AI-Driven Invoice Processing as a Foundation for Smart Sustainability Reporting

Eugene Westerhof, Hans Vrolijk, Robbert Robbemond, Cor Verdouw
Wageningen Social & Economic Research, December 2025
KB-project 'AI for Future Food Systems' (KB-53-000-013), D2.1

The 'AI for Future Food Systems' knowledge base project at WUR investigates the potential and challenges of AI-enabled research infrastructures for shaping pathways towards Future Food Systems. The research focuses on three themes:

A. AI for automation of the research data chain
B. AI for developing, testing and using data-driven research models
C. AI for unlocking research knowledge in natural language.

This paper presents findings from a use case project within the first theme, including two Proof of Concepts (PoCs). Specifically, it explores the use of AI to automate the data chain for farm sustainability monitoring. The study concentrates on the initial stage of this chain: the automated extraction and processing of sustainability accounting data from invoices for the Dutch Farm Sustainability Data Network (FSDN).

Problem Context

Wageningen Social & Economic Research has a decades-long tradition of collecting and curating financial-economic data from agricultural, horticultural, forestry and fisheries enterprises. These data serve as the empirical backbone for European policy evaluation, enabling ex-ante and ex-post analyses of sectoral performance, structural trends, and regulatory impacts. As societal expectations shift toward integrated assessments of economic, environmental and social sustainability, Wageningen Social & Economic Research confronts a widening **gap between the growing informational needs of policymakers and the practical limitations of manual or semi-automated data processing systems**. The expansion of economic bookkeeping into comprehensive environmental and sustainability reporting demands data of far greater granularity and timeliness than existing workflows were designed to deliver. At the same time, it requires the methodological rigor, transparency and quality assurance associated with economic statistics, particularly when sustainability indicators are linked to financial incentives or regulatory compliance.

The **current reliance on manual invoice processing within the Dutch Farm Sustainability Data Network (FSDN)** exemplifies the tension between increasing informational demands and constrained operational capacity. The Dutch FSDN processes hundreds of thousands of invoices every year, many of which contain precisely the kinds of financial, material and contextual data required for both economic and environmental accounting. Yet the **classification, interpretation and extraction of relevant attributes remain labour-intensive, time-consuming and costly**. Traditional Optical Character Recognition (OCR) tools have proven inadequate, especially when confronted with the heterogeneity of invoice formats and the variability of language used in agricultural trade. The resulting bottlenecks impose structural limits on the scalability and responsiveness of sustainability monitoring frameworks such as FSDN. They also generate **administrative burdens for farmers** and other business operators, undermining the political and societal imperative to reduce compliance costs while simultaneously increasing data availability.

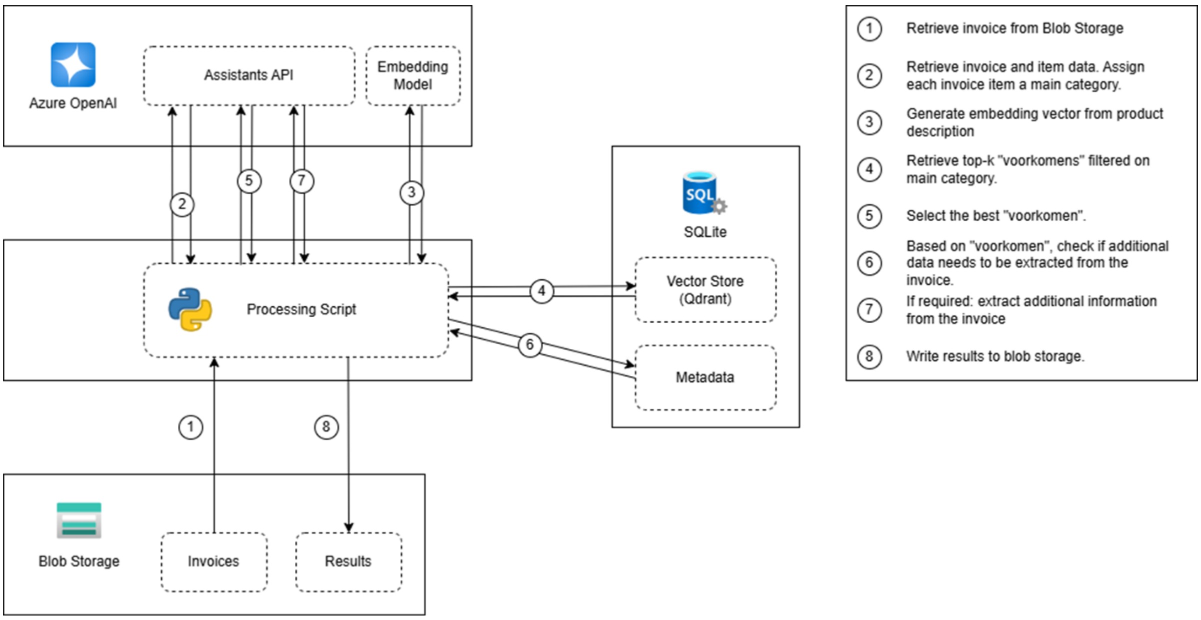

Figure S.1 *Technical architecture of the Proof of Concepts for AI-Driven Invoice Processing*

In this context, Wageningen Social & Economic Research initiated a use-case project that carried out two Proofs of Concept (PoCs): the first focused on AI-driven invoice data extraction, and the second extended this work to an AI-driven invoice processing chain. Figure S.1 visualizes the technical architecture of the PoCs, that are introduced below.

1st Proof of Concept: AI-Driven Invoice Data Extraction

The first PoC has examined the potential of artificial intelligence, document intelligence and semantic search technologies to automate and enhance invoice data extraction. The Proof of Concept demonstrated conclusively that modern Document Intelligence systems can **reliably extract structured information from unstructured invoices**, thereby replacing a substantial portion of manual data entry. It further revealed that Large Language Models (LLMs) can normalise product descriptions, inferring contextual meaning and resolving linguistic variation, tasks historically dependent on expert domain knowledge. The integration of vector search through Qdrant expanded the semantic capabilities of the system, allowing for **accurate classification even in the absence of article codes or standardised descriptions**. Together, these technologies exhibited a maturity and robustness that surpasses earlier generations of OCR-based automation, signalling the viability of a new class of AI-enabled data processing pipelines.

The empirical outcomes of the first Proof of Concept are based on **hands-on experimentation with real-world farm invoices from the Dutch FSDN context**, encompassing heterogeneous invoice layouts, suppliers, and product descriptions. Initial exploratory testing was performed on several dozen invoices comprising hundreds of invoice lines to evaluate end-to-end functionality and identify methodological bottlenecks. In addition, **a large-scale empirical validation was conducted on a set of 10,425 invoices**, which were automatically processed by the system and systematically compared with the same invoices after manual entry and validation by data collectors. Document Intelligence demonstrated high reliability in extracting structural invoice elements and invoice line data, enabling consistent downstream processing at scale. At the same time, the evaluation confirmed that highly granular numerical extraction at the invoice line level remains more challenging and requires further methodological refinement. From an operational perspective, end-to-end processing times ranged between approximately 2 and 4 seconds per invoice, depending on invoice complexity and the need for AI-based interpretation, while average variable costs were estimated at approximately €0.06 per invoice, comprising document ingestion (~€0.03), data extraction output (~€0.02), and semantic categorisation (~€0.01). According to a conservative expert estimate, this is at least 30 times faster and cheaper than manual invoice processing, not counting the training time for new staff. Ongoing runtime costs of the AI components were limited and predictable, with Azure OpenAI services incurring approximately €0.03 per hour when active.

Table 1 summarises the principal findings, which collectively illustrate that the technological foundations are sufficiently mature to justify a second, more operationally oriented PoC, while also highlighting the specific components that require targeted methodological improvement.

Table 1 *Empirical outcomes of the first Proof of Concept*

Metric	Result	Interpretation
Accuracy of recognising product categories	Approximately 80% correct	Demonstrates that the combined LLM and vector search approach can classify most invoice lines reliably despite heterogeneous linguistic descriptions.
Accuracy of price extraction at line-item level	Approximately 40% correct; 60% of values deviate by more than €1	Indicates that detailed numerical extraction remains challenging and demands improved OCR, refined domain constraints and additional model tuning.
Document Intelligence (DI) field extraction (seller, buyer, totals, invoice ID, VAT lines and related attributes)	High reliability across tested invoices	Confirms that DI can replace large portions of manual data entry through consistent structured extraction.
Estimated cost per processed invoice	Approximately €0.06	Suggests that automated invoice processing is economically viable, especially when hybrid classification techniques reduce unnecessary AI calls.
Cloud-based operational cost of AI inference	Roughly €0.03 per hour	Illustrates predictability and manageability of running costs under controlled workloads.

Organisational and legal lessons learnt

Beyond the empirical performance, the PoC also served as a **practical education in operational AI deployment**. Wageningen Social & Economic Research gained hands-on experience with orchestrating a complete invoice-processing pipeline using Azure OpenAI services, learning how to design workflows that route documents through extraction, normalisation, classification and storage phases. The project confirmed that **all processing can be conducted entirely within European data centres**, thereby meeting GDPR requirements and ensuring that confidential farm and financial data do not leave the

European Union. Although the CLOUD Act introduces a theoretical residual risk due to the U.S. jurisdiction of the cloud provider, this risk is mitigated by the fact that no persistent data storage occurs within the AI services themselves and that processing is strictly transient. Internal access is governed by role-based authorisation, which provides an additional layer of control and auditability. Together, these safeguards created confidence **that AI-driven invoice processing can meet Wageningen Social & Economic Research's stringent privacy, security and compliance requirements**.

2nd Proof of Concept: AI-Driven Invoice Processing Chain

The promising results of the initial PoC provided justification for a second, more ambitious study focused not merely on evaluating technological feasibility but on understanding the operational implications, methodological choices and long-term governance requirements of an AI-enabled processing ecosystem. This second PoC seeks to determine **how the entire data processing chain (extraction, normalisation, categorisation and enrichment) can function cohesively** within a production environment. It aims to evaluate the practical trade-offs between different architectural strategies, to assess the reliability and acceptance of AI-generated classifications, and to test whether the system can support the rigorous standards of transparency and traceability expected in official data infrastructures.

Central to the second PoC is the **comparison between two alternative approaches: an AI-first approach and a hybrid approach** that combines deterministic rules with AI and vector search.

The ***AI-first approach*** entrusts the interpretation and categorisation of invoice lines almost entirely to LLMs and semantic search. In this framework, Document Intelligence provides a structural parse of the invoice, after which the LLM handles normalisation, disambiguation and classification, drawing on Qdrant to retrieve semantically relevant product prototypes. This approach is **elegant in its simplicity and highly adaptable to new or unusual cases**. Because it does not rely heavily on hand-crafted rules or tables, it scales efficiently and learns primarily through the accumulation of examples in the vector database. However, its reliance on probabilistic reasoning raises **legitimate concerns regarding transparency, auditability and cost**. Each classification involves an AI inference step, making operational expenditures sensitive to invoice volume. Moreover, the reasoning processes of LLMs, while increasingly structured, remain less interpretable than deterministic decision pathways.

The ***hybrid approach*** addresses these concerns by combining rule-based mechanisms and semantic AI in a layered architecture. In this system, known and recurring product variants are normalized through deterministic SQL tables, article codes provide direct classification cues, supplier–product relationships contribute context, and fixed attribute tables supply domain-specific information such as fat percentages or packaging units. These rules handle a substantial portion of **predictable cases without invoking AI**, thereby reducing costs, accelerating processing and ensuring full transparency. **AI is reserved for genuinely ambiguous or novel cases**, where deterministic approaches fail or where linguistic variation exceeds the expressive capacity of rule-based systems. In such scenarios, Qdrant identifies semantically similar product records, and the LLM synthesises contextual signals to determine the most plausible classification.

The **hybrid system offers clear advantages in domains where explainability and consistency are critical**. Sustainability indicators derived from invoice data may influence public policy, subsidy allocation and regulatory enforcement. Under these conditions, decisions must be traceable and subject to verification. Deterministic rules offer exactly this form of epistemic clarity. Nevertheless, hybrid systems introduce their **own challenges, most notably in the governance of the knowledge base**. Rules must be curated, revised and expanded as new product variants emerge. Qdrant must accommodate new exemplars, and decision criteria must be designed to determine when a human expert should intervene. Without careful governance, the system risks becoming unwieldy, with overlapping rules, inconsistent exemplars and unclear decision boundaries. The hybrid approach therefore requires not only technical sophistication but also institutional discipline and long-term stewardship.

The **second PoC** operationalizes these considerations by formulating explicit evaluation criteria. It seeks a minimum of ninety percent accuracy in document extraction and eighty-five percent accuracy in line-item categorization, with clear differentiation between cases handled deterministically and those handled by AI. It imposes strict constraints on processing time to ensure system scalability, mandates at least a fifty percent reduction in manual processing effort, and requires that human corrections be systematically incorporated into the knowledge base. The PoC also emphasizes the acceptance of domain experts, recognizing that technological success alone is insufficient if the system cannot earn the trust of those responsible for data quality.

Conclusions

The two PoCs chart a coherent pathway toward **smart sustainability reporting**: a data ecosystem in which financial, management, technical and environmental information can be extracted, interpreted and integrated with minimal administrative burden and maximal analytical value. The integration of AI into invoice processing is not merely a technological upgrade—it is a **strategic transformation** that supports the next generation of sustainability metrics required by European policy frameworks, life cycle assessment methodologies, true cost accounting and farm-level sustainability benchmarking. By **reducing manual labour, accelerating data availability and enhancing granularity**, AI lowers barriers to participation for agricultural enterprises, thereby improving the representativeness and timeliness of sustainability data. At the same time, the **hybrid architecture preserves auditability, transparency and quality**, which are cornerstones of economic statistics.

In this sense, the adoption of AI-driven invoice processing becomes more than a response to operational inefficiencies. It is a **deliberate strategic choice** that aligns sustainability reporting with global developments in digitalisation and data science, and that helps sustainability reporting institutes to **reinforce their role as a trusted providers of high-quality statistical information**. The PoC trajectory not only demonstrates the capability of leveraging advanced AI technologies but also highlights a strong starting position for integrating them into robust, explainable and adaptive systems that meet the demands of both policymakers and producers. In this context, **combining AI intelligence with institutional safeguards will be essential** as sustainability reporting in the context of the FSDN moves toward real-time, high-frequency and multi-dimensional data streams. The work undertaken in these PoCs establishes the conceptual and technological foundation for such a system and signals **the readiness to take a leading role in shaping the future of sustainability data infrastructures**.

More information
Eugene Westerhof
T +31 (0)70 335 82 04
E eugene.westerhof@wur.nl
www.wur.eu/social-and-economic-research

2026-019