Causal Inference in the Innovation Lifecycle of AI Technologies for Agriculture

Giani Gradinaru[1,†], Madalina Zurini[1,*,†], Bogdan Florin Matei[1,*,†] and Diana-Ioana Petre[1,†]

[1] *Bucharest University of Economic Studies, Bucharest, Romania*

Abstract

The main objective of the current research is to create a solid time-series database in order to aim the time impact of five impact factors within the lifecycle of AI in agriculture. The study verifies a potential causality in this innovation lifecycle, where advancements in one pillar drive progress in the subsequent pillar in the following order: scientific research, projects, patents, start-ups and public acceptance.

The first research question that we can derive checks if a causality can be determined from comparing the 5 directions or the timeline sequence can be interpreted just as a coincidence. The resulted hypotheses imply that (H1) there is no sequential causality in the innovation lifecycle of AI in agriculture; (H2) increased scientific research output causes an increase in related projects; (H3) an increase in related projects causes an increase in patent filings; (H4) increase in patent filings causes an increase in start-up formation; (H5) an increase in start-up formation causes greater public acceptance and interest in AI technologies in agriculture.

The ultimate aim of the research consists of creating structured time series for each direction with comparable dimensions.

As the research horizon is defined by the two main areas, namely agriculture and artificial intelligence, clear borders had to be set for both to be able to reduce any biases to a minimum and ensure comparable datasets across pillars. As a result, the scope has been narrowed down to the most significant keywords. Multiple simulations are made and an objective function is desired to evaluate the relevance of the searching criteria.

Nonetheless, through this abstract, we aim to explore the potential of our research and discuss methodological approaches for developing high-quality outputs suitable for publication in top-tier journals.

Keywords

Innovation lifecycle, causality, artificial intelligence, agriculture

1. Introduction

The current research is established within CauseFinder project, financed under the Romanian National Recovery and Resilience Plan, by the Romanian Government, entitled “CAUSEFINDER-CAUSALITY IN THE ERA OF BIG DATA”.

The main objectives of CauseFinder project reflect in the following directions:

- Innovation Knowledge Graph Construction
- Causality Modelling

RuleML+RR'24: Companion Proceedings of the 8th International Joint Conference on Rules and Reasoning, September 16--22, 2024, Bucharest, Romania

* Corresponding author.
† These authors contributed equally.
giani.gradinaru@csie.ase.ro (G. Gradinaru); madalina.zurini@csie.ase.ro (M. Zurini); florin.matei@csie.ase.ro (B.F. Matei); diana.petre@csie.ase.ro (D.I. Petre)
0000-0003-3336-1737 (G. Gradinaru); 0000-0001-9872-8566 (M. Zurini); 0000-0003-2598-0220 (B.F. Matei); 0009-0009-9893-586X (D.I. Petre)

- Scalable Probabilistic Reasoning
- Representation Learning for Anomaly Detection
- Validation through Innovation Ecosystem Use Cases.

The inception of the project relies on building and proposing the elements for describing an innovative ecosystem.

In order to reach the desired outcomes of the CauseFinder project, this current proposed research is used as a case study of a micro part of the big designed ecosystem. The area of Artificial Intelligence is of an exponential increase in the domain of agriculture. Narrowing the causality analysis for this sub domain of interest is having a great impact in obtaining answers to the overmentioned hypothesis. The main steps designed for this research are consisted in the next chapters of this paper. The second chapter is describing the main databases used in order to best describe the pillars of innovation. Those databases are integrated within a methodology that is described in chapter 3. Following these, chapter 4 contains some directions of business scenarios that can derive using the proposed methodology. Conclusions are withdrawn in the last chapter.

2. Databases

Starting from the innovation lifecycle steps defined by the existing literature, the present study focuses on the first five: academic world, projects started around innovation, possible patents resulting from the innovation, companies established around it, public perception. Table 1 describes the data sources for each pillar.

Table 1
Source of databases for each involved pillar

Pillar	**Source**	**Dimensions**
Academia	Web of Science, [2]	Years, Authors, Countries, Affiliation
Projects	Cordis (Horizon 2020 and Horizon Europe), [3]	Years, Authors, Countries, Affiliation
Patents	WIPO, [4]	Years, Authors, Countries, Affiliation
Companies	CrunchBase, [1]	Years, Countries, Companies, Investment plan
Public perception	Google Trends, [5]	Years, Countries

The five main sources used for building these time series are given by Web of Science, Cordis, WIPO, CrunchBase and Google Trends.

Web of Science is used for extracting relevant data referring to scientific articles written in the topics of artificial intelligence in agriculture. By using advanced search criteria, the data extracted includes publication dates, citation counts, and journal impact factors, which are crucial for understanding the evolution and influence of scientific research in this domain. The robust indexing system of Web of Science ensures that only high-quality, peer-reviewed articles are included, providing a solid foundation for assessing academic progress and trends in AI applications in agriculture.

Cordis is added for gathering data across different calls and programmes about projects written starting from the core of artificial intelligence in agriculture. Cordis is used to identify and analyze the progression and outcomes of projects that focus on AI in agriculture. This includes detailed project descriptions, funding amounts, project durations, consortium members, and final reports. The data from Cordis helps in understanding how EU policies and funding impact AI innovation cycles in the agricultural sector.

WIPO is the international database of patents used for extracting the number of patents regarding AI in agriculture. By analyzing patent filings, one can gauge the technological advancements and the regions or entities that are leading in innovation. The database offers insights into the type of AI technologies being patented, the rate of patent filings over time, and the collaborative networks forming around these innovations.

CrunchBase is used for data about companies that were formed under the main description of AI in agriculture. It provides information on company funding rounds, investor activities, mergers and acquisitions, and leadership changes. For this research, CrunchBase is used in tracking the emergence and growth of startups that are specifically categorized under AI applications in agriculture. This data helps in understanding market dynamics, investment trends, and the overall health and vibrancy of the entrepreneurial ecosystem in this field.

Google Trends is introduced as the fifth dimension of the desired data from the innovation lifecycle for picking public impact of the aforementioned innovation ideas. This tool helps in identifying peaks in interest which may correlate with major product releases, policy changes, or other significant events. The temporal and geographical data provided by Google Trends offer insights into how and where AI in agriculture is gaining traction.

3. Methodology

The research follows a structured approach to address the previously formulated research questions. The initial step involves defining a theoretical model of the innovation lifecycle, drawing from both scientific literature and business practices. Each key element will be incrementally added to test multiple use cases and identify patterns. The research then progresses to refining the search query for optimal accuracy. As the research horizon is defined by the two main areas, namely agriculture and artificial intelligence, clear borders had to be set for both to be able to reduce any biases to a minimum and ensure comparable datasets across pillars. As a result, the scope has been narrowed down to the most significant keywords. Multiple simulations are made and an objective function is desired to evaluate the relevance of the searching criteria. In the absence of a method that can include absolutely all the values to fully satisfy the applicability of AI in agriculture, a classification of the extracted data was considered based on the most comprehensive search query, which is, however, limited. Besides the general direction of application of artificial intelligence in agriculture evaluated within the five pillars of innovation, a more detailed investigation is also taken into consideration by a more granulated classification on 12 sub-domains of agriculture where artificial intelligence is present. Using statistical methods, a comparison is made between the overall trend of AI in agriculture and the trend of AI for each from the 12 sub-domains proposed. Therefore, the next step involved comparing the means of the resulting series to determine if there are significant differences. If no such differences are found, it can be established that the definition of the search query does not significantly influence the temporal evolution of the analysed dimensions, thus negating the need to identify the most accurate query.

In the continuation of the study, we focused on defining the data structure by selecting the dimensions necessary for creating time series, particularly the temporal dimension and the measure under analysis. This resulted in a series of frequencies representing the number of articles, projects, patents, startups, and the annual average of popularity scores. The resulting time series, corresponding to each stage of the innovation lifecycle, were smoothed and min max normalized to mitigate the cyclic effects caused by external events, or, in the case of projects, the conclusion of one program and the initiation of another, as well as to make the series comparable. The next step involved determining the knee points, from which each series experiences a sudden change, to gain an initial understanding of the order of processes in the innovation cycle and whether it aligns with the formulated hypotheses. The second derivative approach was used to identify the points of maximum curvature. The knee point is where the curve changes from a gradual to a steep slope, indicating exponential growth.

To explore the causal relationships among the different stages of the innovation lifecycle - namely, articles, projects, patents, popularity, and startups - Granger causality analysis was employed. This method was used to determine whether one time series could predict another, indicating potential causal influences among the stages. The analysis was conducted for each pairwise combination of the five variables. For instance, we tested whether the number of articles Granger-causes the number of projects, and vice versa, to understand the directional influence between research publications and the initiation of new projects. Similarly, the relationship between patents and startups was examined to identify whether patent activity precedes the emergence of new startups, suggesting a causal link between intellectual property development and entrepreneurial ventures. Furthermore, the role of popularity, as measured by annual average scores, was verified in relation to each of the other four variables to assess whether shifts in public interest or media coverage can predict changes in research output, project initiation, patent filings, or startup formation. Each Granger causality test was conducted using lagged values appropriate for the temporal dynamics of the data, ensuring that the tests accounted for potential time delays between the stages.

Alternatively, the cross-lagged panel analysis was performed to understand the dynamic interactions between the stages, the Johansen cointegration test to evaluate the null hypothesis of no cointegration against the alternative hypothesis of cointegration, respectively the cross-correlation analysis to identify the lag between the time series.

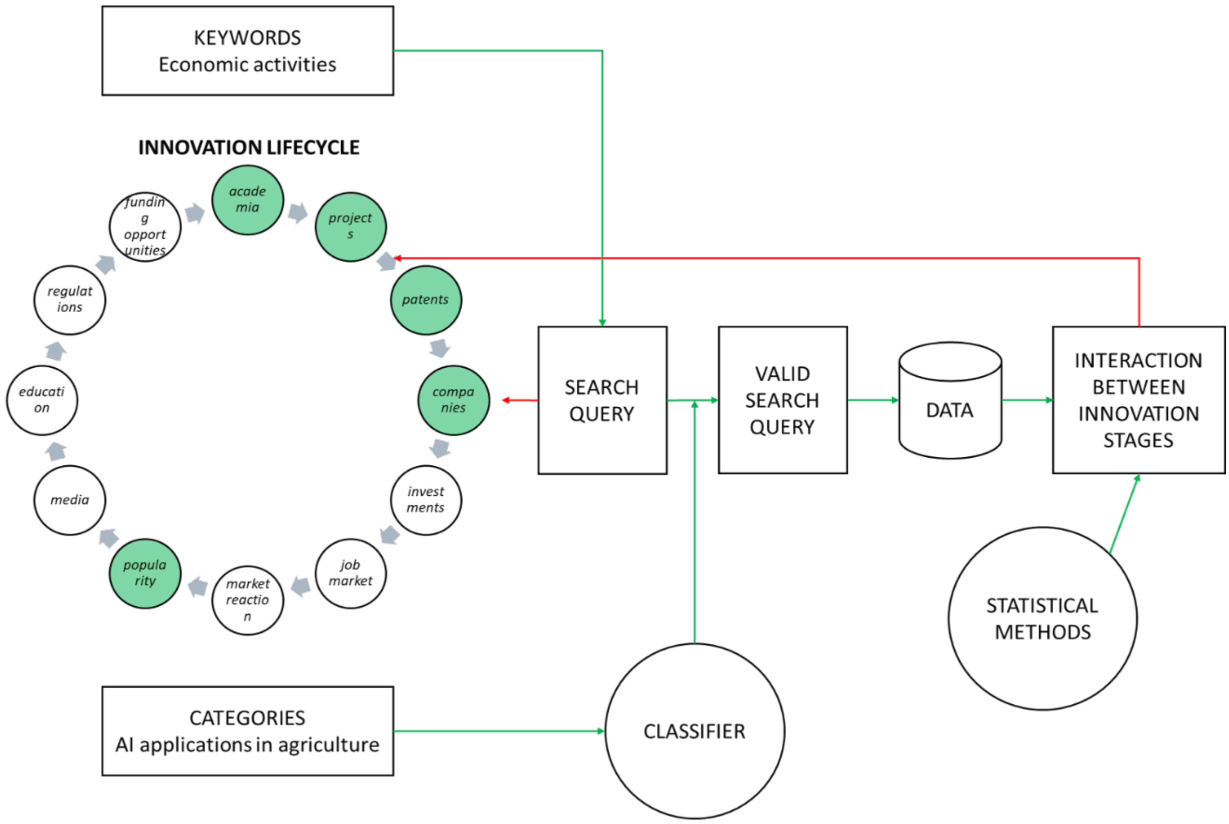

Figure 1: Overview of the main methodological steps.

Figure 1 summarizes the methodological steps undertaken to test the previously formulated hypotheses.

4. Business scenarios

It is of great interest to acknowledge the importance of business scenarios and influential pillars of AI innovations within the agricultural sector. Each pillar brings contributions independently but also interacts dynamically with the others, creating a complex web of influences that determine the overall pace and direction of technological advancements. Those core elements, including scientific research, project implementation, patents, start-up ventures, and public acceptance along with the integration of traditional farming techniques shape agricultural development.

The integration of AI into agriculture extends beyond the technological insertion; it represents a paradigm shift in how data is processed to inform specialized individuals of various decisions ranging from planting, harvesting to market timing and supply chain logistics. Predictive analytics, for example, leverage historical data and AI models to forecast weather patterns or pest infestations, enabling a more controlled and aware environment for everybody involved to mitigate risk and optimize resource use. In addition, the role of government policy and regulatory frameworks cannot be overstated. Effective policies must be implemented to balance the innovation factor with the various concerns the AI domain brings, like privacy, data ownership and the ethical use of those technologies to ensure that the benefits are fairly distributed among all stakeholders, including small-scale farmers and rural communities.

5. Conclusions

The expected outcomes of this research include a set of validated models that can predict the flow of innovation in AI within agriculture based on historical trends. The implications of this study are significant, providing insights that could guide strategic planning and investment in AI applications in agriculture. By understanding the causality and sequentiality in innovation, policymakers, investors, and entrepreneurs can better align their efforts with the most impactful areas. This research not only contributes to academic knowledge but also has the potential to influence real-world applications and practices in agricultural technology.

Nonetheless, through this abstract, we aim to explore the potential of our research and discuss methodological approaches for developing high-quality outputs suitable for publication in top-tier journals.

Acknowledgements

This research is financed under the Romanian National Recovery and Resilience Plan, by the Romanian Government, under the contract number 268/29.11.2022, Entitled “CAUSEFINDER-CAUSALITY IN THE ERA OF BIG DATA”

References

[1] CrunchBase database (AI companies), 2024. URL: https://www.crunchbase.com/

[2] Web of Science database (articles of AI topic), 2024. URL: https://www.webofscience.com/wos/woscc/basic-search

[3] Cordis database (projects having AI topic), 2024. URL: https://cordis.europa.eu/projects

[4] WIPO – World Intellectual Property Organization (database of patens having AI topics), 2024. URL: https://patentscope.wipo.int/search/en/search.jsf

[5] Google Trends (statistics of AI related topics and keywords), 2024. URL: https://trends.google.com/trends/explore?date=2010-07-19%202024-08-19&q=ai%20agriculture&hl=ro