computer Science is to support these profound structural changes by contributing to the structuring of this still young discipline. The LIRIS (Laboratory of Computer Science in Image and Information Systems) is a French research laboratory in computer science. It is affiliated with the CNRS under the label UMR 5205. Comprising around 320 people, including nearly 110 researchers and faculty members, it is supported by four institutions: INSA Lyon, Claude Bernard University Lyon 1, École Centrale de Lyon, and Lumière University Lyon 2.

University of Milan (UniMI - Università degli Studi di Milano)

The University of Milan is a major public institution with more than 60,000 enrolled students. Renowned internationally for excellence in research, teaching, and cultural engagement, it combines creativity, innovation, and a strong global outlook. Located in the heart of one of Europe's most dynamic cities, the University is deeply integrated into Milan's academic, scientific, innovation, and cultural ecosystems.

Milan itself is a vibrant international hub, recognised for its culture, creativity, design, and innovation. It offers an ideal environment for international students: multicultural, well connected, and rich in opportunities for internships and collaborative projects with an extensive network of industries and institutions.

With over 130 Bachelor's and Master's degree programmes and 34 PhD programmes, the University of Milan provides an exceptionally broad and interdisciplinary academic offering. Students benefit from programmes designed to address the challenges of a rapidly evolving world, supported by an international environment, global networks, and a thriving cultural scene that makes Milan an inspiring place for study and personal growth.

As the only Italian member of the League of European Research Universities (LERU), the University consistently ranks highly in major international rankings and attracts substantial competitive funding, including numerous European Research Council (ERC) grants, recognising the quality and impact of its multidisciplinary research.

PhD candidates benefit from an active, multidisciplinary research environment, access to state-of-the-art laboratories and research infrastructures, participation in European and international networks, interdisciplinary doctoral schools, and extensive mobility opportunities.

The University welcomes a diverse global community and offers dedicated services and programmes for international students, alongside cultural and social initiatives, inclusion projects, sports activities, and high-quality artistic events throughout the year.

The University of Milan is also a member of the 4EU+ Alliance, a consortium of six leading European research-intensive universities—Paris-Sorbonne, Charles University in Prague, the University of Copenhagen, Heidelberg University, and the University of Warsaw—committed to developing an integrated European higher-education system through mobility, joint programmes, and shared research infrastructures.

The Department of Agricultural and Environmental Sciences (DiSAA) focuses on the study and research of agriculture, food systems, and environmental sustainability. DiSAA addresses pressing global challenges, including climate change, biodiversity conservation, and sustainable resource management. The department fosters innovation in agricultural practices, environmental protection, and food production, equipping students with both theoretical knowledge and practical skills to address future agricultural and ecological challenges.