Abstract

Is Data Training for Artificial Intelligence Fair Use and Legitimate Interest?

- A Comparative Study of the Double Hurdles in Copyright and Personal Data Protection

Woo Jung JON*, So Won Jung**

This article conducts a comparative analysis of the double hurdle presented by copyright law and personal data protection law concerning the use of training data for generative artificial intelligence (AI). Specifically, it compares the requirements of 'fair use' under copyright law and 'legitimate interests' under personal data protection law. These are exception provisions that become crucial in circumstances where obtaining consent from rights holders is practically infeasible.

The two exception provisions possess distinct philosophical foundations and analytical frameworks. Fair use under copyright law seeks to balance the protection of creators with the advancement of culture, and it involves a holistic consideration of four factors: the purpose and character of the use, the nature of the copyrighted work, the amount of the portion used, and the effect of the use upon the potential market. Conversely, the 'legitimate interests' provision in data protection law is grounded in the protection of personality rights as a fundamental right. It employs a three-step assessment—a legitimacy test for the purpose, a necessity test, and a balancing test—which requires that the interests of the data controller manifestly override the rights of the data subject.

* Korea Advanced Institute of Science and Technology, Graduate School of the Future Strategy, Master Program of Intellectual Property, Associate Professor

** Korea National Open University, Department of Media and Visual Studies, Lecturer

This article finds that the two exception provisions exhibit a significant functional correlation within the context of AI data training. The 'purpose and character of the use' factor in fair use corresponds to the legitimacy of purpose; the 'amount and substantiality of the portion used' aligns with necessity (the principle of minimum infringement); and the 'effect on the market' parallels the concept of balance (a comparative weighing of interests). Furthermore, the 'nature of the copyrighted work' can be analyzed in comparison to the assessment of data sensitivity.

South Korea's Personal Information Protection Commission (PIPC) has indicated the possibility of using publicly available personal data for AI development on the grounds of 'legitimate interests', while simultaneously requiring the implementation of technical safeguards. To resolve the legal uncertainty surrounding copyright, it is necessary to consider the introduction of a statutory text and data mining (TDM) exemption.

Keywords

Generative AI, Copyright, Personal Data Protection, Fair Use, Legitimate Interest