Abstract

The Study on Establishing a Symbiotic Relationship: Protecting Creators in Generative AI Training

Yusun Park*

As AI technology advances and AI-based content becomes more common, there is an ongoing discussion about the need for balanced profit distribution between copyright holders and AI developers. Without proper rights attribution and compensation mechanisms for the creators of works used in AI training, the value of their creative efforts may be undervalued, potentially discouraging them from creating new works and negatively impacting the AI industry's growth. To ensure the sustainability and growth of both human creativity and AI technology, a fair compensation system for copyright holders is essential.

The secrecy in AI model training and content generation, often called the 'black box' problem, makes it difficult to identify the data used for AI training. This makes it challenging for copyright holders to prove substantial similarity in generative AI-related copyright infringement cases without access to the training data. Disclosing AI training data or using metadata to identify works and track their usage can help determine copyright infringement and support compensating copyright holders.

This paper examines the transparency in AI data training by disclosing training data in generative AI copyright infringement cases and various measures to balance the interests of AI technology and copyright holders, such as an opt-out mechanism for copyright holders to reserve their rights, the use of metadata to manage AI

* Professor, Kangwon National University

training data, and the establishment of a centralized platform, like a copyright repository or clearing house, for copyright management.

Keywords

AI Training Data, EU AI Act, AI Copyright Disclosure Act, Metadata, Opt-out, Copyright Reservation, Data Repository, Clearing House