
Multimodal Evidence-Based License Compliance Verification for KOGL Open Government Works on the Web

Seyoung Jang¹, Byeongchan Park¹, Youngmo Kim¹
and Sun-Jib Kim^{2,*}

¹*Department of Computer Science and Engineering, Soongsil University, Seoul, Republic of Korea*

²*Department of Convergence Security, Hansei University, Gunpo, Republic of Korea*

E-mail: Seyjang216@soongsil.ac.kr; pbc866@ssu.ac.kr; ymkim828@ssu.ac.kr; kimsj@hansei.ac.kr

**Corresponding Author*

Received 17 May 2026; Accepted 01 August 2026

Abstract

The reuse of open government works on the web has increased rapidly with the expansion of public data policies and artificial intelligence (AI)-assisted content utilization. In Korea, Korea Open Government Licenses (KOGLs) permit the reuse of public works under conditions such as attribution, non-commercial use, and no-derivatives. However, these conditions are frequently violated or misunderstood in web pages, advertisements, social media posts, and reused digital content. Since compliance-related evidence is distributed across textual descriptions, captions, license metadata, embedded image text, visual changes, and page-level context, manual inspection, and single-modal analysis are not sufficient for scalable web-based compliance review.

This paper proposes a multimodal evidence-based method for KOGL license compliance risk prediction in web environments. The proposed method combines textual context, KOGL metadata, optical character

Journal of Web Engineering, Vol. 25_7, 1419–1448.

doi: 10.13052/jwe1540-9589.2576

© 2026 River Publishers

recognition (OCR) extracted image text, and visual features to classify four risk types: attribution risk, non-commercial use risk, no-derivatives risk, and false KOGL claim risk. A lightweight sequential context model captures the contextual flow of compliance-related cues, while the decision module produces both risk probabilities and evidence-based explanations. Experiments on a KOGL-oriented dataset show that the proposed method achieves higher accuracy, macro-F1, and area under the curve (AUC) than text-only, image-only, and simple fusion baselines. The results demonstrate that multimodal web evidence improves compliance risk prediction and provides practical support for human review in public content management, open government data curation, and AI training data inspection.

Keywords: Web engineering, KOGL, open government works, license compliance, copyright risk prediction.

1 Introduction

The reuse of open government works has increased rapidly with the expansion of public data policies and web-based content distribution. Public-sector content is now widely reused in web pages, online advertisements, social media posts, educational materials, digital archives, and artificial intelligence (AI)-assisted content production. In Korea, the Korea Open Government License (KOGL) provides standardized reuse conditions for public works, including attribution, non-commercial use, and no-derivatives requirements [1]. KOGL is similar in purpose to open licensing schemes such as Creative Commons licenses [11], but it is specifically designed for Korean public-sector content and public works management.

The Korean public works reuse policy has been discussed in legal and institutional studies from various perspectives. Previous studies have examined the legal basis of public works reuse, copyright limitations for public works, institutional improvements for KOGL, and the relationship between public data policy and copyright law [4–10]. In addition, policy reports have addressed the need to improve the KOGL licensing structure and to promote the reuse of public works in practice [2, 3]. These studies show that open licensing can improve public access and reuse, but they also indicate that public works reuse requires consistent copyright information management and clearer compliance support.

However, practical reuse of KOGL-licensed works on the web often leads to compliance problems. Typical cases include missing attribution,

commercial reuse of non-commercial content, unauthorized modification of no-derivatives content, and false KOGL claims for works that contain third-party materials such as stock images, fonts, or externally owned visual elements. These problems do not always appear as explicit violations in a single text field. Instead, compliance-related evidence is distributed across heterogeneous web elements, including page text, captions, source statements, HTML metadata, embedded image text, watermarks, visual modifications, and surrounding page context.

This characteristic makes KOGL compliance verification a web engineering problem rather than a simple document classification task. In web environments, license and ownership information may be represented through structured metadata, page-level descriptions, embedded links, or visual elements. Web metadata vocabularies such as Schema.org provide properties for describing licensing information of creative works [13], while the W3C PROV model defines provenance information related to entities, activities, and agents involved in producing digital objects [12]. Nevertheless, such metadata is often incomplete, inconsistent, or lost during content redistribution. Therefore, metadata alone is insufficient for reliable KOGL compliance risk prediction.

Manual inspection is still widely used to review public content reuse, but it is difficult to scale this process to large volumes of web content. Rule-based approaches can detect simple cases, such as missing source phrases or license labels, but they have difficulty handling ambiguous context. For example, detecting non-commercial condition violations requires understanding whether the content is used in an advertising or revenue-generating context. No-derivatives violations require analyzing visual changes such as cropping, subtitle insertion, logo overlay, or color modification. False KOGL claims are even more difficult because they may involve inconsistencies between metadata, source information, and visual evidence.

Multimodal learning has shown that combining heterogeneous information sources can improve prediction performance when a task depends on both textual and visual evidence [14]. In KOGL compliance analysis, text-only approaches can capture source statements, license phrases, and commercial expressions, but they cannot sufficiently detect visual modification or third-party visual elements. Image-only approaches can capture modification traces, watermarks, or visual inconsistencies, but they cannot fully interpret license conditions or page-level context. Optical character recognition (OCR) techniques can partially bridge this gap by extracting text embedded in images, such as watermarks, subtitles, source marks,

and advertising phrases [17]. Therefore, KOGL compliance risk prediction requires an approach that jointly analyzes textual context, metadata, OCR-extracted text, and visual features.

To address these limitations, this paper presents a multimodal evidence-based approach for KOGL license compliance risk prediction in web environments. The proposed approach integrates textual context, KOGL metadata, OCR-extracted image text, and visual features to classify four types of compliance risk: attribution risk (BY), non-commercial use risk (NC), no-derivatives risk (ND), and false KOGL claim risk (CF). A lightweight sequential context model is used to capture the contextual flow of compliance-related cues, and the final decision process provides both risk probabilities and evidence-based explanations. Since compliance-related decisions may require legal or administrative review, the proposed approach is not intended to replace legal judgment. Instead, it supports human reviewers by prioritizing high-risk web content and presenting interpretable evidence. This direction is consistent with explainable prediction studies that emphasize the need to provide understandable reasons for automated decisions [20]. The main contributions of this paper are summarized as follows.

- **Web-oriented problem formulation:**
This paper formulates KOGL license compliance verification as a web engineering problem involving heterogeneous web evidence, including textual context, license metadata, OCR-extracted text, visual features, and page-level context.
- **Multimodal evidence-based risk prediction:**
This paper proposes a multimodal method that combines text, metadata, OCR, and image features to predict KOGL compliance risk types, including BY, NC, ND, and CF.
- **Evidence-aware decision support:**
The proposed method provides not only risk classification results but also evidence-based explanations that can support practical review by web content managers and public data administrators.
- **Experimental validation:**
Experiments on a KOGL-oriented dataset show that the proposed method achieves better performance than text-only, image-only, and simple fusion baselines in terms of accuracy, macro-F1, and area under the curve (AUC).

The remainder of this paper is organized as follows. Section 2 reviews related work on open government works, web metadata, multimodal web content

analysis, and explainable compliance prediction. Section 3 describes the proposed method for multimodal KOGL compliance risk prediction. Section 4 presents the experimental results and discussion. Section 5 concludes the paper and outlines future research directions.

2 Related Work

2.1 Open Government Works and KOGL License Compliance

Open government works are public-sector content made available for reuse by citizens, researchers, companies, and public institutions. In Korea, the KOGL provides standardized reuse conditions for public works. KOGL defines several license types based on attribution, non-commercial use, and no-derivatives conditions [1]. These conditions are similar in purpose to Creative Commons licenses, which provide standardized legal tools for sharing and reusing creative works under predefined conditions [11].

The legal and institutional basis of public works reuse in Korea has been widely discussed in previous studies. Hong examined copyright limitations for promoting the use of public works [4], while Kim analyzed legal issues related to the use and protection of copyrights in government works [5]. Kim also investigated the public works free-use scheme from the perspective of information law and institutional design [6]. Park and Kim discussed legal problems and countermeasures related to the free use of public works [7], and Kim and Park proposed improvement measures for promoting public works reuse [8]. These studies provide important legal and policy foundations for understanding public works reuse and KOGL license conditions.

Recent studies and reports have also emphasized the need for practical management of copyright information in public-sector content. Yoo analyzed the relationship between the Open Data Act and the Copyright Act in the provision and use of open data containing copyrighted works [9]. Seo and Joung examined copyright information management for increasing the free use of government publications as public works [10]. In addition, policy reports from the Korea Culture Information Service Agency addressed institutional improvements for the KOGL licensing structure and strategies for promoting the use of public works [2, 3].

However, most previous studies have focused on legal interpretation, institutional improvement, and policy-level discussion. Relatively limited attention has been paid to automated compliance risk prediction based on web content, metadata, OCR-extracted image text, and visual evidence. This

paper addresses this gap by treating KOGL compliance risk prediction as a web-based multimodal evidence analysis problem.

2.2 Web Metadata and Content Provenance

License compliance verification in web environments requires the analysis of metadata and provenance information. Web content often contains licensing and ownership information in different forms, such as HTML metadata, structured data, captions, source statements, alt text, and external links. Schema.org provides the license property for describing licensing information associated with creative works on the web [13]. Such structured metadata can help identify whether a content item includes license information and whether that information is consistent with the displayed content.

Provenance is also important for tracking the origin and reuse history of digital content. The W3C PROV data model defines provenance as information about entities, activities, and agents involved in producing or modifying data and digital objects [12]. In the context of KOGL compliance verification, provenance-related information can support the analysis of source institutions, publication history, modification records, and ownership relationships.

Nevertheless, metadata and provenance information are often incomplete or inconsistent in practical web environments. A web page may display a KOGL mark without sufficient attribution information, or an image may contain third-party visual elements that are not represented in metadata. In addition, copied or modified content may lose its original metadata during redistribution. Therefore, metadata alone is insufficient for reliable compliance risk prediction. It should be analyzed together with textual context, OCR-extracted text, and visual features.

2.3 Multimodal Learning for Web Content Analysis

Web content is inherently multimodal. A single web content item may include body text, captions, metadata, images, screenshots, embedded text, logos, watermarks, and surrounding page context. Multimodal machine learning has been studied as a way to integrate heterogeneous information sources and improve prediction performance when a task depends on multiple modalities [14].

In KOGL compliance risk prediction, each modality provides different evidence. Textual context can capture source statements, license phrases,

commercial expressions, and modification-related descriptions. Metadata can provide structured signals such as KOGL type, source URL, agency name, and publication date. OCR-extracted text can reveal embedded source marks, watermarks, subtitles, and advertising phrases inside images.

Visual features can capture cropping, logo overlay, color changes, stock watermarks, and other modification cues.

Several deep learning techniques are relevant to this multimodal analysis. Bidirectional Encoder Representations from Transformers (BERT)-based language models have been widely used to encode contextual information in text [16]. Vision transformer-based models have shown strong performance in visual representation learning [18], and lightweight variants such as TinyViT have been proposed for efficient visual feature extraction [19]. OCR techniques, including the Tesseract OCR engine, have also been used to extract text from document images and visual content [17]. These techniques provide useful foundations for analyzing text, images, and embedded image text in KOGL-related web content.

However, simply applying text or image models independently is not sufficient. Text-only analysis may fail to detect image modification or stock watermarks, while image-only analysis may fail to understand license conditions or commercial page context. Therefore, a multimodal approach is needed to combine textual, metadata, OCR, and visual evidence for compliance risk prediction.

2.4 Explainable Compliance Risk Prediction

KOGL compliance risk is often determined by the contextual relationship among multiple evidence elements rather than by a single keyword or image feature. For example, a commercial phrase may not indicate risk by itself, but it becomes important when it appears together with KOGL type 2 metadata, which restricts commercial use. Similarly, an image modification cue becomes more meaningful when the content is associated with a no-derivatives condition.

Sequential models can be used to capture such contextual dependencies. The gated recurrent unit (GRU) was introduced as an efficient recurrent structure for modeling sequential representations [15]. Compared with more complex recurrent units, GRU provides a practical balance between contextual modeling ability and computational efficiency. This characteristic is useful for web-based compliance review, where many content items may need to be screened repeatedly.

In compliance-related tasks, interpretability is also important. Prediction accuracy alone is not sufficient because human reviewers need to understand why a content item is considered risky. Explainable prediction studies have emphasized the need to provide interpretable evidence for model decisions [20]. In KOGL compliance analysis, useful evidence may include missing attribution phrases, commercial expressions, OCR-detected watermarks, visual modification cues, or inconsistent metadata.

Therefore, this paper combines multimodal evidence analysis, sequential context modeling, and evidence-aware output. The purpose is not to replace legal judgment, but to support human review by predicting compliance risk types and presenting interpretable evidence.

3 Multimodal Evidence-Based KOGL Compliance Risk Prediction

3.1 Overview

KOGL license compliance risk prediction on the web requires the analysis of heterogeneous evidence distributed across textual, visual, and metadata elements. A reused public work may contain a KOGL license label in the page metadata, attribution information in a caption, commercial intent in the surrounding text, and modification traces in the image itself. Therefore, relying on a single type of evidence is insufficient for identifying potential compliance risks in web-based content reuse.

This study presents a multimodal evidence-based approach for predicting KOGL compliance risks by integrating textual context, license metadata, OCR-extracted image text, and visual features. The approach is designed to classify four types of KOGL-related compliance risks: attribution risk, non-commercial use risk, no-derivatives risk, and false KOGL claim risk. These risk types correspond to common problems observed in the reuse of public works on web pages, online advertisements, social media posts, and AI-assisted content reuse environments.

The overall procedure of the multimodal evidence-based KOGL compliance risk prediction is shown in Figure 1.

The procedure consists of five main steps: web evidence representation, text/OCR/metadata/image feature encoding, multimodal feature reduction, sequential context modeling, and risk prediction with evidence output. First, heterogeneous web evidence is collected from textual descriptions, captions, KOGL metadata, image content, OCR-extracted text, and page-level

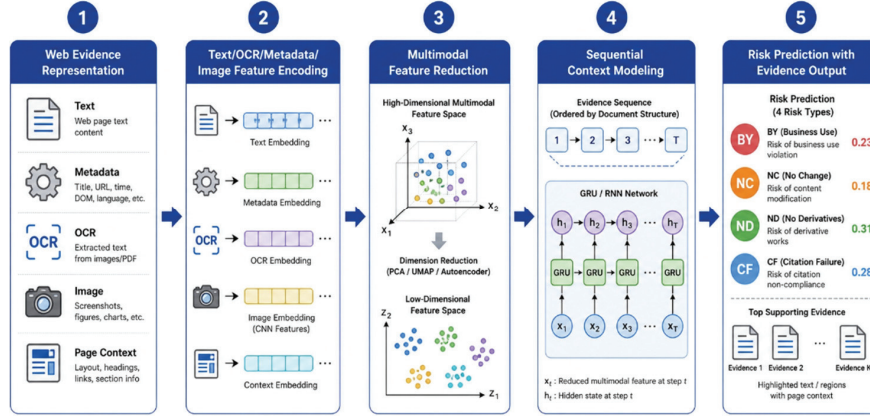


Figure 1 Overall procedure of multimodal evidence-based KOGL compliance risk prediction.

context. Second, each evidence type is preprocessed and encoded into feature representations. Third, redundant and noisy dimensions are reduced while preserving compliance-relevant information. Fourth, a lightweight sequential context model captures the contextual flow among evidence elements. Finally, the risk prediction module outputs the predicted risk type, risk probability, and supporting evidence for human review.

The purpose of the proposed approach is not to make a final legal decision on copyright infringement. Instead, it provides compliance-oriented decision support by prioritizing web content that may require additional review. The output includes not only the predicted risk category but also interpretable evidence, such as missing attribution phrases, commercial expressions, image modification cues, or ownership inconsistencies. This design allows public content managers and web service administrators to review high-risk cases more efficiently.

In this study, a web content item is treated as a collection of multimodal evidence rather than as a single document or image. This design is important because KOGL compliance risks often arise from inconsistencies among different web elements. For example, a page may display a KOGL mark while omitting the required source statement, or an image may be labeled as a public work while containing a stock watermark. By jointly analyzing text, metadata, OCR text, and visual features, the proposed approach aims to improve the reliability of KOGL compliance risk prediction in web environments.

3.2 Web Evidence Representation

A KOGL-related web content item contains multiple types of evidence that may be relevant to license compliance verification. In many cases, the evidence is not located in a single field. Attribution information may appear in a caption or source statement, commercial intent may be implied by surrounding page text, license conditions may be stored in metadata, and modification traces may appear only in the image. Therefore, this study represents a web content item as a multimodal evidence set.

Let W denote a KOGL-related web content item to be reviewed. The web evidence representation is defined as shown in Equation (1).

$$W = \{T, M, O, V, P\} \quad (1)$$

where T represents textual context, M represents license metadata, O represents OCR-extracted image text, V represents visual features, and P represents page-level context. The structure of the web evidence representation is illustrated in Figure 2.

The textual context T includes titles, descriptions, captions, source statements, and surrounding sentences. These elements are mainly used to identify attribution phrases, commercial expressions, and descriptions of content modification. For example, the absence of an agency name or source URL may indicate attribution risk, while expressions such as “purchase,” “discount,” or “promotion” may indicate possible non-commercial condition risk.

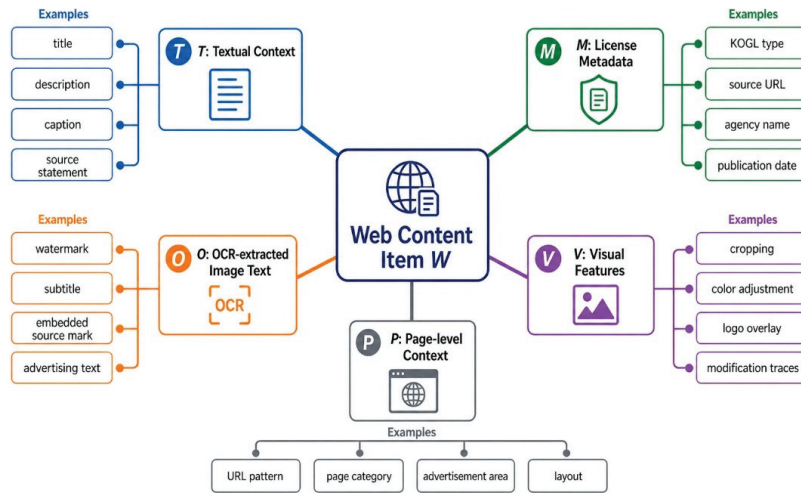


Figure 2 Web evidence representation for KOGL compliance risk prediction.

The license metadata M includes KOGL type, source URL, agency name, author information, publication date, and license label. Metadata is important because KOGL conditions are usually represented as structured or semi-structured information. However, metadata alone is not sufficient because it may be incomplete, inconsistent, or separated from the actual reused content.

The OCR-extracted text O is obtained from images or screenshots. This evidence is useful for detecting embedded source marks, watermarks, subtitles, advertising text, and third-party ownership indicators. In web environments, important compliance-related information is often embedded inside images rather than provided as HTML text. Therefore, OCR text is treated as an independent evidence type.

The visual feature V represents image-level information, including visual modification cues such as cropping, resizing, color adjustment, subtitle insertion, logo overlay, and watermark presence. These features are especially important for identifying no-derivatives risk and false KOGL claim risk.

The page-level context P includes URL patterns, page category, external links, advertisement areas, and surrounding layout information. This evidence helps interpret whether the reused content appears in a commercial, educational, administrative, or informational context.

The output of the prediction task consists of three elements: the predicted compliance risk type, the overall risk probability, and the supporting evidence. These outputs are defined as shown in Equations (2), (3), and (4), respectively.

$$y \in \{BY, NC, ND, CF\} \quad (2)$$

$$P_{\text{risk}} \in [0, 1] \quad (3)$$

$$E = \{e1, e2, \dots, ek\} \quad (4)$$

where y is the predicted compliance risk type, P_{risk} is the overall risk probability, and E is a set of evidence items supporting the prediction.

The four KOGL compliance risk types and their representative web evidence are summarized in Table 1.

To illustrate the relationship between KOGL license types, web content analysis, and compliance risk detection, Figure 3 presents a conceptual overview of the KOGL license compliance verification process.

The diagram shows how different KOGL license types (type 1 through type 4) are monitored across various web environments, and how multimodal evidence analysis enables the detection of four primary risk categories: attribution risk (BY), non-commercial use risk (NC), no-derivatives risk (ND),

Table 1 KOGL compliance risk types and representative web evidence

Risk Type	Description	Representative Web Evidence
BY	Attribution risk	Missing source, missing agency name, incomplete attribution phrase
NC	Non-commercial use risk	Advertisement, sales phrase, monetized context, promotional page
ND	No-derivatives risk	Cropping, editing, subtitle insertion, logo overlay, color modification
CF	False KOGL claim risk	Stock watermark, third-party material, ownership inconsistency, metadata mismatch

and false KOGL claim risk (CF). This conceptual framework provides the foundation for the multimodal evidence-based approach proposed in this study.

This representation allows the model to analyze not only whether a KOGL label exists, but also whether the license condition is consistent with the surrounding web evidence. For example, if a content item has KOGL type 2 metadata but appears on a promotional page with sales-related expressions, the item may be assigned a high NC risk. Similarly, if an image labeled as KOGL content contains a stock watermark or third-party logo, the item may be assigned a high CF risk.

3.3 Text, OCR, Metadata, and Visual Feature Encoding

After defining the web evidence representation, each evidence type is transformed into a feature representation suitable for risk prediction. The encoding process handles textual context, OCR-extracted image text, license metadata, and visual content separately, and then prepares them for multimodal integration.

First, OCR is applied to the input image or screenshot to extract text embedded in visual content. The OCR extraction process is defined as shown in Equation (5).

$$O = \text{OCR}(I) \quad (5)$$

where I denotes an input image or screenshot, and O denotes the text extracted from the image. OCR-extracted text may include source marks, watermarks, subtitles, advertising phrases, agency names, or third-party ownership indicators. This step is important because compliance-related information is often embedded inside images rather than written in the HTML text of a web page.

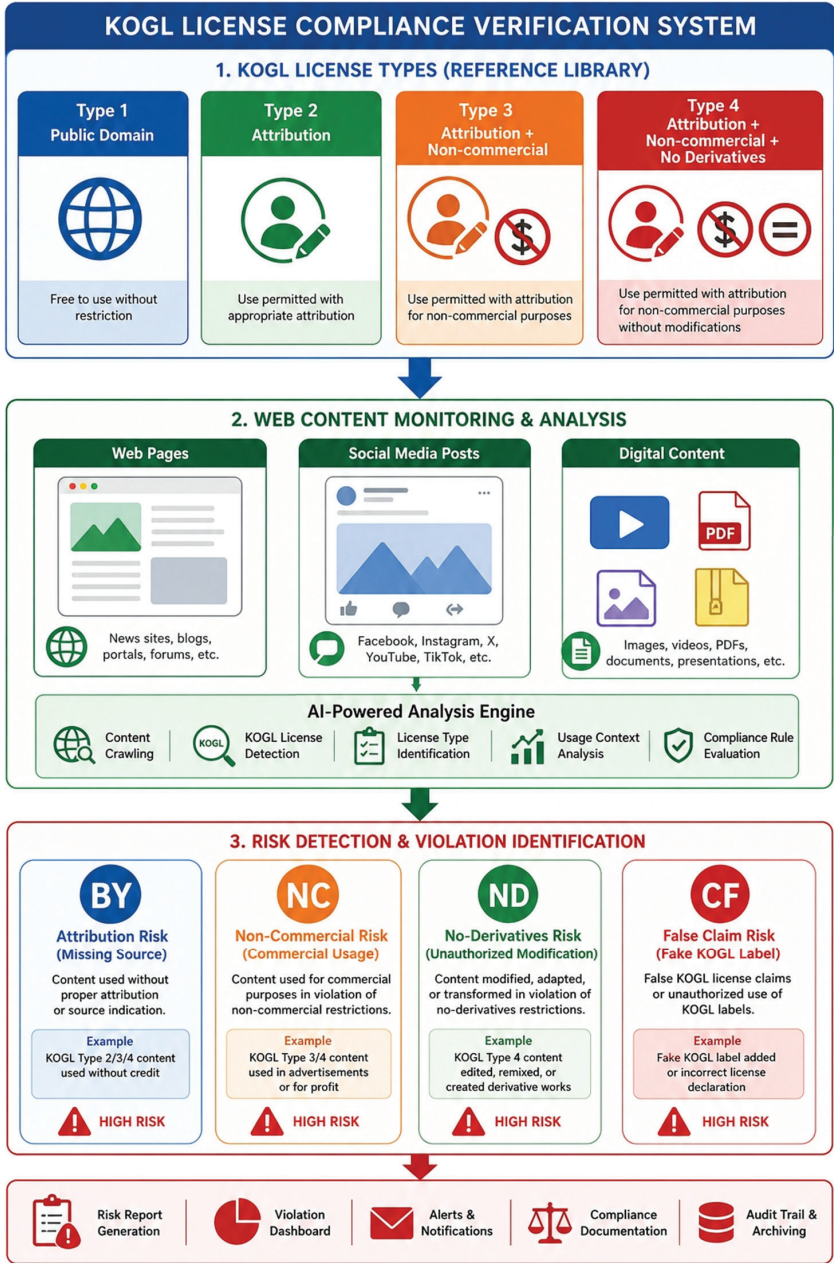


Figure 3 Conceptual diagram of KOGL license compliance verification: from license types through web content analysis to risk detection.

The textual context and OCR-extracted text are then combined and divided into sentence-level units. Let us denote the t -th sentence in the combined textual sequence. Each sentence is tokenized and converted into an embedding vector, as shown in Equation (6).

$$x_t = \text{Embed}(\text{token}(s_i)), \quad t = 1, \dots, n \quad (6)$$

where $\text{token}(\cdot)$ denotes the tokenization function, $\text{Embed}(\cdot)$ denotes the embedding function, and n denotes the number of sentence units. The resulting textual sequence is represented as shown in Equation (7).

$$X = [x_1, x_2, \dots, x_n] \quad (7)$$

The sequence X includes both ordinary web text and OCR-extracted image text. Therefore, it can represent attribution phrases, commercial expressions, modification-related descriptions, and embedded image text in a unified textual representation.

License metadata is encoded separately because it contains structured or semi-structured information related to KOGL conditions. The metadata vector is defined as shown in Equation (8).

$$m = \text{Encode}(M) \quad (8)$$

where M denotes the metadata evidence defined in Equation (1), and denotes the encoded metadata vector. The metadata includes KOGL type, source URL, agency name, author field, publication date, and license label. Categorical fields such as KOGL type are converted into embedding vectors, while binary indicators such as source existence, agency name existence, or KOGL mark existence are represented as numerical features.

For visual content, the input image is first normalized to reduce the effect of pixel-scale variation. The image normalization process is defined as shown in Equation (9).

$$I' = \frac{I - \mu}{\sigma} \quad (9)$$

where I' denotes the normalized image, and μ and σ denote the mean and standard deviation of image pixels, respectively. The normalized image is then passed to an image encoder to obtain a visual feature vector, as shown in Equation (10).

$$v = \text{ImageEncoder}(I') \quad (10)$$

where v denotes the visual embedding. The visual embedding captures image-level cues such as cropping, resizing, color adjustment, subtitle insertion, logo overlay, watermark presence, and other visual inconsistencies. These cues are particularly useful for identifying no-derivatives risk and false KOGL claim risk.

Finally, the encoded evidence representations are prepared for multimodal integration. The encoded textual sequence X , metadata vector m , OCR-derived textual information O , and visual embedding v are used as the main inputs for the next stage. The integrated representation is defined in the following section.

3.4 Multimodal Evidence Fusion and Feature Reduction

After encoding each evidence type, the next step is to integrate the textual, metadata, OCR-derived, and visual representations into a unified multimodal representation. Since KOGL compliance risk is often determined by the relationship among different evidence elements, the encoded features should be analyzed jointly rather than independently.

The multimodal evidence representation is constructed by combining the textual sequence X , metadata vector m , OCR-derived textual evidence O , and visual embedding v . The fusion process is defined as shown in Equation (11).

$$H = [X; O; m; v] \quad (11)$$

where $[:]$ denotes the concatenation operation, and H denotes the integrated multimodal evidence representation. This representation contains compliance-related cues from different web elements, such as source statements, KOGL metadata, image-embedded text, and visual modification traces.

However, direct concatenation may introduce redundant or noisy information. For example, a web page may contain repeated license phrases, irrelevant descriptive text, decorative image elements, or metadata fields that do not directly contribute to compliance risk prediction. Therefore, a feature reduction process is applied to preserve compliance-relevant information while reducing unnecessary dimensions.

Given an input feature vector h , the reduced representation z is computed as shown in Equation (12).

$$z = C(\text{ReLU}(R(h))) \quad (12)$$

where R denotes a dimensionality reduction mapping, C denotes a compliance-relevant feature projection, and $\text{ReLU}(\cdot)$ denotes the activation function. The purpose of this operation is to reduce redundant dimensions while maintaining features related to attribution, commercial context, modification, and ownership inconsistency.

The feature reduction operation is applied to textual and visual representations separately, as shown in Equation (13).

$$z_t = \text{FR}(X), z_v = \text{FR}(v) \quad (13)$$

where $\text{FR}(\cdot)$ denotes the feature reduction operation, z_t denotes the reduced textual representation, and z_v denotes the reduced visual representation. The metadata vector m and OCR-derived text representation are also included in the integrated evidence representation when constructing the final sequence for contextual modeling.

The final reduced multimodal representation is defined as shown in Equation (14).

$$Z = [z_t; O; m; z_v] \quad (14)$$

where Z denotes the reduced multimodal evidence representation. This representation is used as the input to the sequential context modeling step described in the next section.

The feature reduction process plays two roles in the proposed approach. First, it improves computational efficiency by reducing the dimensionality of high-dimensional text and image embeddings. Second, it helps the model focus on compliance-relevant cues rather than general web content features. For example, attribution phrases, commercial expressions, editing-related terms, stock watermarks, and license metadata inconsistencies are more important than unrelated decorative text or background image patterns.

This design is particularly useful for web-based compliance review because web content often contains noisy and heterogeneous information. By reducing redundant features before sequential modeling, the proposed approach can analyze the contextual relationship among evidence elements more effectively.

3.5 Sequential Context Modeling

KOGL compliance risk is not always determined by a single keyword or a single visual cue. In many cases, the risk is identified from the contextual relationship among multiple evidence elements. For example, a content item

may include a KOGL label, but the surrounding page text may indicate commercial promotion. Similarly, a source phrase may be partially provided in a caption, while the image itself may contain editing traces or thirdparty watermarks. Therefore, the temporal and contextual flow of evidence should be considered in the prediction process.

To model these relationships, the reduced multimodal evidence representation Z is processed using a lightweight sequential context model. The input sequence is defined as shown in Equation (15).

$$Z = [z_1, z_2, \dots, z_n] \quad (15)$$

where z_t denotes the multimodal evidence representation at step t , and n denotes the number of evidence units. Each z_t may contain textual, OCR-derived, metadata, or visual information depending on the structure of the reviewed web content.

A GRU-based sequence model is used to capture the contextual dependency among evidence units. The final hidden representation is computed as shown in Equation (16).

$$h_T = \text{GRU}(Z) \quad (16)$$

where h_T denotes the final hidden state summarizing the multimodal evidence sequence. This hidden representation is used for KOGL compliance risk prediction in the next step.

The GRU model is selected because it provides a practical balance between contextual modeling ability and computational efficiency. In web-based compliance review, a large number of content items may need to be screened repeatedly. Therefore, a lightweight sequence model is suitable for prioritizing potentially risky content without requiring excessive computational resources.

The sequential context model captures patterns such as the following:

- CopyKOGL type 2 metadata → promotional text → high NC risk
- KOGL type 3 metadata → image editing cue → high ND risk
- KOGL label exists → stock watermark detected → high CF risk
- missing source phrase → incomplete attribution → high BY risk

These patterns show that KOGL compliance risk depends not only on the presence of individual evidence but also on the relationship among evidence elements. For example, the phrase “promotion” may not always indicate a risk by itself. However, if it appears together with KOGL type 2 metadata, which includes a non-commercial condition, the content may require additional

review. Likewise, image modification cues become more important when the original KOGL condition does not allow derivative works.

The hidden representation h_T therefore serves as a contextual summary of the reviewed web content. It reflects the combined meaning of textual context, metadata, OCR-extracted text, and visual features. This representation is passed to the risk prediction step to estimate the most likely KOGL compliance risk type and the overall risk probability.

3.6 Risk Prediction and Evidence Output

The final step predicts the KOGL compliance risk type and estimates the overall risk probability using the contextual hidden representation. The prediction output is designed to support human review by providing not only a classification result but also interpretable evidence related to the predicted risk.

First, the risk type is predicted using a softmax classifier. The risk type probability is computed as shown in Equation (17).

$$p_{\text{type}} = \text{softmax}(W_{\text{type}}h_t + b_{\text{type}}) \quad (17)$$

where p_{type} denotes the probability distribution over the four KOGL compliance risk types, W_{type} denotes the weight matrix, and b_{type} denotes the bias term. The predicted risk type is selected as shown in Equation (18).

$$\hat{y} = \text{argmax}(p_{\text{type}}) \quad (18)$$

where $\hat{y}$ is the predicted risk type, and $\hat{y} \in \{\text{BY}, \text{NC}, \text{ND}, \text{CF}\}$.

In addition to risk type classification, the overall compliance risk probability is estimated using a sigmoid function. The risk probability is computed as shown in Equation (19).

$$p_{\text{risk}} = \sigma(W_{\text{risk}}h_T + b_{\text{risk}}) \quad (19)$$

where p_{risk} denotes the overall risk probability in the range of $[0, 1]$, W_{risk} denotes the weight matrix, b_{risk} denotes the bias term, and $\sigma(\cdot)$ denotes the sigmoid function. A higher value of p_{risk} indicates that the reviewed content is more likely to require additional compliance review.

The final output consists of the predicted risk type, the risk probability, and supporting evidence. The evidence is extracted from compliance-related cues such as missing attribution phrases, commercial expressions, OCR-detected text, visual modification cues, and metadata inconsistencies. The output format is summarized in Table 2.

Table 2 Output format of KOGL compliance risk prediction

Output Element	Description	Example
Predicted risk type	Most likely KOGL compliance risk category	NC
Risk probability	Overall probability of compliance risk	0.87
Supporting evidence	Evidence contributing to the prediction	Commercial phrase detected in OCR text
Suggested review action	Human-readable recommendation	Review whether the use is commercial

Table 3 Representative evidence and suggested review actions by risk type

Risk Type	Representative Evidence	Suggested Review Action
BY	Source phrase, agency name, or publication year is missing	Check attribution and add source information
NC	Advertisement, purchase, discount, or monetization phrase is detected	Review whether the content is used for commercial purposes
ND	Cropping, subtitle insertion, logo overlay, or color modification is detected	Verify whether derivative use is allowed
CF	Stock watermark, third-party logo, or metadata mismatch is detected	Verify ownership before applying a KOGL label

The evidence output is generated to make the prediction result understandable for human reviewers. For example, when a web content item has KOGL type 2 metadata and includes commercial phrases such as “purchase,” “discount,” or “promotion,” the output may indicate a high NC risk. When a KOGL type 3 content item contains visual modification cues such as cropping, subtitle insertion, or logo overlay, the output may indicate a high ND risk. Similarly, if a KOGL-labeled image contains a stock watermark or third-party logo, the output may indicate a high CF risk.

Representative evidence and review actions for each risk type are summarized in Table 3.

This evidence-based output is important because KOGL compliance verification often requires contextual and administrative interpretation. The proposed approach does not replace legal judgment. Instead, it provides risk-oriented information that helps public content managers and web administrators prioritize content requiring further review.

For example, if the predicted risk type is NC and the risk probability is 0.87, the reviewer can examine the detected commercial expressions and the

KOGL license type before making a final decision. Likewise, if the predicted risk type is CF, the reviewer can verify whether the content contains third-party materials or whether the KOGL metadata is inconsistent with the actual ownership information.

Through this output structure, the proposed approach supports practical compliance review in web-based public content management. It enables reviewers to identify high-risk cases more efficiently while maintaining interpretability of the prediction result.

4 Experiments and Discussion

This section evaluates the effectiveness of the multimodal evidence-based KOGL compliance risk prediction approach. The experiments were designed to answer whether multimodal web evidence improves prediction performance compared with single-modal and simple fusion baselines. The experiments were designed to examine whether multimodal web evidence improves prediction performance compared with rule-based, single-modal, and simple fusion baselines.

4.1 Dataset

The dataset used in this study was constructed from KOGL-related public content, consultation cases, textual descriptions, source statements, images, and modified image samples. Each content item was represented as a multimodal evidence instance consisting of textual context, license metadata, OCR-extracted image text, visual features, and pagelevel context.

The dataset composition is summarized in Table 4.

The textual data include descriptions, captions, source statements, consultation-case sentences, and surrounding page text. These elements were used to identify attribution-related cues, commercial expressions, and modification-related descriptions. The image data include original KOGL-related images, screenshots, and modified image samples. OCR was applied

Table 4 Dataset composition

Data Type	Count/Description
Textual data	12,450
Image and screenshot data	4380
Risk labels	4 classes: BY, NC, ND, CF
Data split	Train 70%/validation 15%/test 15%

to extract embedded text such as watermarks, subtitles, source marks, and advertising phrases.

Each data instance was labeled according to one of the four KOGL compliance risk types: attribution risk (BY), non-commercial use risk (NC), no-derivatives risk (ND), and false KOGL claim risk (CF). The labeling criteria were based on KOGL license conditions and representative web evidence. The risk type definitions used for labeling are consistent with Table 1.

For example, a content item was labeled as BY risk when the required source phrase, agency name, or publication information was missing. A content item was labeled as NC risk when non-commercial KOGL content appeared in a commercial or promotional context. ND risk was assigned when visual modification cues such as cropping, subtitle insertion, logo overlay, or color adjustment were observed in content with no-derivatives conditions. CF risk was assigned when a KOGL label was inconsistent with ownership evidence, such as stock watermarks, third-party logos, or metadata mismatch.

To reduce evaluation bias, the dataset was divided into training, validation, and test sets using a 70:15:15 ratio. The training set was used for model learning, the validation set was used for parameter tuning, and the test set was used only for final performance evaluation.

4.2 Experimental Settings

The experiments were implemented using PyTorch. The experimental environment was configured to reflect a lightweight web-based verification scenario, where large volumes of content may need to be processed without relying on excessive computational resources. The experimental environment is summarized in Table 5.

The text encoder was used to transform textual context and OCR-extracted text into embedding representations. The image encoder was used

Table 5 Experimental environment

Item	Specification
OS	Ubuntu 20.04
Deep learning library	PyTorch 2.3
GPU	NVIDIA RTX 3060 12GB
CPU/RAM	Intel i7 / 32GB
Text processing	KoBERT tokenizer / encoder
Image processing	Tiny-ViT-based image encoder
Training repetitions	1000

Table 6 Baseline models used for comparison

Model	Description
Rule-based method	Uses predefined rules based on keywords, license labels, and metadata fields
Text-only model	Uses textual context and source statements only
Image-only model	Uses image features only
OCR + text model	Uses textual context and OCR-extracted image text
Simple fusion model	Concatenates text and image features without contextual modeling
Proposed approach	Uses text, metadata, OCR text, visual features, feature reduction, and sequential context modeling

to extract visual features from original and modified images. Metadata fields, including KOGL type, source URL existence, agency name existence, and license label, were encoded as categorical or numerical features depending on their data type.

The proposed approach was compared with several baseline methods. The baseline methods were selected to evaluate the contribution of multimodal evidence integration and contextual modeling. The compared models are summarized in Table 6.

The rule-based method was included to represent a practical baseline that can detect simple cases such as missing source phrases or explicit commercial keywords. The text-only and image-only models were used to examine the limitations of single-modal analysis. The OCR + text model was used to evaluate the effect of image-embedded text. The simple fusion model was used to compare direct feature concatenation with the proposed multimodal evidence fusion and sequential context modeling.

4.3 Evaluation Metrics

The performance of each model was evaluated using accuracy, macro-F1, and AUC. Accuracy measures the overall proportion of correctly classified samples. Macro-F1 was used because the four KOGL compliance risk types may have different distributions and difficulty levels. AUC was used to evaluate the discriminative ability of the model in estimating compliance risk probability.

Accuracy is defined as shown in Equation (20).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (20)$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

Precision and recall are defined as shown in Equations (21) and (22).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (21)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (22)$$

F1-score is defined as shown in Equation (23).

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (23)$$

Macro-F1 is calculated by averaging the F1-score of each risk class, as shown in Equation (24).

$$\text{Macro-F1} = \frac{1}{C} \sum_{i=1}^C \text{F1}_i \quad (24)$$

where C denotes the number of classes. In this study, $C = 4$, corresponding to BY, NC, ND, and CF.

In addition to classification metrics, inference latency was also considered to examine whether the proposed approach can be applied to lightweight web-based compliance review environments. Latency was measured as the average prediction time per content item in the test set.

4.4 Overall Performance

The overall performance comparison is summarized in Table 7.

The proposed approach achieved the highest performance among all compared models, recording an accuracy of 0.91, macro-F1 of 0.89, and

Table 7 Overall performance comparison

Model	Accuracy	Macro-F1	AUC
Rule-based method	0.69	0.65	0.70
Text-only model	0.78	0.74	0.79
Image-only model	0.71	0.68	0.72
OCR + text model	0.81	0.78	0.82
Simple fusion model	0.83	0.81	0.84
Proposed approach	0.91	0.89	0.93

AUC of 0.93. These results indicate that combining textual context, metadata, OCR-extracted text, and visual features is more effective than using a single evidence type.

The rule-based method showed limited performance because it mainly depends on explicit keywords and predefined metadata conditions. Although it can detect simple attribution omissions or obvious commercial expressions, it cannot effectively handle ambiguous web contexts or visual modification cues.

The text-only model performed better than the rule-based method, achieving an accuracy of 0.78 and macro-F1 of 0.74. This result shows that textual context contains useful evidence for KOGL compliance risk prediction, especially for BY and NC cases. However, the text-only model had limitations in detecting visual modification and ownership inconsistency because it did not use image-level evidence.

The image-only model showed lower performance than the text-only model. This result indicates that visual features alone are insufficient for understanding license conditions, attribution requirements, and commercial context. However, image features are still important for identifying ND and CF risks, where modification traces, watermarks, and third-party visual elements may appear.

The OCR + text model improved performance compared with the text-only model. This improvement suggests that OCR-extracted image text provides useful additional evidence, such as embedded source marks, subtitles, advertising phrases, and watermarks. In particular, OCR information can help detect cases where important compliance-related text is included inside images rather than in HTML text.

The simple fusion model achieved better performance than single-modal models by combining text and image features. However, its performance was still lower than that of the proposed approach. This result suggests that direct concatenation alone is not sufficient to fully exploit heterogeneous web evidence. In contrast, the proposed approach uses feature reduction and sequential context modeling to capture the relationship among evidence elements.

Overall, the experimental results show that KOGL compliance risk prediction benefits from multimodal evidence integration. The performance improvement is especially meaningful because KOGL risks often arise from inconsistencies among different web elements, such as metadata, page text, OCR text, and image content.

5 Conclusion

This paper presented a multimodal evidence-based approach for KOGL license compliance risk prediction in web environments. The main objective was to support scalable compliance review of open government works reused on the web, where license-related evidence is often distributed across textual descriptions, captions, metadata, image-embedded text, and visual content. Instead of treating KOGL compliance verification as a simple text classification task, this study formulated it as a web engineering problem involving heterogeneous web evidence.

The proposed approach integrates textual context, KOGL metadata, OCR-extracted image text, and visual features to predict four types of compliance risk: attribution risk, non-commercial use risk, no-derivatives risk, and false KOGL claim risk. A feature reduction process was used to reduce noisy and redundant information, and a lightweight sequential context model was applied to capture relationships among different evidence elements. The final output provides the predicted risk type, risk probability, and supporting evidence for human review.

The experimental results showed that the proposed approach outperformed rule-based, text-only, image-only, OCR + text, and simple fusion baselines. In particular, the proposed approach achieved an accuracy of 0.91, macro-F1 of 0.89, and AUC of 0.93. These results indicate that multimodal web evidence is effective for improving KOGL compliance risk prediction. The comparison also showed that OCR-extracted text and visual features provide useful evidence that cannot be fully captured by textual context alone.

The proposed approach is not intended to replace legal judgment. Rather, it provides compliance-oriented decision support by identifying high-risk web content and presenting interpretable evidence. This can help public content managers, web administrators, and open data curators prioritize content that requires further review. The approach can also be useful for reviewing AI training data and reused public content in web-based content management environments.

Future work will extend the proposed approach to additional media types, such as video and audio, where KOGL compliance evidence may appear in subtitles, speech, background images, or temporal editing patterns. Another direction is to incorporate provenance tracking to verify the origin and reuse history of public works. In addition, further evaluation in real-world web-based review environments will be conducted to examine scalability, latency, and practical usability.

Acknowledgment

This study was conducted as part of the 2025 Global Copyright Issue Rapid Response (R&D) project of the Ministry of Culture, Sports and Tourism and the Korea Institute of Culture Technology Evaluation and Planning (KCTEP) (Project Name: Development of Content Analysis and Type Information Determination Technology for Global Expansion of Shared Works, Project Number: RS-2025-02305397, Rate 100%).

References

- [1] Korea Open Government License, “Use Terms Guide for Korea Open Government License,” KOGL Official Guide. Available: <https://www.kogl.or.kr/info/license.do>, Accessed: May 17, 2026.
- [2] Korea Culture Information Service Agency, “A Study on the Improvement of the Open License System for Public Works,” Research Report, 2022. Available: <https://www.kcisa.kr/kr/board/research/boardView.do?bbsIdx=11200>.
- [3] Korea Culture Information Service Agency, “A Study on Establishing Policies for Promoting the Use of Public Works,” Research Report, 2015. Available: <https://www.kcisa.kr/kr/board/research/boardView.do?bbsIdx=3650>.
- [4] J. H. Hong, “A study on copyright limitation for promoting the use of public works,” *Journal of Korean Library and Information Science Society*, vol. 44, no. 2, pp. 315–343, 2013.
- [5] H. K. Kim, “Legal issue on the use and protection of copyrights in government works,” *Sungkyunkwan Law Review*, vol. 25, no. 1, pp. 195–226, 2013.
- [6] H. R. Kim, “A study on the free use system of public works,” *Journal of Korea Information Law*, vol. 19, no. 2, pp. 29–87, 2015.
- [7] Y. G. Park and C. D. Kim, “Legal review of the issue regarding the free use of public works,” *Copyright Quarterly*, vol. 28, no. 1, pp. 31–64, 2015.
- [8] G. S. Kim and Y. G. Park, “Improvement measures for promoting the free use of public works,” *Advanced Commercial Law Review*, no. 76, pp. 163–186, 2016.
- [9] J. H. Yoo, “A study on the provision and use of open data containing copyrighted works: focusing on the open data act and the copyright act,” *Advanced Commercial Law Review*, no. 89, pp. 61–88, 2020.

- [10] H. D. Seo and K. H. Joung, “A study on the management of copyright information for increasing the free use of government publications as public works,” *Journal of Korean Society of Archives and Records Management*, vol. 22, no. 3, pp. 125–141, 2022, DOI: <https://doi.org/10.14404/JKSARM.2022.22.3.125>.
- [11] Creative Commons, “Attribution-NonCommercial-NoDerivatives 4.0 International Public License,” Creative Commons Legal Code, 2013. Available: <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>, Accessed: May 17, 2026.
- [12] W3C, “PROV-DM: The PROV Data Model,” W3C Recommendation, Apr. 30, 2013. Available: <https://www.w3.org/TR/prov-dm/>.
- [13] Schema.org, “Schema.org Property: license,” Schema.org Vocabulary. Available: <https://schema.org/license>, Accessed: May 17, 2026.
- [14] T. Baltrušaitis, C. Ahuja, and L. P. Morency, “Multimodal machine learning: a survey and taxonomy,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 423–443, 2019, DOI: <https://doi.org/10.1109/TPAMI.2018.2798607>.
- [15] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning phrase representations using RNN encoder–decoder for statistical machine translation,” in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1724–1734, 2014, DOI: <https://doi.org/10.3115/v1/D14-1179>.
- [16] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 4171–4186, 2019, DOI: <https://doi.org/10.18653/v1/N19-1423>.
- [17] R. Smith, “An overview of the Tesseract OCR engine,” in *Proceedings of the Ninth International Conference on Document Analysis and Recognition*, vol. 2, pp. 629–633, 2007, DOI: <https://doi.org/10.1109/ICDAR.2007.4376991>.
- [18] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, “An image is worth 16x16 words: transformers for image recognition at scale,” in *Proceedings of the International Conference on Learning Representations*, 2021.
- [19] K. Wu, J. Peng, H. Chen, M. Fu, and H. Chao, “TinyViT: fast pre-training distillation for small vision transformers,” in *Proceedings of the*

European Conference on Computer Vision, Lecture Notes in Computer Science, vol. 13681, pp. 68–85, 2022, DOI: https://doi.org/10.1007/978-3-031-19803-8_5.

- [20] M. T. Ribeiro, S. Singh, and C. Guestrin, ““Why should i trust you?”: explaining the predictions of any classifier,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135–1144, 2016, DOI: <https://doi.org/10.1145/2939672.2939778>.

Biographies



Seyoung Jang received his bachelor’s degree in 2018, received his master’s degree in computer engineering from Soongsil University in 2021, and has been conducting his PhD in computer engineering from 2023 to the present. His research interests include copyright protection and utilization activation.



Byeongchan Park received a bachelor’s degree in computer engineering through the Academic Credit Bank System in 2015, and a master’s and doctor of philosophy degrees in computer science and engineering from Soongsil

University in 2018 and 2023, respectively. He is currently working as a visiting professor at the Department of Computer Science, Soongsil University. His research areas include copyright technology and the promotion of its utilization.



Youngmo Kim received B.S., M.S., and Ph.D. degrees in computer engineering from Daejeon University in 2003, 2005, and 2011, respectively. Since 2012, he has been a professor at Soongsil University. His main research interests include copyright protection and the promotion of its utilization.



Sun-Jib Kim is currently working as a professor at the School of IT, Hansei University. His research areas include information security, the Internet of Things (IoT), cloud computing, and AI system authentication.

