Food Policy 127 (2024) 102684

Contents lists available at ScienceDirect

Food Policy

journal homepage: www.elsevier.com/locate/foodpol

Technology-driven transformations in agri-food global value chains: The role of incumbent firms from a corporate venture capital perspective

Pablo Mac Clay [a,b,*], Roberto Feeney [b], Jorge Sellare [c]

[a] Center for Development Research (ZEF), University of Bonn, Germany
[b] Centro de Agronegocios y Alimentos (CEAg), Universidad Austral, Argentina
[c] Forest and Nature Conservation Policy Group (FNP), Wageningen University & Research, Netherlands

ARTICLE INFO

Keywords:
Agri-food systems
Innovation
Global value chains
Venture capital
Industry structure
Machine learning

ABSTRACT

Agri-food global value chains (GVCs) are failing to provide healthy and affordable diets within planetary boundaries. Many cutting-edge technologies are being developed to address the sustainability challenges throughout agri-food GVCs. While several studies seek to analyze the impacts of specific innovations on socio-economic and environmental dimensions, the relationship between industry structure and the process of technological change has not received much attention in the value chain literature. Here, we focus on the entrepreneurial landscape of innovation in agri-food GVCs to 1) identify which technologies have been receiving most support from investors and 2) analyze the corporate strategies of the largest agri-food multinational companies regarding their investments in new technologies. Using Crunchbase as our primary data source and machine learning for natural language processing, we have identified around 15,500 companies developing innovations linked to agri-food GVCs, from farming to last-mile delivery. However, our analyses show an imbalanced scene in which downstream technologies capture most investors' interest. Then, we use these results to explore the direction of investments by dominant agri-food firms, where we identify three trends: upgrading strategies to improve their core activities, defensive strategies to control technologies competing with their core business, and corporate portfolio strategies to seize profit opportunities.

1. Introduction

Global agri-food systems are failing to provide healthy and affordable diets within planetary boundaries. Food systems are responsible for around one-third of anthropogenic greenhouse gas (GHG) emissions (Crippa et al., 2021) and are one of the main drivers of biodiversity loss (Read et al., 2022), mainly due to land-use change, excessive agrochemical use, and depletion of water resources. Furthermore, from a social perspective, around 3 billion people cannot access affordable and healthy diets (FAO et al., 2022). This situation has worsened with the disruptions experienced in the last few years due to COVID-19 and the Russia-Ukraine war, which exposed the vulnerability of food value chains to shocks (Béné et al., 2021; Deininger et al., 2023).

As a result of these social and environmental pressures, we are currently seeing innovations aiming to steer agri-food systems toward more sustainable pathways and make agri-food global value chains (GVCs) more resilient to various shocks. The convergence of biotechnology with digitalization and artificial intelligence is transforming food systems at the farm level and many other stages, including farming inputs, food processing, logistics, packaging, and retail. This is accelerating the structural transformation of agri-food GVCs, which are becoming longer, capital-intensive, and with a higher relative share of post-farmgate activities in the value added (Reardon, 2015; Yi et al., 2021).

However, the emergence of new technologies is itself insufficient to ensure genuinely sustainable food systems. Agri-food value chains show high market concentration rates (Clapp, 2021; Sexton, 2013), in which a reduced number of large multinational companies set technological and commercial standards. Thus, this new socio-technical regime in which innovation is revolutionizing food systems holds the risk of being held up by lock-ins from the current system (Geels, 2019). We see evidence of industrial consolidation in food retail (Deconinck, 2021), commodity trading (IPES, 2017), seed biotechnology and agricultural inputs (Deconinck, 2020), industrial food and beverage companies (Howard, 2016), and agricultural equipment (Fuglie et al., 2012).

Although a recent and growing number of studies discuss the

* Corresponding author at: Center for Development Research (ZEF), Genscherallee 3, 53113 Bonn, Germany.
E-mail addresses: pmacclay@uni-bonn.de, pmacclay@austral.edu.ar (P. Mac Clay).

https://doi.org/10.1016/j.foodpol.2024.102684
Received 31 August 2023; Received in revised form 8 February 2024; Accepted 5 July 2024
Available online 24 July 2024
0306-9192/

potential of new cross-cutting technologies to transform farm and post-farmgate stages (Bunge et al., 2022; Herrero et al., 2021, 2020; Lezoche et al., 2020; McFadden et al., 2022), this literature has not addressed the political economy behind how specific actors in agri-food GVC can shape the process of developing, scaling, and commercializing these innovations. This strand of literature explores the potential of several technologies to improve food systems but does not discuss how the organizational characteristics of agri-food GVCs may influence the evolution and final success of these promising technologies (Mac Clay and Sellare, 2022). Considering that the nexus between the literature on food value chains and industrial organization in the context of technological change has not received enough attention (Bellemare, 2022; Macchiavello et al., 2022), our paper seeks to fill this gap by exploring the links between technological innovations related to agri-food production, distribution, and consumption and the industry structure in agri-food GVCs. Specifically, we look at the entrepreneurial landscape of innovation in agri-food systems, covering the entire value chain (from farming inputs to last-mile delivery), and explore the potential influence of industry concentration in this technological transition. We use corporate venture capital investments to study the behavior of the large multinational firms that govern agri-food GVCs. With a focus on current cutting-edge technologies, our paper contributes to the rich body of literature that has studied structural transformations in agri-food GVCs (Barrett et al., 2022; Reardon et al., 2019; Reardon and Timmer, 2012; Sexton, 2013).

Our research questions in this paper are: (a) what are the leading technologies and solutions entrepreneurs are taking to the market in agri-food value chains? and (b) what type of investments are incumbent agri-food companies prioritizing as venture capitalists? Following the distinction by Zilberman et al. (2022) between innovation and product supply chains, the first question of our paper targets the innovation supply chain level: we explore the technological solutions that promise to improve agri-food GVCs. However, this cannot be understood isolated from the industrial dynamics of agri-food GVCs, which will determine the success of many of these innovations when moving toward the product supply chain. While many small entrepreneurs (i.e., start-up companies) can come up with new solutions, upscaling and making them suitable for massive markets requires the commitment of incumbent firms in the product supply chain. This is the motivation behind the second question, which explores the role of dominant corporations in front of these new technologies.

The primary data source for our analysis is Crunchbase (2022), a comprehensive database of highly innovative public and private companies. We propose two main goals in this paper. The first is to characterize the innovation landscape in the global agri-food systems based on the value propositions of emergent science- and tech-based firms. We do this by first creating a typology of the technological solutions that these firms are developing and then using machine learning for natural language processing to classify the firms accordingly. This approach is more informative and detailed than the standard industry classification (SIC) codes usually used to categorize industries and companies. Our second goal is to use this typology to explore the direction of multinational agri-food firms' investments from a corporate venture capital perspective and analyze the role these dominant firms play in the technological transformation of agri-food systems.

The remainder of this paper is structured as follows: Section 2 summarizes the process of industrial consolidation experienced by agri-food GVCs in the last decades and why this matters in the context of technological change. Section 3 presents the main characteristics of our database, and section 4 details our methodology. Section 5 presents the results of the classification process and the investment flows by dominant firms in each category, with a discussion of the general trends in the role of incumbent firms as venture capitalists. The last section of the paper includes the main conclusions and business and policy implications.

2. Innovation, sustainable transitions, and the governance of agri-food GVCs

In the face of many environmental and social challenges, the acceleration of technical change in digitalization, biotechnology, and artificial intelligence promises to improve the sustainability of agri-food GVCs by reducing the application of synthetic fertilizers and chemicals in agriculture (i.e., smart farming, farm robotics), improving crop yields with lower development costs (i.e., new gene-edited seeds), reducing the space needed to produce food (i.e., controlled environment agriculture), replacing the production of animal-based proteins (i.e., cellular agriculture, plant-based meats), enhancing food safety and traceability (i.e., blockchain technologies), reducing or reutilizing waste (i.e., biorefineries) and raising food accessibility and reach (i.e., e-commerce and last-mile delivery technologies), among many other promises.

However, technical change is not disconnected from the surrounding institutional environment (Geels and Schot, 2007). Innovations start at a niche level, and then the industrial, political, and social rules create the conditions for new technologies to scale and develop massively. In the last 30 years, agri-food value chains have experienced a process of structural transformation (Barrett et al., 2022; Reardon and Timmer, 2012). These structural transformations were part of a natural modernization process and adaptation to new demands and technological conditions. Still, at the same time, many of these changes have enhanced the role of multinational corporations in moulding the institutional setting and governance rules in every segment of GVCs. In a context where thousands of start-ups are driving innovation in agri-food systems (Klerkx and Villalobos, 2024), understanding industrial dynamics is necessary to elucidate whether technology can deliver the promise of improving agri-food GVCs.

In the upstream segment, we have seen an increase in the potential of biotechnology for the farming sector. Modern seed biotechnology has enhanced the complementarities between seed and agrochemicals (Deconinck, 2020). This race to exploit and monetize these complementarities happened in the context of increased costs and risks of deregulating new biotech events in seeds and agrochemicals (McElroy, 2004). As a consequence, we have seen in the last few decades a consolidation of the sector into a small group of companies (known as the big-six) capturing most of the market share and creating a cross-licensing scheme among them (Deconinck, 2019). This has conditioned the priorities regarding technology development, in which extensive crops such as cotton, maize, or soybean explain most of the traits developed since 1990 (Parisi et al., 2016). That has become even deeper in the last few years, and now the big six have become the big-four.[1] A similar (but somehow less pronounced) dynamic occurred in the agricultural machinery and equipment sector, in which the leading companies increased their market shares in the last quarter of the twentieth century. Significant economies of scale and the need for global manufacturing and distribution facilities have reinforced the leadership of multinational companies (Fuglie et al., 2011).

Agricultural commodity trading has also been consolidated in the last few decades (Clapp, 2015), in which a small group of multinational companies control the largest share of grain exports.[2] The Chinese strategy to move into grain trade and other recent corporate movements[3] have further consolidated the sector (Ballard, 2016; Kelloway, 2023). The first reason behind this process is the natural spatial configuration of trade in GVCs, which requires high investments in fixed

[1] This has happened through the acquisition of Monsanto by Bayer, the purchase of Syngenta by ChemChina and the merger between Dow and DuPont (now Corteva).

[2] These companies are known as the ABCD, in reference to the initials of ADM, Cargill, Bunge and Dreyfus.

[3] In 2017, the Chinese trading company COFCO completed the acquisition of Nidera. Recently, the companies Bunge and Viterra merged.

assets and infrastructure to process and transport grains. The second reason is related to the need for a solid financial structure for market hedging to deal with food price volatility, which has become more evident in the last 20 years mainly due to the nexus between food and energy markets and the massive irruption of investments funds in commodity markets (Salerno, 2017; Wright, 2011).

At the midstream level, many structural changes happened in food processing. These changes are related to changing consumption patterns and the appearance of new processed (and ultra-processed) food and ready-to-eat meals. As a result, farming activities have reduced the share in total food value-added against food processing and marketing (Canning, 2011; Yi et al., 2021). This has also led to a reorganization at the level of the food processing industries, with fewer and bigger actors acting as buyers at a global level (Muehlfeld et al., 2011; Ollinger et al., 2005). Moreover, consumers' demands for more information about their food have brought institutional innovations into the scene to improve food traceability and sustainability. Big food companies have worked to align the value chain by establishing stringent requirements for their suppliers in which private standards have played a relevant role (Lee et al., 2012; Meemken et al., 2021).

Finally, GVCs have also experienced structural changes in the downstream segment, close to final consumers. The expansion of supermarket chains across continents is known as the supermarket revolution (Reardon et al., 2012). Supermarkets have promoted new buying behavior by final consumers (Volpe and Boland, 2022) and have changed the dynamics of agri-food systems, developing a pattern of buyer-driven GVCs (Gereffi, 1994; Gereffi and Christian, 2009). More recently, COVID-19 has also accelerated the e-commerce trend for grocery and food stores, a trend that had been clear for apparel and consumer products but arrived later at agri-food systems (Reardon et al., 2021b).

The structural transformations described in this section explain the morphology of value chains as we know them today. Agri-food GVCs have experienced a consolidation process in most stages,[4] where global multinational corporations now play a central role in establishing governance patterns. Although it is out of the scope of this paper to understand whether this concentration process resulted in a practical exercise of market power, we have to acknowledge the important role that industry structure plays in the process of technical change, which is often presented as one of the pathways through which global agri-food value chains may tackle many sustainability challenges. The success of this set of innovative solutions that promise to revolutionize agri-food systems is highly dependent on industry dynamics since governance rules affect how learning and technology transmission take place in GVCs (Pietrobelli and Rabellotti, 2011). The innovation and technology selection patterns are not independent of industrial characteristics such as concentration rates, firm asymmetries, and market share variability (Dosi et al., 1995; Dosi and Nelson, 2010). Leader firms in agri-food GVCs can prioritize technology canons, control global distribution networks, establish production standards, and cater to consumers' preferences (Béné, 2022; Clapp, 2021).

This paper comprehensively analyzes technical change in agri-food GVCs, comprising every step in the value chain, from farming to final consumers. Agricultural economists have successfully studied dyadic relations in value chains (e.g., farmers-input suppliers, farmers-processing companies), but new methods and perspectives of analysis are needed when it comes to understanding the entire value chain (Bellemare, 2022). And this is especially true when studying innovation, which demands a prospective analysis that cannot rely easily on historical data. Even when we cannot make claims about how industry dynamics will evolve and whether new structural changes will happen, we can still explore and classify the main innovations taking place and delve into the role of agri-food incumbents, getting hints on their strategic intentions through corporate venture investments. The following section describes the data and methods in detail.

3. Data

This paper uses Crunchbase as its primary data source. Crunchbase (2022) is a comprehensive database of highly innovative public and private companies, including data on organizations, people, and investors. The main goal of the database is to facilitate the research process of corporate venture capitalists seeking to invest in young companies, as well as the interaction among founders and workers in those companies. The database includes general information about each company (i.e., country, city, address, date of foundation), a company description paragraph, tags that provide keywords synthesizing the company's activities in one or two words, and information on funding (e.g., number of funding rounds, total accumulated funding). It also has information on investors, acquisitions, and initial public offering of stocks (IPOs). Each company has a unique identification code (called *uuid*). Through these identifiers, it is possible to match, for example, a company with the funding rounds in which this company was involved or a funding round with the investors that participated in it.

The database is updated daily with new information through different channels: the Crunchbase data team that collects and validates information, entrepreneurs who upload information about their own companies, and investors who regularly submit updates of their portfolios (Crunchbase Product Team, 2022). The file for bulk export of the entire database in.csv format is updated daily. This paper used the database updated on January 2nd, 2023.

Crunchbase grants researchers free access to the full database and thus has been increasingly used for academic research, particularly in economics and business management (den Besten, 2020). Many recent papers use this database as its primary source to study different aspects related to technological change and sustainability transitions, such as the role of business ecosystems and networks in innovation (Marra et al., 2017; Tiba et al., 2021), the drivers of investments in new technologies (Holicka and Vinodrai, 2022; Kwon et al., 2018) or the role of different funding sources in the final success of startups (Bertoni and Tykvová, 2015; Croce et al., 2018). To the best of our knowledge, no papers have used Crunchbase with a specific focus on global agri-food GVCs.

As with other commercial databases on companies and venture capital, which are not designed exclusively for research but mainly for business purposes, the quality of Crunchbase coverage may vary across world regions and economic sectors (Breschi et al., 2018; den Besten, 2020). Regarding geographic coverage, firms based in the United States (US) may be better represented in the database in terms of the quantity and quality of the information (Crunchbase News, 2019). However, this is not necessarily a bias. Even when the investment landscape is changing fast, with more activity in emerging countries like India or China, the US is still indisputably the central hub of startup creation and venture capital. Half of the top global startup ecosystems are located in the US (Startup Startup Genome, 2022) and the country still gets around half of global venture capital funds (Florida and Hathaway, 2018; CB Insights, 2023a), with more than half of all "unicorn companies"[5] located in the US (CB Insights, 2023b). Moreover, even when some companies have been founded outside the US, recent trends show they seek to open facilities or legally establish themselves in the US to access talent, knowledge networks, and funding (Bucak, 2022; Dibner, 2020). In this sense, some companies may be listed in the US but have research labs, production facilities, or substantial commercial operations in other regions.

[4] We have some exceptions to this consolidation process. For example, in the midstream we still observe some local transport, logistics and processing activities that show low degrees of concentration but are highly dynamic (Reardon, 2015).

[5] Companies worth more than 1 billion USD.

Some studies have compared Crunchbase with competing databases and confirmed that Crunchbase is not significantly different (or even more comprehensive) than other sources. Dalle et al. (2017) compare this database with the OECD Entrepreneurship Financing Database. The authors do not find significant differences either in the evolution of the funding across time or the geographical distribution of funds. According to Breschi et al. (2018), Crunchbase coverage is exhaustive not only for OECD members but also for the main large emerging economies (i.e., Brazil, Russia, India, and China). This is confirmed by Ferrati and Muffatto (2020), who affirm that more than half of the companies in the database belong to 10 countries, which include the US, Canada, Brazil, India, China, and five European economies. According to these authors, there are companies from more than 206 different countries in Crunchbase.

Crunchbase covers primarily private companies, with an orientation toward startups. While there is no undisputed definition of what a startup is, we usually understand these types of companies as having an innovative profile and being scalable, with a promise of exponential growth for investors. (Connolly et al., 2018; Vergara and Barrett, 2023). In this sense, the database may not capture some small, local, or microentrepreneurs (Crunchbase News, 2019). Even acknowledging that Crunchbase is becoming a primary database for investors, and many firms want to upload information and have a presence in the database (Breschi et al., 2018), some smaller businesses may not be a target of venture capital so the owners may have fewer incentives to keep their information updated in the database. This could lead to some underrepresentation of this type of firm in the database. However, one of the advantages of Crunchbase is that it not only includes information on venture-backed companies but also companies that have not yet been funded, which are mostly highly incipient and small in size (Ferrati and Muffatto, 2020). Moreover, the comparative work done by Retterath and Braun (2020) shows that Crunchbase coverage is still better than other publicly available databases. The authors compare eight different databases of venture capital, including Crunchbase.[6] They analyze a series of funding rounds for more than a hundred companies and evaluate the accuracy of these databases in three dimensions: company data, information on founders, and funding rounds. Crunchbase is among the top-performing databases considering all the dimensions. Therefore, despite potential shortages, biases, and coverage variability in terms of regions and sectors described in this section, Crunchbase still allows us to identify and classify companies innovating in agri-food GVCs and map incumbent firms' investments in these companies. In the next section, we describe how we use the information in the database, explaining our methodology step by step.

4. Methods

4.1. Deductive stage

We start our methodological approach by listing the most prominent startups and young companies working with innovative technologies at different stages of agri-food GVCs. This set of around 500 companies (detailed in Supplementary Material 1) works as a gold standard that guides us in identifying all other firms in the dataset and then classifying them.[7] These companies were selected following specialized industry reports (AgFunder, 2022; Good Food Institute, 2021; Navarro et al., 2022; Forward Fooding, 2022; SVG Ventures, 2020),[8] which also provided information to classify these firms.

We then produced a preliminary-deductive typology of innovations in agri-food systems based on an exploratory literature review (mainly through Web of Science and Scopus), including reference snowballing to identify further relevant articles. We looked for papers discussing innovations and technologies in different segments of food value chains to summarize and consolidate them in our own typology of innovations. We did not focus on specific technologies but rather on "solutions," considering that every technological platform can become the building block for multiple solutions in different value chain stages.[9] The complete list of papers we consulted to build the typology is detailed in Supplementary Material 2. This preliminary typology went through several rounds of fine-tuning at later stages. We started with a more detailed classification of around 35 solutions and then merged the ones that were more similar, or that could be placed under a broader category. This fine-tuning process aimed to balance the need for a descriptive and informative typology (which demanded a higher number of categories) and the efficiency of the subsequent machine learning classification (which required reducing the number of categories). The final version of the typology is presented in Fig. 1 (See Appendix A for the details and definitions of each category).

4.2. Database query

Next, we ran a first query in Crunchbase to preliminarily identify companies working on solutions related to agri-food GVCs. The initial number of organizations in Crunchbase is above 2 million. In this step, we sought to reduce this number to capture companies providing solutions for agri-food systems (following the definitions presented in Section 4.1). We did this preliminary query for two reasons. The first goal was to facilitate algorithm training, having a more manageable number of companies in the manual labeling stage that ensures a balance between the target dataset for classification and the training dataset. Second, we did this to ensure the quality and homogeneity of the data in the classification stage over the "unseen" units (companies that do not belong to the training dataset). Both issues are crucial to reduce the variance in the data and minimize generalization error when moving from the training set to the actual classification process on unseen data (Halevy et al., 2009; Rajput et al., 2023).

Using the R package *corpustools* (Welbers and Van Atteveldt, 2022), this search combined two fields: company descriptions and tags (see Supplementary Material 3 for more details). In this first exploration of the Crunchbase database, we identified around 26,400 potentially relevant companies. We used the Gold Standard built in the previous step to validate the quality of this query (i.e., to ensure that we captured the type of company that is the target of our analysis). About 86 % of the companies listed in the Gold Standard appeared in the search, indicating that the query successfully brought us the type of company we were expecting in the first place. In addition, companies in the Gold Standard that did not appear in the query were added manually. The process is summarized in Fig. 2.

4.3. Data preparation and cleaning

For these 26,400 preselected companies, we combined the long and short description fields in Crunchbase and turned them into a text

[6] The other ones are Angellist, CB-Insights, Dealroom, Pitchbook, Preqin, Tracxn and VentureSource, which was recently acquired by CB-Insights.

[7] The condition to be included in the gold standard is that these companies had to be listed in the Crunchbase database at the moment of download (January 2nd, 2023).

[8] Among the industry reports that we used to build this list, we can mention the ones by AgFunder (2022), Forward Fooding (2022), SVG Ventures (2020), The Good Food Institute (2021) and The Yield Lab (Navarro et al., 2022).

[9] For example, the Internet of Things (IoT) can be the building block for crop sensors, packaging devices or kitchen appliances. Artificial Intelligence is being applied in farm robotics and other devices at the farming level but also for autonomous delivery vehicles in the downstream segment of the value chain.

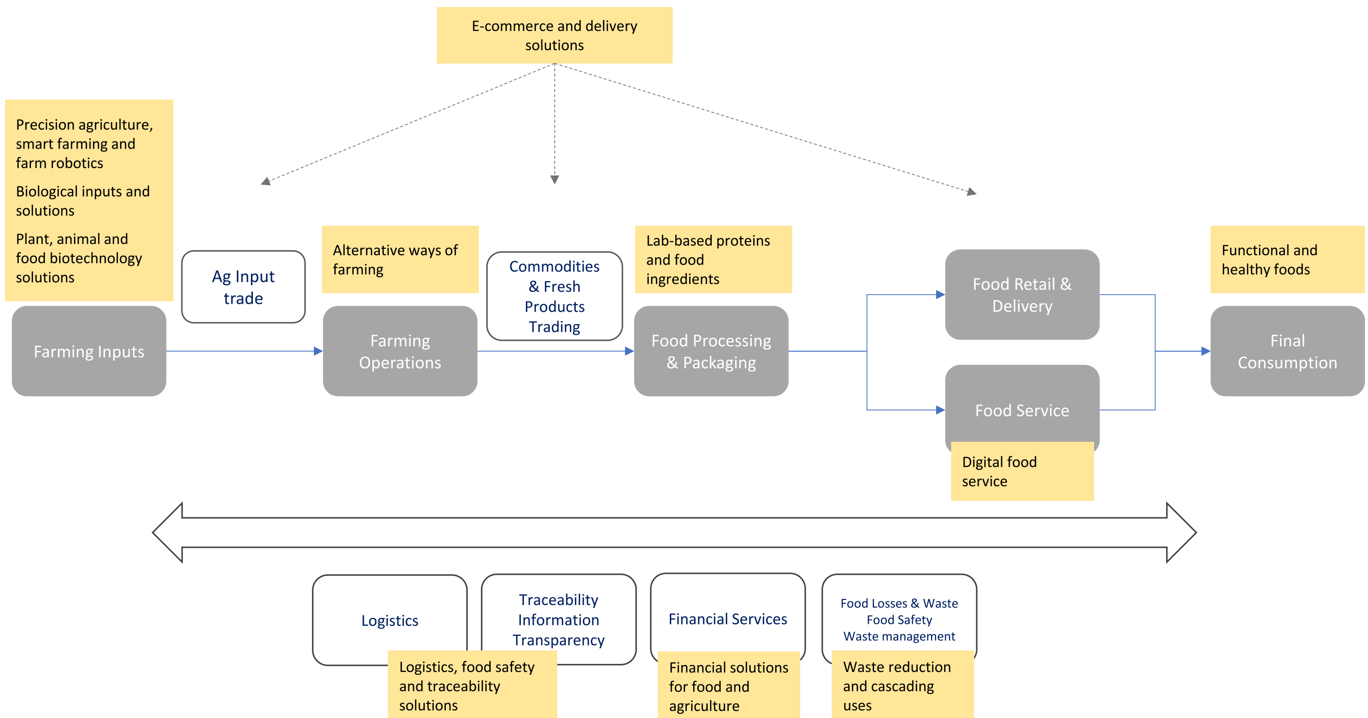


Fig. 1. Typology of innovations in an archetypal agri-food value chain. Eleven different types of innovative solutions are presented in the yellow squares. Gray boxes represent structural functions in a value chain. White boxes represent functions that do not belong to a specific stage but rather connect stages or are performed across stages.

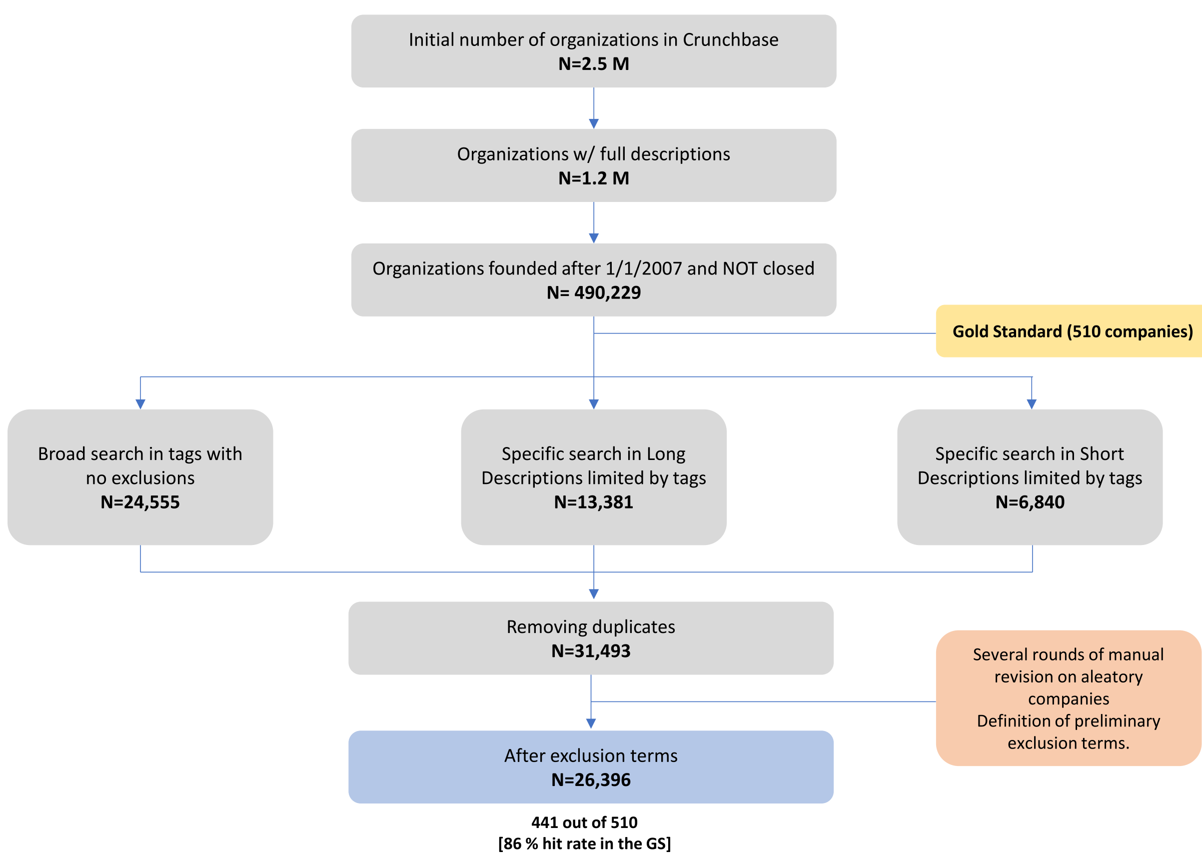


Fig. 2. Database Query.

corpus. Then, we tokenized and cleaned this text corpus (Benoit et al., 2018). Each token is a smaller piece of text, a sequence of characters that serve as units for the analysis (Manning et al., 2009). Tokens can be individual words, compound words, or even symbols, depending on the research goals. We worked with n-grams of size 1 or 2, removing symbols, numbers, punctuations, and URL addresses to turn the corpus into tokens. We also removed English stopwords[10] and tokens of less than three characters to reduce the noise in the text corpus. Every token was then turned to lowercase, and we kept only word stems.

Finally, we built a document-feature matrix (dfm), the base for the analysis, in which documents (i.e., in our case, company descriptions) are listed in rows while tokens are in columns. A dfm is a sparse matrix that shows the frequency for each token in each document. We weight the dfm by the term frequency-inverse document frequency (tf_idf) to reduce the weight of terms that are too frequent (and do not provide much information for the classification stage) and increase the weight of terms that are specific to a document and may provide valuable information (Manning et al., 2009). The tf_idf weighting follows the form:

$$tf\_idf_{t,d} = tf_{t,d}\left(\log\frac{N}{df_t}\right) \tag{1}$$

Where $tf_{t,d}$ is the frequency of term t in document d, N is the total number of documents and df_t is the total frequency of the term t in the text corpus (sum of appearances in all documents). Supplementary material 4 shows a detailed example of this process using the package *quanteda* (Benoit et al., 2018), from building the text corpus to the final weighted dfm matrix.

4.4. Algorithm training

We started the inductive process by manually classifying companies. First, we categorized the companies in the gold standard, and later classified 4,783 additional companies taken randomly from the Crunchbase query. Consequently, we reached a total of 5,293 companies manually labeled (around 20 % of the entire query) (Appendix B shows the final classification for this annotated dataset). This process was based exclusively on Crunchbase descriptions.[11] We also established criteria to exclude companies irrelevant to our classification despite being included in the initial query (Supplementary Material 5 for more details). Finally, we used some of the tools provided by the package *litsearchr* (Grames et al., 2019) to identify the most common compound words and added them manually to help train the algorithms.[12]

4.5. Classification

We used a set of machine learning models to classify companies according to the kind of technology they develop. We used the manually labeled group of 5,293 companies as our base to train the algorithms, with the final goal of accurately predicting the class membership of each company outside the manual review sample. We divided the classification problem into two parts. In the first stage, we worked with a binary classification problem, in which we wanted to automatically exclude companies irrelevant to the analysis despite being included in the initial query (Supplementary Material 5 shows a list of the types of companies that should be excluded). After this stage, we moved to a multilevel classification problem, in which we classified the surviving companies into the 11 categories presented in Fig. 1.

Not every machine learning algorithm works well in a multilevel classification problem with text data as input, so we selected those more suitable for this purpose. Benoit et al. (2018) include Naïve Bayes (NB) and Support Vector Machine (SVM) in the *quanteda* R package. Despite their simplicity, both NB and SPV are efficient for text data and in contexts of small annotated datasets (Riekert et al., 2021).

NB classification is based on Bayes' conditional probabilities theory and uses each token as a feature to predict the outcome. In this algorithm, the probability that a category c is assigned to document d is computed as follows (Frank and Bouckaert, 2006):

$$P(c|d) = \frac{P(c)\prod_{t\epsilon d}P(t|c)^{n_{td}}}{P(d)} \tag{2}$$

in which $P(t|c)$ is the probability of observing a token t given the category c, n_{td} reflects how many times a specific token t appears in the document d, $P(c)$ is the prior probability of category c (estimated by the proportion of documents that belong to category c in the training set), and $P(d)$ is a constant. This algorithm assumes that all predictors are independent and have equal effects on the outcome (Loukas, 2020), which are strong assumptions and not always realistic (this is why it is called "naïve"). Nevertheless, it is a simple and flexible option for machine learning classification and is often used efficiently with text data (Kowsari et al., 2019).

SVM is a classification algorithm used in linear and non-linear settings based on separating hyperplanes to split a set of observations into different categories (Noble, 2006). These hyperplanes officiate as decision lines, splitting points that belong to distinct classes. The points nearest to these division hyperplanes are called support vectors and become references to maximize the distance between categories. The SVM algorithm works based on the idea of structural risk minimization, which creates a classification hypothesis with the lowest error on unseen and randomly selected test data (Joachims, 1998). Additionally, the algorithm specifies a soft margin, allowing some observations to fall on the "wrong side" of the hyperplane (Noble, 2006). This is done to gain flexibility without substantially affecting the final result. The efficiency of this algorithm in dealing with a large set of features and its robustness against overfitting make it especially suitable for working with text data (Joachims, 1998; Kowsari et al., 2019).

Hvitfeldt and Silge (2022) also provide an example of regularized linear models for multilevel classification (LASSO) using the R package *tidymodels*. The LASSO estimates ($\widehat{\beta}$) are based on a quadratic minimization problem of the following form (Tibshirani, 1996):

$$\widehat{\beta} = \text{argmin}\left\{\sum_{i=1}^{N}\left(y_i - \sum_j \beta_j x_{ij}\right)^2\right\} s.t. \sum_j |\beta_j| \leq t \tag{3}$$

where y_i is the response variable, x_{ij} are the predictive features (tokens in our case), and t is the tuning parameter. This tuning parameter helps to shrink some regression coefficients (even turning some of them to 0) to retain only the most important features (Chintalapudi et al., 2022). Despite not being cutting-edge for natural language processing and showing potential drawbacks in terms of the selection of covariates, linear models are still helpful in practice due to their dimension-reducing capabilities and their ease of implementation and interpretation (Freijeiro-González et al., 2022).

We also used an unsupervised technique called topic modeling based

[10] Stopwords are very common terms that operate as connectors or auxiliary words. These words are used so often in the language, that do not add substantial information in natural language processing. Examples of stopwords in English are words like "I", "me", "we", "our", "were", "this", "that", "be" or "do" among many others. R packages oriented to text mining normally include a set of stopwords or the possibility to import it.

[11] We did not consult company websites, social networks, or any other external sources. This is done in order to train the algorithm using the same information that will be used in the classification stage (which is only based on Crunchbase descriptions).

[12] We included compound words that were found very often in each category, but also that meant something in terms of the analysis. For example, we included compound words such as "active ingredient*", "blockchain technolog*", "circular economy", "anaerobic digestion", "cellular agriculture", "plant-based meat", "indoor farm*", "synthetic biolog*", "gene edit*", "soil microbiom*", among many others.

on the Latent Dirichlet Allocation (LDA) algorithm to improve the algorithms' accuracy in the linear setting. LDA generates a series of probabilities for each document belonging to a specific topic or category (Blei et al., 2003). It is based on the underlying idea that each document is a mixture of topics (i.e., document #1 may be 70 % topic A and 30 % topic B), and each topic is characterized by a specific distribution of words (Robinson and Silge, 2017). This technique seeks first to identify the mixture of words behind each topic and then assigns a probability for each document belonging to each topic. We took these probability vectors and included them as input in the linear model to increase its efficiency.

The presented algorithms in this section have already been applied in previous studies dealing with the classification of text information in several fields (Kusumawati et al., 2019; Luo, 2021; Rabby and Berka, 2022; Riekert et al., 2021; Uddin et al., 2022). We compared the performance of these algorithms based on two of the most commonly used metrics in classification problems. The first is prediction accuracy, which measures the rate of correct predictions over total predictions:

$$Accuracy = \frac{True\,Positives + True\,Negatives}{True\,Positives + False\,Positives + True\,Negatives + False\,Negatives} \tag{4}$$

The second metric is the area under the Receiver Operating Characteristics curve (ROC AUC), which is the relationship between two measures:

$$True\ Positive\ Rate\ (TPR) = \frac{True\ Positives}{True\ Positives + False\ Negatives} \tag{5}$$

$$False\ Positive\ \ Rate\ (FPR) = \frac{False\ Positives}{False\ Positives + True\ Negatives} \tag{6}$$

The ROC curve plots the TPR against the FPR at different probabilistic threshold levels, and the ROC AUC is a measure of the area under that curve (James et al., 2013). The closer to 1, the better the algorithm works. A value of 0.5 shows that the algorithm is doing no better than a random assignation.

Crunchbase descriptions show considerable variability in terms of the depth and quality of the information included. While some companies have comprehensive details of what they do, their goals, and which value chain actors they are targeting, other descriptions are short and do not allow a straightforward interpretation of what the company does (see Supplementary Material 6 for specific examples). Appendix D presents a word cloud and a co-occurrence network for the most common tokens in our text corpus. Given the nature and size of our data and considering that each machine learning model prioritizes different parameters and is likely to classify companies differently, we used and compared different models to check how well they agree. Thus, we made the final classification based on a multi-model agreement, with manual control in various stages, to reduce the chance of misclassification. In the results section, we present the details of the final classification.

Since the advent of neural networks in machine learning, deep-learning tools have advanced rapidly in several fields (Vaswani et al., 2017). During 2023, there was a massive popularization of Large-Language Models (LLMs) based on deep-learning neural networks, which can recognize, interpret, and elaborate responses based on human language. The flexibility and simplicity of access and use of these tools[13] make them suitable for various purposes, including classification tasks. In the last year, many academic works have incorporated LLMs (ChatGPT in particular) to solve classification problems (An et al., 2023; Mellon et al., 2023; Zhao et al., 2023). Despite its considerable potential and promising applications, the novelty of the tool and its variability in terms of results still demand caution in its use (Ollion et al., 2023). As a robustness check, to validate the quality of our classification, we used the *openai* package (Rudnytskyi, 2023) to access the Open AI application program interface (API) and perform the classification task using GPT-3.5 (OpenAI, 2023).

4.6. Analysis of venture capital flows

Finally, based on the reports by IPES (2017) and ETC Group (2022), we listed a group of incumbent industry leaders in agri-food systems. These companies have the highest market shares and dominate each stage of agri-food GVCs. Many young, innovative firms may undergo several years without genuine income before monetization starts, so they rely on venture capital inflows to support the transition (Kaul, 2021). Large corporate investments are one of the possible sources these firms can use, which also constitutes an opportunity for incumbents to participate in smaller (and sometimes more dynamic) science- and tech-based firms.

Based on our previously defined typology and classification, we analyzed the funding rounds in which these companies have been involved as investors to describe the direction of their corporate investments.[14] These companies belong to different industries, such as agrochemicals and seed, veterinary pharma, farm machinery, commodity trading, food and beverage, meat and proteins, and grocery retail (Appendix C shows the complete list of incumbents considered in our analysis). This way, we can understand the strategic focus of these companies as investors, exploring which categories in our typology they are addressing with more interest.

5. Results and discussion

5.1. Classification of companies

The purpose of the classification process was to arrive at a set of companies considered relevant in terms of our research goal (i.e., companies that create innovations in different stages of the agri-food value chains), each of them assigned into one of the 11 categories that we proposed in our typology of solutions. In the first step of the classification problem, we started with 26,465 companies. Of this group, 5,293 were manually classified (in this case, as either relevant or non-relevant) and operated as the training set. We selected the three algorithms with the best performances (NB, SVM, and LASSO plus topics from the LDA analysis) and used these algorithms both for the binary and the further multiclass steps.[15] Table 1 summarizes the performance measures for these three algorithms, which outperform the null model.[16]

Out of the 21,172 companies in the classification set, the three algorithms matched for 15,872 cases. In 5,300 cases, two algorithms agreed, while the third did not. In these cases, the companies were classified according to the agreeing algorithms. Fig. 3 summarizes this process.

After the first classification stage, 15,560 companies were considered relevant to continue in the analysis and moved to the next multilevel

[13] The most popular and massively adopted tool in the LLM sphere is ChatGPT, a free artificial intelligence platform developed by the company OpenAI (https://chat.openai.com/).

[14] We do not capture innovations that incumbent firms may pursue within their internal R&D departments or through alliances or partnerships with other firms or research institutions. These innovation efforts certainly contribute toward improving the sustainability of agri-food systems but are outside the scope of this work.

[15] We also tried more sophisticated machine learning models seeking to improve the accuracy in the classification. However, these performed worse than the three algorithms selected. Appendix E summarizes the results for all these other models and compares them with the three models that were selected for the analysis.

[16] The null model represents a random classification of the companies.

Table 1
Results for the classification problems (binary and multiclass).

	Binary stage				Multiclass stage			
	Null model	NB	SVM	LASSO + Topics LDA	Null model	NB	SVM	LASSO + Topics LDA
		quanteda	*quanteda*	*tidymodels*		*quanteda*	*quanteda*	*tidymodels*
Accuracy	0.570	0.806	0.795	0.776	0.160	0.731	0.750	0.753
ROC AUC	0.500	0.787	0.783	0.856	0.500	0.784	0.803	0.955
Training set size	5293 companies				3056 companies			

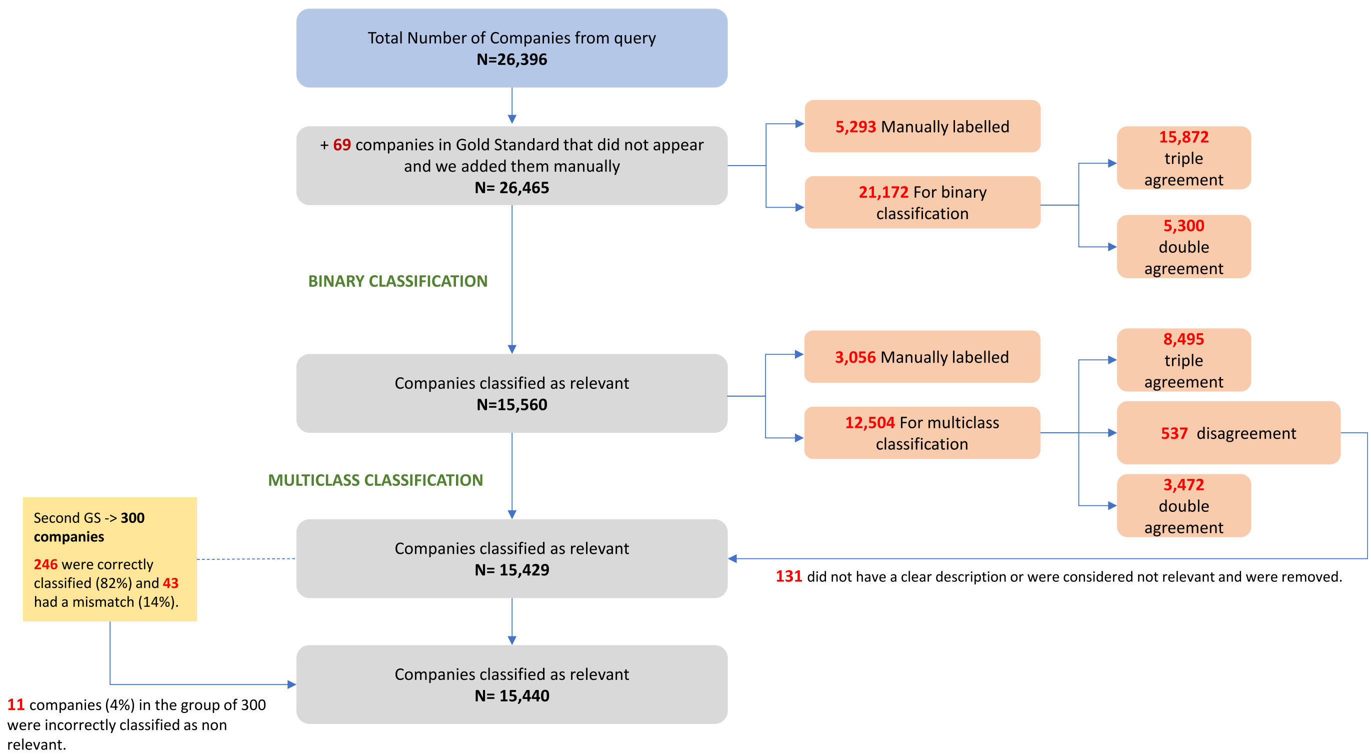


Fig. 3. Classification of companies based on different machine learning algorithms. Details of the control process.

classification problem. Out of this set, 3,056 were already classified in the manually annotated group, whereas 12,504 were the target for classification. We applied the same three algorithms as in the first-stage classification problem (the results are in Table 1). The results of the three selected algorithms (NB, SVM, and LASSO plus LDA probabilities) show that the accuracy rates are above 0.7 but do not go beyond 0.75. Considering the particularities of our text corpus and that the accuracy rate in the null model (i.e., random classification) is around 0.16, we can say that the algorithms are doing a fair job despite potential mistakes in the final classification. Moreover, the performance levels of our algorithms are comparable to those found in other studies that use similar approaches (Qureshi and Sabih, 2021; Rabby and Berka, 2022; Sebők and Kacsuk, 2021).

Out of the 12,504 target companies, 8,495 had a triple agreement between the three algorithms. In 3,472 cases, there was a double agreement and one disagreement, so these companies were classified according to the agreeing algorithms. A number of 537 companies presented a triple conflict (each algorithm placed them under a different category) and were manually reviewed and classified, removing 131 due to incomplete or inaccurate descriptions. We also created a second manually classified gold standard based on industry reports. We classified these companies according to Crunchbase descriptions and checked with these external sources to validate the classification status. This gave us an additional stage of control for the automated process. Out of the 300 companies in this second gold standard, 82 % were accurately classified.[17] The fact that we are simultaneously applying multiple algorithms does not eliminate the possibility of misclassification but helps reduce the individual error rate of each algorithm. The final classified set for analyzing investments and funding rounds comprises 15,440 companies.

5.2. The innovation landscape in agri-food global value chains

Regarding our first research question, which seeks to understand the landscape of innovations in agri-food GVCs, we notice that most companies in our sample (and most of the investment funds) flow toward downstream segments, more related to final consumption. Table 2 shows that around 60 % of the companies in our sample belong to downstream activities such as digital food service, functional and healthy foods, and e-commerce solutions. These activities are more

[17] We manually corrected the ones that were misclassified (14% of the total) and reintroduced 11 companies (4%) that had been incorrectly excluded in the first stage of the classification problem.

Table 2
Final classification of the relevant set of companies.

Class	n (%)
Digital food service	3,225 (21 %)
Functional and healthy foods	3,109 (20 %)
E-commerce and delivery solutions	2,791 (18 %)
Precision agriculture, smart farming and farm robotics	2,072 (13 %)
Alternative ways of farming	1,143 (7.4 %)
Logistics food safety and traceability solutions	950 (6.2 %)
Waste reduction and cascading uses	910 (5.9 %)
Lab-based proteins and food ingredients	479 (3.1 %)
Plant, animal and food biotechnology	408 (2.6 %)
Financial solutions for food and agriculture	182 (1.2 %)
Biological inputs and solutions	171 (1.1 %)
Total	**n = 15,440**

related to the final consumer and the last-mile stage of value chains. Then, around 24 % of the companies belong to activities related to the initial stages of the value chain and the farmer (precision agriculture, alternative ways of farming, biotechnology, and biological inputs). Finally, 16 % of companies belong to midstream segments or activities that provide solutions across the value chains (which include food processing, logistics, waste management, and financial solutions).

We ran a robustness check over these results using the OpenAI API based on GPT 3.5 to see whether a classification based on an LLM performs differently. We found no substantial differences in the distribution of the categories (see Appendix F for the details). In Appendix G, we also include an analysis of the geographic distribution of this final set of companies (in terms of the number of companies and funds). As expected, a large share of the companies in the sample are based in the US (35 % of all companies). Other emerging economies like India, China, Brazil, and several OECD economies (e.g., the UK, Canada, Italy, and Israel) also count with a large number of companies in the sample.

Fig. 4 shows the accumulated and average funding (in millions of dollars) in each category. E-commerce solutions are clearly at the top of both rankings. These companies, mostly oriented to the last-mile delivery stage, are raising the highest interest from investors and capturing most of the funds. Digital food service is the second category for accumulated investments in funding rounds. Upstream-oriented technologies such as biotechnology, biological inputs, precision agriculture, and smart farming are still behind. Moreover, upstream and downstream startups seem to show noticeable differences in scale: the size of a medium or small funding round in the segment of e-commerce or digital food service may even outperform the largest financing round in upstream companies, such as seed biotech or farming technology (AgFunder, 2022).

Three main reasons explain these differences. The first relates to technological maturity: solutions for the downstream segments lean on building blocks (i.e., digital technologies for mobile phones) that are more mature than other technologies in the upstream segment, such as farming robotics, bioinputs, or alternative proteins. In this sense, technologies downstream in the value chain appear as the low-hanging fruit for investment funds, with lower risks, faster investment returns, and more startups offering opportunities to channel funds. A second reason is that the potential of digital technologies to boost sales and provide quick access to massive markets in the downstream segment is high compared to other upstream technologies, in which the challenges for developing markets are multiple (e.g., selling new technology to farmers entails complex technical and commercial challenges compared to selling groceries to final consumers in a medium or large city). Finally, a third reason is linked to regulatory aspects. Many novel technologies around farming still hold several regulatory challenges, which add additional layers of uncertainty and generate caution among investors. These challenges include critical definitions that are still in dispute, such as in the case of new gene-editing technologies (FAO, 2022), different speed and regulatory approaches between countries, which is noticeable in the case of biological inputs (Caldwell and Fife, 2023), and issues related to confidentiality in data sharing, highly discussed in artificial intelligence applications for agriculture (Gardezi et al., 2023).

This initial picture of the innovations in the agri-food systems leads to a question: are innovation efforts and investment funds flowing toward the value chain segment with the highest potential for improving sustainability in GVCs? Venture capital is crucial to support innovative startups but, as reflected in Table 2 and Fig. 4, it also presents biases toward innovations that fit the requirements of investors for fast and relatively secure exits through IPOs or acquisitions (Lerner and Nanda, 2020). Companies related to e-commerce, last-mile delivery, and restaurant technology more easily fulfill those investors' requirements.

The potential of downstream technologies for improving the environmental sustainability in agri-food systems seems limited. For example, food delivery technologies promise to reduce energy use and GHG emissions compared to individual research trips. However, the evidence for a solid comparison of the mitigation potential with a baseline scenario is still scarce (Bunge et al., 2022). On the other hand,

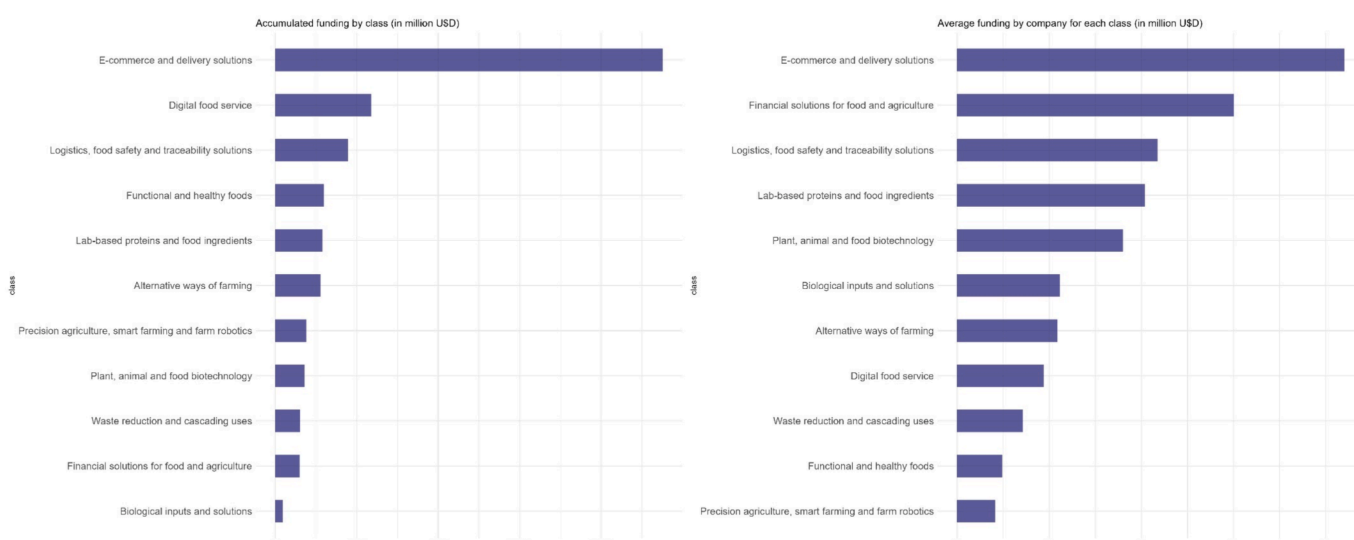


Fig. 4. Accumulated funding and average funding in each category.

technologies that support upstream activities for more sustainable food and biomass production (either by increasing yields or enhancing carbon capture) hold a high mitigation potential (IPCC, 2023), considering that activities related to the farming stage explain around 61 % of the anthropogenic emissions created by food supply chains and are also the primary source of other environmental adverse effects, such as eutrophication and acidification (Poore and Nemecek, 2018).

Upstream technologies oriented to the more efficient use of space, such as replacing animal proteins and producing in controlled environments and management technologies for producing livestock under more regenerative practices (combined with afforestation or reforestation) promise considerable carbon offsets and hold a mitigation potential (Bunge et al., 2022; Costa et al., 2022). Similarly, digital technologies at the farming level improve weed and insect control, which could imply substantial reductions in the application of synthetic herbicides and fertilizers (Finger, 2023). Finally, investments in new seed hybrids to reduce abiotic stress in crops (i.e., drought or severe temperature variations) seem crucial, considering the projected impacts of climate change on agricultural yields (Wing et al., 2021). The evidence shows that investments in technology to improve agricultural productivity at the upstream level hold considerable potential to improve welfare, particularly in developing countries (Fuglie and Echeverria, 2024).

Midstream technologies, including activities related to transportation, storage, wholesale, and food processing, have the potential to improve the quality of agri-food GVC, especially in developing economies where there is a substantial margin for capital intensification of many of these activities (Reardon et al., 2021c; Reardon and Vos, 2023). This broad sector of intermediary activities is usually composed of small firms and fragmented in market size (Ambler et al., 2023). Given this proliferation of small and medium companies in this midstream segment, these are likely to be underrepresented in our database, as explained in Section 3. This could explain the lower number of companies in midstream activities compared to upstream and downstream stages. However, looking at Fig. 4, we see that this midstream segment shows attractiveness for investment funds: two typically midstream categories such as logistics (that includes transport, wholesaling, storage, and traceability) and lab-based proteins and food ingredients (closely related to food processing and the food industry) are the third and fifth category respectively in terms of total investment funds and are the third and fourth category in terms of the average funding size.

5.3. *Investments by agri-food incumbent firms: Rationale and main strategies*

Our second research question was related to the role of agri-food multinational corporations as investors in technological firms. In this subsection, we describe the direction of incumbents' corporate investments and explain their main strategies and the rationale driving their behavior. The approximate size of the investments by this group of firms reached around 4.2 billion dollars[18] in the last 15 years, which is small compared with the total size of approximately 220 billion dollars[19] in our sample. The leading investors in our final selected group of companies are not incumbent firms but private equity funds, investment management firms, sovereign funds, and banks. These investors target startups entering a growth stage with proven products or services and markets developed. Despite not being the leading investors, the investments by dominant firms have an interest given their strategic nature, considering these multinational firms have substantial market power to set the rules in agri-food GVCs.

In Fig. 5, we summarized the direction of the investments from the incumbents' sector to each of the 11 categories in our classification. In this figure, we have pondered each investment according to the size of the funding round they belong to, and we notice that e-commerce has the largest share. Fig. 6 is similar but without ponderation, showing the results regarding the number of investments regardless of their size. This analysis is essential since many investments may not be significant in terms of money but are strategically relevant. Both figures help describe the landscape of incumbents' corporate venture capital movements. As Supplementary Material 7, we include the supporting tables for these figures, with the values for each of the flows.

Fig. 5 shows that investments by grocery retail corporations in the e-commerce and delivery segment look clearly more prominent than the rest of the flows. This is related to the substantial investments made by Walmart in developing e-commerce activities in India. The most significant investments were in the equity rounds in Flipkart (later acquired) and Ninjacart. These examples belong to a set of pivoting strategies that Covid-19 brought to the main scene, in which traditional companies owning physical facilities have turned massively into digital operations (Reardon et al., 2021a, 2021b). Other big supermarket chains have invested in e-commerce and logistics companies, such as Rewe in Flink or Kroger in Nuro. Although smaller in size, we also see a relatively high number of deals involving agricultural input and commodity trading companies in technologies that will improve their ways of commercializing their products and increasing their base of customers, mainly agribusiness marketplaces based on e-commerce technologies (see Fig. 6). For example, the synthetic fertilizer company Yara, the agricultural input leader Syngenta, and the commodity trading company Bunge are among the investors in the agribusiness marketplace Agrofy. Commodity trading companies Cargill, ADM, and LDC are investors in the Brazilian Grão Direto, where Bayer is also participating. These strategies are part of what we have defined as ***upgrading strategies***, when incumbent firms invest in technologies that may likely boost or streamline their core business operations. These strategies represent around 57 % of the investments by incumbents in startups. Fig. 7 shows the main strategic trends that we identify in the incumbents' investment behavior and that we will describe in this subsection. Appendix H includes a detailed list and examples of these strategies.

The logic behind upgrading strategies is that incumbents invest in smaller actors with higher innovation rates instead of doing research and development internally. By doing this, dominant firms secure a future option for introducing those technologies in their companies and avoid circumscribing their innovation process to their own capacities (Dushnitsky and Lenox, 2005). We can mention another two salient cases of upgrading strategies (not in terms of size, but in terms of the number of deals). One is related to agrochemical and seed companies investing in precision agriculture to provide additional services to farmers based on more efficient and personalized applications (e.g., investments by BASF in farm robotics companies such as EAVision and Ecorobotix or Bayer has invested in the carbon mapping company EarthOptics). The second case is traditional food companies investing in new functional and enhanced foods (e.g., investments by Danone in companies like Harmless Harvest or Laird Superfood or Coca-Cola in the nutrition and hydration company BodyArmor).

Fig. 6 shows that another segment that has captured the attention of investors is lab-based proteins and food ingredients. Meat, veterinary, and seafood companies are investing in plant-based and cellular agriculture technologies. For instance, three leading players in the beef

[18] This is an approximate number that comes assigning each investor the same share of a funding round (total of the funding round divided by the number of investors registered in the funding round). This is a strong but necessary assumption since Crunchbase does not provide information about the individual share of each investor in the funding round.

[19] This is an approximate number considering all those funding rounds that have information in terms of quantity in Crunchbase. Part of the funding rounds are registered in the database but do not present information of their size.

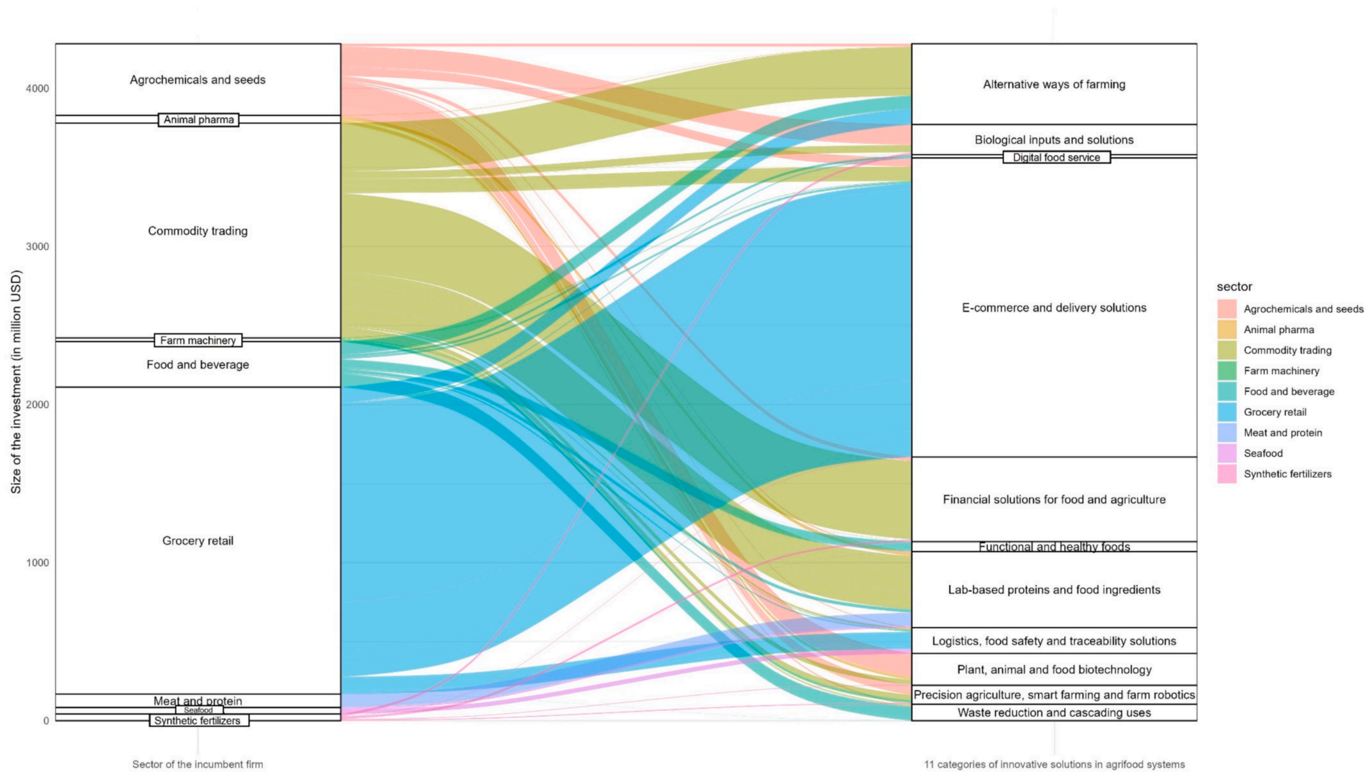


Fig. 5. The direction of investments by incumbents (pondered by the average size of the funding round). In the left column, we have the sector of the incumbent firm, while in the right column, we list the 11 categories of innovative solutions that we use in this paper. The relative size of the flows represents the magnitude of the investment (in million USD).

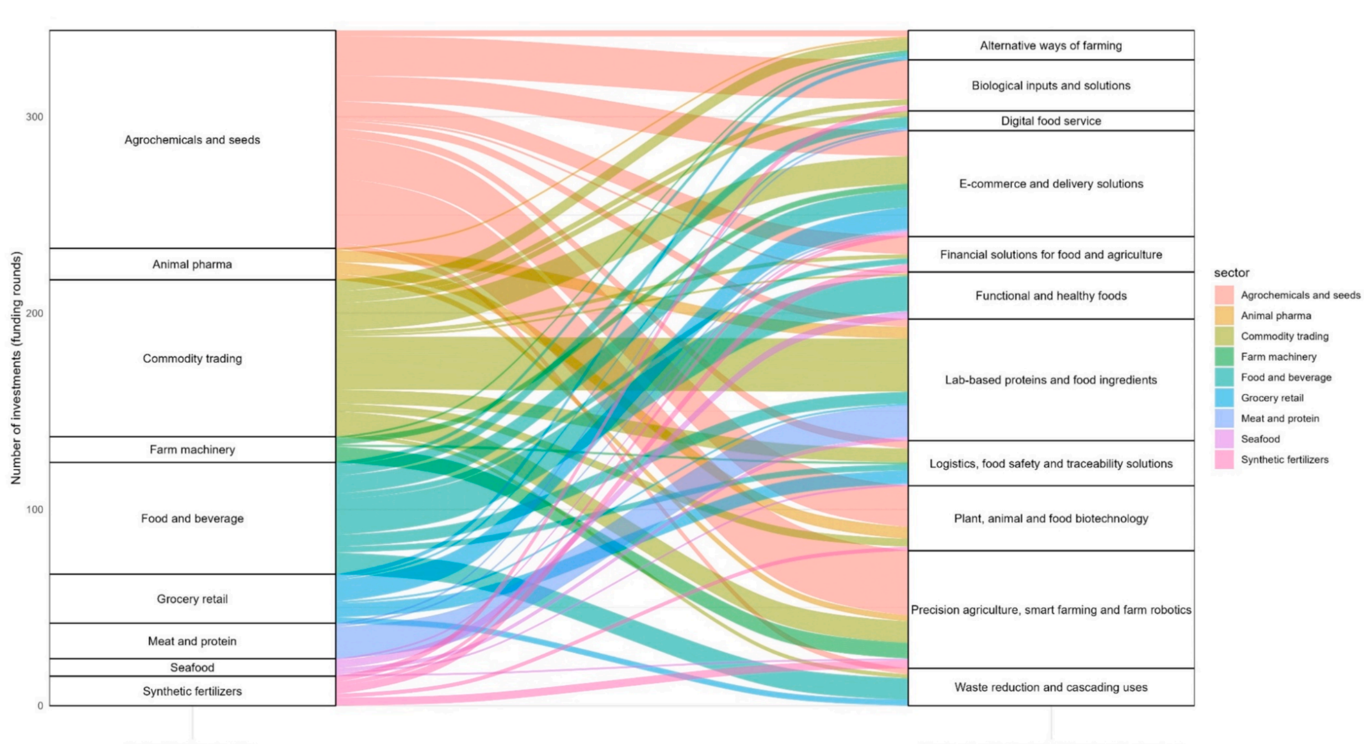


Fig. 6. The direction of investments by incumbents (without ponderation, only in number of investments). In the left column, we have the sector of the incumbent firm, while in the right column, we list the 11 categories of innovative solutions that we use in this paper. The flows represent the number of investments (i.e., funding rounds in which one of the incumbent firms acted as an investor).

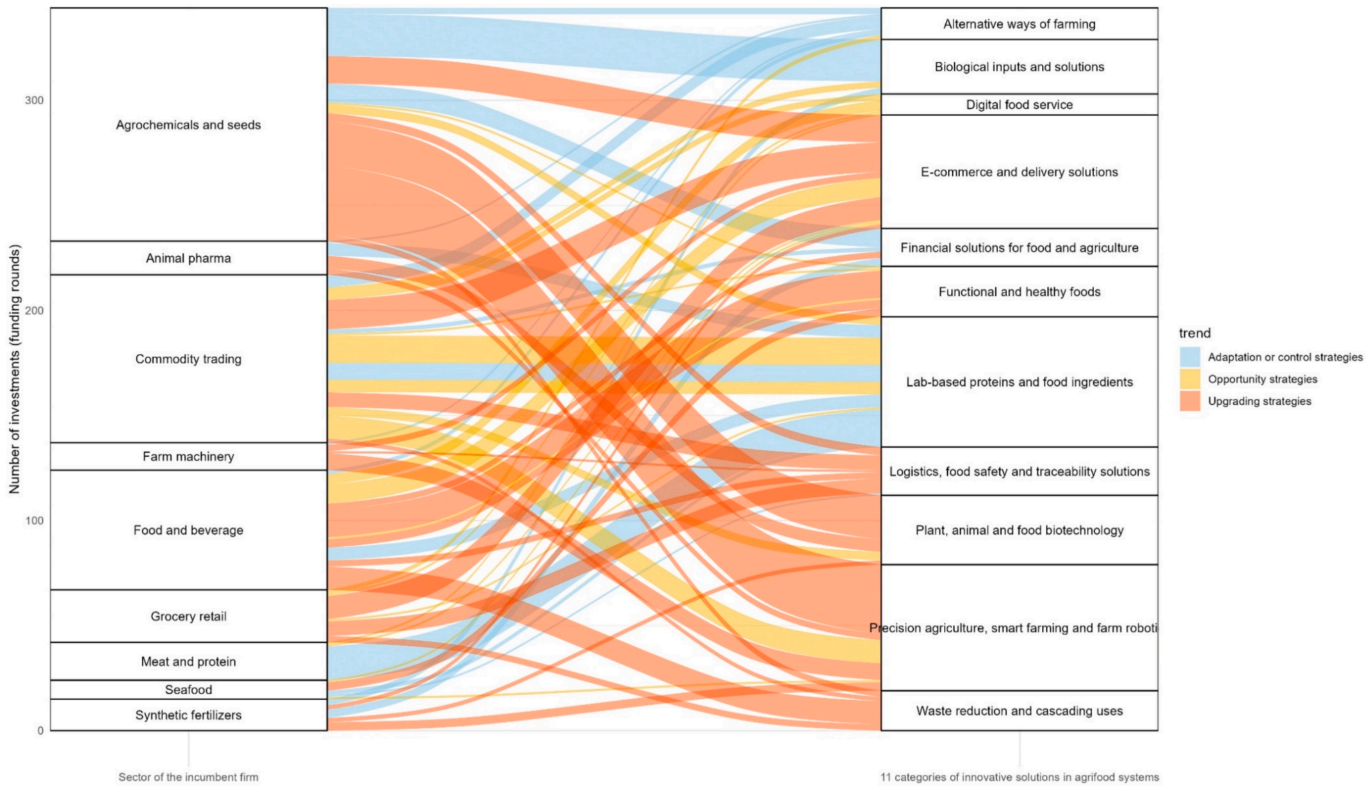


Fig. 7. Three trends of investments by dominant firms: upgrading strategies, adaptation or control strategies, and opportunity strategies. In the left column, we have the sector of the incumbent firm, while in the right column, we list the 11 categories of innovative solutions that we use in this paper. The flows represent the number of investments.

packing industry, Tyson Foods, Cargill,[20] and BRF, are among the leading investors in alternative protein companies like UPSIDE Foods, Calysta, Aleph Farms, Myco Technologies, and Believer Meats. These strategies are illustrative of what we defined as ***adaptation or control strategies.*** This second group of investment strategies (around 31 % of the total) is related to dominant firms funding technologies that compete and may (potentially) replace their core business. These technologies have a disruptive potential over the incumbents' core businesses and may force them to eventually re-arrange their business strategies. We use the broad name "adaptation or control" because, in some cases, this might be related to the desire to adapt to new consumer demands and requirements for healthier and less environmentally-harmful products, while in other cases, the underlying reasoning may be controlling the pace and scope of development of these potentially competing technologies. A second example within this group of strategies is traditional players in the synthetic ag-input industry investing in biological inputs, which are less standardized principles based on natural compounds and microorganisms (e.g., Bayer investing in Joyn Bio, BASF in Provivi and Groundwork or Sumimoto in TerraMera, among many others).

A third group of strategies comprises investment deals toward indirectly related sectors (or even entirely unrelated). We call these ***opportunity strategies*** since incumbents decide to invest in these companies, foreseeing they will have a dynamic development in the future, but are not explicitly related to their core business. They are seizing opportunities to increase their strategic portfolio's revenue without necessarily looking to incorporate any of these technologies into their operations. Fig. 7 shows different trends in this regard. For example, commodity trading companies have been considerably active in investing in different businesses such as cellular agriculture and plant-based, biological inputs or precision agriculture, which are only indirectly related to their core operation in agri-food value chains (e.g., Bunge in PivotBio, MycoTechnologies or Seedcorp, LDC in Motif, Gathered Foods and ADM in Believer Meats and New Culture). Something similar happens with retail grocery firms' investments in controlled environment agriculture or food and beverage companies in digital food service.

6. Conclusions

Global agricultural value chains have undergone several structural transformations in the last three decades, in which global multinational corporations have acquired a central role in establishing technological and commercial standards. These structural transformations have taken place from the beginning (i.e., farming inputs) to the end (i.e., food consumers) of agri-food value chains. Now, food systems are experiencing a large wave of technological change founded in biology, artificial intelligence, and digitalization driven by young emergent science- and tech-based companies. Technological change promises to improve agri-food GVCs in the face of the multiple environmental and social challenges ahead, and novel products and processes are incipiently transforming food production, processing, and distribution. Whether

[20] In Fig. 5 and Fig. 6 Cargill is shown within the Commodity Trading segment, which is its main activity. However, Cargill is also one of the main worldwide players in the beef packing industry.

these technologies will succeed in reaching markets and delivering the promises of transforming food systems largely depends on the business strategies of multinational firms that lead agri-food GVCs.

In this paper, we first presented a comprehensive overview of the innovation landscape in different stages of agri-food GVCs, from farming inputs to last-mile delivery. This comprehensive set of new technological solutions holds promises in every segment of global value chains. Land use change explains most of the GHG emissions from agricultural production (Alexander et al., 2015), so solutions in the upstream segment aimed at increasing agricultural productivity and streamlining space in agriculture hold substantial mitigation potential. Technological change in midstream segments such as transportation, logistics, and traceability has the potential to improve sustainability in agrifood GVC, considering how emissions from domestic food transport and food disposal have increased since 1990, especially in developing economies (Tubiello et al., 2021). Downstream technologies, such as e-commerce, have played a significant role in mitigating value chain disruptions during the peak of Covid-19 (Chenarides et al., 2021) and can also contribute toward improving the nutritional quality of households through better food accessibility (Shen et al., 2023).

However, it seems a daunting challenge for the many young companies developing these promising technologies to cover every step successfully, from scientific discovery to markets. The risks involved in research and development, the financial requirements to manufacture and commercialize new products, and the challenge regarding distribution channels and market intelligence may likely grant a relevant role to incumbent firms in the long run. From a Shumpeterian perspective, innovation in agri-food GVCs is happening at two levels: small firms and entrepreneurs propose new solutions at the discovery level, whereas big firms are needed to upscale these innovations and carry them to massive markets. In this paper, we analyzed the role of large multinational companies dominating agri-food systems and showed that these companies are actively exploring new technologies developed by smaller and more technologically dynamic firms. Many dominant agrochemical, seed, and fertilizer firms use these investments to move from selling standardized inputs to offering personalized services for farmers (Birner et al., 2021). At the same time, many e-commerce and agribusiness marketplace solutions will contribute to the disintermediation of food value chains, helping reduce transaction costs but also facilitating dominant firms access to farmers and final consumers. This improved market reach will likely help increase their sales and market shares, reinforcing industrial consolidation.

Our research has implications for policymakers and business managers. From a policy perspective, the first point to consider is that although substantial investments in technology are needed to reduce environmental impacts and improve diets, there are also risks involved in this process. The displacement of the smallest actors, technology accessibility for consumers, and uneven distribution of innovation rents between world regions are issues in a technological transition. Technological innovation must be coupled with changes at the policy level especially in a context in which corporate influence may create institutional lock-ins that lead to replicating many of the current problems in food value chains. A second point from a policy perspective is that stewardship strategies are needed to stimulate technologies with promising environmental or social sustainability potential. Venture capital is necessary for innovation, but investors tend to be risk-averse and seek the low-hanging fruit that provides fast and attractive returns. As we saw in our analysis, most of the innovation efforts and funds are channeled toward the downstream segment of the value chain (e.g., e-commerce solutions, in which technologies are more mature and have more profit potential). As such, many initiatives with a strong GHG mitigation potential upstream in the value chain (e.g., digital farming, crop biotechnology, or alternative proteins) receive less attention since they present more technical and commercial risks and require a longer timeframe to show investment returns. The idea of a mission-oriented approach, with articulation between public and private initiatives, could favor the crowding in from private funds in the long run (Mazzucato, 2022).

Regarding business implications, our study presents a complete landscape of what is happening in the entrepreneurial world in agri-food systems. This is essential to understanding the direction of technology investments and identifying innovation gaps. While many small and medium technological firms are leading the transition toward more sustainable food systems, incumbent firms seem to play a critical role in the transition from lab to market. The more disruptive a technology is, the more challenges developers face regarding organizational capabilities, technical skills, and value chain building. Small venture capital investors, such as seed funds, startup accelerators, science venture funds, and government agencies, play a critical role when looking at the problem from a network perspective. These actors create innovation ecosystems by placing smaller investments in many companies, which leads to sharing information and lessons from the entrepreneurial experience. In the long run, this helps to accelerate innovation while minimizing failure rates among young firms (Kulkov et al., 2020).

This paper is a first approximation to describing the changing sociotechnical regime in agri-food value chains. The main contribution of our work is to introduce an industrial organization perspective to the food value chain literature in the context of technological change. Despite presenting a comprehensive map of the landscape of innovations and explaining the role of incumbents, we do not measure how concentration or market power rates are changing in agri-food systems. Given the novelty of the problem we are studying, we present some preliminary insights regarding the role of corporate venture capital without offering an impact assessment of these corporate strategies. Moreover, this paper does not address the potential consequences of industrial concentration on farmers and consumers. Both farmers and consumers are the weakest actors in the value chain but also the critical ones in adopting many of the technologies we have discussed in this paper. Further research should seek to provide empirical evidence on how the role of incumbent firms may foster or hinder the market reach of these technologies and whether there are welfare impacts in terms of prices and accessibility for farmers and consumers.

CRediT authorship contribution statement

Pablo Mac Clay: Conceptualization, Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. **Roberto Feeney:** Conceptualization, Supervision, Writing – review & editing. **Jorge Sellare:** Conceptualization, Methodology, Project administration, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was funded by the German Federal Ministry of Food and Agriculture (BMEL), grant number 2219NR144. We want to thank Crunchbase for granting a researcher full access to the database, without which we would not have been able to carry out this paper. We want to especially thank Tomás Peña, María José Soler, Guillermo Bort, Pablo Feldman and José Gobbeé for sharing with us their views and knowledge on venture capital in the agri-food system in a very open and generous way. Their inputs were of great help for the interpretation of our results. Finally, we thank Hendrik Zeddies for commenting on an earlier manuscript version.

Appendix A. Typology of solutions. Details

Solution	Details
Precision agriculture, smart farming, and farm robotics	Data collection, imagery, and information at the farm level: Crop sensors integrating IoT, satellite and drone imagery, yield monitors, telemetry systems, soil maps, and yield maps.
	Precision Agriculture: Automated Guidance, GPS/GNSS, variable rate technologies, smart irrigation systems, smart water management.
	Farm Robotics: Robots for weed control, harvesting, and disease detection; drones for fertilizer or pesticide spraying; autonomous farm technologies.
	Digital Advisory: Digital advisory and extension services (via smartphones, SMS) based on farm data analysis and processing.
	Farm Management Software: Software solutions to improve management at the farm level. Cloud-based solutions and information systems oriented to farm management.
	Precision Technologies for Cattle Raising: Accelerometers, automated heat detection, geofencing and virtual fencing, digital pasture management. Automated milking robots.
Biological inputs and solutions	Biological inputs: biocontrollers, biostimulants, biofertilizers, effective microorganisms, microbiome-based solutions
	Agri-nanotechnology: nano fertilizers, nano pesticides.
	Other solutions: natural pollination services.
Plant, Animal and food Biotechnology solutions	Gene Editing: Seed traits based on genome editing technologies (CRISPR-cas and similars)
	New Genetically Modified (GM) Traits: crop protection, abiotic stress, nitrogen fixation based on genetic engineering. Other crop varieties which are expressing enhanced product features (e.g., improved oil content).
	General Biotech Services: Companies providing biotechnology services in several verticals without a specific field, such as genomic services, bioinformatics, or genetic testing services (Probably part of these companies are producing improved crop varieties, but if this is not the main focus of their company description they are included in this category). We also include general bio-based raw materials, analytical services for manufacturing companies, biotechnology support services for food industries, and the development of nanomaterials and nano ingredients for the food industry.
	Animal Genetics: Animal breeding and genetic improvement in livestock; seafood genetics.
	Pharmaceutical Products: pharma and veterinary products for animal production; vaccines
Alternative ways of farming	Controlled-environment solutions: Indoor farming, vertical farming, urban farming, modular farming, aquaponics, hydroponics, and aquaculture. In general, production methods reducing the reliance on specific soil or weather conditions. Includes the production of equipment for hydroponics, aquaponics, and indoor farming (e.g., modular systems, greenhouse building, led lights for indoor farming). It also includes companies selling inputs for this type of production.
	Nature-based Solutions: Practices that handle production in a natural way, reducing chemical applications and restoring soil health, such as regenerative agriculture, holistic management, permaculture, agroecology, production of organic food, or natural forestry farms. We include reforestation and afforestation. Also, we include here farming strategies that seek the recovery of carbon credits from sustainable practices.
	Alternative productions: algae farms in a controlled environment, seaweed production at an industrial scale, insect rearing farms at an industrial scale. Also, we include companies providing feed solutions to aquaculture based on insects or algae.
Lab-based proteins and food ingredients	Lab-based and Cultured Food: Cellular agriculture, lab-grown foods, synthetic biology, cell-based meat and ingredients, precision fermentation technologies, engineered proteins, protein design, seaweed or algae used as a base for precision fermentation processes (mimicking meat and other proteins' taste and quality)
	Plant-based Proteins: technologies to replace meat, dairy, and seafood products (animal-free alternatives) based on plants (not cultured food or engineered proteins). [More traditional plant-based ingredients not replacing animal proteins and traditional vegan/vegetarian food go into "Functional and healthy foods"].
	Molecular technologies: molecular farming, other molecular technologies.
	Other Lab-based Food Science Tools: food nanotechnology, nanostimulation, and CO2-based proteins.
E-commerce and delivery solutions	e-commerce platforms: e-grocery, e-retail, digital marketplaces for groceries and fresh foods, online shopping, mobile app shopping.
	Delivery intermediaries and solutions: Last-mile delivery, B2B and B2C delivery solutions, delivery intermediaries, vending machine technologies (does not include autonomous or robot delivery that is in other categories).
	Digital agribusiness marketplaces: ag input trading platforms, app-based procurement of ag inputs, and apps for buying, hiring, or sharing machinery. We also include here Uber-type tools (like Uber for tractors).
	Apps connecting farmers to buyers and final consumers: apps seeking to shorten value chains, connecting farmers with buyers and consumers. B2B e-platforms seeking to help large and SME retailers source food products. Apps for fresh food wholesale directly from producers.
Digital food service	Technologies for Restaurants: restaurant management solutions, catering and restaurant automation systems, online and mobile ordering, booking and e-reservation systems, solutions for the hospitality industry, CRM platforms, virtual, cloud & dark kitchen for delivery-only restaurants.

(continued on next page)

(*continued*)

Solution	Details
	Food social networks: restaurant marketplaces, virtual restaurants, food social networks; chef networks, apps to split accounts, customer information apps, customer engagement, food blogging, digital restaurant directory.
	Personalized foods: services delivering ready-to-eat homemade meals, services delivering custom ingredients to cook pre-defined meals, online recipes; corporate food services, home-cooking kits, apps for home chefs to sell their foods, home-cooked food, meal planning services, meal delivery, home-cooked and pre-cooked meals, food sold by home-chefs, pre-portioned ingredients or meal kit delivery for cooking at home.
	Cooking & kitchen technologies: Household smart kitchen devices, automation technologies for kitchens, robotics, and IoT applied to cooking devices. Includes other devices for cooking, like 3D printers.
Functional and healthy foods	Functional Foods: nutrient-enhanced foods, nutraceuticals; food designed to improve diseases such as diabetes and other medical conditions, food designed for people with allergies (allergen-free). Improved infant nutrition and baby foods, dietary or nutritional supplements and bioactive components such as probiotics and prebiotics.
	Healthy and Alternative Food Ingredients: protein supplements and powders; plant-based food (that are not aimed at being perfect substitutes to meat); healthy snacks; organic or non-GMO foods; gluten-free food and ingredients; sustainably-sourced foods; energy foods; nutrient-enhanced and superfoods; ayurvedic supplements or ingredients; healthy food alternative oriented to children; vegan and vegetarian alternative ingredients; sugar-free or sugar-reduced food alternatives; vitamin-infused foods; food with improved nutritional balance; food free of artificial ingredients, colorants or preservatives; hydration solutions; electrolyte drinks; natural cold-pressed juices and drinks; paleo ingredients; seaweed or algae healthy snacks; enriched food and nutritional products based on algae; insect flour and other insect-based products with enriched protein characteristics; tapping water systems to improve water quality. [This category includes vegan/vegetarian and plant-based food that is not aimed at creating meat, egg, or other protein substitutes. These are more generic or traditional foods that are not based on precision or biomass fermentation] [Online marketplaces, apps, or online websites that sell organic, vegan, etc., go in the category of e-commerce]
	Precision Nutrition: personalized nutritional recommendations; digital health advisory services; education and awareness services; apps for weight reduction; food designed according to personal characteristics; personal health or gut microbiome test to develop specific diets; nutritional health advisory.
Logistics, food safety, and traceability solutions	Warehousing Solutions: automated and digital warehousing; warehousing rental; robot palletization; robots for merchandise handling.
	Freight & Transportation Solutions: online freight management; digital shipping carrier solutions; clean transportation technologies; delivery & cargo robots; autonomous vehicles for food delivery and transportation; contactless delivery
	Food Safety and Preservation Technologies: cold storage; refrigeration; thermal control; automation monitoring; ultrasound; pulsed and ultraviolet light, high-pressure processing technologies (HPP), drying technologies at the farm level; tools to improve post-harvest management; food testing
	Traceability & Tracking: Blockchain and distributed ledger technologies; verification services; tags; tracking services
	Food Information: labeling; carbon credits measure & validation, food rating apps, calculation of food GHG emissions; platforms providing information about product attributes; real-time measuring of emissions and pollution; food recognition by image or scanning
	In-store Solutions for Grocery Stores & Supermarkets: Technology gadgets to brick-and-mortar stores; digital shelf technologies; in-store automation solutions for supermarkets, technologies for the reduction in waiting lines. Self-checkout technologies, frictionless technologies.
	Digitalized Fulfillment Solutions: micro-fulfillment and micro-retail solutions; tools for the digitalization of the procurement process; B2B sourcing marketplaces and management solutions oriented to SMEs grocery retailers.
Financial solutions for food and agriculture	Digital Payment Services: payment platforms; e-payment; sms payment; cryptocurrencies oriented to food and agriculture; token services; currency management solutions.
	Fintech Solutions: Digital credit access; credit risk assessment; crowdfunding & crowd farming for farm investment; insurtech solutions.
	Micro-Financial Solutions for Farmers: apps that support financial inclusion, farmer empowerment, flexible loan programs for rural women farmers or small-scale farms
Waste reduction and cascading uses	Cascading Use of Waste from Agriculture and Industrial Production: Companies using waste from industrial and agricultural operations. This includes bioenergy; bioproducts; biorefineries; microbial proteins from organic waste; alternative uses for food byproducts; environmental remediation services; applications from spent or upcycled materials; feed produced from food waste and other residues; use of insects as natural upcyclers for residues; bioconversion processes; upcycling of food residues into new uses.
	Food Losses & Waste Reduction: Monitoring of food waste; apps for planning food needs; rescue of ‘ugly“ or imperfect food; platforms that connect people and businesses to share food that is about to expire and reduce food waste; apps and solutions to reduce food waste at home [Logistics solutions that are oriented to avoid food waste, such as post-harvest technologies, remote control of temperature or cold chain solutions, go in a different category]
	Packaging Technologies: Alternatives to reduce single-use and fossil-based plastics. Smart packaging. Bioplastics; biodegradable and recyclable packaging sources. Invisible packaging for fruits such as natural and biological coatings; edible coatings; nano-coatings. Water soluble packaging. Biodegradable packaging; reusable packaging.

Appendix B. Classification of the annotated set

Class	n (%)
Not relevant	**2237 (42 %)**
Relevant	**3056 (58 %)**
Digital food service	551 (18 %)
E-commerce and delivery solutions	504 (16 %)
Functional and healthy foods	503 (16 %)
Precision agriculture, smart farming and farm robotics	365 (12 %)
Alternative ways of farming	262 (8.6 %)
Logistics food safety and traceability solutions	256 (8.4 %)
Waste reduction and cascading uses	205 (6.7 %)
Lab-based proteins and food ingredients	174 (5.7 %)
Plant, animal and food biotechnology	100 (3.3 %)
Financial solutions for food and agriculture	73 (2.4 %)
Biological inputs and solutions	63 (2.1 %)
Total	**n = 5,293**

Appendix C. List of incumbent firms considered for the analysis

investor_name	sector
Leaps by Bayer	agrochemicals_and_seeds
Bayer	agrochemicals_and_seeds
BASF	agrochemicals_and_seeds
BASF Venture Capital	agrochemicals_and_seeds
UPL	agrochemicals_and_seeds
UPL	agrochemicals_and_seeds
FMC Ventures	agrochemicals_and_seeds
Sumitomo Chemical	agrochemicals_and_seeds
Sumitomo Corporation	agrochemicals_and_seeds
Sumitomo Corporation Equity Asia	agrochemicals_and_seeds
Sumitomo Corporation Europe Limited	agrochemicals_and_seeds
Nufarm	agrochemicals_and_seeds
Bayer Pharmaceuticals	agrochemicals_and_seeds
Syngenta Ventures	agrochemicals_and_seeds
Syngenta	agrochemicals_and_seeds
Syngenta Group	agrochemicals_and_seeds
Yara Growth Ventures	synthetic_fertilizers
Yara International	synthetic_fertilizers
Nutrien Ag Solutions	synthetic_fertilizers
ICL Group	synthetic_fertilizers
ICL Planet Startup Hub	synthetic_fertilizers
EuroChem Group AG	synthetic_fertilizers
K+S Group	synthetic_fertilizers
John Deere	farm_machinery
CLAAS	farm_machinery
AGCO	farm_machinery
KUBOTA Corporation.	farm_machinery
Kubota	farm_machinery
Mahindra Agri Solutions	farm_machinery
Cargill	commodity_trading
Cargill Ventures	commodity_trading
Louis Dreyfus	commodity_trading
Louis Dreyfus Company	commodity_trading
Bunge	commodity_trading
Bunge Ventures	commodity_trading
COFCO	commodity_trading
ADM Venture Capital	commodity_trading
ADM Ventures	commodity_trading
ITOCHU Corporation	commodity_trading
Itochu-Shokuhin	commodity_trading
Marubeni	commodity_trading
Archer Daniels Midland Company	commodity_trading
PepsiCo	food_and_beverage
Nestlé	food_and_beverage
Nestlé Health Science	food_and_beverage
Anheuser-Busch InBev	food_and_beverage
Tyson Foods	meat_and_protein
Tyson Ventures	meat_and_protein
Mars	food_and_beverage
The Coca-Cola Company	food_and_beverage
Coca-Cola Enterprises	food_and_beverage
Coca-Cola Venturing & Emerging Brands	food_and_beverage
Coca-Cola Amatil	food_and_beverage

(*continued on next page*)

(*continued*)

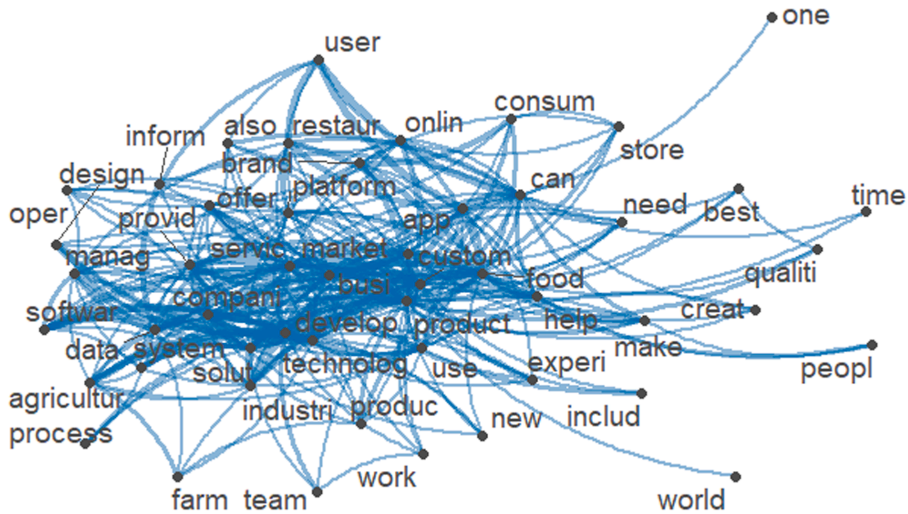
investor_name	sector
Danone	food_and_beverage
Danone Manifesto Venture	food_and_beverage
Mondelez International	food_and_beverage
The Kraft Group	food_and_beverage
BRF	meat_and_protein
NH Foods	meat_and_protein
Marfrig	meat_and_protein
Charoen Pokphand Group	seafood
Trident	seafood
Thai Union Group	seafood
Walmart	grocery_retail
Walmart Foundation	grocery_retail
Kroger	grocery_retail
Kroger Co. Zero Hunger	grocery_retail
ALDI	grocery_retail
Carrefour	grocery_retail
Rewe Group	grocery_retail
Ahold Delhaize	grocery_retail
Tesco	grocery_retail
Merck Animal Health	animal_pharma
Merck Animal Health Ventures	animal_pharma
Novartis	animal_pharma
Elanco	animal_pharma
Nutreco	animal_pharma

Sources:

etc Group(2022) Food Barons 2022: Crisis Profiteering, Digitalization and Shifting Power.

IPES-Food (2017). Too big to feed: The short report. Mega-mergers and the concentration of power in the agri-food sector: How dominant firms have become too big to feed humanity sustainably.

Appendix D. Wordcloud and co-occurrence matrix for the most common tokens in the text corpus



Appendix E. Summary of the performance measures for machine learning models

In addition to the three algorithms that were finally selected as the best performing (i.e., NB, SVM, and LASSO plus topics from the LDA analysis), we also tried other more sophisticated algorithms. We run the Random Forest (RF) algorithm using the randomForest package. RF models create random decision trees based on a bootstrapping approach. The problem with this type of algorithm is that they present high demands regarding computational resources, especially with a high number of trees.

For the binary classification problem, we also explore more complex deep learning algorithms. These neural-based algorithms are built over word embeddings, considering the context in which a word is used for classification. Instead of using the bag-of-word approach, they take advantage of the order of the words in a text corpus based on a multi-layer approach. The general architecture of these neural-based algorithms starts with an input layer (which can be, for example, features of a TF-IDF matrix or a sequence of word embeddings), then we have several hidden layers that capture the relationship between words and information on the structure of the sentences, and finally, we have an output layer that is our set of target categories. In the process, each layer receives information from a previous layer and passes that information to the next one. We apply the simple and convoluted neural network algorithms using the package tidymodels. Although these are more complex models, they need substantial data to perform well (which is not our case since we are working with a relatively small dataset).

Finally, for the multiclass problem, we also try a seeded LDA approach, which is a semi-supervised technique that works similarly to topic models, but the researcher has to pre-define the number of topics and feed the algorithms with keywords associated with that topic.

The results for all these algorithms are summarized in the following two tables:

Binary Stage	NB *quanteda*	SVM *quanteda*	LASSO *tidymodels*	LASSO+Topics LDA *tidymodels*	LASSO+Topics LDA+Hyperp *tidymodels*	RF *randomForest*	Simple Neural Network *tidymodels*	Convolutional Neural Network (CNN) *tidymodels*
Accuracy	0.806	0.795	0.774	0.776	0.760	0.765	0.733	0.673
ROC AUC	0.787	0.783	0.843	0.856	0.836	0.758	0.813	0.814
Precision (rate of true positives)	0.911	0.797	0.768	0.781	0.772	0.901	0.714	0.643
Neg. Pred Value (rate of true negatives)	0.663	0.792	0.786	0.767	0.740	0.614	0.711	0.870

Base: 5,293 companies in the labeled dataset
Null model: Accuracy = 0.57; ROC AUC=0.5

Multiclass problem	NB *quanteda*	SVM *quanteda*	LASSO *tidymodels*	LASSO+Topics LDA *tidymodels*	LASSO+Topics LDA+Hyperp *tidymodels*	RF *randomForest*	Seeded LDA *quanteda*
Accuracy	0.731	0.750	0.732	0.753	0.729	0.658	0.553
ROC AUC	0.784	0.803	0.947	0.955	0.946	0.720	0.748

Base: 3,056 companies labeled as relevant in the original dataset
Null model: Accuracy = 0.16; ROC AUC=0.5

Appendix F. Robustness check based on Large-Language models (LLMs)

We perform a classification task based on GPT-3.5 through the *openai*, an R wrapper package for OpenAI API. This task first requires creating a model fine-tuning, which is the training stage. For this purpose, we used our training dataset of 5,293 companies manually labeled for the binary stage and the subset of 3,056 relevant companies for the multiclass stage (see Appendix B). GTP-3.5 has different models available for this fine-tuning process. In our case, we used the model 'ada,' which is suitable for basic language tasks and has the fastest processing time. We set four iterations ("number of epochs" according to the package) to fine-tune the model hyperparameters.

The accuracy rates for the binary model reached 0.856 (with 14,300 companies classified as "relevant" for the next stage) and 0.843 for the multiclass model. After the fine-tuning, we applied the resulting trained model over the unclassified dataset. In the following table, we show the results of this classification process in terms of the 11 categories and compare them to the distribution of categories we reached following our own methodology in the paper.

class	Original classification		GPT classification	
	n	*freq*	*n*	*freq*
Digital food service	3,225	21 %	2,798	20 %
Functional and healthy foods	3,109	20 %	2,722	19 %
E-commerce and delivery solutions	2,791	18 %	2,490	17 %
Precision agriculture, smart farming and farm robotics	2,072	13 %	1,605	11 %
Alternative ways of farming	1,143	7 %	1,169	8 %
Logistics, food safety and traceability solutions	950	6 %	1,155	8 %
Waste reduction and cascading uses	910	6 %	894	6 %
Lab-based proteins and food ingredients	479	3 %	519	4 %
Plant, animal and food biotechnology	408	3 %	443	3 %
Financial solutions for food and agriculture	182	1 %	259	2 %
Biological inputs and solutions	171	1 %	246	2 %

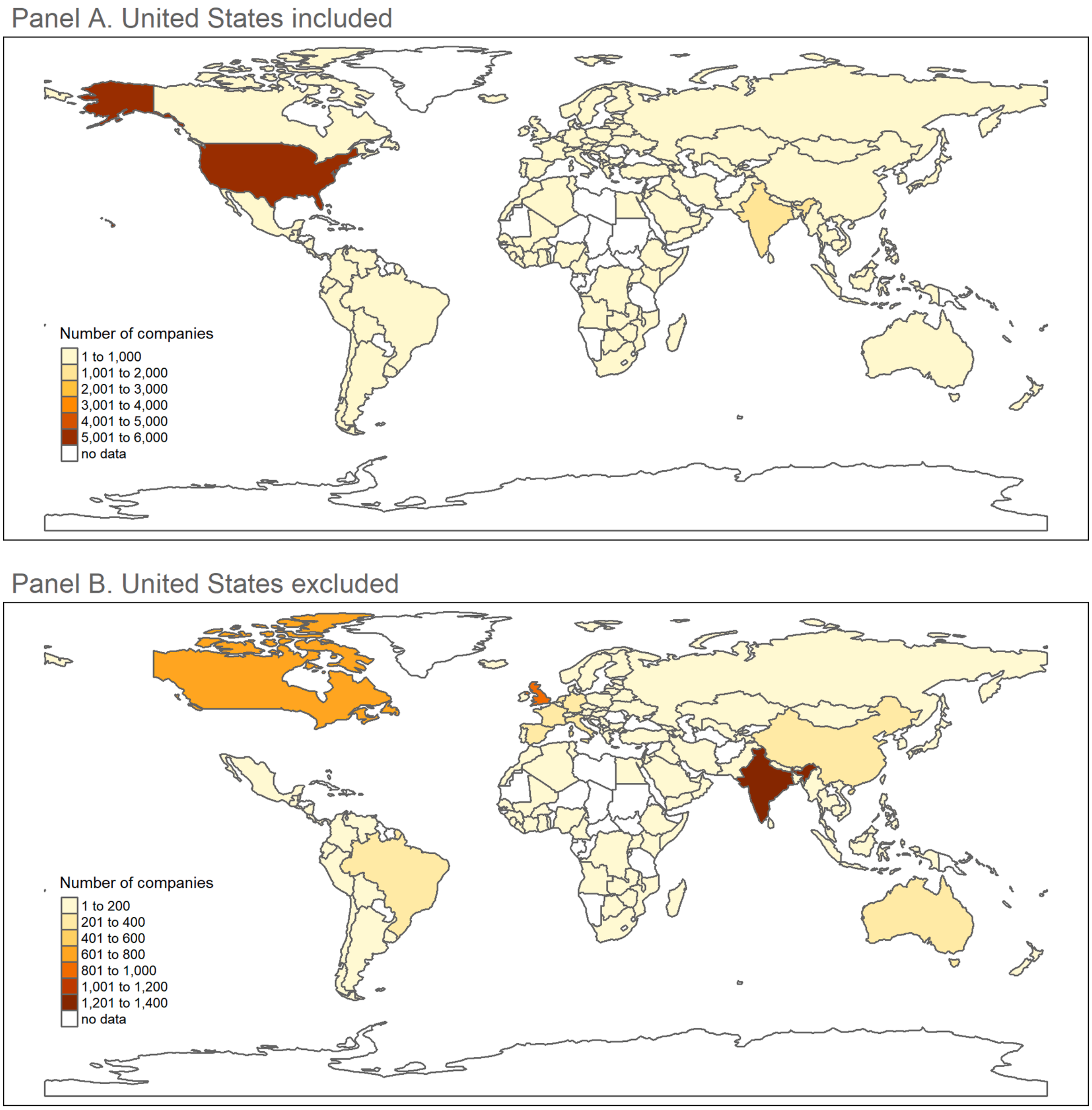
Even though there are differences in the classification between our original analysis and the classification based on chatGPT (which is expected considering no algorithm has perfect accuracy), the final results in terms of the distribution of the categories do not show substantial differences or changes in the relative weight of the categories. Around 75 % of companies in the second Gold Standard (300 companies) were accurately classified by ChatGPT (the accuracy of the original classification in terms of this second Gold Standard was higher, reaching 82 %).

Appendix G. Geographic distribution of companies and funds

Geographic distribution of companies in the final sample (main 20 countries).

#	name	continent	income_grp	N of companies	freq
1	United States	North America	High income: OECD	5405	35.0
2	India	Asia	Lower middle income	1215	7.9
3	United Kingdom	Europe	High income: OECD	973	6.3
4	Canada	North America	High income: OECD	620	4.0
5	Brazil	South America	Upper middle income	385	2.5
6	Italy	Europe	High income: OECD	368	2.4
7	Israel	Asia	High income: OECD	365	2.4
8	Spain	Europe	High income: OECD	351	2.3
9	China	Asia	Upper middle income	337	2.2
10	France	Europe	High income: OECD	333	2.2
11	Germany	Europe	High income: OECD	329	2.1
12	Australia	Oceania	High income: OECD	280	1.8
13	Netherlands	Europe	High income: OECD	239	1.5
14	Switzerland	Europe	High income: OECD	170	1.1
15	Japan	Asia	High income: OECD	166	1.1
16	Sweden	Europe	High income: OECD	131	0.8
17	Nigeria	Africa	Lower middle income	125	0.8
18	Indonesia	Asia	Lower middle income	116	0.8
19	Ireland	Europe	High income: OECD	114	0.7
20	United Arab Emirates	Asia	High income: nonOECD	104	0.7

Geographic distribution of companies in the final sample.



These maps depict the number of companies by country for our final sample. Panel A includes the full set of countries. Panel B presents the same map but excludes the United States (which explains around 35 % of the total). We do this to highlight the differences in scale and improve the visualization.

Appendix H. A detailed list of investment strategies by incumbent firms

Strategy	Investor	Target Sector
Upgrading strategies	Farm machinery	Precision agriculture and smart farming
	Agrochemicals and Seeds Synthetic Fertilizers	Precision agriculture and smart farming
	Agrochemicals and Seeds Synthetic Fertilizers Veterinary Pharma	Plant Biotechnology
	Food & Beverage	Functional foods
	Food & Beverage	Waste reduction
	Grocery Retail Commodity Trading	Logistics, food safety & traceability solutions

(*continued on next page*)

(*continued*)

Strategy	Investor	Target Sector
	Commodity Trading Agrochemicals and Seeds Synthetic Fertilizers Farm Machinery	E-commerce (agribusiness marketplaces)
	Grocery Retail	E-commerce
Adaptation or Control Strategies	Meat and Protein Seafood Food & Beverage Animal Pharma	Cellular agriculture and plant-based
	Agrochemicals and Seeds Synthetic Fertilizers	Biological inputs
	Commodity Trading	Alternative ways of farming
	Agrochemicals and Seeds Synthetic Fertilizers Commodity Trading	Financial solutions for food & agriculture
Opportunity Strategies	Commodity Trading	Cellular agriculture and plant-based
	Commodity Trading	Biological inputs
	Commodity Trading	Precision agriculture and smart farming
	Commodity Trading	Plant biotechnology
	Grocery Retail	Alternative ways of farming
	Food & Beverage	E-commerce
	Food & Beverage	Digital food service

Appendix I. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodpol.2024.102684.

References

AgFunder, 2022. 2022 AgFunder AgriFoodTech Investment Report. AgFunder, San Francisco (CA).

Alexander, P., Rounsevell, M.D.A., Dislich, C., Dodson, J.R., Engström, K., Moran, D., 2015. Drivers for global agricultural land use change: The nexus of diet, population, yield and bioenergy. Global Environmental Change 35, 138–147. https://doi.org/10.1016/j.gloenvcha.2015.08.011.

Ambler, K., de Brauw, A., Herskowitz, S., Pulido, C., 2023. Viewpoint: Finance needs of the agricultural midstream. Food Policy 121, 102530. https://doi.org/10.1016/j.foodpol.2023.102530.

An, R., Yang, Y., Yang, F., Wang, S., 2023. Use prompt to differentiate text generated by ChatGPT and humans. Machine Learning with Applications 14, 100497. https://doi.org/10.1016/j.mlwa.2023.100497.

Ballard, E., 2016. Chinese Grain Trading Firm Cofco Buys Out Nidera. Wall Street Journal.

Barrett, C.B., Reardon, T., Swinnen, J., Zilberman, D., 2022. Agri-food Value Chain Revolutions in Low- and Middle-Income Countries. Journal of Economic Literature 60, 1316–1377. https://doi.org/10.1257/jel.20201539.

Bellemare, M.F., 2022. Agricultural value chains: towards a marriage of development economics and industrial organisation? Australian Journal of Agricultural and Resource Economics 66, 241–255. https://doi.org/10.1111/1467-8489.12467.

Béné, C., 2022. Why the Great Food Transformation may not happen – A deep-dive into our food systems' political economy, controversies and politics of evidence. World Development 154, 105881. https://doi.org/10.1016/j.worlddev.2022.105881.

Béné, C., Bakker, D., Chavarro, M.J., Even, B., Melo, J., Sonneveld, A., 2021. Global assessment of the impacts of COVID-19 on food security. Global Food Security 31, 100575. https://doi.org/10.1016/j.gfs.2021.100575.

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., Matsuo, A., 2018. quanteda: An R package for the quantitative analysis of textual data. JOSS 3, 774. https://doi.org/10.21105/joss.00774.

Bertoni, F., Tykvová, T., 2015. Does governmental venture capital spur invention and innovation? Evidence from young European biotech companies. Research Policy 44, 925–935. https://doi.org/10.1016/j.respol.2015.02.002.

Birner, R., Daum, T., Pray, C., 2021. Who drives the digital revolution in agriculture? A review of supply-side trends, players and challenges. Applied Economic Perspectives and Policy 43, 1260–1285. https://doi.org/10.1002/aepp.13145.

Blei, D.M., Ng, A., Jordan, M., 2003. Latent Dirichlet Allocation. Journal of Machine Learning Research 3, 993–1022.

Breschi, S., Lassébie, J., Menon, C., 2018. A portrait of innovative start-ups across countries (OECD Science, Technology and Industry Working Papers No. 2018/02), OECD Science, Technology and Industry Working Papers. Doi: 10.1787/f9ff02f4-en.

Bucak, S., 2022. Unicorn drain: Europe is still losing its most valuable startups to the US. Sifted.

Bunge, A.C., Wood, A., Halloran, A., Gordon, L.J., 2022. A systematic scoping review of the sustainability of vertical farming, plant-based alternatives, food delivery services and blockchain in food systems. Nat Food 1–9. https://doi.org/10.1038/s43016-022-00622-8.

Caldwell, B., Fife, J., 2023. Biologicals Regulation: The Right Way, 3Bar Bio. accessed 1.5.24.

Canning, P., 2011. A Revised and Expanded Food Dollar Series: A Better Understanding of Our Food Costs. Economic Research Service USDA.

CB Insights, 2023a. State of Venture 2022 Report.

CB Insights, 2023b. List of Unicorn Startups & Markets.

Chenarides, L., Grebitus, C., Lusk, J.L., Printezis, I., 2021. Food consumption behavior during the COVID-19 pandemic. Agribusiness 37, 44–81. https://doi.org/10.1002/agr.21679.

Chintalapudi, N., Angeloni, U., Battineni, G., di Canio, M., Marotta, C., Rezza, G., Sagaro, G.G., Silenzi, A., Amenta, F., 2022. LASSO Regression Modeling on Prediction of Medical Terms among Seafarers' Health Documents Using Tidy Text Mining. Bioengineering 9, 124. https://doi.org/10.3390/bioengineering9030124.

Clapp, J., 2015. ABCD and beyond: From grain merchants to agricultural value chain managers. Canadian Food Studies / La Revue Canadienne Des Études Sur L'alimentation 2, 126–135. https://doi.org/10.15353/cfs-rcea.v2i2.84.

Clapp, J., 2021. The problem with growing corporate concentration and power in the global food system. Nat Food 2, 404–408. https://doi.org/10.1038/s43016-021-00297-7.

Connolly, A.J., Turner, J., Potocki, A.D., 2018. IGNITE your corporate innovation: insights from setting up an ag-tech start-up accelerator. IFAM 21, 833–846. https://doi.org/10.22434/IFAMR2017.0089.

Costa, C., Wollenberg, E., Benitez, M., Newman, R., Gardner, N., Bellone, F., 2022. Roadmap for achieving net-zero emissions in global food systems by 2050. Sci Rep 12, 15064. https://doi.org/10.1038/s41598-022-18601-1.

Crippa, M., Solazzo, E., Guizzardi, D., Monforti-Ferrario, F., Tubiello, F.N., Leip, A., 2021. Food systems are responsible for a third of global anthropogenic GHG emissions. Nat Food 2, 198–209. https://doi.org/10.1038/s43016-021-00225-9.

Croce, A., Guerini, M., Ughetto, E., 2018. Angel Financing and the Performance of High-Tech Start-Ups. Journal of Small Business Management 56, 208–228. https://doi.org/10.1111/jsbm.12250.

Crunchbase, 2022. Database behing API Permission. Research. Researrch Access.

Crunchbase News, 2019. How the Crunchbase news team uses data in its reporting [WWW Document]. Crunchbase News. URL https://news.crunchbase.com/methodology/ (accessed 1.2.24).

Crunchbase Product Team, 2022. Where does Crunchbase get their data? [WWW Document]. accessed 1.8.24 Crunchbase | Knowledge Center. https://support.crun chbase.com/hc/en-us/articles/360009616013-Where-does-Crunchbase-get-their-data-.

Dalle, J.-M., den Besten, M., Menon, C., 2017. Using Crunchbase for economic and managerial research (OECD Science, Technology and Industry Working Papers No. 2017/08), OECD Science, Technology and Industry Working Papers. Doi: 10.1787/6c418d60-en.

Deconinck, K., 2019. New evidence on concentration in seed markets. Global Food Security 23, 135–138. https://doi.org/10.1016/j.gfs.2019.05.001.

Deconinck, K., 2020. Concentration in Seed and Biotech Markets: Extent, Causes, and Impacts. Annu. Rev. Resour. Econ. 12, 129–147. https://doi.org/10.1146/annurev-resource-102319-100751.

Deconinck, K., 2021. Concentration and market power in the food chain (No. 151), OECD Food, Agriculture and Fisheries. OECD, Paris. Doi: 10.1787/3151e4ca-en.

Deininger, K., Ali, D.A., Kussul, N., Shelestov, A., Lemoine, G., Yailimova, H., 2023. Quantifying war-induced crop losses in Ukraine in near real time to strengthen local and global food security. Food Policy 115, 102418. https://doi.org/10.1016/j.foodpol.2023.102418.

den Besten, M., 2020. Crunchbase Research: Monitoring Entrepreneurship Research in the Age of Big Data. https://doi.org/10.2139/ssrn.3724395.

Dibner, G., 2020. The Startup Nation goes to America. Angular Ventures.

Dosi, G., Nelson, R.R., 2010. Chapter 3 - Technical Change and Industrial Dynamics as Evolutionary Processes, in: Hall, B.H., Rosenberg, N. (Eds.), Handbook of the Economics of Innovation, Handbook of The Economics of Innovation, Vol. 1. North-Holland, pp. 51–127. Doi: 10.1016/S0169-7218(10)01003-8.

Dosi, G., Marsili, O., Orsenigo, L., Salvatore, R., 1995. Learning, Market Selection and the Evolution of Industrial Structures. Small Business Economics 7, 411–436.

Dushnitsky, G., Lenox, M.J., 2005. When do incumbents learn from entrepreneurial ventures?: Corporate venture capital and investing firm innovation rates. Research Policy 34, 615–639. https://doi.org/10.1016/j.respol.2005.01.017.

ETC Group, 2022. Food Barons 2022: Crisis Profiteering. Digitalization and Shifting Power.

Fao, 2022. Gene editing and agrifood systems. FAO, Rome, Italy. https://doi.org/10.4060/cc3579en.

FAO, IFAD, UNICEF, WFP, WHO, 2022. The State of Food Security and Nutrition in the World 2022. FAO, Rome. https://doi.org/10.4060/cc0639en.

Ferrati, F., Muffatto, M., 2020. Using Crunchbase for Research in Entrepreneurship: Data Content and Structure. Presented at the 19th European Conference on Research Methodology for Business and Management Studies, 10.34190/ERM.20.120.

Finger, R., 2023. Digital innovations for sustainable and resilient agricultural systems. European Review of Agricultural Economics 50, 1277–1309. https://doi.org/10.1093/erae/jbad021.

Florida, R., Hathaway, I., 2018. How the Geography of Startups and Innovation Is Changing. Harvard Business Review.

Forward Fooding, 2022. The FoodTech 500 - 2022 Edition.

Frank, E., Bouckaert, R.R., 2006. Naive Bayes for Text Classification with Unbalanced Classes, in: Fürnkranz, J., Scheffer, T., Spiliopoulou, M. (Eds.), Knowledge Discovery in Databases: PKDD 2006, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 503–510. Doi: 10.1007/11871637_49.

Freijeiro-González, L., Febrero-Bande, M., González-Manteiga, W., 2022. A Critical Review of LASSO and Its Derivatives for Variable Selection Under Dependence Among Covariates. International Statistical Review 90, 118–145. https://doi.org/10.1111/insr.12469.

Fuglie, K.O., Echeverria, R.G., 2024. The economic impact of CGIAR-related crop technologies on agricultural productivity in developing countries, 1961–2020. World Development 176, 106523. https://doi.org/10.1016/j.worlddev.2023.106523.

Fuglie, K.O., Heisey, P., King, J.L., Day-Rubenstein, K., Schimmelpfennig, D., Wang, S.L., Pray, C.E., Karmarkar-Deshmukh, R., 2011. Research Investments and Market Structure in the Food Processing, Agricultural Input, and Biofuel Industries Worldwide (No. ERR-130). U.S. Dept. of Agriculture, Econ. Res. Serv., Washington D.C.

Fuglie, K.O., King, J.L., Heisey, P.W., Schimmelpfennig, D.E. (Eds.), 2012. Rising Concentration in Agricultural Input Industries Influences New Farm Technologies. Amber Waves: The Economics of Food, Farming, Natural Resources, and Rural America. Doi: 10.22004/ag.econ.142404.

Gardezi, M., Joshi, B., Rizzo, D.M., Ryan, M., Prutzer, E., Brugler, S., Dadkhah, A., 2023. Artificial intelligence in farming: Challenges and opportunities for building trust. Agronomy Journal agj2.21353. https://doi.org/10.1002/agj2.21353.

Geels, F.W., 2019. Socio-technical transitions to sustainability: a review of criticisms and elaborations of the Multi-Level Perspective. Current Opinion in Environmental Sustainability 39, 187–201. https://doi.org/10.1016/j.cosust.2019.06.009.

Geels, F.W., Schot, J., 2007. Typology of sociotechnical transition pathways. Research Policy 36, 399–417. https://doi.org/10.1016/j.respol.2007.01.003.

Gereffi, G., 1994. The Organization of Buyer-Driven Global Commodity Chains: How U.S. Retailers Shape Overseas Production Networks. Commodity Chains and Global Capitalism 95–122.

Gereffi, G., Christian, M., 2009. The Impacts of Wal-Mart: The Rise and Consequences of the World's Dominant Retailer. Annual Review of Sociology 35, 573–591.

Good Food Institute, 2021. 2021 State of the Industry Report - Cultivated meat and seafood. The Good Food Institute.

Grames, E.M., Stillman, A.N., Tingley, M.W., Elphick, C.S., 2019. An automated approach to identifying search terms for systematic reviews using keyword co-occurrence networks. Methods in Ecology and Evolution 10, 1645–1654. https://doi.org/10.1111/2041-210X.13268.

Halevy, A., Norvig, P., Pereira, F., 2009. The Unreasonable Effectiveness of Data. IEEE Intell. Syst. 24, 8–12. https://doi.org/10.1109/MIS.2009.36.

Herrero, M., Thornton, P.K., Mason-D'Croz, D., Palmer, J., Benton, T.G., Bodirsky, B.L., Bogard, J.R., Hall, A., Lee, B., Nyborg, K., Pradhan, P., Bonnett, G.D., Bryan, B.A., Campbell, B.M., Christensen, S., Clark, M., Cook, M.T., de Boer, I.J.M., Downs, C., Dizyee, K., Folberth, C., Godde, C.M., Gerber, J.S., Grundy, M., Havlik, P., Jarvis, A., King, R., Loboguerrero, A.M., Lopes, M.A., McIntyre, C.L., Naylor, R., Navarro, J., Obersteiner, M., Parodi, A., Peoples, M.B., Pikaar, I., Popp, A., Rockström, J., Robertson, M.J., Smith, P., Stehfest, E., Swain, S.M., Valin, H., van Wijk, M., van Zanten, H.H.E., Vermeulen, S., Vervoort, J., West, P.C., 2020. Innovation can accelerate the transition towards a sustainable food system. Nature Food 1, 266–272. https://doi.org/10.1038/s43016-020-0074-1.

Herrero, M., Thornton, P.K., Mason-D'Croz, D., Palmer, J., Bodirsky, B.L., Pradhan, P., Barrett, C.B., Benton, T.G., Hall, A., Pikaar, I., Bogard, J.R., Bonnett, G.D., Bryan, B. A., Campbell, B.M., Christensen, S., Clark, M., Fanzo, J., Godde, C.M., Jarvis, A., Loboguerrero, A.M., Mathys, A., McIntyre, C.L., Naylor, R.L., Nelson, R., Obersteiner, M., Parodi, A., Popp, A., Ricketts, K., Smith, P., Valin, H., Vermeulen, S. J., Vervoort, J., van Wijk, M., van Zanten, H.H., West, P.C., Wood, S.A., Rockström, J., 2021. Articulating the effect of food systems innovation on the Sustainable Development Goals. The Lancet Planetary Health 5, e50–e62. https://doi.org/10.1016/S2542-5196(20)30277-1.

Holicka, M., Vinodrai, T., 2022. The global geography of investment in emerging technologies: the case of blockchain firms. Regional Studies, Regional Science 9, 177–179. https://doi.org/10.1080/21681376.2022.2047769.

Howard, P.H., 2016. Concentration and power in the food system: who controls what we eat?, Contemporary food studies: economy, culture and politics. Bloomsbury Academic, London New York Oxford New Delhi Sydney.

Hvitfeldt, E., Silge, J., 2022. Supervised Machine Learning for Text Analysis in R. Chapman & Hall.

Ipcc, 2023. IPCC AR6 Synthesis Report: Climate Change 2023. Summary for Policymakers, IPCC - Intergovernmental panel on climate change.

Ipes, 2017. Too big to feed: Exploring the impacts of mega-mergers, consolidation, and concentration of power in the agri-food sector. International Panel of Experts on Sustainable Food Systems.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning, Springer Texts in Statistics. Springer, New York, New York, NY, 10.1007/978-1-4614-7138-7.

Joachims, T., 1998. Text categorization with Support Vector Machines: Learning with many relevant features, in: Nédellec, C., Rouveirol, C. (Eds.), Machine Learning: ECML-98, Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 137–142. Doi: 10.1007/BFb0026683.

Kaul, S., 2021. Venture capital accelerates food technology innovation. Nat Food 2, 909–910. https://doi.org/10.1038/s43016-021-00433-3.

Kelloway, C., 2023. Bunge and Viterra's Mega Merger Would Dramatically Consolidate Global Grain Trade. Food & Power.

Klerkx, L., Villalobos, P., 2024. Are AgriFoodTech start-ups the new drivers of food systems transformation? An overview of the state of the art and a research agenda. Global Food Security 40, 100726. https://doi.org/10.1016/j.gfs.2023.100726.

Kowsari, K., Jafari Meimandi, K., Heidarysafa, M., Mendu, S., Barnes, L., Brown, D., 2019. Text Classification Algorithms: A Survey. Information 10, 150. https://doi.org/10.3390/info10040150.

Kulkov, I., Hellström, M., Wikström, K., 2020. Identifying the role of business accelerators in the developing business ecosystem: the life science sector. European Journal of Innovation Management 24, 1459–1479. https://doi.org/10.1108/EJIM-04-2020-0139.

Kusumawati, R., D'arofah, A., Pramana, P.A., 2019. Comparison Performance of Naive Bayes Classifier and Support Vector Machine Algorithm for Twitter's Classification of Tokopedia Services. J. Phys.: Conf. Ser. 1320, 012016 https://doi.org/10.1088/1742-6596/1320/1/012016.

Kwon, O., Lim, S., Lee, D.H., 2018. Acquiring startups in the energy sector: a study of firm value and environmental policy. Business Strategy and the Environment 27, 1376–1384. https://doi.org/10.1002/bse.2187.

Lee, J., Gereffi, G., Beauvais, J., 2012. Global value chains and agrifood standards: Challenges and possibilities for smallholders in developing countries. PNAS 109, 12326–12331. https://doi.org/10.1073/pnas.0913714108.

Lerner, J., Nanda, R., 2020. Venture Capital's Role in Financing Innovation: What We Know and How Much We Still Need to Learn. Journal of Economic Perspectives 34, 237–261. https://doi.org/10.1257/jep.34.3.237.

Lezoche, M., Hernandez, J.E., del Alemany Díaz, M.E., Panetto, H., Kacprzyk, J., 2020. Agri-food 4.0: A survey of the supply chains and technologies for the future agriculture. Computers in Industry 117, 103187. https://doi.org/10.1016/j.compind.2020.103187.

Loukas, S., 2020. Text classification using Naive Bayes: theory & a working example [WWW Document]. Towards. accessed 6.7.23 Data Science. https://towardsdatascience.com/text-classification-using-naive-bayes-theory-a-working-example-2ef4b7eb7d5a.

Luo, X., 2021. Efficient English text classification using selected Machine Learning Techniques. Alexandria Engineering Journal 60, 3401–3409. https://doi.org/10.1016/j.aej.2021.02.009.

Mac Clay, P., Sellare, J., 2022. Value chain transformations in the transition to a sustainable bioeconomy. SSRN Journal. https://doi.org/10.2139/ssrn.4193583.

Macchiavello, R., Reardon, T., Richards, T.J., 2022. Empirical Industrial Organization Economics to Analyze Developing Country Food Value Chains. Annual Review of Resource Economics 14, 193–220. https://doi.org/10.1146/annurev-resource-101721-023554.

Manning, C., Raghavan, P., Schuetze, H., 2009. Introduction to Information Retrieval.

Marra, A., Antonelli, P., Pozzi, C., 2017. Emerging green-tech specializations and clusters – A network analysis on technological innovation at the metropolitan level.

Renewable and Sustainable Energy Reviews 67, 1037–1046. https://doi.org/10.1016/j.rser.2016.09.086.
Mazzucato, M., 2022. Financing the Green New Deal. Nat Sustain 5, 93–94. https://doi.org/10.1038/s41893-021-00828-x.
McElroy, D., 2004. Valuing the product development cycle in agricultural biotechnology—what's in a name? Nat Biotechnol 22, 817–822. https://doi.org/10.1038/nbt0704-817.
McFadden, J., Casalini, F., Griffin, T., Antón, J., 2022. The digitalisation of agriculture: A literature review and emerging policy issues. OECD, Paris. https://doi.org/10.1787/285cc27d-en.
Meemken, E.-M., Barrett, C.B., Michelson, H.C., Qaim, M., Reardon, T., Sellare, J., 2021. Sustainability standards in global agrifood supply chains. Nat Food 2, 758–765. https://doi.org/10.1038/s43016-021-00360-3.
Mellon, J., Bailey, J., Scott, R., Breckwoldt, J., Miori, M., Schmedeman, P., 2023. Do AIs Know What the Most Important Issue is? Using Language Models to Code Open-Text Social Survey Responses at Scale. https://doi.org/10.2139/ssrn.4310154.
Muehlfeld, K., Weitzel, U., van Witteloostuijn, A., 2011. Mergers and acquisitions in the global food processing industry in 1986–2006. Food Policy 36, 466–479. https://doi.org/10.1016/j.foodpol.2011.05.002.
Navarro, A.I., Camusso, J., Soler, M.J., Garzón Moresi, M.C., Galiano, A.E., 2022. Opportunities and challenges for the generation of sustainable AgTech Startups in Latin America and The Caribbean. Universidad Austral & Yield Lab Institute.
Noble, W.S., 2006. What is a support vector machine? Nat Biotechnol 24, 1565–1567. https://doi.org/10.1038/nbt1206-1565.
Ollinger, M., Nguyen, S.V., Blayney, D., Chambers, B., Nelson, K., 2005. Structural Change in the Meat, Poultry, Dairy, and Grain Processing Industries. Department of Agriculture, Economic Research Service, U.S.
Ollion, E., Shen, R., Macanovic, A., Chatelain, A., 2023. ChatGPT for Text Annotation? Mind the Hype! (preprint). SocArXiv. https://doi.org/10.31235/osf.io/x58kn.
OpenAI, 2023. OpenAI API Reference Guide [WWW Document]. URL https://platform.openai.com (accessed 1.9.24).
Parisi, C., Tillie, P., Rodríguez-Cerezo, E., 2016. The global pipeline of GM crops out to 2020. Nat Biotechnol 34, 31–36. https://doi.org/10.1038/nbt.3449.
Pietrobelli, C., Rabellotti, R., 2011. Global Value Chains Meet Innovation Systems: Are There Learning Opportunities for Developing Countries? World Development, Special Section (pp. 1204-1270): Foreign Technology and Indigenous Innovation in the Emerging Economies 39, 1261–1269. Doi: 10.1016/j.worlddev.2010.05.013.
Poore, J., Nemecek, T., 2018. Reducing food's environmental impacts through producers and consumers. Science 360, 987–992. https://doi.org/10.1126/science.aaq0216.
Qureshi, K.A., Sabih, M., 2021. Un-Compromised Credibility: Social Media Based Multi-Class Hate Speech Classification for Text. IEEE Access 9, 109465–109477. https://doi.org/10.1109/ACCESS.2021.3101977.
Rabby, G., Berka, P., 2022. Multi-class classification of COVID-19 documents using machine learning algorithms. J Intell Inf Syst. https://doi.org/10.1007/s10844-022-00768-8.
Rajput, D., Wang, W.-J., Chen, C.-C., 2023. Evaluation of a decided sample size in machine learning applications. BMC Bioinformatics 24, 48. https://doi.org/10.1186/s12859-023-05156-9.
Read, Q.D., Hondula, K.L., Muth, M.K., 2022. Biodiversity effects of food system sustainability actions from farm to fork. Proceedings of the National Academy of Sciences 119, e2113884119. Doi: 10.1073/pnas.2113884119.
Reardon, T., 2015. The hidden middle: the quiet revolution in the midstream of agrifood value chains in developing countries. Oxf Rev Econ Policy 31, 45–63. https://doi.org/10.1093/oxrep/grv011.
Reardon, T., Echeverria, R., Berdegué, J., Minten, B., Liverpool-Tasie, S., Tschirley, D., Zilberman, D., 2019. Rapid transformation of food systems in developing regions: Highlighting the role of agricultural research & innovations. Agricultural Systems, Agricultural Research for Rural Prosperity: Rethinking the Pathways 172, 47–59. https://doi.org/10.1016/j.agsy.2018.01.022.
Reardon, T., Belton, B., Liverpool-Tasie, L.S.O., Lu, L., Nuthalapati, C.S.R., Tasie, O., Zilberman, D., 2021a. E-commerce's fast-tracking diffusion and adaptation in developing countries. Applied Economic Perspectives and Policy 43, 1243–1259. https://doi.org/10.1002/aepp.13160.
Reardon, T., Timmer, C.P., Minten, B., 2012. Supermarket revolution in Asia and emerging development strategies to include small farmers. Proceedings of the National Academy of Sciences 109, 12332–12337. Doi: 10.1073/pnas.1003160108.
Reardon, T., Heiman, A., Lu, L., Nuthalapati, C.S.R., Vos, R., Zilberman, D., 2021b. "Pivoting" by food industry firms to cope with COVID-19 in developing regions: E-commerce and "copivoting" delivery intermediaries. Agricultural Economics 52, 459–475. https://doi.org/10.1111/agec.12631.
Reardon, T., Liverpool-Tasie, L.S.O., Minten, B., 2021c. Quiet Revolution by SMEs in the midstream of value chains in developing regions: wholesale markets, wholesalers, logistics, and processing. Food Sec. 13, 1577–1594. https://doi.org/10.1007/s12571-021-01224-1.
Reardon, T., Timmer, C.P., 2012. The Economics of the Food System Revolution. Annual Review of Resource Economics 4, 225–264. https://doi.org/10.1146/annurev.resource.050708.144147.
Reardon, T., Vos, R., 2023. How resilience innovations in food supply chains are revolutionizing logistics, wholesale trade, and farm services in developing countries. International Food and Agribusiness Management Review 1–12. https://doi.org/10.22434/IFAMR2022.0138.
Retterath, A., Braun, R., 2020. Benchmarking Venture Capital Databases. https://doi.org/10.2139/ssrn.3706108.
Riekert, M., Riekert, M., Klein, A., 2021. Simple Baseline Machine Learning Text Classifiers for Small Datasets. SN COMPUT. SCI. 2, 178. https://doi.org/10.1007/s42979-021-00480-4.
Robinson, D., Silge, J., 2017. Text Mining with R, 1st Ed. ed. O'Reilly Media.
Rudnytskyi, I., 2023. openai: R Wrapper for OpenAI API.
Salerno, T., 2017. Cargill's corporate growth in times of crises: how agro-commodity traders are increasing profits in the midst of volatility. Agric Hum Values 34, 211–222. https://doi.org/10.1007/s10460-016-9681-8.
Sebők, M., Kacsuk, Z., 2021. The Multiclass Classification of Newspaper Articles with Machine Learning: The Hybrid Binary Snowball Approach. Political Analysis 29, 236–249. https://doi.org/10.1017/pan.2020.27.
Sexton, R.J., 2013. Market Power, Misconceptions, and Modern Agricultural Markets. American Journal of Agricultural Economics 95, 209–219. https://doi.org/10.1093/ajae/aas102.
Shen, J., Zhu, Z., Qaim, M., Fan, S., Tian, X., 2023. E-commerce improves dietary quality of rural households in China. Agribusiness 39, 1495–1511. https://doi.org/10.1002/agr.21864.
Startup Genome, 2022. The Global Startup Ecosystem Report. Global Entrepreneural Network.
SVG Ventures, 2020. Thrive Top 50 AgTech & FoodTech.
Tiba, S., van Rijnsoever, F.J., Hekkert, M.P., 2021. Sustainability startups and where to find them: Investigating the share of sustainability startups across entrepreneurial ecosystems and the causal drivers of differences. Journal of Cleaner Production 306, 127054. https://doi.org/10.1016/j.jclepro.2021.127054.
Tibshirani, R., 1996. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society. Series B (methodological) 58, 267–288.
Tubiello, F.N., Rosenzweig, C., Conchedda, G., Karl, K., Gütschow, J., Xueyao, P., Obli-Laryea, G., Wanner, N., Qiu, S.Y., Barros, J.D., Flammini, A., Mencos-Contreras, E., Souza, L., Quadrelli, R., Heiðarsdóttir, H.H., Benoit, P., Hayek, M., Sandalow, D., 2021. Greenhouse gas emissions from food systems: building the evidence base. Environ. Res. Lett. 16, 065007 https://doi.org/10.1088/1748-9326/ac018e.
Uddin, M.N., Hafiz, M.F.B., Hossain, S., Islam, S.M.M., 2022. Drug Sentiment Analysis using Machine Learning Classifiers. International Journal of Advanced Computer Science and Applications (IJACSA) 13. https://doi.org/10.14569/IJACSA.2022.0130112.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is All you Need. Advances in Neural Information Processing Systems. Curran Associates Inc.
Vergara, J.M., Barrett, C.B., 2023. Private capital investments in agrifood-tech startups in South America, 2007-2022.
Volpe, R., Boland, M.A., 2022. The Economic Impacts of Walmart Supercenters. Annual Review of Resource Economics 14, 43–62. https://doi.org/10.1146/annurev-resource-111820-032827.
Welbers, K., Van Atteveldt, W., 2022. Corpustools.
Wing, I.S., De Cian, E., Mistry, M.N., 2021. Global vulnerability of crop yields to climate change. Journal of Environmental Economics and Management 109, 102462. https://doi.org/10.1016/j.jeem.2021.102462.
Wright, B.D., 2011. The Economics of Grain Price Volatility. Applied Economic Perspectives and Policy 33, 32–58.
Yi, J., Meemken, E.-M., Mazariegos-Anastassiou, V., Liu, J., Kim, E., Gómez, M.I., Canning, P., Barrett, C.B., 2021. Post-farmgate food value chains make up most of consumer food expenditures globally. Nat Food 1–9. https://doi.org/10.1038/s43016-021-00279-9.
Zhao, B., Jin, W., Del Ser, J., Yang, G., 2023. ChatAgri: Exploring potentials of ChatGPT on cross-linguistic agricultural text classification. Neurocomputing 557, 126708. https://doi.org/10.1016/j.neucom.2023.126708.
Zilberman, D., Reardon, T., Silver, J., Lu, L., Heiman, A., 2022. From the laboratory to the consumer: Innovation, supply chain, and adoption with applications to natural resources. Proceedings of the National Academy of Sciences 119, e2115880119. Doi: 10.1073/pnas.2115880119.