Green Field Data

Exceptional
36-months Double Degree PhD <u>Scholarship</u>
Position C
<u>Phd in Computer Science</u>
Powering data-driven sustainability assessment tasks in Agri-food systems with IoT-data Data-lakes and large language models

EU Recruiting institutions	Aarhus University, Electrical and Computer Engineering (ECE), Aarhus, Denmark (18 Months), Supervisor: Clause Aage Grøn Sørensen
	Université Libre de Bruxelles, Dept. of Computer and Decision Eng., Data Science and Engineering lab, Belgium (18 Months), Supervisor: D. Sacharidis
Keywords	Data Science, ML & AI, Digital Agriculture, Remote Sensing (IoT and IoRT), Sustainability, Impact assessment, IoT data, Data lakes, 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

Gross amount

Living Allowance	Mobility Allowance*	Family Allowance**
EUR 5325	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 a 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 C – Powering data-driven sustainability assessment tasks in Agri-food systems with IoT-data "Data Lakes" and "Large Language Models"

Context: Digitization and the interconnection of agriculture processes are key to delivering national and international sustainability goals such as climate-smart and environmentally sustainable Agri-food systems sectors. Digitalization is the necessary precondition for scalable and sustainable Agri-food systems. In this context, Internet of Things (IoT) services [6], not only the connection and integration of devices that monitor the physical world (e.g., temperature, pollution, energy consumption, emissions) but also analytics over information including data represented as a collection of time sequences of events (i.e., time series data [7]) that those devices create to derive insights and take appropriate actions. However, this relies on creating an environment where data is curated, standardized, integrated, and organized while Ag-tech stakeholders can create new value from this data effectively and efficiently.

More specifically, data-driven assessment tasks and processes (e.g., LCA-Life-Cycle Assessment-based processes [4,5]) in accounting for effects, trade-offs, and synergies between different mitigation activities in the efforts to reduce climate impact and increase sustainability in Agri-food systems may leverage analytics over data lakes [1,2] for extracting knowledge and deriving insights from the vastly growing amounts of local, external and open data nonetheless. In contrast, advances in data analytics enable tremendous automation and scalability opportunities, and new productivity and usability challenges emerge, especially in large-scale data lakes involving structured, semi-structured, and unstructured time series data. Specifically, agricultural data is diverse in terms of format and content, fragmentation, metadata description, standards, and semantics, which means that the interoperability problem exists both at the technical and semantic levels. Firstly, the same repetitive, error-prone, and time-consuming data curation and integration work needs to be done each time a developer wants to curate knowledge from data sources, to make them easily accessible to applications. Secondly, the current generation of technologies in data analytics primarily targets data scientists, i.e. professional data curators. Much of these available technologies presuppose technical expertise comparable to that of professional programmers, including skilfully employing different programming languages such as SQL and Python and low-level APIs to access various data sources, together with procedural data flow constructs to create and maintain complex data curation and integration scripts. For instance, mainstream assessment techniques and tools rely heavily on experts' knowledge to determine which parts of the data lake(s) to analyse and perform assessment tasks (e.g., answer assessment questions). As a result, using these approaches is difficult for enterprise users and expensive as well. For instance, small and medium-sized organizations may not have the resources to leverage LCA techniques and tools. In a nutshell, LCA assumes the availability of expert knowledge to understand what should be measured (purpose-driven data collection), what data sets to use or collect and use, etc.

Consequently, measurements are expensive and not reusable.

Objectives: In this PhD, we aim to investigate novel techniques to scale and simplify data-driven impact assessment processes making them intuitive, powerful, and accessible to end users, by leveraging large language models (LLMs) [3]. LLMs are pre-trained on broad data and can be adapted to diverse tasks, from question answering, and summarization, to data wrangling [REFs]. More specifically, we aim to provide users with a single-entry point where they define and structure express assessment tasks in a natural way (e.g., text) and interact with data lakes to perform such tasks. All the mentioned objectives can be listed as follow:

1. Develop methods for assessing climate impacts (e.g. GHG emissions) in the robotized cereals production system by means of Life Cycle Assessment (LCA) and innovative generic methods (Task 5.2)
2. Make them intuitive, powerful, and accessible to end users, by leveraging gender inclusive large language models (LLMs) (Task 3.2).
3. LLMs are pre-trained on broad data and can be adapted to diverse tasks, from question answering, summarization, to data wrangling (Task 1.1). More specifically, we aim at providing users with a single-entry point where they define and structure express assessment tasks in a natural way (e.g. text) and interact with data lakes to perform such tasks.

These objectives will be achieved using the following work planning to grant their feasibility.

Work plan:
1. Develop and target feasible methods for assessing climate impacts (e.g., GHG emissions) in the robotized plant production system by means of Life Cycle Assessment (LCA) and innovative generic methods. [Duration: Months 6-9]
2. Develop and implement methods for specific aspect of sustainability assessment (Economic, Operational, Environmental) relevant for the project objectives. [Duration: Months 9-21]
3. Applying large language models (LLMs) to analyse and interpret sustainability data, translate numeric IoT insights into actionable narratives, and simplify reporting making the assessments intuitive, powerful, and accessible to end users. This task will be assisted by PhD students at ULB supported by the HE projects Wise Food (GA 101181895) and who bring expertise on reliable and explainable LLM-based systems. [Duration: Months 19-25]
4. Develop user-centric tools that can automatically generate reports and charts based on user queries. [Duration: Months 24-30]
5. Develop scenario building and applying what if tools to explore different LCA scenarios. This task will be assisted by researchers at ULB with expertise at counterfactual reasoning. [Duration: Months 30-33]

Expected Results

1. Climate change assessment (GHG emissions) by means of Life Cycle Assessment (LCA) and innovative generic methods
2. Gender inclusive query language based LLM
3. Tools that can automatically generate reports and charts based on user queries
4. What if tools to explore different LCA scenarios. These results will make data-driven assessment techniques more user-friendly, gender+ inclusive, and accessible, allowing faster and more efficient analysis of plant production data, and promoting transparency and accountability in agricultural sustainable plant production.

References

[1] Yeye He, Jie Song, Yue Wang, Surajit Chaudhuri, Vishal Anil, Blake Lassiter, Yaron Goland, Gaurav Malhotra. Auto-Tag: Tagging-Data-By-Example in Data Lakes. CoRR abs/2112.06049 (2021)

[2] Rihan Hai, Christoph Quix, Matthias Jarke. Data lake concept and systems: a survey. CoRR abs/2106.09592 (2021)

[3] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Nee-lakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. Advances in neural information processing systems, 33:1877–1901, 2020.

[4] Kronborg Jensen, J. (2012). Product carbon footprint developments and gaps. International Journal of Physical Distribution & Logistics Management 42, 338–354, 2012

[5] Curran, M.A., Laboratory (US), N.R.M.R., Research, O. of, and Development, U.S., Environmental Protection Agency. Life-cycle assessment: principles and practice, 2006.

[6] Athman Bouguettaya, Quan Z. Sheng, Boualem Benatallah, Azadeh Ghari Neiat, Sajib Mistry, Aditya Ghose, Surya Nepal, Lina Yao. An Internet of Things service roadmap. Commun. ACM 64(9): 86-95 (2021)

[7] Quang-Khai Pham, Guillaume Raschia, Noureddine Mouaddib, Régis Saint-Paul, Boualem Benatallah.Time sequence summarization to scale up chronology-dependent applications. CIKM 2009: 1137-1146

PRACTICAL INFORMATION

Recruiting and host institutions	• Aarhus University, Electrical and Computer Engineering (ECE), Aarhus, Denmark (18 Months) (Recruiting institution) • Université Libre de Bruxelles, Dept. of Computer and Decision Eng., Data Science and Engineering lab, Belgium (18 Months)
Doctoral schools	• GSTS @ Aarhus University, Aarhus, Denmark • EPB @ Université Libre de Bruxelles, Belgium
Supervisors	• Pr. Claus Aage Grøn Sorensen (Aarhus University, Aarhus, Denmark) • Pr. Dimitris Sacharidis (Université Libre de Bruxelles, Belgium)
Non-academic mentors	• M. Pedersen (ImpactSphere Inc., Denmark) • B. Benatallah (DCU, France) • M. Hammal (CGI, France)
Secondments (1 to 6 hosting months)	• ImpactSphere, Denmark, months 12-18, Work and train using generic sustainability tools like ImpactSphere facilitating system analysis and model building for sustainability assessment in Agri-food systems with IoRT data lakes. • DCU, France, month 19, Work and train using LLM research technologies • CGI, France, months 20-23, Work on leveraging LLMs for Agri-food systems.
Contact information	• Claus.soerensen@ece.au.dk • Dimitris.Sacharidis@ulb.be

RECRUITMENT CRITERIA

General criteria

- MSCA Mobility Rule: researchers must not have resided or carried out their main activity (work, studies, etc.) in **Denmark** 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. The master degree must be granted by a university recognized by the International Association of Universities.
- Scientific excellence to fit the PhD project
- Fluent (oral and written) English skills as the project operates in English language (English language requirement is comparable to a minimum of TOEFL 83 or IELTS 6.5)
- Knowledge of the language of the host country may be considered a merit
- Team-mindedness

Required skills

- Advanced software developing and programming skills
- Advanced data analysis skills
- Familiarity with LLM-based system skills
- Familiarity with Life Cycle Assessment (LCA) methods
- Familiarity with sustainability assessment methods

APPLICATION

How to apply?

- **All information is provided [here](#).**

Deadline: **15th April 2026**

Other information

Aarhus University

Aarhus University is a dynamic, modern and highly international university. Since its founding in 1928, it has grown to become a leading public university with international impact and reach across the entire research spectrum. Aarhus is a great place to study and a great place to work. Here are some of the top reasons to come work or study with us.

We are a Top 100 University: AU is consistently ranked as one of the world's top universities. It was ranked number 85 in the latest Shanghai ranking and is among the world's 100 best universities in 17 out of 48 subjects in the latest QS World University Rankings by Subject.

Denmark is a great place to live.

Safe, secure and equal: Denmark is widely cited as one of the world's most liveable places for a variety of reasons. It has the world's highest level of income equality, according to the OECD. Furthermore, Denmark is widely cited as one of the world's most liveable places. And based on the Corruption Transparency Index, Denmark is the least corrupt country in the world. Levels of public trust are high and crime is low.

We're always here for you: AU offers all the practical support and guidance you need to feel at home here. For students, the International Centre offers a full induction and introduction programme as well as professional, friendly support and guidance throughout your time at AU.

For researchers and PhD students, Aarhus University offers a full range of services to make your transition to the university as smooth as possible. IAS (International Academic Staff) assists researchers and PhD students from abroad with all practical matters, including visa, residence and work permits, housing services, child care, etc.

Top reasons to study in Aarhus

Our teachers are world-class lecturers: At AU, classes are taught by active researchers in an informal, stimulating atmosphere. All degree programmes are deeply rooted in the latest research and are subject to a rigorous quality assurance programme to ensure they meet the highest global quality standards.

English is our second language: AU offers more than 50 full degree programmes in English at Bachelor's and Master's level. All PhD programmes are in English. And Danes were recently ranked as among the best non-native English speakers in the world on the on the annual English Proficiency Index (EPI), so it's easy for international students to feel at home in Denmark from day 1.

Our student body is diverse and highly international: Around 12 per cent of AU's 36,000 students are internationals – from over 120 countries.

Denmark offers an attractive green card scheme: The green card residence permit granted to international university students in Denmark is valid for an additional six months after completion of the degree, which gives graduates time to look for work in Denmark.

Top reasons to work at AU

We have state-of-the-art research facilities: Aarhus University offers world-class research facilities and laboratories in a wide range of subjects. We have a strong tradition of multidisciplinary research for instance in one of our 42 major research centres.

We offer attractive working conditions: When asked about what they appreciate most about AU, our international staff at AU emphasise favourable working conditions as an important motivation for working at Aarhus University. These include an attractive salary, a generous pension scheme and parental leave benefits. What's more, international academic staff members are eligible for tax breaks in many cases.

We are world champions in work-life balance: Danish workplace culture stresses the importance of work-life balance, and AU is no exception. Researchers enjoy flexible working conditions and a high level of autonomy and self-determination. What's more, the municipality offers a range of high-quality social services that enhance the quality of life for internationals, including free healthcare and subsidised childcare and international schools.

Université Libre de Bruxelles (ULB)

Université Libre de Bruxelles (ULB) is a major public research university in Brussels with a broad international profile and multidisciplinary teaching and research across science, engineering, social sciences and humanities. ULB runs numerous Master and PhD programs in computer science,

engineering and data science and hosts many research partnerships with national and European research networks. ULB participates in this project with the *Data Science & Engineering Lab*, which consists of three professors (e.g., E. Zimanyi, D. Sacharidis, M. Sakr) and several researchers working on data management, data engineering, and machine learning within the CoDE umbrella and the Brussels School of Engineering. The teams work on both theoretical and applied problems, related to databases, spatio-temporal data, and responsible AI.

Brussels is an excellent place to live and work

A cosmopolitan capital: Brussels is an international hub for science, technology, arts, and policymaking, hosting the EU institutions, major research organizations, and global companies.

Safe, multicultural, and well connected: The city offers high-quality healthcare, efficient public transport, and easy access to the rest of Europe.

English-friendly environment: Although French is the main local language, English is widely spoken in academia, industry, and daily life.

Top reasons to work at ULB

Research Facilities and Innovation Networks: ULB provides access to extensive research facilities and innovation networks that support high-level work in AI, data engineering, and related fields. Researchers benefit from computing resources, specialized laboratories for AI and data-intensive experimentation, and participation in regional access-to-labs initiatives that connect academic and industry partners. ULB is also a founding member of FARI – AI for the Common Good Institute, a Brussels-based interdisciplinary centre that brings together ULB and VUB experts in artificial intelligence, data governance, and digital policy. Through FARI, ULB researchers gain access to shared infrastructures, applied research programmes, and collaboration pathways with public institutions, companies, and societal stakeholders across Brussels and Europe.

ULB supervisors are active researchers: ULB's academic staff are active researchers who drive cutting-edge work across computer science, data engineering, and AI. Their research expertise directly shapes the university's programmes, ensuring that students engage with the latest scientific developments and methodologies. Teaching is grounded in rigorous quality assurance processes and is closely integrated with research laboratories, creating a strong link between advanced research activity and graduate training.

Attractive working conditions and research support: ULB offers: Structured doctoral training through its doctoral schools. Support for grant applications, EU projects, technology transfer, and industry collaboration. Access to state-of-the-art computing resources, specialized labs, and innovation infrastructures (e.g., Sustain.brussels and access-to-labs networks). Career development workshops, mobility grants, and international exchange opportunities

Participation in major European networks: ULB is a core partner in initiatives such as DEDS (Data Engineering for Data Science) and coordinates the DEAI Erasmus Mundus Joint Master, reflecting a strong focus on responsible AI, data engineering, and end-to-end data lifecycle research and training.

Balanced academic life in a vibrant city: Researchers benefit from flexible working conditions and access to high-quality public services. Brussels's cultural and social life offers rich opportunities outside the lab, combined with excellent transport links and a strong international community.