A Survey of Intellectual Property Rights Protection in Big Data Applications

Rafik Hamza and Hilmil Pradana

1 Institute for International Strategy, Tokyo International University, Kawagoe 350-1197, Japan
2 National Institute of Information and Communications Technology (NICT), Big Data Integration Research Center, Tokyo 184-8795, Japan
* Correspondence: rhamza@tiu.ac.jp
§ This paper is an extended version of our paper published in Hamza, R.; Dao, M.-S.; Ito, S.; Koji, Z. Towards Intellectual Property Rights Protection in Big Data. In Proceedings of the 3rd ACM Workshop on Intelligent Cross-Data Analysis and Retrieval, Newark, NJ, USA, 27–30 June 2022.

Abstract: Big Data applications have the potential to transform any digital business platform by enabling the analysis of vast amounts of data. However, the biggest problem with Big Data is breaking down the intellectual property barriers to using that data, especially for cross-database applications. It is a challenge to achieve this trade-off and overcome the difficulties of Big Data, even though intellectual property restrictions have been developed to limit misuse and regulate access to Big Data. This study examines the scope of intellectual property rights in Big Data applications with a security framework for protecting intellectual property rights, watermarking and fingerprinting algorithms. The emergence of Big Data necessitates the development of new conceptual frameworks, security standards, and laws. This study addresses the significant copyright difficulties on cross-database platforms and the paradigm shift from ownership to control of access to and use of Big Data, especially on such platforms. We provide a comprehensive overview of copyright applications for multimedia data and a summary of the main trends in the discussion of intellectual property protection, highlighting crucial issues and existing obstacles and identifying the three major findings for investigating the relationship between them.

Keywords: Big Data; applications; copyright; intellectual property; security

1. Introduction

Japan’s ‘Society 5.0’ initiative envisions a people-centric society combining economic progress through the use of artificial intelligence and Big Data capabilities [1]. Under Society 5.0, the Japanese government and industry IT are collaborating to effectively integrate technological developments, particularly in AI and Big Data, through cloud platforms [2,3]. To develop these technologies, the Japanese government provides various ways to improve and integrate them from different types of data such as images, videos, historical data, and signals. These data types are commonly used in various application technologies to solve their problems. The initiative from Japanese government was launched primarily to align economic progress with solving social problems. By following the age of 5G and AI, Society 5.0 is a potential technological framework to revolutionize the world in terms of information technology, especially in the areas of fin-tech (financial technology), healthcare, mobility, and other infrastructures. The future goal for Japan is to build the “Super Smart Society” that will improve people’s lives more sustainably [4].

The integration of cutting-edge technology into all facets of human life is to create datasets impacting people and institutions in many ways. For the benefit of data, many researchers can create their models to improve their technologies and achieve significant impact in all application aspects. All these data are being brought together to create Big Data that can be used in a variety of industries, such as medicine, banking, and Internet
advertising. The ability to take in unusually large amounts of information, effectively analyze and understand it, and then arrive at insights and interpretations is referred to in this context as “Big Data Analytics”.

Data have a value or an expected value and are documented or should be documented. This is referred to as “information value” [5]. This area is about techniques such as collecting data needed for a “service” and sending the information to the appropriate party. It is not practical to require uniform security methods for all information assets, as it is likely that IoT devices and the data processed will differ depending on the service offered. Data security and privacy are key concerns for both the government and corporations when processing massive data. The majority of IT companies regularly gather, transport, store, and analyze large datasets, which raises substantial privacy problems [6, 7]. The implementation of the “Society 5.0” initiative is complicated by these issues. A great deal of attention has recently been paid to data protection in order to improve privacy in use, storage, and transmission [8–10]. Thanks to cryptographic security approaches, such as homomorphic encryption and blockchain, datasets of any size can currently be kept secure and effective as they travel across the network and are stored in the data store [11–13].

In this paper, we examine intellectual property rights protection in Big Data applications [11]. The main contributions are threefold:

• This paper will take a closer look at the unclear boundaries of intellectual property rights in Big Data applications and present different viewpoints on copyright in cross-data platforms.
• This paper addresses real-world case studies of the underlying technology of cross-data analytics with a security policy framework to protect intellectual property rights.
• This paper highlights the main technical solutions for intellectual property protection, including the latest copyright algorithms for multimedia data.
• This paper discusses some important aspects of copyright protection and identifies the main problems and difficulties.

The paper is organized as follows: Section 2 defines, organizes, and interprets related intellectual property rights protection works. Non-technical solutions for protecting intellectual property rights are discussed in more detail in Section 3. An overview of technical solutions is discussed in Section 4. We then discuss current copyright applications for multimedia data in Section 5. In the last section, we conclude this paper.

2. Related Work

In contrast to today’s digital rights management technology, which has evolved into a systems engineering approach that covers the entire lifecycle of digital works, traditional data rights management focused exclusively on encryption and authorization. An intellectual property rights protection strategy based on a region of interest with mathematical polynomial transformation has been described by Murali et al. [14]. In their work, they investigated their proposed framework and discovered that it can improve the resistance to image manipulation and other types of attacks and replace the main method of sharing secret data of the presented property model.

Devi et al. [15] presented a steganography method based on visual cryptography, shuffled singular value decomposition, and applied it to copyright protection of digital images. The method was shown to outperform the visual cryptography algorithm in its ability to independently authenticate the copyright of digital images and resist various types of image processing attacks. Liang et al. [9] developed a security system with verification methods that provide effective data protection using the decentralized mesh node approach and multimedia data security technology. However, few studies have addressed the copyright protection of information assets, with most current research focusing on the copyright protection of multimedia datasets or a specific type of data.

We address the problems and difficulties of securing intellectual property rights in information goods to fill the gaps mentioned above. A review of the literature revealed various definitions of intellectual property rights and the categories under which they fall.
The main categories of intellectual property rights are shown in Figure 1: database rights, patents, designs, performers’ rights, copyrights, and trademarks. Table 1 provides some examples of information assets that need to be protected. A security asset in this context is described as “a valuable resource that must be protected” [16].

Figure 1. Intellectual property rights, adopted from [11].

Table 1. Examples of protected information assets, adopted from [11].

| Information Assets            | Description                                                                 |
|------------------------------|-----------------------------------------------------------------------------|
| Contents                     | Audio, image and visual inputs information (copyright, relevant information, and confidential information when using economic material), consumer history. |
| User information             | Private information about the customer (identification, zip code, contact details, date and place of birth, account number, etc.), digital information about the customer, previous transactions, etc. |
| Equipment information        | Communication device information (manufacturer, identification, unique driver’s license, etc.), device account credentials, and so on. Operating system monitoring (operating statistics, connection usage status, etc.) specific to each system. |
| Software settings            | Configuration information (procedure setting, connection setting, authorization setting, version) and other information specific to each program; computer operating system, middleware, etc. |
| Design data and internal logic | In the planning/design processes, design information of the requirements/design documentation is developed. |

When applied effectively, intellectual property rights can maximize the benefits and value of creativity while supporting the development, protection, and commercialization of breakthrough technologies, even if they appear to offer only a modest degree of certainty. For this reason, we chose to focus our study on copyright protection and patent protection as the two most important types of intellectual property protection.

2.1. Copyright

When a work is created, copyright immediately arises. This has numerous benefits for the people who contribute or create it, but it also raises problems for those who control the data and want to see the public’s interest in their creations, which are also available to the public. Until recently, creators had the option of allowing artists to publish their work. Customers can freely and securely use the protected content in this domain. Intellectual property owners always have the option to defend their rights if they are accused by other users of violating their rights [17].
As with all new goods, copyright claims should arise naturally once the information assets are developed [18]. In the case of software protection, it is more likely to be supported by legislation because the copyright mode of protection takes effect at any time, and the purpose and role of legal protection is greater [19]. Cai et al. [20] offered a deep learning strategy using blockchain as an example of the goal of copyright protection for digital music. The authors claim that their copyright protection system for digital music processes queries with an error rate of 0% in the presence of multiple concurrent users, ensuring a high level of intellectual property rights protection.

Computer software is one of the most significant information assets. It must be protected by copyright because it has been creatively developed, mostly through the use of terms of use agreements [21]. Additionally, computer scientists define computer programs using a variety of digital codes and programming languages before simulating them on computers for consumers to download and use [18]. In order to protect software, copyright measures must be implemented in the program code.

Software, a type of copyrighted material, is considered high technology, so laws protecting intellectual property rights in software are often inadequate. Most countries have copyright laws that protect software intellectual property, and patent protection can also be sought for software design ideas that are closely related to hardware. The security of computer software and systems may be largely covered by existing software protection methods, but gaps in protection do exist.

2.2. Patent

In order to promote innovation and ensure the protection of R & D rights, patent rights should be preserved. The goal of developing information goods is to perform information management functions and replace traditional labor with information technology. In general, copyright law deals only with the recognized categories of information goods and is not able to adequately protect the way these goods are used or processed. Therefore, research and development must use patent-protected techniques to protect its operating procedures, development processes, and formulated thoughts. According to Eduardo et al. [22], patent rights are believed to promote both innovation and the diffusion of knowledge.

Algorithms link various data structures to successfully achieve the desired result using abstract control approaches, which should fully meet intellectual property protection requirements [18,21,23,24]. In this section, we would like to describe a security policy framework for defending intellectual property rights in Big Data. According to the authors, the three most important intellectual property rights in the context of Big Data are copyright, database right, and confidentiality. As the name implies, copyright is a legal remedy that prohibits unauthorized copying. However, copyright can be particularly helpful when there are widely used message formats, interfaces, protocols, and other standards that dictate that data must be in a certain format, as is the case in the Big Data world.

3. Non-Technical Solutions for Protecting Intellectual Property Rights

In this section, we address the dimensions of a Big Data value ecosystem, focusing primarily on the areas of law and compliance. The primary concern is to provide a non-technical framework for dealing with Big Data intellectual property rights that links data with various information assets. We believe that one of the most important issues to be resolved in any Big Data system is copyright. We can clearly state that despite legal agreements and the use of security policies as administrative methods to protect copyright, intellectual property rights still present some difficult problems and need to be further protected [25–28]. Confidentiality obligations associated with the substance of data that is not in the public domain might occasionally constitute the most valuable intellectual property right, since copyright and database regulation protect only the expression and form of information, not its content. A non-technical response could be critical in this case. In contrast, there are contractual rights with respect to data. People are bound by contract
law to strong, enforceable obligations and strong, enforceable rights. The financial market data industry, a USD one billion industry, has evolved around an ecosystem that licenses and controls data use, with virtually all of the risk associated with the data governed by contract. From our observations of industry contracts, we can conclude that data contracts are a reliable and theoretically sound method of defending intellectual property rights. However, these contracts have limited utility, especially when they do not bind a user who is not a party to the contract. In a contractual shell, contractual intellectual property can impose obligations that are similar to rights. The drafter of the contract must therefore evaluate the two components—the contractual intellectual property and the actual intellectual property regime—indepedently. These policy provisions will eventually conflict with each other. Researchers have addressed these concerns and successfully resolved several difficulties [26,29]. For example, Hoeren et al. [30] deals with the question “what happens if the new property rights in data conflict with data protection laws”?

Data protection, which grants rights and establishes rules for the processing of personal data, is the most important aspect of data regulation. EU competition authorities have become much more interested in corporate practices, licensing, and contracting for data in a number of industries over the past five years, particularly industrial market data [31–33].

Many industries are also tightening their privacy policies. These policies governing client confidentiality and privilege have been a cornerstone of the legal profession. The computerization of data, on the other hand, has radically changed the picture. For example, healthcare companies should aggregate only anonymized clinical outcome data from patients [34] and data collection companies with data localization laws, i.e., cross-border transfers from non-EU countries [35].

3.1. Legal Compliance

All computer software is original, distinctive, and self-contained, having been created by researchers. Unlike other legal tasks, a software program conveys the researcher’s design using a variety of data and coding languages [18]. The researcher’s design, research, and language skills are critical to software development. Before it can be officially launched, it must go through countless iterations and rounds of testing and debugging. The computer researcher identifies intellectual activities in social contexts, develops actionable research results, and then communicates those results in computer language. The effective use and understanding of Big Data as a business asset holds tremendous opportunities for the global economy and society. As shown in Figure 2, created by Cavanillas et al. [36], the barriers to the development of a Big Data ecosystem in Europe have been categorised into a number of important factors. To support the development of a Big Data ecosystem, the world must overcome these various challenges [36]. Of course, the results of exploration are logical, and eventually a number of mature operating software products did emerge. Therefore, the intellectual property rights of computer software should be put on par with the intellectual property rights of other materials, and the legal protection system required for any artificial Big Data operation should be further built and clarified.

Digital piracy is a major threat to the development of IT businesses and to the expansion of the digital media industry [37]. Software piracy, the unauthorized copying and use of software in a manner that violates applicable terms of use, threatens the legitimate rights and interests of researchers. Pirated software, unlike real software, has limited technical substance; therefore, the cost of citation and further development is comparatively low. Usually, the development of new software involves immense and time-consuming effort. If it is pirated, the innovative power of the research and development work is inevitably destroyed.
Figure 2. The dimensions of a Big Data value ecosystem (Adapted from (Cavanillas et al. [36])).

3.2. The Intellectual Property Rights of the Various Information Assets

There are different considerations in intellectual property rights based on different information values. Therefore, we look at the practices we use in the context of Big Data, including data rights. Table 2 lists common practices in the context of Big Data to ensure legal rights to databases.

Table 2. Data legal rights in the context of Big Data practices.

| Management       | Security, practices, policy, Process standards: example information security management ISO 27001 |
|------------------|----------------------------------------------------------------------------------------------------|
| Regulation       | Non-industry relevant: data protection, competition rules Industry-relevant: commercial, professional services, etc. |
| Contracting for datasets | Contract is priority Effective but restricted (contractual parties only) |
| Intellectual property rights in Big Data | Copyright, database right, confidentiality, patents, trademarks Weak but extensive, intellectual property rights in data uncertain |

First, private information goods: copyright also protects the materials accompanying proprietary software. The accompanying documents of proprietary software are considered unpublished works under copyright law because the source code of private software is not publicly available [18]. The author of the disputed document must have his own privilege, but he cannot acquire ownership because it has not yet been released for sale. Substantial similarity and interaction are two criteria that can be used to judge whether software copyright infringement has occurred in this situation.

Second, public information assets: “Substantial similarity” refers to the availability on the market of products substantially similar to the documentation describing the protected software. Substantial similarity does not exist if the required documentation for the program is not publicly available, and other organizations or individuals are free to point out the similarities of the application and file separate copyright applications. This is because open software and open source code not only reduce the cost of developing new software and promote technological progress in the development of software products, but also meet the document reading and learning needs of most industries, businesses, groups, and even individuals in society. Authorities should create a protection system that takes
into account the method, appropriate fines, and balance of benefits to successfully protect software copyrights from infringement.

Third, there is patent protection: In establishing a framework for software patent protection, the two components of private software and public software should be considered.

National governments determine the validity and legality of the protection of software patents and their use from a legal position [38]. Under current patent law, an innovation patent must be practical, creative, and unique. The term “novelty” refers to the fact that the software is not prior art and that there are no competing applications on the market [39]. However, the threats to patent regulation are growing, especially with the advent of artificial intelligence (AI) [39]. We would like to briefly explain and state the consequences of intellectual property and patent theft. Currently, central government agencies and offices control intellectual property. When a rights holder wishes to exercise his rights, it is these agencies that uphold his claims.

3.3. Limitations

The government or executive agency in the country where a rights holder wishes to exercise his rights enforces intellectual property regulations. Consequently, the physical boundaries of the system are the most important. The ease with which these rights are violated is increasing daily. One of the rights most affected by the development of the Internet and artificial intelligence is intellectual property [39]. However, certain more current issues, such as the nexus of privacy issues and copyright issues (sometimes known as “multi-party copyright disputes”), are rapidly evolving with cross-data. Although there are numerous non-technical possibilities, databases are protected in much the same way as strongly other forms of intellectual property are protected. Cross-data and the problem of overlapping legal and security terms have expanded the problem of intellectual property protection today. As shown in Figure 3, we therefore intend to provide a clear legal solution for managing copyrights with cross-information assets via a user agreement.

![Figure 3](image.png)

**Figure 3.** Manage copyrights with cross-assets using terms of use agreements.

It is a challenge for an integration designer to ensure that all security standards are met because today’s information resources, and especially software, are not capable of enforcing or managing security policies and processes. Due to a mistake by the designer or improperly instantiated software, third parties can gain entry to security resource control and observation, leading to confidentiality and integrity issues [16]. Moreover, it is difficult to observe the overlap of intellectual property rights raised during the process of negotiating legal agreements (see Figure 3), where various security policy conflicts and intellectual property rights issues may arise repeatedly.

In our opinion, the system of legal agreements is not adequate for managing intellectual property. The need for a comprehensive technical solution is compelling. Nowadays,
business structures are becoming more and more digital, which means that there are fewer and fewer geographical and physical constraints associated with them. As a result, blockchain is seen as a viable alternative for a “legal agreement system” [40–42].

4. Technical Solutions for Protecting Intellectual Property Rights

In today’s world, where cybercrime is on the rise and unauthorized users can easily access priceless cultural media, preservation of multimedia is one of the most important issues [43]. There are many researchers who apply watermarking algorithms to protect their data from theft and to use or modify it without the owner’s permission [44]. In this section, we explore technical solutions to intellectual property rights protection in two different applications. First, we study fingerprinting algorithms to protect the privacy and copyright of digital products from unauthorized distribution and to prevent attacks. Then, we explain watermarked and non-watermarked copyright protection algorithms to solve the application of copyright to multimedia data.

4.1. Fingerprinting Algorithms

The use of digital fingerprinting technology has proven to be a reliable way to protect the privacy and copyright of digital products from unauthorized distribution and to prevent adversarial attacks [45–47]. It is also used to track the user of multimedia content delivered via the cloud and its unauthorized distribution. Any design of algorithm fingerprinting used to generate these fingerprinting codes should consider the following properties: robustness, discriminability, compactness, and unpredictability [45]. For fingerprinting algorithms to work, an ensemble of feature vectors must be extracted from the source content, and these feature vectors must then be compared to a database of known features vectors associated with copyrighted content. A source pattern and a library pattern are considered identical if there are enough matches between them. Most audio, image, and video fingerprinting techniques either extract manually created features or train a neural network to extract fingerprint features [48].

A digital signature is also defined as a digital fingerprint [49]. Digital signatures are used to authenticate that a particular entity has submitted or approved a particular piece of data. Digital fingerprints are used to verify the accuracy of the data provided. However, a different signature is applied to each copy of the document. It provides the maximum level of protection that a digital signature can provide and can be used to track down intruders.

Monga and Evans [50] developed a fingerprinting method that uses feature points for photo identification and authenticity. The authors limited the removal or addition of relatively large elements in their performance and robustness studies. Their decision process is based on the idea that the altered image regions represent a mismatch of several derived feature points. In studies such as (Li et al. [46]), robust fingerprints were created by training networks on typical distortions such as adding an edge noise or rotating the video. Such networks are resistant to predefined distortions, but they are no longer robust [48].

Due to the growing number of videos posted on social media and video platforms, localization of video clips is crucial for practical applications such as detecting copyright infringement in videos [51]. Lahouari [45] proposed a reliable and perception-based fingerprinting method that can be used for both authentication and identification of video data. The resistance of the proposed fingerprint codes to content modification and geometric attacks makes them useful for content identification, but their sensitivity to malicious attacks makes them suitable for forgery detection and verification. A novel sequence normalization idea based on the circular shift properties of the discrete cosine and sine transforms enables this dual use (DCT and DST).

Jialuo et al. [47] presented a unique testing method that statistically compares two deep learning models—a victim model and a suspect model—to protect copyright. To assess whether a suspicious model is a replica of the victim model, a wide range of test metrics and effective test case generation algorithms are used to build a chain of evidence. The authors expect this to have several advantages, including that their proposed work
is noninvasive because it works directly with the model and does not interfere with the training process. Second, quick scans of the two models and a limited selection of test cases make the procedure efficient. Finally, deep learning model copyright protection is highly resistant to model extraction and adaptive attacks, i.e., it can easily incorporate additional test metrics or test case generation methods to achieve a more secure and robust judgment [47]. Through extensive testing on image classification and speech recognition datasets with a range of model architectures, the authors demonstrate the effectiveness of deep learning models for copyright protection in three common copyright infringement cases, including model fine-tuning, pruning, and extraction [47].

To protect digital properties in relational databases for digital privacy management, Fatima et al. [49] proposed a fingerprinting approach for the cloud environment. The new fingerprinting method used in the proposed solution makes it reliable and effective. It can solve problems, such as securely embedding data in the cloud, which is necessary for relational database security. Fatima et al. [49] claimed that their approach can prevent unauthorized distribution of intellectual content and copyright infringement.

4.2. Watermarking Algorithms

The process of embedding data into a multimedia element, such as a picture, audio, or video file, is known as watermarking. For security reasons, this encoded data may later be retrieved from or discovered in the multimedia. The embedding algorithm, extraction algorithm, and detection method make up a watermarking algorithm.

Based on the scores of watermarking and non-watermarking algorithms, there are several approaches, such as precision, recognition, peak signal-to-noise ratio (PSNR), structural similarity index metric (SSIM), normalized cross-correlation (NC), correlation coefficient (CC), and bit error rate (BER). PSNR, SSIM, and NC are used to estimate the differences between the original image and the result after noise reduction or deformable image. One of the applications of PSNR, SSIM, and NC is to calculate the similarity measure between the original image and the watermarked image to understand the robustness of the watermarking method to the invisibility of the watermark, and the unit is dB. To evaluate the robustness of watermarking in audio applications, SNR [52] and ODG [53] are used. SNR is a standard evaluation by measuring perceived transparency based on a statistical standard introduced by IFPI to measure the similarity rate between main and watermarked signals. ODG is a standard parameter measured by the EAQUAL [53] open source software.

Watermarks can be incorporated into a transform domain or a pixel domain. Embedded watermarks in multimedia applications should be undetectable, reliable, and big.

Most recent works [54–59] applied copyright of multimedia data to the problem of color image watermarking. Table 3 summarizes the important overview of copyright applications for multimedia data based on state-of-the-art techniques. Ref. [54] proposed a color image watermarking algorithm that has high performance and can be applied in various watermarking schemes based on discrete cosine transform (DCT) and discrete Hartley transform (DHT). Unfortunately, the effectiveness of this method is low when the loss of pixel values occurs after a zoom-out attack. Schur decomposition was first proposed by [55] where it was applied to color image watermarks. The idea of this method is to develop an algorithm for color image watermarking that has high robustness, large watermarking capacity, and high security. However, the effectiveness of this method is low when the image is attacked with salt-and-pepper noise. The multi-watermarking technology for color images was presented by [56]. His method finds the optimal embedding region solutions with several different embedding methods when applied to a color multi-watermark using adaptive multiple embedding factors (AMEF), particle swarm optimization and gray wolf optimizer (PSO-GWO), discrete wavelet transform (DWT), and singular value decomposition (SVD). The weakness of this method is that the total size of all watermarks should be smaller than the size of the selected regions in the image, making the method unreliable for larger watermarks. One of the recent works, ref. [57] focused on simple geometric attacks, such as cropping, shifting, rotating and distorting, which give good
results compared to the state-of-the-art methods using fusion domain color watermarking based on Haar transformation and image correction. However, the size of the input image and the watermark image must still be the same. To integrate the color domain into the RGB channel, the quaternion QR decomposition (QQRD) proposed by [58] is used to find an effective way to apply the watermark. However, the resilience of the proposed method is limited by Gaussian noise and scaling correction up to 50%.

Table 3. An overview of watermarking algorithms for copyright applications on multimedia data based on state-of-the-art schemes.

| Ref ID | Applications | Goals | Approaches | Evaluation Metric | Limitations |
|--------|--------------|-------|------------|-------------------|-------------|
| [54]   | Images       | To design a color image watermarking algorithm which has high performance and can be applied into different watermarking schemes. | Discrete cosine transform (DCT) and discrete Hartley transform (DHT) | PSNR, SSIM, and NC | The effectiveness of proposed method is low when the loss of pixel value after the zoom-out attack. |
| [55]   | Images       | To design a color image watermarking algorithm which has strong robustness, large watermark capacity, and high security. | Schur decomposition | NC | The proposed system has low accuracy on salt-and-pepper noise attack. |
| [56]   | Images       | To find the optimal embedding region solutions with multiple different embedding in application to a color multi-watermarking. | Adaptive multiple embedding factors (AMEF), particle swarm optimization and gray wolf optimizer (PSO-GWO), discrete wavelet transform (DWT), singular value decomposition (SVD) | PSNR and CC | The total of the sizes of all the watermarks should be lower than the size of the selected regions in the image. |
| [57]   | Images       | To focus on simple geometric attacks: cropping, translation, rotation and distortion. | A fusion-domain color watermarking based on Haar transform and image correction | PSNR, SSIM, and NC | the size of both input image and watermark image still needs to be the same. |
| [58]   | Images       | To integrate RGB channels to obtain the effective way for watermark extraction for increasing computing performance. | Quaternion QR decomposition (QQRD) | PSNR, BER, and NC | The resistance of proposed method is limited up to 50% by Gaussian noise and scaling correction. |
| [59]   | Audios       | To create a compromise method by following robustness, transparency, and capacity. | Fuzzy inference system, singular value decomposition (SVD), and discrete cosine transform (DCT). | SNR and ODG | Proposed scheme is weak against Addbrumm_2100 attack type. |

In the application of audio watermarking, ref. [59] used the fuzzy inference system, singular value decomposition (SVD), and discrete cosine transform (DCT) to create a compromise method for robustness, transparency, and capacity. The results show that the proposed method has a great advantage over other state-of-the-art methods.

4.3. Non-Watermarking Algorithms

Image correction and eigenvalue decomposition, reported by [60], uses color image watermarking to design a color image watermarking algorithm that has strong resilience, large capacity, and high security and can be applied to various geometric attacks. However, the computational time is high because the proposed architecture is very complex. On the other hand, ref. [61] uses precision and recall to evaluate the performance of its architecture for solving code plagiarism copyright protection. Precision is computed by dividing correctly classified positives by classified positives, while recall is computed by dividing correctly classified positives by actual positives.

Some recent works [61,62] use code plagiarism detection and video and audio applications to solve various copyright applications. Ref. [61] used full nodes and lightweight nodes to solve the problem of copyright confirmation and copyright protection in code plagiarism applications. Video watermarking has a similar path to image watermarking applications. Ref. [62] proposed a 2D DFT (two-dimensional discrete Fourier transform) to
create a simplification algorithm by direct extraction without synchronization for simple geometric attacks.

On the other hand, blockchain technology has been applied to exclude the attack from DDOS and a key generation center. Some applications on blockchain, such as images, music and videos, are presented on Table 4. Ref. [63] applied images, while [64,65] worked on music and videos. To create the features of de-trusted third parties by combining the fairness and process automation of smart contracts to make up for the shortcomings of the zero-watermarking algorithm, ref. [63] used blockchain and zero-watermark to solve the problem of data loss but relied more on the trusted third party than a traditional digital watermark, which makes its prospects limited. Ref. [64] addressed the problem of illegal distribution of copyright-protected music files without the consent of the owners, which has negative consequences in the music industry, while [65] addressed poor robustness, weak imperceptibility, and difficulty in traceability of the current protection schemes for video copyright.

Table 4. An overview of non-watermarking and hybrid algorithms for copyright applications on multimedia data.

| Ref ID | Applications | Goals | Approaches | Evaluation Metric | Limitations |
|--------|--------------|-------|------------|-------------------|-------------|
| [61]   | Multimedia   | To solve the problem of copyright confirmation and protection. | Full nodes and lightweight nodes | Precision and recall | Optimizing the process during extracting eigenvalues to decrease computation time. |
| [62]   | Videos       | To create a simplification algorithm by direct extraction without performing synchronization for simple geometric attacks. | 2-D DFT (two-dimensional discrete Fourier transform) | PSNR and SSIM | The effectiveness of proposed method is low against frame dropping attacks. |
| [63]   | Images       | To create the features of de-trusted third parties with combining the fairness and process automation of smart contracts to make up for the shortcomings of the zero-watermarking algorithm. | Blockchain and Zero-Watermark | NC | Due to cost constraints, it is almost impossible to make the absolute credibility of the third party in the digital watermarking technology. |
| [64]   | Musics       | To address the problem of illegal distribution of copyright-protected music files without the consent of the owners, which has negative consequences in the music industry. | A public-permission-less blockchain | File size vs uploading cost | Smart contracts cannot be able to pull data from off-chain resources; instead, that data should be “pushed” to the smart contract. |
| [65]   | Videos       | To address poor robustness, weak imperceptibility and difficulty in traceability of the current protection schemes for video copyright | blockchain with two layers: “on-chain” and “off-chain” | NC and Precision ratio | The proposed system has low accuracy on noise attack. |

5. Discussion

To promote accountability and responsible data sharing for all stakeholders, a different technical framework based on fiduciary legal concepts is urgently needed. Our investigation also found that data ownership and protection techniques are limited and may not solve the problems with cross-data platforms and Big Data applications. As stated in this paper, we do not believe that essential shared data should be protected as individual intellectual property because there are no specific rights for cross-database platforms. Data sharing becomes much more difficult under this type of ownership, especially when using cross-data services [11].

Small R&D companies may not benefit sufficiently from existing patents and face various obstacles to using or developing innovative technologies because of the issue of intellectual property protection for inventions. In this context, the problem that it is difficult for a R&D company to retain ownership of innovation was raised. The reason is that other IT companies that have not invested as much in the development could duplicate the idea and offer the new product at a lower price, putting the inventor out of business.
The problem of appropriability can be solved with intellectual property rights that grant companies a reasonable degree of exclusivity in the use of their innovation(s) [23]. This provides both an incentive to innovate and an opportunity to develop new products and processes. Ownership of intellectual property rights also gives companies the ability to negotiate licensing agreements with other companies to bring new products to market [66,67]. The ability to raise money from investors, especially business angels and venture capitalists, can occasionally be critical. The major categories of intellectual property rights are (1) patents and models, (2) trademarks, (3) databases, (4) valuable trade secrets or nonpublic information, and (5) copyrights and related rights [11].

Consequently, the basic idea that justifies the protection of intellectual property can promote innovation [68]. It is clear that intellectual property and innovation are closely related. As information assets have become the foundation of the knowledge economy, intellectual property has become increasingly important to companies worldwide [69]. Intellectual property rights have become a critical tool for gaining control over a company’s information assets, leveraging creative R&D results, facilitating technology licensing, and promoting improved or new products and services based on innovation and creativity in cross-platform services.

We have entered a new era in the world of intellectual property, where a whole new form of constantly emerging partnerships and products allows people around the world to collaborate in an ecosystem. In general, we can see that there are two main solutions to intellectual property rights. The first is with non-technical solutions, such as legal agreements (terms of use, etc.), and the second is with technical solutions, such as fingerprinting algorithms and cryptographic techniques. Each solution is tailored to a specific intellectual property problem, and we can clearly see that there is no comprehensive solution. A new solution based on both solutions is needed.

Blockchain can be revolutionary in terms of proof of ownership and various transactions, as it does not require an intermediary and can store immutable data, including records and proof of ownership. Centralized registration of intellectual property can be completely eliminated thanks to blockchain technology. Blockchain-based copyright records can be incredibly useful for protecting artwork such as images or music. By using this technology, owners can instantly track their works and maximize their revenue. It is possible for certain information and data containing trade secrets to be accessible, monitorable, and trackable. The question is: What makes blockchain technology unique and suitable for intellectual property protection? Instead of the traditional client–server approach, blockchain uses a “peer-to-peer” model where there is no central authority. As a result, data in older blocks cannot be changed, making the blockchain a pure record. Finally, the blockchain focuses on a decentralized information and data ledger with a high level of security.

The idea of smart contracts is another exciting component of blockchain technology. Technically, the smart contract is executed on the blockchain rather than through the usual legal negotiations. You set guidelines for the norms and penalties of the agreement and execute it automatically. Anyone can specify a particular function, and other computers are virtualized in intellectual property. However, due to its various shortcomings, blockchain technology may not be the optimal method for identifying and protecting digital intellectual property. Comparing digital media is a challenging and individual task. Much more investigation and creativity is needed to achieve intellectual property rights in cross-data applications.

6. Conclusions

This paper focuses on intellectual property rights in Big Data applications and addresses copyright issues in cross-data platforms. We have explored the problem of intellectual property in the context of cross-data platform services. The research in this paper attempts to look at intellectual property rights from two different angles. In the first part of the paper, we looked at the dimensions of the Big Data ecosystem, focusing mainly on legal and compliance issues, using non-technical solutions such as legal agreements (terms
of use, etc.). The second part deals with technical solutions, such as fingerprinting and watermarking algorithms, that can be used to enforce intellectual property rights. Here, each solution should be tailored to a specific intellectual property rights difficulty and based on a concrete study case. However, the main problem is to provide a non-technical security framework for dealing with intellectual property rights in Big Data, where data rights intersect with various information assets.

There is no one-size-fits-all approach to managing intellectual property rights in Big Data applications. For example, to enable equitable access to Big Data while ensuring regulatory compliance, any security framework must be able to manage the intellectual property rights of information assets and enforce those rights through technical processes. A homogeneous solution based on both is required to enforce intellectual property rights in cross-data applications. Finally, it is important to consider the elements of copyright regulations to avoid conflicts with intellectual property rights in cross-data platforms.

**Author Contributions:** Writing—original draft preparation: R.H.; writing—review and editing: H.P.

All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by the Personal Research Fund of Tokyo International University.

**Data Availability Statement:** Not applicable.

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

**References**

1. Hamza, R.; Dao, M.S. Privacy-preserving deep learning techniques for wearable sensor-based Big Data applications. *Virtual Real. Intell. Hardw.* 2022, 4, 210–222. [CrossRef]

2. Aquilani, B.; Piccarozzi, M.; Abbate, T.; Codini, A. The role of open innovation and value co-creation in the challenging transition from industry 4.0 to society 5.0: Toward a theoretical framework. *Sustainability* 2020, 12, 8943. [CrossRef]

3. Fukuda, K. Science, technology and innovation ecosystem transformation toward society 5.0. *Int. J. Prod. Econ.* 2020, 220, 107460. [CrossRef]

4. Holroyd, C. Technological innovation and building a ‘super smart’society: Japan’s vision of society 5.0. *J. Asian Public Policy* 2022, 15, 18–31. [CrossRef]

5. Miragliotta, G.; Sianesi, A.; Convertini, E.; Distante, R. Data driven management in Industry 4.0: A method to measure Data Productivity. *IFAC-PapersOnLine* 2018, 51, 19–24. [CrossRef]

6. Taeihagh, A.; Lim, H.S.M. Governing autonomous vehicles: Emerging responses for safety, liability, privacy, cybersecurity, and industry risks. *Transp. Rev.* 2019, 39, 103–128. [CrossRef]

7. Nair, M.M.; Tyagi, A.K.; Sreenath, N. The future with industry 4.0 at the core of society 5.0: Open issues, future opportunities and challenges. In Proceedings of the 2021 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 27–29 January 2021; pp. 1–7.

8. Hamza, R.; Zettsu, K. Investigation on Privacy-Preserving Techniques For Personal Data. In Proceedings of the 3rd ACM Workshop on Intelligent Cross-Data Analysis and Retrieval, Taipei, Taiwan, 21 August 2021; pp. 62–66.

9. Liang, G.; Weller, S.R.; Luo, F.; Zhao, J.; Dong, Z.Y. Distributed blockchain-based data protection framework for modern power systems against cyber attacks. *IEEE Trans. Smart Grid* 2018, 10, 3162–3173. [CrossRef]

10. Kuzminykh, L.; Carlsson, A.; Yevdokymenko, M.; Sokolov, V. Investigation of the IoT device lifetime with secure data transmission. In *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 16–27.

11. Hamza, R.; Dao, M.S.; Ito, S.; Koji, Z. Towards Intellectual Property Rights Protection in Big Data. In Proceedings of the 3rd ACM Workshop on Intelligent Cross-Data Analysis and Retrieval, Newark, NJ, USA, 27–30 June 2022; pp. 50–57.

12. Lu, Y.; Zhu, M. Privacy preserving distributed optimization using homomorphic encryption. *Automatica* 2018, 96, 314–325. [CrossRef]

13. Pulido-Gaytan, B.; Chernykh, A.; Cortés-Mendoza, J.M.; Babenko, M.; Radchenko, G.; Avetisyan, A.; Drozdov, A.Y. Privacy-preserving neural networks with Homomorphic encryption: Challenges and opportunities. *Peer-to-Peer Netw. Appl.* 2021, 14, 1666–1691. [CrossRef]

14. Murali, P.; Sankaradass, V. An efficient ROI based copyright protection scheme for digital images with SVD and orthogonal polynomials transformation. *Optik* 2018, 170, 242–264. [CrossRef]

15. Devi, B.P.; Singh, K.M.; Roy, S. A copyright protection scheme for digital images based on shuffled singular value decomposition and visual cryptography. *SpringerPlus* 2016, 5, 1091. [CrossRef] [PubMed]
16. Contreras, G.K.; Nahiyan, A.; Bhunia, S.; Forte, D.; Tehranipoor, M. Security vulnerability analysis of design-for-test exploits for asset protection in SoCs. In Proceedings of the 2017 22nd Asia and South Pacific Design Automation Conference (ASP-DAC), Chiba, Japan, 16–19 January 2017; pp. 617–622.

17. Zhang, D.Y.; Li, Q.; Tong, H.; Badilla, J.; Zhang, Y.; Wang, D. Crowdsourcing-based copyright infringement detection in live video streams. In Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Barcelona, Spain, 28–31 August 2018; pp. 367–374.

18. Hou, K.; Zhang, M. Discussion on legal model of intellectual property of computer software. Proc. J. Phys. Conf. Ser. 2021, 1883, 012011. [CrossRef]

19. Gürkaynak, G.; Yılmaz, I.; Yeşilaltay, B.; Bengi, B. Intellectual property law and practice in the blockchain realm. Comput. Law Secur. Rev. 2018, 34, 847–862. [CrossRef]

20. Cai, Z. Usage of deep learning and blockchain in compilation and copyright protection of digital music. IEEE Access 2020, 8, 164144–164154. [CrossRef]

21. Karjala, D.S. Intellectual Property Rights in Japan and the Protection of Computer Software. In Intellectual Property Rights in Science, Technology, and Economic Performance; Routledge: London, UK, 2019; pp. 277–289.

22. Melero, E.; Palomeras, N.; Wehrheim, D. The effect of patent protection on inventor mobility. Manag. Sci. 2020, 66, 5485–5504. [CrossRef]

23. Mansfield, E. Intellectual property, technology and economic growth. In Intellectual Property Rights in Science, Technology, and Economic Performance; Routledge: London, UK, 2019; pp. 17–30.

24. Fang, L.H.; Lerner, J.; Wu, C. Intellectual property rights protection, ownership, and innovation: Evidence from China. Rev. Financ. Stud. 2017, 30, 2446–2477. [CrossRef]

25. Lundqvist, B. Big data, open data, privacy regulations, intellectual property and competition law in an internet-of-things world: The issue of accessing data. In Personal Data in Competition, Consumer Protection and Intellectual Property Law; Springer: Berlin/Heidelberg, Germany, 2018; pp. 191–214.

26. Gervais, D. Exploring the interfaces between Big Data and intellectual property law. J. Intellect. Prop. Inf. Technol. Electron. Commer. Law 2019, 10, 3. [CrossRef]

27. Liu, C.; Hou, C.C. Challenges of Credit Reference Based on Big Data Technology in China. Mob. Netw. Appl. 2022, 27, 47–57. [CrossRef]

28. Liu, Z.; Shestak, V. Issues of crowdsourcing and mobile app development through the intellectual property protection of third parties. Peer-to-Peer Netw. Appl. 2021, 14, 2618–2625. [CrossRef]

29. Wang, S. The Solution of Network Intellectual Property Infringement under the Background of Big Data. Proc. J. Phys. Conf. Ser. 2020, 1533, 042048. [CrossRef]

30. Hoeren, T. Big data and the ownership in data: Recent developments in Europe. Eur. Intell. Prop. Rev. 2014, 36, 751–754.

31. Goddard, M. The EU General Data Protection Regulation (GDPR)—An Implementation and Compliance Guide; IT Governance Ltd.: Ely, UK, 2020.

32. Yoon, C. Theory of planned behavior and ethics theory in digital piracy: An integrated model. J. Bus. Ethics 2011, 100, 405–417. [CrossRef]

33. Gagliani, G. Cybersecurity, Technological Neutrality, and International Trade Law. J. Int. Econ. Law 2020, 23, 723–745. [CrossRef]

34. Shuaib, M.; Daud, S.M.; Alam, S.; Khan, W.Z. Blockchain-based framework for secure and reliable land registry system. Telkomnika 2020, 18, 2560–2571. [CrossRef]

35. Hon, W.K. Data Localization Laws and Policy: The EU Data Protection International Transfers Restriction through a Cloud Computing Lens; Edward Elgar Publishing: Cheltenham, UK, 2017.

36. Cavanillas, J.M.; Curry, E.; Wahlster, W. The Big Data value opportunity. In New Horizons for a Data-Driven Economy; Springer: Berlin/Heidelberg, Germany, 2016; pp. 3–11.

37. Alshammas, A.; Alzahrani, A.; Khan, W.Z. Blockchain-based framework for secure and reliable land registry system. Peer-to-Peer Netw. Appl. 2020, 13, 2618–2625. [CrossRef]

38. Alshammas, A.; Alzahrani, A.; Khan, W.Z. Blockchain-based framework for secure and reliable land registry system. Peer-to-Peer Netw. Appl. 2020, 13, 2618–2625. [CrossRef]

39. Alshammas, A.; Alzahrani, A.; Khan, W.Z. Blockchain-based framework for secure and reliable land registry system. Peer-to-Peer Netw. Appl. 2020, 13, 2618–2625. [CrossRef]
44. Parah, S.A.; Sheikh, J.A.; Loan, N.A.; Bhat, G.M. Robust and blind watermarking technique in DCT domain using inter-block coefficient differencing. *Digit. Signal Process.* 2016, 53, 11–24. [CrossRef]
45. Ghouti, L. A new perceptual video fingerprinting system. *Multimed. Tools Appl.* 2018, 77, 6713–6751. [CrossRef]
46. Li, Y.; Wang, D.; Tang, L. Robust and secure image fingerprinting learned by neural network. *IEEE Trans. Circuits Syst. Video Technol.* 2019, 30, 362–375. [CrossRef]
47. Chen, J.; Wang, J.; Peng, T.; Sun, Y.; Cheng, P.; Ji, S.; Ma, X.; Li, B.; Song, D. Copy, right? A testing framework for copyright protection of deep learning models. In Proceedings of the 2022 IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 22–26 May 2022; pp. 824–841.
48. Saadatpanah, P.; Shafahi, A.; Goldstein, T. Adversarial attacks on copyright detection systems. In Proceedings of the International Conference on Machine Learning, PMLR, Virtual Event, 13–18 July 2020; pp. 8307–8315.
49. Fatima, M.; Nisar, M.W.; Rashid, J.; Kim, J.; Kamran, M.; Hussain, A. A novel fingerprinting technique for data storing and sharing through clouds. *Sensors* 2021, 21, 7647. [CrossRef] [PubMed]
50. Monga, V.; Evans, B.L. Perceptual image hashing via feature points: Performance evaluation and tradeoffs. *IEEE Trans. Image Process.* 2006, 15, 3452–3465. [CrossRef]
51. Wei, S.; Zhao, Y.; Zhui, C.; Xu, C.; Zhu, Z. Frame fusion for video copy detection. *IEEE Trans. Circuits Syst. Video Technol.* 2011, 21, 15–28. [CrossRef]
52. Mohsenfar, S.; Mosleh, M.; Barati, A. Audio watermarking method using QR decomposition and genetic algorithm. *Multimed. Tools Appl.* 2013, 74, 759–779. [CrossRef]
53. Larch, A. Zplane Development, EAQUAL Evaluate Audio QUALity, Version: 0.1.3. 2012. Available online: http://www.mp3-tech.org/programmer/misc.html (accessed on 20 September 2022).
54. Yuan, Z.; Su, Q.; Liu, D.; Zhang, X. A blind image watermarking scheme combining spatial domain and frequency domain. *Vis. Comput.* 2021, 37, 1867–1881. [CrossRef]
55. Liu, D.; Su, Q.; Yuan, Z.; Zhang, X. A color watermarking scheme in frequency domain based on quaternion coding. *Vis. Comput.* 2021, 37, 2355–2368. [CrossRef]
56. Shen, Y.; Tang, C.; Xu, M.; Chen, M.; Lei, Z. A DWT-SVD based adaptive color multi-watermarking scheme for copyright protection using AMEF and PSO-GWO. *Expert Syst. Appl.* 2021, 168, 114414. [CrossRef]
57. Liu, D.; Su, Q.; Yuan, Z.; Zhang, X. A fusion-domain color image watermarking based on Haar transform and image correction. *Expert Syst. Appl.* 2021, 170, 114540. [CrossRef]
58. Chen, Y.; Jia, Z.G.; Peng, Y.; Peng, Y.X.; Zhang, D. A new structure-preserving quaternion QR decomposition method for color image blind watermarking. *Signal Process.* 2021, 185, 108088. [CrossRef]
59. Mosleh, M.; Setayeshi, S.; Barekatain, B.; Mosleh, M. A Novel Audio Watermarking Scheme Based on Fuzzy Inference System in DCT Domain. *Multimed. Tools Appl.* 2021, 80, 20423–20447. [CrossRef]
60. Liu, D.; Su, Q.; Yuan, Z.; Zhang, X. A blind color digital image watermarking method based on image correction and eigenvalue decomposition. *Signal Process. Image Commun.* 2021, 95, 116292. [CrossRef]
61. Jing, N.; Liu, Q.; Sugumaran, V. A blockchain-based code copyright management system. *Inf. Process. Manag.* 2021, 58, 102518. [CrossRef]
62. Sun, X.C.; Lu, Z.M.; Wang, Z.; Liu, Y.L. A geometrically robust multi-bit video watermarking algorithm based on 2-D DFT. *Multimed. Tools Appl.* 2021, 80, 13491–13511. [CrossRef]
63. Wang, B.; Jiawei, S.; Wang, W.; Zhao, P. Image Copyright Protection Based on Blockchain and Zero-Watermark. *IEEE Trans. Netw. Sci. Eng.* 2022, 9, 2188–2199. [CrossRef]
64. Halgamuge, M.; Guruge, D. Fair Rewarding Mechanism in Music Industry using Smart Contracts on Public-Permissionless Blockchain. *Multimed. Tools Appl.* 2022, 81, 1523–1544. [CrossRef]
65. Wu, X.; Ma, P.; Jin, Z.; Wu, Y.; Ou, W.; Han, W. A Novel Zero-Watermarking Scheme Based on NSCT-SVD and Blockchain for Video Copyright. *EURASIP J. Wirel. Commun. Netw.* 2022, 20. [CrossRef]
66. Tien, N.H.; Ngoc, N.M. Comparative Analysis of Advantages and Disadvantages of the Modes of Entering the International Market. *Int. J. Adv. Res. Eng. Manag.* 2019, 5, 29–36.
67. Chandra, G.R.; Liaqat, I.A. Commercialization of intellectual property; an insight for technocrats. In Proceedings of the 2019 International Conference on Automation, Computational and Technology Management (ICACTM), London, UK, 24–26 April 2019; pp. 373–378.
68. Drahos, P. *A Philosophy of Intellectual Property*; Routledge: London, UK, 2016.
69. Drahos, P.; Braithwaite, J. *Information Feudalism: Who Owns the Knowledge Economy?* Routledge: London, UK, 2017.