Strategic and Competitive Intelligence

Report
Precision Agriculture

Brasile Pierluigi
Iannolo Benedetta Amalia
Manai Eva
Volpi Federico

AY 2022/2023

Summary

Introduction

"The intent of precision agriculture is to match agricultural inputs and practices to localized conditions within a field to do the right thing, in the right place, at the right time, and in the right way".

(Pierce et al., 1994)

"Precision Agriculture is a management strategy that gathers, processes and analyzes temporal, spatial and individual data and combines it with other information to support management decisions according to estimated variability for improved resource use efficiency, productivity, quality, profitability and sustainability of agricultural production."

(Official definition of the International Society for Precision Agriculture (ISPA), January 2021)

Two definitions of AP, in 1994 and 2021 [01]

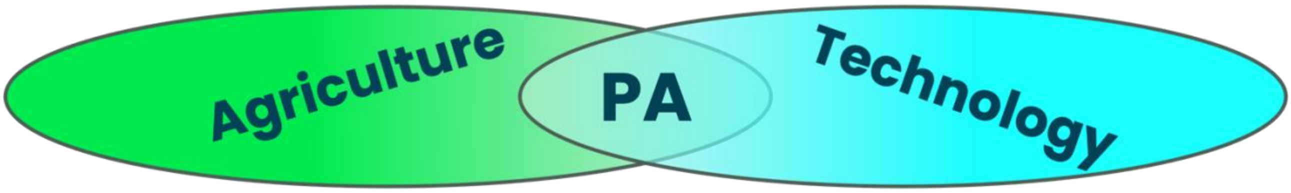


Figure 1: Precision Agriculture: an intriguing intersection between Agricultural and Technological macro-areas.

In this report we examine the evolution of Precision Agriculture over time from different perspectives. Precision Agriculture (PA) caught our attention because it is an innovative approach that has made possible the creation of efficient and sustainable agricultural systems, resulting from the intriguing intersection of two important sectors: agriculture and technology.

The purpose of this report is to discover various aspects of PA that can drive on a favorable path both company's plans and strategies, which want to implement a data-driven approach for decision making processes in the wide field of Precision Agriculture. We try to highlight how the transformative power of technological advancements is (and will potentially be in the future) world widespread and known. We research key insights, trends, potential gaps, to support stakeholders (as researchers, entrepreneurs, investors, policymakers) in making informed decisions.

Motivations

'Agriculture touches on every single aspect of human life'

Agriculture; Rudolf Steiner, Catherine E. Creeger, 1993, SteinerBooks [0]

Agriculture is one of the oldest and most important human activities, which provides food, fiber, and essential materials for the survival and well-being of humanity. It includes specific activities all directed to the accomplishment of the main goal: to produce sufficient and high-quality food. Since the time of ancient Egypt, agriculture has played a crucial role in the economic growth of many countries through four forms: in terms of product, in resources, in markets and in terms of foreign exchange [02].

Technology is the study and practical application of scientific knowledge for problem-solving. It encompasses the use of scientific approaches and advanced methodologies that can be applied to improve the performance of productive activities and enables the development of innovative solutions to meet and simplify human needs.

The integration of technology into agriculture is of utmost importance as it enables us to effectively address significant challenges. With the growing global population, there is an increasing demand for food production. However, this needs to be accomplished while ensuring the preservation of resources and the environment. By harnessing advanced technologies and optimized techniques, we can meet these challenges head-on. The integration of technology empowers us to enhance agricultural productivity, optimize resource utilization, and minimize environmental impact, thereby creating a sustainable and resilient food system for the present and future generations.

The following subcategories represent integral components that drive the development of precision agriculture. These subcategories encompass key domains that are pivotal in advancing the principles of precision agriculture, ultimately fostering improved efficiency and sustainability in modern farming practices.

- Crop Monitoring: Real-time data collection on soil conditions, plant health, and growth, to allow farmers to make informed decisions to optimize crop management.
- Resource management: precise use of agricultural resources such as water, fertilizer, and pesticides, minimizing waste and environmental impact and maximizing efficiency.
- Pest and Disease Control: early detection and targeted response using advanced technologies, such as sensors and data analysis to ensure the health of crops.
- Drone-based precision agriculture: integration of drones equipped with sensors and cameras for high-resolution field mapping, real-time monitoring, and proactive identification of problems.
- Agricultural data management: organization and analysis of data from different sources, which leads farmers to an optimized decision-making process.
- Automation and robotics: Integration of intelligent machines and robots into specific farming tasks.

Methodology

Next, we want to summarize and clarify the procedure adopted to lead this research. It can be divided into four main steps, inspired by the KDD process structure [03], that seems to us a customizable framework that easily fits our necessities along this study, given that our main challenge is drive it relying on information embedded into patents: so, we are going to face a big amount of data from which we want to retrieve useful, intelligent, and strategic knowledge to support decision making processes that can guide some specific kinds of stakeholders. In this section, we present detailed descriptions of the first two steps, then we dispose separate sections for the last two steps.

Figure 2:Four main step of general methodology followed.

Field Understanding

As first step, we want to frame the topic "Precision Agriculture" to get a raw idea of our challenging task, trying to answer some simple trivial questions emerged during the first brainstorm session we take. Moreover, we need to give some bounds to preliminary information needed, avoiding becoming experts in the sector of Precision Agriculture and inevitable miss our main objective.
To do this, we apply an iterative process summarized in *figure 3*, that allow us to calibrate and better address the research of information at each iteration taken.
The process starts by selecting reading sources (see below), on which we test the actual KITs and KIQs sets. The selection phase is based on reliability and representativeness of those sources, meaning that for example, we consider significant only readings from trustworthy references (official, well-known, authoritative, scientifically approved, …), in a determined interval of time ("last ten years") and obviously that are involved in our matter of study, or in one of its subsectors. This, give us a set of boundaries that we can adjust, if necessary, during the process.
Then, these sources are used to frame KITs, answer KIQs, improve, refine, or change the definition of them, and also build a query that it is our tool to retrieve data from patent databases. As well as KITs, KIQs, and Query, also Readings are objective of improvements; basing on what gathered by queries results and the level of fitting with our KIQs and KITs, we can search more appropriate readings to get meaningful insights. The entire process can output a set of KITs and KIQs and a Query at each round, but only after some iteration we can reach significant outcomes, with desired significance and granularity levels.

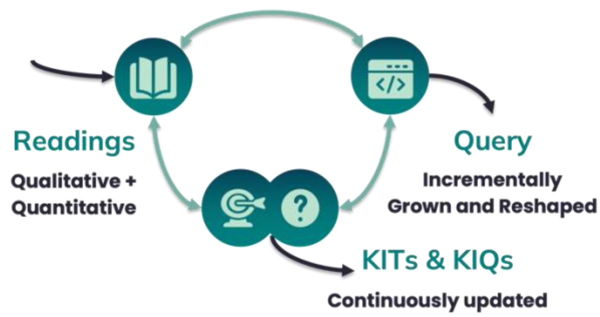


Figure 3: Field understanding step: an iterative process.

Let’s see in detail what these nodes do:

Readings

The readings employed in this first phase are both qualitative and quantitative, following “data sources” type of Triangulation Method as presented by Denzin in [04] with also the application of the traditional model “Convergence” later defined by Clark in [05]. This framework, establish to collect, and get results from qualitative data separately from quantitative ones, and then let them converge in an interpretation phase where “outcomes have the purpose to validate, confirm, give value to quantitative results with qualitative findings” [06]. So, it fits our necessity to have a wider idea of the task and a manageable quantity of data, and we employ it both for collection and evaluation of readings gathered.

Different kinds of sources were examined:

- **commercial, informative and scientific papers**: Wikipedia [07], Ecowatch [08], Science [09], Medium [10]), trends and surveys (CropLife [11], EIP-Agri E[12];
- **professionals’ websites:** FinScience [13], Regrow Ag [14], Small Robot Company [15], Soiltech [16], Precision AI [17], Agri Intelligence [18];
- **video-advertisements**: available on YouTube [19];
- **university of Pisa researches:** Biomedical & Health Informatics Lab [20], and learning material of Agricultural science courses, Department of Agricultural science [21].

So, the couple we find more interesting is composed by:

- **Government guidelines** and legislative acts, like the one published by Italian Ministries of Agricultural in December 2017 [22]; the policy framework [23] published on European Climate Adaptation Platform (Climate-ADAPT) [24], that represent a partnership between the European Commission and the European Environment Agency (EEA [25].
- **National and International associations** interested in the themes of: “Food” (FAO [26], FDA [27], EFSA [28]), “Agriculture and Precision Agriculture” (ISPA [29], IFAMA [30], AES [31]), and “Environment” (IPCC [32], Greenpeace [33]).

In our opinion the worth these two last wells have in this study is very impactful: by concentrate our attention on specialistic associations that have an high and detailed level of criticism, where many people interact quickly and produce a very update set of information, we can focus on the “state of art” of the matter, collecting clues both on existing realities and foresights for future, and last but not least, future legal requisites that companies will be soon required to comply.

KITs & KIQs

Through first raw mappings of the subject, we have a starting point from which we want to frame better the topic, by setting our own matching definition, boundaries, subsectors of applications, actors involved, knowing that a well-defined set of Key Intelligence Topic must represent team’s

intelligence needs and provide it a guidance that can make the difference between an effective program and one that waste resources on low-value activities.
To limit the depth of the growth of our discoveries, we need to define bounds and goals of our study, to do that we support ourselves by framing some Key Intelligence Topic (also accounting the three main categories proposed in [34]: Strategic Decisions and Actions, Early Warning Topics, Descriptions of Key Players) and we want to analyze each topic by a set of investigative questions (Key Intelligence Question) we want to answer.
So, KITs and KIQs are used mainly for evaluation and communication purposes: by evaluating the degree of fitting of information actually gathered (by readings and query), we can reshape the query (empowering also the data collection step) and look for more specific readings. Moreover, in intermediate phases of this study, these KITs and KiQs help to understand if content and results actually found in readings and gathered by patent analysis contains valuable information to be communicated as interesting clues, respecting the guidance of KITs and KIQs. In the final phases, KITs and KIQs are also used to organize contents of results.
At the beginning, we detect manually a bunch of interesting topics, that, by applying the enhancement process in *figure 3*, were filtered ending up with three main topics: Robotics, Artificial Intelligence, and Image Recognition.
For the choice of Key Intelligent Questions, it was difficult to select, but more difficult to discard questions. So, we decide to let us drive by "4W's" Method [35] customized by us into "6W's" to simplify the procedure of questions choice. We want that each step taken in the iterative model that involves the node "KITs & KIQs" produce a new question, or modify, or discard an existing one. The initial set of key intelligent questions reached is shown below as example, in *figure 4*.

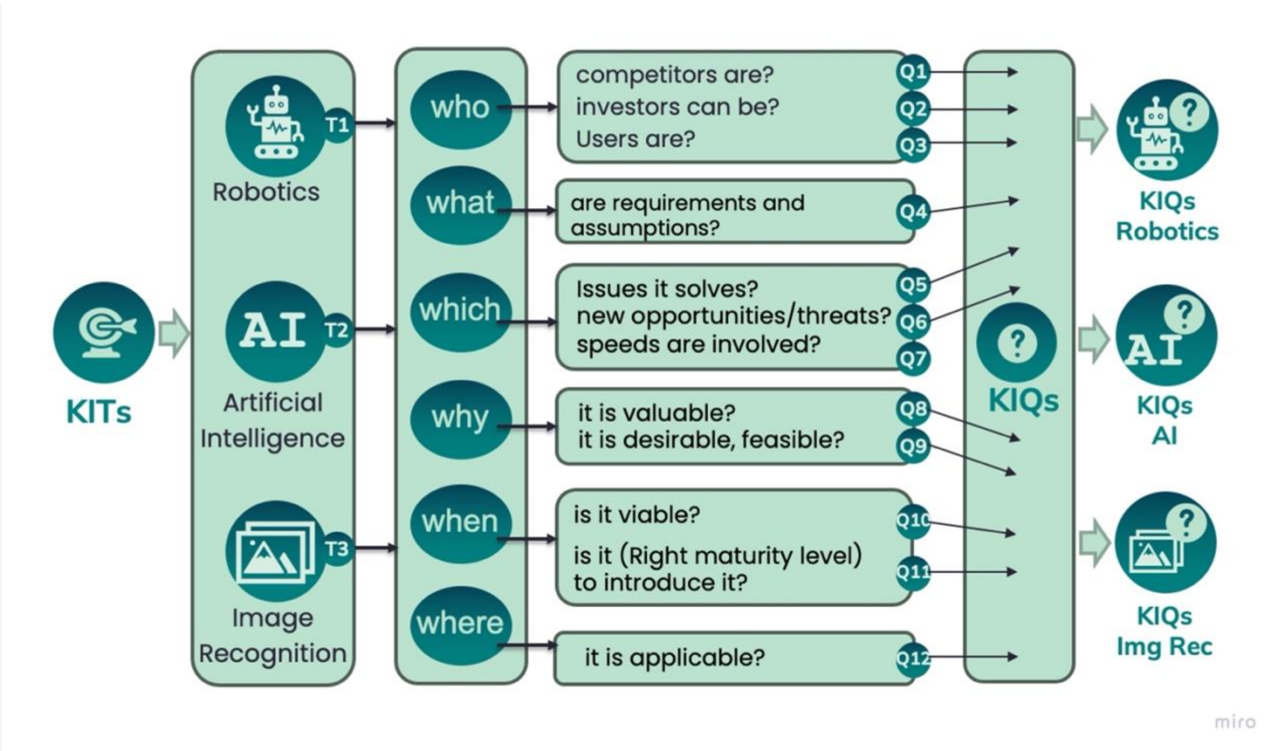


Figure 4: We derive KIQs from KITs using "6W's" Method.

Query

At the beginning, the query we incrementally build is used as explorative tool, for the definition of KITs and KIQs and as guidance for the selection of readings, analogously of what done in the other two nodes. Here, we present first the quality criteria we adopt to decide if the query needs to be improved, by pruning, adding or substitute terms, and in the second part of this paragraph we explain details on its final structure. Hence, we use a lot of different queries as evaluation method to frame and enhance KITs and KIQs, and search for significative readings in firsts phases, and only when we get a satisfying quality of feedbacks in patents retrieved by the query, we truly use the "elected" one to collect patents, to further submit them to the next data analysis step.

Figure 5: Incremental growth of "The" query

Quality criteria & Precision evaluation

The abovementioned quality is assessed through a customized method based on precision and recall, given that we couldn't find a more objective and precise way to measure them, and due to the impossibility to retrieve the total number of relevant elements in data pool as precise measure.

The two standard metrics are defined as follows:

$$Precision = \frac{Tot.\ Relevant\ patents\ in\ output}{Tot.\ Patents\ in\ output} \qquad Recall = \frac{Tot.\ Relevant\ patents\ in\ output}{Tot.\ Relevant\ Patents\ in\ data\ pool}$$

Formulae 1: Precision and Recall as usually defined on Patents set.

From the definition of recall, it's easy to understand why we need to adapt this metrics to our strengths: we can't really know how many patents are relevant in whole data pool.

Queries are run on three famous patent databases platforms, Espacenet [36], Google Patent [37] and Lens [38]. Once gathered and ordered by relevance the set of *N* patents retrieved (option available directly on platforms), we submit the subset of *n* most relevant patents to our personal evaluation: by reading the fields "*Title*" and "*Abstract*" we search for the first insights of "match", then we move to "*Claims*" and "*Description*" fields to confirm correspondence or finally reject the patent, by using the three following rules, obtaining quantitative measures to employ in the precision formula. So, $n=20$ patents are manually evaluated in a [-5; +5] interval of goodness, called by us PGS (Patent Grading Scale) by giving importance scores (positive and negative weights) following these three rules:

R1) If a patent's field is judged positively (negatively) respondent to our criteria (defined in KITs and KIQs, and in meaning of terms used to structure the query), the relative Patent gets a positive (negative) grade equal to the weight associate to it, as defined in *table 1*:

Field in patent	weight
Title	± 0.5
Abstract	± 1
Description	± 1.5
Claims	± 2

Table 1: Importance assigned to patent's fields.

R2) Patents that obtain an ambiguous grade (i.e., in the neighborhood of *0*: the two-sided "ambiguous region" of bounds *±2.5*), need to be further investigate by reading wider portions of fields that contains the keywords searched, and if grade doesn't change, this patent can be swapped with another one, randomly drawn from the set of top *2n+1* more relevant patents, of relevance rank in [*n+1; 2n+1*].

R3) Patents that reach a grade more extreme than *2.5* can be immediately classified as "Accepted" or "Rejected".

Figure 6: Patent Grading Scale to evaluate goodness of patents.

Once computed the number of "Accepted" patents among the top-*n*, we simply use the above formula of precision to score the set of patents retrieved with actual query. Starting from values of Precision that vary from *30%* to *50%* for the initial attempts, that involves hundreds of thousands of patents, we subsequently refine the queries until we get a significant result of *80%*, keeping this last one as "The" query, that was able to retrieve few dozens of thousands of patents (*N*), dense of useful information.

To ease the reading of this big amount of data, we rely on tools available in most common pdf viewer applications: once downloaded the pdf files we perform an automatic search of desired terms, to let us read only the neighborhood of them: often, the specific local context is useful to take the final decision.

We also face some issues, for example when our query contains the bigram "Neural Network", we have a lot of patents that seems to be "respondent" to some of our criteria, but only by looking to the neighborhood of this bigram and then to its separate monograms in pdf files we are able to notice that we retrieve a lot of patents related to Neuroscience or neurological issues, or again network-based neurological stuff, confirming our suspicions of noise introduction in data.

Query Structure

The underneath logic structure with which we grow the query since firsts attempts substantially doesn't change through each refinement step but only contents and dimensionality are affected by significative changings over developing time. Thinking at the KITs as deeper intersection from the original one presented in the motivation chapter, we get inspired to consider further aspects for a more holistic analysis, so we decide to analyze some areas of interest.

Starting from Agriculture and Technology, the initial intersection, we notice that is very challenging to avoid explosion of branches in the query, so we take a brainstorm session to decide five main areas in which classify keywords based on their similarity of meaning, inspired by the "Divide et Impera" approach to a problem. Thanks to the qualitative information gathered by readings, and feedbacks given in the entire process, we navigate pretty confidently in this challenge, understanding which areas are mainly involved. Here we briefly list the main aspects that characterize terms of each area chosen, also summarized as example in *figure 7*:

Agriculture: the type, the location of activity, the subsectors of application; actions, techniques, and basic agricultural terms.

Technology: complete names and acronyms of different kinds of technologies already used in agriculture or nowadays in constant development; specifical techniques terms.

Problem/Solution: known issues of agricultural sector, terms that indicate generic solutions to them.

Economics/ Law/Environment: concepts drawn from economics, job and food safety, legal requirements and environmental respect and preservation.

CPC/IPC classes: once identified which CPC/IPC classes reflects more our KITs, we add them to the query to better filter the set of patents to retrieve.

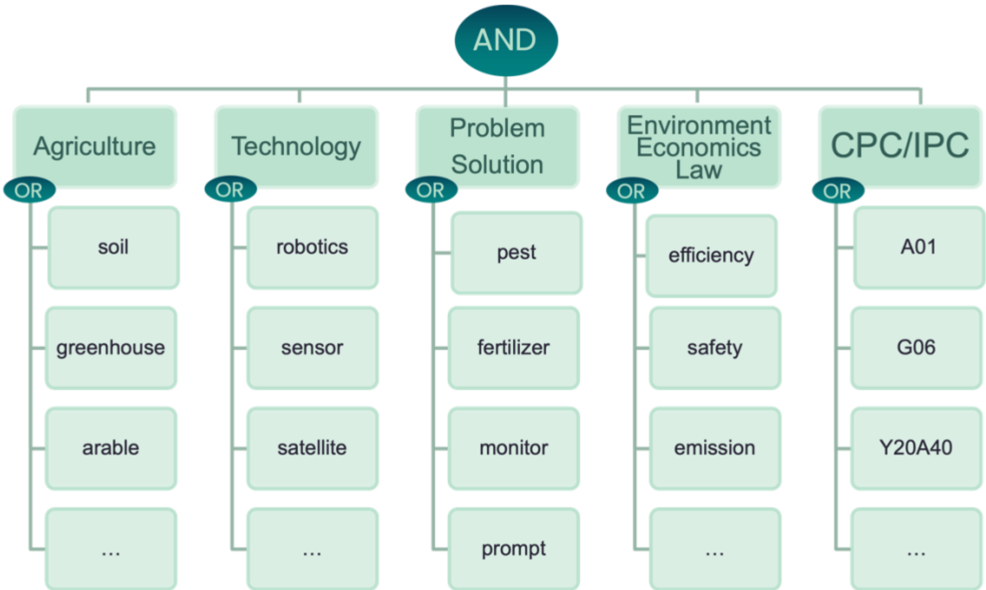


Figure 7: Five main categories used to build our queries.

As described in *figure 7,* we put these five main categories in an *AND* branch, meaning that we want to find at least one keyword in each area. Those areas are separately grown as an *OR* branch, with which we require to find at least one keyword among all terms in the *OR*. Moreover, also time constraints were added to the query, not reported in *figure* 7 for simplicity.

By performing this part of the process, field understanding, we prepare the basements to exploit the second step, Data Collection and Preparation.

Data Collection & Preparation

This second step begins when we obtain “The” query we want to use for our analysis. By exploring different patents databases available on internet in the previous step, we notice that each one has its own characteristics: the required syntax of queries, different ways to show results, sometimes different set of patents retrieved with same (translated) query, ranked with same criteria. Moreover, we found constraints regarding the number of downloadable patents, or restrictions on their field’s availability. We need to decide and, by comparing Lens with our initial choice of Espacenet, we get that the sets of patents retrieved with “The” query are comparable, and by performing again the Precision evaluation mentioned in paragraph [Field Understanding - Query], Precision value didn’t change. This platform also allows a greater number of downloadable patents and supply a lot of useful graphic visualizations, as we discover during query refinement period.
A comprehensive data cleaning and pre-processing phase followed to ensure data quality and relevance. This phase involved removing rows with missing values and duplicates, as well as eliminating irrelevant features. Also applicants' names were corrected to remove orthographic errors using the FuzzyWuzzy [39] library.

The resulting dataset can be summarized as follows:

- **Publication_Number**: Unique publication number assigned to a patent upon release.
- **Publication_Year**: Year of publication for each patent.
- **Title**: Title assigned to each patent.
- **Applicants**: Entity or person who submitted the patent application.
- **CPC_Classifications**: List of classification IDs assigned to a single patent.

To enhance the dataset and gather additional information, a data crawler was implemented using Python. Leveraging the BeautifulSoup [40] library, the crawler scanned Google Patents [E18] by searching for the publication number of each patent. This process enabled the retrieval of the following information:

- **Abstract**: Written content of the patent.
- **Claims_count**: Number of claims within the patent.
- **Citation_count**: Number of citations received by the patent.
- **Citation_ids**: List of publication numbers that cited the patent.
- **Citation_dates**: List of dates when patents in Citation_ids cited the patent.
- **Cited_count**: Number of patents cited by the patent.
- **Cited_ids**: List of publication numbers cited by the patent.

Patent Analysis

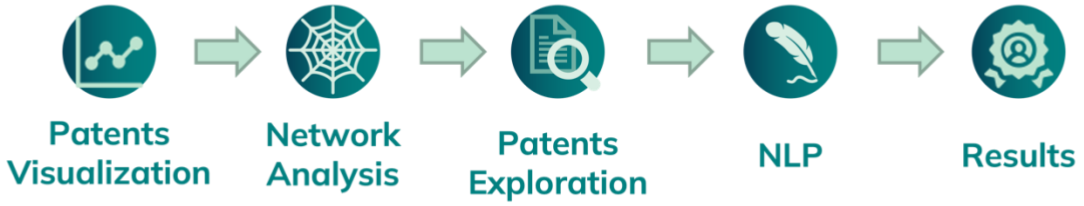


Figure 8: Patent analysis main steps.

Here we present how we conduct analysis on our elected type of data: patents. We focus both on methodological aspects jointly with intermediate findings gained, that strongly characterize this third step of our total workflow. We conduct Patent analysis as summarized in *figure 8.*

Patents Data Analysis & Visualization

Our initial approach involved obtaining a temporal overview of patent publications over the past decade. Through *Figure 9*, we observed distinct patterns that provided valuable insights. From *2013* to *2018*, patent publications remained relatively stable. However, starting from 2018, there was a significant exponential increase in the number of patents published. This shift prompted further in-depth analyses.

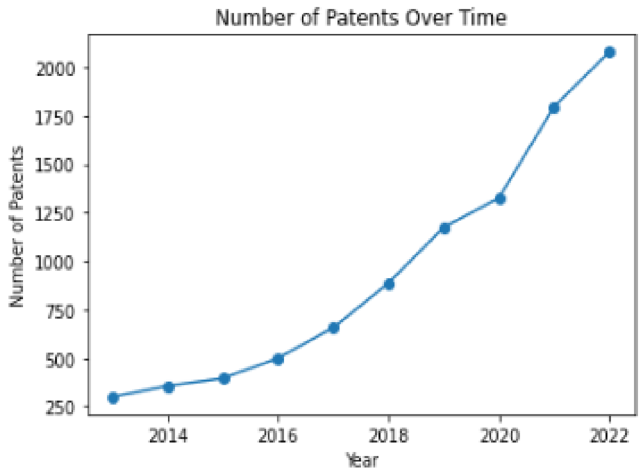


Figure 9: Patent publication distribution over time.

To gain a better understanding of the subsectors with the highest number of published patents, we created a bar plot visualization based on the CPC classifications. This visualization reported in *Figure 10* allowed us to identify the subsectors that exhibited a significant number of patents.

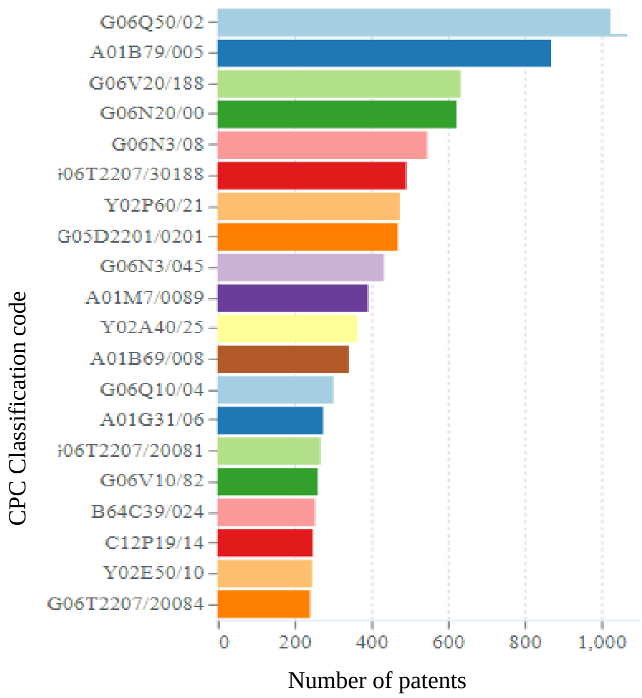


Figure 10: Number of patents for each sector.

Following the visualization of CPC classifications, our focus shifted to identifying the top sectors with the highest number of published patents. Two sectors emerged as the most prominent:

G06Q50/02:
Systems or methods specially adapted for specific business sectors, such as Agriculture, Fishing, Mining.

A01B79/005:
Methods for working soil; Precision agriculture.

We proceeded to identify the top companies within these sectors based on two temporal spans. The bar plots representing these companies were normalized to the total number of patents in each sector from *2013* to *2023* (*Figure 11*). The differences in the number of published patents among the companies were relatively insignificant due to the low number of patents in the initial years. Nonetheless, the graph provided insights into the presence of companies within each sector.

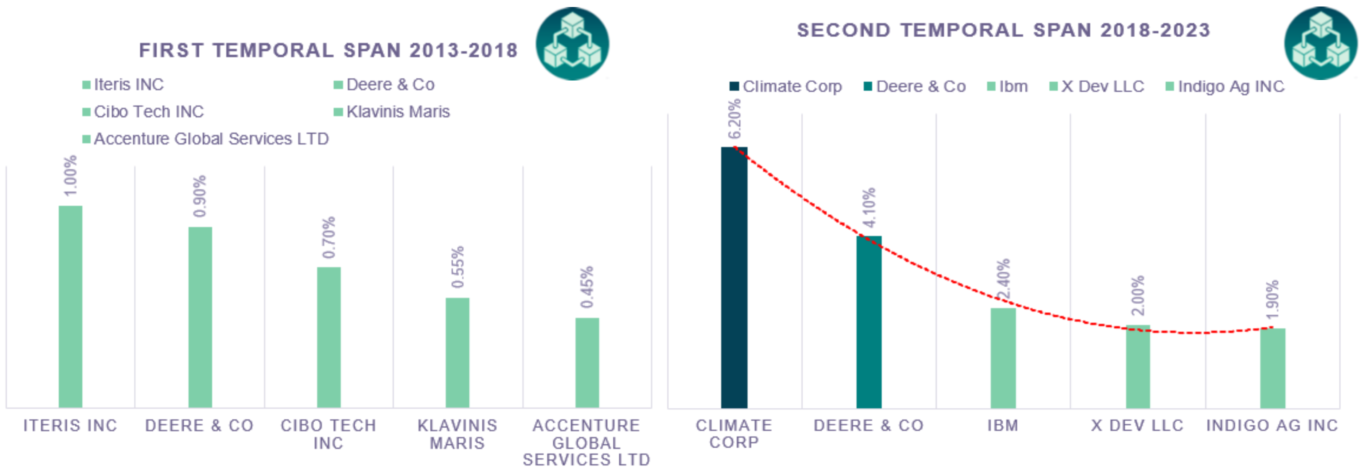


Figure 11: Company dynamic analysis for G06Q50/02

For the first sector, Iteris INC and Deere & Co emerged as the top companies. In the second temporal span, Deere & Co showed a 3.20% increase in the number of published patents and remained the only company in the top five. Accenture Global Services LTD and IBM also emerged as significant contributors.

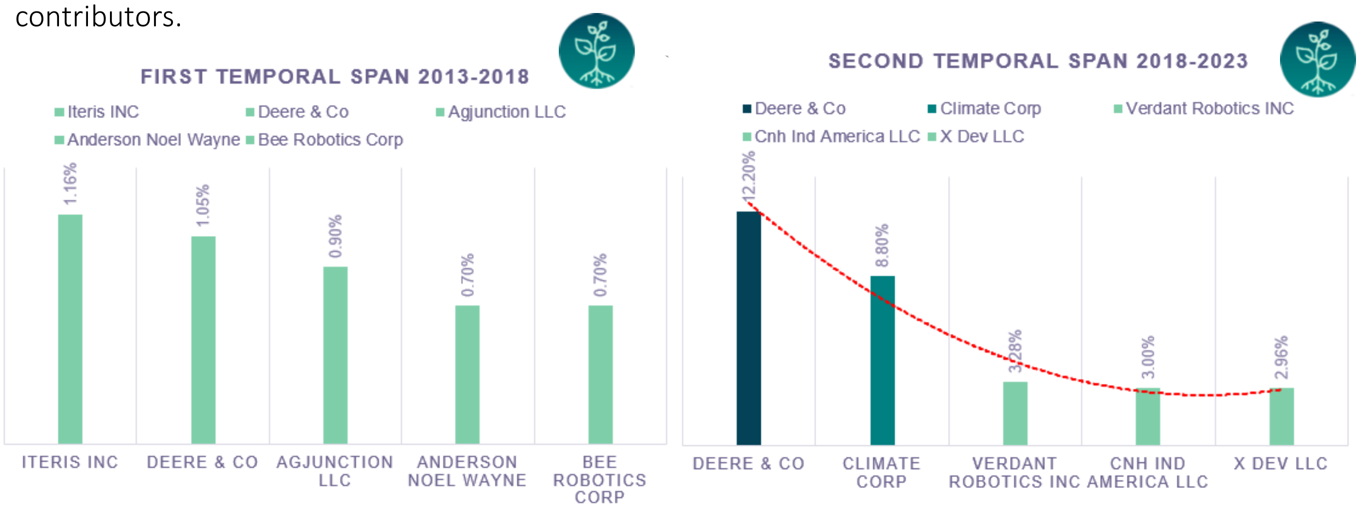


Figure 12: Company dynamic analysis for A01B79/005

Similarly, for the second sector, Iteris INC and Deere & Co were identified as the top companies, consistent with the first sector (*Figure 12*). However, in the second temporal span, Deere & Co demonstrated a remarkable *10.2%* increase in the number of published patents, accompanied by Climate Corporation.

We propose a model to better understand the temporal dynamics of firms within a specific sector.

The model aims to classify companies based on two key factors: the general level of patent affluence and the share percentage of patents published by the top companies under analysis. Drawing inspiration from the well-known BCG Matrix shown in *Figure 13* [41], we have adapted its structure to the patent context.

To construct the matrix, we plotted the number of patents registered in a certain sector during two distinct temporal spans along the y-axis. The x-axis represents the number of patents associated with each company. By adopting this approach, we allocated companies below or above the x-axis, considering the affluence of patents published during the relevant temporal span. Furthermore, companies were positioned on the left or right side of the y-axis based on their share of patent publications within the sector.

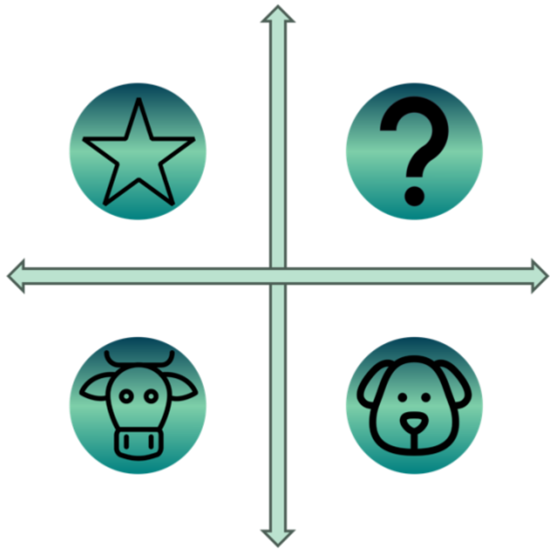


Figure 13: BCG Matrix

This modified BCG Matrix offers valuable insights into the entry, confirmation, and exit of companies over the past 15 years. By examining the positioning of companies within the matrix, we can understand the temporal dynamics and strategic trajectories of firms in the sector. The dynamic allocation of firms follows the following criteria: we define Patent Supply based on a certain time frame, which is normalized to the total number of patents retrieved in a specific sector.

$$Patent\ Supply\ (\tau) = \frac{N.Patents\ (\tau)}{Tot.N.Patents}$$

Formulae 2: Patent supply linked to a certain temporal span.

We then compare this value to a threshold, which is set at half of the total number of patents.

$$T = \frac{Tot.N.Patents}{2}$$

Formulae 3: Threshold of the two patents supply scenario.

Next, we selected one of two possible scenarios using the following formula. If the patent supply within a specific time frame exceeds the threshold, it indicates a high patent supply scenario, and we allocate the company over the x-axis. Conversely, if the patent supply falls below the threshold, it indicates a low patent supply scenario, and we allocate the company below the x-axis.

$$Scenario = \begin{cases} High\ Patent\ Supply & if\ Patent\ Supply\ (\tau) > T \\ Low\ Patent\ Supply & otherwise \end{cases}$$

Formulae 4: Assignment of Patent supply scenario

This data-driven approach allows us to classify a company in the appropriate quadrant, taking into account the potential patent supply scenario on the y-axis and the impact of published patents per company in a specific sector on the x-axis. Below there is an example that illustrates the aforementioned case study:

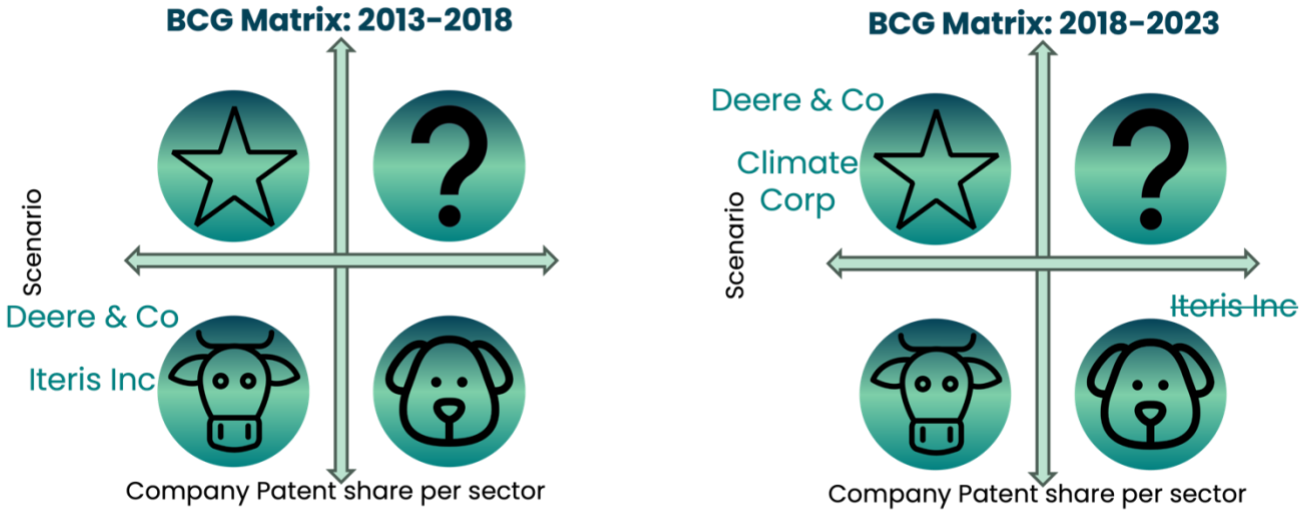


Figure 14: BCG Matrices of Companies in two different time spans

The application of the BCG Matrix, adapted to the patent context, enables a comprehensive assessment of the temporal dynamics and strategic behavior of companies within a specific sector. By analyzing the interplay between patent affluence and the share percentage of patent publications, this model provides valuable insights into companies' innovation strategies, market presence, and overall competitiveness. The matrix facilitates the identification of companies that have successfully entered and consolidated their positions within the sector, as well as those facing challenges or potentially exiting the market.

By leveraging the BCG Matrix's analytical framework, businesses and strategic decision-makers can gain a deeper understanding of the evolving landscape within the sector and make informed decisions regarding innovation, intellectual property management, and market positioning.

Our analysis revealed interesting findings regarding the top companies within the two identified sectors of precision agriculture patents. Deere & Co, being an agricultural machinery company, was expected to have a prominent presence in both sectors due to its specialization in mechanical components.

We also made the decision to investigate why Iteris Inc., an American company specializing in software services for infrastructure management, withdrew from the BCG Model. In 2015 and 2016, this company played a significant role in the development of a crop monitoring system based on weather simulation. Iteris Inc. was particularly active in this field, as it was the sole company utilizing land surface models for climate and weather prediction. In 2015, they even obtained a patent for their system. While it may not currently be recognized as an actively operating company in patent publications, it continues to provide IT solutions for the agriculture industry.

Additionally, the presence of IBM, a multinational technology company, publishing patents in the agricultural business sectors suggests the importance of this subsector and the need to monitor its developments closely. These findings provide valuable insights for strategic decision-making in the precision agriculture domain.

Network Analysis

In this task, we employed a two-step approach to identify the most correlated fields in terms of patent citations. Firstly, we extracted a list of subclasses to determine the fields that show the highest correlation with patent citations. Subsequently, we conducted a network analysis based on the co-classification of CPC codes, which provided further insights into the relationships between different sectors.
To identify the fields most correlated with patent citations, we extracted a list of subclasses. This allowed us to determine which subclasses had the highest frequency of citations. By analyzing the occurrences of these subclasses in relation to patent citations, we assessed their level of correlation. We performed a network analysis using the occurrences of co-classification CPC codes. Co-classification refers to the set of all CPC codes assigned to a patent. By counting the occurrences of each co-classification CPC code pair for each patent, we constructed an undirected network. In this network, the nodes represent the sectors, and the arcs represent the links between them.
For the network analysis, we considered two key measures to gain insights from the network:

- **Betweenness Centrality**: This measure helps identify the sectors that act as key connectors or bridges between other sectors in the network.
- **Degree Centrality**: This measure indicates the number of connections (or links) a sector has with other sectors in the network, thereby identifying sectors with higher overall connectivity.

The network analysis provided valuable insights into the interconnections and relationships between different sectors. By considering the Betweenness Centrality and Degree Centrality measures, we identified the most meaningful sectors within the network.To visualize the network analysis results, we used the Gephi [42] tool for *Figure 15*, generating a graph visualization. For clarity, we filtered the network based on the highest weight degree of each node. This approach allowed us to select the most significant sector interactions based on the frequency of their occurrences in patents.

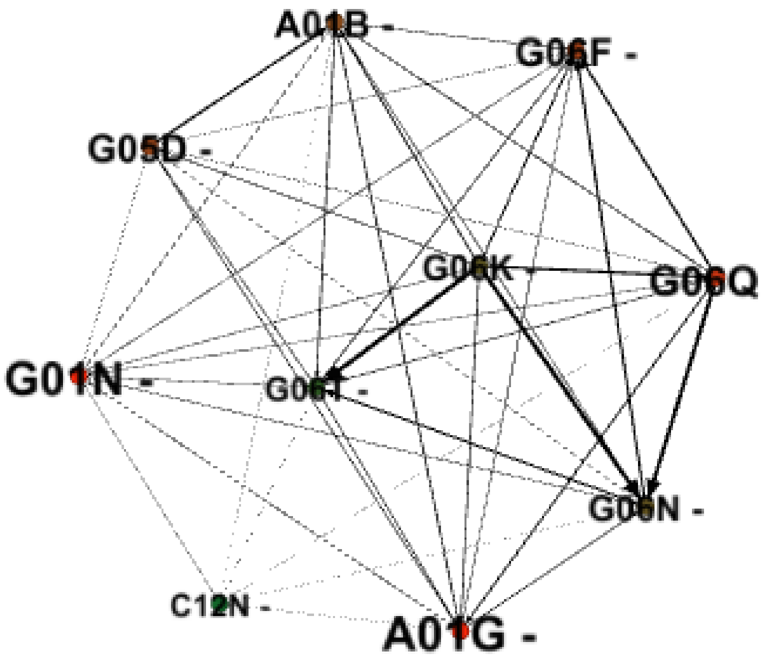


Figure 15: Weighted graph of most meaningful CPC, obtained with Gephi tool.

To summarize, we compiled a table that presents the top subsectors associated with each CPC group. This involved converting the CPC codes resulting from the graph analysis into words, providing a clear representation of the most relevant subsectors within each group.

Source_cpc	Target_cpc	Subsectors
G06K	G06N, G06T	Graphical data reading based on computational model
A01B	A01C	Horticolture, planting fertilising based on vegetable, rice, flower
C12P	C12N	Fermentation or enzyme using process and genetic engeneering

Table 2: Results of subsector detection.

By combining subclass analysis and network analysis, we identified the most correlated fields and sectors in terms of patent citations. This information is crucial for understanding the relationships between different fields and sectors, enabling businesses to make informed decisions regarding patent strategies.

The findings presented in the graph visualization and the table provide a comprehensive overview of the significant sectors within the network, highlighting the areas with the highest correlation to patent citations. This knowledge can drive strategic actions and resource allocation, allowing businesses to capitalize on the identified correlated fields and sectors for a competitive advantage.

Data Exploration

General Exploration

As first part of our data exploration process, we started from statistics given by Espacenet itself. We filtered by publications and by a time horizon considering the last ten years, from *2013-01-01* to *2023-05-31.*

As we can see in *figure 16* the countries with most patents' publications are *the US* (*3644*), followed by China (*3325*) and the World Intellectual Property Organization (*2286*). Other countries that also fare well are Australia (*1102*) and Japan (*876*).

We have decided to focus only on papers in English, because as literature showed for example referring to Zhang and Toral (2019), the translation from Chinese to English can lead to significant changes on training and test sets.

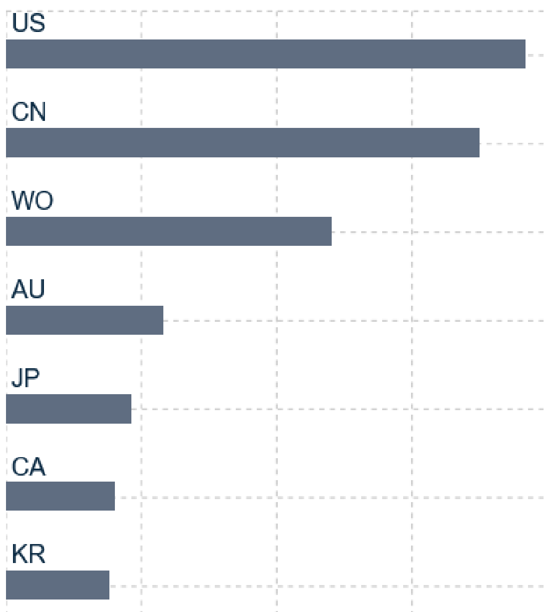


Figure 16: On the right: Number of patents per Country: US 3644, China 3325, World Intellectual Property Organization 2286.

Other interesting graphs for data exploration on our query regard the legal status. The Pending status is the most common (*3549*), immediately followed by Active (*3072*).

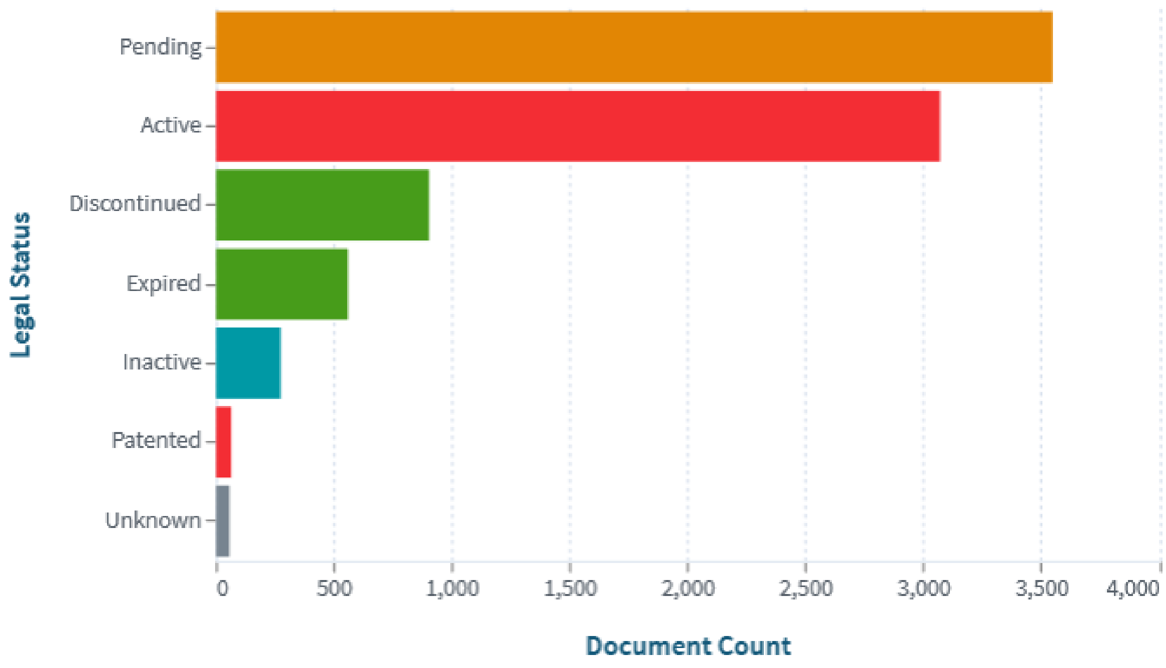


Figure 17: Number of patents per legal status.

Patent Applications count for *6079* of the document types, while Granted Patents are *2324*. With these numbers we understand that the patents are in large part applications themselves, with a minority being already granted.

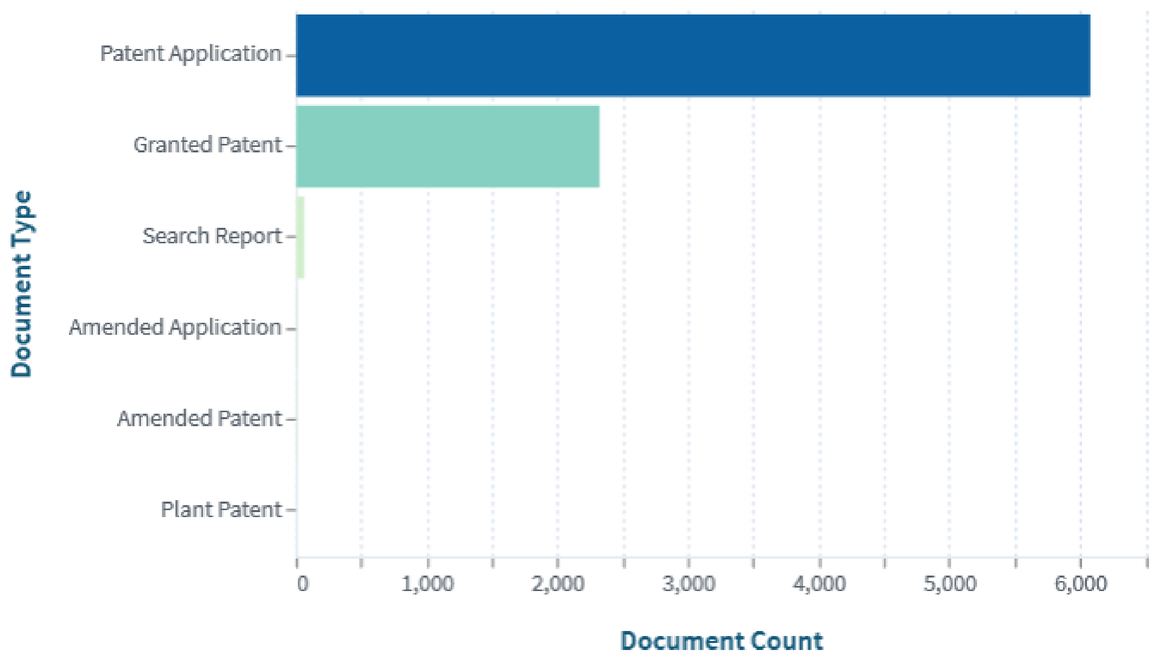


Figure 18: Number of patents per application status

Coherently with the Porter model [43], we proceed with an analysis of the competitors, inferring them from the frequency of the patent applicants. First, it is relevant to understand the percentage of patents published by firms or universities or others. We have found that firms constitute the 91% of patents from applicant types, with universities just 3% and inventors as applicants 5,9%.

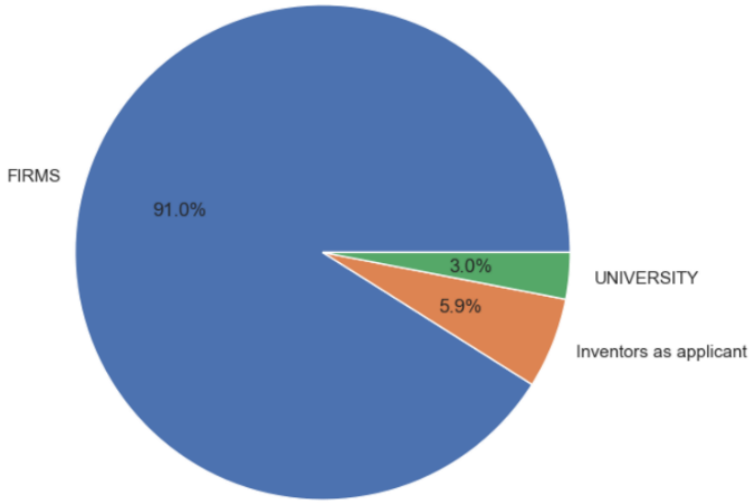


Figure 19: Percentage of patents released from applicant's types: 91% Firms, 5.9% Inventors, 3.0% Universities.

Focusing on the applicants themselves we have extracted, ranked, and plotted the top players of the niche we have selected in the query.

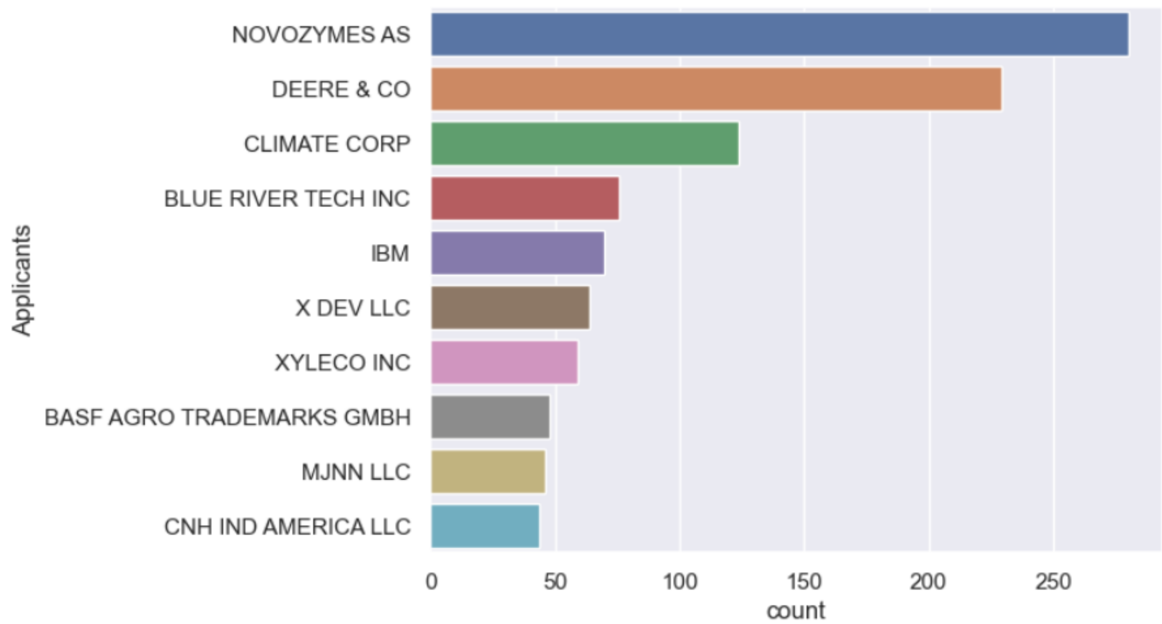


Figure 20: Number of patents per application status.

The presence of competitors like Deere & Co, an American corporation that focuses on agricultural machinery with $109 billion dollars of market capitalization, is an interesting fact to underline. It is relevant to know that the most important players have high bargaining power, and this may be evaluated when new projects are introduced to avoid building up on patents that these portfolios may hold strategically.

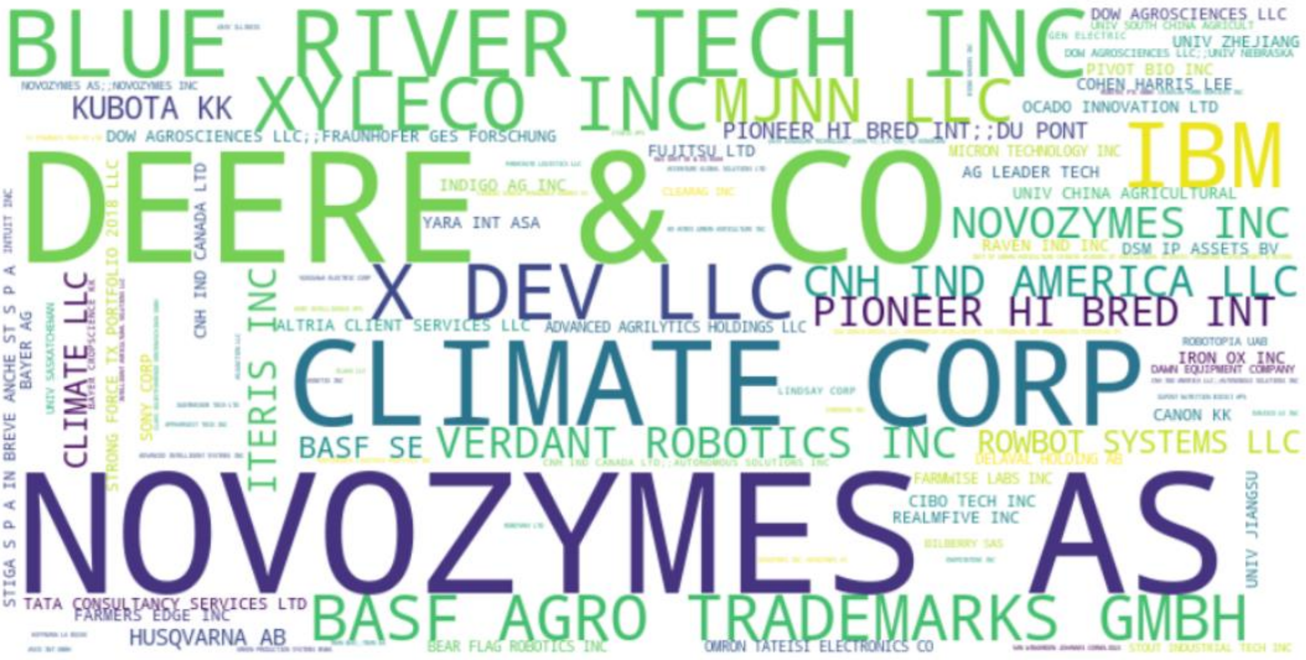
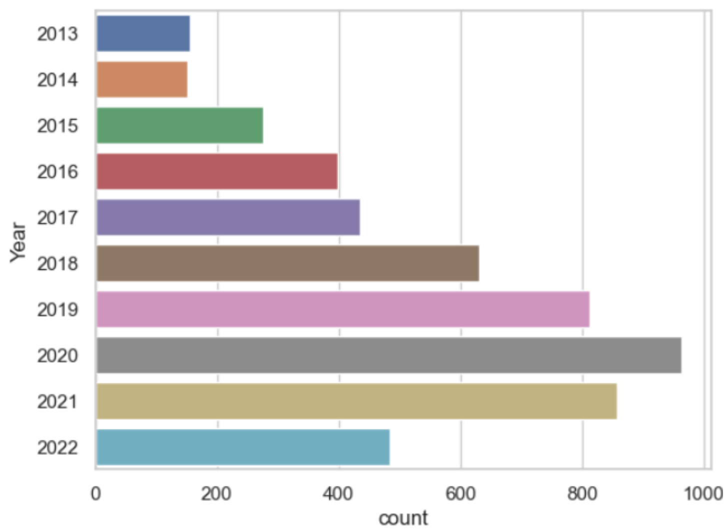


Figure 21: Word Cloud of competitors.

In the figure above, we have plotted the WordCloud of the various competitors present in the patents we have exported, in larger characters those that appear more often and in smaller ones those that appear less.

We have decided to study in more details the trends of patenting in the perimeter that we have defined. To do so, we have decided to retrieve the number of patents by year from their application date. The result, considering them in absolute terms, *as figure 22*:

Figure 22: On the right: Number of patents per application year.



In absolute terms, meaning counting the frequency of patents each year, we can clearly see a peak of interest in the sector starting in the second part of the 2010s, with a decline afterwards.

The fact that the sector in the perimeter of the query we have created had such a relevant peak can make us infer an increasing trend, but to refine better the analysis we have also decide to overlap these results with a plot of the growth rate of the frequency of current year and Y-1. We have of course to mention the fact that when plotting relative growth rate, the variance of the distribution will be larger at the beginning of the distribution and will stabilize over time, as variance is higher when data is less and tends to decrease as data increase.

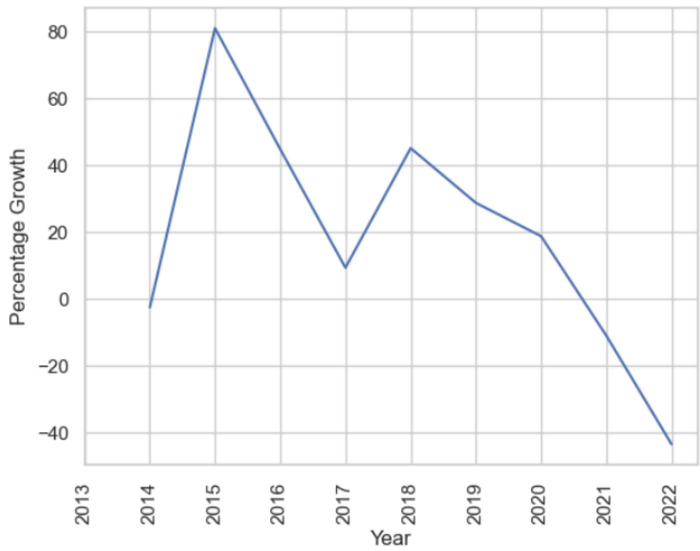


Figure 23: Patent frequency per year.

With this graph we can have a better understanding of the trends. We have on the y-axis the percentage growth and the year on the x-axis, with relevant peaks as the world economy started recovering from the 2008 crisis and stagnation in the years of the COVID-19 pandemic. To have a better understanding of the trend, we plotted separately the rate of growth of both China and USA.

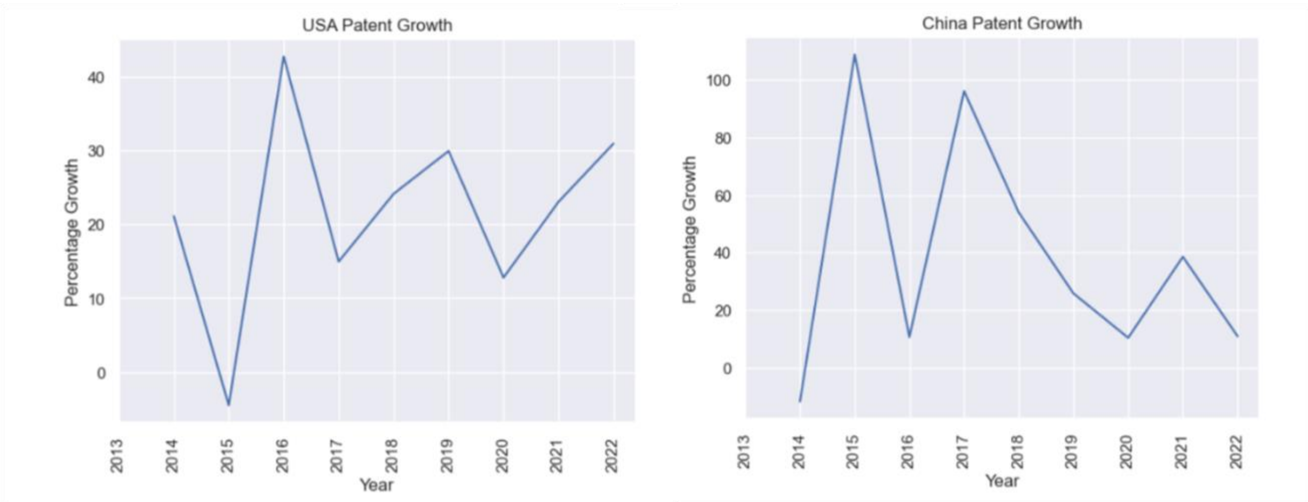


Figure 24: Patent activity growth in USA and China.

On average, China Patent Growth has been historically higher in the last decade than the USA, but both countries have still registered very high rates, nonetheless. It is worth mentioning that the USA even managed to increase their patent growth in the years of the COVID-19 pandemic and its aftermath.

To deepen our knowledge on the industry, we have done an NLP analysis. We have chosen a pipeline that fit our goals of extracting the most relevant and frequent words in the abstracts.

NLP analysis

First, we have converted all the content of the "abstract" column lowercase, then we have tokenized the column, in other words splitting each sentence in single words. Since there are certain words, such as pronouns or prepositions that recur very often, we have cleaned the stop words. Then, to increase the accuracy we have lemmatized the tokens, meaning that we have brought each token to its root word in the dictionary to make recognition easier. We have them computed the frequencies of the tokens, ranked them, and plot them through a horizontal bar plot.

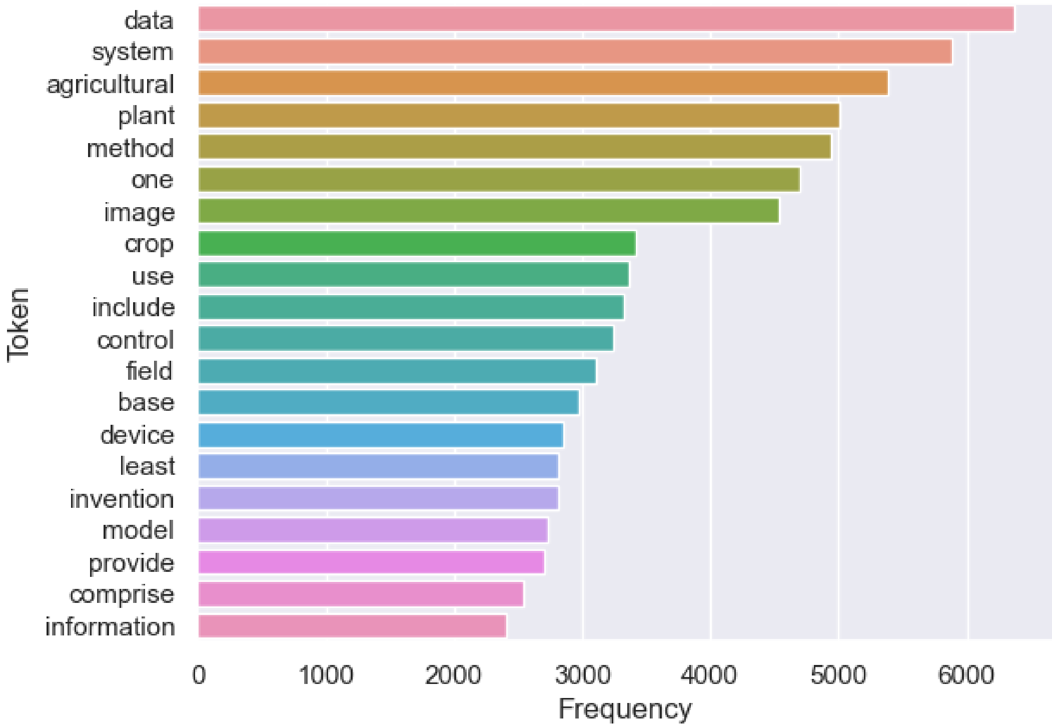


Figure 25: Most common words in patent's abstracts.

The result of the NLP analysis was the one of the figures above. The most common word is "data", that can point to a data-driven approach, or just only to the empirical side of the analysis. These results can be ambiguous, even if certain trends can be sketched: a stress on the sector on new methods, inventions, and systems to introduce new way of management in the agricultural sector.

To confirm our hypotheses, we have written a data-driven lexicon filtering for the top ranked words that deal with technology because of the NLP analysis, adding also other terms that emerged in papers on the topic. Then, we have computed the frequency of each word in the lexicon and plot the result for two different periods: 2013-2017 and 2018-2022.

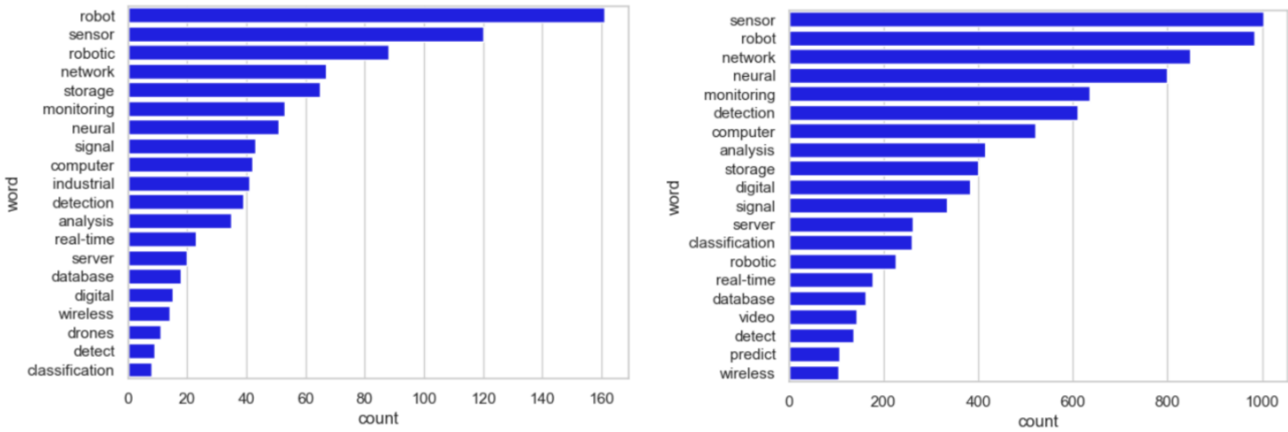


Figure 26: Most common words in patent's abstracts compared to a lexicon.

From these results clearer trends appear; "robot" and "sensor" dominate the list, confirming themselves as two of the most relevant technologies implemented in the agricultural sector. From the plot we can also infer the main goals of which the users are most interested in detection, analysis, monitoring, prediction. The bar chart also shows the emergence of new technologies, such as those related to AI.

It is worth noting that these plots reflect only unigrams so certain important bigrams, such as "neural networks", are not present. However, we can infer the higher importance of such AI-related technologies in the last time bucket looking at how the word "neural" increased in its importance relative to the word "network".

Deploy and Communicate Results

To let our findings be available and our procedures repeatable, we synthetize in this report the main steps taken, and methods adopted. We encourage get inspiration from our work, to design further customized procedures. We also did preliminary presentation of our work during intermediate phases of the entire process, to gain feedback from teachers and colleagues' teams, actively involved in studying the same topic, in the same time span.

Results

As KITs we have chosen three relevant technologies of which we wanted to understand the impact in the current trends in the subfields selected. These three were robotics, artificial intelligence, and image recognition. We respectively propose a data-driven evaluation of each, starting from the KIQs we had formulated to fill the gaps in our knowledge.

We focused on the following KIQs, that emerged as the most fitting after our data-driven analysis:

- Who are the competitors?
- What are requirements and assumptions?
- Which issues does it solve?
- Why is it desirable?
- Where it is applicable?

Robotics emerged as the most trending technology of the ones we have chosen, and even overall between the technology lexicon that we have extracted. As with the others three technologies, competitors in the subfields we have selected consist also of big players such as Deere & Co, that have the financial capacity to invest in the R&D. The assumptions that brought us to these results are of course technical, as regarding first of all of the consistency of our methodology. We have found the desirability of this technology to absolve the most important tasks of current agriculture such as detection and monitoring. Robotics is revolutionized itself year after year, but the viability of solutions in the field seems to be quite robust, as it maintained its top status for the entire last decade.

Artificial intelligence shares the same competitors of Robotics, and the same goals but had a different story. The viability of AI-related technology became apparent in the last five years, where a peak in the new patents of AI-related technologies in agriculture appeared. Its field of application involve crop monitoring and techniques of outlier detection.

While image recognition may not be as prominent as robotics and AI, it has proven to have its own distinctive value.

Both iteris Inc and Climate Corp have made significant contributions to the advancement of image recognition in weather forecasting and crop modeling niches.

By analyzing imagery data obtained from satellites, drones, and ground-based sensors, image recognition algorithms generate insights and predictions related to crop health, growth patterns, and yield potential. This enables farmers and agricultural stakeholders to optimize their practices, reduce input wastage, and increase overall productivity.

The adoption of image recognition technologies in weather forecasting and crop modeling offers several significant benefits. By automating the process of analyzing large volumes of visual data, image recognition enhances the efficiency of weather forecasting and crop modeling. This automation allows for faster and more accurate predictions, enabling stakeholders to make timely decisions and take appropriate actions.

One of the notable advantages of image recognition in this domain is the potential for minimizing inputs such as fertilizers, irrigation, and other applications. By providing accurate insights into crop health and weather conditions, image recognition aids in optimizing resource allocation. This optimization helps reduce waste and minimize the environmental footprint associated with agricultural practices.

The increasing number of patents in climate change mitigation technologies is a testament to the growing recognition of the importance of sustainable practices. The adoption of image recognition in weather forecasting and crop modeling aligns with this trend and further contributes to the development of technologies aimed at mitigating climate change.

This was the last concept of our report. We started from a particular sector of AI with a specific query, and we tried to answer to the most KIQ possible that we established at the beginning. Unfortunately, a couple of them were not deeply explored yet, like question 7 and 9, that need further investigations to be accurately answered.

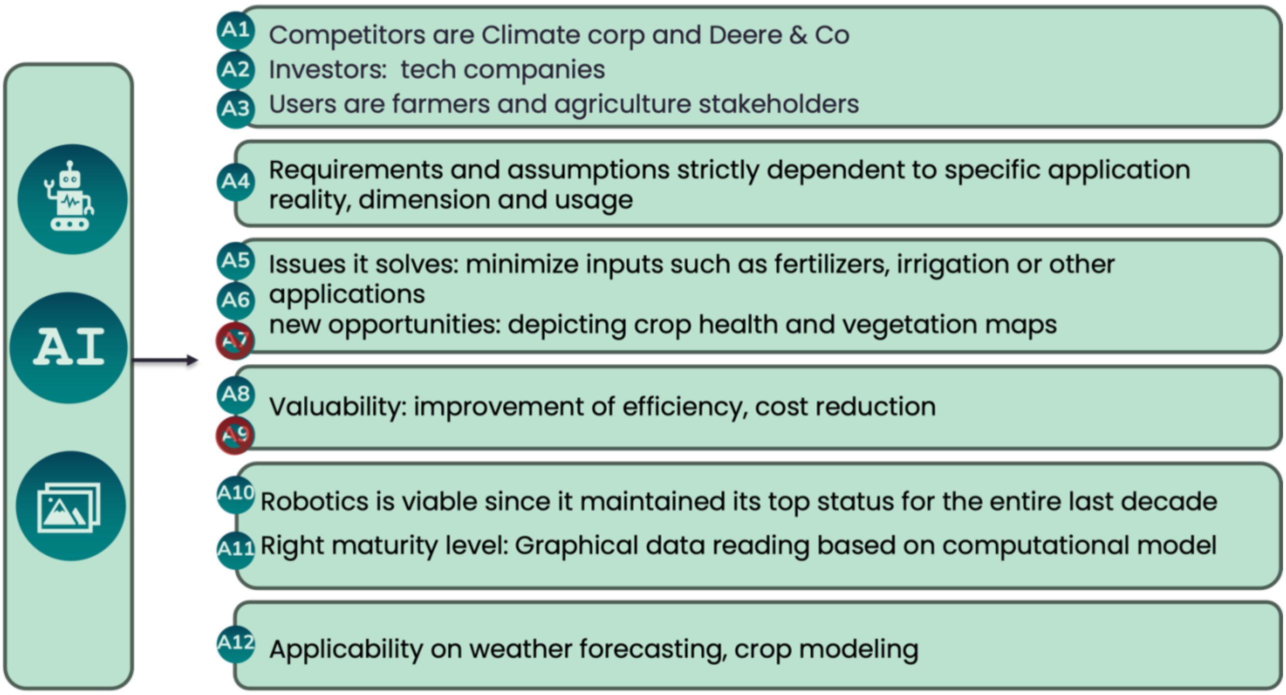


Figure 26: Majority of initial questions were answered, a couple need further investigations.

Conclusions

It is evident that agriculture is rapidly transforming into a high-tech industry, driven by advancements in technology and the integration of knowledge from complementary fields. Specifically, we anticipate a significant shift towards the adoption of artificial intelligence (AI) in agriculture, with the aim of reaping environmental benefits.

The aforementioned insights stem from our comprehensive analysis. However, it is crucial to acknowledge that the precise outcomes of our analysis may be subject to uncertainty. This is primarily due to the absence of domain experts in our research process.
Nevertheless, we sought to adopt a data-driven approach, ensuring a thorough understanding of the field and employing methodologies that bridged knowledge gaps. Our perspective was guided by our expertise in graph theory and natural language processing, with a strong emphasis on addressing business needs and objectives.

References

[0] https://books.google.it/books/about/Agriculture.html?id=VAy11h30A5sC&redir_esc=y

[01] https://www.ispag.org/about/definition

[02] (*L'agricoltura nello sviluppo economico*; Donato Romano; September 2010; University of Firenze, Department of Scienze per l'Economia e per l'Impresa)

[03] KDD: https://en.wikipedia.org/wiki/Knowledge_extraction#Knowledge_discovery

[04] Denzin, N. K. (1970). The research act: A theoretical introduction to sociological methods. Chicago, IL: Aldine Transaction.

[05] Creswell, Clark. (2017) Designing and Conducting Mixed Methods Research, chap. 4, third edition. Thousand Oaks, CA: Sage Publications. Retrieved at: https://www.sagepub.com/sites/default/files/upm-binaries/10982_Chapter_4.pdf

[06] Manai E. 2022 Design method review, Triangulation. Available on demand at 25728026@studenti.unipi.it

[07] https://en.wikipedia.org/wiki/Precision_agriculture

[08] https://www.ecowatch.com

[09] https://www.science.org

[10] https://medium.com

[11] https://www.croplife.com/precision-tech/2022-precision-survey-as-ag-economy-shifts-so-does-technology-use-by-retailers/

[12] https://ec.europa.eu/eip/agriculture/en/news/precision-agriculture-survey.html

[13] https://finscience.com

[14] https://www.regrow.ag

[15] https://www.smallrobotcompany.com

[16] https://www.soiltech.it

[17] https://www.precision.ai

[18] https://www.agrointelligence.co

[19] www.youtube.com

[20] https://biomedinfo.di.unipi.it/ict-methods-for-precision-agriculture/

[21] https://elearning.agr.unipi.it/course/view.php?id=1167

[22] https://www.politicheagricole.it/flex/cm/pages/ServeBLOB.php/L/IT/IDPagina/12069

[23] https://climate-adapt.eea.europa.eu/en/eu-adaptation-policy/sector-policies/agriculture/index_html/#policy-framework

[24] https://ec.europa.eu/eip/agriculture/en/digitising-agriculture/developing-digital-technologies/precision-farming-0.html

[25] https://www.eea.europa.eu/it

[26] https://www.fao.org/home/en

[27] https://www.fda.gov

[28] https://efsa.onlinelibrary.wiley.com

[29] https://www.ispag.org

[30] https://www.ifama.org

[31] https://www.aes.ac.uk

[32] https://www.ipcc.ch

[33] https://www.greenpeace.org/italy/tag/agricoltura/

[34] Competitive Intelligence Review, Vol. 10(4) 10–19 (1999)© 1999 John Wiley & Sons, Inc.CCC 1058-0247/99/04010-10

[35] https://medium.com/@moyukhmukherjee/from-whining-to-winning-the-4w-approach-d1178bb499b1

[36] https://worldwide.espacenet.com

[37] https://patents.google.com

[38] https://www.lens.org

[39] https://pypi.org/project/fuzzywuzzy/

[40] https://beautiful-soup-4.readthedocs.io/en/latest/

[41] https://it.wikipedia.org/wiki/Matrice_BCG

[42] https://gephi.org

[43] https://it.wikipedia.org/wiki/Modello_delle_cinque_forze_competitive_di_Porter

All images are edited on miro platform: https://miro.com