agriculture

Article

Analysis of Farm Data License Agreements: Do Data Agreements Adequately Reflect on Farm Data Practices and Farmers' Data Rights?

Jasmin Kaur and Rozita Dara *

School of Computer Science, University of Guelph, Guelph, ON N1G 2W1, Canada
* Correspondence: drozita@uoguelph.ca

Abstract: Farm data license agreements are legal documents that play an important role in informing farmers about farm data processing practices such as collection, use, safeguarding, and sharing. These legal documents govern the exchange, access, and dissemination of farm data and are expected to provide legal protection against misuse of data. Despite their significant influence on farm data processing and governance, there is limited understanding of the content of farm data license agreements and standards for drafting them. Although online privacy policy content has been extensively studied, farm data agreements' evaluation and analysis have been overlooked. This study aims to investigate the structure, content, and transparency of farm data licenses. We collected 141 agricultural terms of use agreements and used natural language processing methods such as keyword and keyphrase analysis to perform text feature analysis, Flesch Readability Ease Score and Flesch Grade Level readability analysis, transparency analysis, and content analysis to gain insight into common data practices adopted by the agriculture technology providers. We also manually reviewed these agreements to validate the results and strengthen the observations. The findings show that data agreements are long, complex, and difficult to read and comprehend. The results suggest that 95% of the agreements fall under the difficult-to-read category and close to 75% of the policies require university-level education to understand the content. Furthermore, it is noted that some of the data management practices are not given adequate attention and are not as frequently mentioned in the agreements as expected. Finally, our analysis enabled us to provide recommendations on the content of farm data license agreements and strategies to improve them.

Keywords: farm data license agreement; farm data practice analysis; natural language processing; privacy regulations; farm data codes of conduct

Citation: Kaur, J.; Dara, R. Analysis of Farm Data License Agreements: Do Data Agreements Adequately Reflect on Farm Data Practices and Farmers' Data Rights? *Agriculture* **2023**, *13*, 2170. https://doi.org/10.3390/agriculture13112170

Academic Editor: Sanzidur Rahman

Received: 14 October 2023
Revised: 17 November 2023
Accepted: 17 November 2023
Published: 20 November 2023

1. Introduction

The growth in the use of advanced digital technology in the agricultural domain has led to a transformation of the agricultural sector. Digital and robotic technologies are being used to improve productivity and efficiency at farms, which has led to the phenomenon referred to as 'smart farming' [1]. Smart farming uses digital technology, sensors, the Internet of Things (IoT), and Artificial Intelligence (AI) to automate and address issues on farms. For example, the sensors can monitor the soil conditions and determine the water levels of the farm and can communicate with the IoT-connected sprinkler system to water the farm accordingly. There are many applications of smart farming which include monitoring crop fields (based on humidity, temperature, and soil moisture), automating irrigation systems, greenhouse automation, livestock monitoring, and autonomous milking robots [1].

With the use of these technologies in smart farming, a large amount of data is being generated and collected from farms. It is estimated that there will be over four million data points generated on an average farm per day by 2035 [2]. The data collected from the farms can include information related to the farms (e.g., location, weather), farming operations (e.g., production, harvesting), and farm management (e.g., fertilizers, animal feed). It can include agronomic data, land data, farm management data, machine data, and farmers' personal data. To obtain the value from the data collected from farms, the data are analyzed and shared with other stakeholders to develop/improve technologies and facilitate decision-making to enhance productivity and profitability [3,4].

The success of smart farming depends on the willingness of farmers to share their farm data with other stakeholders such as the Agricultural Technology Providers (ATPs) or the agribusinesses that develop the agricultural technologies using data collected from the farms [5–9]. The increasing collection of farm data, which include farmers' personal data, has raised concerns about privacy of farmers and misuse of information without farmers' awareness.

A survey conducted by the American Farm Bureau in 2014 found that 77.5% of farmers are concerned about sharing their data without their permission [10]. This survey illustrated that farmers have privacy concerns with sharing of farm data [11]. Farmers also expressed their lack of trust in the ATPs since there is a lack of transparency in the collection, sharing, and processing of farm data [6]. Farmers' responses showed the tension between farmers and ATPs regarding data ownership, i.e., who owns their farm data. The issue of data ownership arises because the farmers assume that they own the farm data and have some level of control over them since they are collected on their farms; however, some ATPs consider themselves the owner of the collected data [12]. Lastly, farmers feel at risk of being exploited by the ATPs since the farmers' data could be used for other business purposes and would be profitable to the ATPs only [11]. This lack of appetite for farm data sharing can cause hurdles in optimizing smart farming practices and, in turn, slow down food production and security.

The other implications resulted from the use of new technologies in smart farming include social, legal, and ethical issues. These issues can affect various stakeholders in the agri-food chain including farmers. Some of the socio-ethical challenges include inequitable development of farms, lack of farmers' control over data and technology, and the power imbalance. The unequal distribution of benefits from technology, leading to inequitable development of big vs. small farms in urban and rural regions, has led to concerns related to justice, equity, and fairness which has resulted in deepening the digital divide [6,13,14]. Furthermore, there are differences in the level of digital knowledge and expertise between ATPs and farmers. The ATPs use farmers' data to develop digital farming technologies and, therefore, have a higher level of digital knowledge and expertise in collecting, aggregating, and analyzing farm data. This can lead to a power imbalance where the ATPs assert power by being dominant in establishing processes and protocols as compared to the farmers [6,15].

There are other legal challenges of smart farming which include the lack of regulatory frameworks for the protection of agricultural data. Farm data protection and terms of use are usually established through legal contracts which we refer to as data license agreements in this paper. These contracts are usually written in complex legal language and can be one-sided to support the ATPs' mandate.

With the lack of appropriate legal solutions to deal with smart farming data issues, there have been attempts to formulate guidelines and protocols to improve agricultural data management practices. There are voluntary codes of conduct and best practices that are commonly used in different regions around the globe. For example, the European Union Code of Conduct on agricultural data sharing by contractual agreement ("the EU Code of Conduct") was developed to provide guidance on the use of agricultural data. The EU Code of Conduct has instructions for drafting the content of farm contractual agreements or data license agreements [16].

Data license agreements are documents that have information on what data are being collected from farms, with what organizations data are being shared, for what purpose, and for how long data are kept, and in what format. A data license agreement is a formal contract that should explain data governance and exchange practices and the appropriate use of the data. These data agreements, if developed appropriately using the guidelines provided by codes of conduct, can help in building trust between farmers, ATPs, and other stakeholders. If the terms and conditions of data collection, processing, retention, and sharing in the license agreements are explained transparently and clearly, it can help improve the trust and confidence of farmers in the technology and can facilitate more effective data use and exchange [5,6].

Past research has shown that farmers rarely read the terms and conditions in the data license agreements before agreeing to them [16–19]. This is due to the fact that these agreements are long, complex, and difficult to read because of the legal terminologies used in these documents [20]. Furthermore, farmers do not have a thorough knowledge of terms and conditions related to farm data due to the use of legal terminologies in these legal documents and their limited legal and technology literacy [21]. Research has also suggested that farmers do not trust the ATPs because of a lack of transparency and clarity on data practices and are also skeptical about sharing their farm data with the ATPs or other farmers and stakeholders [6]. Moreover, there are no standardized practices for writing the data license agreements and their content. These issues have highlighted the need for analyzing and understanding the content of farm data license agreements (FDLAs) and investigating common practices in drafting them. To solve some of these issues, some codes of conduct (e.g., the Australian Farm Data Code, Ag Data's Core Principles) have put emphasis on the need for transparency and clarity of data practices in the agreements and how they should be written in plain language.

Adding to the complexity, there is limited knowledge of the content of FDLAs. There is limited information on how closely ATPs follow these and other recommendations in their data agreements. This reinforces the need for analyzing the content of data license agreements in the agricultural domain. We also aim to investigate the length, complexity, and readability of these agreements and examine the level of digital/legal literacy required to understand them. There have been some efforts to review and understand the codes of conduct and their impact on digital agriculture. However, to the best of our knowledge, there have been limited efforts to analyze FDLAs to evaluate data practices, common data standards and procedures, transparency, and other textual features of these documents. This paper aims to address this gap in the digital agriculture domain.

This paper is aimed at evaluating these agreements to understand their structure, strengths, and deficiencies. Our analysis is guided by these two questions: RE 1: How do the text features, including sentence complexity, ambiguous words, and length, influence the readability and transparency of FDLAs? RE 2: To what degree do the key topics and terminologies used in FDLA comply with and meet the content requirements by regulations and best practices? The findings of this study can help enhance data governance in the agriculture sector by developing better data agreements and policies that promote the responsible use of farm data which benefits all the stakeholders involved.

We analyzed 141 agreements by using Natural Language Processing (NLP) to perform text feature analysis, readability analysis, transparency analysis, and content analysis. We also analyzed the agreements based on a set of predefined questions to find out how commonly the ATPs refer to important topics recommended by regulations and codes of conduct such as data collection, ownership, data portability, data sharing, and data security in the agreements.

The objective of our analysis is to understand how the FDLAs are written, what data practices are included, and what practices/recommendations are not satisfied in the agreements. To the best of our knowledge, this research is unique in the following ways: (1) this is the first attempt that examines the content of FDLAs; (2) we proposed a novel framework for the analysis of the content of these agreements. This framework was developed based on well-established best practices and regulations; (3) we used NLP to perform a comprehensive keyword topic analysis. The proposed approach also evaluates readability and other features of the agreements. We have provided examples of the content to further enhance our analysis; and (4) the findings and recommendations provided can be used by agri-food technology providers and policymakers to create more effective data agreements to improve the legal and ethical data collection and usage practices in the agricultural sector.

The paper is divided into the following: Section 2 discusses the related work. Furthermore, Sections 3–5 explain motivation, proposed approach, and methodology, respectively. The results are presented in Section 6. Section 7 includes a discussion and recommendations. Finally, we conclude with the conclusion and future work.

2. Related Work

In emerging economies, data contracts in the agricultural supply chain help in managing the socio-ethical challenges by enabling farmers to make decisions, collaborate, and trust other stakeholders. Data contracts in the agricultural supply chain with provisions for inputs, credits, and output procurement are crucial. This helps in empowering small farmers and linking them to the agricultural supply chain [22]. There has been work carried out in designing contracts and using transaction cost theory to resolve the socio-ethical challenges. Dutta et al. developed a decision-making model to help design contracts for firms or companies for procuring small farmers in emerging economies [23]. The authors suggest that their analysis could help firms and policymakers design socioeconomically viable agri-supply chain contracts. Sgroi et al. used transaction cost theory to explore solutions to help small farms recover competitiveness and other territorial imbalances and marginalization of small agricultural activities in developed economies [24]. It is noted that cultivation contracts are a valuable tool for addressing the marginalized problems faced by many agricultural companies.

Despite the effort, there have been no attempts to investigate, analyze, and understand how FDLAs are written and whether they satisfy recommended practices. On the other hand, online privacy policies have been the subject of attention in recent years [25–27]. Kaur et al. analyzed privacy policies from diverse corpora of policies which included health, financial, and other application domains and policies from different countries including Canada, USA, and Europe. The authors extracted keywords using a topic modeling method to compare the content of policies. The authors also performed a comprehensive analysis of keywords for each of the data practices, coverage of data privacy practices, and ambiguity analysis of words in the privacy policies. The authors found that privacy policies use ambiguous words frequently which results in less transparent policies. The authors also noticed that some of the privacy categories such as data collection are given more extensive attention in the policies while important topics such as security and choice have been overlooked [25]. Amos et al. conducted an automated analysis of privacy policies to understand the privacy policies over time. The authors investigated the readability and text features from 2009 to 2019. The authors concluded that the privacy policies are long, hard to read, and lack transparency in some areas such as tracking cookies and data sharing with third parties [28]. Bateni compared the privacy policies before and after the General Data Protection Regulation (GDPR) [24,27,29]. The authors concluded that GDPR enforcement helped in improving the content of privacy policies, however, many policies did not satisfy the GDPR requirements for these legal documents.

Although online privacy policies have been extensively analyzed, the evaluation of farm data practices has been overlooked. We aim to address this gap by developing an evaluation framework and using methods such as NLP to analyze FDLAs.

3. Proposed Approach

In this paper, we developed a framework to evaluate different aspects of FDLAs. Figure 1 provides an overview of the proposed framework. The different components of this framework are developed by taking into consideration the privacy regulations and best practices and recommendations provided by the farm codes of conduct such as Farm Data Code [30]. Many privacy regulations and best practices have guidelines for data protection and how the data agreements should be written. For example, Article 12 of the GDPR mentions the use of clear and plain language so that it is understandable to a child [31]. Hence, we intend to perform a readability analysis to understand how FDLAs score in terms of readability. Furthermore, we intend to investigate the length of the sentences, since many data protection organizations such as the FTC recommend using short and effective privacy protection including disclosure [32]. Additionally, transparency is an important issue emphasized in privacy regulations and farm codes of conduct. For instance, the GDPR emphasizes the importance of avoiding ambiguity in the agreements [31]. To this end, we investigated the transparency of FDLAs, by analyzing the use of ambiguous words that can hamper transparency. Also, the privacy regulations (e.g., the GDPR) and data protection guidelines (e.g., the OECD [33], the Federal Trade Commission's Fair Information Practices (the FTC FIP)) have recommended important topics (e.g., data practices) that should be included in the data agreements such as online privacy policies. Some of the important sections in the data agreements are notice, data sharing, data collection, data access, data retention, and cross-border data transfer.

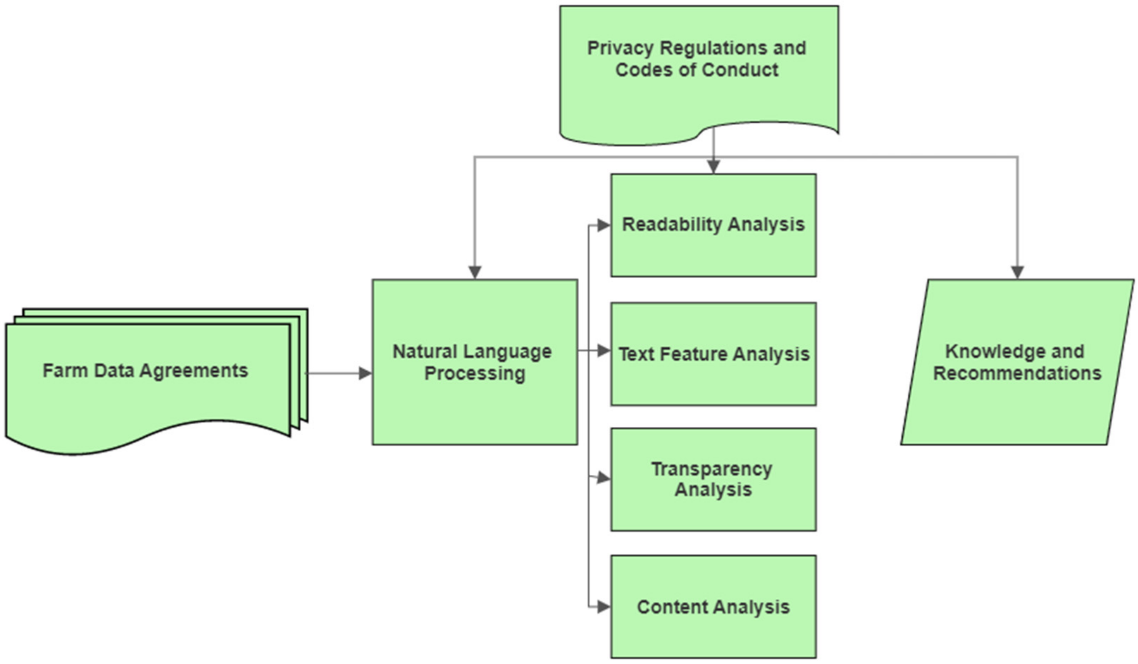

Figure 1. Overview of the framework for data license agreement evaluation.

Similarly, some codes of conduct have recommendations on the farm data management practices and data protection principles for ATPs and what should be included in the data agreements. The American Farm Bureau Federation (AFBF), for example, has formulated a set of principles called the Privacy and Security Principles for Farm Data (or Core Principles) in which the companies or ATPs should follow the principles related to data ownership, data collection, data access and control, data portability, and farmers' choice/control over the data management. In addition to these practices, the EU Code of Conduct has recommended including information related to data protection, data privacy, and security in the contracts. All the farm data governance best practices, including the agriculture codes of conduct, have put emphasis on transparency of data contracts. Therefore, we tend to investigate whether these important topics are discussed in the FDLAs and to what extent they have been given attention.

Through a comprehensive literature review of existing best practices, prior work by the OECD, the FTC FIP, privacy regulations, farm codes of conduct, and related publications, we developed the proposed framework. The core steps of the framework include four components, readability analysis, text feature analysis, transparency analysis, and content analysis, which are discussed in the following Section. We collected and analyzed 141 ATP data license agreements that were available online using the proposed framework. These data were collected between May to June 2021. Our approach included the use of NLP to pre-process, tokenization of the FDLAs and preparing the dataset for further analysis.

The proposed framework has four components that are designed to evaluate high-level structure of data agreements, ease of comprehension, legal terminologies, ambiguous language, and existence of recommended topics in the agreements:

1. Text feature analysis: Text feature analysis assists with evaluating the appearance of data agreements by analyzing the average length of sentences, average length of data agreements in terms of words, and passive voice index;
2. Readability analysis: Readability is defined as the ease of reading and understanding the text. This measure depends on the complexity of the text and considers features such as length of sentence, number of words, number of syllables, number of letters, etc. It gives insight into how easy or difficult it is for the reader to understand the text;
3. Transparency analysis: The concept of transparency can be broad. According to the GDPR, any information addressed to the individuals or public should be easy to understand, clear, in plain language, and concise. Some of these attributes cannot be easily measured. However, we have limited the analysis to the use of ambiguous words as they can be identified and measured. Generally, using ambiguous words can hamper the clarity of the reading materials. Several privacy regulations have also recommended reducing the use of ambiguous words to ensure transparency;
4. Content analysis: The analysis of the content of FDLAs helped us investigate the common data practices for processing farm data and the ones that are overlooked. This was achieved by 12 pre-defined questions that were created using recommended topics for online privacy policies by best practices and regulations such as the OECD, the FTC FIP, and the GDPR, and the content recommended by farm codes of conduct and Ag Data Transparent [34].

4. Methodology

4.1. Dataset

For our analysis, we manually compiled the data license agreements by visiting the online websites of ATPs or companies from diverse sectors such as crops (e.g., Climate FieldView, San Francisco, CA, USA), livestock (e.g., Afimilk, HerdDogg), agricultural machinery (e.g., John Deere, Maureen, IL, USA), security/monitoring IoT companies, farm management software (CattleMax, Farm at Hand by TELUS Agriculture, College Station, TX, USA), IoT-based/remote sensing (e.g., METOS, Precision Hawk, Raleigh, NC, USA). In several cases, we contacted the organizations and validated whether their online agreements are in fact the agreements that they share with farmers who use their services and technologies. In this process, we had to discard a number of policies that were not relevant. We collected used FDLAs in our experiments.

4.2. Natural Language Processing

NLP is commonly used for automated computational processing of human languages. With NLP techniques, algorithms can analyze and model massive amounts of unstructured data such as text and voice [35]. Data preprocessing is the key step in text processing and includes tokenization, stop word removal, and stemming. We used Natural Language Tool Kit (NLTK 3) and Python 3.7 for the data pre-processing.

Tokenization: This is a process in which the text is segmented into words or phrases which are referred to as tokens. These tokens are then used to analyze the text. For tokenization, data cleaning is performed to convert words into lowercase and punctuation marks are removed [35].

Stop word removal: The words which are not useful for analysis such as conjunctions, some verbs, and names are referred to as stop words. Since these words do not help in providing useful information related to objective of analysis, they are removed from future steps [36]. A list of stop words is created and these words are deleted while processing the text.

Stemming: This is the process in which the word is condensed to its root or the base words. This helps in removing redundancy and helps in improving the reliability of findings [36].

4.2.1. Text Feature Analysis

Text feature analysis measures appearance of text such as the average number of sentences, average length of FDLAs in terms of words, syllable count, and passive voice index. These metrics illustrate the high-level structure of the FDLAs.

1. Average number of sentences: This calculates the average number of sentences in the agreements;
2. Average length of data agreements in terms of words: This calculates the average number of words in a farm data license agreement;
3. Syllable count: This measures the number of syllables in the agreement;
4. Passive Voice Index: This provides information regarding the sentences that contain passive verb forms. We used the Spacy library to perform dependency parsing and identify passive sentences. We considered the pattern of Nsubjpass (that is Nominal subject (passive)), followed by aux (Auxiliary), and then followed by Auxpass (Auxiliary (passive)) to compute the passive voice.

4.2.2. Readability Analysis

Flesch Readability Ease Score (*FRES*): The Flesch Readability Ease Score assesses the ease by which the user can understand the text [37,38]. It considers the average number of syllables per word, average length of sentences, and other predefined factors to measure the readability of the text and outputs a number between 0 and 100 [21]. The *FRES* is calculated using Equation (1) shown below. The building blocks of this score are sentence length, the average number of words in a sentence, and word length, the average number of syllables in a word. A higher score means that a document is easy to read, and a lower score means that it is more difficult to comprehend.

$$FRES = 206.835 - 1.015\left(\frac{words}{sentences}\right) - 84.6\left(\frac{syllables}{words}\right) \tag{1}$$

Based on the *FRES*, the user can evaluate the readability of text. For example, if the score is in the range of 90–100 then the text is very easy, a score in the range of 80–89 is easy, 70–79 score is considered fairly easy, 60–69 score is considered standard, 50–59 score means the text is fairly difficult, 30–49 means that the text is difficult and 0–29 score signifies that the text is very confusing.

Flesch Grade Level: This is a widely used readability measure that assesses the reading grade level of a text [39]. It considers the sentence length and number of syllables and other factors in order to calculate the *FGL* score [39]. The *FGL* score predicts the grade level required to comprehend the text (Table 1). For example, an *FGL* score of 7.5 means that a seventh grader would be able to read the document. The *FGL* score is calculated as shown in Equation (2):

$$FGL = 0.39\left(\frac{words}{sentences}\right) + 11.8\left(\frac{syllables}{words}\right) - 15.59 \tag{2}$$

Table 1. Flesch Grade Level.

Flesch Grade Level	Readability Level
1–6	10 (Very easy)
7	9
8	8
9	7 (Easy)
10	6
11	5
12	4 (Difficult)
College/University (Freshman year)	3
College/University (Sophomore year)	2
College/University	1 (Very difficult)

4.2.3. Transparency Analysis

To evaluate the transparency of data agreements, we counted the frequency of ambiguous words in the data agreements and their occurrence with the important data management practices. We also used the ambiguous words that were taken from [25]. These words are listed in Table 2. Important keywords are summarized in Table 3.

Table 2. List of ambiguous words for transparency analysis.

Ambiguous Words
occasional, will, perhaps, such, some, certain, various, reasonable, like, example, sometimes, depending, necessary, appropriate, inappropriate, generally, mostly, widely, general, commonly, usually, normally, typically, largely, often, may, might, can, could, would, likely, possible, possibly, unsure, anyone, certain, everyone, numerous, some, most, few, much, many, various

Table 3. List of questions and keywords used for content analysis.

Questions	Keywords
What categories of farm-related data does the product or service collect from my farm?	soil, fertility, topographical, elevation, watershed, drainage, geospatial, tillage, conservation, plant, seed, yield, disease, pest, fertilization, financial, tax, employ, commodity, supply, chain, fuel, genetic, animal, health, reproduction, mortality, feed, health, machine, agronomic, land, nutrient, asset, camera, imagery, sensor, debt, history, fax, bank, equipment, irrigation, nonidentifiable
What categories of personal (sensitive) data does the product or service collect from me?	collect, person, equipment, phone, number, address, name, age, phone, zip, postal, code, contact, product, financial, voluntarily, location, identifiable, email
Do the service provider agreements address ownership of my farm data after the data are collected and/or transferred? (ownership as a power to control)	owner, ownership, right, grower, customer, originate, license, access, control, change, sell, download, delete, disclose, transfer, retain
If the service provider gets into a contract with other companies to provide data-related services, does the service provider require these companies to adhere to the original data agreements (privacy policies) that I have agreed to?	third, party, confidential, share, agreement, partner, abide, comply, obligation, vendor, accordance, terms, protect

Table 3. *Cont.*

Questions	Keywords
Will the service provider obtain my consent before providing other companies with access to my data?	consent, access, authorize, third, party, option, opt-in, opt-out, permission, consult, obtain, purpose, notice, choice, disclose, disclosure
After I upload data to the service provider's servers, will it be possible to retrieve my original complete dataset in an original or equivalent format?	download, original, convert, access, retrieve, control, portable, digital, format, view, request, electron, copy, available, written, request, export, readable
Will the service provider notify me when its agreements change?	change, notice, consent, inform, notify, update, customer, agreement, term, time, email, log, regularly, effect, publish, promptly, announce, amend, revise, periodically, aware
Will the service provider notify me if a breach of data security occurs that causes disclosure of my data to an outside party?	secure, compromise, unauthorize, attack, threat, breach, notice, notify, contact, disclose, ransomware, promptly, inform, aware, email, compromise, disaster, unforeseen, regulation
Upon my request, can my original dataset be deleted when my contract with the service provider terminates?	request, delete, remove, termination, choose, retention, cancel, backup, erase, erasure, comply, obligation
Do the service provider agreements establish how long my original datasets will be retained?	request, delete, store, retain, record, profile, keep, backup, discard, terminate, destroy, retain, period, time, retention, close
Do the service providers agreements address what happens to my data if the service provider is sold to another company?	notice, sell, sold, agreement, adhere, abide, condition, term, notify, consent, merger, acquisition, sale, asset, transfer, disclose, option, delete, comply, transaction, inform
Who can we contact in the company if we have questions about farm data?	contact, custom, help, question, request, inquiry, concern, direct

4.2.4. Content Analysis

The content analysis was performed using two experiments: by analyzing the words and phrases (unigram, bigram, trigram) and by analyzing the keywords that represented possible answers to important questions that arose from recommended legal/data practice topics. We further analyzed relevant text to better understand the use of these keywords/keyphrases in the agreements.

1. Analysis of words and phrases

We used NLTK 3 to gather the unigram, bigram, and trigrams for the FDLAs after splitting the sentences into tokens. We extracted the most relevant and/or frequent unigrams, bigrams, and trigrams along with the number of data agreements they are used in;

2. Keyword Analysis for Predefined Questions

We formulated a set of 12 questions that were inspired by recommended topics for online privacy policies by best practices and regulations including the OECD, the FTC FIP, and the GDPR [25] and the questions proposed by Ag Data Transparent [40]. These questions are related to how the ATPs collect, share, use, and protect data collected from farms. The list of 12 questions is given below in Table 3. We ensured that these questions covered all the recommended data practices by the farm codes of conduct and privacy best practices.

We extracted keywords related to the questions as recommended by past research [25,41] and also by manually reviewing 70 agreements and selecting the most relevant keywords related to the question. For complex legal questions such as the ones related to data ownership, we selected words that directly relate to owning farm data (e.g., ownership, owner) as well as the choices/options the farmers are given to control and manage their farm data (e.g., keywords such as access, control). The other questions were somewhat more straightforward in terms of identifying relevant keywords that represent the topic/potential answers. The keywords are listed below in Table 3.

We also reviewed several data agreements for each of the 12 questions to gain an understanding of how relevant data practices are addressed in the document. We then extracted sample text from the data agreements using keywords associated with the questions to investigate the topics that are given the most attention, the ones that are neglected, or statements that lack consistency or are ambiguous. Results are presented in the following sections.

5. Results

We present the results of FDLAs analysis in this section. Results are presented for each component of the framework separately in the following subsections.

5.1. Readability Results

We used the FRES to test the reading difficulty level of the agreements. Table 4 shows the readability results of the FRES. The results show that around 95% of policies fall under the fairly difficult, difficult and very confusing in terms of readability. Only 4.9% of policies are readable according to the FRES.

Table 4. Readability results using Flesch Readability Ease Score.

Flesch Reading Ease Score	Number of Agreements
90–100 (Very Easy)	1
80–89 (Easy)	2
70–79 (Fairly Easy)	3
60–69 (Standard)	1
50–59 (Fairly difficult)	8
30–49 (Difficult)	87
0–29 (Very confusing)	39

We also used the FGL, a readability metric that assesses the approximate reading grade level required to understand the agreements. Table 5 shows the results of FGL scoring. The results from the FGL score show that around 75% of the policies require university-level education to understand the text. Only 7.09% are easy to read and require a grade 9 level education to understand the text.

Table 5. Readability results using the Flesch Grade Level.

Flesch Grade Level	Readability Level	Number of Farm Data Agreements
1–6	10 (Very easy)	5
7	9	3
8	8	0
9	7 (Easy)	2
10	6	3
11	5	7
12	4 (Difficult)	14
College/University (Freshman year)	3	22
College/University (Sophomore year)	2	28
College/University	1 (Very difficult)	57

The FRES and FGL scores suggest that the majority of data agreements are difficult to read and comprehend due to their length and complexity. The use of long sentences with long and complex words has made reading and understanding the agreements more difficult.

Observation: We compared the agreements based on the FRES and FGL scores. The following are two sample texts related to data security taken from two different agreements. It is observed that different word choices and lengths of sentences can impact the readability measure. The FGL scores here represent the grade level required to understand the data agreements (see Tables 4 and 5).

Sample text 1 (FRES: 26.34 and FGL score: 16.5): "Our company" has implemented an information and data security program that contains administrative, technical and physical controls that are designed to reasonably safeguard Personal Information and Machine Data. For example, we use industry-standard encryption technology to secure your Financial Account Information and other sensitive Personal Information when it is being collected and transmitted over the Internet.

Sample text 2 (FRES: 63.61 and FGL score: 9.8.): All data is encrypted when transmitted from our servers to your browsers—this is the same connection method used by banks. The database backups are also encrypted.

5.2. Text Feature Analysis

The results of the text features such as syllable count, sentence count, average sentence length, and passive voice index are given below in Table 6. It is noted that the average number of sentences in the data agreements is around 122 with 26 the average number of words in the sentence.

Table 6. Results of Text Feature Metrics.

Text Metrics	Average
Syllable count	11,029
Average Sentence Length	26.21
Sentence count	122.81
Passive Voice Index	15.53

We compared our results with the results of [27] where the authors analyzed and compared data agreements before and after the GDPR was established (Table 7). The GDPR came into effect on 25 May 2018 and had a significant impact on the content of online privacy policies. The GDPR puts emphasis on the readability of online privacy policies. One aspect of readability is shorter sentences. GDPR also mandates transparency for these policies. We compared our results with the findings of [27] and noted that the FDLAs have extremely large sentence lengths.

Table 7. Comparing results with the data agreements before and after the GDPR [27].

Features	Privacy Policies Before GDPR	Privacy Policies after GDPR	Farm Data Agreements
Average number of sentences	106.52	213.22	122.81
Average length of agreements in terms of words	1549.47	1601.10	3101.43

Generally, lengthy sentences can be difficult to understand [42]. The length of sentences is known to be correlated with the difficulty of comprehending text. On the other hand, shorter and simpler sentences can be easy to understand.

We also analyzed the results of the passive voice index for each agreement. Passive voice index helps identify the sentences that contain passive voices. The passive voice generally creates unclear, wordy, and unconcise sentences which can lead to ambiguity in the meaning of the sentences. The results of the passive voice index show that the average of passive voice sentences is 15.53 per agreement. The GDPR has recommended avoiding the use of passive voice sentences or ambivalent sentences that could lead to different interpretations which, in turn, hampers transparency of the agreements [43].

We also investigated whether there is any correlation between the FRES and Passive Voice index. The correlation between the two metrics is shown in Figure 2. The figure illustrates that there is a slight negative relationship between FRES and Passive voice. These results are similar to Li et al. [44]. This suggests that the use of passive voice in the data agreements decreases the readability of the agreements. For instance, the number of passive sentences in Farmbeats (Microsoft) data agreement is large, and the FRES is 40.69 which suggests that this agreement is difficult to read. But in general the correlation is not large.

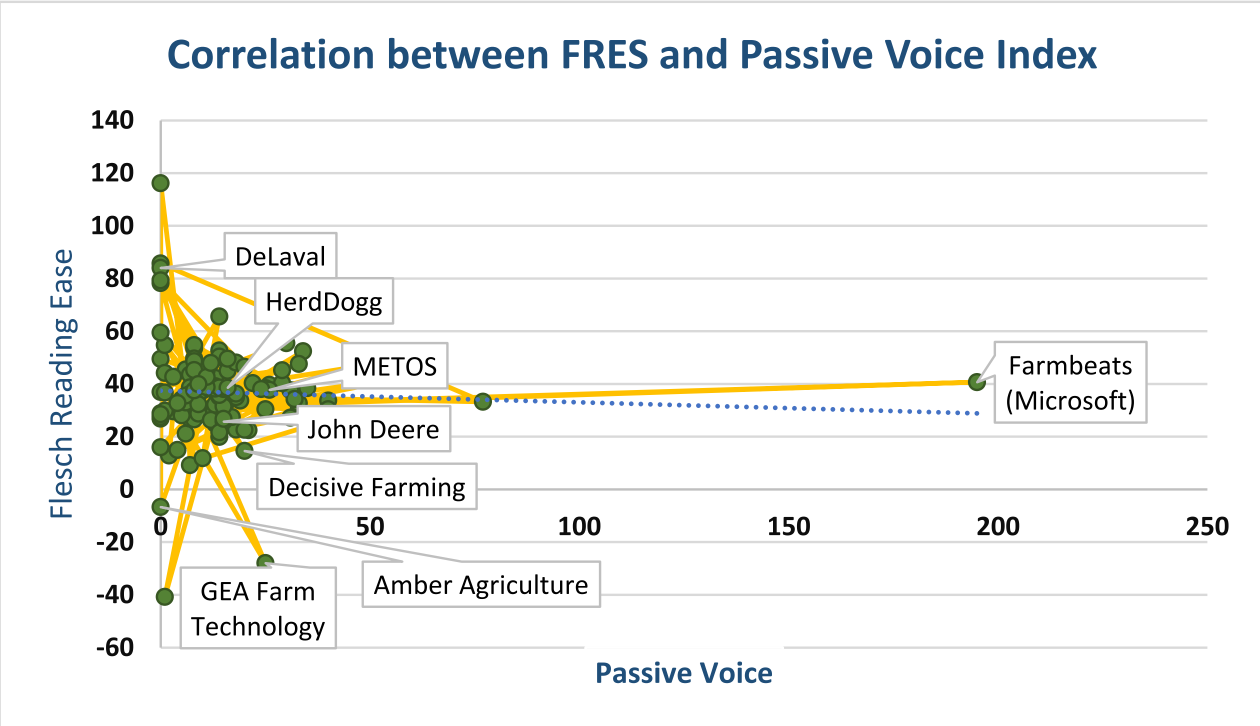

Figure 2. Analysis of correlation between Flesch Reading Ease Score and Passive Voice Index.

5.3. Transparency Analysis

The GDPR recommends limiting the use of ambiguous words as they decrease the transparency of online privacy policies. Most farm codes of conduct also put emphasis on the transparency of farm data contracts. To this end, we analyzed the use of ambiguous terminologies in the corpus of data license agreements that we collected as one of the indicators of transparency. The results of the most frequent ambiguous words are presented in Table 8. It can be observed that *may*, *will*, *such*, *can*, and *certain* are some of the most frequent ambiguous words used in the agreements.

The following are sample texts with ambiguous words that are taken from different agreements. The sample text shows that the agreements use ambiguous words. The use of these words will affect the ability of the farmers to make informed decisions about ATPs reliability and transparency of service.

Table 8. Most frequent ambiguous words.

Ambiguous Words	Frequency	Number of Agreements
may	3697	125
will	2070	124
such	1847	120
can	1589	123
certain	542	104
example	512	89
some	401	97
necessary	382	102
like	263	69
reasonable	216	88

Sample text 1 for transparency analysis: We **may** also disclose personal data as part of a corporate transaction such as a merger or sale of assets.

Sample text 2 for transparency analysis: While we make all reasonable efforts to ensure that such information is anonymized, it is **possible** that your Personal Data may be disclosed.

Sample text 3 for transparency analysis: We take **reasonable** steps to ensure that the data we collect is reliable for its intended use, accurate, complete, and up to date.

5.4. Content Analysis

In this section, we present the results of our effort to perform content analysis of data license agreements through both quantitative and qualitative approaches. We first extracted frequently used relevant words/phrases and then reviewed some samples.

5.4.1. Frequent Words and Phrases

We analyzed the most frequent words (unigrams) and phrases (bigrams, trigrams) used in the agreements to obtain insights into the content of data agreements.

The results of most frequent words (unigrams) are shown in Table 9. It is noted that the most frequent words used in the agreements are *may*, *collect*, *contact*, *third*, *party*, *secure*, *access*, *request*, *email* and *purpose.* "Collect" is one of the most frequent words that potentially refers to "collection" practices of the ATP. Similarly, words such as *cookie*, *secure*, *access*, *access* and *third party* are used frequently. Each of these important words refers to a different data practice [25].

The results of some of the most frequent phrases (bigrams and trigrams) are shown in Table 10. Additional results of the frequent phrases (bigrams and trigrams) are presented in the Supplementary Sections S1 and S2. The most frequently used bigrams are third party (in 83.6% of agreements), contact us (in 80.85% of agreements), email address (in 75.88% of agreements), personally identifiable (in 47.51% of agreements). The results of trigrams show that the agreements discuss various data practices such as sharing personal information, disclosing personal information, protecting personal information, policy change, and using and collecting personal information.

Table 9. Results of most frequent words.

Words (Unigram)	Frequency	Number of Agreements
may	3697	125
collect	2472	123
contact	1313	122
party	1294	115
secure	1287	90
access	1263	120
request	1128	119
email	1120	113
purpose	1106	116
third	1097	121

Table 10. Results of frequent phrases (bigram and trigram).

Bigram	Frequency	Number of Agreements	Trigrams	Frequency	Number of Agreements
third, party	1073	118	personal, information, collect	147	51
service, provider	560	98	disclose, personal, information	145	52
contact, us	526	114	please, contact, us	128	68
may, use	383	98	inform, third, party	115	63
email, address	348	107	share, personal, information	111	37
use, information	345	93	use, personal, data	96	19
collect, use	329	102	disclosure, personal, information	84	41
information, may	316	91	third, party, service	76	34
personal, identifiable	214	67	access, personal, information	75	37
IP, address	222	84	protect, personal, information	72	40

Below are samples of text taken from various agreements to show the context in which some of these phrases such as *third party*, *share personal information*, and *email address* are used.

Sample text related to 'third party': If we transfer any personal information to **a third party** subcontractor, we will provide the subcontractors only with the information needed to perform the subcontracted service, and will use appropriate contractual or other means to provide a comparable level of protection while the information is being used by them. However, you agree not to hold us liable for the actions of **any third party** subcontractor, even if we would normally be held vicariously liable for their actions, and understand that you must take legal action against them directly should they commit any tort or other actionable wrong against you.

Sample text related to 'Share Personal Information': We may need to share personal information with our third-party service providers in order to provide products or services to you, and we may share your personal information only to the extent that it is related to such transaction or service.

Sample text related to email address: We will collect information such as your name, job title, **email address**, phone number, postal address, company information (such as business name, farm name, website information), your industry sector, your communication preferences, your reason for contacting us, and the contents of your communication to us.

5.4.2. Keyword Analysis for Questions

As per discussed above, we created 12 questions (see Table 3) regarding ATPs' data practices and used these queries to further analyze the agreements. The following are the results from the analysis of keywords for these questions. Analysis of some of the questions is presented in the Supplementary Materials Section S3.

Question 1: Do the service provider agreements address ownership of my farm data after the data are collected and/or transferred?

The results presented in Table 11 show that *ownership* was mentioned in 19 policies only (13.47% of policies) which suggests that the keyword "*ownership*" is not frequently used in FDLAs. The codes of conduct (e.g., the EU Code of Conduct on agricultural data sharing by contractual agreement) and best practices (e.g., Ag Data's Core Principles) suggest that farmers are the owners of the data since data are generated on their farms or during their farming operations and the farmers must have the right to access and control the data. This recommendation was not reflected in the data agreements. However, ownership is a complex legal concept, and a clear definition does not exist. It is also impossible to clearly define ownership of farm data as many farm technologies are usually owned/accessed by the ATP, farm data reside in cloud storage providers, farm data are transferred through the network provided by the communication technologies, and also third parties have access and process farm data.

Table 11. Results of keyword analysis for Question 1.

Keyword	Frequency	Number of Farm Data Agreements
access	1263	122
control	591	103
change	565	117
disclose	562	110
delete	489	94
transfer	463	97
retain	199	73
ownership	28	19
own	582	116

In addition to *ownership*, we analyzed the keywords that represented control over data such as *control*, *access*, *transfer*, *change*, *delete,* and other related words. These keywords illustrate whether ATPs allow farmers to have some level of control over their data, e.g., the ability to access or manage their data. It is noted that keywords such as *access (86%)*, *control (73%)*, *disclose*, *delete*, *transfer*, and *retain* are more frequently used in agreements in comparison to ownership. We also noted the word "own" is used in more than 70% of the agreements.

Observations: The following samples are taken from different data agreements which demonstrate how the term *ownership* is used in the agreements. It is noted that *ownership* is used in other contexts which are not related to ownership of farm data. For example, Sample text 2 mentions ownership of land and field boundaries. Another interesting example is Sample text 1. Although the service provider claims that farmers own their data, this claim does not include ownership of anonymized data. Interestingly, the service provider does not provide a clear explanation on what constitutes data ownership from legal and practical perspectives in their policies and also what constitutes anonymized data.

Sample text 1 related to Q1: Ownership of Your Data. At "our company", we strongly believe that You **own** Your Data. "Your Data" means all data or information, electronic or otherwise, that is submitted by you in connection with the use of our Services and is identifiable to you. You exclusively own all rights, title and interest in and to all of Your Data. We do maintain a license to include Your Data in our anonymized, aggregated database in which individually identifiable information is masked. Your **ownership** does not include any **anonymized databases** we create.

Sample text 2 related to Q1: Land data, such as Your ownership and lease information related to field boundaries.

Question 2: If the service provider gets into a contract with other companies to provide data-related services, does the service provider require these companies to adhere to the original data agreements (privacy policies) that I have agreed to?

The results illustrate that *third party* is the most frequent keyphrase used in the agreements (86.52%) (Table 12). *Share* is used 872 times in 77.30% of the policies. *Obligation* is used in 60% of agreements and *adhere* is used 31 times in around 14% of the contracts.

Table 12. Results of keyword analysis for Question 2.

Keyword	Frequency	Number of Farm Data Agreements
party	1294	124
third	1097	122
terms	889	123
share	872	109
agreement	369	79
partner	336	96
obligation	283	85
comply	241	104
responsible	161	64
adhere	31	20

Observations: Some of the agreements, such as in Sample text 1, require that third parties should adhere to the data agreement that the farmers have agreed to. Other service providers (such as in Sample text 2) have clearly mentioned that they do not take any responsibility for the data practices of the third parties and also that farmers should consult the third party’s data practices. Such statements are very concerning as they suggest third parties have no legal constraints over use and misuse of farm data. It is observed that most of the data agreements did not have comprehensive third-party data sharing policies.

Sample text 1 related to Q2: This Privacy Statement does not address, and we are not **responsible** for, the privacy, information or other practices of any **third parties**, including any **third party** operating any website or service to which the Services link. The inclusion of a link on the Services does not imply endorsement of the linked site or service by us or by our affiliates.

Sample text 2 related to Q2: Handling of information: Your Data is held by our related companies and **third party** service providers in accordance with this Privacy Policy and on servers operated by **third party** service providers.

Question 3: Will the service provider notify me when its agreement changes?

Among these sets of words, *email* is mentioned 1120 times in around 86% of the policies (Table 13). *Notify* is used 158 times in 58% of policies and *notice* is used in 70% of the policies. *Communicate* is used 554 times in 75% of the policies. *Update* and *log* are the other relevant frequent words related to this category.

Table 13. Results of keyword analysis for Question 3.

Keyword	Frequency	Number of Farm Data Agreements
email	1120	122
terms	889	123
consent	754	111
change	565	117
communicate	554	106
update	464	104
notice	424	99
agreement	369	79
effect	203	82
log	166	122
notify	158	82

Observations: The agreements do not refer to "change notifications" as frequently as expected and recommended. Instead, farmers are asked to read the current or updated version of the policies. For example, Sample text 3 mentions checking the policy to be aware of the changes. This is in contrast with many codes of conduct (e.g., the Australian Farm Data Code) and data privacy regulations (e.g., the GDPR) recommendations in which they put emphasis on "notice" and obtaining consent for any changes in data collection processing. This is one of the core steps ATPs can take to foster transparency. The GDPR goes one step further and recommends to companies to communicate the changes in such a way that the user will notice them [45]. This could be achieved in the form of sending notifications through email or adding information to the privacy settings, apps, and through other means. Despite these recommendations, some ATPs assume acceptance of new/revised agreements after "an update" without providing a user-friendly notice (Sample text 3).

Sample text 1 related to Q3: From time to time, we may **update** this Privacy Statement to reflect new or different privacy practices. We will place a **notice** on our homepage when we make material **changes** to this Privacy Statement. Additionally, if the changes will materially affect the way we use or disclose previously collected Personal Information, we **may also notify** you about the **change** by sending a **notice** to you.

Sample text 2 related Q3: This Policy was last updated on 24 May 2018. We reserve the right to **revise** this Policy at any time and, as such, you should review its terms each time that you use our website or application. Any **changes** to this Policy will be promptly communicated on this page, but will not go into effect until at least five (5) days after they are posted.

Sample text 3 related to Q3: The "Last **Updated**" legend at the top of this Privacy Statement indicates when this Privacy Statement was last revised. Any **changes** will become effective when we post the revised Privacy Statement on the Services. Your use of the Services following these changes means that you accept the revised Privacy Statement.

Question 4: Will the service provider notify me if a breach of data security occurs that causes disclosure of my data to an outside party?

The most frequent keywords related to this query are *secure*, *disclose*, and *unauthorize*. Keywords such as *threat*, *compromise*, *attack*, *disaster*, and *ransomware* are used very rarely in policies (Table 14). These keywords only exist in 1% to 17% of the policies. *Breach* is used 96 times in only 27% of the policies.

Table 14. Results of keyword analysis for Question 4.

Keyword	Frequency	Number of Farm Data Agreements
contact	1313	127
secure	1287	121
email	1120	122
disclose	562	110
notice	424	99
event	307	112
notify	158	82
unauthorize	114	65
breach	96	39
compromise	17	14

Observation: Data security is another important data practice that is recommended by all data privacy regulations and several codes of conduct. This data practice requires that ATPs clearly explain what security measures and safeguards they use in order to keep the farm data secure. This also includes notifying the farmers and other stakeholders about data breach incidents, in particular, if data breaches compromise farmers' personal/sensitive data. From our analysis, we have found that the ATPs mention data security measures in 86% of agreements. For example, Sample text 1 and Sample text 2 have referred to some data security practices such as encryption and password protection. However, many of the agreements do not mention notifying farmers/users about data breaches or data disclosure. Only a few companies (such as in Sample text 3) have mentioned that they would notify farmers about data breaches within two weeks. Furthermore, some of the companies (shown in Sample text 4) mention data security practices but do not guarantee that personal information will be secure.

Sample text 1 related to Q4: "Our company" stores your personal data on **secure** on-site servers or **secure** cloud servers. These servers are further secured by administrative restricted, **password** protected precautions.

Sample text 2 related to Q4: For example, we use industry-standard encryption technology to **secure** your Financial Account Information and other sensitive Personal Information when it is being collected and transmitted over the Internet.

Sample text 3 related to Q4: In order to be in line with Fair Information Practices, we will take the following responsive action, should a data **breach** occur: We will **notify** you via email within 2 weeks.

Sample text 4 related to Q4: We seek to use reasonable organizational, technical and administrative measures to protect Personal Information within our organization. Unfortunately, no data transmission or storage system can be guaranteed to be 100% secure. If you have reason to believe that your interaction with us is no longer **secure**, please immediately **notify** us in accordance with the "Contacting Us" section below.

Question 5: Upon my request, can my original dataset be deleted when my contract with the service provider terminates?

For this query, *request* is the most frequent keyword which was used 1128 times in 85% of the policies (Table 15). Other frequent keywords are *delete*, *remove*, *terminate*, *cancel*, *backup*, *erase*, and *erasure*.

Table 15. Results of keyword analysis for Question 5.

Keyword	Frequency	Number of Farm Data Agreements
request	1128	121
delete	489	94
obligations	283	85
choose	262	85
comply	241	104
remove	181	61
terminate	148	72
retention	111	59
cancel	57	17
backup	41	26
erase	26	38
erasure	25	19

Observations: Acceptance of requests for data deletion is recommended in most of the data privacy regulations (e.g., the right to erasure in the GDPR) and codes of conduct (e.g., the Australian Farm Data Code). Some ATPs (e.g., Sample text 1) allow requests for data deletion. However, words such as '*make our best effort*' and '*where permissible*' are used which allow them to decline or ignore such requests without a clear justification. Also, some agreements (Sample text 2 and Sample text 3) have mentioned that personal information may continue to exist in the backup even after the deletion of personal information. This can be due to the fact that deleting every single record in the backup can be challenging. However, there are many best practices and standards that can be implemented for tracking backup data in the cloud environment. ATPs should ensure that personal information is deleted as per farmers' and other legitimate users' requests.

Sample text 1 related to Q5: Where permissible, we will also **delete** your personal data upon **request**; please see the Contact Us information below to make such a **request**.

Sample text 2 related to Q5: If you **request** that your name be **removed** from our databases, it may not be possible to completely **delete** all your Personal Information due to technological and legal constraints.

Sample text 3 related to Q5: Personal information that is no longer required will be destroyed, erased or made anonymous, although copies of **deleted** information may continue to exist on backup media. When **destroying** personal information, we **delete** electronically stored personal information and share any tangible materials containing personal information. While we will endeavour to **destroy** all copies of personal information, you acknowledge that **deleted** information may continue to exist on back-up media but will not be used unless permitted by law.

Sample text 4 related to Q5: We **retain** information for as long as reasonably necessary to deliver our Services to you or to fulfill the purposes described in this Policy, or as required by law. To request the **deletion** of your personal information, please send us an email at. . .. We will make our best effort to **delete** all personal information about you. However, please note that certain information, including other information, may be **retained** in backup databases.

Question 6: Will the service provider obtain my consent before providing other companies with access to my data?

Our observation shows that *third party* and *access* are the most frequent keyphrases which are used in 87.94% and 86.52% of the policies, respectively (Table 16). The other frequent words are *purpose* which is used in 63.68% of the policies, *disclose*, *consent* is mentioned in around 77% of the policies, *disclosure* is used in 65.95% of the policies.

Table 16. Results of keyword analysis for Question 8.

Keyword	Frequency	Number of Farm Data Agreements
party	1294	124
access	1263	122
purpose	1106	118
third	1097	122
account	959	101
consent	754	111
disclose	562	110
notice	424	99

The following are samples of text which are related to the query. 'Consent' is mentioned in these texts.

Observation: Obtaining consent is an important data practice which is recommended and, in some cases, enforced by the data protection legislation (e.g., the GDPR) and codes of conduct (e.g., the Australian Farm Data code, the EU code of conduct on agricultural data sharing on contractual agreement). It is observed that consent is mentioned in most of the agreements. However, it is expected that the ATPs collect farm data through informed consent to ensure transparency. Some sample texts are included below. Sample text 1 mentions obtaining consent before transferring information to third parties. Sample text 2 mentions that the individuals can withdraw their consent to processing of data and the service provider collects information from third parties with the consent of the individual.

Sample text 1 related to Q8: We do not **disclose** any personal information to any third party without your informed **consent** unless we are required by law to **disclose** the information.

Sample text 2 related to Q8: Individuals can deny or withdraw their **consent** to AGI's collection, use and **disclosure** of their personal information at any time upon reasonable **notice**, subject to any legal or contractual requirements. However, if **consent** is denied or withdrawn, AGI may not be able to provide certain products or services.

Occasionally, "our company" may collect personal information from **third party** sources, but only with the knowledge and **consent** of the individual or where otherwise authorized by law.

Question 7: What categories of farm-related data do the product or service collect from my farm?

Our results suggest that information related to farms such as land, animal health, machinery, or other non-identifiable information is not frequently mentioned in data agreements (Table 17). For example, keywords such as *camera* and *sensors* are used in 15.60% and 25.53% of the agreements, respectively. Other relevant keywords are *machine* (used in 27.65% of the agreements), *land* (occurred in 43.26% of the agreements), *equipment* (occurred in 24.82% of the agreements), and *non-identifiable* (occurred in 7.09% of the agreements).

Table 17. Results of keyword analysis for Question 9.

Keyword	Frequency	Number of Farm Data Agreements
camera	168	22
financial	155	47
history	138	45
machine	123	39
asset	115	56
health	78	45
sensor	71	36
bank	58	44
animal	49	17
land	43	61
nonidentifiable	23	10
equipment	227	35

Observations: Some of the codes of conduct (e.g., the Australian Farm Data Code) recommend that the ATPs clearly explain what type of data is collected from the farms and farming operations to ensure transparency. However, only some of the agreements mention data collection from farms. The data agreements, in general, do not extensively mention the information regarding farm data.

Sample text 1 related to Q9:

Machine Data is data generated by, collected by, or stored in your **equipment** or any hardware or device interfacing with your equipment. **Machine** data includes the location of your **equipment**, the number of engine hours of your **equipment**, data regarding **equipment** operation (such as quantity of fuel used), and **equipment** diagnostic data.

Sample text 2 related to Q9: Livestock Information: We collect data on livestock you select for tagging and monitoring through the Services, including information on **health** and **animal history**, when you voluntarily install our DoggTag and DoggBone products in your herd and/or subscribe to the Services.

Sample text 3 related to Q9: The Services also may receive certain data from a drone that you connect to the mobile device that uses our mobile application, such as motion **sensor** data, visual and sonic **sensor** data, **GPS sensor** data, battery level and images captured by a **camera**. This Privacy Policy may refer to any such data collected from your device or drone as "device data".

Question 8: What categories of personal (sensitive) data does the product or service collect from me?

It is observed that the collection of personal information is discussed in the majority of the agreements. For example, *email* is used in 86.52% of the agreements, and *address* is used in 85.10% of the agreements. Location data is a very sensitive data point that is used in 70% of the data agreements. *Name* is used in 80% of the data agreements and *age* is used in 89% of the agreements. *Phone* is used in 56% of the data agreements (Table 18).

Observations: The codes of conduct (e.g., the Australian Farm Data Code) recommends being transparent about the collection of personal information. Transparency about collection of personal data is also encouraged by many privacy regulations. In our analysis, it is observed that the collection of personal information is discussed more extensively in the agreements as compared to farm data collection.

Table 18. Results of keyword analysis for Question 10.

Keyword	Frequency	Number of Farm Data Agreements
personal	4585	125
email	1120	122
address	977	120
identifiable	772	112
location	550	100
name	480	114
number	431	108
phone	251	80
equipment	227	35
financial	155	47

Sample text 1 related to Q10: The type of information that AGI may collect from individuals depends upon the Authorized Purpose for which it is being collected. This information may include, for example, **name**, email **address**, **address**, **telephone number**, fax **number**, information regarding your use of AGI's website, products or services, and other information individuals choose to provide.

Sample text 2 related to Q10: We may use your device's physical location to provide you with personalized **location-based** services and content.

5.4.3. Coverage Analysis of the Content Related to Questions

We performed a coverage analysis to examine how popular topics were in the FDLAs we collected. Figure 3 illustrates the average of related keywords per agreement. The least referred to topics were farm data collection and data deletion. Data collection, obtaining consent, and notifying policy changes were the most popular topics in the agreements. These results are dependent on the number of keywords for each topic/question and also depend on the outcome of stemming.

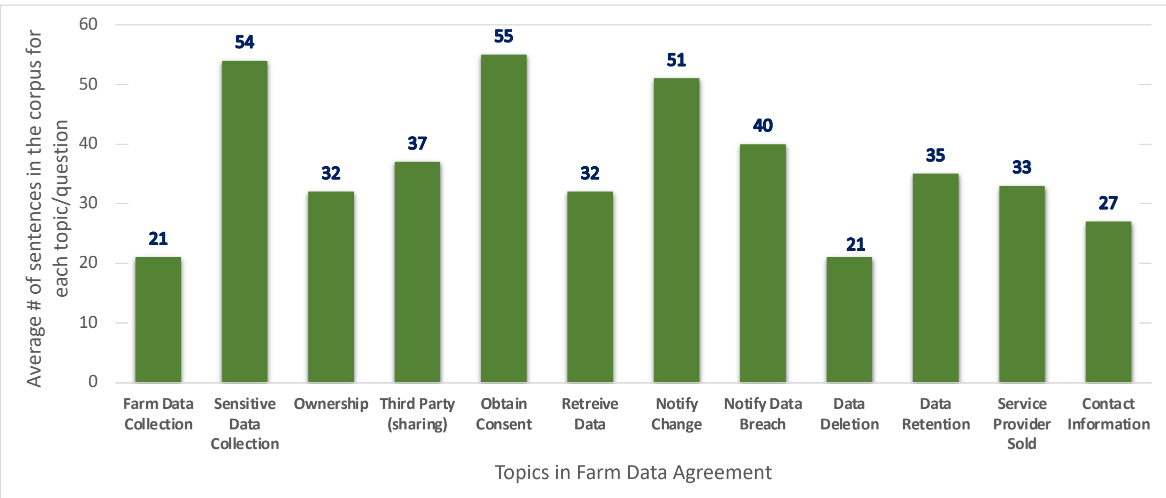

Figure 3. Topic coverage in 141 FDLAs—Average of related Keywords Per Agreement.

To further elaborate on these results, we checked the distribution of related keywords for each topic per agreement. Here are some of the interesting statistics:

- Farm Data Collection: In total, 16 agreements did not have any keywords related to farm data collection and 50% have less than 11 keywords in this category;

- Sensitive Data Collection: In total, 13 agreements did not have any keywords related to sensitive data collection and 50% have less than 51 keywords in this category;
- Ownership: In total, 14 agreements did not have any keywords related to ownership and 50% have less than 29 keywords in this category;
- Third-Party Sharing: In total, 14 agreements did not have any keywords related to third-party sharing and 50% have less than 31 keywords in this category;
- Obtaining Consent: In total, 15 agreements did not have any keywords related to obtaining consent and 50% have less than 44 keywords in this category;
- Retrieve Data: In total, 11 agreements did not have any keywords related to retrieving farm data and 50% have less than 26 keywords in this category;
- Notify Change: In total, 14 agreements did not have any keywords related to notifying change and 50% have less than 46 keywords in this category;
- Notify Data Breach: In total, 14 agreements did not have any keywords related to notifying a data breach and 50% have less than 36 keywords in this category;
- Data Deletion: In total, 15 agreements did not have any keywords related to data deletion and 50% have less than 15 keywords in this category;
- Data Retention: In total, 14 agreements did not have any keywords related to data retention and 50% have less than 28 keywords in this category;
- Contact Information: In total, 13 agreements did not have any keywords related to data deletion and 50% have less than 20 keywords in this category.

It is important to note that contact information has the smallest number of keywords and farm data collection has the largest number of relevant keyword.

6. Discussion and Recommendations

This paper proposes a novel framework for the evaluation of FDLA in terms of their content, structure, and shortcomings. The paper also uses this framework to evaluate certain features of a sample of agreements that have been collected online. To the best of our knowledge, this is the first study and the most comprehensive evaluation of FDLAs.

6.1. Contribution to Research and Findings

The results of the readability analysis illustrated that 95% of the FDLAs are difficult to read. It is also shown that 75% of the agreements need a college degree in order to understand the content. These findings are concerning and emphasize the need for standards and best practices to draft FDLAs that would make these legal agreements more readable and easier to understand. Additionally, the results of the text feature analysis suggest that the sentences are very long in the FDLAs. Long sentences can make the document complex and can be difficult to understand.

Furthermore, the collection of data is discussed in most of the agreements. However, they do not transparently explain what data are collected from the farmers and their farms. We have also observed that FDLAs mention some personal information such as email, location and name. But they do not mention farm data collection such as equipment information, or production information (livestock). Our findings suggest other important considerations for the content of FDLAs. There were very few agreements that had relevant content on other organizations (third parties) that were given access to farm and farmers' data. Very few of the agreements clearly explained for what purposes they were given access to data, and whether those organizations were in agreement with the terms and conditions of the ATPs' data agreements.

The GDPR and other privacy regulations and best practices recommend using shorter sentences instead of using overly complex sentences and avoiding passive voice and ambivalent terminology in the agreements. Moreover, transparency analysis shows that the FDLAs use ambiguous words quite frequently. 'May', 'such', and 'certain', for instance, were frequent terminologies in the FDLAs. This observation suggests that sentences in the agreements are written in an ambiguous manner which can make it difficult to understand the data license agreements [46].

We analyzed some of the important data practices using predefined questions. We also reviewed sample text extracted from the agreements. In our analysis, we found some gaps in the content of agreements that need to be addressed to improve the content. For example, only a few agreements mention that ATPs will notify the user about the changes in the policy in the form of a "notice". Some ATPs even expect the farmers to check the most updated agreement frequently to obtain the latest information about the changes in the agreements. Some of those ATPs even go to the extent of assuming that farmers agree to the changes because farmers continue to use the technologies. Such data practices violate recommendations by the GDPR, agriculture codes of conduct, and other best practices as they require organizations to follow procedures to notify and obtain explicit consent from the users. Our analysis also showed that the agreements refer to data deletion and data retention; however, some of the agreements use vague language such as 'where permissible'. Also, most ATPs do not refer to backups or anonymized records in their policy. A few of the agreements have generalized information related to data retention without giving appropriate details such as the reason or purpose for retaining information and the length of retention of data. Furthermore, we found that some of the ATPs mention data security practices such as encryption of data. The majority of the agreements do not address any proactive measures such as notifying the farmers and other stakeholders in the case of data and security breaches, which is very concerning.

6.2. Practical and Managerial Recommendations

There are many recommendations that can help in improving the FDLAs. For improving readability, longer sentences and longer words must be avoided, and documents must be simplified in terms of structure, by breaking the text into meaningful paragraphs and adding appropriate headings. Word choices should also be selected carefully to avoid complex legal and technology-related terminologies that are not familiar to the farmers and other users. Additionally, when creating the agreements, particular attention should be paid to the textual features to effectively communicate and ease the comprehension of the agreements. Furthermore, it is recommended by several codes of conduct such as the Australian Farm Data, Ag Data's Core Principles, and privacy regulations such as the GDPR that the agreements should be transparent and should be written in clear and plain language [31].

Some regulations recommend avoiding the use of ambivalent and ambiguous words and suggests using plain language [31]. Likewise, the Australian Farm Data Code recommends using plain language and being transparent about the data processing practices. This requirement should be reflected in the FDLA and they must be written using unambiguous words. The "transparency" principle also encourages that ATPs should be upfront and clear about their data practices such as retention, data access, data collection, data sharing, data safeguarding mechanisms, and many other practices. Additionally, the use of consistent and standardized terminologies can improve the transparency of data agreements.

To improve the content of FDLAs, it is recommended that the key concepts such as data ownership, data portability, personal information (including Personally Identifiable Information), and other important data practices such as data anonymization and de-identification must be defined clearly in the agreements. This will help in establishing a common understanding of terminologies and legal frameworks and their associated data practices used in the agreements. For example, data ownership is a complex legal terminology in the current technology ecosystem. It will be helpful if ATPs clearly define the concept of "ownership" in their legal and technology ecosystem and the rights and responsibilities of multiple owners or stewards. Furthermore, it is highly recommended that ATPs provide notice to the farmers and other users about changes in the agreements and data practices, in a transparent and easy-to-understand manner, so that they can make informed choices about their options.

ATPs should be transparent about data deletion and data retention policies and allow farmers more control over this option. Farmers must be able to delete some of their data, e.g., personal data and personally identifiable data, when terminating the contract. ATPs should also be cognizant of the fact that anonymization and de-identification do not fully protect farmers' privacy. Furthermore, the service providers must give the necessary details about retention of data, since it is recommended by several agriculture codes of conduct and privacy regulations.

Codes of conduct such as the Australian Farm Data Code, the EU Code of Conduct on agricultural data sharing by contractual agreement, Ag Data Core principles, and privacy regulations such as the GDPR have all strongly recommended data security and safeguards for protecting personal data and some farm information. The service providers must take appropriate measures to ensure security of data. Methods such as encryption, multi-factor authentication, and anonymization can better protect farm data. Additionally, if the farmers or other stakeholders are informed about breaches in a timely manner, they can take action to protect their data and potentially a network of connected devices. They can also seek clarification about the harm to their farm and assets. The service providers must take prudent steps in notifying the user in case of a data breach.

According to privacy legislation such as the GDPR, data minimization must be followed by the service providers. This data practice requires that the service providers only collect information that is necessary for processing purposes and must limit the collection of personal information. This should also be reflected in the FDLAs by mentioning the data attributes that are collected from farms. Furthermore, ATPs should be transparent about third-party data sharing. ATPs should clearly mention how their data agreements, or the third-party data agreement can be found and accessed and whether the third parties are in agreement with the original data agreements or have their own policy. Finally, FDLAs should be made available to all stakeholders in a user-friendly manner. A hardcopy of the data agreements should be provided to the farmers and other stakeholders. The agreements should also be accessible in digital format on a website. The layout of these documents should be user-friendly and should have appropriate headings and fonts.

7. Conclusions and Future Work

The data agreements play a vital role in describing data practices that ATPs use and implement for farm data processing. In this paper, we presented the analysis of 141 FDLAs to provide insight into common data management practices included in data agreements and most likely used by ATPs. Our findings also highlighted the common practices for drafting FDLAs. We performed text feature analysis, readability analysis, transparency analysis, and content analysis to evaluate the structure and loopholes of the agreements. The analysis showed that the sentences are long, difficult to read, and have ambiguous and passive voice sentences in the data agreements. We also analyzed the content of the agreements by examining the most frequent words and phrases and analyzing the content by using predefined questions which represented different data practices. We observed that the data agreements address some of the data privacy concepts in great detail. However, some other data practices such as notifying the farmers and other stakeholders about collected data attributes, data security breaches, and changes in the agreements have not been given enough attention. Furthermore, the purpose and period of data retention and data control choices for farmers are not discussed in the agreements to the extent that is expected.

FDLAs, if developed appropriately using the guidelines provided by codes of conduct and our recommendations in this paper, can help in building trust between farmers, ATPs, and other stakeholders. Trust instills confidence among the actors in agriculture and encourages data sharing. In return, data sharing enables innovation, transparent communication, and collaboration.

Several limitations may have impacted the findings of this work. This study used 141 agreements that we were able to find online. A larger corpus of data agreements can be collected and analyzed. In addition, these data agreements do not accurately reflect the organization's internal data and governance practices. We provided a comprehensive analysis of the content of agreements using NLP. NLP methods can show the existence or absence of topics. However, they cannot evaluate how transparently or completeness of a topic have been discussed. We have shown examples to strengthen our analysis. These agreements can be further reviewed and analyzed by subject matter experts such as lawyers. In the future, a more in-depth analysis of FDLAs can be performed by applying topic modeling approaches and machine learning algorithms to examine the semantics of data policies. Another possible direction is to compare data agreements in other sectors that are closely related to agriculture to understand their differences and examine ways to strengthen and improve agreement content. Examining ATPs' internal practices and understanding how and whether they are aligned with the recommended standards will be another interesting direction for this research.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/agriculture13112170/s1, list of bigrams and trigrams collected from the FDLA and keywords and content analysis related to other remaining questions in Table 3. Section S1: Results of frequent bigrams. Section S2: Results of frequent trigrams. Section S3: Keyword Analysis for Predefined Questions.

Author Contributions: Conceptualization, R.D.; methodology, J.K. and R.D.; validation, J.K. and R.D.; formal analysis, J.K. and R.D.; investigation, J.K. and R.D.; resources, R.D.; data curation, J.K.; writing—original draft preparation, J.K.; writing—review and editing, J.K. and R.D.; supervision, R.D.; funding acquisition, R.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to ongoing research related to this study.

Acknowledgments: This research was funded by a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant and Ontario Ministry of Agriculture Food and Rural Affairs, New Directions; funding awarded to Rozita Dara.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Quy, V.K.; Van Hau, N.; Van Anh, D.; Quy, N.M.; Ban, N.T.; Lanza, S.; Randazzo, G.; Muzirafuti, A. IoT-Enabled Smart Agriculture: Architecture, Applications, and Challenges. *Appl. Sci.* **2022**, *12*, 3396. [CrossRef]
2. Global Aginvesting. Available online: https://www.globalaginvesting.com/contributed-content-iot-agriculture/ (accessed on 12 July 2022).
3. Wolfert, S.; Ge, L.; Verdouw, C.; Bogaardt, M.-J. Big Data in Smart Farming—A review. *Agric. Syst.* **2017**, *153*, 69–80. [CrossRef]
4. Wysel, M.; Baker, D.; Billingsley, W. Data sharing platforms: How value is created from agricultural data. *Agric. Syst.* **2021**, *193*, 103241. [CrossRef]
5. van der Burg, S.; Wiseman, L.; Krkeljas, J. Trust in farm data sharing: Reflections on the EU code of conduct for agricultural data sharing. *Ethic. Inf. Technol.* **2020**, *23*, 185–198. [CrossRef]
6. Jakku, E.; Taylor, B.; Fleming, A.; Mason, C.; Fielke, S.; Sounness, C.; Thorburn, P. "If they don't tell us what they do with it, why would we trust them?" Trust, transparency and benefit-sharing in Smart Farming. *NJAS Wagening. J. Life Sci.* **2018**, *90–91*, 100285. [CrossRef]
7. Durrant, A.; Markovic, M.; Matthews, D.; May, D.; Enright, J.; Leontidis, G. The role of cross-silo federated learning in facilitating data sharing in the agri-food sector. *Comput. Electron. Agric.* **2022**, *193*, 106648. [CrossRef]
8. Top, J.; Janssen, S.; Boogaard, H.; Knapen, R.; Şimşek-Şenel, G. Cultivating FAIR principles for agri-food data. *Comput. Electron. Agric.* **2022**, *196*, 106909. [CrossRef]
9. Zhang, A.; Heath, R.; McRobert, K.; Llewellyn, R.; Sanderson, J.; Wiseman, L.; Rainbow, R. Who will benefit from big data? Farmers' perspective on willingness to share farm data. *J. Rural Stud.* **2021**, *88*, 346–353. [CrossRef]

10. American Farm Bureau Survey Shows Big Data Use Increasing—Global Ag Tech Initiative. Available online: https://www.globalagtechinitiative.com/digital-farming/data-management/american-farm-bureau-survey-shows-big-data-use-increasing/ (accessed on 12 July 2022).
11. Regan, A. Smart farming' in Ireland: A risk perception study with key governance actors. *NJAS-Wagening. J. Life Sci.* **2022**, *90*, 100292. [CrossRef]
12. van der Burg, S.; Bogaardt, M.-J.; Wolfert, S. Ethics of smart farming: Current questions and directions for responsible innovation towards the future. *NJAS Wagening. J. Life Sci.* **2019**, *90–91*, 100289. [CrossRef]
13. Hackfort, S. Patterns of Inequalities in Digital Agriculture: A Systematic Literature Review. *Sustainability* **2021**, *13*, 12345. [CrossRef]
14. Fleming, A.; Jakku, E.; Lim-Camacho, L.; Taylor, B.; Thorburn, P. Is big data for big farming or for everyone? Perceptions in the Australian grains industry. *Agron. Sustain. Dev.* **2018**, *38*, 24. [CrossRef]
15. Glavee-Geo, R.; Engelseth, P.; Buvik, A. Power Imbalance and the Dark Side of the Captive Agri-food Supplier–Buyer Relationship. *J. Bus. Ethics* **2021**, *178*, 609–628. [CrossRef] [PubMed]
16. EU Code of Conduct on Agricultural Data Sharing by Contractual Agreement. Available online: https://fefac.eu/wp-content/uploads/2020/07/eu_code_of_conduct_on_agricultural_data_sharing-1.pdf (accessed on 14 October 2022).
17. Fadul-Pacheco, L.; Wangen, S.R.; da Silva, T.E.; Cabrera, V.E. Addressing Data Bottlenecks in the Dairy Farm Industry. *Animals* **2022**, *12*, 721. [CrossRef]
18. Digital Technologies and the Big Data Revolution in the Canadian Agricultural Sector: Opportunities, Challenges, and Alternatives. 2021. Available online: http://www.usaskstudies.coop (accessed on 21 July 2022).
19. Cue, R.; Doornink, M.; George, R.; Griffiths, B.; Jorgensen, M.W.; Rogers, R.; Saha, A.; Taysom, K.; Cabrera, V.E.; Wangen, S.R.; et al. Data Governance in the Dairy Industry. *Animals* **2021**, *11*, 2981. [CrossRef]
20. Visualizing the Length of the Fine Print, for 14 Popular Apps—Visual Capitalist. Available online: https://www.visualcapitalist.com/terms-of-service-visualizing-the-length-of-internet-agreements/ (accessed on 20 July 2022).
21. Wiseman, L.; Sanderson, J.; Zhang, A.; Jakku, E. Farmers and their data: An examination of farmers' reluctance to share their data through the lens of the laws impacting smart farming. *NJAS Wagening. J. Life Sci.* **2019**, *90–91*, 100301. [CrossRef]
22. U./Fao; /Ifad. Synthesis of the Legal Aspects of Contract Farming Agreements Legal Guide on Contract Farming. Available online: http://www.unidroit.org/ (accessed on 20 April 2023).
23. Dutta, N.; Kaur, A. A socially responsible decision-making model for firms contracting with constrained farmers. *Int. Trans. Oper. Res.* **2022**, *30*, 2094–2121. [CrossRef]
24. Sgroi, F.; Sciancalepore, V.D. Dynamics of structural change in agriculture, transaction cost theory and market efficiency: The case of cultivation contracts between agricultural enterprises and the food industry. *J. Agric. Food Res.* **2022**, *10*, 100396. [CrossRef]
25. Kaur, J.; Dara, R.A.; Obimbo, C.; Song, F.; Menard, K. A comprehensive keyword analysis of online privacy policies. *Inf. Secur. J. A Glob. Perspect.* **2018**, *27*, 260–275. [CrossRef]
26. Wilson, S. The Creation and Analysis of a Website Privacy Policy Corpus. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016—Long Papers, Berlin, Germany, 7–12 August 2016; Volume 3, pp. 1330–1340. [CrossRef]
27. Bateni, N.; Kaur, J.; Dara, R.; Song, F. Content Analysis of Privacy Policies Before and After GDPR. In Proceedings of the 2022 19th Annual International Conference on Privacy, Security & Trust (PST), Fredericton, NB, Canada, 22–24 August 2022; pp. 1–9. [CrossRef]
28. Amos, R.; Acar, G.; Lucherini, E.; Kshirsagar, M.; Narayanan, A.; Mayer, J. Privacy Policies over Time: Curation and Analysis of a Million-Document Dataset. In Proceedings of the The Web Conference 2021—Proceedings of the World Wide Web Conference, WWW 2021, New York, NY, USA, 19–23 April 2021; Volume 12, pp. 2165–2176. [CrossRef]
29. General Data Protection Regulation (GDPR)—Official Legal Text. Available online: https://gdpr-info.eu/ (accessed on 12 July 2022).
30. Farm Data Code. Available online: https://nff.org.au/wp-content/uploads/2020/02/Farm_Data_Code_Edition_1_WEB_FINAL.pdf (accessed on 14 October 2022).
31. Art. 12 GDPR—Transparent Information, Communication and Modalities for the Exercise of the Rights of the Data Subject—General Data Protection Regulation (GDPR). Available online: https://gdpr-info.eu/art-12-gdpr/ (accessed on 20 April 2023).
32. Federal Trade Commission. FTC Issues Final Commission Report on Protecting Consumer Privacy. Available online: https://www.ftc.gov/news-events/news/press-releases/2012/03/ftc-issues-final-commission-report-protecting-consumer-privacy (accessed on 20 April 2023).
33. OECD Legal Instruments. Available online: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0188 (accessed on 20 April 2023).
34. Core Principles—Ag Data Transparent. Available online: https://www.agdatatransparent.com/principles (accessed on 3 April 2022).
35. Nadkarni, P.M.; Ohno-Machado, L.; Chapman, W.W. Natural language processing: An introduction. *J. Am. Med. Inform. Assoc.* **2011**, *18*, 544–551. [CrossRef]
36. Vijayarani, S.; Ilamathi, J.; Nithya, S. Preprocessing Techniques for Text Mining-An Overview. *Int. J. Comput. Sci. Commun. Netw.* **2015**, *5*, 7–16.

37. Flesch, R. A new readability yardstick. *J. Appl. Psychol.* **1948**, *32*, 221–233. [CrossRef] [PubMed]
38. How to Write. Available online: https://pages.stern.nyu.edu/~wstarbuc/Writing/Flesch.htm (accessed on 22 July 2022).
39. Kincaid, J.; Fishburne, R.; Rogers, R.; Chissom, B. Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel. Institute for Simulation and Training: Orlando, FL, USA, January 1975. Available online: https://stars.library.ucf.edu/istlibrary/56 (accessed on 22 July 2022).
40. Certified—Ag Data Transparent. Available online: https://www.agdatatransparent.com/certified2 (accessed on 3 April 2022).
41. Guntamukkala, N.; Grewal, G.; Dara, R. A Machine-Learning Based Approach for Measuring the Completeness of Online Privacy Policies Collaboration with The Digital Identification and Authentication Council of Canada View project A Machine-Learning Based Approach for Measuring the Completeness of Online Privacy Policies. In Proceedings of the 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), Miami, FL, USA, 9–11 December 2015. [CrossRef]
42. Translation and Comprehensibility—Google Books. Available online: https://books.google.ca/books?hl=en&lr=&id=5XerBwAAQBAJ&oi=fnd&pg=PA263&dq=lengthy+sentences+can+be+difficult+to+understand&ots=wo9gxBEpNb&sig=8BbvxUTC8bioLx_zMnKTmRDoeUw#v=onepage&q&f=false (accessed on 27 July 2022).
43. Top 10 Operational Responses to the GDPR—Part 6: Transparency and Privacy Notices. Available online: https://iapp.org/news/a/top-10-operational-responses-to-the-gdpr-part-6-transparency-and-privacy-notices/ (accessed on 27 July 2022).
44. Li, Y.; Stweart, W.; Zhu, J.; Ni, A. Online Privacy Policy of the Thirty Dow Jones Corporations: Compliance with FTC Fair Information Practice Principles and Readability Assessment. *Commun. Iima* **2014**, *12*, 5. [CrossRef]
45. Art. 5 GDPR—Principles Relating to Processing of Personal Data—General Data Protection Regulation (GDPR). Available online: https://gdpr-info.eu/art-5-gdpr/ (accessed on 27 July 2022).
46. Lebanoff, L.; Liu, F. Automatic Detection of Vague Words and Sentences in Privacy Policies. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018, Brussels, Belgium, 31 October–4 November 2018; pp. 3508–3517. [CrossRef]