Estruturação de dados e inteligência artificial generativa na produção de feijão: impactos socioambientais na tomada de decisões agrícolas

Data structuring and generative artificial intelligence in bean production: socio-environmental impacts on agricultural decision-making

Herica Fernanda de Sousa Carvalho
Centro de Tecnologia da Informação Renato Archer (CTI), Campinas/SP, Brasil
Robson Carlos Bosse
Centro Universitário Campo Limpo Paulista (FACCAMP), CL. Paulista/SP, Brasil
Ferrucio de Franco Rosa
Centro de Tecnologia da Informação Renato Archer (CTI), Campinas/SP, Brasil
Luiza Maria Capanema Bezerra
Instituto Agronômico, SAA-SP, Brasil

GT 12 - Impactos socioambientais das novas tecnologias no agronegócio

Resumo: O estudo analisa o uso de tecnologias digitais, especialmente a estruturação de dados e a Inteligência Artificial generativa, na geração de impactos socioambientais na cadeia produtiva do feijão no Brasil. A proposta integra modelagem conceitual, banco de dados relacional, dashboards interativos e modelos de linguagem para qualificar a tomada de decisão agronômica. A organização de dados do RNC/MAPA reduziu inconsistências e aumentou a confiabilidade das análises. Os resultados indicam ganhos na identificação de padrões produtivos e adaptação regional contribuindo para a eficiência produtiva, a segurança alimentar e a sustentabilidade.
Palavras-chave: inteligência artificial generativa, tecnologias digitais, modelagem de dados, impactos socioambientais, cadeia produtiva do feijão.

Abstract: This study analyzes the use of digital technologies, particularly data structuring and generative artificial intelligence, to generate socio-environmental impacts within the Brazilian bean production chain. Our proposal integrates conceptual data modeling, a relational database, interactive dashboards, and large language models to support agronomic decision-making. Structuring data from the Brazilian National Cultivar Registry reduced inconsistencies and improved analytical reliability. Results show gains in identifying productive patterns and regional adaptation, contributing to productive efficiency, food security, and sustainability in agricultural systems.
Keywords: generative artificial intelligence; digital technologies; data modeling; socio-environmental impacts; production chain.

1. Introduction

Digital transformation in agriculture has expanded the strategic use of data and intelligent decision-support systems, fostering new ways of organizing, analyzing, and applying information in the sector (Sharma et al., 2020). However, the adoption of these technologies remains concentrated in more capitalized production chains, revealing structural inequalities in access to digital innovation (Rotz et al., 2021).

In this context, crops of high socio-economic relevance, such as the common bean (Phaseolus vulgaris), remain relatively underexplored regarding the use of digital technologies for data analysis and decision support, despite their importance for food security and family farming, especially in developing countries (Petry et al., 2021; Oliveira & Wander, 2023).

Part of this limitation is associated with the low availability of structured and standardized data. Information on cultivars registered in the Brazilian National Cultivar Registry (RNC) (Brasil, 2025), is fragmented, poorly standardized, and difficult to analyze, restricting its use in evidence-based decision-making.

In response, we propose developing a digital infrastructure, AgriBeans, that integrates conceptual data modeling, relational databases, interactive dashboards, and generative artificial intelligence (GenAI).

The main objective is to analyze how data structuring and GenAI can improve agronomic decision-making while generating socio-economic and environmental impacts in the bean production chain, contributing to the Sustainable Development Goals (SDGs 1, 2, and 13).

2. Methodology

The proposed methodology was structured into integrated stages of data collection, processing, modeling, and analysis, as shown in Figure 1.

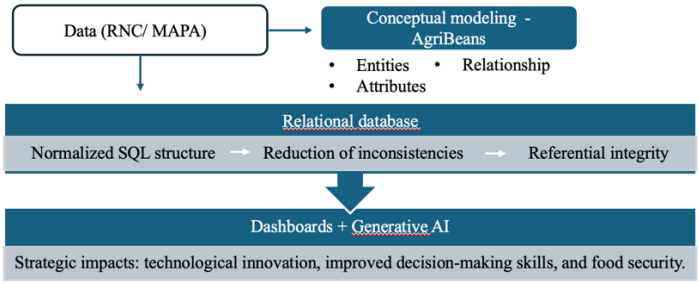

Figure 1: AgriBeans system workflow

Initially, we used data from RNC (Brasil, 2025), updated up to May 2025, comprising 348 records and 10 variables. This dataset was subsequently expanded to 33 attributes by incorporating information from scientific literature and research institutions, including the Agronomic Institute of Campinas (IAC), the Brazilian Agricultural Research Corporation (Embrapa), the Rural Development Institute of Paraná (IDR-Paraná), and the Agronomic Institute of Pernambuco (IPA). Data standardization and inconsistency correction procedures were performed at this stage.

Next, conceptual data modeling was developed using the Entity-Relationship (ER) model, with the cultivar as the central entity, connected to morphological, productive, phytopathological, and geographical adaptation attributes. The database was implemented in a well-known database management system (MySQL), ensuring referential integrity, normalization, and scalability. Missing data were treated using the MissForest algorithm (Random Forest), preserving phenotypic diversity and ensuring analytical consistency.

The analytical stage included the development of interactive dashboards (Power BI) and experiments with large language models (GPT-5 and Gemini 2.5 Flash) to assess how structured data influence response quality. Two scenarios were compared: (i) without structured data and (ii) with access to the structured AgriBeans database.

3. Results and Discussion

Data structuring and standardization within the AgriBeans system enabled us to develop an integrated database capable of organizing and relating morpho-agronomic and phytopathological information on bean cultivars. This structure supports analytical intelligence, enabling pattern identification and the generation of strategic insights for decision-making.

Interactive dashboards allow the visualization of approximately 340 cultivars across 15 bean species registered in Brazil, highlighting productive, morphological, and regional adaptation patterns (Carvalho et al., 2025).

Experiments with large language models demonstrated that structured data considerably improves the response quality. While models operating without structured context produced generic and imprecise outputs, those using the AgriBeans database generated more accurate, consistent, and analytically robust responses. For instance, structured queries enabled precise outputs such as the cultivar growth-habit distribution and the flower-color frequencies.

From a socio-environmental perspective, the technology contributes across multiple dimensions. Economically, it supports the identification of high-performing regionally adapted cultivars, improving production efficiency. Socially, it enhances food security and broadens access to qualified information for different actors of the production chain. Environmentally, it supports more sustainable decisions by enabling the selection of cultivars better adapted to local agroecological conditions and reducing production risks.

4. Final Remarks

The AgriBeans system demonstrates that integrating conceptual data modeling, relational databases, interactive dashboards, and GenAI creates a strategic digital infrastructure for strengthening the bean production chain.

The proposed approach integrates heterogeneous data into a unified analytical structure, enabling their effective use in AI-based systems. By enhancing data organization, transparency, and analytical capacity, the system supports more efficient and sustainable decision-making.

Beyond a technological tool, AgriBeans acts as a mechanism for democratizing agronomic knowledge and promoting the responsible use of Artificial Intelligence in

agriculture. The study highlights the potential of digital technologies to generate positive socio-environmental impacts and reinforces the role of innovation in building resilient, knowledge-based agri-food systems.

Acknowledgment

This study was funded by the São Paulo Research Foundation (FAPESP), grant no. 2020/01721-7. We also thank CNPq (National Council for Scientific and Technological Development) for the technical support grant to the first author, grant 300573/2026-0.

References

BRASIL, Ministério de Agricultura e Pecuária - MAPA, **Registro Nacional de Cultivares - RNC, Cultivares ou espécies registradas,** https://www.gov.br/agricultura/pt-br/assuntos/insumos-agropecuarios/insumos-agricolas/sementes-e-mudas/registro-nacional-decultivares/cultivares-ou-especies-registradas, Accessed: may, 2025.

CARVALHO, H. F. S. et al. **Dashboard operacional para construção de conhecimento sobre variedades de feijão.** In: Congresso da sociedade brasileira de economia, administração e sociologia rural (sober), 63., 2025, Passo Fundo. *Anais [...]* Passo Fundo: Universidade de Passo Fundo, 2025. Disponível em: https://www.even3.com.br/anais/sober2025/1103233-DASHBOARD-OPERACIONAL-PARA-CONSTRUCAO-DE-CONHECIMENTO-SOBRE-VARIEDADES-DE-FEIJAO. Acesso em: 6 mar. 2026.

OLIVEIRA, G. M.; WANDER, A. E. **Mapeamento da cadeia produtiva do feijão comum no Brasil.** *Revista Economia Política do Desenvolvimento,* Maceió, v. 14, n. 32, p. 96 - 22, jul.-dez. 2023.

PETRY, N.; BOY, E.; WIRTH, J. P.; HURRELL, R. **The potential of the common bean (*Phaseolus vulgaris*) as a vehicle for iron biofortification.** *Nutrients,* v. 7, n. 2, p. 1144 -1173, 2015.

ROTZ, S. et al. **The politics of digital agricultural technologies: a preliminary review**. *Sociologia Ruralis,* v. 59, n. 2, p. 203 - 229, 2019.

SHARMA, A. et al. **Machine learning applications for precision agriculture: a comprehensive review**. *IEEE Access,* v. 9, p. 4843 - 4873, 2020.