CHALLENGING THE "FAIR USE" EXCEPTION: A CRITICAL EVALUATION OF TRANSNATIONAL AI SCRAPING WITHIN SINHALA AND TAMIL LINGUISTIC DATABASES

Dona Nethmi Yashodya Weerakoon

Faculty of Law, University of Colombo, Sri Lanka

Abstract

Transnational artificial intelligence companies routinely execute automated scraping of regional web repositories to train large language models, exposing low-resource Sinhala and Tamil linguistic databases to uncompensated structural extraction. Challenging the 'fair use' exception traditionally invoked by these entities, this paper provides a critical evaluation of the statutory omissions within the Sri Lankan Intellectual Property Act No. 36 of 2003. The primary objective is to analyse how the open-ended, four-factor balancing test under Section 11 fails to govern automated data ingestion, permitting developers to claim domestic research immunities for mass commercial scraping operations. Utilizing a comparative doctrinal methodology, the research directly contrasts Sri Lanka's exposure against the legal frameworks of India and Singapore. It highlights that while the Delhi High Court's landmark 24th July 2026 interim ruling in *ANI Media Pvt Ltd v OpenAI* interpreted AI training as a form of research under Section 52(1)(a)(i) of Indian Copyright Act, 1957, Singapore's statutory model under Sections 243 and 244 of the Singapore Copyright Act 2021 explicitly immunizes computational data analysis but conditions it strictly on "lawful access." This evaluation demonstrates that technical tools like robots.txt lack statutory binding power. This leaves local creators without a civil cause of action against offshore, cloud-hosted training networks. To mitigate these extraterritorial enforcement limitations, this paper outlines a practical blueprint for a domestic statutory collective licensing model administered by a state-backed clearinghouse. This satisfies the international "three-step test" under Article 9(2) of Berne Convention and Article 13 of TRIPS Agreement by isolating mass commercial ingestion without impairing standard human exploitation. Finally, the paper defines an enforceable, three-part cross-border compliance mechanism. It conditions local digital market access on licensing adherence, mandates domestic corporate representation for foreign tech entities, and levies targeted digital service taxes on non-compliant developers. Thus, it safeguards indigenous cultural property while maintaining technological integration.

Keywords: Data Sovereignty, Text and Data Mining, Fair Use Doctrine, Linguistic Equity, Collective Licensing