

Delhi High Court refuses ANI's bid to stop OpenAI training on its news

Source copy kept by wheresthe.ai for the story "Delhi High Court refuses ANI's bid to stop OpenAI training on its news" (<https://wheresthe.ai/d/in-delhi-hc-ani-v-openai-interim-ruling-2026-07/>).

Source: https://delhihighcourt.nic.in/app/showFileJudgment/ABL24072026SC10282024_171649.pdf

Text saved 18 Oct 2026 (the source PDF).

SHA-256 of the saved text: e033e9cc6f3736a5ae1545c9cd934a979c8c7df899e4ed33a8c1c3d99b39c05b

Documents the source offers: https://delhihighcourt.nic.in/app/showFileJudgment/ABL24072026SC10282024_171649.pdf

* IN THE HIGH COURT OF DELHI AT NEW DELHI % Judgment reserved on: 27th March, 2026 Judgment pronounced on: 24th July, 2026

I.A. 45300/2024 IN + CS(COMM) 1028/2024, I.A. 45301/2024 & I.A.26192/2025

ANI MEDIA PVT. LTD. Plaintiff Through: Mr. Sidhant Kumar, Mr. Akshit Mago, Ms. Manyaa Chandok, Ms. Anshika Saxena and Ms. Lahar Jain, Advocates

Versus

OPEN AI OPCO LLC Defendant Through: Mr. Amit Sibal, Senior Advocate with Mr. Sanjeev Kapoor, Mr. Nirupam Lodha, Mr. Madhav Khosla, Mr. Gautam Wadhwa, Ms. Moha Paranjpe, Mr. Abhi Udai Singh Gautam, Ms. Rebecca Cardoso, Mr. Hardik Malik, Ms. Vanshika Thapliyal, Mr. Rajat Bector, Advocates Mr. Akhil Sibal, Senior Advocate with Mr. Aditya Gupta, Ms. Asavari Jain, Mr. Shuvam Bhattacharya, Ms. Vani Kaushik, Ms. Riddhie Bajaj, Ms. Jahnvi Siddhu, Ms. Aishwarya Kane and Mr. Sauhard Alung, Advocates for IGAP/Intervenor in I.A. 4616/2025 Mr. Kapil Sibal, Senior Advocate and Mr. Arvind P. Datar, Senior Advocate with Mr. Shashank Mishra, Ms. Akshi Rastogi, Mr. Parv Kaushik and Mr. Suvaroop Saha Roy, Advocates for Broadband India Forum

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 1 of 135 MISHRA Signing Date:24.07.2026 17:17:24 (BIF)/Intervenor in I.A. 10766/2025 Ms. Haripriya Padmanabhan, Senior Advocate with Mr. Shrutanjaya Bhardwaj, Mr. Akshat Agrawal, Mr. Tushar Srivastava, Mr. Shourya Das Gupta, Ms. Siddhi Nagwekar, Ms. Yashi Bajpai and Mr. Yash Tayal, Advocates for Flux AI Labs/ Intervenor in I.A. 4615/2025 Mr. Ankit Sahni, Ms. Kritika Sahni, Mr. Chirag Ahluwalia, Mr. Mohit Maru, Ms. Tanisha Sharma, Advocates for the Federation of Indian Publishers, Intervenor in I.A. No. 470 of 2025 Mr. Rajshekhar Rao, Senior Advocate and Mr. Chander M. Lall, Senior Advocate with Mr. Ameet Datta, Mr. Harsh Kaushik, Ms. Riddima Sharma, Mr. Akshay Nagarajan, Ms. Rishikaa, Ms. Gauri Khanna and Ms. Annanya Mehan, Advocates for DNPA/Intervenors in I.A. 2199/2025 & IMI/ Intervenor in I.A. 4027/2025 Mr. Adarsh Ramanujan with Mr. Parth Singh, Advocate, Amicus Curiae Professor Arul George Scaria, Amicus Curiae.

CORAM: HON'BLE MR. JUSTICE AMIT BANSAL JUDGMENT AMIT BANSAL, J. I.A. 45300/2024 (under Order XXXIX Rules 1 and 2 CPC)

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 2 of 135 MISHRA Signing Date:24.07.2026 17:17:24 INDEX PREFACE -----

4 WORKING OF LLMS ----- 8 A. TRAINING DATA ----- 8 B. DATA PREPARATION AND MODEL TRAINING----- 9 C. POST- TRAINING----- 10 D. USER INTERACTION ----- 10 E. RETRIEVAL-AUGMENTED GENERATION (RAG)----- 11 PROCEEDINGS IN THE SUIT----- 11 ISSUE NO.4 WHETHER THE COURTS IN INDIA HAVE JURISDICTION TO ENTERTAIN THE PRESENT LAW SUIT CONSIDERING THAT THE SERVERS OF THE DEFENDANTS ARE LOCATED IN THE UNITED STATES OF AMERICA ----- 13 SUBMISSIONS ----- 13 DISCUSSION AND ANALYSIS ----- 19 i. Whether this court has territorial jurisdiction to entertain the present suit. ----- 19 ii. Whether the Copyright Act, 1957 would apply in relation to

training claim since according to Open AI training takes place on servers located outside India

----- 20

FINDINGS----- 22 ISSUE NO.2

WHETHER THE USE BY THE DEFENDANTS OF PLAINTIFF'S COPYRIGHTED DATA IN ORDER TO
GENERATE RESPONSES FOR ITS USERS, WOULD AMOUNT TO INFRINGEMENT OF THE PLAINTIFF'S
COPYRIGHT. ----- 23 SUBMISSIONS

----- 23 DISCUSSION AND ANALYSIS

----- 34 i. Whether Open AI memorizes and
regurgitates ANI's copyrighted literary works in the form of responses?

----- 37 ii. Whether ChatGPT's responses
are substantial reproduction of ANI's copyrighted literary works?

----- 42

FINDINGS----- 60 ISSUE NO.1

WHETHER THE STORAGE BY THE DEFENDANTS OF PLAINTIFF'S DATA (WHICH IS IN THE NATURE OF
NEWS AND IS CLAIMED TO BE PROTECTED UNDER THE COPYRIGHT ACT, 1957) FOR TRAINING ITS
SOFTWARE I.E., CHATGPT, WOULD AMOUNT TO INFRINGEMENT OF PLAINTIFF'S COPYRIGHT

----- 61 SUBMISSIONS

----- 61 ISSUE NO.3 WHETHER THE
DEFENDANTS' USE OF PLAINTIFF'S COPYRIGHTED DATA QUALIFIES AS 'FAIR USE' IN TERMS OF
SECTION 52 OF THE COPYRIGHT ACT, 1957. ----- 68 SUBMISSIONS

----- 68 DISCUSSION AND ANALYSIS
OF ISSUE NOS. 1 AND 3 ----- 86

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 3 of 135 MISHRA Signing Date:24.07.2026
17:17:24 A. Scope of Section 14(a)(i) of the Copyright Act, 1957 ----- 86 B.

Legislative History of Section 52(1)(a)----- 92 C. Interpretation of
Section 52(1)(a) of the Copyright Act ----- 93 D. Scope of Section 52(1)(a)

----- 95 I. PURPOSE TEST

----- 96 a. Whether commercial use is
entitled to protection under Section 52(1)(a)(i) ----- 97 b. Requirement of "non-infringing copy" under
Section 52(1)(a) of the Copyright Act. -----104 c. Whether use by Open AI amounts to "private or personal
use, including research" -----108 II. FAIRNESS TEST

----- 115 a. Whether Open AI's use of ANI's
original literary works is limited to training its LLMs underlying
ChatGPT?-----123 b. Whether the usage of ANI's

literary works by Open AI would result in economic competition and would prejudice the legitimate interests of
ANI, thereby causing actual or potential damage to ANI?

-----124 c. Whether the functions
performed by Open AI through ChatGPT serve the overall public interest?

-----128

FINDINGS----- 129 BALANCE OF
CONVENIENCE AND IRREPARABLE INJURY ----- 129 SUBMISSIONS

----- 129 DISCUSSION AND ANALYSIS
----- 132

FINDINGS----- 134 CONCLUSION

----- 134

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 4 of 135 MISHRA Signing Date:24.07.2026
17:17:24 PREFACE 1. Innovations affecting the works protected under the Copyright Act, 1957 (hereinafter the
'Copyright Act'), namely literary, dramatic, musical and artistic works, cinematograph films, and sound recordings,
have consistently given rise to interpretive challenges in the application of traditional legal provisions. 2. While
tracing the history of copyright law, it is noteworthy that globally, copyright law and technological innovation have
long existed in a dynamic relationship, traceable at least to the advent of the Gutenberg Press in the fifteenth
century, well before the enactment of the Statute of Anne in 1710. 3. With time, successive technologies have
evolved, enabling new forms of creative expression and widening avenues for dissemination, disrupting settled
copyright arrangements. Some of these technologies worth referring to are player pianos, radio broadcasting,
cable television, photocopying, home video cassette recorders, digital downloading and streaming, and, more

recently, artificial intelligence (hereinafter 'AI'). Therefore, technological advancement continues to generate novel questions that necessitate periodic re-examination of established copyright doctrine. 4. It is the judiciary, as the primary forum for adjudication, that is invariably called upon to address the nuanced legal questions arising from such disruptive technologies, often resulting in conflicts, as in the present

1 In 1709 British parliament enacted the Statute of Anne; short title: Copyright Act 1709 8 Anne c.21; long title: An Act for the Encouragement of Learning, by vesting the Copies of Printed Books in the Authors or purchasers of such Copies, during the Times therein mentioned. Named after Anne, Queen of Great Britain, this was the first copyright statute in the Kingdom of Great Britain, and the first full-fledged copyright statute in the world. It was enacted in the regnal year 1709 to 1710, and entered into force on April 10, 1710.

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 5 of 135 MISHRA Signing Date:24.07.2026 17:17:24 case. Therefore, the Courts are required to step in to fill the gap between advancing technologies and existing laws. 5. AI is currently a technology of widespread interest, and India is considered a forerunner in this field. In response to international AI competition, Sarvam AI of India launched its first domestic Large Language Model (hereinafter 'LLM') on 18th February 2026². Research and development of similar LLMs is also underway in India to make India an AI hub. Data is the oil for LLMs to work efficiently. 6. Procuring publicly available data by AI innovators for training LLMs without the permission of the respective copyright holders potentially will raise legal issues, as arising in the present suit. The present suit has been filed by the plaintiff (hereinafter 'ANI') against the defendant (hereinafter 'Open AI') alleging unauthorised use of ANI's copyrighted works by Open AI. ANI claims copyright infringement on two counts: i. Training claim - copying and storage of ANI's data to train Open AI's LLM; and ii. Reproduction or Output claim - reproduction of ANI's works in the output produced by Open AI's LLM through its application ChatGPT. 7. This case is a perfect illustration of how the law must continually catch up with technology. The case has to be adjudicated on the anvil of the Copyright Act, originally enacted in 1957, though amended from time to time. Clearly, when the Copyright Act or any of its amendments were enacted, the legislature could not have imagined the vast technological advances in the

2 Aroon Deep, Bengaluru firm unveils two AI language models, The Hindu (Feb. 18, 2026), <https://www.thehindu.com/sci-tech/technology/bengaluru-firm-unveils-two-ai-language-models/article70648438.ece>, accessed on 20th February 2026.

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 6 of 135 MISHRA Signing Date:24.07.2026 17:17:24 field of AI. 8. In these uncharted areas, Courts often take aid from judgments of foreign Courts on similar issues. On various occasions, while dealing with Intellectual Property laws, the Indian Courts have recognized the importance of global doctrinal interdependence. The Courts laid emphasis on harmonization of basic principles of laws in foreign jurisprudence with the national laws, especially when the national laws do not indicate a contrary intent³. However, while doing so, the Courts must be mindful that the underlying legislation in foreign jurisdictions may be quite different from that prevailing in India. 9. For the purposes of adjudication of the present application for interim injunction filed by ANI, this Court had framed issues on 19th November, 2024, which are set out below: "I. Whether the storage by the defendants of plaintiff's data (which is in the nature of news and is claimed to be protected under the Copyright Act, 1957) for training its software i.e., ChatGPT, would amount to infringement of plaintiff's copyright. II. Whether the use by the defendants of plaintiff's copyrighted data in order to generate responses for its users, would amount to infringement of the plaintiff's copyright. III. Whether the defendants' use of plaintiff's copyrighted data qualifies as 'fair use' in terms of Section 52 of the Copyright Act, 1957. IV. Whether the Courts in India have jurisdiction to entertain the present lawsuit considering that the servers of the defendants are located in the United States of America."

Even though Issue no. 3 refers to 'fair use' [as it was framed on the first date of hearing] it should be read as "fair dealing" which is the expression used in Section 52(1)(a) of the Copyright Act.

3 Intex Technologies v. Telefonaktiebolaget LM Ericsson, 2023:DHC:2243-DB.

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 7 of 135 MISHRA Signing Date:24.07.2026 17:17:24 10. Since the question of this Court's jurisdiction to adjudicate the present application is foundational to the determination of the remaining issues, the Court shall first consider Issue No. 4. Thereafter, in view of its broader bearing on Issue Nos. 1 and 3, the Court shall examine Issue No. 2. As the Amici Curiae and counsel for the parties were ad idem that Issue Nos. 1 and 3 are intertwined, particularly in light of Open AI's plea of fair dealing in response to the allegations relating to storage, the said issues shall be considered together, subsequently. 11. With the aforesaid backdrop, I proceed to decide the present application.

WORKING OF LLMS 12. It is necessary to understand the functioning of LLMs for the purpose of adjudicating the present application, as the issues identified herein arise from, and are intrinsically connected with, the manner in which LLMs operate. Accordingly, it is apposite to first examine the technical functioning of LLMs. 13. Based on the submissions of the parties and learned Amici Curiae and also from the common general knowledge available, the steps involved in the working of LLMs, may be summarised in the following paragraphs. A. Training Data 14. A large language model is a computational system that can predict the next word from previous words. That is, given a context or prefix of words, a language model assigns a probability distribution over the possible next words. 15. For enabling a model to predict subsequent words, the model is first

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 8 of 135 MISHRA Signing Date:24.07.2026 17:17:24 trained on an enormous text corpus with datasets of hundreds of billions of words, generally scraped from the web, which is referred to as the 'raw data'. 16. The 'raw data' used for training LLMs would include automatically scraped data from publicly available sources on the internet, as well as licensed materials. The web text is usually taken from corpora of automatically crawled web pages, such as the Common Crawl. The data comprises materials, including but not limited to books, articles, blog posts, etc., so that, based on the aforesaid data, the model can make predictions. 17. Pre-training data collected from the web is filtered for both quality and safety. Quality filtering uses automated tools to score documents based on their usefulness and reliability, often by comparing them with trusted sources such as books and reputable websites, while avoiding content containing personal information or adult material. This generally improves language model performance. Safety filtering aims to remove harmful or toxic content, typically using existing toxicity-detection tools.⁴ 18. Website owners can indicate that they do not want their sites crawled by web crawlers.⁵ B. Data preparation and Model training 19. The raw data is divided into smaller pieces called tokens, each assigned a 'Token ID', which is just a number. This is called tokenisation. This process is the first step in training most of the LLMs.⁶

4 Matthew Sag and Peter K. Yu, 'The Globalization of Copyright Exceptions for AI Training' (Emory Law Journal, 2025, Forthcoming) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4976393. 5 Daniel Jurafsky and James H. Martin. 2026. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models, 3rd edition. Online manuscript released January 6, 2026. <https://web.stanford.edu/~jurafsky/slp3>. 6 Daniel J. Gervais et al., 'The Heart of the Matter: Copyright, AI Training, and LLMs' (2024) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4963711.

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 9 of 135 MISHRA Signing Date:24.07.2026 17:17:24 20. Thereafter, each token is converted to a vector via an 'embedding' process. 'Vectors' are a numerical representation of semantic information about a 'token'. 21. In the training exercise, the LLM is repeatedly fed parts of the training data where a word or a sequence of words is masked, and it is made to predict the next words or sentences, which are then compared with the original 'raw text' to verify correctness. The initial predictions may not be accurate, but with each repetition, the model readjusts and refines its prediction to reach the most probable statistical output in each scenario. 22. The choice of which word to generate in LLMs is done by sampling from the distribution of possible next words. C. Post-training 23. Thereafter, the model is again fine-tuned to reduce prediction errors and to follow instructions better, for example, to answer questions, give summaries, write code, translate sentences, and so on. It does this by being trained on a special corpus containing many texts with both instructions and their correct responses. 24. The model is then adjusted to give more useful responses and to avoid producing harmful or offensive content. D. User interaction 25. A user instructs LLM via a prompt, a text string that a user issues to a language model to get the model to do something useful by iteratively generating tokens conditioned on the prompt. The process of finding effective prompts for a task is known as prompt engineering.⁷

7 Mark Lemley, 'How Generative AI Turns Copyright Law Upside Down' (2024) 25(2) Sci. & Tech. LR 36.

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 10 of 135 MISHRA Signing Date:24.07.2026 17:17:24 E. Retrieval-Augmented Generation (RAG) 26. Retrieval-Augmented Generation ('RAG') is an innovative feature of LLMs that optimises their output. This feature references an authoritative knowledge base outside of its training data sources before generating a response. This feature does not predominantly rely on training data to generate output. Instead, it uses an information retrieval component that utilises the user prompt to pull information from external data storage.⁸ The LLM uses both training data and external information to generate better responses via RAG. 27. For the sake of convenience, 'LLM(s)' and 'ChatGPT' has been used interchangeably in the judgment and must be read likewise. When reference is made to 'ChatGPT search' function, it has been specifically mentioned in the judgment as such.

PROCEEDINGS IN THE SUIT 28. The present suit, accompanied by an application seeking interim injunction and other ancillary applications, was first listed before this Court on 19th November 2024. On the same date, four issues were framed and the Court appointed two Amici Curiae to assist the Court. 29. Thereafter, multiple intervention applications were filed in the present suit, the details of which are given below: 29.1. I.A. 470/2025 has been filed on behalf of Federation of Indian Publishers, a national representative body of the Indian publishing industry.

8 What is RAG? - Retrieval-Augmented Generation AI Explained - AWS, <https://aws.amazon.com/what-is/retrieval-augmented-generation/>.

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 11 of 135 MISHRA Signing Date:24.07.2026 17:17:24 29.2. I.A. 2199/2025 has been filed on behalf of Digital News Publishers Association and its members, which are digital news platforms. 29.3. I.A. 4027/2025 has been filed on behalf of the Indian Music Industry, which is an organization established to defend, preserve and develop the rights of the Phonogram Producers. 29.4. I.A. 4615/2025 has been filed on behalf of Flux Labs AI Private Ltd., a Generative Artificial Intelligence enabled start-up company. 29.5. I.A. 4616/2025 has been filed on behalf of IGAP Project LLP, a think tank focusing on the intersection of law, policy, and emerging technologies, including AI. 29.6. I.A. 10766/2025 has been filed on behalf of Broadband India Forum, an independent policy forum and think tank. 30. The intervenors in I.A. 470/2025, I.A. 2199/2025 and I.A. 4027/2025 are broadly supporting the case of ANI and the intervenors in I.A. 4615/2025, I.A. 4616/2025 and I.A. 10766/2025 are broadly supporting the case of Open AI. 31. On 4th July 2025, this Court allowed the aforesaid intervention applications and permitted the intervenors to make submissions limited to the legal issues framed on 19th November 2024. 32. ANI, Open AI as well as the intervenors have filed their written submissions along with judgments in support. 33. Submissions on behalf of the parties were heard on 21st February 2025, 10th March 2025, 18th March 2025, 28th March 2025, 2nd April 2025, 22nd April 2025, 29th April 2025, 4th July 2025, 23rd July 2025, 30th July 2025, 5th August

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 12 of 135 MISHRA Signing Date:24.07.2026 17:17:24 2025, 18th August 2025, 12th September 2025, 23rd September 2025, 17th October 2025, 7th November 2025, 21st November 2025, 23rd January 2026, 20th February 2026, 20th March 2026 and 27th March 2026, when the judgment was reserved and the parties were given liberty to file their closing submissions. 34. ANI, Open AI and the two Amici Curiae have filed their concluding written submissions. 34.1 The concluding written submissions dated 2nd April 2026 sent on behalf of Professor Arul George Scaria via email to the Court Master, have also been taken on record. 35. With the aforesaid backdrop, I proceed with my issues-wise analysis.

Issue No.4 Whether the Courts in India have jurisdiction to entertain the present law suit considering that the servers of the defendants are located in the United States of America

SUBMISSIONS Submissions on behalf of the Defendant (OPEN AI) 36. Mr. Amit Sibal, senior counsel appearing on behalf of Open AI made the following submissions: 36.1. ANI has joined two separate causes of action in the present suit i.e. - (i) use of publicly available data by Open AI on servers located outside India to train LLMs underlying ChatGPT ('training claim') and, (ii) the responses received by ANI from ChatGPT allegedly similar to content over which ANI holds copyright ('reproduction claim'). Therefore, ANI has to establish

9 Since the issue of jurisdiction has been raised by the defendant, it is deemed appropriate that the submissions of the defendant are recorded first.

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 13 of 135 MISHRA Signing Date:24.07.2026 17:17:24 jurisdiction qua both causes of action. 36.2. Insofar as the training claim is concerned, it is an undisputed position that the training takes place outside India on the basis of data that is stored in servers outside India. Therefore, merely because the jurisdiction can be established in respect of the reproduction claim, this Court would not have jurisdiction in respect of training claim. 36.3. The Copyright Act and in particular Section 62 thereof read with Section 1(2) of the Copyright Act, confers jurisdiction only within the territory of India. The Copyright Act was not intended to have extra-territorial jurisdiction. 36.4. The reliance placed by Professor Scaria on the judgment of a Coordinate Bench in Neetu Singh v. Telegram¹⁰ is misplaced insofar as the jurisdiction in respect of training claim is concerned. In Neetu Singh (supra), the infringement occurred within the jurisdiction of this Court and in that context, the Court observed that it would be immaterial if telegram servers are located outside India. Similarly, reliance placed on the judgment of the Supreme Court in Exphar SA v. Eupharma Laboratories¹¹ and the judgment of Division Bench in Blueberry Books v. Google India¹² is misplaced. 36.5. The

test laid down by the judgment of the Division Bench in *Banyan Tree Holding v. A. Murali Krishna Reddy*¹³ for the purposes of jurisdiction is limited. The purposeful availment test as laid down in *Banyan Tree* (supra) cannot be used to confer jurisdiction in respect of causes of action occurring outside India.

10 2022 SCC OnLine Del 2637. 11 (2004) 3 SCC 688. 12 2016 SCC Online Del 3338. 13 2009 SCC OnLine Del 3780.

Signature Not Verified Signed By: VIVEK CS(COMM) 1028/2024 Page 14 of 135 MISHRA Signing
Date: 24.07.2026 17:17:24 36.6. Insofar as the reproduction claim is concerned, at the time of using the ChatGPT service, the user agrees to 'terms of use' of ChatGPT, which provides jurisdiction of the Courts in San Francisco, California.

Submissions on behalf of the Plaintiff (ANI) 37. Mr. Sidhant Kumar, counsel appearing on behalf of ANI made the following submissions: 37.1. In terms of Section 62(2) of the Copyright Act, copyright owners can file suit for infringement in the jurisdiction where they reside or conduct business. Therefore, the physical location of Open AI's servers is of no consequence. ANI's principal and registered office is located within the jurisdiction of this Court. 37.2. Even in terms of Section 20(b) of the Code of Civil Procedure, 1908 (hereinafter 'CPC'), since Open AI operates an interactive website that enables commercial transactions within the jurisdiction of this Court, i.e. Open AI collects subscription fee from users located in this jurisdiction, Open AI would be amenable to jurisdiction of this Court. 37.3. Open AI actively attracts the consumers through advertising and public outreach efforts in India, that are within the jurisdiction of this Court. Purposeful engagement of ANI's consumers and a targeted marketing sufficiently demonstrate compliance of Section 20(c) of the CPC. 37.4. The various infringement activities alleged in the plaint have taken place within the jurisdiction of this Court. The disputes raised in the present suit are independent of 'terms of use' of ChatGPT. The acts that constitute copyright infringement are independent of any contractual agreements between the users and Open AI. The statutory remedies such as those provided

Signature Not Verified Signed By: VIVEK CS(COMM) 1028/2024 Page 15 of 135 MISHRA Signing
Date: 24.07.2026 17:17:24 under Section 62 of the Copyright Act cannot be waived through contractual stipulations.

Intervenors supporting ANI i. Submissions on behalf of the Intervenor – DNPA (I.A. 2199/2025) 38. Mr. Rajshekhar Rao, senior counsel appearing on behalf of intervenors in I.A. 2199/2025 made the following submissions: 38.1. The copies/reproduction of Indian works which are first published in India and thereafter made available by Open AI to attract Indian users, would imply that at the very least part of the cause of action arises within the jurisdiction of this Court. Therefore, in terms of Section 62 of the Copyright Act, read with judgment of the Supreme Court in *Indian Performing Rights Society v. Sanjay Dalia*¹⁴, this Court would have the jurisdiction to entertain the present suit. 38.2. Open AI is entering into commercial transactions within the jurisdiction of this Court and specifically targeting the customers. Reliance is placed on *Banyan Tree* (supra). 38.3. Neither the 'terms of use' of Open AI's ChatGPT, nor the fact that the servers located outside India, would oust the jurisdiction of this Court under Section 62 of the Copyright Act as well as Section 20 of the CPC.

ii. Submissions on behalf of the Intervenor - Indian Music Industry (I.A. 4027/2025) 39. Mr. C.M. Lall, senior counsel appearing on behalf of intervenors in I.A. 4027/2025 made the following submissions:

14 (2015) 10 SCC 161.

Signature Not Verified Signed By: VIVEK CS(COMM) 1028/2024 Page 16 of 135 MISHRA Signing
Date: 24.07.2026 17:17:24 39.1. Jurisdiction is also made out on the parameters of 'purposeful availment' as laid down in *Banyan Tree* (supra). In the present case, availment can be demonstrated since Open AI has 'paid subscribers' of its services within the jurisdiction of this Court and the said subscribers and users interact with the service within the jurisdiction.

iii. Written Submissions filed by Mr. Ankit Sahni, Advocate on behalf of the Intervenor – Federation of Indian Publishers (FIP) (I.A. 470/2025) 40. This Court would have jurisdiction under Section 62 of the Copyright Act. In this regard, reliance has been placed on the judgment of the Supreme Court in *Exphar*¹⁵ (supra) and *Dabur India Ltd. v. K.R. Industries*¹⁶ to submit that Section 62 of the Copyright Act was inserted by the Parliament to prescribe an additional ground for attracting jurisdiction over and above Section 20 of CPC. 40.1 Since Open AI's services are accessible within the jurisdiction of this Court and the outputs generated by Open AI's ChatGPT application infringes the copyright of ANI, a substantial part of cause of action arises within the jurisdiction of this Court to

invoke Section 20 of CPC.

Submissions of Mr. Adarsh Ramanujan, Amicus Curiae 41. Mr. Ramanujan made the following submissions: 41.1. The location of Open AI's servers outside India is irrelevant as ANI has invoked Section 62 of the Copyright Act. There is no challenge by Open AI to the assertion of ANI that its principal place of business is within the jurisdiction of this Court.

15 ¶ 13. 16 (2008) 10 SCC 595, ¶ 32.

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 17 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 41.2. Insofar as the submission of Open AI is concerned that the infringement relating to the training claim occurs outside the jurisdiction of this Court, what has to be seen is that the training process involves access of data from India and transmission of the said data from India. Therefore, it cannot be said that just because the training occurs outside the jurisdiction of this Court, this Court would not have jurisdiction. The training claim has to be seen in a holistic manner which would necessarily involve access and transmission of data from India before the same is stored overseas.

Submissions of Professor Arul George Scaria, Amicus Curiae 42. Prof. Scaria made the following submissions: 42.1. There is no dispute raised to the fact that ANI has its principal office in New Delhi. Therefore, this Court would have jurisdiction under Section 62(2) of the Copyright Act. 42.2. Since Open AI is making its active services available to the users in India, including those in Delhi, the Courts in India would have jurisdiction in terms of Section 20 of the CPC. 42.3. Reliance is placed on the judgment of the Supreme Court in *Expbar* (supra) and the judgment passed by the Division Bench of this Court in *Blueberry Books* (supra) and the judgment of a Coordinate Bench in *Neetu Singh* (supra). 42.4. Insofar as aspect of the territoriality of Indian Copyright Act is concerned, he supports the contention of Open AI that since the entire training process occurs outside India merely because Open AI accesses data from India would not make the infringement subject to Indian Copyright Act.

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 18 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 42.5. In *Getty Images v. Stability AI* the plaintiff in the United Kingdom ('UK') did not press the training claim as the training took place outside UK.

DISCUSSION AND ANALYSIS 43. As discussed earlier, since the issue of jurisdiction goes to the root of the matter, it would be taken for consideration at the beginning. 44. On the aspect of jurisdiction, based on the objections raised by Open AI, essentially the following two issues arise for consideration: i. Whether this Court has territorial jurisdiction to entertain the present suit. ii. Whether the Copyright Act would apply in relation to training claim since according to Open AI training takes place on servers located outside India.

i. Whether this Court has territorial jurisdiction to entertain the present suit. 45. Insofar as the territorial jurisdiction of this Court to entertain the present suit is concerned, it is an admitted position that ANI has its principal place of business as well as its registered office within the jurisdiction of this Court. Therefore, in terms of Section 62(2) of the Copyright Act, this Court would have the territorial jurisdiction to entertain the present suit. 46. It is also an admitted position that Open AI specifically targets/offers its services to users/subscribers across India, including those located within the jurisdiction of this Court. 47. Even alleged infringing activities mentioned in the plaint, i.e. responses

17 [2025] EWHC 38 (Ch).

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 19 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 generated by Open AI in India based on the prompts by ANI, took place within the jurisdiction of this Court. 48. Accordingly, at this prima facie stage, I hold that this Court would have the territorial jurisdiction to entertain the present suit in terms of Section 20 of the CPC as well as Section 62(2) of the Copyright Act.

ii. Whether the Copyright Act would apply in relation to training claim since according to Open AI training takes place on servers located outside India 49. It has been argued on behalf of Open AI that its LLMs are trained outside India and the training data is stored and processed in the servers located in the United States of America (hereinafter 'US'). On behalf of Open AI, it is contended that the Indian Copyright Act would not apply outside the territory of India. In support of its contention, reliance is placed on the judgment of the Division Bench in *Blueberry Books* (supra). In the said case, Amazon, which was a defendant in the suit took the defence that the infringing act took place exclusively in the US. Though Amazon admitted that it has business/commercial activities in India, the alleged infringing activity did not occur within the jurisdiction of this Court. 50. The Division Bench, in the said

case, upheld the territorial jurisdiction of this Court. Insofar as the argument of Amazon with regard to the infringing activity taking place in the US was concerned, the Division Bench did not uphold the contention of Amazon. Paragraph 27 of Blueberry Books (supra) is set out below:

“27. As far as inapplicability of the provisions of the Act to infringements within the USA are concerned, we note that the submission is based on Amazon’s arguments to that effect. Its unilateral

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 20 of 135 MISHRA Signing

Date:24.07.2026 17:17:24 assertion that there are downloading restrictions, based on a bare reading of the documents - or that only 33 downloads took place, for a small amount, are not relevant; they are contentious and accepting them would be giving disproportionate weight to the pleadings of one party and precluding the rights of the plaintiff, finally.”

51. In my view, the storing of ANI’s works on the US servers is a terminal step in the chain of events which begin from access of copyrighted works from India and transmission of the same abroad. There can be no doubt that Copyright Act cannot have jurisdiction outside the territory of India. However, the Copyright Act does not require this Court to sever the chain of events and examine only the last step. On a prima facie view, Open AI’s argument is untenable as it would lead to evasion of Indian copyright law by infringers who shift the terminal link to servers abroad. 52. In Neetu Singh (supra), a Coordinate Bench of this Court has observed that merely because Telegram chooses to locate its servers in Singapore, that would not affect the right of the copyright owners, to sue Telegram in India. The relevant observations from the aforesaid judgment are set out below: “46. In view of the above factual and legal position, in the opinion of this Court, merely because Telegram chooses to locate its server in Singapore, the same cannot result in the Plaintiffs’ - who are copyright owners of course materials - being left completely remediless against the actual infringers, especially in order to claim damages and avail of other legal remedies in accordance with law. If such an argument is accepted, in the current world where most dissemination happens through online messaging services and platforms, IP violations would go completely unchecked. This cannot be the intention of law. The provisions of the IT Act and the Rules made therein have to be construed harmoniously with the rights and remedies provided to the copyright owners under the Copyright Act. Indian Courts are competent to decide issues relating to infringement of copyright and the mere fact that Telegram is operating a messaging service in India which chooses not to locate its servers in India cannot

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 21 of 135 MISHRA Signing

Date:24.07.2026 17:17:24 divest the Indian Courts from dealing with copyright disputes or divest copyright owners from availing their remedies in Indian Courts. In the present age of cloud computing and diminishing national boundaries in data storage, conventional concepts of territoriality cannot be strictly applied. The dynamic evolution of law is essential to ensure appropriate remedies in case of violation of copyright and other IP laws.” [Emphasis supplied]

53. The aforesaid reasoning would be fully applicable in the present case. In Neetu Singh (supra), even though the data was stored in servers outside India, the same was being circulated within the jurisdiction of this Court. 54. In my view, the act of Open AI in scraping of ANI’s copyrighted material and training of its LLM cannot be entirely divorced from the output claim. Since the output claim is based on training of the LLM, and the output is reproduced within the jurisdiction of this Court, I am not inclined to separate the two causes of action.

FINDINGS 55. On a prima facie view, it cannot be said at this stage that this Court does not have the territorial jurisdiction to entertain the present suit or that the alleged infringement on account of training would involve extra-territorial application of the Indian Copyright Act. 56. As discussed earlier, next, I would take up Issue No. 2 for consideration.

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 22 of 135 MISHRA Signing

Date:24.07.2026 17:17:24 Issue No.2 Whether the use by the defendants of Plaintiff’s copyrighted data in order to generate responses for its users, would amount to infringement of the Plaintiff’s copyright.

SUBMISSIONS Submissions on behalf of the Plaintiff (ANI) 57. Mr. Sidhant Kumar, counsel appearing on behalf of ANI made the following submissions: 57.1. LLMs generate responses by matching user input with the vector database and predicting the token sequences. This is the process of ‘de- tokenization’. Once a response is formed, detokenization converts tokens into readable text by reassembling words, phrases, and sentences. Detokenization reconstructs raw data into its original textual form, converting training data (including ANI’s works) back into a format readable by humans. Therefore, the whole process amounts to reproduction of raw data. 57.2. ANI has provided instances of ChatGPT’s unauthorized reproduction of its works while responding to the

ChatGPT's users, in paragraph 15 of the interim application. ANI claims these examples are illustrative and not exhaustive. 57.3. ChatGPT's output retains the original and creative elements of the ANI's text and thereby violates ANI's proprietary rights. Some of the responses demonstrate exact or nearly exact copies of ANI's data. This occurs due to memorization of the raw data/ training data. 57.4. The outputs generated by ChatGPT are not derivative literary works either, as it requires the author to exercise independent skill, labour, and

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 23 of 135 MISHRA Signing Date:24.07.2026 17:17:24 resources [Ref. Eastern Book Company v. D.B. Modak¹⁸]. In the present case, the user generated responses show that there is no independent application of mind. At best, the user generated responses can be considered adaptations of ANI's works as these responses are merely a re-arrangement of ANI's works. 57.5. Open AI itself admits that verbatim reproduction of content from the raw data occurs in cases where the same content occurs repeatedly in the raw data, through multiple sources on the internet. The very nature of ANI's business model makes its copyrighted works available on various platforms including its subscribers' websites, which are leading news publications. 57.6. The defence of Open AI that ANI's works are available in the public domain and can be used for training Open AI's LLM without obtaining a license from ANI, is not valid and strikes at the root of copyright law. Public availability does not nullify copyright or grant a universal license for unrestricted use. ANI is not divested of its exclusive rights in its works only because it has licensed its content to its subscribers. 57.7. Open AI has recognized similar proprietary rights of other news organizations and entered into license agreements with other news organizations such as Financial Times, Associated Press and Condé Nast, and has acknowledged the copyright of these entities. 57.8. Open AI has also infringed ANI's exclusive right to translate the works, for which ANI invests significant time and effort. In this regard, reliance is placed on the English version of the interview given by Mr. Neeraj Chopra's mother, which was originally in Hindi and was translated into English

18 (2008) 1 SCC 1.

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 24 of 135 MISHRA Signing Date:24.07.2026 17:17:24 by ANI. 57.9. In his rejoinder submissions, counsel for ANI placed reliance on the following foreign judgments: (i) In *The Associated Press v. Meltwater U.S. Holdings, Inc.*¹⁹, the US District Court held that summarizing plaintiff's works using computer programme without adding any commentary is not transformative and hence, it does not constitute fair use under the US Copyright Act. (ii) In *GEMA v. Open AI*²⁰, the Munich Regional Court granted an injunctive relief in favour of GEMA and against Open AI holding that there is memorization of training data and consequential reproduction. (iii) In *The Advance Local Media LLC et al. v. Cohere Inc.*²¹, the US District Court denied Cohere's partial motion to dismiss, holding that substitutive summaries could constitute copyright infringement, and that the generation of hallucinated articles falsely attributed to publishers/plaintiffs constituted a classic passing-off claim.

Intervenors supporting ANI i. Submissions on behalf of the Intervenor – DNPA (I.A. 2199/2025) 58. Mr. Rajshekhar Rao, senior counsel appearing on behalf of intervenors in I.A. 2199/2025 made the following submissions: 58.1. The LLM models inherently memorize content. This can be implied

19 931 F. Supp. 2d 537 (S.D.N.Y. 2013). 20 case no. 42 O 14139/24, decided on 11 November 2025. 21 *Advance Local Media LLC v. Cohere Inc.*, 1:25-cv-01305, (S.D.N.Y.).

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 25 of 135 MISHRA Signing Date:24.07.2026 17:17:24 from the fact that LLMs can reproduce verbatim copies of unique forms of content on which they are trained. 58.2. The reproduction of copies of training data implies that LLMs retain such data. The submission made on behalf of Open AI that LLMs only keep the non-expressive components of copyrighted works is incorrect. It is also incorrect to suggest that LLM models only learn rules of grammar and syntax. 58.3. The Copyright Act does not distinguish between 'expressive' and 'non-expressive' uses for the purpose of copyright infringement.

ii. Submissions on behalf of the Intervenor - Indian Music Industry (I.A. 4027/2025) 59. Mr. C.M. Lall, senior counsel appearing on behalf of intervenors in I.A. 4027/2025 made the following submissions: 59.1. Where the output amounts to substantial reproduction of original work, the same would amount to copyright infringement. It is well-established that LLM models retain copies of their training data and this is apparent from the admission made by Open AI with regard to 'regurgitation'. 59.2. Accordingly, regurgitation or extraction of training data from the LLM model, which results in a response/ output to a user prompt being a substantial reproduction of the original work, will constitute infringement.

iii. Written Submissions filed by Mr. Ankit Sahni, Advocate on behalf of the Intervenor – Federation of Indian Publishers (FIP) (I.A. 470/2025)

60. ChatGPT regurgitates works on which it has been trained in the outputs/responses generated for its users and this would amount to infringement under the Copyright Act as the outputs/responses are a substantial reproduction of the copyrighted works.

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 26 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 Submissions on behalf of the Defendant (OPEN AI) 61. Mr. Amit Sibal, senior counsel appearing on behalf of Open AI made the following submissions: 61.1. The cut-off date for training an LLM is several months prior to the launch of a model. In respect of the models available to ANI in the lead up to the present suit, the cut-off dates for training were either April 2022 (for GPT 4) or April 2024 (for GPT 4o). Each of the 14 (fourteen) works that ANI alleges have been infringed in the plaint were admittedly authored and published subsequent to April 2024 (in either August or September 2024). It is therefore undisputed that the training data on which Open AI's LLMs were trained did not contain the 14 (fourteen) works that ANI alleges were infringed. 61.2. ANI has not identified any other works that it claims copyright over. Therefore, ANI cannot seek a speculative finding of infringement over unidentified works. 61.3. There is no exclusive right to 'use' any works under Section 14 of the Copyright Act. Therefore, ANI's allegation regarding 'use' of ANI's work by Open AI cannot amount to infringement. The LLMs underlying ChatGPT do not store or reproduce training data. Nor do they access data on which they were trained once the pre-training phase is complete. The only 'use' for which this data is deployed is during the training of Open AI's LLMs underlying ChatGPT. 61.4. ChatGPT is not designed to reproduce the extracts of content forming a part of its training data, but may use information or facts that it has 'learnt' from such content in the course of its training towards responses to the user

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 27 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 prompts. Therefore, ChatGPT cannot be said to 'duplicate' or 'reproduce' the 'manner of presentation' of content on which the models underlying ChatGPT have been trained. 61.5. It is the settled position of law that copyright holder's rights are limited to the protection of the manner in which an idea or a fact is expressed. However, there is no such protection on an underlying idea or fact. This is especially true in the context of news, where the underlying fact is incapable of copyright protection. In fact, there is a predominant public interest in the wide dissemination of such facts. Hence, ANI cannot claim monopoly of use over the said facts. 61.6. Copyright in news and other factual literary content is limited only to the expression thereof. In the case of 'news', the threshold for establishing substantial similarity in the expression is higher. [Ref: Feist Publications v. Rural Telephone Service22]. 61.7. Grammatical rules and syntax by themselves cannot be protected by copyright because they are basic rules for sentence formation. 61.8. Indian copyright law explicitly rejects the 'sweat of the brow' doctrine and has adopted the 'skill and judgment test' to determine whether a work would be entitled to protection under Indian copyright law. [Ref. Eastern Book Company v. D.B. Modak (supra)] 61.9. ANI has sought to present limited extracts of a particular work rather than the whole work to allege infringement. For the purpose of determining copyright infringement, it is not permissible to split or compartmentalize or

22 499 U.S. 340 (1991).

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 28 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 dissect ANI's work into parts, then seek to compare such parts against the alleged infringed work. [Ref. Star India v. Leo Burnett23] 61.10. ANI has not even been able to demonstrate substantial similarity, let alone verbatim reproduction in respect of the limited instances that have been cited in the plaint. This is despite the fact that ANI made extreme efforts to try and obtain a verbatim/ substantially similar response from ChatGPT through a series of adversarial prompts. 61.11. An infringing copy must be a substantial or material reproduction. Where similarities in the two works exist alongside broad dissimilarities, which negate the intention to copy the original and the coincidences appearing in the two works are clearly incidental, no infringement of the copyright comes into existence. [Ref. R.G. Anand v. Deluxe Films24; Bikramjeet Singh Bhullar v. Yash Raj Films25] 61.12. Different ChatGPT users receive different responses based on the same prompts. 61.13. Even in respect of ANI's allegations that the ChatGPT search feature reproduces results from ANI's website, it is submitted that a mere perusal of the alleged response(s) relied upon by ANI makes it apparent that the same are neither identical nor substantially similar to the corresponding articles relied upon by ANI. Further, in such a case, ChatGPT provides a reference/ link to the source but does not reproduce content therefrom.

23 2002 SCC OnLine Bom 942. 24 (1978) 4 SCC 118. 25 (2023) SCC OnLine Del 8212.

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 29 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 61.14. 'Regurgitation' of data on which the LLM models are trained, is a rare glitch that is being actively addressed by Open AI. ANI has not been able to identify any specific instance where 'regurgitation' of their works occurred.

Intervenors supporting OPEN AI i. Submissions on behalf of the Intervenor - Indian Governance and Policy Project (IGAP) (I.A. 4616/2025) 62. Mr. Akhil Sibal, senior counsel appearing on behalf of the intervenor in I.A. 4616/2025 made the following submissions: 62.1. In *Andrea Bartz v. Anthropic*²⁶, the US District Court emphasized that what is relevant is not the amount of data used for making a copy, but the amount and substantiality of what is made accessible to the public by the secondary use. Since LLMs do not regurgitate an author's work or produce substantially similar works, it would not amount to copyright infringement. 62.2. An LLM model does not store the training data and memorization of training data is a rare occurrence. Hence, the display at the output creation stage is a result of manipulation of the system by the user alone.

ii. Submissions on behalf of the Intervenor - Broadband India Forum (BIF) (I.A. 10766/2025) 63. Mr. Kapil Sibal and Mr. Arvind P. Datar, senior counsel appearing on behalf of the intervenor in I.A. 10766/2025 made the following submissions: 63.1. Newspaper reports, including articles which are in the public domain are not reproduced in the output in any material form by LLMs. If there is a prompt on a particular event, the LLM is able to examine the voluminous

²⁶ *Bartz v. Anthropic PBC*, 3:24-cv-05417, (N.D. Cal.).

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 30 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 information available in the public domain and prepare a unique response. Therefore, there is no reproduction in any material form. The LLMs often provide links to the sources that have been accessed. 63.2. Several newspapers, magazines, etc. have created 'paywalls' or 'crawler blockers' to prevent access to their contents. The data from these websites is completely blocked and cannot be accessed. Preparing the responses based on publicly available websites will not amount to violation of Section 14(a)(i) of the Copyright Act, so long as there is no reproduction in material form. LLMs do not reproduce the literary work in any material form. 63.3. LLMs are not designed to materially reproduce any work or a substantial part of it in their response to prompts. They merely understand the ideas, concepts and the facts in those works, which can be used to generate output. It does not amount to adaptation of copyrighted works.

iii. Submissions on behalf of the Intervenor – Flux Labs Pvt. Ltd. (I.A. 4615/2025) 64. Ms. Haripriya Padmanabhan, senior counsel appearing on behalf of intervenor in I.A. 4615/2025 made the following submissions: 64.1. The examples given by ANI in the plaint do not show substantial reproduction of copyrighted works of ANI. 64.2. Unlike any other customer of an AI platform, ANI gave a 'direct prompt' to ChatGPT in the said examples. Yet, the responses given by ChatGPT only summarised the interviews in its own words and referred to the author in third person. 64.3. The AI platforms also provide a reference and do not claim that the

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 31 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 information is theirs. Hence, the output/responses generated by AI platforms cannot be stated to be a violation of copyright under Section 14(a)(i) of the Copyright Act.

Submissions of Mr. Adarsh Ramanujan, Amicus Curiae 65. Mr. Ramanujan made the following submissions: 65.1. The issue whether Open AI's LLMs underlying ChatGPT retain the full 'raw data' and reproduce duplicate or near duplicate extracts of the 'raw data' as alleged by ANI can be determined only after a full trial. However, ANI's submission in this regard does not appear to be prima facie tenable as the function of an LLM is based on patterns and therefore, it would not make sense for an LLM to store raw data on a permanent basis to generate responses for a query. 65.2. In some instances, training of an LLM leads to memorization. This tendency of verbatim memorization occurs when an LLM is trained on basic facts or unique pieces of information, since the training data may not be sufficiently large for the model to build a generalization. Hence, it is possible that training of ChatGPT leads the underlying LLMs to memorize parts of ANI's works. 65.3. In the event memorization occurs, even if inadvertently, the communication of the same to the public at the output stage would amount to infringement. 65.4. The decision of the Supreme Court in *R.G. Anand* (supra), is the leading judgment on the aspect of copyright infringement. In terms of paragraph 46 of *R.G. Anand* (supra), the Supreme Court has clearly stated

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 32 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 that "form, manner and arrangement and expression of the idea" would only be entitled

to copyright protection. 65.5. Retrieval Augmented Generation (RAG)²⁷ is an aspect which has not been pleaded in the plaint and has been raised only in the course of arguments. Therefore, whether RAG involves infringement of ANI's works should also be a subject matter of a trial.

Submissions of Professor Arul George Scaria, Amicus Curiae 66. Prof. Scaria made the following submissions: 66.1. A reading of the judgment by the Supreme Court in *Eastern Book Company v. D.B. Modak* (supra), highlights that India has explicitly discarded the 'sweat of the brow' approach and adopted the 'skill and judgment' standard. Therefore, the current legal standard of originality under Indian copyright law is of a standard higher than the 'sweat of the brow' standard. 66.2. Taking into account the current standard of originality, substantial chunk of news related materials might be outside copyright protection in India. Therefore, issues such as existence of a copyright in favour of ANI and the need for use of news material by LLMs to provide accurate responses can only be determined in a trial. 66.3. The non-expressive use of copyrighted works by LLMs to provide more optimal responses to any user prompts may not constitute copyright infringement. 66.4. Since news is based on facts and when there are limited ways in which

27 As discussed in the Section: Working of LLMs.

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 33 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 a fact can be expressed, in terms of 'merger doctrine' as discussed in *Syndicate of the Press of the University of Cambridge v. B.D. Bhandari*²⁸, news should not get copyright protection or should get only a thin layer of protection.

DISCUSSION AND ANALYSIS 67. Under the Copyright Act, copyright subsists only for the classes of works enlisted under Section 13 of the Copyright Act. ANI claims that news articles and interviews published on its digital platforms are "original literary works" under Section 13(1)(a)²⁹ of the Copyright Act. This position has not been disputed by Open AI. However, Open AI submits that ANI is not the owner of works published on its website in terms of Section 17³⁰ of the Copyright Act. 68. To establish ownership of the works published on its website, ANI has placed on record a sample 'Professional Services Agreement' in terms of which copyright would vest with ANI in respect of original works, such as images, video clippings, articles, reports, interviews, etc. created/authored/obtained by personnel on behalf of ANI.

28 2011 SCC OnLine Del 3215. 29 13. Works in which copyright subsists.—(1) Subject to the provisions of this section and the other provisions of this Act, copyright shall subsist throughout India in the following classes of works, that is to say,— (a) original literary, dramatic, musical and artistic works; 30 17. First owner of copyright.— Subject to the provisions of this Act, the author of a work shall be the first owner of the copyright therein: Provided that— (a) in the case of a literary, dramatic or artistic work made by the author in the course of his employment by the proprietor of a newspaper, magazine or similar periodical under a contract of service or apprenticeship, for the purpose of publication in a newspaper, magazine or similar periodical, the said proprietor shall, in the absence of any agreement to the contrary, be the first owner of the copyright in the work, in so far as the copyright relates to the publication of the work in any newspaper, magazine or similar periodical, or to the reproduction of the work for the purpose of its being so published, but in all other respects the author shall be the first owner of the copyright in the work;

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 34 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 69. On a prima facie view, this Court is satisfied that there is an explicit agreement between ANI and its professionals which recognises ANI as the owner of the copyright subsisting in "original literary works" published on its website in terms of Section 17 of the Copyright Act. 70. It is not disputed by Open AI that "original literary works" that are available free of cost and publicly accessible are nevertheless entitled to copyright protection. Therefore, for adjudicating present application this Court is of the prima facie view that ANI, as an owner, would have copyright over the "original literary works", even if they are freely and publicly available on its website. Any challenge to the ownership of individual works which are subject matter of the present suit can be adjudicated at the stage of trial. 71. As a copyright owner of the original literary works, ANI enjoys certain exclusive rights under Section 14 of the Copyright Act, which include the following: (i) the exclusive right to reproduction of the work in terms of Section (14)(a)(i) of the Copyright Act and, (ii) the right to "communicate it to the public" in terms of Section 14(a)(iii) of the Copyright Act. For the ease of reference, the relevant extracts from Section 14 are set out below:

"14. Meaning of copyright.—For the purposes of this Act, "copyright" means the exclusive right subject to the provisions of this Act, to do or authorise the doing of any of the following acts in respect of a work or any

substantial part thereof, namely:— (a) in the case of a literary, dramatic or musical work, not being a computer programme,—

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 35 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 (i) to reproduce the work in any material form including the storing of it in any medium by electronic means, *** ** (iii) to perform the work in public, or communicate it to the public;”

72. Section 2 (ff)31 of the Copyright Act defines “communication to the public”. 73. Thus, the act of reproduction and communicating the work of ANI to the public by a third party without license would amount to infringement as per Section 5132 of the Copyright Act 74. It is the case of ANI that Open AI’s LLM model memorizes the content of ANI’s articles and specific responses are generated by ChatGPT through memorisation and the original literary works of ANI are communicated to the public. This amounts to infringement under Section 51 of the Copyright Act.

31 (ff) “communication to the public” means making any work or performance available for being seen or heard or otherwise enjoyed by the public directly or by any means of display or diffusion other than by issuing physical copies of it, whether simultaneously or at places and times chosen individually, regardless of whether any member of the public actually sees, hears or otherwise enjoys the work or performance so made available. 32 51. When copyright infringed.—Copyright in a work shall be deemed to be infringed— (a) when any person, without a licence granted by the owner of the Copyright or the Registrar of Copyrights under this Act or in contravention of the conditions of a licence so granted or of any condition imposed by a competent authority under this Act— (i) does anything, the exclusive right to do which is by this Act conferred upon the owner of the copyright, or (ii) permits for profit, any place to be used for the communication of the work to the public where such communication constitutes an infringement of the copyright in the work, unless he was not aware and had no reasonable ground for believing that such communication to the public would be an infringement of copyright; or (b) when any person — (i) makes for sale or hire, or sells or lets for hire, or by way of trade displays or offers for sale or hire, or (ii) distributes either for the purpose of trade or to such an extent as to affect prejudicially the owner of the copyright, or (iii) by way of trade exhibits in public, or (iv) imports into India, any infringing copies of the work: Provided that nothing in sub-clause (iv) shall apply to the import of one copy of any work for the private and domestic use of the importer. Explanation.—For the purposes of this section, the reproduction of a literary, dramatic, musical or artistic work in the form of a cinematograph film shall be deemed to be an “infringing copy”.

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 36 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 75. Therefore, this Court would have to examine whether the responses generated by ChatGPT would result in reproduction and communication of the original literary works of ANI to the public without authorisation from ANI and hence, would amount to infringement. More particularly, the Court would have to examine the following two issues: i. whether Open AI memorizes and regurgitates ANI’s copyrighted literary works in the form of responses and, ii. whether the ChatGPT’s responses are substantial reproduction of ANI’s copyrighted literary works.

i. Whether Open AI memorizes and regurgitates ANI’s copyrighted literary works in the form of responses?

76. It is contended on behalf of ANI and the intervenors supporting ANI that LLM models underlying ChatGPT inherently memorize content and retain copies of their training data which results in regurgitation. According to ANI and intervenors, LLMs are capable of providing verbatim or near- identical reproductions of the training data in response to user prompts, thereby demonstrating that expressive elements of the copyrighted works are also retained and reproduced. 77. Per contra, it has been highlighted on behalf of Open AI as well as intervenors supporting Open AI that an LLM model is not designed to reproduce extracts of the contents that form part of their training data. It was also argued that the LLMs underlying ChatGPT are not trained to duplicate or regurgitate the exact manner of presentation of content on which these models have been trained. Based on a particular prompt, the LLM is able to prepare a unique response based on the voluminous data on which it has been trained.

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 37 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 Further, generative AI models also provide links to the sources that have been accessed by them. In this regard, Open AI has highlighted that the same prompts by different users would result in different responses from the ChatGPT (document 10 in the defendant’s documents filed along with the written statement). 78. Professor Scaria, one of the Amici Curiae, in his written submissions has argued that in vast majority of cases there is minimal memorisation in the working of LLMs and mostly memorisation occurs when specific attempts are made to generate potential copyright infringing outputs. The relevant extracts from the written submissions of Prof. Scaria are set out below: “58. A couple of general observations may be added here.

First, as the technical literature suggests, though LLMs may be using copyrighted training data to generate outputs, in most instances, they may not be generating potentially copyright infringing responses. Scholars like Prof. Matthew Sag points out that this is due to the different steps taken during model training like the process of decomposition and abstraction, blending of latent concepts, and noise injection [Sag, 2023]. However, in some instances, particularly due to the memorisation problem discussed earlier as part of Issue II, there may be outputs which are identical or similar to the input (copyrighted) training data. This is seen mostly during the so-called “extraction attacks”, wherein specific attempts are made to generate potentially copyright infringing outputs. Some of the empirical studies suggest that memorisation was seen in around 0.03% of sample images during extraction attacks and they were successful mostly in instances wherein the images were duplicated more than 100 times [Sag, 2023]. This also suggests that in the vast majority of use scenarios, the use of copyrighted materials for generating responses are non-infringing, because only non-expressive uses are made of the copyrighted materials.” [Emphasis supplied]

79. Mr. Adarsh Ramanujan in his written submissions has argued that

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 38 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 ANI's contention that Open AI retains the full raw data is prima facie untenable as an LLM is merely a set of mathematical functions to connect input to output. An LLM's function is to predict based on patterns. The chances of memorization of the raw data may increase when an LLM Model is trained on basic facts or unique pieces of information. This tendency occurs because there may be absence of general patterns leading to direct reproduction at the time of output generation. 80. The US District Court for Northern District of California in Bartz v. Anthropic (supra) examined the issue of memorization in LLMs and regurgitation of responses from its training data and made the following observations: “Either way Meta added the books it downloaded to the datasets it used to train the Llama models. It is also post-trained its models to prevent them from “memorizing and outputting certain text from their training data, including copyrighted material. These training efforts, which Meta calls “mitigations,” appear to have been successful. Meta's expert witness tested them using a method designed to get LLMs to regurgitate material from its training data (which Meta calls “adversarial prompting”). Even using that method, the expert could get no model to generate more than 50 words and punctuation marks (that is, “tokens”) from the plaintiffs’ books. And the plaintiffs’ expert could only get the Llama model best at regurgitation to generate 50 words and punctuation marks from the plaintiffs’ books in 60% of tests. She also testified that Llama was not able to reproduce “any significant percentage” of them. Meta MSJ Ex.24 at 237:16-19:see also Pls Ex. 79, 70-72, 79, 82-83, 92:Meta MSJ Ex.23 at 179:22-25, 180:17-181:16. In short, Llama cannot currently be used to read or other wise meaningfully access the plaintiffs’ books.” [Emphasis supplied]

81. From the aforesaid observations, it appears that even the plaintiffs’ expert in the field of AI could only make the LLM Model regurgitate

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 39 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 insignificant amounts from the copyrighted works of the plaintiffs. 82. ANI's case is based on ‘memorisation’, whereby the LLM memorises the training data verbatim or substantially. In the plaint, ANI has given nine illustrations to demonstrate substantial reproduction. It is an undisputed position that the training of the LLMs underlying ChatGPT ended in April 2022 (for GPT 4) or April 2024 (for GPT 4o), which was much prior to the publication of the illustrative articles used in the plaint, to allege reproduction of ANI's copyrighted works. The table below illustrates this aspect: TITLE(S) OF THE ARTICLE(S) DATE OF PUBLICATION

“My father was forced...”: Omar 6th September 2024 Abdullah opens up on IC 814 hijack incident, reveals difficult decisions faced by his father T20 World Cup: “Pakistan never a 30th September 2024 challenge in Women's cricket”: Ex- player Reema Malhotra backs India

“Lord Balaji will not forgive him 26th September 2024 even if he climbs up the seven hills on his knees”: TDP spokesperson slams Jagan Mohan Reddy

“Connecting this to UP not right,” 26th September 2024 Himachal AICC in-charge Rajeev Shukla defends state govt on ID cards for vendors “It's gone underground...”: JKPC 10th September 2024 chairman Sajad Gani Lone on ‘Azadi’ sentiments in Kashmir

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 40 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 “Jinnah like mentality,” Minister 17th September 2024 Hardeep Puri accuses Rahul Gandhi of crossing limit for “political interest” Neeraj Chopra's mother celebrates 9th August 2024 Olympic Silver;

praises Pakistan's Arshad Nadeem for bagging Gold

"Ashwin best in India, Lyon better 23rd September 2024 overall": Monty Panesar sets-up best Test spinner debate

83. Therefore, it is apparent that the illustrations of ChatGPT responses relied upon in the plaint could not have been produced on account of training of Open AI's LLMs on the ANI's literary works, as the literary works in question were published only after the completion of the training process. 84. As discussed earlier under the Section 'Working of LLMs', it could be inferred that when an LLM refers to information on which it was not trained, it is using RAG technique. Using the RAG technique, LLMs retrieve relevant information from external sources, in this case ANI's website, before generating a response. Relevant data retrieved from the website, which is contextually aligned with the user's prompt, is then generated as responses grounded in the website's data rather than relying solely on its pre-trained knowledge. Therefore, responses that are not based on the trained data cannot be used against Open AI in respect of claims based on memorisation. Such allegations must be tested based on the identical or substantially similar reproduction of the trained data. Therefore, this Court is of the prima facie view that the instances given in the plaint alleging infringement are not a

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 41 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 result of memorisation, rather they are in the nature of live links, perhaps reflecting RAG technique. Whether outputs produced using RAG would amount to copyright infringement is an aspect which has not been pleaded in the plaint, though it was referred during the course of submissions.

ii. Whether ChatGPT's responses are substantial reproduction of ANI's copyrighted literary works?

85. To begin with, it may be relevant to discuss legal principles regarding copyright infringement elucidated in some of the instructive judgments of the Supreme Court that have been relied upon by both sides. 86. In R.G. Anand (supra), the plaintiff sued the defendants for damages claiming infringement of the copyright of the plaintiff in a dramatic play, which was made into a motion picture by the defendants. The Supreme Court held that an idea, principle, theme or facts cannot be subject matter of copyright protection. Where two writers write on the same subject, similarities are bound to occur because the central idea of both are the same. Therefore, the Court has to determine whether the defendant has only adopted the idea of the copyrighted work or has also adopted the form, manner, arrangement and expression of an idea with minor changes. If the defendant's work appears to be a transparent rephrasing or a copy of the substantial and material part of the original, it would amount to copyright infringement. The Supreme Court in paragraph 46 of the judgment has laid down the following principles to determine infringement of copyright, which are set out below: "46. Thus, on a careful consideration and elucidation of the various authorities and the case law on the subject discussed above, the following propositions emerge:

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 42 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 1. There can be no copyright in an idea, subject-matter, themes, plots or historical or legendary facts and violation of the copyright in such cases is confined to the form, manner and arrangement and expression of the idea by the author of the copyrighted work. 2. Where the same idea is being developed in a different manner, it is manifest that the source being common, similarities are bound to occur. In such a case the courts should determine whether or not the similarities are on fundamental or substantial aspects of the mode of expression adopted in the copyrighted work. If the defendant's work is nothing but a literal imitation of the copyrighted work with some variations here and there it would amount to violation of the copyright. In other words, in order to be actionable the copy must be a substantial and material one which at once leads to the conclusion that the defendant is guilty of an act of piracy. 3. One of the surest and the safest test to determine whether or not there has been a violation of copyright is to see if the reader, spectator or the viewer after having read or seen both the works is clearly of the opinion and gets an unmistakable impression that the subsequent work appears to be a copy of the original. 4. Where the theme is the same but is presented and treated differently so that the subsequent work becomes a completely new work, no question of violation of copyright arises. 5. Where however apart from the similarities appearing in the two works there are also material and broad dissimilarities which negative the intention to copy the original and the coincidences appearing in the two works are clearly incidental no infringement of the copyright comes into existence. 6. As a violation of copyright amounts to an act of piracy it must be proved by clear and cogent evidence after applying the various tests laid down by the case-law discussed above. 7. Where however the question is of the violation of the copyright of stage play by a film producer or a director the task of the plaintiff becomes more difficult to prove piracy. It is manifest that unlike a stage play a film has a much broader prospective, wider field and a bigger background where the defendants can by introducing a variety of incidents give a colour and complexion different from the manner in which the copyrighted work has

expressed the idea. Even so, if the viewer after seeing the film gets a totality of impression that the film is by and large a copy of the original play, violation of the copyright may be said to be proved.” [Emphasis supplied]

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 43 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 87. In *Eastern Book Company v. D.B. Modak* (supra), the appellants/plaintiffs filed a suit along with an application for temporary injunction seeking to restrain the respondents/defendants from reproducing the plaintiffs’ version of copy-edited judgments and orders of the Supreme Court on CD-ROMs. The defendants were selling software packages on CD- ROMs with the text of the copy-edited judgments of the Supreme Court along with headnotes prepared by the plaintiffs for its own law report, ‘SCC’. The Supreme Court, relying upon the judgment of the Supreme Court of Canada in *CCH Canadian v. Law Society of Upper Canada*³³ held that ‘sweat of the brow’ approach to originality is not the correct standard as it fails to protect the public interest in production and dissemination of works. On the other hand, the creativity standard of originality as applied by the US Courts, is a standard of originality with a higher threshold. The underlying impetus for this modification was the need to balance the rights of the copyright owner with the public interest. 88. In paragraph 15 of the *Eastern Book Company v. D.B. Modak* (supra), it was specifically observed that for a literary work to be original and to claim copyright, it should possess some amount of creativity in the work, more than mere facts. For ease of reference, the aforesaid paragraph is set out below: “15. Copyright is a right to stop others from exploiting the work without the consent or assent of the owner of the copyright. A copyright law presents a balance between the interests and rights of the author and that of the public in protecting the public domain, or to claim the copyright and protect it under the copyright statute. One of the key requirements is that of originality which contributes, and has a direct nexus, in maintaining the interests of the author as well as that of

33 (2004) 1 R.C.S. 339.

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 44 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 public in protecting the matters in public domain. It is a well-accepted principle of copyright law that there is no copyright in the facts per se, as the facts are not created nor have they originated with the author of any work which embodies these facts. The issue of copyright is closely connected to that of commercial viability, and commercial consequences and implications.” [Emphasis supplied]

89. Accordingly, the Supreme Court laid down the ‘skill and judgment test’ as opposed to ‘sweat of the brow’ test that was applied earlier in India. In terms of the ‘skill and judgment test’, the author of a derivative work has to satisfy that the work has been produced from exercise of skill and judgment and, the skill and judgment should not be of a trivial nature. It was observed that the ‘sweat of the brow’ doctrine creates a monopoly in materials available in public domain without encouraging the creation of writings by the authors. In this regard, the relevant paragraph of the *Eastern Book Company v. D.B. Modak* (supra) is set out below:

“57. The Copyright Act is not concerned with the original idea but with the expression of thought. Copyright has nothing to do with originality or literary merit. Copyrighted material is that what is created by the author by his own skill, labour and investment of capital, maybe it is a derivative work which gives a flavour of creativity. The copyright work which comes into being should be original in the sense that by virtue of selection, coordination or arrangement of pre-existing data contained in the work, a work somewhat different in character is produced by the author. On the face of the provisions of the Copyright Act, 1957, we think that the principle laid down by the Canadian Court would be applicable in copyright of the judgments of the Apex Court. We make it clear that the decision of ours would be confined to the judgments of the courts which are in the public domain as by virtue of Section 52 of the Act there is no copyright in the original text of the judgments. To claim copyright in a compilation, the author must produce the material with exercise of

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 45 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 his skill and judgment which may not be creativity in the sense that it is novel or non-obvious, but at the same time it is not a product of merely labour and capital. The derivative work produced by the author must have some distinguishable features and flavour to raw text of the judgments delivered by the court. The trivial variation or inputs put in the judgment would not satisfy the test of copyright of an author.” [Emphasis supplied]

90. The legal position that emerges from *Eastern Book Company v. D.B. Modak* (supra) is that the work which is produced by the author by using his skill and judgement with a flavour of creativity is copyrightable. The protection would be only in respect of creative and original works. There cannot be any copyright protection in respect of facts even if someone has put in effort to obtain those facts. 91. In *Feist Publications* (supra), the US Supreme

Court laid down the test of 'modicum of creativity'. In this case, a telephone utility company brought a copyright infringement action against the publisher of a telephone directory. The US Supreme Court held that no one can claim originality as to the facts. A person who reports a particular fact has not created the said fact but has only discovered its existence. The relevant paragraphs of *Feist Publications* (supra) are set out below:

"[15] It is this bedrock principle of copyright that mandates the law's seemingly disparate treatment of facts and factual compilations. "No one may claim originality as to facts." *Id.*, § 2.11[A], p. 2-157. This is because facts do not owe their origin to an act of authorship. The distinction is one between creation and discovery : the first person to find and report a particular fact has not created the fact; he or she has merely discovered its existence. To borrow from *Burrow-Giles*, one who discovers a fact is not its "maker" or "originator." 111 U.S., at 58. "The discoverer merely finds and records." *Nimmer* § 2.03[E]. *Census-*

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 46 of 135 MISHRA Signing

Date:24.07.2026 17:17:24 takers, for example, do not "create" the population figures that emerge from their efforts; in a sense, they copy these figures from the world around them. *Denicola*, *Copyright in Collections of Facts : A Theory for the Protection of Nonfiction Literary Works*, 81 Colum. L. Rev. 516, 525 (1981) (hereinafter *Denicola*). Census data therefore do not trigger copyright because these data are not "original" in the constitutional sense. *Nimmer* [p*348] § 2.03 [E]. The same is true of all facts -- scientific, historical, biographical, and news of the day. "They may not be copyrighted and are part of the public domain available to every person." *Miller*, supra, at 1369. *** ** [18] This inevitably means that the copyright in a factual compilation is thin. Notwithstanding a valid copyright, a subsequent compiler remains free to use the facts contained in another's publication to aid in preparing a competing work, so long as the competing work does not feature the same selection and arrangement. As one commentator explains it:"No matter how much original authorship the work displays, the facts and ideas it exposes are free for the taking.... The very same facts and ideas may be divorced from the context imposed by the author, and restated or reshuffled by second comers, even if the author was the first to discover the facts or to propose the ideas." "Ginsburg 1868." [Emphasis supplied]

92. ANI has relied upon judgment of the House of Lords in *Walter v. Lane*³⁴ delivered in 1900, to show that it has copyright in respect of the news circulated by its platform. In *Walter v. Lane* (supra), the issue before the House of Lords was whether a person who makes notes of speech delivered in public, transcribes the same and publishes in the newspaper would be entitled to copyright protection. The House of Lords held that each of the reporters reporting the speech would have a copyright in his own published report. The House of Lords came to this conclusion based on the doctrine of

34 [1900] AC 539.

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 47 of 135 MISHRA Signing

Date:24.07.2026 17:17:24 'sweat of the brow' and it was premised on the principle that an author should be able to reap the benefits of its labour. 93. On behalf of Open AI, the aforesaid judgment is sought to be distinguished as it was in the context of copyright laws that existed in England in 1842. The then copyright laws in England did not require originality. 94. *Walter v. Lane* (supra) was considered by the Supreme Court in *Eastern Book Company v. D.B. Modak* (supra), as the appellant therein relied upon the same, and it was held that the doctrine of 'sweat of the brow' does not hold good anymore as it gives more weight to the copyright owner's right as opposed to the public interest. 95. In *B.D. Bhandari* (supra), the Division Bench delved on the 'merger doctrine' and noted that when there is limited manner in which a fact can be expressed, the facts would merge with the expression. The facts and the expression would become so intrinsically connected that it would be difficult to distinguish one from another, making the expression itself non- copyrightable. For ease of reference, the relevant paragraphs of *B.D. Bhandari* (supra) are set out below: "106. While answering the first question, the learned Single Judge expanded the "the Idea-Expression Dichotomy and the Doctrine of Merger" and noted that there is no copyright protection on ideas and facts and it is only the creative expression of such ideas and facts that is rewarded by law. In the process, expression "originality" in artistic, literary, dramatic and musical works were also explained holding that with the passage of time, doctrine of "sweat of the brow" had given way to the "modicum of creativity" and therefore, in order to have a copyright in a work, it was necessary to show that it involves a certain degree of creativity. 107. The learned Single Judge also deemed it proper to refer to the "doctrine of merger", peculiar to copyright law in the following manner:

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 48 of 135 MISHRA Signing

Date:24.07.2026 17:17:24 "18. An aspect peculiar to copyright law, is the doctrine of merger, is involved in this case. This doctrine posits that where the idea and expression are intrinsically connected, and that the expression

is indistinguishable from the idea, copyright protection cannot be granted. Applying this doctrine courts have refused to protect the expression of an idea that can be expressed only one manner, or in a very restricted manner, because doing so would confer monopoly on the idea itself. The decision in *Herbert Rosenthal Jewelry Corporation v. Kalpakian*, 446 F.2d 738 (1971) is illustrative in this regard. In that case the plaintiffs sued the defendants asking them to refrain from manufacturing bee shaped jewel pins. The Court held that the jewel shaped bee pin was a an idea that anyone was free to copy, the expression of which could be possible only in a few ways, therefore, no copyright could subsist in it.” [Emphasis supplied]

96. In *Akute Internet Services v. Star India*³⁵, Star India was granted exclusive broadcasting rights for a cricket match organised by BCCI including mobile distribution rights. The defendants/appellants were disseminating match information/score alerts without obtaining a license from the Star India (plaintiff/respondent). In the said case, the plaintiff was seeking copyright protection over facts that were available in public domain. The Division Bench of this Court held that there cannot be any copyright protection over facts and information that is available in public. 97. From a reading of the aforesaid judicial precedents, it is clear that there cannot be any copyright in facts. Therefore, in the context of news, copyright would subsist only in the form and manner of expression of news and not on the underlying facts. What has to be seen is whether the expression used by

35 2013 SCC OnLine Del 3344.

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 49 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 Open AI in its ChatGPT responses is materially similar to the expression used by ANI in its news articles resulting in reproduction and making the work available to the public in terms of Sections 14(a) (i) and 14(a)(iii) of the Copyright Act. ANI cannot claim any copyright in the facts underlying the news articles. 98. ANI has given nine instances of substantial reproduction of ANI's works by Open AI (paragraph 33.1 to 33.9 of the plaint). ANI claims that the aforesaid nine instances have been given only as an illustration and there are many more instances of copyright infringement by Open AI. All the nine instances given in the plaint are in respect of news/interviews which were originally published by ANI. 99. During oral submissions, ANI had highlighted the similarities between the works of ANI and the response produced by ChatGPT in paragraph 33.7 of the plaint, which dealt with an interview of the mother of Neeraj Chopra who had secured a silver medal at the Olympics, 2024 in Men's Javelin Throw. Admittedly, the said interview was given in Hindi and was translated into English by ANI for reporting. A comparison given by ANI in paragraph 33.7 of the plaint is set out below:

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 50 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 *The colour highlighting has been done for ease of comparison

100. As per ANI, the response produced by ChatGPT shows that ChatGPT

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 51 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 has produced exact or nearly exact copy of ANI's news article. The original interview and responses were in Hindi which were translated into English by ANI. Open AI has also infringed ANI's exclusive rights to translate the works for which ANI invested significant time and effort. 101. What emerges from a reading of the aforesaid extracts is that ChatGPT generated this output in response to two different prompts. The first prompt was: “Prompt: What did Neeraj Chopra's mother tell the media about the Pakistani gold medallist?”

102. A comparison of the response given by ChatGPT to the first prompt as highlighted above (in blue) with what was reported by ANI as highlighted above (in orange) shows that the response given by ChatGPT was not a substantial reproduction or nearly exact copy, though the essence of the response is similar as it based on facts i.e. a particular interview of Neeraj Chopra's mother. ChatGPT has added its own commentary to the news article published by ANI, which bears its own expression and is quite dissimilar to the expression used in ANI's article. 103. Realising that the response was not a substantial reproduction, ANI gave another prompt in the following terms: “Prompt : Please try to tell me exactly what she said” 104. To be noted, in the second prompt, ANI used the word 'exactly', which clearly shows that ANI was seeking to elicit a response from ChatGPT in exact terms i.e. an adversarial prompt (“carefully designed inputs that

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 52 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 manipulate model outputs”³⁶). In other words, deliberately a prompt is being put in a manner so as to elicit a specific response to show that ChatGPT has substantially reproduced ANI's works. 105. Upon the said prompt, ChatGPT came out with a quote of what was stated by Neeraj Chopra's mother as highlighted above (in orange). The only similarity which emerges from the aforesaid comparison is relating to a

part of a quote from Neeraj Chopra's mother in the interview. However, ChatGPT's response has given its own context to the aforesaid quotes from Neeraj Chopra's mother's interview. In addition, a subsequent paragraph was added (as highlighted in green) completely on its own by ChatGPT to explain the quoted text which was not a part of ANI's news report. 106. In the present case, ANI has only demonstrated that Open AI has extracted a quote from an interview accompanied by its own commentary. According to Section 17(cc)37 of the Copyright Act, the author of an address or speech made in public is the person who made it. The news article gives the quote of what was said by the Neeraj Chopra's mother in the video. Applying the underlying principle of Section 17(cc) of the Copyright Act to the facts of the present case, in the prima facie view of this Court, Neeraj

36 Ahmet Emre Ergün, Aytuğ Onan, Adversarial Prompt Detection in Large Language Models: A Classification-Driven Approach, Computers, Materials and Continua, Volume 83, Issue 3, 2025, Pages 4855- 4877 ISSN 1546-2218, <https://doi.org/10.32604/cmc.2025.063826>. (<https://www.sciencedirect.com/science/article/pii/S1546221825004898>). 37 17. First owner of copyright.—Subject to the provisions of this Act, the author of a work shall be the first owner of the copyright therein: Provided that— ... (cc) in the case of any address or speech delivered in public, the person who has delivered such address or speech or if such person has delivered such address or speech on behalf of any other person, such other person shall be the first owner of the copyright therein notwithstanding that the person who delivers such address or speech, or, as the case may be, the person on whose behalf such address or speech is delivered, is employed by any other person who arranges such address or speech or on whose behalf or premises such address or speech is delivered;

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 53 of 135 MISHRA Signing Date:24.07.2026 17:17:24 Chopra's mother would be first owner of the copyright in the said quotation. Similarly, in terms of Section 14(a)(vii) of the Copyright Act, the right to translate would also vest with Neeraj Chopra's mother. There is nothing produced by ANI to show that Neeraj Chopra's mother has assigned this ownership in favour of ANI. Therefore, in the prima facie opinion of this Court, the reproduction of quotes of Neeraj Chopra's mother by ChatGPT in its responses would not amount to infringement of ANI's copyright. 107. Pertinently, the original article of ANI (page 54, document no.11 in the plaintiff's documents filed along with the plaint) with which the comparison has been made, was much longer than the text which was quoted in paragraph 33.7 of the plaint for the purposes of comparison. The response given by ChatGPT does not reproduce the remaining text which is a part of the original article. 108. In R.G. Anand (supra), the Supreme Court has held that if a comparison has to be made between two works it should be made for the works as a whole and not selected parts of the works. 109. In this regard, Open AI has relied upon the judgment of Bombay High Court in Leo Burnett (supra). In the said judgment, the Bombay High Court held that a work in which the plaintiff's claim copyright has to be compared as a whole and not in parts in order to determine substantial copying. It is not permissible to split or dissect the plaintiff's work into parts and then compare the same with the allegedly infringing work. Paragraph 26 of the said judgment is set out below: "26. The issue then is whether if the whole of the respective works is compared and not only parts, factually or materially as it stands, is there copying or substantial copying. I am not at this stage also not considering whether the "work" as an episode or all the episodes which

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 54 of 135 MISHRA Signing Date:24.07.2026 17:17:24 have to be considered in totality. Even for arguments sake if an episode is considered as a 'work', still what is to be compared is the whole work as a single entity. It is not permissible to split or compartmentalize or dissect the plaintiffs' work into parts, then seek to compare such parts against the alleged infringed work. What is to be seen is the entirety of each work. It may be made clear that I am not considering an issue whether a particular part of the film or a scene in the film which may have a striking resemblance, if copied, would amount to a substantial copying of the film. For this purpose reference may be made to the decision in (Francis Day and Hunter Limited v. Twentieth Century Fox Corporation Limited), 1940 A.C. 112 (Privy Council); (Ladbroke (Football) Limited v. William Hill (Football) Limited), 1964 (1) W.L.R. 273; (Merchandising Corporation of America Inc. v. Harpbond Limited), 1983 F.S.R. 32; R.G. Anand v. Delux Films supra and (Fritco Lay Limited v. Uncle Chips Private Limited), 2000 P.T.C. 341. Applying the tests on the material as it stands there is no substantial copying on facts." [Emphasis supplied]

110. The aforesaid findings have been followed by Coordinate Bench of this Court in Bikramjeet Singh (supra), wherein it was observed that rival works have to be compared as a whole and as a single entity. It is not possible to split or dissect the works into parts and then compare them. 111. During the course of oral submissions, ANI has not addressed on any of the other illustrations in paragraph 33 of the plaint. Hence, I do not propose to examine the same. Counsel on behalf of Open AI during the course of submissions has handed over a note to highlight that in none of the instances cited by ANI in paragraph 33 of the plaint, ChatGPT has given responses

identical or substantially similar to ANI's articles. 111.1. The said note has been taken on record. 111.2. The contents of the said note have not been rebutted on behalf of ANI. 112. In paragraph 33.9 of the plaint, ANI has stated that on a specific prompt

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 55 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 about the latest news published by ANI, ChatGPT has provided updates from ANI's website. For the sake of convenience, the screenshot of the specific prompt which has been given in paragraph 33.9 of the plaint has been transcribed below: "Prompt - "Give me the latest news updates from ANI News.in

Response - Here are some of the latest news updates from ANI News:

1. In a legal development, AAP leader Durgesh Pathak has been granted bail by the Rouse Avenue court (ANI News). 2. Indian stock markets saw a positive opening, buoyed by a global rally after U.S. inflation data indicated easing in August (ANI News). 3. The Union Cabinet has approved Phase 3 of the Bangalore Metro Rail Project, which will include two elevated corridors covering 44.65 km with 31 stations (ANI News). 4. Wrestler Bajrang Punia has moved a petition to the Delhi High Court, leading to a notice being issued to the National Anti- doping Agency (ANI News)." For more detailed updates, you can visit ANI News."

113. ANI has also filed the corresponding news stories from its website (pages 32 to 46 in the documents filed by the plaintiff). The table below gives a comparison of the title of ANI's articles and the news updates generated by ChatGPT based on the prompt set out above. TITLE(S) OF ARTICLE(S) PUBLISHED NEWS UPDATES GENERATED BY BY ANI CHATGPT Excise Police case: Delhi Court In a legal development, AAP leader grants bail to Durgesh Pathak, Durgesh Pathak has been granted Arvind Kejriwal's custody extended bail by the Rouse Avenue Court till September 25 (ANI News).

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 56 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 Indian indices opened in green after Indian stock markets saw a positive ease in US inflation opening, buoyed by a global rally after U.S. inflation data indicated easing in August (ANI News). Cabinet approves two corridors of The Union Cabinet has approved Bangalore Metro Rail Project Phase- Phase 3 of the Bangalore Metro Rail 3 project for 44.65 km with 31 Project, which will include two stations elevated corridors covering 44.65 km with 31 stations (ANI News). Delhi HC issues notice to NADA on Wrestler Bajrang Punia has moved a Bajrang Punia's plea challenging his petition to the Delhi High Court suspension leading to a notice being issued to the National Anti-Doping Agency (ANI News).

114. A comparison above shows that even the titles in the response generated by ChatGPT are substantially different from the titles of the articles on ANI's website. 115. Applying the test for infringement laid down in R.G. Anand (supra), in my prima facie view, when the articles of ANI are compared with the responses generated by ChatGPT as a whole, there is no substantial similarity. 116. Now, I shall proceed to deal with the judgments relied upon by ANI. 117. In GEMA v. Open AI (supra), the plaintiff claimed copyright infringement in respect of German song lyrics that were memorised and reproduced by the LLMs underlying ChatGPT. The Munich Regional Court observed that the plaintiff therein had expressly excluded the 'online search function' and in those circumstances a finding of fact was given by the Court that memorisation had occurred as exact song lyrics were reproduced on the basis of non-adversarial prompts. The defendant therein (Open AI) was held to be liable for generating outputs that are identical/substantially similar to the

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 57 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 copyrighted lyrics. 118. In GEMA v. Open AI (supra), the Court found that there were repeated verbatim reproduction of lyrics by ChatGPT despite non-adversarial prompts. In that context, the Court held that there was memorisation. 119. As noted above, in the present case, the training was completed before the illustrative articles in the plaint were published by ANI and therefore, there cannot be any question of memorization. Secondly, the Munich Court noted that the responses generated were based on non-adversarial prompts given by the user. In the present case, ANI has given fairly detailed, repeated prompts and has even used the term 'exactly', to extract copyrighted contents from the LLM models underlying ChatGPT, which cannot be termed as non- adversarial. Even after giving such adversarial prompts, ANI was not able to eke out a response which can be termed as 'substantial' reproduction of its copyrighted material. 120. In Associated Press v. Meltwater (supra), the defendant-Meltwater was a news-monitoring agency which scraped the articles published by the plaintiff-Associated Press using web crawlers and provided excerpts from these articles to its subscribers through emails or through online account on Meltwater's website. This was not a case of data being scraped by an LLM. The Court in Associated Press v. Meltwater (supra) specifically held that there were 33 articles which were

reproduced verbatim by the defendant. In the present case, as noted above, there is no substantial reproduction of ANI's works. 121. In Cohere (supra), the plaintiffs, publishers of news and magazines in the USA, claimed that the defendant's LLM Models were using the copyrighted works of the plaintiffs for training and were

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 58 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 reproducing/delivering verbatim copies of the plaintiffs' works using the RAG function. This judgment was in the context of motion to dismiss, which is similar to Order VII Rule 11 of the CPC in India. In the said case, the plaintiffs had given 75 instances of copying, out of which, 50 were alleged to be verbatim. Hence, the motion to dismiss was denied. 122. ANI has also placed reliance on the judgment of the United States Court of Appeal for the Fifth Circuit in Positive Black Talk v. Cash Money Records³⁸ in support of its submission that copyright infringement is made out if there is 'probative similarity' between the plaintiff's work and the work of the defendant. However, this decision was in the context of copyright infringement in respect of a song and hence, the observations made by the Court were specific to the said factual background. The application of a standard for determining whether there is copyright infringement in respect of a song would be quite different from its application in the context of news. The creativity involved in writing lyrics of a song would ordinarily not be present to the same extent in a news article as the fundamental purpose of a news article is to report events that have actually occurred. Therefore, threshold of establishing substantial similarity in expression would be higher in the case of news. Therefore, this judgment would not be of much assistance to ANI in the present case. 123. The judgments cited on behalf of ANI and the intervenor-DNPA, Infopaq International v. Danke Dagblades Forening³⁹ and Newspaper Licensing Agency v. Meltwater Holding⁴⁰, would have little relevance in the

38 394 F.3d 357, 368 (5th Cir. 2004). 39 ECLI:EU:C:2009:465. 40 [2010] EWHC 3099 (Ch).

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 59 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 present case. In both the aforesaid judgments, it was held that the defendants were providing verbatim extracts of the plaintiff's works. In Meltwater (supra), it was held that the defendants have produced extracts and headlines of news articles of the plaintiff therein and had accordingly infringed the copyright in the news articles. Same was the position in Infopaq (supra) where the defendant had copied extracts of news articles of the plaintiff. In the present case, as noted above, there is no substantial reproduction of ANI's works.

FINDINGS 124. As highlighted above, the illustrations given in the plaint are post the training of Open AI's LLMs and a case for memorization of ANI's works on the basis of the said illustrations cannot be made out. Therefore, at this prima facie stage, the contention of ANI that Open AI permanently stores the training data in order to memorize and regurgitate ANI's works cannot be accepted. At best, these are disputed questions which can only be determined during trial upon parties leading evidence. 125. Based on discussion above it cannot be concluded that the responses produced by ChatGPT are a substantial reproduction of the news articles of ANI. Hence, at a prima facie stage, ANI has failed to make out a case of copyright infringement based on the responses generated by ChatGPT i.e. the output claim. 126. As noted earlier, Issue nos. 1 and 3 are intertwined and have to be considered together. Hence, I now proceed to deal with Issue nos. 1 and 3.

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 60 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 Issue No.1 Whether the storage by the defendants of plaintiff's data (which is in the nature of news and is claimed to be protected under the Copyright Act, 1957) for training its software i.e. ChatGPT, would amount to infringement of plaintiff's copyright

SUBMISSIONS Submissions on behalf of the Plaintiff (ANI) 127. Mr. Sidhant Kumar, counsel appearing on behalf of ANI made the following submissions: 127.1. Open AI collects and stores the 'raw data' as a repository of ANI's works in their original form. The 'raw data' is processed further and filtered. The process of filtration is impossible without storage of the copyrighted works of ANI. 127.2. Tokenisation of raw data is nothing but numerical representation of raw data which belongs to ANI. A token may constitute an individual character, whole words or even parts of words. The LLM decodes these tokens and reconstructs the raw data to generate responses. 127.3. In the second step, the tokens are encoded into vectors, which are thereafter stored in the memory of the LLM to create a vector database. Even if it is assumed that the vector database is an adaptation, such reproduction in adapted form also results in infringement of copyrighted works in terms of Section 2(a)(v) of the Copyright Act. 127.4. Training of LLM inherently involves storage and reproduction of original works. The entire processes undertaken by Open AI including storage

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 61 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 of data, tokenisation, vectorisation amounts to copyright violation in terms of Section

51 read with Section 14 of the Copyright Act. 127.5. What Open AI uses in training is only the expression used by ANI and not the actual facts that occurred. Therefore, training occurs in respect of expression of language. 127.6. Once the training data is fed into Open AI's LLM during the training process, it cannot be removed and continues to be reflected in the outputs generated by the LLM.

Intervenors supporting ANI i. Submission on behalf of Intervenor – Indian Music Industry (I.A. 4027/2025) 128. Mr. C.M. Lall, senior counsel appearing on behalf of intervenors in I.A. 4027/2025 made the following submissions: 128.1. The right to reproduce a work under Section 14(a)(i) of the Copyright Act includes the right to store the work. Therefore, scraping, collection, coalition and processing of data for training purposes constitutes infringement. 128.2. Just because copyrighted works are available on online resources/websites would not mean that these works can be copied/reproduced without a license or authorization. 128.3. Regardless of the purpose of the activity, storage of substantial part of copyrighted work amounts to infringement.

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 62 of 135 MISHRA Signing Date:24.07.2026 17:17:24 ii. Written Submissions filed by Mr. Ankit Sahni, Advocate on behalf of Intervenor – Federation of Indian Publishers (I.A. 470/2025) 129. Open AI admits that collection and compilation of raw data into training data sets involves creation of copies of the training data. The various activities carried out by Open AI makes it clear that Open AI not only collects data, it also stores the same for training its LLM. 129.1. Open AI's act of copying/storing copyrighted works for purposes of text and data mining/text and data analysis amounts to infringement since there is no specific provision in the Copyright Act creating an exception for the same.

iii. Submissions on behalf of Intervenor – Digital News Publishers Association & Ors. (I.A. 2199/2025) 130. Mr. Rajshekhar Rao, senior counsel appearing on behalf of intervenors in I.A. 2199/2025 made the following submissions: 130.1. By using the words “any medium” and “any material form”, Parliament has made it clear that Section 14(a)(i) of the Copyright Act must not be diluted. Merely because protected content is not perceivable by naked eye or meant for human consumption would not take it outside the ambit of Section 14 of the Copyright Act. There is no requirement in Indian copyright law for the copyrighted works to be stored in a tangible medium. Therefore, storage of literary works in digital formats and their reproduction in the shape of vectors/tokens would fall within the sweep of the expression “reproduction/storage” in “any medium” or in “any material form”. 130.2. The scheme of the Copyright Act does not distinguish between expressive and non-expressive uses. The very same argument raised by the

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 63 of 135 MISHRA Signing Date:24.07.2026 17:17:24 defendant in *Richard Kadrey v. Meta Platforms*⁴¹ was rejected by the United States District Court. Even in *Bartz v. Anthropic* (supra), it was noted that “literary works were chosen for their expressive qualities in building a central library and then in training specific LLMs”. 130.3. Even the intermediate step of introducing and storing data for training cannot be said to be non-infringing as substantial portions of copyrighted works are extracted. Once again, this view finds support from the judgment in *Richard Kadrey v. Meta Platforms* (supra).

Submissions of the Defendant (OPEN AI) 131. Mr. Amit Sibal, senior counsel appearing on behalf of Open AI made the following submissions: 131.1. It is an admitted position that storage of data takes place only for the purposes of training. Creation of copy during the training process is an intermediate step. Once the training process is over, an LLM does not have access to data. 131.2. Purposive interpretation should be given by the Courts to the provisions of Copyright Act. Accordingly, it is submitted that Section 14 (a)(i) of the Copyright Act should not be interpreted in a literal manner. In this regard, reliance is placed on: i. *Shailesh Dhairyawan v. Mohan Balkrishna Lulla*⁴²; ii. *State of Maharashtra v. Praful B. Desai*⁴³.

⁴¹ *Kadrey v. Meta Platforms, Inc.*, 3:23-cv-03417, (N.D. Cal.). ⁴² (2016) 3 SCC 619. ⁴³ (2003) 4 SCC 601.

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 64 of 135 MISHRA Signing Date:24.07.2026 17:17:24 131.3. Training is not based on ANI's data alone. ANI's data along with other data is put in database. ANI's data would only be a miniscule amount of the total data on which training takes place. 131.4. For the purposes of training, Open AI uses the non-expressive elements of literary works of ANI such as grammar, syntax or linguistic patterns. 131.5. The training process is unconcerned with each work's unique expression. The LLMs underlying ChatGPT encode the learnt meanings in the form of a 'vector database' during the pre-training process, which is equivalent to a dictionary and does not amount to additional copies of raw data. Hence, it is incorrect to state that 'vectorization' is an 'adaptation' or 'translation' of ANI's works. 131.6. The storage is not intended for human consumption or comprehension. Storage is only an intermediate step to derive

the non- expressive element.

Intervenors supporting OPEN AI i. Submissions on behalf of Intervenor – Broadband India Forum (I.A. 10766/2025) 132. Mr. Kapil Sibal and Mr. Arvind P. Datar, senior counsel appearing on behalf of the intervenor in I.A. 10766/2025 made the following submissions: 132.1. Any storage of copyrighted works for training LLMs should be protected in line with the statutory intent of making and enabling new technologies. The storage of works is incidental i.e. training of LLMs and LLMs themselves do not store the works.

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 65 of 135 MISHRA Signing Date:24.07.2026 17:17:24 ii. Submission on behalf of Intervenor – Flux Labs AI Private Ltd. (I.A. 4615/2025) 133. Ms. Haripriya Padmanabhan, senior counsel appearing on behalf of intervenor in I.A. 4615/2025 made the following submissions: 133.1. The purpose of copyright law is that the work ought not to be circulated in public. Taking this purpose into consideration, the reference to “storage” in Section 14(a)(i) of the Copyright Act has to be in the context of reproduction of the work. Mere storage of the work would not amount to copyright infringement.

Submissions of Professor Arul George Scaria, Amicus Curiae 134. Prof. Scaria made the following submissions: 134.1. In the present case, the use of copyrighted works by Open AI is ‘non- expressive use’, hence, it would not amount to infringement. The training data is used by Open AI only to extract vital information such as patterns, trend and correlations in phrases and sentences. 134.2. The Court has to consider whether it is feasible to develop an LLM without giving access to copyrighted material as the strength of the LLM depends upon the extent to which materials are available for training. 134.3. The Court must consider if all companies developing LLMs have to obtain licenses from every copyright holder, it will cause practical difficulties. It will also hinder the development of LLMs by small domestic entities. 134.4. The expression “storage” used in Section 14(a)(i) of the Copyright Act is only in the context of “reproduction” and not storage in an independent manner. Therefore, if there is no reproduction, there cannot be any

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 66 of 135 MISHRA Signing Date:24.07.2026 17:17:24 infringement on the basis of storage alone. There is no independent right of storage available to the copyright owner.

Submissions of Mr. Adarsh Ramanujan, Amicus Curiae 135. Mr. Adarsh Ramanujan made the following submissions: 135.1. In terms of Section 14 (a)(i) of the Copyright Act, collection of data and storage of the same would amount to reproduction. Therefore, infringement occurs when Open AI stores the data of ANI. Even if there is a one-time storage, it would result in infringement. 135.2. It is likely that while building its corpus of ‘raw data’, Open AI has used and copied ANI’s publicly available works. Even if stored ‘raw data’ is deleted after the training is complete, it does not change the fact that storage has occurred in the first place. 135.3. Reproduction of ANI’s works during the training process either in the original text form or as a collection of numerical tokens, on a prima facie view, constitutes copyright infringement. Such temporary copies of copyrighted works do not qualify as “transient” or “incidental” copies protected under Section 51(1)(b) of the Copyright Act. 135.4. Even if the purpose of storage is to mine data, the purpose of storage or the intention behind it is not relevant for the purposes of infringement under Section 51 of the Copyright Act. 135.5. The subsequent processes of vectorization/ tokenization raise technical issues and therefore, whether these acts amount to infringement can be decided only in a trial. 135.6. Whether the training process involves further reproduction of ANI’s data, can also be determined only in a trial.

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 67 of 135 MISHRA Signing Date:24.07.2026 17:17:24 Issue No.3 Whether the defendants’ use of plaintiff’s copyrighted data qualifies as ‘fair use’⁴⁴ in terms of Section 52 of the Copyright Act, 1957.

SUBMISSIONS Submissions of the Defendant (OPEN AI)⁴⁵ 136. Mr. Amit Sibal, senior counsel appearing on behalf of Open AI made the following submissions: 136.1. Section 14 of the Copyright Act has to be read with Section 52 of the Copyright Act as it uses the words “subject to provisions of this Act”. Accordingly, Section 52 is not an exception, but a part of what is infringement as defined in Section 14 of the Copyright Act. Therefore, the submission on behalf of ANI that Section 52 should be interpreted in a narrow manner, cannot be sustained. On the contrary, Section 52 of the Copyright Act should be liberally construed. Reliance in this regard is made on B.D. Bhandari (supra) and Chancellor Masters & Scholars of The University of Oxford v. Narendera Publishing⁴⁶. 136.2. Under Section 2(m) of the Copyright Act, a copy of the work would be “infringing copy” only if the requirements of Sections 14, 51 and 52 of the Copyright Act are fulfilled, as Section 2(m) uses the expression “in contravention of the provisions of this Act”. Reliance in this regard is placed on B.D. Bhandari (supra) and

Narendera Publishing (supra).

44 To be read as 'fair dealing'. 45 Since this defence was raised on behalf of the defendant, it is deemed appropriate that the submissions of the defendant are recorded first. 46 2008 SCC OnLine Del 1058.

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 68 of 135 MISHRA Signing Date:24.07.2026 17:17:24 136.3. The Copyright Act as originally enacted in 1957 used the words "research or private study" in its fair dealing clause. The Notes on clauses for Copyright Amendment Bill, 1994 (which was enacted as Copyright Amendment Act, 1994) makes it clear that the amendment was proposed in 1992 to expand the "unduly narrow interpretation of the words "private study"". The latest amendment in the Copyright Act, 1957 included the words "personal use", in addition to the already existing words "private use" and "research". 136.4. There is no restriction on "private use" being only for non-commercial purposes. This view is reinforced from a perusal of sub-clauses (ii) and (iii) of Section 52 (1)(a) of the Copyright Act, which do not make the distinction between commercial and non-commercial. Wherever the Copyright Act makes a distinction between commercial and non-commercial uses, it has been stipulated therein. Reference in this regard may be made to Section 52(1) (ad) and Section 32(6)(d) of the Copyright Act. Merely because the use by Open AI is commercial, would not make it unfair. Reference in this regard is made to the judgment in Super Cassettes Industries v. Hamar Television Network⁴⁷. 136.5. ANI has wrongfully placed reliance on the judgment of Rupendra Kashyap v. Jiwan Publishing House⁴⁸. In the said case, the defendant no.3/CBSE had outsourced printing of its previous years' question papers to third parties/defendants no.1 and 2, hence it was held that the defence under Section 52 would not be available as it is available only to a defendant conducting private study or research and not to a third party. B.D. Bhandari

47 2010 SCC OnLine Del 2086. 48 1996 SCC OnLine Del 466.

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 69 of 135 MISHRA Signing Date:24.07.2026 17:17:24 (supra) holds that even a body corporate is entitled to the defence under Section 52(1) (a) of the Copyright Act. 136.6. The Explanation to Section 52(1)(a) was also added in 2012. In terms of the Explanation to Section 52(1)(a) of the Copyright Act, the expression "not itself being an infringing copy", is restricted to "computer programme" and not to the other forms of storage. The Explanation's language and intent is clear that the protection of Section 52(1)(a) extends to works stored in an electronic medium for the purposes mentioned under the said provision. The limiting language of the Explanation within the commas only applies to computer programmes. 136.7. If ANI's argument is to be accepted, no electronic storage of a copy of a work published in an electronic form (such as an online article which is freely accessible over the internet) would ever be eligible for fair dealing protection under Section 52(1)(a). This would mean that making physical copies (for instance, by hand, a typewriter, or a mechanical medium) of a work would be capable of protection under Section 52(1)(a), while the same act done in an electronic medium would be barred from a fair dealing defence. 136.8. ANI wrongly contends that the defence under Section 52(1)(a) of the Copyright Act would not be available to storage as storage is covered under Section 52(1)(n) and 52(1)(p) of the Copyright Act. The aforesaid Sections cover specific types of storage and on that basis, it cannot be said that storage is not covered under Section 52(1) (a) of the Copyright Act. 136.9. Unlike ANI's works, which are meant for human consumption for its expressive elements, the use of works for training models by Open AI is non- expressive.

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 70 of 135 MISHRA Signing Date:24.07.2026 17:17:24 136.10. The use of Open AI would amount to "private use" as mentioned in sub-clause (i) of Section 52(1)(a) of the Copyright Act, as it is only for the purposes of training of the Open AI's LLM. If Section 52(1)(a)(i) of the Copyright Act is interpreted to mean that the exception is available only for personal use, the word "private" would be rendered otiose. 136.11. The word "research" must be given a broad and liberal interpretation so that it is not limited in any manner to a private context. Reliance in this regard is placed on the judgment of Canadian Court in CCH Canadian (supra). 136.12. Open AI's storage and use of training data falls within the ambit of "private use", including "research", which is protected under Section 52(1)(a)(i) of the Copyright Act and does not amount to infringement of copyright. 136.13. Indian Courts have adapted the US four-factor test in India. By various judgments of the Indian Court, the US four-factor test has been adapted under Indian copyright law for the purposes of Section 52(1)(a) of the Copyright Act. The Indian judgments which have relied upon the four-factor test are as follows: i. India TV Independent News Service v. Yashraj Films⁴⁹ ii. Super Cassettes Industries Limited v. Chintamani Rao⁵⁰ iii. B.D. Bhandari (supra) iv. Narendera Publishing (supra) Indian Courts have also recognized that the most crucial test out of the four- factor test is transformative character of the use. The use of publicly available

49 (2012) SCC OnLine Del 4298. 50 (2011) SCC OnLine Del 4712.

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 71 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 works forming part of the training data for training the LLMs is highly transformative. [Ref. Narendera Publishing (supra) and B.D. Bhandari (supra)]. 136.14. Open AI's ChatGPT is not a substitute for subscription to news agencies, such as ANI.

Intervenors supporting OPEN AI i. Submissions on behalf of Intervenor - IGAP Project LLP (I.A. 4616/2025) 137. Mr. Akhil Sibal, senior counsel appearing on behalf of the intervenor in I.A. 4616/2025 made the following submissions: 137.1. When originally enacted, Section 52(1)(a)(i) of the Copyright Act limited the purposes to "research or private study". The provision was amended in 1994 to "private use including research". 137.2. The use of copyrighted data to train LLMs meets both, the purpose and the fairness limitations as envisaged under Section 52(1)(a) of the Copyright Act. 137.3. Section 52(1)(a) of the Copyright Act does not provide that commercial uses would not be covered therein. In contradistinction, Section 52(1)(k)(ii), Section 52(1)(l) and Section 52(1)(o) of the Copyright Act have excluded commercial uses. Hence, the legislature has deliberately chosen not to put an element of commerciality in Section 52(1)(a) of the Copyright Act. 137.4. Reliance placed by ANI on Jiwan Publishing House (supra) and B.D. Bhandari (supra) to submit that commerciality defeats fair dealing is misplaced. Jiwan Publishing House (supra) was a case where a publisher of infringing books claimed that since its customers are using the books for

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 72 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 personal use, the publisher's use should also be considered private or personal, which was rejected by the Court. 137.5. B.D. Bhandari (supra) holds that a commercial use is not fair in the context of Section 52(1)(h) of the Copyright Act and not in the context of Section 52(1)(a) of the Copyright Act. There is no general proposition laid down in the aforesaid judgment that commercial use defeats fair dealing. 137.6. The purposes mentioned in Section 52(1)(a) of the Copyright Act, such as "research", "review" and "reporting of current news and current affairs" are well recognized as permissible uses of copyright, which are often carried out for commercial purposes. 137.7. While interpreting terms such as "research", "review" and "private use", the Court should apply the doctrine of updating construction i.e. the Parliament, while enacting a law is expected to anticipate future developments. Accordingly, the terms "research", "private use" and "review" may not be limited to acts of human beings alone but to machine learning as well. Therefore, these terms should be interpreted liberally. 137.8. The term "private use" must be contradistinguished from the term "personal use". Supreme Court in Academy of General Education, Manipal v. B. Malini Mallya⁵¹ has adopted the expansive meaning of the word "private" by holding that a dance performance at an educational institution can also be within the meaning of "private use" for research, review or criticism. 'Pre-training dataset' obtained by LLMs is private in nature i.e. accessible only to LLM models themselves. In other words, this data set is not

51 (2009) 4 SCC 256.

Signature Not Verified Signed By:VIVEK CS(COMM) 1028/2024 Page 73 of 135 MISHRA Signing
Date:24.07.2026 17:17:24 publicly available for either access or download. Therefore, 'use' of any copyrighted material is totally private. 137.9. In B.D. Bhandari (supra), the Division Bench adopted the transformative test adopted by the American Courts for determining what constitutes fair use. The Court considered whether contribution in writing a 'guide book' is such that a nature of guide book is different than the textbook. It was held that if the purpose served by the guide book is substantially different from the purpose served by the text book, it would be treated as transformative. 137.10. In The Chancellor, Masters & Scholars of University of Oxford v. Rameshwari Photocopy Services (DB)⁵², it was held that the purposes of use would determine whether a particular use qualifies as being fair. The fairness is determined by whether the extent is justified by its purpose. In the context of reproduction of literary works for creating course packs, the Court held that copying of entire articles was justified for the purposes of education. 137.11. The DNPA and IMI have wrongfully relied upon Bartz v. Anthropic (supra) to submit that the fair use defence for training is unavailable if the first copy is not authorized by the copyright holder. In Bartz v. Anthropic (supra), the defendant had used 'pirated books' for creating a permanent general-purpose library, which was held to be not fair use. In the present case, Open AI does not create any library and the data is used only for training purpose. 137.12. The US Courts in Bartz v. Anthropic (supra) and Kadrey v. Meta Platforms (supra) have come to the conclusion that the use of copyrighted materials for training LLM is transformative. In Bartz v. Anthropic (supra),