EXHIBIT 1

(FILED UNDER SEAL)

HIGHLY CONFIDENTIAL – ATTORNEYS' EYES ONLY

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

SAN JOSE DIVISION

CONCORD MUSIC GROUP, INC., ET AL., Plaintiffs, v. ANTHROPIC PBC, Defendant.	Case Number: 5:24-cv-03811-EKL-SVK Judge Eumi K. Lee Magistrate Judge Susan van Keulen

EXPERT REPORT OF DR. SHAWN SHAN

December 19, 2025

Contains Information Designated as
Confidential and/or Highly Confidential—Attorneys' Eyes Only

standard copyright symbol ©) and \u2117 (the sound recording copyright symbol ℗).

iv. Plaintiff Identification: For every text block confirmed to contain a copyright symbol, I executed a targeted search to identify the Plaintiffs' specific names. I utilized a precise text-matching technique (known technically as "Regular Expressions") to scan the content. In this context, to "detect" means the software searched for any text string that contained the exact sequence of characters for the following entities: "concord music," "capitol," "universal music," "polygram," or "abkco." Crucially, I applied a "whole word" constraint to this search. This ensured that the system would only report a match if the name appeared as a complete, distinct word (*e.g.*, correctly matching "Capitol"), while automatically rejecting cases where those letters were merely a fragment hidden within a longer, unrelated word (*e.g.*, ignoring "Capitol" if it appeared inside the word "Capitoline").

58. My analysis identified a massive volume of copyright symbols linked to names associated with Publishers within Anthropic's training data. The results, broken down Publisher, and appended as **Exhibit F** are as follows:

Keywords corresponding to Publishers	**Number of copyright symbol matches**
"Universal Music"	[redacted]
"Capitol"	[redacted]
"Concord"	[redacted]
"ABKCO"	[redacted]
"Polygram"	[redacted]

59. In my opinion, these findings confirm that the raw training data ingested by Anthropic contained explicit copyright management information identifying the relevant rights holders. In the raw data files I inspected, the "©" and "℗" symbols and accompanying text remained intact. While the precise figures in the table above are likely to be overinclusive in certain

respects and underinclusive in other respects, based on the nature of Anthropic's data and the limitations of my searches, the results nevertheless reflect that the initial copies of many of the Works in Suit in the training data included this type of CMI. While Anthropic may attempt to trim or process this text in later pipeline stages, the training data as acquired and stored definitively includes these copyright notices. To the extent Anthropic's AI models later generated output containing Publishers' lyrics without this corresponding CMI (as reflected in the specific Claude output records described in more detail in the reports of Publishers' experts Joshua Dennis and Robert Leonard, and as reflected in Ex. B to Publishers' First Amended Complaint and in various Claude output records produced by Anthropic), such CMI was removed from the copies of Publishers' lyrics later in the process of Anthropic's AI training or in the process of generating output based on the training data.

X. Guardrail Examination

60. As discussed above, I understand that Anthropic has disclosed as part of this litigation that it has implemented three guardrail systems, which it states are aimed at preventing output that copies Publishers' Works in Suit and/or other copyrighted content.

61. As part of my inspections, I investigated the "Content Compare System" ("CCS"), also referred to as "Copyfairy," which Anthropic has described as a "recitation checker" that checks the AI models' output against an external dataset to determinate whether the output is the same or similar to the external dataset's contents.[36] I understand that the [redacted] were not added to this system until [redacted], and the [redacted] were added to this system on [redacted][37]

62. To operate this system, Anthropic has collected a large number of known copyrighted lyrics and processed them into digital "hashes" (unique fingerprints). These hashed values are then stored in a [redacted]. The system functions as an active output filter. As the model generates a response to a user, the system performs checks throughout

[36] Nov. 12, 2025 Interrogatory Responses at 12.
[37] *See id.* at 14.

Dated: December 19, 2025

Shawn Shan