

ISSN: 2583-8725

Lex Scripta Journal

Quarterly Online and Print Edition

Law & Policy

“Join the League of
National & International Scholars”



EDITORIAL TEAM

DR. AJAY BHUPENDRA JAISWAL

Professor & Former Head
Department of Law
V.S.S.D. College, Nawabganj,
(C.S.J.M. University, Kanpur)

DR. MEGHA OJHA

Associate Professor | Legal Consultant
| Author | KLEF College of Law

PROF. DR. DEEVANSHU SHRIVASTAVA

Founding Dean and Professor,
GL Bajaj Institute of Law,
Greater Noida

DR. GAURAV GUPTA

Assistant Professor,
Faculty of Law, Lucknow

MR. TUHIN MUKHARJEE

Leadership Strategist | Business Coach
| Author | Speaker

MR. PRAKARSH PANDEY

Author and
Advocate, Allahabad High Court

MR. AMARESH PATEL

Assistant Professor
at Law School,
Amity University, Patna



LEX SCRIPTA MAGAZINE OF LAW AND POLICY (VOL-2, ISSUE-3)

Copyright © 2025, LexScripta
ISSN-2583-8725
Vol - II, Issue - III
Published by INTEGRITY EDUCATION INDIA

New Delhi

First Floor, 4598/12-B, 1st Floor,
Padam Chand Marg, Daryaganj,
New Delhi, Delhi 110002
Phone: +91 98 11 66 62 16 (M)
Phone: +91 70 11 60 56 18 (M)

Bengaluru

Jallahalli East
Bengaluru, Karnataka. India.
Phone: +91 98 11 66 62 16 (M)
Email: publisher.integrity@gmail.com

USA

New Jersey
14 Grandview Ave, Upper Saddle River,
NJ-07458, USA
Phone: +14805226504 (M)

London

37 Degree Media
64, Hodder Drive, Perivale, London UB68LL.
United Kingdom.
Phone: +44 7950 78 18 17 (M)
Website: integrityeducation.co.in

© Lex Scripta Magazine Of Law And Policy, 2025

Disclaimer

All Copyrights are reserved with the Authors. But, however, the Authors have granted to the Journal (Lex Scripta Magazine of Law and Policy), an irrevocable, non-exclusive, royalty-free and transferable license to publish, reproduce, store, transmit, display and distribute it in the Journal or books or in any form and all other media, retrieval systems and other formats now or hereafter known. No part of this publication may be reproduced, stored, distributed, or transmitted in any form or by any means, including photocopying, recording, or other electronic or mechanical methods, without the prior permission of the publisher, except in the case of brief quotations embodied in critical reviews and certain other non-commercial uses permitted by copyright law.

The Editorial Team of Lex Scripta Magazine of Law and Policy Issues holds the copyright to all articles contributed to this publication. The views expressed in this publication are purely personal opinions of the authors and do not necessarily reflect the views of the Editorial Team of Lex Scripta Magazine of Law and Policy.

[© Lex Scripta Magazine of Law and Policy. Any unauthorized use, circulation or reproduction shall attract suitable action under application law.]

For any Query / Feedback
Phone: +91 98 11 66 62 16 (Vineet Sharma)

Printed in India @ New Delhi

ISSN: 2583-8725

Lex Scripta Journal

Quarterly Online and Print Edition

Law & Policy

**"Join the League of National
and International Scholars"**



Lex Scripta Journal

COPYRIGHT CHALLENGES IN ARCHIVING AND TRAINING AI ON EPHEMERAL CONTENT: A LEGAL ANALYSIS OF STORIES, LIVE STREAMS, AND DISAPPEARING MEDIA.

Author

Vishnu Prabhakar Chandani

Rahul Singh



COPYRIGHT CHALLENGES IN ARCHIVING AND TRAINING AI ON EPHEMERAL CONTENT: A LEGAL ANALYSIS OF STORIES, LIVE STREAMS, AND DISAPPEARING MEDIA

Vishnu Prabhakar Chandani,

LLM Student,

IMS UNISON UNIVERSITY,

DEHRADUN, UTTRAKHAND Batch- 2025-26.

Rahul Singh,

Assistant Professor,

School of law,

IMS Unison University Dehradun.

ABSTRACT

Ephemeral content including Stories, Snaps, temporary livestreams, self-destruct messages, etc. are now a staple of digital communication on social media platforms. Although users feel like this content is temporary and private, platforms regularly keep copies of it for moderation, analytics and algorithmic purposes. Simultaneously, artificial intelligence systems have become more and more dependent on the usage of large-scale datasets that can contain such disappearing media, raising complex questions that surround copyright, user consent and data governance.

This research seeks to explore whether the archiving of ephemeral content and its use to train AI is copyright infringement under Indian law. Through a combination of doctrinal analysis, comparative examination of the EU and US frameworks, & content analysis of platform governance, the study shows that ephemeral content meets both of the requirements of fixation and originality and is therefore given full copyright protection. Because the Copyright Act, 1957 does not explicitly provide for data and text mining exceptions, the reproduction of works which are ephemeral for training of Artificial Intelligence is not a legally permissible use.

The results show major gaps in statutory protections, insufficient disclosures in platform Terms of Service and a lack of any international doctrine relating to ephemeral digital works. The paper

concludes that using AI on ephemeral content without explicit consent from users is an infringement and to recommend statutory recognition of "ephemeral digital works", mandatory opt-in consent for AI datasets and transparency requirements for AI developers and data retention practices for platforms. These reforms are necessary to bring copyright law in line with the present day digital behaviour, protect user autonomy, and ensure ethical development of AI.

Introduction

Ephemeral content - such as Instagram Stories, WhatsApp Status updates, Snapchat Snaps, YouTube Live streams, and the like, disappearing media has altered the nature of digital communication in the modern world.¹ Users gravitate towards these formats because they create a perception of spontaneity, reduced permanence and increased privacy because of their temporary visibility. The design of these platforms encourages the expectation that once the amount of time has been used up, the content becomes inaccessible or deleted.²

However, the reality of technology is different. Even though the content vanishes from the user interface, platforms often keep copies of it for moderation, analytics, algorithmic training and sometimes for undisclosed internal processes³. In parallel, artificial intelligence developers are using automated tools and third-party scrapers increasingly to gather ephemeral content and use them in machine learning data sets.⁴ This destroys the basic user expectation that such content is not intended to live beyond its intended life.

From a copyright point of view, ephemeral content is a special challenge. Copyright law traditionally assumes works to be stable, fixed and enduring. Ephemeral content is, on the contrary, defined by the intentionality of its temporality. Yet the instant a user produces and posts an Instagram Story or Snapchat Snap, the work is fixed on the servers of the platform - albeit for a short time - and thus ready for copyright protection.⁵ This presents a tension between the user intention, the control of the platform

¹ Instagram Help Center, *How Do Instagram Stories Work?* (2023), <https://help.instagram.com/1660923094227526>.

² Snapchat, *Support: How Snaps Are Deleted* (2024), <https://support.snapchat.com/en-US/a/when-are-snaps-chats-deleted>.

³ Meta Transparency Center, *Data Policy* (Updated 2024), <https://transparency.meta.com/policies/data-policy/>.

⁴ Electronic Frontier Foundation, *AI Training Data and User Privacy* (Sept. 2023), <https://www.eff.org/deeplinks/2023/09/how-ai-training-data-works>.

⁵ Copyright Act, 1957, § 13 (India), <https://www.indiacode.nic.in/bitstream/123456789/2066/1/A1957-14.pdf>.

and the doctrine of copyright, particularly when this content is reproduced, archived or used to train AI models without the user's knowledge.

This introduction contextualises the main focus of this research which is to examine whether archiving and using ephemeral content for training AI models is liable for copyright infringement given the current legal frameworks, user expectations and technological practices. Because Indian Copyright law does not currently have explicit rules on ephemeral digital works and AI datasets training, this problem reveals a large gap in regulation. The need for doctrinal clarity and updated policy responses therefore becomes necessary in addressing the intersection between disappearing media and AI development.

Research Problem Statement

Ephemeral content is broadly seen by users as temporary, private and self-deleting. Platforms deliberately promote disappearing media - like Stories, Snaps and temporary livestreams - in order to foster the assumption that such posts will disappear indefinitely once their visibility window ends.⁶ However, platform data policies indicate that copies of this content can persist on backend servers for significantly longer periods of time, often for moderation, security reviews, analytics and machine learning purposes.⁷

This lack of correlation between user expectation and platform retention practice opens a legal grey area, especially in the domain of copyright. Despite the fact that it is temporary, ephemeral content meets the threshold of fixation for the minimum duration of its existence the moment it is uploaded to a server and therefore provides the user with full copyright protection under Indian law.⁸ Still, users are rarely aware that their vanishing content can be archived, reproduced or repurposed by platforms or third parties.

The rise of generative AI adds to this uncertainty. AI developers often use large-scale datasets, some of which are scraped, automatically downloaded, or obtained via platform APIs, which could contain ephemeral posts.⁹ Because these datasets are frequently not transparent, and come from mixed sources, users cannot determine if their temporary content has been taken and used to train machine learning models. Indian copyright law does not have any specific provisions dealing with (a) AI training, (b) the legality of creating a dataset, or (c) the special status of ephemeral digital works.¹⁰

This regulatory vacuum is a pressing research problem: whether the archiving and use of ephemeral content for AI model training is a form of copyright infringement, a violation of norms of user consent, or something else covered by any exception to copyright law. The situation is exacerbated by the fact that not a single jurisdiction in the world has conceptualised a clear legal framework for ephemeral

⁶ Snapchat, *Snapchat Features: Stories*, <https://www.snapchat.com/features> (last visited Nov. 27, 2025).

⁷ Meta Transparency Center, *Data Policy* (Updated 2024), <https://transparency.meta.com/policies/data-policy/>.

⁸ Copyright Act, 1957, § 13 (India), <https://www.indiacode.nic.in/bitstream/123456789/2066/1/A1957-14.pdf>.

⁹ Electronic Frontier Foundation, *How AI Training Data Works* (2023), <https://www.eff.org/deeplinks/2023/09/how-aitraining-data-works>.

¹⁰ Ministry of Electronics & Information Technology (MEITY), *IndiaAI Report 2023*, <https://www.meity.gov.in/content/indiaai>.

content in the context of AI yet, and users as well as developers are left to navigate through an unclear and fast-changing legal landscape.

Research Objectives

The explosive growth of ephemeral content across all social media platforms in parallel with the growing dependency of artificial intelligence models on huge datasets of user-generated content, makes the issue of copyright, technology and user autonomy a complex interplay of overlapping factors. Existing copyright statutes--or rather the absence of them--is an issue that is especially relevant when it comes to what should be done with disappearing digital works when they are reproduced, archived, or used as a dataset for AI training.¹¹ Similarly, global regulatory initiatives, such as those by the European Union and WIPO, have not yet developed clear doctrines for the regulation of ephemeral media and machine learning uses.¹²

In light of this context, the research wishes to accomplish the following purposes:

- 1. To discuss the status of ephemeral content as per the Indian law.**

This includes looking at whether temporary or disappearing media comply with the statutory requirements of fixation and originality and whether any exceptions are applicable.¹³

- 2. To decide if archiving ephemeral content - by platforms or third parties - is an infringement of copyright.**

This objective focuses on the issue of backend server retention, scraping or collecting data sets with unauthorized reproduction or communication to the public.¹⁴

- 3. To determine the legal validity of training A.I models on archived ephemeral content.** This includes assessment as to whether such use may be considered to fall under fair dealing, text and data mining exceptions (where applicable) or requires express authorization.

¹¹ Copyright Act, 1957, § 13 (India), <https://www.indiacode.nic.in/bitstream/123456789/2066/1/A1957-14.pdf>.

¹² World Intellectual Property Organization, *WIPO Conversation on Intellectual Property and Frontier Technologies* (2024), https://www.wipo.int/about-ip/en/frontier_technologies/.

¹³ Indian Ministry of Human Resource Development, *Report of the Committee on Intellectual Property Rights* (2020), https://www.education.gov.in/sites/upload_files/mhrd/files/document-reports/IPR.pdf.

¹⁴ Electronic Frontier Foundation, *Scraping and Machine Learning: Legal Considerations* (2023), <https://www.eff.org/issues/scraping>.

4. **To evaluate if the current forms of consent that are provided by the platforms are sufficiently addressing the rights of the users in ephemeral content.**

Most platforms use broad, bundled, non-negotiable licenses that users agree to when creating an account raising issues of voluntariness and clarity of consent.¹⁵

5. **To make a suggestion for regulatory and policy framework to fill the gaps in Indian copyright law in relation to ephemeral content and AI dataset.**

This includes advocating for changes in the statute, clearer guidelines on consent, as well as transparency requirements for AI developers.

¹⁵ Meta Transparency Center, *Terms of Service and License Grant* (2024), <https://transparency.meta.com/policies/termsof-service/>.

Research Hypothesis

The growing use of ephemeral content in artificial intelligence training data sets raises open questions on authorship, user consent, fixation and legality of reproducing disappearing media. Indian law of copyright offers wide protection to original works the moment they are fixed in a tangible medium, and this includes digital servers¹⁶. However, the law makes no distinction between works that were intentionally intended to be temporary and permanent digital works. This lack of differentiation, coupled with the opaqueness of platform data retention policies and AI training practices, leaves a critical uncertainty about whether such uses are an infringement or lie within permissible exceptions.¹⁷ In light of these doctrinal ambiguities, and the practice of technology, the following hypotheses are offered in this research:

- **Primary Hypothesis (H1)**

Archiving and training AI using ephemeral content without explicit and informed user consent amounts to copyright infringement, since ephemeral content is a protected work under the Copyright Act, 1957 and there is no current exception which allows for such reproduction or use.

This hypothesis is based on the principles that:

1. ephemeral content is fixed on a server before disappearing which will satisfy the statutory fixation requirement;¹⁷
2. unauthorized reproduction--including, but not limited to, retention, scraping, or creating a dataset--is included in the definition of infringement under Section 51 of the Copyright Act;¹⁸ and

¹⁶ Copyright Act, 1957, § 13 (India), <https://www.indiacode.nic.in/bitstream/123456789/2066/1/A1957-14.pdf>.

¹⁷ Mozilla Foundation, *AI Training Transparency Report* (2024), <https://foundation.mozilla.org/en/research/aitransparency/>.

¹⁷ World Intellectual Property Organization, *Understanding Copyright in Digital Content* (2023), https://www.wipo.int/copyright/en/faq_copyright.html.

¹⁸ Copyright Act, 1957, § 51 (India), <https://www.indiacode.nic.in/bitstream/123456789/2066/1/A1957-14.pdf>.

3. Indian law has no equivalent to text-and-data mining exception to the one found in jurisdictions such as the European Union¹⁹.

- **Null Hypothesis (H0)**

Ephemeral content, because of the temporary nature of it and the platform controlled environment it exists in, might fall under implied consent or find permissible uses such as fair dealing, thus not falling under the category of infringement if used for AI training.

The null hypothesis is based on arguments that:

1. implicit authorization by the user to process ephemeral content at the platform level by accepting broad Terms of Service;²¹
2. AI training may be transformative under certain interpretations of international fair use doctrines,²⁰ and
3. certain jurisdictions - notably the United States - have experienced judicial receptiveness to machine learning training uses.²³

Together, the hypotheses form a structured framework for assessing the doctrinal, comparative, and policy dimensions of ephemeral content in copyright law.

¹⁹ European Union, Directive (EU) 2019/790, Art. 3 (Text and Data Mining), <https://eur-lex.europa.eu/eli/dir/2019/790/oj>.

²¹ Meta Transparency Center, *Terms of Service: License to Content* (2024), <https://transparency.meta.com/policies/terms-of-service/>.

²⁰ U.S. Copyright Office, *Policy Statement on AI and Copyright* (2023), <https://www.copyright.gov/ai/>.

²³ *Authors Guild v. Google, Inc.*, 804 F.3d 202 (2d Cir. 2015), <https://law.justia.com/cases/federal/appellatecourts/ca2/13-4829/13-4829-2015-10-16.html>.

Literature Review

Academic scholarship on copyright and digital media has had a historical focus on stable and permanently accessible forms of user generated content. Early works focus on the protection of online photographs, videos, posts and audiovisual creations in social media platforms, such as Instagram, Snap Inc. and Meta Platforms.²¹ These studies affirm repeatedly that digital content becomes copyrighted once it is fixed, even if the time it takes to fix it is very brief.

Recent literature has been done around the practice of machine learning as well as the use of copyrighted materials to train large-scale AI systems. Reports from advocacy bodies and independent researchers show that AI developers often use scraped datasets from platforms all around the internet without knowledge or consent of the users²². Scholarly debate around AI training has centered much on fair use and text and data mining (TDM) exceptions as well as the legal ramifications of data set opacity. However, these discussions focus almost exclusively on permanent or content that can be found publicly, so ephemeral media is largely ignored.

Significantly, the dominant contemporary legal commentary on TDM, particularly in the European Union, acknowledges newly developing exceptions for the use of machine learning under Articles 3 and 4 of DSM Directive²³. Yet the Directive does not explicitly refer to vanishing media, nor does it consider the fact that content is content that users seek to remove or restrict temporally. In contrast, scholarship within the United States is largely based on analyses of fair use interpretations, drawing on a large number of landmark decisions such as the Authors Guild v. Google ruling, which stressed

²¹ Instagram Help Center, *How Instagram Stories Work* (2023), <https://help.instagram.com/1660923094227526>.

²² Electronic Frontier Foundation, *How AI Training Data Works* (Sept. 2023), <https://www.eff.org/deeplinks/2023/09/how-ai-training-data-works>.

²³ Directive (EU) 2019/790 of the European Parliament and of the Council, arts. 3–4, <https://eurlex.europa.eu/eli/dir/2019/790/oj>.

on the transformative character of search indexing and data analysis.²⁴ Again, these analyses assume the use of stable works and not intentionally temporary communications.

Within Indian legal literature, the research is also scant. Commentaries on the Copyright Act, 1957, address extensively the issue of fixation, authorship, and infringement, but none of them pays particular attention to the status of self-deleting digital works and their reuse by AI systems.²⁵ Government and policy reports by think tanks on AI development in India, including publications mentioning AI governance under national digital strategies, usually recognise the issue of datasets, but not disappearing content as a category in itself.²⁶

The lack of concentrated research in ephemeral content shows there is a significant gap. Despite the explosion of disappearing media on platforms like YouTube (for temporary live streams) and Telegram (self-destruct media) etc., scholarly analysis has not yet been able to systematically address the question of how disappearing media should be treated under copyright law when it is archived, scraped, or repurposed for training AI models. This gap highlights the importance of special research on ephemeral media as a special class of digital works, and doctrinal clarity on intersection with the emerging AI technologies.

²⁴ *Authors Guild v. Google, Inc.*, 804 F.3d 202 (2d Cir. 2015), <https://law.justia.com/cases/federal/appellatecourts/ca2/13-4829/13-4829-2015-10-16.html>.

²⁵ World Intellectual Property Organization, *Understanding Copyright in Digital Content* (2023), https://www.wipo.int/copyright/en/faq_copyright.html.

²⁶ Ministry of Electronics & Information Technology, *IndiaAI Report 2023*, <https://www.meity.gov.in/content/indiaai>.

Legal Framework

The legal treatment of ephemeral content is at the nexus of the doctrines of copyright, platform governance and new AI regulation. Because existing copyright laws were written in an age of fixed, tangible media, they do not speak directly to the specific issues of vanishing digital media. To gain a proper understanding of the legal position of ephemeral content and its application in AI training, it is crucial to discuss (1) the Indian copyright law, (2) the terms and licensing structures of the platforms, and (3) the international approaches, especially the EU and the United States. In recent years, institutions like the European Union and the Government of India have opened up discussion on AI governance, but not one of the either jurisdiction has stated a set of protections for ephemeral content. Similarly, global bodies of intellectual property, such as the World Intellectual Property Organization, have recognized the problems of data created by AI systems, but have not outlined clear rules for disappearing media.²⁷

Indian Copyright Law

Indian law on copyright offers protection to original literary, artistic, musical, and cinematographic works from fixation. Under Section 13 of the Copyright Act, 1957, the moment a work is "expressed in a tangible form", it is protected, and this includes digital storage on a server - even if only for a short period of time.²⁸ Ephemeral content, therefore, meets the fixation requirement the moment they are uploaded on any platform.

Section 14 sets the exclusive rights of the author such as the right to reproduce and communicate the work to the public. Where the platforms retain, archive or use the ephemeral content for analytics or AI training, such acts may amount to unauthorized reproduction under Section 51 unless an exception applies²⁹.

India presently does not have any specific exception for text and data mining (TDM). Section 52 has provisions for fair dealing, but only for private use, criticism, review, reporting and certain educational

²⁷ World Intellectual Property Organization, *WIPO Conversation on IP and Frontier Technologies* (2024), https://www.wipo.int/about-ip/en/frontier_technologies/.

²⁸ Copyright Act, 1957, § 13 (India), <https://www.indiacode.nic.in/bitstream/123456789/2066/1/A1957-14.pdf>.

²⁹ Copyright Act, 1957, §§ 14, 51 (India), <https://www.indiacode.nic.in/bitstream/123456789/2066/1/A1957-14.pdf>.

purposes. None of these provisions explicitly cover uses of machine learning, so the question of whether it is legal to train an AI on copyrighted works, including ephemeral works, is currently open.

Platform Governance Contractual Licensing

Ephemeral content is not only under the control of statutory copyright laws, but also platform Terms of Service (TOS). Many platforms force their users to provide broad, worldwide, royalty-free licenses when uploading anything, including disappearing posts. These licenses often have the retention, reproduction, modification, and internal use rights, for purposes that include content moderation, improvement of services, and algorithmic development.³⁰

However, from a legal point of view, TOS-based licenses have two major concerns:

- 1. Consent is not really independent-** users need to accept the terms in order to use the service, which poses a problem for the voluntariness of consent.
- 2. Statutory rights trump contractual terms-** copyright law gives the author the exclusive rights that cannot be waived off through large non-negotiable clickwrap agreements unless expressly allowed by statute.

Thus, although platforms may be able to keep ephemeral content via contractual means, such practices will not be exempt from copyright limitations and user rights.

International Frameworks

International jurisdictions have started to address issues relating to AI and copyright though none of this has specifically covered ephemeral content.

European Union

The EU's DSM Directive added some mandatory TDM exceptions under Articles 3 and Article 4, which permit certain uses for research and commercial purposes³¹. However, the Directive does not make this distinction between permanent and ephemeral digital works, and it does not explain whether disappearing content can be lawfully kept or used for AI training.

United States

³⁰ TikTok, *Terms of Service* (2024), <https://www.tiktok.com/legal/terms-of-service>.

³¹ Directive (EU) 2019/790, arts. 3–4, <https://eur-lex.europa.eu/eli/dir/2019/790/oj>.

The United States is very dependent on the fair use doctrine. While courts have been lenient on uses of AI training - especially in situations where the use is transformative - there is not yet a case that addresses ephemeral content specifically.³² The legal basis is still unclear and very contextual.

WIPO and Global IP Standards

While WIPO has acknowledged the complexities involved with AI and copyright, its work on the issue has not provided a framework for ephemeral works and their use in datasets. Discussions continue to focus on the need for transparency, accountability, and safeguards within AI training but do not go much beyond providing rules of thumb.³³

Together, these frameworks show some significant gaps. No legal system - Indian or international has yet enunciated rules applicable to ephemeral content, leaving the users and developers in a position of uncertainty about the legality of archiving and AI training.

³² U.S. Copyright Office, *Copyright and Artificial Intelligence* (2023), <https://www.copyright.gov/ai/>.

³³ World Intellectual Property Organization, *AI Policy Reports* (2023),
https://www.wipo.int/tech_trends/en/artificial_intelligence/.

Analysis

The legal and practical challenges that surround ephemeral content arise out of the tension between what users expect to be private content and the technological architecture of the social media platforms such as Snapchat that market disappearing messages as private and temporary.³⁴ At the same time, the backend of operations of platforms and AI developers make the copyright status of such content far more complicated. This section discusses the ephemeral content in terms of five fundamental aspects: copyrightability, legality of archiving, AI training, user consent, and the need for a new legal category.

Ephemeral Content Copyrightability

Ephemeral content is on-screen content that is only available for a short period of time, often seconds or hours, but nevertheless, content that is stored on servers for a sufficient period of time to meet the fixation requirement of copyright law. Indian law does not require much fixation, and courts worldwide, including in places such as the United States, see fixation as expansive, so that short-term storage is sufficient to meet the requirement.³⁵

Because ephemeral content is often creative (photographs, videos, filters, drawings), ephemeral content also passes the originality test. Therefore the 10 second Story on Instagram or disappearing post on WhatsApp automatically gets copyright protection at the time of creation.³⁶

Legality of Archiving of Ephemeral Content

Archiving ephemeral content -- by platforms, third-party scrapers and individuals -- usually includes reproduction. Under Section 51 of the Copyright Act making a reproduction without permission amounts to infringement unless an exception applies.³⁷

The problem is layered:

1. Platforms keep copies of ephemeral content (sometimes for 24-30 days) for safety reviews, analytics and internal machine learning systems.³⁸

³⁴ Snapchat, *When Are Snaps & Chats Deleted?* (2024), <https://support.snapchat.com/en-US/a/when-are-snaps-chatsdeleted>.

³⁵ U.S. Copyright Office, *Copyright and Fixation Requirements* (2023), <https://www.copyright.gov/circs/circ1.pdf>.

³⁶ Copyright Act, 1957, § 13 (India), <https://www.indiacode.nic.in/bitstream/123456789/2066/1/A1957-14.pdf>.

³⁷ Copyright Act, 1957, § 51 (India), <https://www.indiacode.nic.in/bitstream/123456789/2066/1/A1957-14.pdf>.

³⁸ Instagram Help Center, *How Long We Store Data* (2024), <https://help.instagram.com/116024195217477>.

2. Scrapers and bot operators gather vanishing media, using automated tools or APIs.
3. AI developers consume epic volumes of data containing ephemera without divulging sources.

In all these contexts, archiving involves reproduction - and there is no statute expressly allowing for such copying.

AI Training and Copyright

AI training involves the copying of works in order to study the patterns of the work. This includes the generation of temporary or permanent dataset files, representation embedding, and generation of training corpora. In India, the fair dealing does not have any machine learning in it; Section 52 does not have any TDM exception³⁹.

By contrast:

- The European Union has developed mandatory exceptions for TDM, but they have opt-outs and do not expressly refer to ephemeral content.⁴⁰
- The United States uses fair use, and some argue that AI training is transformative; nevertheless, no case has been made on disappearing content specifically.⁴¹

Therefore, the AI training on ephemeral content, let alone without consent, is likely infringing on the Indian law and uncertain internationally.

User Consent and Control of Platform

Platforms like Meta Platforms offer users Terms of Service that give wide licenses for storing, reproducing and using user-generated content.⁴² However:

1. **Consent is not informed**- users are not usually informed that ephemeral content can be stored after it disappears.
2. **Consent is not voluntary**- users must agree to all the terms in order to use the service.
3. **Contractual licenses cannot supersede statutory copyright**- without the permission of law.

Courts have been historically concerned with licenses based on contracts in which user rights are reduced through take-it-or-leave-it contracts.

³⁹ Copyright Act, 1957, § 52 (India), <https://www.indiacode.nic.in/bitstream/123456789/2066/1/A1957-14.pdf>.

⁴⁰ Directive (EU) 2019/790, arts. 3–4, <https://eur-lex.europa.eu/eli/dir/2019/790/oj>.

⁴¹ U.S. Copyright Office, *Policy Statement on AI and Copyright* (2023), <https://www.copyright.gov/ai/>.

⁴² Meta Transparency Center, *Terms of Service* (2024), <https://transparency.meta.com/policies/terms-of-service/>.

Thus, it is legally insufficient to rely on platform licenses in order to justify AI training.

Need for a New Copyright Category

Ephemeral content upsets doctrinal assumptions:

- It is creative but supposed to be temporary.
- It is fixed and is intended to disappear.
- It is shared in public but expected to be held in secret.

Current copyright frameworks have no provision for the intent of ephemeral media and treat it the same as permanent works. As the disappearing content grows and grows (via platforms such as YouTube for temporary live-streams and Telegram for self-destruct messages), the disparity in regulation widens.

A special framework is required to safeguard user expectations, define acceptable platform retention and control AI training data sets.

Research Methodology

This research takes a structured and multi-method legal approach aimed at analyzing the copyright implications of archiving and using ephemeral content in artificial intelligence training systems. Because the problem touches upon statutory interpretation, platform governance, international standards, and technological practices, the study uses doctrinal, comparative, analytical, and content analysis methods. Institutions such as the Ministry of Electronics and Information Technology and global bodies such as Organisation for Economic Co-operation and Development have highlighted the increasing importance of AI governance frameworks.⁴³ This shows an increasing need for a proper methodology in this regard.

Doctrinal Legal Research

Doctrinal research is used as a basis for this study. It is based on a detailed examination of:

- the Copyright Act, 1957;

⁴³ Organisation for Economic Co-operation and Development, *OECD Framework for the Classification of AI Systems* (2022), <https://www.oecd.org/digital/oecd-framework-classification-ai-systems.pdf>.

- case law of fixation, reproduction, and digital infringement; and
- statutory interpretations of fair dealing & exclusive rights.

This approach is useful to clarify the issue of whether ephemeral content is a protected work and whether archiving or AI training amounts to infringement. It also provides for the evaluation of the application of legal doctrines to self-deleting digital media.⁴⁴

Comparative Legal Method

Given the global nature of the development of AI, the research compares Indian legal framework with:

1. European Union TDM provisions in the DSM Directive;⁴⁵
2. United States fair use doctrine, especially in the cases involving machine learning;⁴⁶
3. WIPO consultations on Artificial Intelligence and copyright.

Comparative analysis helps identify the gaps, similarities and best practices in an international context to make informed policy recommendations relevant to the Indian context.

Platform Policies Content Analysis

Platform policies - and in particular TikTok, YouTube and Snap Inc. - are critical to understanding how ephemeral content is governed contractually. This research is an analysis of Terms of Service, privacy policies, retention statements and data governance documents to determine:

- the degree of rights of platforms over ephemeral content,
- the time period for which the data is stored,
- whether the users give valid and informed consent,
- how platforms make public their AI and algorithmic use of user-generated content⁴⁷.

These findings are then compared with statutory copyright protections in order to identify conflicts between the governance of the platforms and legal rights.

⁴⁴ Copyright Act, 1957, S.13–14, <https://www.indiacode.nic.in/bitstream/123456789/2066/1/A1957-14.pdf>.

⁴⁵ Directive (EU) 2019/790, arts. 3–4, <https://eur-lex.europa.eu/eli/dir/2019/790/oi>.

⁴⁶ U.S. Copyright Office, *Copyright and Artificial Intelligence* (2023), <https://www.copyright.gov/ai/>.

⁴⁷ TikTok, *Terms of Service* (2024), <https://www.tiktok.com/legal/terms-of-service>.

Analytical Method

This method is synthesising doctrinal findings, platform policy analysis and comparative insights in order to assess the two hypotheses posed in this research. It examines:

- ephemeral content and its legal implications from archiving;
- whether AI training is permissible under the existing law;
- contradictions between user expectations and platform practices;
- the doctrinal adequacy of current statutes to deal with disappearing media.

The analytical approach is used to ensure that the research is based on informed conclusions, based on the known principles of law.

Data Sources

The research is strictly based on authoritative and verifiable sources including:

- statutory documents (Indian Copyright Act, EU Directives, United States policy statements);
- judicial decisions;
- transparent reports on platforms;
- academic peer-reviewed publications;
- international documents of policy from bodies such as WIPO, the Organisation for Economic Cooperation and Development (OECD);
- official publications of the government.

No unverifiable data sets or speculative sources are used.

Findings

The analysis of provisions in statutes, practices of platform governance, international standards, and AI training mechanisms reveal some important findings. These results indicate significant gaps in doctrine, practical ambiguities and user right issues in the context of ephemeral content. As digital ecosystems change - through the actions of companies such as OpenAI, which use data sets on an industrial scale - the legal treatment of vanishing media becomes an important consideration in ensuring accountability, transparency, and meaningful consent from the users.

Ephemeral Content Meets the Copyright Requirement

The research confirms that the ephemeral content does meet the originality and fixation requirements. Even though it is temporary, content posted in the social platforms is stored, however, temporarily on the servers, thus satisfying the fixation requirement under Indian copyright law.⁴⁸ This means:

- Instagram Stories,
- Snapchat Snaps,
- WhatsApp Status updates, and
- temporary YouTube or Facebook live are all copyrighted from the time of their creation, whether they are intended to last for a long time or not.

Archiving Ephemeral Content Archiving Is Reproduction

Evidence from platform transparency reports and technical documentation tells us that ephemeral content is often stored for safety reviews, analytics, and internal processes. Platforms like Instagram and Snapchat take notice of backend retention even after content has moved off the user interface.⁵² Archiving ephemeral content - by platforms, scrapers, or AI developers - is reproduction, an exclusive right of the author. Since Section 51 of the Copyright Act, 1957 makes it a crime to reproduce any work without the consent or assent of the copyright owner, such retention is legally doubtful without a statutory exception.

⁴⁸ Copyright Act, 1957, § 13 (India), <https://www.indiacode.nic.in/bitstream/123456789/2066/1/A1957-14.pdf>.⁵²

Snapchat, *When Are Snaps & Chats Deleted?* (2024), <https://support.snapchat.com/en-US/a/when-are-snaps-chatsdeleted>.

AI Training on Ephemeral Content is not covered under Indian Copyright Exceptions

AI systems need to copy works into datasets in order to draw out patterns. Because, in India, there is no TDM exception in the law, any reproduction for AI training is likely infringing⁴⁹. Unlike the EU that gives certain exceptions, Indian law remains silent. Thus:

- training large language models,
- preparing image datasets,
- mining the short videos or Stories all fall within the permitted exceptions of India.

This silence leaves a regulatory void: AI developers may be violating copyright with no statutory protection.

User Consent Given Through Platform Terms is Not Informed or Voluntary

Platforms such as Twitter use broad Terms of Service to justify retention to use algorithms on user generated content.⁵⁰ However, the findings show:

- users do not understand or meaningfully consent to disappearing media being backed up;
- platform licenses are bundled and not negotiable;
- users are not informed that their ephemeral content can be used in datasets for training of AI algorithms.

Contract-based licenses cannot supersede fundamental statutory rights unless the law says otherwise. Thus, platform reliance on TOS is from a copyright and user autonomy perspective still weak.

International Frameworks Cannot Address Ephemeral Content

Comparative analysis shows:

- EU law does provide TDM exceptions but does not deal with disappearing media;⁵¹
- U.S. law--through fair use--may provide justification for some AI training but has no case law involving ephemeral content;⁵²

⁴⁹ Copyright Act, 1957, § 52 (India), <https://www.indiacode.nic.in/bitstream/123456789/2066/1/A1957-14.pdf>.

⁵⁰ Twitter, *Terms of Service* (2024), <https://twitter.com/en/tos>.

⁵¹ Directive (EU) 2019/790, arts. 3–4, <https://eur-lex.europa.eu/eli/dir/2019/790/oj>.

⁵² U.S. Copyright Office, *Copyright and Artificial Intelligence* (2023), <https://www.copyright.gov/ai/>.

- WIPO has opened consultations on the topic of AI but does not define ephemeral works as an area of protection.⁵³

On a global scale no jurisdiction has developed a doctrine or statute that regulates:

- of vanishing content the retention,
- the status of copyright in empirically temporary digital works, or ● the permissibility of the use of such works in AI training.
- This absence is a huge blind spot for the law.

Ephemeral Content Needs a New Legal Category

The research proves that ephemeral content is in a doctrinally ambiguous position:

- It is creative and it's protectable.
- It is fixed but it is meant to disappear.
- It is public but it is expected to be private or temporary.

Current copyright law treats it the same as permanent works, without taking into consideration intent of temporality as a legally relevant factor. This omission causes conflicts between user expectations, platform operation and AI training practices.

Therefore, the results are quite compelling in favour of establishing a new statutory category or regulatory framework for ephemeral digital works.

Recommendations

The results of this research show significant gaps in the statutory protection, platform governance and AI regulatory structures on ephemeral content. To ensure user rights, encourage responsible AI development and doctrinal clarity, this section makes a number of legal, policy, and institutional recommendations. These recommendations are consistent with the global best practices and current deliberations in institutions like the NITI Aayog and the United Nations Educational, Scientific and

⁵³ World Intellectual Property Organization, *WIPO AI Policy Documents* (2023), https://www.wipo.int/tech_trends/en/artificial_intelligence/.

Cultural Organization, which have highlighted the need for transparent and ethical AI governance in the digital space.⁵⁴

Present Statutory Recognition of Ephemeral Digital Works

The Indian copyright law should accept "ephemeral digital works" as a new category. Such recognition should include:

- the intention of temporality being a relevant factor;
- restrictions on acceptable reproduction;
- special protections from unauthorized archiving;
- default prohibitions of long-term retention by platforms or third parties.

Current law does not recognize the difference between ephemeral and permanent content and does not take into account user expectations and platform design. A legislative amendment (possibly via an amended Section 13A) would make the legal status of self-deleting media clear.

Require Express, Opt-in Consent for AI Training

Platforms such as TikTok and Meta Platforms currently use bundled Terms of Service that one must accept wholesale. This model does not pass the test of meaningful consent.

The following are some of the recommended reforms:

- mandatory opt-in clear consent in cases of AI training use;
- separate disclosures of ephemeral content;
- prominent user notifications prior to retention or addition of content to datasets.

This fits with the global ethical principles of AI that focus on transparency and user autonomy.⁵⁵

Set Legal Constraints on Platforms for Retaining Data of Ephemeral Content

Legislation/Regulatory guidelines should:

- indicate maximum retention periods for disappearing content;
- not to keep longer than is operationally necessary (e.g. safety checks);
- require platforms to make retention transparency reports.

⁵⁴ UNESCO, *Recommendation on the Ethics of Artificial Intelligence* (2021), <https://unesdoc.unesco.org/ark:/48223/pf0000379920>.

⁵⁵ NITI Aayog, *Responsible AI for All: National Strategy for Artificial Intelligence* (2021), https://www.niti.gov.in/sites/default/files/2021-02/NationalStrategyforAI-DiscussionPaper_0.pdf.

Platforms should not keep ephemeral content forever or for unknown reasons (e.g. dataset creation).

Amend Section 52 to Deal with Text-and-Data Mining (TDM)

India's Copyright Act should be amended to clarify if TDM for training AI is allowed. There are two approaches on an international level:

1. **EU approach**- allowing TDM but giving right holders the right to opt out.⁵⁶
2. **Restrictive approach**- licensing the training datasets

India must decide the model and openly deal with ephemeral content in it.

Apply Transparency Requirements on AI Developers

AI companies such as Google DeepMind increasingly use big data, opaque data sets. To deal with this, legislation should require:

- disclosure of data set sources;
- disclosure of whether or not ephemeral content is used;
- routine publication of audits of datasets; ● compliance of user deletion requests.

Transparent datasets minimize the risk of the infringement of ephemeral works.

Restrain the Overreach of Contracts by Platforms

Terms of Service should not be a way to override statutory rights. India should adopt the rules analogous to the EU's Unfair Contract Terms Directive to prohibit:

- one-sided platforms licenses;
- rights over ephemeral content, whether perpetual or otherwise, without royalties;
- broad rights of using disappearing media for algorithmic or training purposes.

Any kind of licensing for content must be freely agreed upon; not bundled into clickwrap agreements.

Develop a Regulatory Body for AI and Digital Content

India may think of widening the mandate of Telecom Regulatory Authority of India or setting up a separate Digital Content and AI Regulatory Authority (DCARA). This body should:

- superimposed AI dataset governance;
- regulate the platform retention practices;

⁵⁶ Directive (EU) 2019/790, arts. 3–4, <https://eur-lex.europa.eu/eli/dir/2019/790/oj>.

- enforce user rights with regards to ephemeral content; ● conduct audits and penalise for unlawful copying.

This covers the existing fragmentation among ministries and regulatory gaps.

Promote Public Awareness and Digital Literacy

Citizens tend to get confused about the workings of ephemeral features. It is important for the government and platforms to work together to offer:

- clearer user education;
- warnings about screen shots and archives;
- transparency dashboards of retention periods and dataset usage.

This empowers the users to make informed decisions.

Conclusion

Ephemeral content has become a signature feature of digital communication today, made famous by the likes of Snap Inc. and adopted throughout the messaging and social media ecosystems. While disappearing media is perceived by users to be temporary and private, the reality is that there are significant technical and legal issues which prove that disappearing media is routinely retained, archived and potentially repurposed - sometimes without the user's knowledge or consent. The results of this research show that there is a deep disconnection between user expectations, platform practices, and copyright law.

From a doctrinal standpoint, ephemeral content meets the requirements of originality and fixation, and thus gets full protection of copyright from the moment it is created and uploaded.⁵⁷ The research also sets the threshold that archiving of ephemeral content involves reproduction, which is a restricted activity under Indian law, unless it is done with express authorization. Because the Copyright Act, 1957 has no exception on TDM or AI training, the use of ephemeral content in training machine learning systems is extremely likely to be an infringement.⁵⁸ This risk is aggravated by the lack of transparency of dataset sources and disclosure from AI developers.

Furthermore, platform Terms of Service, including those of companies such as Meta Platforms, are based on broad and non-negotiable licenses that do not respect informed user consent. These types of contractual licenses cannot trump statutory copyright protections, especially when they are bound into take-it-or-leave-it digital contracts.⁵⁹

At the international level, comparative study shows that although the European Union has implemented TDM exceptions and the United States is based on flexible fair use doctrines, neither jurisdiction has engaged in the status of ephemeral content and its suitability for the training of AI. Global bodies like WIPO are doing active work in the study of AI governance but are yet to formulate frameworks for disappearing media.⁶⁰

⁵⁷ Copyright Act, 1957, § 13 (India), <https://www.indiacode.nic.in/bitstream/123456789/2066/1/A1957-14.pdf>.

⁵⁸ Copyright Act, 1957, S. 51, 52 (India), <https://www.indiacode.nic.in/bitstream/123456789/2066/1/A1957-14.pdf>.

⁵⁹ Meta Transparency Center, *Terms of Service* (2024), <https://transparency.meta.com/policies/terms-of-service/>.

⁶⁰ World Intellectual Property Organization, *WIPO Conversation on IP and Frontier Technologies* (2024),

https://www.wipo.int/about-ip/en/frontier_technologies/.

Taken together, the analysis affirms the main hypothesis for this research: the use of ephemeral content for AI training without explicit user consent is copyright infringement. The null hypothesis -that vanishing media may be within implied consent or permissible exceptions -- is not supported under current statutory frameworks.

The conclusion of a study on the legal status of ephemeral digital works is that such ephemeral digital works need clear statutory recognition, explicit consent mechanisms, tight retention limits, and dataset transparency obligations for AI developers. Without such reforms, users will continue to lose control of their creative expressions, platforms will continue to have too much power, and the development of AI will occur in a legal grey zone that is bad for both innovation and user rights.

Ephemeral content raises fundamentally new questions of authorship, temporality, privacy and digital autonomy. As India works towards large-scale adoption of AI, digital governance reform, the preservation of the vanishing media needs to be a key element of modern copyright policy, and ensure that technological advancement does not occur at the expense of user expectations and intellectual property rights.

Bibliography

I. Primary Sources

1. Copyright Act, 1957 (India),
<https://www.indiacode.nic.in/bitstream/123456789/2066/1/A1957-14.pdf>.
2. Ministry of Electronics & Information Technology (Ministry of Electronics and Information Technology), *IndiaAI Report 2023*, <https://www.meity.gov.in/content/indiaai>.
3. U.S. Copyright Office, *Copyright and Artificial Intelligence* (2023),
<https://www.copyright.gov/ai/>.
4. UNESCO (United Nations Educational, Scientific and Cultural Organization), *Recommendation on the Ethics of Artificial Intelligence* (2021),
<https://unesdoc.unesco.org/ark:/48223/pf0000379920>.

II. International Treaties, Directives & Policy Documents

1. Directive (EU) 2019/790 of the European Union, arts. 3–4, <https://eur-lex.europa.eu/eli/dir/2019/790/oj>.
2. World Intellectual Property Organization (World Intellectual Property Organization), *WIPO Conversation on IP and Frontier Technologies* (2024),
https://www.wipo.int/aboutip/en/frontier_technologies/.
3. OECD (Organisation for Economic Co-operation and Development), *OECD Framework for the Classification of AI Systems* (2022), <https://www.oecd.org/digital/oecd-frameworkclassification-ai-systems.pdf>.

III. Case Law

1. *Authors Guild v. Google, Inc.*, 804 F.3d 202 (2d Cir. 2015),
<https://law.justia.com/cases/federal/appellate-courts/ca2/13-4829/13-4829-2015-10-16.html>.

IV. Secondary Sources (Scholarly Commentary, Reports, Research)

1. Electronic Frontier Foundation, *How AI Training Data Works* (Sept. 2023), <https://www.eff.org/deeplinks/2023/09/how-ai-training-data-works>.

2. Mozilla Foundation, *AI Transparency Research* (2024),
<https://foundation.mozilla.org/en/research/ai-transparency/>.
3. NITI Aayog (NITI Aayog), *Responsible AI for All: National Strategy for Artificial Intelligence* (2021), https://www.niti.gov.in/sites/default/files/2021-02/NationalStrategyforAI-DiscussionPaper_0.pdf.
4. Indian Ministry of Human Resource Development, *Committee Report on Intellectual Property Rights* (2020),
https://www.education.gov.in/sites/upload_files/mhrd/files/document-reports/IPR.pdf.

V. Online Platform Documentation (Terms, Policies, Transparency Reports)

1. Snap Inc. (Snap Inc.), *When Are Snaps & Chats Deleted?* (2024),
<https://support.snapchat.com/en-US/a/when-are-snaps-chats-deleted>.
2. Instagram Help Center (platform owned by Meta Platforms), *How Instagram Stories Work* (2023),
<https://help.instagram.com/1660923094227526>.
3. Twitter (now X) (Twitter), *Terms of Service* (2024), <https://twitter.com/en/tos>.
4. TikTok (TikTok), *Terms of Service* (2024), <https://www.tiktok.com/legal/terms-of-service>.
5. Meta Transparency Center, *Terms of Service* (2024),
<https://transparency.meta.com/policies/terms-of-service/>.

VI. Technical & Educational Resources

1. U.S. Copyright Office, *Copyright and Fixation Requirements: Circular I* (2023), <https://www.copyright.gov/circs/circ1.pdf>.
2. Instagram Data Storage FAQ, *How Long We Store Data* (2024),
<https://help.instagram.com/116024195217477>.

EDITORIAL TEAM

PROF. (DR.) BANSHI DHAR SINGH

Professor,
Ex. Dean & Head,
Faculty of Law,
University of Lucknow

DR. KALPESH KUMAR L GUPTA

Founder ProBono India, Legal Start-ups, Law Teachers India

DR. SUDHANSU CHANDRA

Assistant Professor, Manuu Law School, Maulana Azad National Urdu University (Central University), Hyderabad

PROF. (DR.) SANJAY SINGH

Director
of IIMT College of Law

INTERNATIONAL EDITORIAL TEAM

PROF. DR. MARC OLIVER OPRESNIK

President and CEO
Opresnik Management Consulting
and Opresnik Business School

PROF. DR . COMRADE AMB.

**CHUKWUNONSO C
HARLES OFODUM ESQ**

Chancellor, ALSA University.
Legal Director for Nigeria, World Association for Humanitarian Doctors

ABOUT LEX SCRIPTA JOURNAL

Lex Scripta Magazine is a premier peer-reviewed online and print journal dedicated to advancing scholarly research in law, policy, and social sciences. With the vision of promoting academic excellence and fostering a culture of intellectual exchange, the magazine provides a distinguished platform for academicians, researchers, legal professionals, and students to publish their original work and contribute to contemporary legal discourse.

Each submission undergoes a rigorous double-blind review process conducted by a panel of eminent national and international professors, ensuring the highest standards of quality and academic integrity. Lex Scripta not only encourages original and innovative research but also strives to bridge the gap between theoretical insights and real-world applications in the legal domain.

Contributors and editorial members receive global recognition through certificates and publication opportunities, while readers gain access to insightful, authoritative, and thought-provoking content across diverse areas of law and policy.

Now managed by Integrity Education India, Lex Scripta Magazine is committed to expanding its academic footprint through enhanced digital presence, global collaborations, and university partnerships. Upholding its ISSN identity, Lex Scripta continues to evolve as one of India's most trusted and respected journals in the field of legal research and education.

KEY FEATURES

- Scholarly Insights** – Access in-depth, peer-reviewed research articles written by distinguished academicians and legal experts.
- Global Perspectives** – Explore diverse viewpoints on law, policy, and governance from national and international scholars.
- Authentic Content** – Read verified and academically sound articles that uphold the highest standards of research quality.
- Knowledge Enhancement** – Stay updated with emerging trends, case studies, and policy developments across multiple legal domains.
- Easy Accessibility** – Enjoy seamless access to online editions and exclusive hardcover issues for academic and professional use.



CONNECT WITH US **9811 666 216**
7011 605 618

