Bridging Global Language Models and Local Spatial Data: The JackDaw Approach to Context-Aware Agriculture and Rural Planning

Karel Charvát [1], (Stein) Runar Berhgeim [2], Raul Palma [3], Adam Aron Rynkiewicz [3], Matúš Botek [5], Alexander Kovalenko [4], Pavel Kordík [4], Antonín Kubíček [5], Šárka Horáková [1], Markéta Kollerová [5]

[1] Czech Center for Science and Society, Radlická 663/28, Smíchov, 150 00 Praha, Czech Republic
Email: charvat@ccss; horakova@ccss.cz

[2] Asplan Viak Internet Ltd, Havnegården (Kystveien 14), Postboks 701 Stoa, Arendal, Aust-Agder 4808, Norway
Email: rb@avinet.no

[3] Poznan Supercomputing and Networking Center, ul. Jana Pawła II 10, 61-139 Poznan, Poland
Email: rpalma@man.poznan.pl; adam.rynkiewicz@doctorate.put.poznan.pl

[4] Faculty of Information Technology, Czech Technical University in Prague, Thákurova 9, 160 00 Praha 6, Czech Republic
Email: alexander.kovalenko@fit.cvut.cz; Pavel.Kordik@fit.cvut.cz

[5] Plan4all z.s., K Rybníčku 557, 33012 Horní Bříza, Czech Republic
Email: matus.botek@gmail.com, tony.kubicek@gmail.com; marketa.kollerova@plan4all.eu

Abstract—**Large Language Models (LLMs) excel at synthesising globally documented knowledge but lack the fine-grained, real-time awareness required for field-level agricultural and rural-planning decisions. This paper introduces JackDaw, a spatially enabled chat-agent architecture that couples foundation-model reasoning with multi-modal geospatial data streams and a retrieval-augmented generation (RAG) pipeline. JackDaw implements a tool-prefiltering mechanism that selects only those data connectors whose topical, temporal and spatial metadata match the current query, thereby mitigating the diminishing returns observed when LLMs are exposed to large, flat toolsets. Through LangChain-based orchestration the platform dynamically assembles workflows that range from lightweight natural-language processing models to domain-specific analytic kernels, while a value-engineering strategy allocates computationally intensive models (e.g., GPT-4-class) only to tasks that require broad contextual reasoning. Benchmark experiments on forestry-asset discovery and vineyard-site assessment demonstrate that JackDaw delivers location-specific, traceable answers that outperform a standalone proprietary LLM, which provides only generic or spatially misattributed responses. The results confirm that bridging global language models with local spatial intelligence markedly reduces hallucination rates and enhances the operational readiness of AI for sustainable agriculture and rural development.**

Index Terms—**Large language models; geospatial AI; retrieval-augmented generation; context-aware agriculture; spatial decision support; tool prefiltering; JackDaw system; rural planning.**

I. Introduction

Artificial intelligence (AI) has transitioned from a niche computational technique to a pervasive epistemic lens through which agriculture and rural development are increasingly interpreted. In agronomic contexts, AI is not merely a collection of algorithms but a socio-technical apparatus that converts raw environmental signals into actionable knowledge, thereby mediating the long-standing tension between global food-system efficiency and local stewardship of land, labour and biodiversity. By formalising inductive reasoning at unprecedented spatial and temporal scales, AI promises to reconcile high-frequency decision support for producers with broader rural-development directives—resilience, equity, sustainability—provided its deployment remains transparent, uncertainty-aware and ethically aligned [1], [2].

AI pipelines now ingest heterogeneous data streams from in-situ sensors (soil-moisture probes, automated weather stations), passive and active satellite constellations (multi-spectral Sentinel-2 MSI, C-band Sentinel-1 SAR, DESIS hyperspectral, LiDAR), low-altitude unmanned-aerial systems, and physics-based or statistical climate re-analyses (GFS, ERA-5-Land). Early studies relied on single-sensor inputs and shallow classifiers; contemporary frameworks fuse multi-temporal, multi-modal observations via machine-and deep-learning architectures optimized for high-

This work was is supported by the EU project PoliRuralPlus

dimensional spatiotemporal covariance structures. Sentinel-2 reflectance cubes harmonised with Sentinel-1 backscatter underpin continental land-use-intensification inventories [3] and sub-field crop-type delineation [4]. Ensemble decision-tree learners, notably Random Forest, remain baseline regressors for temperate-cereal yield [5] and national wheat statistics [6], while deep neural networks gain traction where label density is sufficient: One study [7] stacked convolutional, gated-recurrent, and transformer encoders on fused Sentinel-1/2 sequences for pixel-level classification; The method proposed in [8] introduced the 3-D → 2-D HypsLiDNet that couples DESIS voxels with LiDAR-derived canopy structure; Research in [9] demonstrated multimodal convolutional fusion of Sentinel-2, edaphic layers and mesoscale meteorology for field-scale yield forecasting. Object-based learning exploiting explicit geostatistical segmentation persists for landscape stratification—findings from [10] combined Sentinel-2A, ALOS DSM, and PALSAR layers with Random Forest to derive agro-ecological strata—and temporal self-attention mechanisms have been applied to integrated crop-livestock delineation in the subtropics [11].

Operational deployment is constrained by observation gaps (cloud cover, revisit frequency) and by aleatoric uncertainty in multi-sensor fusion. Research described in [12] mitigated these issues by embedding Sentinel-1 SAR into the UnCRtainTS cloud-removal network, lowering spectral MAE to ≈ 0.025 reflectance units, and enabling continuous NDVI trajectories across Central-European croplands. On the prognostic front, high-resolution climatic forcing has been advanced by up-scaling 0.25° GFS products to ERA-5-Land resolution via a U-Net super-resolution model, merging the output with HadISD station records, and applying CatBoost regression to reduce 24-h temperature MAE to ~1 °C [13]. Nevertheless, most agronomic AI studies still rely on internal cross-validation; spatially independent or inter-annual robustness tests remain rare—examples include paddock-out validation [14] and temporally stratified ten-fold splits [15]. Consequently, the methodological frontier is defined by scalable multi-sensor fusion, uncertainty-aware ensembles or transformers, and rigorous cross-region benchmarking to support operational use in dynamic agroecosystems.

Large-language-model (LLM) research entered agricultural informatics only recently, yet adoption is rapid. A systematic scan identified 26 geospatial-LLM papers up to early 2024—almost all post-November 2022—describing the field as "embryonic but rapidly accelerating" [16]; a parallel survey of multimodal ag-LLMs reached similar conclusions and noted rudimentary benchmarks [17]. Developments have progressed from demonstration chatbots to end-to-end systems: ChatGeoAI converts plain-language queries into PyQGIS code [18]; OmniGeo demonstrates zero-shot reasoning over satellite imagery, vector polygons, and free text using an instruction-tuned vision-language model [19]. Text-focused agronomic advisers include a retrieval-augmented fruit-and-vegetable system coupling Baichuan-2 with RAG to reduce hallucination rates by 10–40 % [20] and a cotton-soil engine where a fine-tuned GPT-2 outperforms LLaMA-2 for nutrient recommendations [21]. Synthetic data generation is exemplified by an LLM-driven agentic workflow that fabricates realistic fault patterns for smart-tractor telemetry, enabling predictive maintenance testing). Document intelligence pipelines empower domain-agnostic LLMs to extract structured pest information from agronomic literature in zero-shot mode [22].

Across these studies, shared hurdles are evident. High-quality labeled agri-text corpora and paired image–text datasets remain scarce, forcing reliance on generic foundation models or few-shot prompting [20] [17]. Factual reliability is fragile, motivating retrieval-augmented generation [16]. Multimodal grounding across satellite, drone, sensor, and text streams is largely heuristic, as shown in [17] [19], while the compute and energy costs of fine-tuning multi-billion-parameter models raise sustainability concerns. Explainability, data governance, and ethical compliance are critical yet under-explored prerequisites for commercial farm deployment [16].

Within two harvest seasons, the community has advanced from isolated demonstrations to retrieval-augmented, multimodal assistants addressing yield prediction, soil management, pest intelligence, and equipment maintenance. Consolidating open data resources, curbing hallucinations, mastering efficient geo-text fusion, and delivering transparent, farmer-oriented explanations now constitute the central agenda for AI-driven rural development.

II. Motivation

Large Language Models (LLMs) have significantly impacted various domains by leveraging vast datasets to provide comprehensive responses on diverse topics. Their capability to process global and well-documented information allows them to efficiently recognize general trends, interpret historical data patterns, and detail well-known geographical areas. However, despite these substantial strengths, LLMs frequently exhibit limitations concerning the provision of precise, real-time insights required for location-specific and dynamic contexts, such as agriculture and rural development.

In agricultural practices, where decisions rely heavily on real-time data about local environmental conditions, precise field-level analytics, and seasonal variability, the limitations of standalone LLMs become particularly pronounced. For instance, while an LLM may comprehensively discuss global agricultural trends or extensively documented practices, it may lack critical, timely insights into crop conditions in a less-documented rural area or the subtle seasonal variations affecting local agricultural productivity.

Integrating LLMs with geospatial data offers a robust solution to bridge these gaps. Real-time geospatial data, encompassing remote sensing imagery from satellites, drones, weather stations, and in-situ sensors, can complement the generalized insights of LLMs with detailed, actionable information at specific locations and moments in time. By coupling spatial analysis technologies—already established through remote sensing platforms such as Sentinel-1 and Sentinel-2, as well as advanced AI-driven methodologies—with the broad contextual knowledge offered by LLMs, systems like JackDaw have demonstrated the potential to significantly enhance rural-urban planning and agricultural decision-making.

The JackDaw prototype has underscored the value of this integrative approach by successfully combining the power of LLM reasoning with spatial analysis to provide context-aware insights critical for agriculture and rural management. Its approach exemplifies how synthesizing multi-modal data streams with deep-learning architectures can mitigate the limitations stemming from gaps in sensor observations or uncertainties in data fusion. Furthermore, JackDaw exemplifies the potential of multi-modal integration to improve practical agricultural outcomes by offering more precise yield forecasts, targeted resource management recommendations, and improved environmental monitoring.

Thus, developing an integrative platform that combines the strengths of LLMs with standard agricultural AI tools and GIS solutions represents a strategic opportunity. Such integration promises substantial advancements in rural agricultural practices, enabling more robust, evidence-based decision-making, enhancing sustainability, and ultimately improving rural livelihoods and economic development.

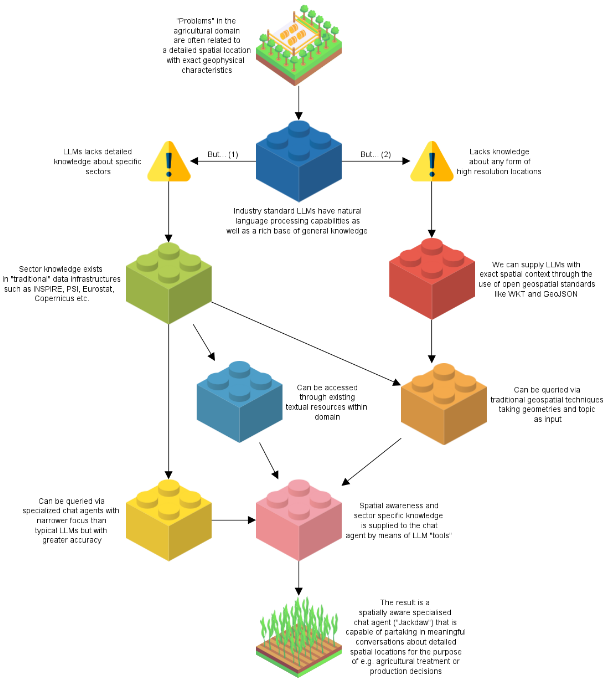

Fig 1. The motivation for developing the JackDaw chat agent

III. Solution

The JackDaw solution is designed to enhance the functionality of LLMs by addressing their inherent limitations in accessing and interpreting localized, real-time data. This enhancement is significant in agricultural and rural planning professions, where precise localized knowledge is crucial. The gap between the global knowledge embedded in LLMs and the need for detailed local contextual information is significant. JackDaw attempts to bridge this gap by integrating advanced geospatial data analysis and specialised AI tools into conversational AI interfaces. This integration enables more accurate and contextually aware responses from LLMs than what is possible with the foundation models.

The conceptual solution underlying JackDaw is facilitating meaningful conversations about geographical locations through LLMs. To achieve this, it is essential for LLMs to be not only aware of the location being discussed but also to possess an understanding of what specific data is relevant to that place. Furthermore, to provide useful responses, JackDaw requires the model to discern which data sources should be utilised based on the topic, purpose, and spatial location of the conversation. These LLMs are expected to engage iteratively with various data sources to foster an incremental understanding of the context-specific to each inquiry.

Conceptual workflow of a conversation with the JackDaw chat agent

JackDaw operates using a systematic workflow that begins when a user initiates a query related to a specific geographic location through a user message to the agent. The system interprets this query and tries to match it against a large number of available "tools" that connect the LLM to a wide range of relevant data sources. These sources include spatial datasets, such as those defined by the Infrastructure for Spatial Information in Europe (INSPIRE), public sector information (PSI), and specialised AI models optimised for agricultural analytics.

Upon identifying which "tools" may be relevant to answering the question, JackDaw proceeds to pass them to the language model, which collects both spatial and tabular data sets pertinent to the user's query. This includes sensor-derived field data, satellite imagery, current weather forecasts, soil properties, and agronomic statistics. Specialised AI models process these datasets to generate analytical insights, which are then contextualised and embedded into the chat interface.

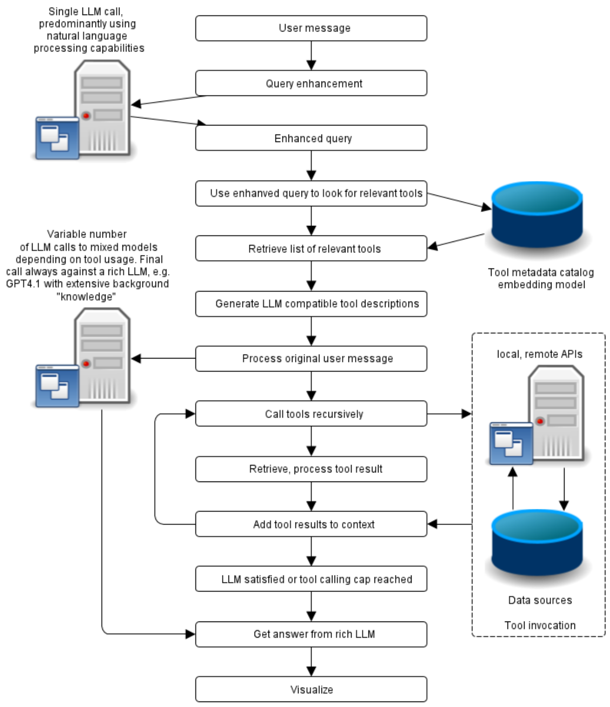

Fig 2. Conceptual flow of a conversation with the JackDaw spatial-enabled chat agent

Visualisations derived from the processed datasets are integrated into the user interface, providing graphical context to the inquiries. These visual representations facilitate clearer insights and more actionable decision-making. By embedding enriched data insights into the LLM's reasoning process, JackDaw effectively bridges the gap between general global knowledge and the need for localized, actionable information. This approach offers robust and informed answers to user inquiries, promoting enhanced decision-making processes tailored specifically to rural and agricultural contexts.

Through the innovative integration of geographic information systems (GIS) and advanced AI-driven data analytics, JackDaw capitalises on existing expertise within these domains. This harnessing of advanced technologies promotes a nuanced and effective decision-making process, significantly enhancing LLM capabilities in understanding and responding to location-specific questions. Consequently, JackDaw transforms the application of LLMs in agricultural planning and rural development, making them more responsive and accurate in addressing context-specific inquiries.

Enhancing Chat Conversations with Spatial Context

Integrating an area of interest into a chat query can significantly augment user interaction by providing spatial context to the conversation. This enhancement involves adding spatial features that encapsulate the geographical areas users wish to inquire about. The spatial data is represented using Well-Known Text (WKT) geometry features, a format allowing a detailed description of geometric shapes in a human-readable text string. By incorporating WKT geometry, users can define specific areas of interest directly within the chat interface.

Ways of Incorporating Spatial Features

Users can add spatial features to a chat query in two primary ways: through a client interface or via natural language processing (NLP). When utilising a client interface, users can directly draw the areas of interest onto a map. This method relies on graphical tools that allow precise selection and adjustment of shapes to represent the desired geographical space accurately. These shapes are then converted into WKT format, maintaining their geometric characteristics while enabling text-based communication in the chat. This method provides users with a tangible and intuitive way to specify areas of interest.

Alternatively, spatial features can be derived from NLP queries. In this approach, a tool or algorithm processes the user's natural language inputs to extract relevant geographical information and convert it into WKT geometry. This method leverages advances in NLP to interpret user intent and spatial references embedded within the chat. Such tools can identify place names, landmarks, or specific directions mentioned by users and translate them into spatial features. This capability allows users to describe their areas of interest using everyday language, simplifying the process and making it accessible to a broader range of users.

Facilitating Function and Tool Integration

Once the spatial features are integrated into the chat, they are encapsulated within code blocks. These code blocks are clearly delineated to distinguish them from other elements of the chat. This demarcation ensures that the data is accurately identified and processed by subsequent tools or functions that the chat agent may call upon. Code blocks serve not only to organise information but also to maintain data integrity when transferring inputs across different operational environments.

Incorporating spatial features as part of chat queries empowers chat agents to utilise this data in subsequent interactions. The defined area of interest can be used as a parameter for various function or tool calls. For instance, a chat agent might access weather data, perform geographical analysis, or retrieve demographic information specific to the identified area. This capacity expands the functionality of chat agents, enabling more sophisticated queries and responses tailored to the user's spatial context.

Adding areas of interest to chat queries through WKT geometry represents a meaningful advancement in interactive technology. By allowing users to define spatial parameters using both graphical interfaces and natural language processing, the system becomes more flexible and usercentric. Furthermore, the encapsulation of data within distinct code blocks facilitates seamless integration with other tools and functions. Such enhancements lead to more effective and targeted interactions, bridging the gap between spatial awareness and conversational interfaces. This integration serves as a transformative enhancement in fields where geographical context is critical, such as urban planning, logistics, and environmental monitoring, thereby expanding the utility and applicability of chat-based systems in professional and research environments.

Connecting data sources to LLMs

Implementing knowledge acquisition in large language models (LLMs) involves equipping these models with an understanding of various "things," or categories of information, which can be used to improve the applicability and accuracy of chat agents. This methodology necessitates the development of versatile tools that connect LLMs with relevant data sources. LLM tools serve a crucial function in facilitating this connection, providing the necessary scaffolding to bridge the gap between an LLM's inherent language skills and its access to a structured knowledge base.

Versatility is a fundamental requirement when constructing these tools. They must be able to interlink effectively with multiple data sources, thereby expanding the functional capabilities of chat agents. To achieve this, tool prompt engineering is an essential process. This involves designing specific prompts that guide the retrieval and processing of data in a manner that aligns with the operational goals of the LLM. Making each tool available through a web service enhances accessibility and usability, allowing these tools to be integrated seamlessly into web-based applications.

Developing numerous specialised tools can significantly enhance the LLM's ability to handle complex inquiries. In the realm of geophysical information, for instance, tools can analyse terrain by examining landforms, elevation, aspect, slope, and surface roughness. This data provides essential context for applications ranging from environmental monitoring to urban planning. Similarly, climatic tools that gather information about temperature, wind patterns, cloud cover, precipitation, and solar radiation contribute to a comprehensive understanding of environmental conditions across different regions and times, enriching predictive models and research studies.

Demographic information represents another vital category, enabling insights into population characteristics and trends. Tools that analyse demographic data must consider population size, growth projections, historical trends, and the composition of populations based on gender and age. This data supports social dynamics, policy making, and economic forecast analyses. Economic tools can assess business demography and market sizes, offering valuable input for economic modeling, investment analysis, and strategic planning.

Temporal tools are indispensable for temporal analysis, which is critical in assessing real-time changes and zeitgeist trends. When coupled with historical data, a snapshot of current weather conditions enables detailed climatic assessments. Moreover, zeitgeist tools, such as those utilising the Global Database of Events, Language, and Tone (GDELT), are instrumental in capturing contemporary socio-political dynamics, offering insights into prevailing public sentiment and events affecting international relations and market behaviours.

Incorporating statistical data with georeferenced identifiers is instrumental for deriving meaningful insights based on location-specific features. By identifying Nomenclature of Territorial Units for Statistics (NUTS) identifiers for particular locations, it is possible to gather relevant demographic and business demography statistics. This georeferenced approach enables tailored analyses considering local variations, which is essential for targeted policymaking and regional development strategies.

The utilisation of raster and grid data complements this analysis by providing detailed spatial information. Sources such as the European Space Agency (ESA), the Group on Earth Observations (GEO), and Sentinel data offer high-resolution satellite imagery and derived raster datasets. These datasets are optimised for large-scale multidimensional queries, allowing users to extract information that spans multiple dimensions and scales. This capability is particularly useful in applications like precision agriculture, land use planning, and disaster management.

Vector data, on the other hand, involves using line and polygon data formats typical of traditional Geographic Information System (GIS) infrastructures. Adhering to standards such as those established by the Infrastructure for Spatial Information in the European Community (INSPIRE), vector data provides a detailed locational context crucial for detailed mapping and spatial analysis. This infrastructure supports a wide array of applications, from infrastructure development to environmental conservation efforts.

Document data serves an integral role in the holistic use of LLM tools, particularly in contexts that require the retrieval of legal documents, plans, reports, and other contextual information. Access to such documents allows LLMs to provide informed responses that are grounded in the regulatory and historical framework of the area in question. For instance, in urban development scenarios, understanding zoning laws and historical city planning documents is essential to delivering relevant insights and recommendations.

Limiting the number of tools made available to the LLM for improved accuracy and efficiency

A significant aspect of working with these models involves the invocation of external tools to enhance their problem-solving capabilities. However, as the number of tools accessible to an LLM increases, the efficiency and effectiveness of these tool invocations tend to diminish. To address this issue, it is critical to develop a robust method for prefiltering tools based on the specific requirements of a query or question.

LLMs are inherently equipped with flat tool invocation capabilities. This means that they can be supplied with a limited set of tools to assist in fulfilling a task. However, as the array of available tools expands, the efficiency of the model's capacity to appropriately call upon these tools declines. This is mainly due to the overwhelming number of potential options, which can obscure the model's decision-making process and reduce its ability to select the most pertinent tools for a given task. Therefore, a prefiltering step becomes necessary to manage tool selection effectively.

Prefiltering tools is an essential process that relies on structural metadata to determine which tools should be considered by the model in relation to the question at hand. This step ensures that only relevant tools are available for use, thereby optimising the model's performance and reducing computational overhead. The prefiltering step can be executed by employing several criteria, each aimed at ensuring the model's proficiency in addressing the specific nuances of the question.

The initial criterion for filtering tools involves sorting by topic relevance. This means identifying tools that align with the subject matter of the inquiry. Questions often pertain to specific topics, and tools designed with an understanding of these subjects can offer more accurate and precise insights. For instance, if a question relates to environmental science, the prefiltering system would bypass tools irrelevant to this field and focus on those offering insights or functionality that complement environmental studies.

Another critical filtering criterion is temporal resolution. Questions may address events or data related to particular time periods, and not all tools possess the capability to handle temporal queries comprehensively. The prefiltering system must, therefore, identify and prioritise tools that are equipped to deal with the specified temporal context. For example, if a query involves historical climate data from the 1980s, the system should filter out tools that only provide contemporary data, ensuring the most temporally relevant tools are selected.

In addition to temporal relevance, the selection of tools may also hinge on determining the best source of information, especially in scenarios with overlapping data coverage. This means the system should favour tools that offer the most precise or contextually appropriate data. For example, if a model for Norway offers superior data specificity compared to a broader European model for a particular question, the prefiltering mechanism should select the Norwegian model. This ensures that the data used is both relevant and of the highest quality possible for the query.

Spatial resolution is another determinant in the prefiltering process. Many questions necessitate specific spatial data resolutions, and tools may vary greatly in the granularity of their data. It is essential to filter tools based on the spatial requirements specified in the question. If a question demands high-resolution spatial data, such as a 1-meter resolution for detailed topographic analysis, the prefiltering mechanism should prioritise those tools capable of fulfilling such needs. Conversely, for less detailed queries, tools offering broader, 100-meter resolution data may suffice.

The implementation of a structured prefiltering step enhances the ability of LLMs to select and utilise tools effectively, leading to more accurate and efficient outcomes. By categorising tools based on topic, temporal and spatial capabilities, and data quality, the system reduces the cognitive load on the model and ensures that only the most relevant tools are engaged. This not only improves the model's response time and accuracy but also conserves computational resources.

Agent implementation using Langchain for Python

The integration and orchestration of calls to various machine learning models constitute a key component in the efficient deployment and utilisation of these models in real-world applications. Among the technologies facilitating this process, Langchain emerges as a pivotal tool. It offers an interface designed specifically for dynamic interaction with language models, significantly enhancing applied solutions' flexibility and maintainability.

Langchain serves as an abstraction layer that simplifies the setup of client calls to large language models (LLMs). Through its design, Langchain provides a mechanism by which these models can be interacted with seamlessly, ensuring that the underlying complexities of each model are effectively managed. The abstraction layer implemented by Langchain allows developers to perform model orchestration in a manner that separates the business logic from model-specific code. This, in turn, supports the ease of integration and modular development, facilitating a more structured approach to software development that incorporates these advanced models.

Further, the abstraction layer not only supports seamless communication with a single LLM but also allows for the straightforward replacement of one model with another. This feature is essential in today's rapidly evolving machine-learning landscape, where the rapid development of new models can often leave practitioners with outdated tools. By using Langchain, developers can efficiently swap out LLMs at any workflow stage to achieve improved performance, accuracy, or other desirable traits inherent in more advanced models.

The utility of this abstraction layer extends beyond the wholesale replacement of models. It also permits granular control over individual steps within the model workflow. This means that each component of the process can be optimised by integrating the most suitable model for that specific task. Such granular control is invaluable, as it provides the opportunity to tailor each step of the interaction with the LLM to align with specific requirements or constraints. Consequently, applications deploying these models become more robust and targeted in their outcomes.

Langchain's design enhances the portability and scalability of machine learning systems, allowing for simple expansion in scope and scale. Abstracting model interactions enables systems to cater to increased demand or incorporate complex functionalities without significantly reengineering existing systems. The ability to efficiently manage and organise various LLMs through Langchain allows developers to create more sophisticated applications that can handle a wider range of tasks and challenges.

Moreover, Langchain plays a significant role in encouraging best practices in software engineering and model deployment. Enforcing separation of concerns ensures that the model logic does not hinder the development of other software components. Developers are encouraged to engage in clean code practices and modular development, which enhances the maintainability and longevity of software systems. Using an abstraction layer thus aids in creating a more streamlined development process, where different teams can work on separate components without interference, leading to a more collaborative and efficient workflow.

In addition to improving development processes, Langchain may also contribute to the reduction of deployment risks. It allows testing new models within existing infrastructures without extensive overhauls, facilitating smoother transitions and updates. This capacity to swap models quickly and without significant downtime is especially critical in environments where continuous integration and deployment are essential requirements.

Fostering an ecosystem where evolving models can be integrated without friction is fundamental to maintaining a competitive edge in fields relying heavily on machine learning. Langchain thus acts as a linchpin in the operational management of these models, aligning technological capabilities with business objectives. Organisations are thereby empowered to harness the full potential of LLM advancements with minimal disruption to their existing systems.

Value engineering in the back-end solution

Back-end value engineering is a crucial aspect in the development of computational tools, particularly tools like JackDaw that involve complex model interactions. This discipline focuses on the optimal allocation of computational resources to enhance performance and reduce operational costs. Within the context of back-end value engineering, a nuanced approach to model selection and implementation can lead to substantial improvements in efficiency and effectiveness.

At its core, the process of back-end value engineering involves the strategic use of different models for varying steps within a computational sequence. By tailoring the choice of model to the specific requirements of each task, developers can achieve better resource allocation, ultimately enhancing the tool's performance.

For initial tasks requiring basic language processing capabilities, models possessing natural language processing (NLP) capacity are sufficient. Such tasks often involve identifying appropriate arguments to be passed to a function. NLP models excel in understanding linguistic input, making them ideal for these preliminary steps. Their use ensures that computational resources are not expended unnecessarily on more complex models when simpler ones suffice.

In contrast, certain steps necessitate the employment of a base model with a comprehensive understanding of diverse knowledge areas. Models like GPT-4po serve this purpose effectively. These models are designed to possess a wide-ranging repository of generic knowledge, thereby enabling them to tackle tasks that demand a broad understanding. Their utility lies in their ability to provide insights and execute functions that require a generalist perspective.

Moreover, some stages in a computational process necessitate the application of bespoke models with domain-specific expertise. An example might be a model engineered to handle species tolerance parameters. These custom models are tailored to address specialised tasks, offering insights and solutions that generic models might overlook. Their usage is particularly advantageous in niche applications where precision and specialised knowledge are critical.

By implementing an optimised model selection strategy, the overall cost associated with complex tool-calling chains can be significantly mitigated. This approach entails identifying the most suitable model for any given task and deploying it accordingly. Such strategic allocation not only reduces computational expenses but also enhances the tool's efficacy in performing its designated functions.

The process of determining the optimal model for each step involves several considerations. These include the computational requirements of the task, the desired level of accuracy, and the potential impact on overall performance. By weighing these factors, developers can make informed choices that align with both technical objectives and budgetary constraints.

There is a crucial balance to be struck between leveraging the simplicity and efficiency of NLP models and harnessing the expansive capabilities of base models like GPT-4po. Additionally, the precision offered by customised models with specialised knowledge must be integrated judiciously to maximise effectiveness. Such decisions require careful assessment and strategic planning to ensure that the right model is deployed at the right stage.

In scenarios where tasks are relatively straightforward, deploying simpler models can free up computational resources, thereby allowing more complex models to be reserved for tasks that truly necessitate their capabilities. Conversely, when tasks demand a more in-depth understanding or specialised insight, more sophisticated models are utilised to achieve the desired outcomes.

Furthermore, this tiered approach to model selection also facilitates scalability. As computational demands evolve, new models can be integrated, or existing models can be updated to reflect the latest advancements and requirements. This adaptability ensures the system remains robust and responsive to changing needs without requiring complete overhauls or excessive expenditure.

Integrating diverse models in back-end processes also fosters innovation, as developers are encouraged to explore novel combinations and applications to enhance performance. By considering how different models can interact and complement one another, new solutions can be devised that push the boundaries of current capabilities, all while keeping resource use and costs in check.

Through careful analysis and planning, back-end value engineering plays an instrumental role in the development and deployment of computational tools. By strategically selecting and implementing models tailored to specific tasks, developers can transform their systems into more efficient and cost-effective solutions. This calculated approach not only optimises performance but also paves the way for ongoing innovation and refinement.

While of lesser scientific value, value engineering is critical to ensuring adoption and a realistic path to a real and tangible market.

RAG module

One of the key components that enables JackDaw to deliver more credible and less hallucination-prone responses is the Retrieval Augmented Generation (RAG) module, illustrated in Figure X. This module enhances JackDaw's responses by grounding them in external references and supporting the ingestion of new data sources. The RAG module consists of three main components:

1. **Embedding Model -** Transforms input text into high-dimensional vector embeddings, enabling semantic search within a vector database.
2. **PDF-to-Markdown Conversion Tool** - Parses user-submitted PDF files and extracts both content and structure. It converts the documents into clean, structured Markdown, preserving key elements such as headings, paragraphs, tables, and lists to maintain readability and information integrity.
3. **RAG Application -** The core service that handles user queries, retrieves relevant content from the vector database, and compiles the results for response generation.

The module uses a microservices architecture to support scalability and simplify maintenance. At its core is the RAG Application, which orchestrates the query processing and retrieval workflow.

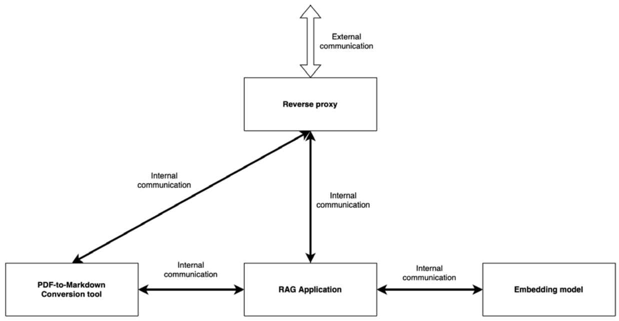

Fig 3. Overview of the RAG module architecture

The RAG module interfaces with the JackDaw system through a reverse proxy, facilitating secure external communication. Internally, the reverse proxy routes requests to two main components: the RAG Application and the PDF-to-Markdown Conversion tool. The RAG Application acts as the central component, orchestrating interactions with both the PDF-to-Markdown Conversion tool and the Embedding Model. All communication between these components occurs internally, ensuring efficient data flow within the system.

Before deployment, the system was seeded with trusted content - policy documents and scientific articles. These documents were processed using the same PDF-to-Markdown tool described above. This tool utilized Docling Library [23] to perform the parsing. The resulting Markdown files were segmented into smaller units of approximately 500 tokens, referred to as chunks.

Each chunk was embedded using the **Alibaba-NLP/gte-Qwen2-1.5B-instruct Model** [24], a high-performing multilingual language model. The resulting embeddings, along with their corresponding text chunks, were stored in **ChromaDB**, an open-source vector database optimized for similarity search.

At runtime, when a user submits a query, the RAG Application first generates an embedding of the query. This embedding is then used to retrieve semantically similar chunks from the vector database. These passages are returned to JackDaw along with metadata that includes source information, enabling more accurate responses and full traceability. This approach ensures that users can access the original documents from which answers are derived, promoting transparency and trust.

Currently, the PDF-to-Markdown tool is limited to processing individual PDF files and returning the parsed content directly to JackDaw. This setup is designed for scenarios where users want assistance understanding or extracting information from specific documents.

In the future, the tool will be expanded to support the creation of personalized knowledge bases. Users will be able to upload multiple PDF files, effectively building a private vector database similar to the existing global one. This personal database can then be used to enrich the responses generated by the language model, resulting in more context-aware, tailored, and useful interactions. Moreover, this setup will support the continuous growth of the data source, enabling users to expand and evolve their assistant's knowledge over time incrementally.

IV. Results

We tested the performance of JackDaw on various scenarios related to geospatial data tasks in agriculture and forestry and compared its solutions with standard proprietary LLMs to determine its benefits and limitations. For each user's query, our system invokes tools with information about the geographical area of interest and reasons over the retrieved content to provide a final answer. To make the comparison more balanced, we include the geographical area in the form of bounding box coordinates in the input query so that the standard LLM has some information about the location. Without this information, the standard LLM would have no way of providing relevant information. The following scenarios showcase that JackDaw is able to provide location-specific relevant information, whereas the standard LLM can only give general advice, sometimes not even related to the area of interest. We used GPT-4o-mini as the LLM.

Scenario 1

In this scenario, we asked about a partially forested area; the user query was: "Is this area forested, and are there any specific utilities regarding forestry?". Figure 4 shows responses from JackDaw and a standard LLM. It also contains the selected area of interest. GPT-4o-mini managed to offer general advice and derived the location from the coordinates as close to Pilsen, which is true, but it was unable to determine the real location. JackDaw retrieved data by utilizing tools and was able to list existing facilities related to forestry and also determine the distribution of land in the area.

Scenario 2

In this scenario, we asked whether an area in the Czech Republic is suitable for vineyards. The user query was: "Determine if this area is suitable for a vineyard and provide a concise answer." Figure 5 shows both responses together with the area of interest. This time, JackDaw provided a concise answer consisting of climate, land cover, and elevation analysis. The standard LLM wrongly determined the area to be in Austria, with a general recommendation for the area for vineyards without any backing data.

Fig 4. Comparison of responses regarding a query about forestry

Context and Visualisation Imperatives

In the years leading up to generative AI becoming "public property", the volume of text emerging from argumentative sources such as consultancy reports was already expanding noticeably. The increase was often characterised by esoteric industry language and overcomplicated sentence structures, far beyond what was necessary for effective communication. Reports usually offered vague reasoning and inconclusive advice. Consequently, readers, typically non-scientific, technically oriented professionals had to develop new approaches to extract key ideas or recommendations from these wordy passages, which again led to altered reading habits and gave root to the scientifically unjustified and yet not entirely unfounded maxim that "nobody reads".

Following the advent of the Language Model (LLM), the creation of well-structured text is now within reach of literally anyone, leading to a further accelerated growth in well-padded textual content, resulting in a situation where now not only "nobody reads" but neither does anyone write. While this is anecdotal and should not be considered a scientific finding, it does describe a situation where visualisation remains as important as ever to reveal and convey trends and exceptions in data in a manner that makes it more easily accessible to readers who may find a good-looking block of text deceptively attractive without taking into consideration what it actually says.

Effective visualisation in this context means applying the building blocks derived from the more traditional analytical tool suites, such as geoportals (maps, 2D, 3D) and business intelligence (charts, tables), as well as more advanced forms of visualisation.

LLMs generate text and often use the "markdown" language to include formatting. This includes the capability of outputting code blocks that can contain structured data intermixed with the text. A code block may contain any text-based or text-representable format, such as GeoJSON or GeoCSV, making rendering LLM output as any of the above visual indicators easy. JackDaw utilises custom markdown renderers to accomplish render tables, charts, and maps.

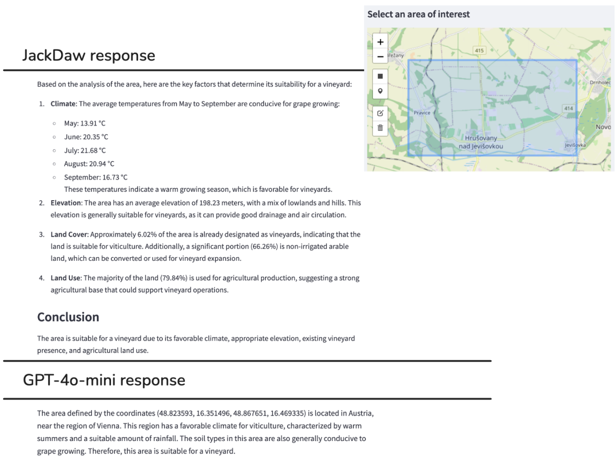

Fig 5. Comparison of responses to an area analysis query related to vineyards

V. Conclusion

This paper has demonstrated that a language-model agent can be made spatially competent by constraining its reasoning to a query-specific subset of geospatial tools and by grounding each answer in an external retrieval layer. The proposed JackDaw architecture contributes three principal advances:

1. Structured tool pre-filtering. Tools are surfaced to the LLM only when their topical, temporal and spatial metadata intersect the user's query, avoiding the accuracy and latency penalties that arise from flat tool lists.

2. Value-engineered model orchestration. A LangChain abstraction allocates lightweight NLP models to syntactic steps, reserves foundation models for broad-context reasoning, and inserts domain-specific kernels where specialised knowledge is required.

3. Retrieval-augmented generation with transparent provenance. A micro-service RAG layer embeds trusted documents in a vector store and returns source passages with every response, enabling fact-checking and reducing hallucinations.

Benchmark experiments on forestry-asset discovery and vineyard-site assessment show that JackDaw delivers concise, location-specific answers, whereas a baseline proprietary LLM produces only generic or spatially misattributed guidance. These findings confirm that coupling global language understanding with local spatial intelligence materially improves decision support in agriculture and rural planning.

Limitations. The evaluation covered two use-cases within Central Europe and a single growing season; cross-regional transferability, additional thematic domains and longer temporal windows remain untested. The current RAG pipeline ingests individual PDF documents but does not yet support user-specific knowledge bases or large-scale cloud retrieval, and cost modelling excluded vector-query overheads.

Future work. Planned extensions include (i) adaptive ontology-driven tool selection to further shrink the action space, (ii) ingestion of multimodal sensor streams (e.g., hyperspectral cubes, IoT telemetry), (iii) private vector databases for institutional users, and (iv) benchmarking against emerging geospatial-LLM test suites. Governance requirements—explainability, bias auditing and energy accounting—will be addressed to ensure regulatory compliance and sustainable operation.

In summary, the results substantiate that an architecture which bridges foundation-scale language models with dynamically selected, high-resolution spatial data streams is a viable path toward reliable, context-aware AI services for agriculture and rural development.

Acknowledgment

Work is also supported by the EU project PoliRuralPlus – Fostering Sustainable, Balanced, Equitable, Place-based and Inclusive Development of Rural-Urban Communities' Using Specific Spatial Enhanced Attractiveness Mapping ToolBox (Grant agreement ID: 101136910).

References

[1] Ryan, M., Isakhanyan, G., & Tekinerdogan, B. (2023). An interdisciplinary approach to artificial intelligence in agriculture. NJAS: Impact in Agricultural and Life Sciences, 95(1), 2168568, https://doi.org/10.1080/27685241.2023.2168568

[2] Ruiz-Real, J. L., Uribe-Toril, J., Torres Arriaza, J. A., & de Pablo Valenciano, J. (2020). A look at the past, present, and future research trends of artificial intelligence in agriculture. Agronomy, 10(11), 1839, https://doi.org/10.3390/agronomy10111839

[3] Bolfe, É. L., Parreiras, T. C., Sano, E. E., Bettiol, G. M., & Vicente, L. E. (2023). Mapping agricultural intensification in the Brazilian Savanna: A machine-learning approach using harmonised Landsat–Sentinel-2 data. ISPRS International Journal of Geo-Information, 12(7), 263, https://doi.org/10.3390/ijgi12070263

[4] Dakir, A., Barramou, F., & Bachir Alami, O. (2023). Towards a Machine Learning-based Model for Automated Crop Type Mapping. International Journal of Advanced Computer Science and Applications (IJACSA), 14(1). https://doi.org/10.14569/IJACSA.2023.0140185

[5] Dhillon, M. S., Dahms, T. C., Kübert-Flock, C., Rummler, T., Arnault, J., Steffan-Dewenter, I., & Ullmann, T. (2023). Integrating random forest and crop modelling improves the yield prediction of winter wheat and oil-seed rape. Frontiers in Remote Sensing, 3, 1010978. https://doi.org/10.3389/frsen.2022.1010978

[6] Gumma, M. K., Nukala, R. M., Panjala, P., Bellam, P. K., Gajjala, S., Dubey, S. K., & Deevi, K. C. (2024). Optimizing Crop Yield Estimation through Geospatial Technology: A Comparative Analysis of a Semi-Physical Model, Crop Simulation, and Machine Learning Algorithms. *AgriEngineering*, 6(1), 786–802, https://doi.org/10.3390/agriengineering6010045

[7] Qi, H., Wang, L., Zhu, H., Gani, A., & Gong, C. (2023). The barren plateaus of quantum neural networks: review, taxonomy and trends. *Quantum Information Processing*, 22(12), 435. https://doi.org/10.1007/s11128-023-04188-7

[8] Farmonov, N., Esmaeili, M., Abbasi-Moghadam, D., Sharifi, A., Amankulova, K., & Mucsi, L. (2024). HypsLiDNet: 3D-2D CNN model and spatial-spectral morphological attention for crop classification with DESIS and LiDAR data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. https://ieeexplore.ieee.org/document/10571369, DOI: 10.1109/JSTARS.2024.3418854

[9] Miranda, M., Pathak, D., Nuske, M., & Dengel, A. (2024, July). Multi-modal fusion methods with local neighborhood information for crop yield prediction at field and subfield levels. In *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium* (pp. 4307–4311). IEEE. https://doi.org/10.1109/IGARSS53475.2024.10640993

[10] Pavlova M.A., Sidorchuk D.S., Bocharov D.A. Classification of crops by NDVI time series of reduced dimensionality // Sensornye sistemy. - 2023. - Vol. 37. - N. 2. - P. 306-312. doi: 10.31857/S023500922302004X

[11] Toro, A. P., Bueno, I. T., Werner, J. P., Antunes, J. F., Lamparelli, R. A., Coutinho, A. C., ... & Figueiredo, G. K. (2023). SAR and optical data applied to the early-season mapping of integrated crop-livestock systems using deep and machine learning algorithms. Remote Sensing, 15(4), 1130. DOI: https://doi.org/10.3390/rs15041130

[12] Pihrt, J., Šimánek, P., Kovalenko, A., Kvapil, J., & Charvát, K. (2024). AI-Based Spatiotemporal Crop Monitoring by Cloud Removal in Satellite Images. In Proceedings of the 19th Conference on Computer Science and Intelligence Systems (FedCSIS), Annals of Computer Science and Information Systems, Volume 39, pp. 485–492, DOI: 10.15439/2024F5446

[13] F. Koutenský, J. Pihrt, M. Čepek, V. Rybář, P. Šimánek, M. Kepka, K. Jedlička, and K. Charvát, "Combining Local and Global Weather Data to Improve Forecast Accuracy for Agriculture," in *Communication Papers of the 19th Conference on Computer Science and Intelligence Systems (FedCSIS)*, 2024, pp. 77–82. DOI:10.15439/2024F5990

[14] Filippi, P., Jones, E., Bishop, T., Acharige, N., Dewage, S., Johnson, L., Ugbaje, S., Jephcott, T., Paterson, S., & Whelan, B. (2017, October). A Big Data Approach to Predicting Crop Yield. In *Proceedings of the 7th Asian-Australasian Conference on Precision Agriculture* (pp. 16–18). Hamilton, New Zealand. Zenodo. https://doi.org/10.5281/zenodo.893668

[15] M. Hosseini, I. Becker-Reshef, R. Sahajpal, L. Fontana, P. Lafluf, G. Leale, and C. Justice, "Crop Yield Prediction Using Integration of Polarimetric Synthetic Aperture Radar and Optical Data," in *Proceedings of the 2020 IEEE India Geoscience and Remote Sensing Symposium (InGARSS)*, 2020, pp. 17–20. DOI: 10.1109/InGARSS48198.2020.9358978

[16] Wang, S., Hu, T., Huang, X., Li, Y., Zhang, C., Ning, H., Zhu, R., Li, Z., & Ye, X. (2024). GPT, large language models (LLMs) and generative artificial intelligence (GAI) models in geospatial science: a systematic review. *International Journal of Digital Earth*, 17(1), Article 2353122.LinkedIn+2University of Bristol+2University of Bristol+2, DOI: 10.1080/17538947.2024.2353122

[17] Sapkota, R., Qureshi, R., Hassan, S. Z., Shutske, J., Shoman, M., & Sajjad, M. (2024). *Multi-Modal LLMs in Agriculture: A Comprehensive Review*. TechRxiv Preprint. DOI: 10.36227/techrxiv.172651082.24507804/v1

[18] Mansourian, A., & Oucheikh, R. (2024). *ChatGeoAI: Enabling Geospatial Analysis for Public through Natural Language, with Large Language Models*. ISPRS International Journal of Geo-Information, 13(10), 348.Lund University, DOI: 10.3390/ijgi13100348

[19] L. Yuan, F. Mo, K. Huang, W. Wang, W. Zhai, X. Zhu, Y. Li, J. Xu, and J.-Y. Nie, "OmniGeo: Towards a Multimodal Large Language Models for Geospatial Artificial Intelligence," *arXiv preprint arXiv:2503.16326*, Mar. 2025. https://arxiv.org/html/2503.16326v1

[20] L. Wang, J. Chou, A. Tien, X. Zhou, and D. M. Baumgartner, "AviationGPT: A Large Language Model for the Aviation Domain," in *Proceedings of the AIAA AVIATION Forum and ASCEND 2024*, 2024, p. 4250. DOI: 10.2514/6.2024-4250

[21] Ali. S. H, Shahid. M. F, Tanveer. M. H, Rauf. A, "Integrating LLM for Cotton SoilAnalysis in Smart Agriculture System", IJIST, Special Issue pp 283-294, (2024) *(PDF) Integrating LLM for Cotton Soil Analysis in Smart Agriculture System*. Available from: https://www.researchgate.net/publication/389436000_Integrating_LLM_for_Cotton_Soil_Analysis_in_Smart_Agriculture_System [accessed May 23 2025].

[22] Peng, B., Li, C., He, P., Galley, M., & Gao, J. (2023). Instruction tuning with GPT-4. *arXiv preprint arXiv:2304.03277*. https://doi.org/10.48550/arXiv.2304.03277

[23] Auer, C., Lysak, M., Nassar, A., Dolfi, M., Livathinos, N., Vagenas, P., Berrospi Ramis, C., Omenetti, M., Lindlbauer, F., Dinkla, K., Mishra, L., Kim, Y., Gupta, S., Teixeira de Lima, R., Weber, V., Morin, L., Meijer, I., Kuropiatnyk, V., & Staar, P. W. J. (2024). *Docling Technical Report* (Version 1.0.0). arXiv preprint arXiv:2408.09869. https://doi.org/10.48550/arXiv.2408.09869

[24] Li, Z., Zhang, X., Zhang, Y., Long, D., Xie, P., & Zhang, M. (2023). Towards general text embeddings with multi-stage contrastive learning. *arXiv preprint arXiv:2308.03281*. https://doi.org/10.48550/arXiv.2308.03281