
BEYOND THE SCOREBOARD: PROTECTING GENERIC VS. SPORTS AI-GENERATED DATA UNDER INDIAN IP LAW

Neilotpla Palak & Akshat Pugalia, LL.B. (Hons.), National Law University Odisha

ABSTRACT

Artificial intelligence now plays a growing role in generating and processing sports data, from real-time analytics to predictive modelling and biomechanical tracking. Indian intellectual property law was framed around human authorship and remains ill-equipped to address content created or compiled by non-human agents. This article examines Indian copyright law as they apply to AI-generated data, using sports data as a lens to explore the tension between generic and sport-specific challenges. It begins by analysing the human authorship requirement under section 2(d) of the copyright act, 1957, and the modicum of creativity, standard established in *Eastern Book Company V D.B. Modak*, both of which limit copyright protection for ai-generated databases and outputs. The article then argues that sports data presents a distinct and more acute problem. It is built on uncopyrightable facts (*publici juris*), lacks any *sui generis* database right as exists in the European Union, and implicates athletes' personality and publicity rights when biomechanical or predictive models are trained on individual player data. In the absence of adequate statutory protection, the article demonstrates how broadcasting rights under Section 37 of the Copyright Act, contractual restrictions, and licensing arrangements rather than copyright itself have emerged as the principal tools for safeguarding commercial interests in sports data, as illustrated by *Star India Pvt Ltd V Piyush Agarwal*. The article concludes that until Indian law evolves to address AI-generated works and database rights specifically, and how contractual exclusivity will remain the most effective mechanism for protecting investment in sports-tech innovation.

INTRODUCTION

Artificial intelligence is playing an increasingly prominent role in the collection, processing, and commercialization of sports data, ranging from game statistics to predictive analytics and biomechanics. Indian law, however, was born in an age when human authorship was the norm, and its copyright and related laws have been ill-equipped to address scenarios wherein a disembodied entity such as ai is at the helm of generation or compilation of data. This article engages with various lacunae and intersections in Indian copyright law concerning AI-generated works by focusing on the area of sports data. By examining issues of copyrightability, database rights, broadcasting, and personality rights, this article attempts to bring out the multifaceted nature of legal challenges posed by ai-driven data creation while focusing on the confluence of uncopyrightable facts, commercial investment, and athletes' personality rights in sports data.

THE HUMAN AUTHORSHIP BARRIER

the copyright act, 1957 was drafted specifically for human authored work. Section 2(d)¹ of the act defines author in terms that assume a human agent is behind a literary work. The person who creates it, for a computer-generated work, the person who causes the work to be created. This phrase was intended to cover a human using a tool like a photographer using a camera, not a scenario where a generative artificial intelligence system independently produces the output with minimal to no human input. When a large language model or image generator produces text, code, or synthetic data with only a prompt as human contribution, it becomes legally uncertain whether any author in conventional terms exists at all and without an author, there is no copyright.

Where the artificial intelligence (ai) output takes the form of a database or compilation like a structured dataset of synthetic customer records, section 2(o)² treats such compilations as literary works. As the supreme court has established in its ruling in *Eastern Book Company V. D.B. Modak* (2008)³ this lenient standard protection is not satisfactory to establish originality. The judgement laid down that mere investment of intellectual labour in arranging the data for processing by computer cannot categorise it as literary work. There should be some element of

¹ Copyright Act, No. 14 of 1957, § 2(d) (India).

² Copyright Act, No. 14 of 1957, § 2(o) (India).

³ *Eastern Book Co. v. D.B. Modak* (2008) 1 S.C.C. 1 (India).

independent selection or arrangement which involves at least a modicum of creativity; mechanical sorting or collation, extensive as it may be, is unlikely to meet the test of originality.

This being the position, it is not surprising that attempts have been made to rely on other laws like trade secrets to protect the data bases which are fed into the ai engines. The information technology act, 2000, particularly sections 43A⁴ and 72A⁵, dealing with compensation for unauthorized access to and disclosure of confidential information and contractual restrictions, particularly the end-user license agreements (EULAs) and application programming interfaces (APIs), play a significant role in safeguarding the secrets of algorithms. It is the terms of these agreements, and not any copyright protection, which prohibit scraping or redistribution of the outputs.

THE DISTINCTION: WHY AI SPORTS DATA IS LEGALLY UNIQUE

Indian copyright law has never protected facts available to the public at large. A cricket score, a batsman's strike rate, the speed of a bowling machine are all considered to be *publici juris* in India, that is, not eligible for copyright protection, being mere facts. The problem with ai is that while it seeks to go beyond simple facts, it only does so in a way that is processed. A pitch map or win-probability curve or expected goals model are all forms of value-added information, but which do not constitute matters subject to copyright. Unlike a chatbot which is trained on publicly available information, sports ai is not independent of the game-data upon which it is trained. It is, after all, a predictive model which can only make predictions based on facts that already exist or are postulated to exist in the game (the ai cannot predict a six that is not hit, a save that is not made). So, while the surface of an ai sports graphic might look like something independently created, it always has a deep structure of facts that are not independently creatable or copyrightable.

The fundamental issue that the layer of ai adds on top of data that is publicly available is not enough to make it "original" for copyright purposes. A related issue is the complete lack of sui generis database rights in India. The database directive in the EU⁶ grants additional copyright-like protection for databases that required "investment" in their creation beyond normal copyright norms. This protection is not available in India, which means that the investment in

⁴ Information Technology Act, No. 21 of 2000, § 43A (India).

⁵ Information Technology Act, No. 21 of 2000, § 72A (India).

⁶ Council Directive 96/9/EC, of 11 March 1996 on the Legal Protection of Databases, art. 7, 1996 O.J. (L 77) 20.

AI-driven databases (in terms of time, money, and effort) is not protected either. Even if, for instance, a sports entity creates an investment-heavy ai system which tracks player movement in a certain manner, biomechanics of a limb, or something else completely unique to them, it cannot use copyright law to protect such a system. There is no “protection for facts” in India, for sports ai or otherwise. The law of protection for facts in India comes in the form of contract law, licensing, paywalls, API access, and other technical barriers to entry, not copyright protection.

RIGHTS OF PUBLICITY & PERSONALITY RIGHTS (ATHLETE NAME, IMAGE, LIKENESS)

This is where sports AI data is distinctly different from generic AI. While a common large language model (LLM) is unlikely to infringe upon personality rights, training a sports ai model on the biomechanics of an athlete or using its own ai to predict them inevitably crosses into the realm of personality rights.

Indian courts have long held that a person’s name, image, and other attributes cannot be used for commercial gain without permission, under the common law tort of passing off, as well as publicity rights. One such case is *Dm Entertainment V. Baby Gift House*.⁷ If an athlete’s name, image, voice, or other identifying features are used to create a biomechanical model of their predicted movements, it crosses into the realm of using their personality for commercial gain, which is a violation of their publicity rights, unless they have given permission for the same.

ENFORCEMENT MECHANISMS & COMMERCIAL REALITIES

Copyright affords little or no protection to data generated by ai systems, so another parallel legal regime has developed around sports data. This is based on the broadcasting rights and not on copyright in the original content. Section 37 of the copyright act⁸ provides that the assignee of broadcasting rights shall have the exclusive right to re-broadcast, re-show, record and communicate to the public any broadcast programme, independently of any copyright in the work itself, and such rights subsist for twenty-five years from the date of first broadcast. These rights have been repeatedly invoked, and have figured in numerous cases, to safeguard the data from unauthorized use and unauthorized re-broadcast. It is this regime which has proved to be

⁷ *DM Entm't Pvt. Ltd. v. Baby Gift House* CS (OS) No. 893/2002 (Del. H.C. Apr. 29, 2010).

⁸ Copyright Act, No. 14 of 1957, § 37 (India).

the most effective in combating data theft.

In *Star India Pvt Ltd v. Piyush Agarwal* (Delhi High court, 2012)⁹, the court restrained unauthorized sensational news dissemination of live cricket scores gathered through unauthorized means. Therefore, effectively treating rapid unlicensed extraction and redistribution of match data as an infringement of the broadcaster's rights and investment, even though the underlying scores themselves remained uncopyrightable facts. This case is frequently cited as India's closest analogue to a hot news misappropriation doctrine.

Where broadcast rights do not reach like in case where data is collected independently by third parties inside sports venue, organizations fall back on contract access restrictions. Stadium's ticketing terms include explicit no data scraping or no commercial data collection clauses. As a result, making unauthorized telemetry being categorised as a breach of contract and potentially a trespass rather than a copyright matter. Leagues such as in the Indian premier league (IPL), hosted by BCCI further formalizes this through exclusive data intermediary agreements, licensing official real-time statistics and analytical feeds to certain partners for, fantasy sports platforms or broadcasters. While contractually barring unlicensed scraping or redistribution by competitors. Altogether, these mechanisms show that Indian sports-tech does not rely on intellectual property law's substantive protections at all. Instead, exclusivity is manufactured commercially through broadcast rights, venue access conditions, and tightly drafted licensing agreements filling the vacuum left by the absence of copyright protection and the lack of a sui generis database right.

CONCLUSION

An exercise in redistribution of the areas of influence: where copyright was intended to protect, ai-generated sports data threatens to undermine. Left in a state of limbo, the Indian copyright law has failed to protect itself from the disruption, with rights of broadcasting and contract law filling in the perceived gaps more efficiently due to their very nature. This, in turn, should inform the debate about the protection of investments in the sphere of ai-generated sports analytics. The battle for IP rights for such works may well be an exercise in futility for the time being, while redistribution of rights between various stakeholders occurs, as well as the shift in competitive advantage to whomever possesses tighter control of the source of information,

⁹ *Star India Pvt. Ltd. v. Piyush Agarwal* CS (OS) No. 2722/2012 (Del. H.C. Nov. 8, 2012).

be it broadcast, stadium, or other. While database rights and rights of reproduction may well become more contested in the future as the technology of approximation of individual player's data from aggregated data improves, for the time being, personality rights and rights of use remain more relevant. In short, until the Indian law catches up with the innovation, be it through sui generis protections for databases or expanded interpretations of copyright to include reproduction of the sports statistics, the most efficient way to protect one's position in the market will be to acquire and enforce the most protective contracts possible.