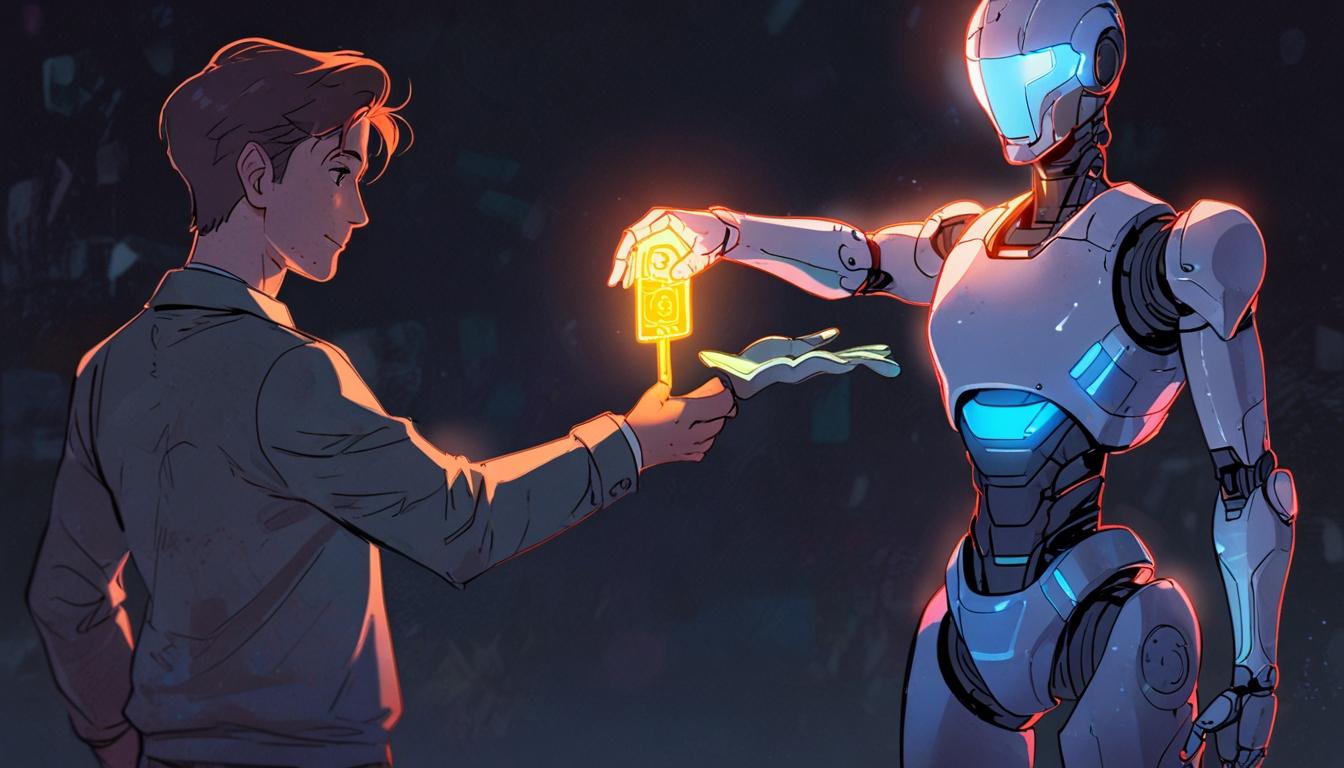
# Start-ups drive $67.5bn AI content licensing market amid legal battles



As artificial intelligence continues to evolve, a burgeoning network of start-ups is stepping in to bridge the gap between content creators and AI developers. These companies are facilitating the licensing of creative work to tech giants like OpenAI, Meta, and Google, which face mounting scrutiny over their data sourcing practices. In recent years, as concerns escalate around copyright infringement, investors have poured significant capital—around $215 million since 2022—into firms such as Pip Labs, Vermillio, Created by Humans, ProRata, Narrativ, and Human Native. Their goal is to establish marketplaces and tools that ensure writers, musicians, publishers, and filmmakers can monetise their content for AI training.

The rise of the AI licensing market reflects shifting dynamics in the creative industries. Dan Neely, the chief executive of Vermillio, states that the licensing business will be integral to future content monetisation strategies. With major partnerships with Sony Pictures and Sony Music, Vermillio anticipates a staggering growth in the AI licensing sector, projecting it to expand from around $10 billion in 2025 to an estimated $67.5 billion by 2030. This growth is not isolated; the number of AI licensing agreements has surged, peaking at a record 16 in December 2024.

Notably, the urgency for media companies to formalise licensing agreements stems from a broader landscape of legal anxiety. OpenAI and Perplexity have signed over 20 deals with media groups since 2023, particularly targeting news organisations. However, this momentum is overshadowed by legal challenges, including lawsuits from the New York Times and other media outlets that accuse major AI firms of misappropriating copyrighted materials for their models. As indicated by James Smith, chief executive of Human Native, AI firms are rapidly realising the importance of content acquisition, stating, "They’re just getting around [to] spending millions on the third," referring to data as an essential component alongside talent and compute power.

In the United States, the political climate surrounding copyright enforcement is fractious. Just recently, Shira Perlmutter was dismissed as the director of the U.S. Copyright Office following political upheaval, an event linked to a report that detailed potential loopholes for tech companies regarding the use of copyrighted material for AI training. This has raised alarms among critics, including Democratic senators, who argue that such dismissals threaten the independence of copyright regulation in the face of corporate interests.

Internationally, similar tensions are unraveling. In India, a panel has been formed to review copyright laws in light of the prevailing legal disputes involving OpenAI, with leading news outlets alleging that their materials were used without consent. OpenAI maintains it operates within the bounds of public data and offers opt-out mechanisms; nevertheless, an undercurrent of discontent persists among content creators who feel their rights are being undermined.

While there is a clear momentum in the establishment of content licensing frameworks, hurdles remain. The primary challenge is garnering enough high-quality data providers to fuel a sustainable business model. Many online datasets still contain harmful content, which could expose companies to reputational risks. Furthermore, convincing artists that selling their content to train AI models is beneficial is a formidable task. According to Gina Neff, a professor of responsible AI, there exists a significant lack of trust among creators regarding existing technical solutions, leading many to perceive such licensing as a poor trade-off.

Nonetheless, industry innovators are advocating for alternative models rather than waiting for legislative changes. Jason Zhao, co-founder of Pip Labs, posits that the focus should be on creating solutions that benefit both AI firms and content creators, suggesting that technological advancements could provide a more promising path forward. Meanwhile, Prem Akkaraju, chief executive of Stability AI, is exploring the establishment of a licensing marketplace where artists can manage their intellectual property rights more effectively, thereby granting permission for AI training in a way that adds value to all parties involved.

In this rapidly evolving landscape, the question of how to ethically incorporate copyrighted material into AI training processes remains paramount. As major legal cases loom and creative industries grapple with the implications of AI, discourse surrounding the fair use of content continues to intensify. The road ahead for these start-ups and their tech counterparts will likely require innovative collaboration and a commitment to respecting the rights of creators.

### Reference Map

1. Paragraphs 1, 2, 4, 5, 7
2. Paragraph 3
3. Paragraph 6
4. Paragraph 6
5. Paragraph 6
6. Paragraph 6
7. Paragraph 6

Source: [Noah Wire Services](https://www.noahwire.com)

## Bibliography

1. <https://www.ft.com/content/dc1225e1-22ce-4d6f-a343-a15bf360bf3c> - Please view link - unable to able to access data
2. <https://www.ft.com/content/dc1225e1-22ce-4d6f-a343-a15bf360bf3c> - Investors are increasingly supporting start-ups that facilitate content licensing for AI training, as major tech firms like OpenAI, Meta, and Google face legal and regulatory scrutiny over the use of copyrighted material. Since 2022, start-ups such as Pip Labs, Vermillio, Created by Humans, ProRata, Narrativ, and Human Native have secured $215 million in funding to develop marketplaces and tools allowing creatives to monetize their content. Vermillio, which partners with major studios like Sony, forecasts the AI licensing market to rise from $10 billion in 2025 to $67.5 billion by 2030. The number of licensing agreements surged, with a record 16 in December 2024. Companies like OpenAI and Perplexity have made over 20 media licensing deals since 2023. Despite this momentum, challenges remain, including attracting quality data providers, mitigating content-related risks, and convincing creators of the value. Legal actions against companies like Meta and Stability AI highlight the stakes, while ongoing policy discussions in the UK, US, and EU underscore the evolving regulatory landscape. Start-ups like Pip Labs advocate for technological rather than legislative solutions to fairly compensate content creators and support sustainable AI model development.
3. <https://www.reuters.com/legal/government/trump-fires-head-us-copyright-office-2025-05-12/> - On May 10, 2025, the Trump administration dismissed U.S. Copyright Office director Shira Perlmutter via a White House email, following the previous day's removal of Librarian of Congress Carla Hayden. The Copyright Office, a department within the Library of Congress, had recently released a report under Perlmutter’s leadership suggesting that technology companies' use of copyrighted materials to train artificial intelligence systems may not be covered under U.S. copyright law, particularly when AI-generated content competes with original creators. This stance is contested by tech companies like OpenAI and Meta, who argue that requiring payments to copyright holders could hinder AI development. Critics of the firing, including Democratic Senators Adam Schiff and Chuck Schumer and Representative Joe Morelle, claim the dismissal is politically motivated and unlawful, asserting that the role of the Copyright Office was intended to be insulated from political influence. Morelle also suggested a connection between Perlmutter’s refusal to support Elon Musk’s use of copyrighted content for AI training and her abrupt termination. Musk, who oversees the Department of Government Efficiency and owns AI firm xAI, has yet to comment on the matter.
4. <https://www.reuters.com/sustainability/boards-policy-regulation/india-panel-review-copyright-law-amid-legal-challenges-openai-2025-05-06/> - India has formed a panel to review its copyright laws in response to rising legal challenges involving artificial intelligence (AI), particularly concerning OpenAI. A confidential memo reveals that the commerce ministry established a group of eight experts last month to examine how current copyright laws, specifically the Copyright Act of 1957, apply to AI-generated content. This move follows a high-profile court case in New Delhi, where major Indian news outlets, including NDTV, Indian Express, and Hindustan Times, allege that OpenAI has used their copyrighted material without consent to train its ChatGPT chatbot. OpenAI denies any wrongdoing, stating it only uses public data and provides a mechanism for websites to opt out. The expert panel, comprised of intellectual property lawyers, government officials, and industry leaders, will propose recommendations to address the legal and policy implications of AI in copyright. Globally, technology companies face similar lawsuits from content creators who argue that their copyrighted work is being exploited without permission or compensation for AI training purposes.
5. <https://apnews.com/article/61fb43f20d945753a8c86881aa631d65> - A global coalition of media organizations, including the European Broadcasting Union (EBU) and the World Association of News Publishers (WAN-IFRA), is urging artificial intelligence (AI) developers to collaborate in combating misinformation and safeguarding fact-based journalism. Announced at the World News Media Congress in Krakow, Poland, the "News Integrity in the Age of AI" initiative encompasses thousands of media groups and outlines five core principles for ethical AI use in news. Key demands include requiring prior authorization for using news content in AI models, ensuring transparency in attribution, and making original sources clearly identifiable. The initiative involves major media associations such as the Asia-Pacific Broadcasting Union, North American Broadcasters Association (which includes Fox, NBC Universal, PBS, and others), and the Latin American broadcasters association AIL. The call to action comes amid rising tension between traditional media and AI developers, with some outlets—such as The New York Times—pursuing lawsuits against OpenAI and Microsoft over copyright concerns. Meanwhile, other organizations have entered content licensing agreements with AI firms. The debate continues over whether using copyrighted content to train AI models falls under "fair use" provisions.
6. <https://www.ft.com/content/b1f4965f-6ea6-4afd-968b-76bb2f7acfc4> - Meta faces a significant legal challenge as it defends itself in court against a lawsuit filed by a group of U.S. authors, including Ta-Nehisi Coates and Richard Kadrey. The authors allege that Meta unlawfully used copyrighted books from the shadow library LibGen to train its Llama artificial intelligence models. The case is one of the first major tests to determine whether AI companies can claim "fair use" of copyrighted material for model training, especially when sourced from pirated databases. Meta contends that its usage was transformative and thus permissible, asserting that there was no viable market to license such works. However, court documents revealed internal Meta emails suggesting the company knowingly entered legal grey areas and sought to avoid public disclosure of their actions. The authors claim Meta bypassed licensing discussions to cut costs, resulting in a loss of authors’ control and income. Similar lawsuits are pending against other tech giants like Microsoft and OpenAI. The outcome of this case could set a precedent with far-reaching implications for copyright law and AI development. Meta continues to defend its stance, emphasizing the importance of fair use in advancing generative AI technology.
7. <https://www.reuters.com/business/publisher-ziff-davis-sues-openai-copyright-infringement-2025-04-24/> - On April 24, 2025, digital media publisher Ziff Davis filed a copyright infringement lawsuit against OpenAI in Delaware federal court. The lawsuit alleges that OpenAI, supported by Microsoft, unlawfully used Ziff Davis's copyrighted content to train its artificial intelligence systems, including the widely used chatbot ChatGPT. Ziff Davis, whose media portfolio includes ZDNet, PCMag, CNET, IGN, and Lifehacker, claims that OpenAI has systematically and deliberately exploited its copyrighted materials without permission. The case joins a growing number of legal challenges from publishers, authors, and artists who accuse major AI firms like OpenAI, Google, and Meta Platforms of using their intellectual property without authorization to develop generative AI technologies. OpenAI defends its practices by stating it trains models on publicly available data and operates within fair use standards. A spokesperson for Ziff Davis declined to comment further on the ongoing litigation. This lawsuit underscores the increasing tension between content creators and tech companies over the boundaries of copyright in the age of AI.