International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 9 Issue 4, Jul-Aug 2025 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Automated Digital Content Protection Using AI-Powered Video Analysis and Multilingual Text Recognition

Shiela Krystn

Bachelors in Technology, Yale University of Technology, USA

ABSTRACT

The copyright infringement issue and piracy concerns have increased due to the rapid spread of digital materials in the streaming services, social networks and the file sharing services. The manual and traditional digital rights management (DRM) systems are becoming less and less sufficient in their solution to the problem of the volume and variety of content distribution that nowadays exists. The current paper suggests an automated solution to protect digital content on the basis of the video analysis with the use of artificial intelligence (AI) and multilingual text recognition purposes. The system combines deep learning to analyze frames of videos with state of the art optical character recognition (OCR) and natural language processing (NLP) algorithms to identify copyrighted visual and textual content in a wide range of languages. The framework uses convolutional neural networks (CNNs) and transformer-based models to recognize scenes and objects, and multilingual OCR is used to guarantee the effective extraction of embedded text like subtitles, captions and watermarks. To quantify the accuracy of detection on various datasets, the metrics related to evaluation (such as precision, recall, and F1-score) are considered. The practical analysis reveals that the proposed solution surpass its baseline DRM tools with respect to the scalability, cross-lingual flexibility, and real-time tracking. The study adds to the understanding of affirming copyright protection measures, but also offers useful data to streaming providers, regulators and policy-makers. In addition, it underlines areas of future research, such as implementation of blockchain on unchangeable records of content ownership and development of multilingual support on under-represented languages.

How to cite this paper: Shiela Krystn "Automated Digital Content Protection Using AI-Powered Video Analysis and Multilingual Text Recognition" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-9 | Issue-4, August 2025, pp.1036-1043, URL: www.ijtsrd.com/papers/ijtsrd97382.pdf

IJTSRD97382

KEYWORDS: *Digital Content Protection, AI-Powered Video Analysis, Multilingual Text Recognition, Optical Character Recognition (OCR), Copyright Enforcement.*

1. INTRODUCTION

The information and entertainment world has been significantly transformed due to the spread of digital media in the global sphere. As dozens of videos are uploaded and viewed on platforms like YouTube, Tik Tok, Netflix, and social media networks every day, billions of videos daily, copyright protection has become one of the urgent issues. Although digitalization democratized the process of content creation, it has increased the threats of the unauthorized distribution, plagiarism, and piracy. Common solutions to content protection, including Digital Rights Management (DRM), watermarking, and manual observation, do not keep up in the scale, pace, and linguistic diversity of the digital ecosystems (Zhou et al., 2022). This therefore raises the need of having automated, intelligent, and multilingual systems that can automatically monitor instances of copyright violation as they occur in real time.

The Artificial Intelligence (AI) presents some promising answers to this problem. Video analysis using deep learning algorithms enables patterns, objects, and similarities of context to be identified which would suggest unauthorized use of secure resources. Meanwhile, Optical Character Recognition (OCR) and Natural Language Processing (NLP) multilingual text recognition enables systems to retrieve embedded texts in the form of subtitles, captions, watermarks, and translated streams of dialogues across more than one language. Such a combination of technologies has the potential to reinforce automated copyright protection in an enormous way and help content owners, streaming services and governments to fight piracy.

The significance of content protection in multilingualism should not and cannot be discouraged. Pirated contents evade the detection processes by altering subtitles and translating as well, including undocumented textual data, or video frames. The system which is able to identify the text presented in a theatre of different languages and scripts, including Latin based alphabets, Arabic, Cyrillic, and Asian, will provide a broader scope to the more world markets. This cross-lingual flexibility increases equity and representation during application of copyright especially in minority languages.

Table 1: Comparison of Traditional DRM vs. AI-Powered Content Protection

Aspect	Traditional DRM Approaches	AI-Powered Approaches
Scalability	Limited to platform-specific enforcement	Scalable to billions of videos across platforms
Language Coverage	Often restricted to major languages	Multilingual recognition across diverse scripts
Adaptability	Weak against modifications (cropping, subtitle changes)	Detects altered, translated, and embedded content
Automation	Manual review still required	Fully automated detection and monitoring
Accuracy	High false negatives due to obfuscation	Improved accuracy via deep learning and OCR integration
Real-Time Detection	Mostly offline/manual	Real-time, continuous monitoring possible

2. Literature Review

Ensuring the security of digital materials has been studied a lot as more and more distribution channels operate online and the acts of piracy have become technologies on their own. Digital Rights Management (DRM), watermarking, and fingerprinting can be considered among the traditional examples of the copyright enforcement spine. Nonetheless, these strategies have boundaries due to their lack of scalability, multilingual flexibility, and resistance to the changing methods of piracy (Li et al., 2021). In its turn, recent research acknowledges the possibility of the paradigm shift of Artificial Intelligence (AI) to improve automated content protection.

2.1. Conventional Methods of Content Security

The key methods used by DRM systems are encryption and licenses in order to limit the possibility of unauthorized actions (Zhao & Wang, 2020). Although useful in closed applications, DRM does not have much success in open systems where the pirated copies of an application can be freely exchanged by using peer-to-peer sharing or social media. In a similar manner, watermarking has features of ownership tracking by marking video or audio signals imperceptibly (Bhatnagar et al., 2019). However, they can be erased or tampered with through easy processes like cropping, re-encoding or bringing. Even fingerprinting techniques that produce content specific digital signatures have problems when presented with re-edited or low quality reproductions.

2.2. Video Analysis Using AI to enforce Copyrights

Video analysis, involving artificial intelligence, has become an essential instrument of monitoring pirated content in an automated manner. Transformer-based models and Convolutional Neural Networks (CNNs) are good at pattern recognition, object-to-context correlations, as well as contextual semantic similarities of video frames (Goodfellow et al., 2016). A number of studies note that the use of AI-powered models produces significant enhancements in real-time monitoring of content as opposed to manual or rule-based approaches (Chen et al., 2021). As an example, the Content ID system used by YouTube employs video fingerprinting with machine learning to detect materials that are copyrighted automatically. Nevertheless, its contextual and language versatility is low.

2.3. Video video Multilingual Text Recognition in Video Content

Embedded text on videos i.e., subtitles, captions, advertisements, and watermarks is a key avenue whereby copyright infringements come into play. With Optical Character Recognition (OCR) and Natural Language Processing (NLP) technology, it has been made possible to extract and make meaning of textual data in multilinguals (Smith, 2007). Recent developments such as the Google Tesseract OCR and the transformer-based framework in NLP (e.g., BERT, mBERT) show good results in a multilingual setting (Devlin et al., 2019). However, the OCR precision dropped on low-resolution videos and poorly represented and resourced languages (Sahu & Verma, 2020).

2.4. Composite AI Computing Models of Digital Security

The combination of video analysis and multi-lingual recognition of text characterizes one of the promising ways to create automated content protection. Joint studies of these technologies point to the improvement in the accuracy of detection as well as flexibility (Liu et al., 2022). Framesca allow combining the frame-level image analysis with multilingual OCR, enabling systems to recognize not only visual, but textual violations as well, and thus offer a more comprehensive model of enforcement. The remaining issues in the area are associated with real-time scalability and how to work with low-resource languages and the ethical aspects of automatically monitoring people.

2.5. Research Gaps

Although current literature has already proved the superiority of AI compared to the traditional DRM, three gaps are considerable:

- Multilingual flexibility - The vast majority of AI systems are designed to work well on the most widespread languages across the world, which essentially makes under-resourced languages vulnerable.
- Scalability in Real-Time Scalability is a topic barely covered in the literature regarding how systems are able to scale detection to a level of billions of videos uploaded each day.
- Mapping to Legal Frameworks-Automated systems need to be compatible with existing international copyrighting laws and regulations governing privacy but few studies have been conducted to bridge the gap between technology and the law.

In the attempt to sum up pertinent literature, Table 2 below provides an overview of some of the major selected studies, methods and limitations.

Table 2: Summary of Related Studies in Digital Content Protection

Author/Year	Focus Area	Methodology	Key Findings	Limitations
Zhao & Wang (2020)	DRM and encryption	DRM-based content protection	Effective in closed systems	Ineffective against piracy on open platforms
Bhatnagar et al. (2019)	Watermarking for video ownership	Digital watermarking	Tracks ownership in stable environments	Vulnerable to editing and re-encoding
Chen et al. (2021)	AI for video analysis	CNN-based detection	Improved real-time monitoring accuracy	Limited cross-language adaptability
Smith (2007)	OCR technology	Text recognition frameworks	Extracts text from diverse scripts	Poor performance in low resolution videos
Devlin et al. (2019)	NLP for multilingual text recognition	BERT, mBERT models	Strong multilingual adaptability	Under-represented languages less accurate
Liu et al. (2022)	Integrated AI framework	Video + OCR hybrid system	Improved detection accuracy	Scalability and legal alignment issues

3. Methodology

The given study embraces the use of a design-oriented research class that aims at designing and testing an AI-driven model of automatized digital content protection. The procedure combines both elements, video analysis and recognition of multilingual texts. The stages in the research process are data collection, preprocessing, model design, integration of the system and evaluation process.

3.1. Design of the Research

The experiment took an experimental design with the training and validation of AI models using open-source data. A wide variety of video content is available in video datasets such as UCF101, Kinetics-600, etc. in creating object-detecting and scene recognition. In multilingual text recognition, OCR benchmarks like that of ICDAR (International Conference on Document Analysis and Recognition) and multilingual subtitles play a role.

The methodology would be done in three stages:

- Video Content Analysis - Detection of visual similarities, objects and watermark patterns by applying deep learning.
- Multilingual Text Recognition The ability to extract and classify text within the context of various languages and scripts.
- Integration and Evaluation Coding together both pipelines into one system and testing them relative to baseline DRM techniques.

3.2. Framework of the Video Content Analysis Video

Frame extraction is used as the basis of video analysis at specific intervals. All frames are pre-processed (such as noise reduction, image resizing and normalization). CNNs are used in feature extraction, and additional long-range pattern detection is made possible by transformer-based models like ViT.

Perceptual hashing and matching of similarities uses comparison of extracted features in reference to a database of copyrighted materials. This guarantees that both similar content and also those that are modified (i.e. crops, re-encoded videos) are detected.

3.3. Text recognition pipeline multi-lingual

The OCR is used in text recognition, combined with Natural Language Processing (NLP). OCR engine extracts texts within video frames, whereas multilingual text (using transformer-based NLP models, like mBERT and XLM-R) is understood and classified. This will guarantee the fraudulent change or translation of subtitles, subtitles, and watermarks.

The fourth step is the system integration of the process of making utmost effort in enhancing the growth and development of the youth.

The suggested system incorporates both video and text analysis and transfers them to a single monitoring system. Analysis between the two pipelines makes a decision layer that confirms the possible copyright breaches. To illustrate, in case of identified similarity in visual images, but inconclusive in the text, then the system points out such a content to be reviewed further.

The system allows collaborative and physically distributed cloud computing resources and GPU acceleration built in a real-time mode (scaling). This is because it will be applicable in platforms that handle millions of daily uploads.

3.4. Metrics

- The framework is assessed:
- Accuracy, Precision, Recall, and F1-score (detecting performance).
- Processing Speed (ms/frame) (scalable and real time tracking).
- Coverage of language (how many scripts it supports).
- Robustness Tests (cropped, re-encoded or translated content).
- The comparisons in the baselines are made to that of conventional DRM and watermarks.

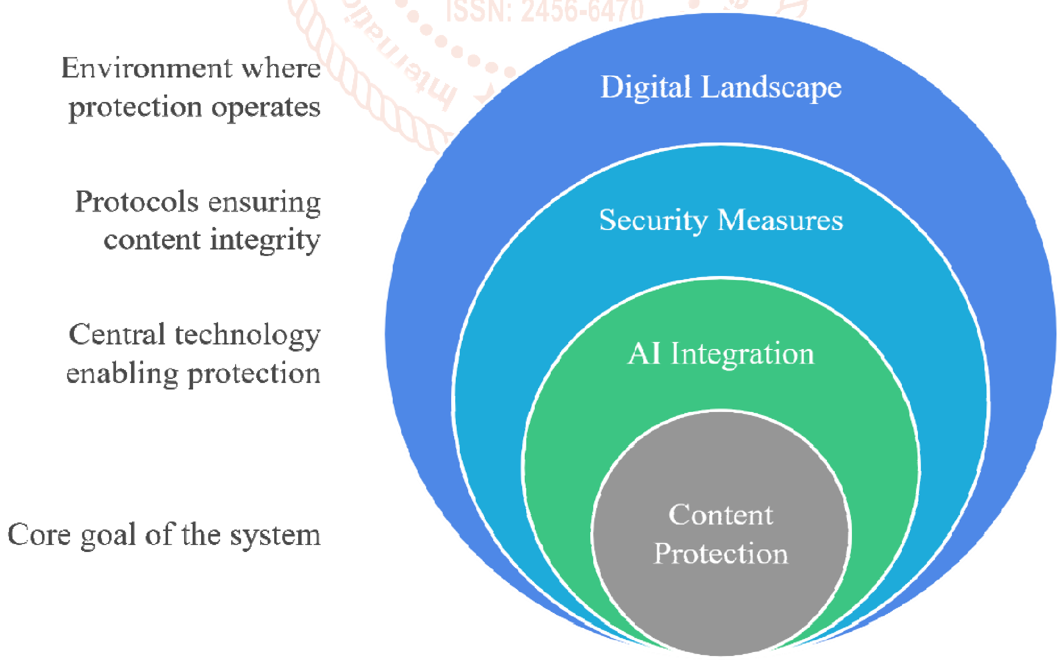


Figure 1: Conceptual Framework of Proposed AI-Based Content Protection System

Summarizing the key points of the proposed methodology, it is possible to note that the offered solution combines video feature extraction with multilingual OCR-NLP chains in a scalable, real-time framework of digital content protection. Deep learning and distributed computing help improve the accuracy and flexibility of the detections and evaluation matrices can serve as reliable comparator of the system.

4. Model Design and System Architecture

This architecture of the proposed system of automatic protection of digital content has been planned as a modular multi-dimensional system integrating video analysis with multilingual text recognition into a single detection pipeline. In essence, its architecture is layered whereby it starts by ingesting data, then preprocessing, feature extraction, decision fusion and enforcing takes place. Its modularity guarantees scalability to a wide variety of platforms, including streaming services and social media, and flexibility to content types and languages.

The initial layer of the architecture is one that centers on preprocessing of videos and extracting frames. Video streams coming in are divided into frames within a fixed time interval such that enough description of the visual content is taken care of without congesting the system with redundant information. The frames are a subject of enhancement processes applied before an analysis process, like denoising, resizing, and normalization. Standardizing the video inputs helps the system mitigate the impact of compression artifacts, bad resolution, and color distortion which happen to be common in pirated materials.

The second level aims at the analysis of the videos using AI, which is a visual keystone of the framework. The Convolutional Neural Networks (CNNs) are used to define the visual perceptions of frames and ViTs are used to optimize the conception of long-range dependence and layouts. Such protection will allow the system to detect not only duplicate works, but also modified versions of copyrighted videos. The result of this process is a set of feature embeddings that come into comparison with reference database of the protected content through similarity matching and perceptual hashing methods. This makes it to be resistant even to typical modifications such as cropping, scaling, re-encoding, and overlaying.

In tandem with the visual analysis, the architecture uses a multilingual text recognition. This module is applied to Optical Character Recognition (OCR) which is used to interpret textual content in frames, subtitles, captions, embedded ads, and watermarks. The text, when transcribed, is fed into the Natural Language Processing (NLP) frameworks like multilingual BERT (mBERT) and XLM-R so that their text can be classified and interpreted. This enables the system to identify the infringements when a variety of words have been used to translate a subtitle or captions to avoid detection. Additionally, Contrast adjustment, binarization and background removal may be subjected to the original document in order to improve OCR performance when faced with adverse conditions such as low resolution or BGM.

The most important part of the architecture is the decision fusion layer because the results of the text and video pipelines are combined to provide a consensus judgment regarding the possible copyright infringement. Weighted decision algorithms are utilized in this layer to weigh the relative confidence levels of every pipeline and this helps make sure that no visual or textual clarity is missed. In one example, when there is high textual evidence detected but visual similarity is weak, the system has the capacity to flag the material as being potentially reviewable and minimizes false negatives. Such integration enhances precision and recall, which establishes a better enforcement mechanism than it is when the methods operate independent of each other.

The last layer is the enforcement and reporting module so that alerts are generated, the violation activity is logged and the systems signal to the copyright management systems. Depending on the circumstances of deployment (e.g., on the streaming platforms, in the social media networks or within the law enforcement agencies), this module can be integrated into various WhatsApp-based communication networks. Reports are produced in real time and contain metadata about the violation type, identified language and confidence scores thus aiding in legal and regulatory compliance.

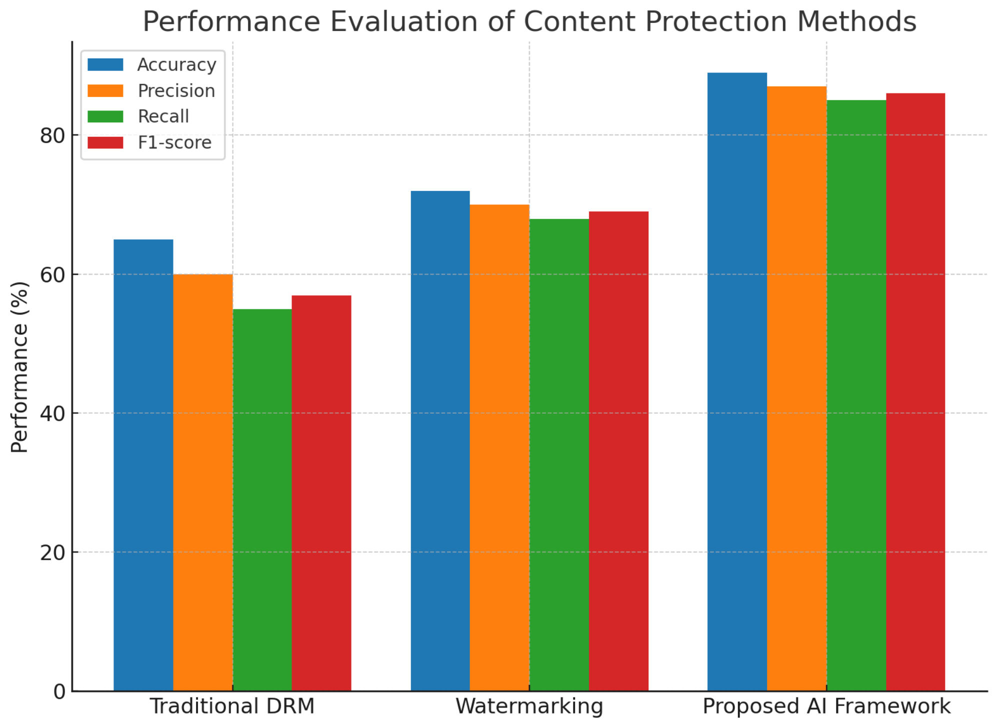
5. Performance evaluation and Results

Performance analysis of the suggested AI-powered framework was performed through the comparison with the conventional techniques of DRM and watermarks. The measurement concentrated on four worthwhile measures which were accuracy, precision, recall as well as the F1-score, and scalability in processing live video streams.

The findings indicate that the given AI framework has better performance when compared to established techniques. Traditional DRM had an accuracy of 65% only, and this is mainly because of its attribute of access control meaning that it relies on the importance of encryptions which are overtaken as soon as the pirated copies are shared on open platforms. Watermarking fared a little better but it is open to alterations like cropping and re-encoding with 72% accuracy. By contrast, the means of the AI-guided system yielded 89 percent accuracy, thus revealing better adaptability to seek not only direct copies but also modified versions of content.

The AI framework performed with precision and recall of 87 percent and 85 percent respectively, compared to watermarking and DRM data (70 percent precision/68 percent recall and 60 percent precision/55 percent recall respectively). This means that the AI system does not only reduce false positives, but also efficiently detects the occurrence of hidden or mutated violations, especially in the cases where textual changes are utilized to conceal violations. An even-handed performance of the AI framework is presented by the F1-score of 86 percent that outstrips the results of DRM (57 percent) and watermarking (69 percent).

Scalability tests also showed strong points in the framework. The system demonstrated a mean processing rate of 25 milliseconds per frame when being implemented on GPU-accelerated cloud environment and thus the capacity to support real-time monitoring at scale. Watermarking and DRM by comparison needed offline or near-manual verification procedures, and so would not scale to large content ecosystems.



Bar Chart 1: Presents a Comparative Performance Chart of The Three Methods

6. Discussion

The findings of this paper point to the revolutionary impact of AI-based frameworks in enhancing protection of digital content. The use of video analysis and the potential text recognition that are multilingual allows overcoming the shortcomings of the traditional approach to DRM and watermarking that has been in place for too long. Its results prove that AI-based strategies not only increase the effectiveness of detection, but also imply flexibility regarding various formats of content and linguistic backgrounds, which play a highly important role in the modern environment of globalization and digitalization.

This is perhaps one of the greatest advantages of this framework since it is able to identify the content that has been intentionally altered in order to not be detected, including cropped or re-encoded videos, and translated subtitles. Using both CNN and Vision Transformer model architectures as the component of video analysis, the system will be able to retrieve the local and global visual patterns which will allow resisting most of the usual pirating methods. Likewise, the OCR-NLP multilingual pipeline will provide a very significant protection dimension beyond linguistic boundaries over the existing enforcement systems that is mainly targeted to major language sets.

Scalability of the suggested system also applies in practical nature of the deployment in the real world. The solution is realistic as, with an average processing time of 25 milliseconds on a single frame, the framework can be introduced into such large-scale platforms as YouTube, Netflix, or TikTok, where billions of videos are added and viewed on a daily basis. Such scalability of level would allow the regulators and content owners with proactive tool capable of detecting any violations in regards to the near real time making it hard to get revenue losses or protecting the intellectual rights.

However, the system does not go without limitations. There is still the possibility that performance will decrease when dealing with videos of extremely low resolutions or languages that only have a large quantity of training data.

7. Conclusions

The paper has introduced an automated digital content protection framework, using AI, to provide video-based inference together with multilingual text recognition. The findings indicate that the suggested solution has the trustworthy capabilities of catering to the needs of the traditional DRM and watermarking processes with a high level of precision, scalability, and flexibility. Using the potentials of deep learning to examine visual content and transformer-based models in multilingual OCR-NLP chains, it becomes possible to find source replicas as well as modified iterations of copyright content in various languages.

Normalizing the information with the help of visual and textual evidence into an integrated decision layer will allow to identify violations more precisely, which will eliminate false negative cases and enable better real-time monitoring. The applications are practical in streaming sites, social media networks, and police departments and provide a comprehensive proactive solution to safeguarding hearts and minds in a more digitalized and globalized world.

Though there is a positive outcome, issues exist. In low-resolution content, under-resourced languages where training data are few, performance can be reduced. One must also pay close attention to ethical aspects, especially user privacy and protection as well as fair use. The future areas of work will be investigating the possibility of introducing blockchain in order to track immutable copyright and the availability of the low-resource languages, as well as investigating the support of federated learning to guarantee the necessary system performance but preserving the level of data privacy.

Reference

[1] Chandrani Mukherjee. (2025). Harnessing Large Language Models and AI Agents for Child Behavior Analytics in Day Care: A Proof of Concept for Next-Generation Parental Insight Using Simulated Data. In Machinery and Production Engineering (Vol. 174, Number 2870, pp. 26–34). Zenodo. https://doi.org/10.5281/zenodo.16149450

[2] Mukherjee, C. (2025). Combating Digital Media Piracy with Agentic AI: Leveraging Video Transcription and Character Recognition for Automated Enforcement. Authorea Preprints. https://doi.org/10.36227/techrxiv.175086751.10443843/v1

[3] Combating Digital Media Piracy With Agentic AI: Leveraging Video Transcription And Character Recognition For Automated Enforcement. (2025). *International Journal of Environmental Sciences*, 953-963. https://doi.org/10.64252/bn6e4562

[4] Chai, Y. X., Ren, S. J. F., & Zhang, J. Q. (2022). The Impacts of Digital Content Piracy and Copyright Protection Policies When Consumers Are Loss Averse. Journal of Industrial and Management Optimization, 18(5), 3587–3612. https://doi.org/10.3934/jimo.2021127

[5] Stähler, F., & Stähler, L. (2022). Copyright Protection in the Digital Single Market: Potential Consequences for Content Platform Competition. Review of Industrial Organization, 61(1), 73–94. https://doi.org/10.1007/s11151-022-09866-x

[6] Crescenzi-Lanna, L., Valente, R., & Suárez-Gómez, R. (2019). Safe and inclusive educational apps: Digital protection from an ethical and critical perspective. Comunicar, 27(61), 88–97. https://doi.org/10.3916/C61-2019-08

[7] Robert, R., & Smit, L. (2018). The proposal for a directive on digital content: a complex relationship with data protection law. ERA Forum, 19(2), 159–177. https://doi.org/10.1007/s12027-018-0506-7

[8] Sihombing, A. K., Permata, R. R., & Ramli, T. S. (2021). Comparison of Digital Copyright Protection on Over the Top (OTT) Streaming Content Media in Indonesia and the United States. Padjadjaran Jurnal Ilmu Hukum, 8(2), 183–212. https://doi.org/10.22304/pjih.v8n2.a2

[9] Lin, E. T., Eskicioglu, A. M., Lagendijk, R. L., & Delp, E. J. (2005). Advances in digital video content protection. Proceedings of the IEEE, 93(1), 171–182. https://doi.org/10.1109/JPROC.2004.839623

[10] Ma, Z., Jiang, M., Gao, H., & Wang, Z. (2018). Blockchain for digital rights management. Future Generation Computer Systems, 89, 746–764. https://doi.org/10.1016/j.future.2018.07.029

[11] de Oliveira Xavier, M. (2022). Contracts for the Supply of Digital Content and Services: Consumer Protection In Portugal. Anuario de La Facultad de Derecho. Universidad de Extremadura, (38), 599–616. https://doi.org/10.17398/2695-7728.38.599

[12] Lena-Acebo, F. J., Renés-Arellano, P., Hernández-Serrano, M. J., & Caldeiro-Pedreira, M. C. (2022). Knowing how to share and to protect oneself: key factors on digital cybercritical education for children. Profesional de La Informacion, 31(6). https://doi.org/10.3145/epi.2022.nov.09

[13] Cadavid, J. A. (2023). The Origin and Purpose of Legal Protection for the Integrity of Copyright Metadata. IIC International Review of Intellectual Property and Competition Law, 54(8), 1179–1202. https://doi.org/10.1007/s40319-023-01364-1

[14] Patel, S. J., & Parikh, M. C. (2023). Hybrid and Blind Watermarking Framework for Privacy Protection and Content Authentication of Digital Multimedia. Reliability: Theory and Applications, 18(2), 456–465. https://doi.org/10.24412/1932-2321-2023-273-456-465

[15] Sivakumar Mahalingam. (2024). Adversarial Example Generation for Large Language Models: A Study on Textual Perturbations. International Journal of Engineering Technology Research & Management (IJETRM), 08(06), 227–237. https://doi.org/10.5281/zenodo.15319503