CHAPTER 6: COPYRIGHT

A. Owning Information

1. Copyright

ASSESSMENT TECHNOLOGIES OF WISCONSIN, LLC V. WIREDATA, INC.
350 F.3d 640 (2003)

Posner, Circuit Judge:

This case is about the attempt of a copyright owner to use copyright law to block access to data that not only are neither copyrightable nor copyrighted, but were not created or obtained by the copyright owner. The owner is trying to secrete the data in its copyrighted program—a program the existence of which reduced the likelihood that the data would be retained in a form in which they would have been readily accessible. It would be appalling if such an attempt could succeed.

Assessment Technologies (AT, we'll call it) brought suit for copyright infringement and theft of trade secrets against WIREdata, and the district court after an evidentiary hearing issued a permanent injunction on the basis of AT's copyright claim alone, without reaching the trade secret claim. ...

The copyright case seeks to block WIREdata from obtaining noncopyrighted data. AT claims that the data can't be extracted without infringement of its copyright. The copyright is of a compilation format, and the general issue that the appeal presents is the right of the owner of such a copyright to prevent his customers (that is, the copyright licensees) from disclosing the compiled data even if the data are in the public domain.

WIREdata, owned by Multiple Listing Services, Inc., wants to obtain, for use by real estate brokers, data regarding specific properties—address, owner's name, the age of the property, its assessed valuation, the number and type of rooms, and so forth—from the southeastern Wisconsin municipalities in which the properties are located. The municipalities collect such data in order to assess the value of the properties for property-tax purposes. Ordinarily they're happy to provide the data to anyone who will pay the modest cost of copying the data onto a disk. Indeed, Wisconsin's "open records" law, Wis. Stat. §§ 19.31-.39, which is applicable to data in digital form, *see id.* at 195-96; Wis. Stat. § 19.32(2), requires them to furnish such data to any person who will pay the copying cost. However, three municipalities refused WIREdata's request. They (or the contractors who do the actual tax assessment for them) are licensees of AT. The open-records law contains an exception for copyrighted materials, *id.*, and these municipalities are afraid that furnishing WIREdata the requested data would violate the copyright. WIREdata has sued them in the state courts of Wisconsin in an attempt to force them to divulge the data, and those suits are pending. Alarmed by WIREdata's suits, AT brought the present suit to stop WIREdata from making such demands of the municipalities and seeking to enforce them by litigation.

The data that WIREdata wants are collected not by AT but by tax assessors hired by the municipalities. The assessors visit the property and by talking to the owner and poking around the property itself obtain the information that we men-

tioned in the preceding paragraph—the age of the property, the number of rooms, and so forth. AT has developed and copyrighted a computer program, called "Market Drive," for compiling these data. The assessor types into a computer the data that he has obtained from his visit to the property or from other sources of information and then the Market Drive program, in conjunction with a Microsoft database program (Microsoft Access), automatically allocates the data to 456 fields (that is, categories of information) grouped into 34 master categories known as tables. Several types of data relating to a property, each allocated to a different field, are grouped together in a table called "Income Valuations," others in a table called "Residential Buildings," and so on. The data collected by the various assessors and inputted in the manner just described are stored in an electronic file, the database. The municipality's tax officials can use various queries in Market Drive or Market Access to view the data in the file.

WIREdata's appeal gets off on the wrong foot, with the contention that Market Drive lacks sufficient originality to be copyrightable. Copyright law unlike patent law does not require substantial originality. *Feist Publications, Inc. v. Rural Telephone Service Co.*, 499 U.S. 340, 345-48 (1991). In fact, it requires only enough originality to enable a work to be distinguished from similar works that are in the public domain, since without some discernible distinction it would be impossible to determine whether a subsequent work was copying a copyrighted work or a public-domain work. This modest requirement is satisfied by Market Drive because no other real estate assessment program arranges the data collected by the assessor in these 456 fields grouped into these 34 categories, and because this structure is not so obvious or inevitable as to lack the minimal originality required, as it would if the compilation program simply listed data in alphabetical or numerical order. *Feist,* 499 U.S. at 362-64. The obvious orderings, the lexical and the numeric, have long been in the public domain, and what is in the public domain cannot be appropriated by claiming copyright. Alternatively, if there is only one way in which to express an idea—for example, alphabetical order for the names in a phone book—then form and idea merge, and in that case since an idea cannot be copyrighted the copying of the form is not an infringement. That is not the situation here.

So AT has a valid copyright; and if WIREdata said to itself, "Market Drive is a nifty way of sorting real estate data and we want the municipalities to give us their data in the form in which it is organized in the database, that is, sorted into AT's 456 fields grouped into its 34 tables," and the municipalities obliged, they would be infringing AT's copyright because they are not licensed to make copies of Market Drive for distribution to others; and WIREdata would be a contributory infringer (subject to a qualification concerning the fair-use defense to copyright infringement, including contributory infringement, that we discuss later). But WIREdata doesn't want the compilation as structured by Market Drive. It isn't in the business of making tax assessments, which is the business for which Market Drive is designed. It only wants the raw data, the data the assessors inputted into Market Drive. Once it gets those data it will sort them in accordance with its own needs, which have to do with providing the information about properties that is useful to real estate brokers as opposed to taxing authorities.

But how are the data to be extracted from the database without infringing the copyright? Or, what is not quite the same question, how can the data be separated from the tables and fields to which they are allocated by Market Drive? One possibility is to use tools in the Market Drive program itself to extract the data and

place it in a separate electronic file; this can be done rapidly and easily with just a few keystrokes. But the municipalities may not have the program, because the inputting of the data, which did of course require its use, was done by assessors employed by firms to do this work as independent contractors of the municipalities. And if the municipalities do have the program, still their license from AT forbids them to disseminate the data collected by means of it—a restriction that may or may not be in violation of the state's open-records law, a question we come back to later. A second extraction possibility, which arises from the fact that the database is a Microsoft file accessible by Microsoft Access, is to use Access to extract the data and place it in a new file, bypassing Market Drive. But there is again the scope of the license to be considered and also whether the method of extraction is so cumbersome that it would require more effort than the open-records law requires of the agencies subject to it. It might take a programmer a couple of days to extract the data using Microsoft Access, and the municipalities might lack the time, or for that matter the programmers, to do the extraction. But that should not be a big problem, because WIREdata can hire programmers to extract the data from the municipalities' computers at its own expense.

From the standpoint of copyright law all that matters is that the process of extracting the raw data from the database does not involve copying Market Drive ... ; all that is sought is raw data, data created not by AT but by the assessors, data that are in the public domain. ... The municipalities would not be infringing Market Drive by extracting the raw data from the databases by either method that we discussed and handing those data over to WIREdata; and since there would thus be no direct infringement, neither would there be contributory infringement by WIREdata. It would be like a Westlaw licensee's copying the text of a federal judicial opinion that he found in the Westlaw opinion database and giving it to someone else. Westlaw's compilation of federal judicial opinions is copyrighted and copyrightable because it involves discretionary judgments regarding selection and arrangement. But the opinions themselves are in the public domain (federal law forbids assertion of copyright in federal documents, 17 U.S.C. § 105), and so Westlaw cannot prevent its licensees from copying the opinions themselves as distinct from the aspects of the database that are copyrighted. *See Matthew Bender & Co. v. West Publishing Co.*, 158 F.3d 693 (2d Cir. 1998); *Matthew Bender & Co. v. West Publishing Co.*, 158 F.3d 674 (2d Cir. 1998).

AT would lose this copyright case even if the raw data were so entangled with Market Drive that they could not be extracted without making a copy of the program. The case would then be governed by *Sega Enterprises Ltd. v. Accolade, Inc.*, 977 F.2d 1510, 1520-28 (9th Cir. 1992). Sega manufactured a game console, which is a specialized computer, and copyrighted the console's operating system, including the source code. Accolade wanted to make computer games that would be compatible with Sega's console, and to that end it bought a Sega console and through reverse engineering reconstructed the source code, from which it would learn how to design its games so that they would activate the operating system. For technical reasons, Accolade had to make a copy of the source code in order to be able to obtain this information. It didn't want to sell the source code, produce a game-console operating system, or make any other use of the copyrighted code except to be able to sell a noninfringing product, namely a computer game. The court held that this "intermediate copying" of the operating system was a fair use, since the only effect of enjoining it would be to give Sega control over noninfringing products, namely Accolade's games. *See also* S*ony Computer Entertainment,*

Inc. v. Connectix Corp., 203 F.3d 596, 602-08 (9th Cir. 2000); *Bateman v. Mnemonics, Inc.*, 79 F.3d 1532, 1539-40 n. 18 (11th Cir. 1996); *Atari Games Corp. v. Nintendo of America, Inc.*, 975 F.2d 832, 842-44 (Fed. Cir. 1992). Similarly, if the only way WIREdata could obtain public-domain data about properties in southeastern Wisconsin would be by copying the data in the municipalities' databases as embedded in Market Drive, so that it would be copying the compilation and not just the compiled data only because the data and the format in which they were organized could not be disentangled, it would be privileged to make such a copy, and likewise the municipalities. For the only purpose of the copying would be to extract noncopyrighted material, and not to go into competition with AT by selling copies of Market Drive. We emphasize this point lest AT try to circumvent our decision by reconfiguring Market Drive in such a way that the municipalities would find it difficult or impossible to furnish the raw data to requesters such as WIREdata in any format other than that prescribed by Market Drive. ...

AT argues that WIREdata doesn't need to obtain the data in digital form because they exist in analog form, namely in the handwritten notes of the assessors, notes that all agree are not covered by the Market Drive copyright. But we were told at argument without contradiction that some assessors no longer make handwritten notes to copy into a computer at a later time. Instead they take their laptop to the site and type the information in directly. So WIREdata could not possibly obtain all the data it wants (all of which data are in the public domain, we emphasize) from the handwritten notes. But what is more fundamental is that since AT has no ownership or other legal interest in the data collected by the assessor, it has no legal ground for making the acquisition of that data more costly for WIREdata. AT is trying to use its copyright to sequester uncopyrightable data, presumably in the hope of extracting a license fee from WIREdata.

We are mindful of pressures, reflected in bills that have been pending in Congress for years, to provide legal protection to the creators of databases, as Europe has already done. The creation of massive electronic databases can be extremely costly, yet if the database is readily searchable and the data themselves are not copyrightable (and we know from Feist that mere data are indeed not copyrightable) the creator may find it difficult or even impossible to recoup the expense of creating the database. Legal protection of databases as such (as distinct from programs for arranging the data, like Market Drive) cannot take the form of copyright, as the Supreme Court made clear in Feist when it held that the copyright clause of the Constitution does not authorize Congress to create copyright in mere data. But that is neither here nor there; what needs to be emphasized in this case is that the concerns that actuate the legislative proposals for database protection have no relevance because AT is not the collector of the data that go into the database. All the data are collected and inputted by the assessors; it is they, not AT, that do the footwork, the heavy lifting. ...

The judgment is reversed with instructions to vacate the injunction and dismiss the copyright claim.

2. The E.U. Database Directive

E.U. DATABASE DIRECTIVE
Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996
on the Legal Protection of Databases
[1996] OJ L 77

art. 1 – ***Scope***

1. This Directive concerns the legal protection of databases in any form.
2. For the purposes of this Directive, 'database' shall mean a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means.
3. Protection under this Directive shall not apply to computer programs used in the making or operation of databases accessible by electronic means. ...

art. 7 – ***Object of protection***

1. Member States shall provide for a right for the maker of a database which shows that there has been qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents to prevent extraction and/or re-utilization of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database.
2. For the purposes of this Chapter:
   (a) 'extraction' shall mean the permanent or temporary transfer of all or a substantial part of the contents of a database to another medium by any means or in any form;
   (b) 're-utilization' shall mean any form of making available to the public all or a substantial part of the contents of a database by the distribution of copies, by renting, by on-line or other forms of transmission. The first sale of a copy of a database within the Community by the rightholder or with his consent shall exhaust the right to control resale of that copy within the Community;

   Public lending is not an act of extraction or re-utilization.
3. The right referred to in paragraph 1 may be transferred, assigned or granted under contractual licence.
4. The right provided for in paragraph 1 shall apply irrespective of the eligibility of that database for protection by copyright or by other rights. Moreover, it shall apply irrespective of eligibility of the contents of that database for protection by copyright or by other rights. Protection of databases under the right provided for in paragraph 1 shall be without prejudice to rights existing in respect of their contents.
5. The repeated and systematic extraction and/or re-utilization of insubstantial parts of the contents of the database implying acts which conflict with a normal exploitation of that database or which unreasonably prejudice the legitimate interests of the maker of the database shall not be permitted.

art. 8 – Rights and obligations of lawful users

1. The maker of a database which is made available to the public in whatever manner may not prevent a lawful user of the database from extracting and/or re-utilizing insubstantial parts of its contents, evaluated qualitatively and/or quantitatively, for any purposes whatsoever. Where the lawful user is

authorized to extract and/or re-utilize only part of the database, this paragraph shall apply only to that part.

2. A lawful user of a database which is made available to the public in whatever manner may not perform acts which conflict with normal exploitation of the database or unreasonably prejudice the legitimate interests of the maker of the database.
3. A lawful user of a database which is made available to the public in any manner may not cause prejudice to the holder of a copyright or related right in respect of the works or subject matter contained in the database.

art. 9 – ***Exceptions to the sui generis right***

Member States may stipulate that lawful users of a database which is made available to the public in whatever manner may, without the authorization of its maker, extract or re-utilize a substantial part of its contents:

(a) in the case of extraction for private purposes of the contents of a non-electronic database;

(b) in the case of extraction for the purposes of illustration for teaching or scientific research, as long as the source is indicated and to the extent justified by the non-commercial purpose to be achieved;

(c) in the case of extraction and/or re-utilization for the purposes of public security or an administrative or judicial procedure.

art. 10 – ***Term of protection***

1. The right provided for in Article 7 shall run from the date of completion of the making of the database. It shall expire fifteen years from the first of January of the year following the date of completion.
2. In the case of a database which is made available to the public in whatever manner before expiry of the period provided for in paragraph 1, the term of protection by that right shall expire fifteen years from the first of January of the year following the date when the database was first made available to the public.
3. Any substantial change, evaluated qualitatively or quantitatively, to the contents of a database, including any substantial change resulting from the accumulation of successive additions, deletions or alterations, which would result in the database being considered to be a substantial new investment, evaluated qualitatively or quantitatively, shall qualify the database resulting from that investment for its own term of protection.

BRITISH HORSERACING BOARD LTD. V. WILLIAM HILL ORGANIZATION LTD.

European Court of Justice
ECLI:EU:C:2004:695

1. This reference for a preliminary ruling concerns the interpretation of Article 7 and Article 10(3) of Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases (OJ 1996 L 77, p. 20, 'the directive').
2. The reference was made in the course of proceedings brought by The British Horseracing Board Ltd, the Jockey Club and Weatherbys Group Ltd ('the BHB and Others') against William Hill Organization Ltd ('William Hill').

The litigation arose over the use by William Hill, for the purpose of organising betting on horse racing, of information taken from the BHB database.

The Main Proceedings and the Questions Referred for a Preliminary Ruling

10. The BHB and Others manage the horse racing industry in the United Kingdom and in various capacities compile and maintain the BHB database which contains a large amount of information supplied by horse owners, trainers, horse race organisers and others involved in the racing industry. The database contains information on inter alia the pedigrees of some one million horses, and 'pre-race information' on races to be held in the United Kingdom. That information includes the name, place and date of the race concerned, the distance over which the race is to be run, the criteria for eligibility to enter the race, the date by which entries must be received, the entry fee payable and the amount of money the racecourse is to contribute to the prize money for the race.
11. Weatherbys Group Ltd, the company which compiles and maintains the BHB database, performs three principal functions, which lead up to the issue of pre-race information.
12. First, it registers information concerning owners, trainers, jockeys and horses and records the performances of those horses in each race.
13. Second, it decides on weight adding and handicapping for the horses entered for the various races.
14. Third, it compiles the lists of horses running in the races. This activity is carried out by its own call centre, manned by about 30 operators. They record telephone calls entering horses in each race organised. The identity and status of the person entering the horse and whether the characteristics of the horse meet the criteria for entry to the race are then checked. Following those checks the entries are published provisionally. To take part in the race, the trainer must confirm the horse's participation by telephone by declaring it the day before the race at the latest. The operators must then ascertain whether the horse can be authorised to run the race in the light of the number of declarations already recorded. A central computer then allocates a saddle cloth number to each horse and determines the stall from which it will start. The final list of runners is published the day before the race.
15. The BHB database contains essential information not only for those directly involved in horse racing but also for radio and television broadcasters and for bookmakers and their clients. The cost of running the BHB database is approximately £4 million per annum. The fees charged to third parties for the use of the information in the database cover about a quarter of that amount.
16. The database is accessible on the internet site operated jointly by BHB and Weatherbys Group Ltd. Some of its contents are also published each week in the BHB's official journal. The contents of the database, or of certain parts of it, are also made available to Racing Pages Ltd, a company jointly controlled by Weatherbys Group Ltd and the Press Association, which then forwards data to its various subscribers, including some bookmakers, in the form of a 'Declarations Feed', the day before a race. Satellite Information

Services Limited ('SIS') is authorised by Racing Pages to transmit data to its own subscribers in the form of a 'raw data feed' ('RDF'). The RDF includes a large amount of information, in particular, the names of the horses running in the races, the names of the jockeys, the saddle cloth numbers and the weight for each horse. Through the newspapers and the Ceefax and Teletext services, the names of the runners in a particular race are made available to the public during the course of the afternoon before the race.

17. William Hill, which is a subscriber to both the Declarations Feed and the RDF, is one of the leading providers of off-course bookmaking services in the United Kingdom, to both UK and international customers. It launched an on-line betting service on two internet sites. Those interested can use these sites to find out what horses are running in which races at which racecourses and what odds are offered by William Hill.

18. The information displayed on William Hill's internet sites is obtained, first, from newspapers published the day before the race and, second, from the RDF supplied by SIS on the morning of the race.

19. According to the order for reference, the information displayed on William Hill's internet sites represents a very small proportion of the total amount of data on the BHB database, given that it concerns only the following matters: the names of all the horses in the race, the date, time and/or name of the race and the name of the racecourse where the race will be held. Also according to the order for reference, the horse races and the lists of runners are not arranged on William Hill's internet sites in the same way as in the BHB database.

20. In March 2000 the BHB and Others brought proceedings against William Hill in the High Court of Justice of England and Wales, Chancery Division, alleging infringement of their *sui generis* right. They contend, first, that each day's use by William Hill of racing data taken from the newspapers or the RDF is an extraction or re-utilisation of a substantial part of the contents of the BHB database, contrary to Article 7(1) of the directive. Secondly, they say that even if the individual extracts made by William Hill are not substantial they should be prohibited under Article 7(5) of the directive. ...

The questions referred

Preliminary observations

23. Article 7(1) of the directive provides for specific protection, called a *sui generis* right, for the maker of a database within the meaning of Article 1(2) of the directive, provided that it 'shows that there has been qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents'.

24. By its second and third questions, which should be considered together, the referring court seeks an interpretation of the concept of investment in the obtaining and verification of the contents of a database within the meaning of Article 7(1) of the directive.

25. Article 7(1) of the directive authorises a maker of a database protected by a *sui generis* right to prevent extraction and/or re-utilisation of the whole or of a substantial part of its contents. Article 7(5) also prohibits the repeated and systematic extraction and/or re-utilisation of insubstantial parts of the contents of the database implying acts which conflict with a normal ex-

ploitation of that database or which unreasonably prejudice the legitimate interests of the maker of the database. ...

The second and third questions, concerning the concept of investment in the obtaining or verification of the contents of a database within the meaning of Article 7(1) of the directive

28. By its second and third questions the referring court seeks clarification of the concept of investment in the obtaining and verification of the contents of a database within the meaning of Article 7(1) of the directive.

29. Article 7(1) of the directive reserves the protection of the *sui generis* right to databases which meet a specific criterion, namely to those which show that there has been qualitatively and/or quantitatively a substantial investment in the obtaining, verification or presentation of their contents.

30. Under the 9th, 10th and 12th recitals of the preamble to the directive, its purpose, as William Hill points out, is to promote and protect investment in data 'storage' and 'processing' systems which contribute to the development of an information market against a background of exponential growth in the amount of information generated and processed annually in all sectors of activity. It follows that the expression 'investment in ... the obtaining, verification or presentation of the contents' of a database must be understood, generally, to refer to investment in the creation of that database as such.

31. Against that background, the expression 'investment in ... the obtaining ... of the contents' of a database must, as William Hill and the Belgian, German and Portuguese Governments point out, be understood to refer to the resources used to seek out existing independent materials and collect them in the database, and not to the resources used for the creation as such of independent materials. The purpose of the protection by the *sui generis* right provided for by the directive is to promote the establishment of storage and processing systems for existing information and not the creation of materials capable of being collected subsequently in a database. ...

33. The 19th recital of the preamble to the directive, according to which the compilation of several recordings of musical performances on a CD does not represent a substantial enough investment to be eligible under the *sui generis* right, provides an additional argument in support of that interpretation. Indeed, it appears from that recital that the resources used for the creation as such of works or materials included in the database, in this case on a CD, cannot be deemed equivalent to investment in the obtaining of the contents of that database and cannot, therefore, be taken into account in assessing whether the investment in the creation of the database was substantial.

34. The expression 'investment in ... the ... verification ... of the contents' of a database must be understood to refer to the resources used, with a view to ensuring the reliability of the information contained in that database, to monitor the accuracy of the materials collected when the database was created and during its operation. The resources used for verification during the stage of creation of data or other materials which are subsequently collected in a database, on the other hand, are resources used in creating a database and cannot therefore be taken into account in order to assess whether there was substantial investment in the terms of Article 7(1) of the directive.

35. In that light, the fact that the creation of a database is linked to the exercise of a principal activity in which the person creating the database is also the creator of the materials contained in the database does not, as such, preclude that person from claiming the protection of the *sui generis* right, provided that he establishes that the obtaining of those materials, their verification or their presentation, in the sense described in paragraphs 31 to 34 of this judgment, required substantial investment in quantitative or qualitative terms, which was independent of the resources used to create those materials.
36. Thus, although the search for data and the verification of their accuracy at the time a database is created do not require the maker of that database to use particular resources because the data are those he created and are available to him, the fact remains that the collection of those data, their systematic or methodical arrangement in the database, the organisation of their individual accessibility and the verification of their accuracy throughout the operation of the database may require substantial investment in quantitative and/or qualitative terms within the meaning of Article 7(1) of the directive.
37. In the case in the main proceedings, the referring court seeks to know whether the investments described in paragraph 14 of this judgment can be considered to amount to investment in obtaining the contents of the BHB database. The plaintiffs in the main proceedings stress, in that connection, the substantial nature of the above investment.
38. However, investment in the selection, for the purpose of organising horse racing, of the horses admitted to run in the race concerned relates to the creation of the data which make up the lists for those races which appear in the BHB database. It does not constitute investment in obtaining the contents of the database. It cannot, therefore, be taken into account in assessing whether the investment in the creation of the database was substantial.
39. Admittedly, the process of entering a horse on a list for a race requires a number of prior checks as to the identity of the person making the entry, the characteristics of the horse and the classification of the horse, its owner and the jockey.
40. However, such prior checks are made at the stage of creating the list for the race in question. They thus constitute investment in the creation of data and not in the verification of the contents of the database.
41. It follows that the resources used to draw up a list of horses in a race and to carry out checks in that connection do not represent investment in the obtaining and verification of the contents of the database in which that list appears. ...

The seventh, eighth and ninth questions, on the terms 'extraction' and 're-utilisation' in Article 7 of the directive

43. By its seventh, eighth and ninth questions, the referring court seeks essentially to know whether use such as that made by William Hill of a database constitutes extraction and/or re-utilisation within the meaning of Article 7 of the directive. The referring court asks, inter alia, whether the protection conferred by the *sui generis* right also covers the use of data which, although derived originally from a protected database, were obtained by the user from sources other than that database.

44. The protection of the *sui generis* right provided for by Article 7(1) of the directive gives the maker of a database the option of preventing the unauthorised extraction and/or re-utilisation of all or a substantial part of the contents of that database, according to the 41st recital of the preamble to the directive. Furthermore, Article 7(5) of the directive prohibits, under certain conditions, the unauthorised extraction and/or re-utilisation of insubstantial parts of the contents of a database.
45. The terms extraction and re-utilisation must be interpreted in the light of the objective pursued by the *sui generis* right. It is intended to protect the maker of the database against 'acts by the user which go beyond [the] legitimate rights and thereby harm the investment' of the maker, as indicated in the 42nd recital of the preamble to the directive.
46. According to the 48th recital of the preamble to the directive, the *sui generis* right has an economic justification, which is to afford protection to the maker of the database and guarantee a return on his investment in the creation and maintenance of the database.
47. Accordingly, it is not relevant, in an assessment of the scope of the protection of the *sui generis* right, that the act of extraction and/or re-utilisation is for the purpose of creating another database, whether in competition with the original database or not, and whether the same or a different size from the original, nor is it relevant that the act is part of an activity other than the creation of a database. The 42nd recital of the preamble to the directive confirms, in that connection, that 'the right to prohibit extraction and/or re-utilisation of all or a substantial part of the contents relates not only to the manufacture of a parasitical competing product but also to any user who, through his acts, causes significant detriment, evaluated qualitatively or quantitatively, to the investment'.
48. It must also be pointed out that, although the Proposal for a Council Directive on the legal protection of databases (OJ 1992 C 156, p. 4), presented by the Commission on 15 April 1992, restricted the scope of the protection conferred by the *sui generis* right, under Article 2(5), to unauthorised extraction or re-utilisation 'for commercial purposes', the absence of any reference in Article 7 of the directive to such purposes indicates that it is irrelevant, in an assessment of the lawfulness of an act under the directive, whether the act is for a commercial or a non-commercial purpose.
49. In Article 7(2)(a) of the directive, extraction is defined as 'the permanent or temporary transfer of all or a substantial part of the contents of a database to another medium by any means or in any form', while in Article 7(2)(b), re-utilisation is defined as 'any form of making available to the public all or a substantial part of the contents of a database by the distribution of copies, by renting, by on-line or other forms of transmission'.
50. The reference to 'a substantial part' in the definition of the concepts of extraction and re-utilisation gives rise to confusion given that, according to Article 7(5) of the directive, extraction or re-utilisation may also concern an insubstantial part of a database. As the Advocate General observes, in point 90 of her Opinion, the reference, in Article 7(2) of the directive, to the substantial nature of the extracted or re-utilised part does not concern the definition of those concepts as such but must be understood to refer to one of

the conditions for the application of the *sui generis* right laid down by Article 7(1) of the directive.

51. The use of expressions such as 'by any means or in any form' and 'any form of making available to the public' indicates that the Community legislature intended to give the concepts of extraction and re-utilisation a wide definition. In the light of the objective pursued by the directive, those terms must therefore be interpreted as referring to any act of appropriating and making available to the public, without the consent of the maker of the database, the results of his investment, thus depriving him of revenue which should have enabled him to redeem the cost of the investment.
52. Against that background, and contrary to the argument put forward by William Hill and the Belgian and Portuguese Governments, the concepts of extraction and re-utilisation cannot be exhaustively defined as instances of extraction and re-utilisation directly from the original database at the risk of leaving the maker of the database without protection from unauthorised copying from a copy of the database. That interpretation is confirmed by Article 7(2)(b) of the directive, according to which the first sale of a copy of a database within the Community by the rightholder or with his consent is to exhaust the right to control 'resale', but not the right to control extraction and re-utilisation of the contents, of that copy within the Community.
53. Since acts of unauthorised extraction and/or re-utilisation by a third party from a source other than the database concerned are liable, just as much as such acts carried out directly from that database are, to prejudice the investment of the maker of the database, it must be held that the concepts of extraction and re-utilisation do not imply direct access to the database concerned.
54. However, it must be stressed that the protection of the *sui generis* right concerns only acts of extraction and re-utilisation as defined in Article 7(2) of the directive. That protection does not, on the other hand, cover consultation of a database.
55. Of course, the maker of a database can reserve exclusive access to his database to himself or reserve access to specific people. However, if he himself makes the contents of his database or a part of it accessible to the public, his *sui generis* right does not allow him to prevent third parties from consulting that base.
56. The same applies where the maker of the database authorises a third party to re-utilise the contents of his database, in other words, to distribute it to the public. According to the definition of re-utilisation in Article 7(2)(b) of the directive, read in conjunction with the 41st recital of the preamble thereto, the authorisation of the maker for the re-utilisation of the database or a substantial part of it implies that he consents to his database or the relevant part of it being made accessible to the public by the third party to whom he gave that authorisation. In authorising re-utilisation, the maker of the database thus creates an alternative means of access to the contents of and of consultation of his database for those interested.
57. The fact that a database can be consulted by third parties through someone who has authorisation for re-utilisation from the maker of the database does not, however, prevent the maker from recovering the costs of his investment. It is legitimate for the maker to charge a fee for the re-utilisation of the

whole or a part of his database which reflects, inter alia, the prospect of subsequent consultation and thus guarantees him a sufficient return on his investment.

58. On the other hand, a lawful user of a database, in other words, a user whose access to the contents of a database for the purpose of consultation results from the direct or indirect consent of the maker of the database, may be prevented by the maker, under the *sui generis* right provided for by Article 7(1) of the directive, from then carrying out acts of extraction and/or re-utilisation of the whole or a substantial part of the database. The consent of the maker of the database to consultation does not entail exhaustion of the *sui generis* right.

59. That analysis is confirmed, as regards extraction, by the 44th recital of the preamble to the directive, according to which, 'when on-screen display of the contents of a database necessitates the permanent or temporary transfer of all or a substantial part of such contents to another medium, that act should be subject to authorisation by the rightholder'. Similarly, as regards re-utilisation, the 43rd recital of the preamble to the directive states that 'in the case of on-line transmission, the right to prohibit re-utilisation is not exhausted either as regards the database or as regards a material copy of the database or of part thereof made by the addressee of the transmission with the consent of the rightholder'.

60. It should, however, be emphasised that the prohibition in Article 7(1) of the directive concerns only extraction and/or re-utilisation of the whole or of a substantial part of a database whose creation required a substantial investment. According to Article 8(1) of the directive, apart from in the cases referred to in Article 7(5) of the directive, the *sui generis* right does not prevent a lawful user from extracting and re-utilising insubstantial parts of the contents of a database.

61. It follows from the foregoing that acts of extraction, in other words, the transfer of the contents of the database to another medium, and acts of re-utilisation, in other words, the making available to the public of the contents of a database, which affect the whole or a substantial part of the contents of a database require the authorisation of the maker of the database, even where he has made his database, as a whole or in part, accessible to the public or authorised a specific third party or specific third parties to distribute it to the public.

62. The directive contains an exception to the principle set out in the previous paragraph. Article 9 defines exhaustively three cases in which Member States may stipulate that lawful users of a database which is made available to the public in whatever manner may, without the authorisation of its maker, extract or re-utilise a 'substantial part' of the contents of that database. Those cases are: extraction for private purposes of the contents of a non-electronic database, extraction for the purposes of illustration for teaching or scientific research and extraction and/or re-utilisation for the purposes of public security or an administrative or judicial procedure.

63. In the case in the main proceedings, the order for reference states that the data concerning horse races which William Hill displays on its internet site and which originate in the BHB database are obtained, first, from newspa-

pers published the day before the race and, second, from the RDF supplied by SIS.

64. According to the order for reference, the information published in the newspapers is supplied to the press directly by Weatherbys Group Ltd, the company which maintains the BHB database. As regards William Hill's other source of information, it must be borne in mind that SIS is authorised by Racing Pages Ltd, which is partly controlled by Weatherbys Group Ltd, to supply information concerning horse races in the form of RDF to its own members, which include William Hill. The data in the BHB database concerning horse races have thus been made accessible to the public for the purpose of consultation with the authorisation of BHB.
65. Although William Hill is a lawful user of the database made accessible to the public, at least as regards the part of that database representing information about races, it appears from the order for reference that it carries out acts of extraction and re-utilisation within the meaning of Article 7(2) of the directive. First, it extracts data originating in the BHB database by transferring them from one medium to another. It integrates those data into its own electronic system. Second, it re-utilises those data by then making them available to the public on its internet site in order to allow its clients to bet on horse races.
66. According to the order for reference, that extraction and re-utilisation was carried out without the authorisation of BHB and Others. Since the present case does not fall within any of the cases described in Article 9 of the directive, acts such as those carried out by William Hill could be prevented by BHB and Others under their *sui generis* right provided that they affect the whole or a substantial part of the contents of the BHB database within the meaning of Article 7(1) of the directive. If such acts affected insubstantial parts of the database they would be prohibited only if the conditions in Article 7(5) of the directive were fulfilled. ...

The first, fourth, fifth and sixth questions, concerning the terms 'substantial part' and 'insubstantial part' of the contents of a database in Article 7 of the directive

68. By its fourth, fifth and sixth questions, the referring court raises the question of the meaning of the terms 'substantial part' and 'insubstantial part' of the contents of a database as used in Article 7 of the directive. By its first question it also seeks to know whether materials derived from a database do not constitute a part, substantial or otherwise, of that database, where their systematic or methodical arrangement and the conditions of their individual accessibility have been altered by the person carrying out the extraction and/or re-utilisation.
69. In that connection, it must be borne in mind that protection by the *sui generis* right covers databases whose creation required a substantial investment. Against that background, Article 7(1) of the directive prohibits extraction and/or re-utilisation not only of the whole of a database protected by the *sui generis* right but also of a substantial part, evaluated qualitatively or

quantitatively, of its contents. According to the 42nd recital of the preamble to the directive, that provision is intended to prevent a situation in which a user 'through his acts, causes significant detriment, evaluated qualitatively or quantitatively, to the investment'. It appears from that recital that the assessment, in qualitative terms, of whether the part at issue is substantial, must, like the assessment in quantitative terms, refer to the investment in the creation of the database and the prejudice caused to that investment by the act of extracting or re-utilising that part.

70. The expression 'substantial part, evaluated quantitatively', of the contents of a database within the meaning of Article 7(1) of the directive refers to the volume of data extracted from the database and/or re-utilised, and must be assessed in relation to the volume of the contents of the whole of that database. If a user extracts and/or re-utilises a quantitatively significant part of the contents of a database whose creation required the deployment of substantial resources, the investment in the extracted or re-utilised part is, proportionately, equally substantial.

71. The expression 'substantial part, evaluated qualitatively', of the contents of a database refers to the scale of the investment in the obtaining, verification or presentation of the contents of the subject of the act of extraction and/or re-utilisation, regardless of whether that subject represents a quantitatively substantial part of the general contents of the protected database. A quantitatively negligible part of the contents of a database may in fact represent, in terms of obtaining, verification or presentation, significant human, technical or financial investment. ...

74. In that regard, it appears from the order for reference that the materials displayed on William Hill's internet sites, which derive from the BHB database, represent only a very small proportion of the whole of that database, as stated in paragraph 19 of this judgment. It must therefore be held that those materials do not constitute a substantial part, evaluated quantitatively, of the contents of that database.

75. According to the order for reference, the information published by William Hill concerns only the following aspects of the BHB database: the names of all the horses running in the race concerned, the date, the time and/or the name of the race and the name of the racecourse, as also stated in paragraph 19 of this judgment.

76. In order to assess whether those materials represent a substantial part, evaluated qualitatively, of the contents of the BHB database, it must be considered whether the human, technical and financial efforts put in by the maker of the database in obtaining, verifying and presenting those data constitute a substantial investment.

77. BHB and Others submit, in that connection, that the data extracted and re-utilised by William Hill are of crucial importance because, without lists of runners, the horse races could not take place. They add that those data represent a significant investment, as demonstrated by the role played by a call centre employing more than 30 operators.

78. However, it must be observed, first, that the intrinsic value of the data affected by the act of extraction and/or re-utilisation does not constitute a relevant criterion for assessing whether the part in question is substantial, evaluated qualitatively. The fact that the data extracted and re-utilised by

William Hill are vital to the organisation of the horse races which BHB and Others are responsible for organising is thus irrelevant to the assessment whether the acts of William Hill concern a substantial part of the contents of the BHB database.

79. Next, it must be observed that the resources used for the creation as such of the materials included in a database cannot be taken into account in assessing whether the investment in the creation of that database was substantial, as stated in paragraphs 31 to 33 of this judgment.

80. The resources deployed by BHB to establish, for the purposes of organising horse races, the date, the time, the place and/or name of the race, and the horses running in it, represent an investment in the creation of materials contained in the BHB database. Consequently, and if, as the order for reference appears to indicate, the materials extracted and re-utilised by William Hill did not require BHB and Others to put in investment independent of the resources required for their creation, it must be held that those materials do not represent a substantial part, in qualitative terms, of the BHB database.

The 10th question, concerning the scope of the prohibition laid down by Article 7(5) of the directive

83. By its 10th question, the referring court seeks to know what type of act is covered by the prohibition laid down by Article 7(5) of the directive. It also seeks to know whether acts such as those carried out by William Hill are covered by that prohibition.

84. On that point, it appears from Article 8(1) and from the 42nd recital of the preamble to the directive that, as a rule, the maker of a database cannot prevent a lawful user of that database from carrying out acts of extraction and re-utilisation of an insubstantial part of its contents. Article 7(5) of the directive, which authorises the maker of the database to prevent such acts under certain conditions, thus provides for an exception to that general rule. ...

86. It follows that the purpose of Article 7(5) of the directive is to prevent circumvention of the prohibition in Article 7(1) of the directive. Its objective is to prevent repeated and systematic extractions and/or re-utilisations of insubstantial parts of the contents of a database, the cumulative effect of which would be to seriously prejudice the investment made by the maker of the database just as the extractions and/or re-utilisations referred to in Article 7(1) of the directive would.

87. The provision therefore prohibits acts of extraction made by users of the database which, because of their repeated and systematic character, would lead to the reconstitution of the database as a whole or, at the very least, of a substantial part of it, without the authorisation of the maker of the database, whether those acts were carried out with a view to the creation of another database or in the exercise of an activity other than the creation of a database. ...

89. Under those circumstances, 'acts which conflict with a normal exploitation of [a] database or which unreasonably prejudice the legitimate interests of the maker of the database' refer to unauthorised actions for the purpose of reconstituting, through the cumulative effect of acts of extraction, the whole or a substantial part of the contents of a database protected by the *sui gener-*

is right and/or of making available to the public, through the cumulative effect of acts of re-utilisation, the whole or a substantial part of the contents of such a database, which thus seriously prejudice the investment made by the maker of the database.

90. In the case in the main proceedings, it is clear, in the light of the information given in the order for reference, that the acts of extraction and re-utilisation carried out by William Hill concern insubstantial parts of the BHB database, as stated in paragraphs 74 to 80 of this judgment. According to the order for reference, they are carried out on the occasion of each race held. They are thus of a repeated and systematic nature.
91. However, such acts are not intended to circumvent the prohibition laid down in Article 7(1) of the directive. There is no possibility that, through the cumulative effect of its acts, William Hill might reconstitute and make available to the public the whole or a substantial part of the contents of the BHB database and thereby seriously prejudice the investment made by BHB in the creation of that database.
92. It must be pointed out in that connection that, according to the order for reference, the materials derived from the BHB database which are published daily on William Hill's internet sites concern only the races for that day and are limited to the information mentioned in paragraph 19 of this judgment. ...
93. As explained in paragraph 80 of this judgment, it appears from the order for reference that the presence, in the database of the claimants, of the materials affected by William Hill's actions did not require investment by BHB and Others independent of the resources used for their creation.
94. It must therefore be held that the prohibition in Article 7(5) of the directive does not cover acts such as those of William Hill. ...

B. Generative AI

IN RE ZARYA OF THE DAWN

United States Copyright Office (Feb. 21, 2023)

https://www.copyright.gov/docs/zarya-of-the-dawn.pdf

[Kristina Kashtanova attempted to register a copyright in a comic book titled *Zarya of the Dawn*. The following is a letter to Kashtanova's attorney, Van Lindberg, from Robert J. Kasunic, the Associate Register of Copyrights and Director of the Office of Registration Policy and Practice at the Copyright Office.]

... I. Description of the Work

As described in the application and accompanying deposit materials provided by Ms. Kashtanova, the Work is a "comic book" consisting of eighteen pages, one of which is a cover. The cover page consists of an image of a young woman, the Work's title, and the words "Kashtanova" and "Midjourney." The remaining pages consist of mixed text and visual material. A reproduction of the cover page and the second page are provided below:

II. Summary of Administrative Record

On September 15, 2022, Ms. Kashtanova submitted an application for the Work and copies of each page of the Work as the deposit copy. In her application, Ms. Kashtanova listed the author of the Work as "Kristina Kashtanova" and stated that she had created a "[c]omic book." The application did not disclose that she used artificial intelligence to create any part of the Work, nor did she disclaim any portion of the Work. The Office reviewed the application on the same day and registered the Work as registration number VAu001480196.

Shortly after registering the Work, the Office became aware of statements on social media attributed to Ms. Kashtanova that she had created the comic book using Midjourney artificial intelligence. Because the application had not disclosed the use of artificial intelligence, the Office determined that the application was incorrect, or at a minimum, substantively incomplete. In a letter dated October 28, 2022, the Office notified Ms. Kashtanova that it intended to cancel the registration unless she provided additional information in writing showing why the registration should not be cancelled.

On November 21, 2022, the Office received a timely response from Ms. Kashtanova's attorney. The letter describes Ms. Kashtanova's creation of the Work, including specific information about her use of Midjourney. Mr. Lindberg argues that the Work's registration should not be cancelled because (1) Ms. Kashtanova authored every aspect of the work, with Midjourney serving merely as an assistive tool, and, (2) alternatively, portions of the work are registrable because the text was authored by Ms. Kashtanova and the Work is a copyrightable compilation due to her creative selection, coordination, and arrangement of the text and images.

III. Discussion

A. Legal Standards

Before turning to our analysis of the Work, we summarize here the legal principles that guide that analysis. The Copyright Act defines the scope of copyright protec-

tion. Under the Act, a work may be registered if it qualifies as an "original work[] of authorship fixed in any tangible medium of expression." 17 U.S.C. § 102(a). The Supreme Court has explained that the term "original" in this context consists of two components: independent creation and sufficient creativity. *See Feist Publ'ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 345 (1991). First, the work must have been independently created by the author. Second, the work must possess sufficient creativity. Only a modicum of creativity is necessary, but the Supreme Court has ruled that some works—such as the alphabetized telephone directory at issue in *Feist*—fail to meet even this low threshold. The Court observed that "[a]s a constitutional matter, copyright protects only those constituent elements of a work that possess more than a *de minimis* quantum of creativity." *Id.* at 363. It found that there can be no copyright in a work in which "the creative spark is utterly lacking or so trivial as to be virtually nonexistent." *Id.* at 359.

Courts interpreting the phrase "works of authorship" have uniformly limited it to the creations of human authors. For example, in *Burrow-Giles Lithographic Co. v. Sarony*, the Supreme Court held that photographs were protected by copyright because they were "representatives of original intellectual conceptions of the author," defining authors as "he to whom anything owes its origin; originator; maker; one who completes a work of science or literature." 111 U.S. 53, 57–59 (1884). In doing so, the Court rejected the argument that a photograph was merely "a reproduction on paper of the exact features of some natural object or of some person" made by a machine. *Id.* at 56. But the Court explained that if photography was a "merely mechanical" process, "with no place for novelty, invention or originality" by the human photographer, then "in such case a copyright is no protection." *Id.* at 59.

In cases where non-human authorship is claimed, appellate courts have found that copyright does not protect the alleged creations. For example, the Ninth Circuit held that a book containing words "'authored' by non-human spiritual beings" can only gain copyright protection if there is "human selection and arrangement of the revelations." *Urantia Found. v. Kristen Maaherra*, 114 F.3d 955, 957–59 (9th Cir. 1997). The *Urantia* court held that "some element of human creativity must have occurred in order for the Book to be copyrightable" because "it is not creations of divine beings that the copyright laws were intended to protect." *Id.*

The Office's registration practices follow and reflect these court decisions. The Office collects its understanding of the law in the *Compendium of U.S. Copyright Office Practices (Third Edition)*, which provides standards for examining and registering copyrighted works. Following the cases described above, the *Compendium* explains that the Office "will refuse to register a claim if it determines that a human being did not create the work." *Compendium* § 313.2 (3d ed. 2021) (providing examples of works lacking human authorship such as "a photograph taken by a monkey" and "an application for a song naming the Holy Spirit as the author of the work").

Having considered the requirements for copyright protection, the Office turns to the elements of the Work as described in your letter.

B. The Work's Text

The Office agrees that the text of the Work is protected by copyright. Your letter states that "the text of the Work was written entirely by Kashtanova without the help of any other source or tool, including any generative AI program." Based on this statement, the Office finds that the text is the product of human authorship. Moreover, the Office finds that the text in the Work contains more than the "mod-

icum of creativity" required for protection under *Feist.* For this reason, the text of the Work is registrable.

C. The Selection and Arrangement of Images and Text

The Office also agrees that the selection and arrangement of the images and text in the Work are protectable as a compilation. Copyright protects "the collection and assembling of preexisting materials or of data that are selected, coordinated, or arranged" in a sufficiently creative way. 17 U.S.C. § 101 (definition of "compilation"). Ms. Kashtanova states that she "selected, refined, cropped, positioned, framed, and arranged" the images in the Work to create the story told within its pages. Based on the representation that the selection and arrangement of the images in the Work was done entirely by Ms. Kashtanova, the Office concludes that it is the product of human authorship. Further, the Office finds that the compilation of these images and text throughout the Work contains sufficient creativity under *Feist* to be protected by copyright. Specifically, the Office finds the Work is the product of creative choices with respect to the selection of the images that make up the Work and the placement and arrangement of the images and text on each of the Work's pages. Copyright therefore protects Ms. Kashtanova's authorship of the overall selection, coordination, and arrangement of the text and visual elements that make up the Work.

D. The Individual Images

Turning to the individual images in the Work, the Office must consider the impact of Ms. Kashtanova's use of Midjourney's artificial intelligence technology in its copyrightability analysis. The majority of the Kashtanova Letter focuses on how Ms. Kashtanova used Midjourney to create these images. Before addressing the question of whether the images are copyrightable, the Office describes its understanding of Midjourney and how it works. ...

1. How Midjourney Works

Midjourney offers an artificial intelligence technology capable of generating images in response to text provided by a user. Midjourney operates on top of an unaffiliated third-party communication service called Discord, which is made up of individual servers operated by its users. In order to use Midjourney, users must first join the Midjourney Discord server, which contains public "channels" where users can enter text. Midjourney primarily operates through an automated account on these channels that reads user-entered text and generates images based on it. ...

Users operate Midjourney through "prompts," which are text commands entered in one of Midjourney's channels. As Midjourney explains, prompts must start with the text "/imagine" and contain text describing what Midjourney should generate. Users also have the option to include (1) a URL of one or more images to influence the generated output, or (2) parameters directing Midjourney to generate an image in a particular aspect ratio or providing other functional directions.

After a user provides Midjourney with a prompt, the technology will generate four images in response. The images are provided in a grid, and buttons underneath the grid allow users to request that Midjourney provide a higher-resolution version of an image (*e.g.*, U1, U2, U3, U4), create new variations of an image (*e.g.*, V1, V2, V3, V4), or to generate four new images from scratch (see light blue circular icon at far right below). For example, entering the prompt "/imagine cute baby

dinosaur shakespeare writing play purple" resulted in the following response from

Midjourney:

It is relevant here that, by its own description, Midjourney does not interpret prompts as specific instructions to create a particular expressive result. Because Midjourney "does not understand grammar, sentence structure, or words like humans," it instead converts words and phrases "into smaller pieces, called tokens, that can be compared to its training data and then used to generate an image." *Prompts*, Midjourney, https://docs.midjourney.com/docs/prompts. Generation involves Midjourney starting with "a field of visual noise, like television static, [used] as a starting point to generate the initial image grids" and then using an algorithm to refine that static into human-recognizable images. *Seeds*, Midjourney, https://docs.midjourney.com/docs/seeds.

The process by which a Midjourney user obtains an ultimate satisfactory image through the tool is not the same as that of a human artist, writer, or photographer. As noted above, the initial prompt by a user generates four different images based on Midjourney's training data. While additional prompts applied to one of these initial images can influence the subsequent images, the process is not controlled by the user because it is not possible to predict what Midjourney will create ahead of time.

2. Application of Copyright Law to Midjourney Images

Based on the record before it, the Office concludes that the images generated by Midjourney contained within the Work are not original works of authorship protected by copyright. Though she claims to have "guided" the structure and content of each image, the process described in the Kashtanova Letter makes clear that it was Midjourney—not Kashtanova—that originated the "traditional elements of authorship" in the images.

Ms. Kashtanova claims that each image was created using "a similar creative process." Summarized here, this process consisted of a series of steps employing Midjourney. First, she entered a text prompt to Midjourney, which she describes as "the core creative input" for the image. [E.g., "dark skin hands holding an old photograph --ar 16:9"] Next, "Kashtanova then picked one or more of these output images to further develop." She then "tweaked or changed the prompt as well as the other inputs provided to Midjourney" to generate new intermediate images, and ultimately the final image. Ms. Kashtanova does not claim she created any visual material herself—she uses passive voice in describing the final image as "created, developed, refined, and relocated" and as containing elements from intermediate images "brought together into a cohesive whole." To obtain the final image, she describes a process of trial-and-error, in which she provided "hundreds or thousands of descriptive prompts" to Midjourney until the "hundreds of iterations [created] as perfect a rendition of her vision as possible."

Rather than a tool that Ms. Kashtanova controlled and guided to reach her desired image, Midjourney generates images in an unpredictable way. Accordingly, Midjourney users are not the "authors" for copyright purposes of the images the technology generates. As the Supreme Court has explained, the "author" of a copyrighted work is the one "who has actually formed the picture," the one who acts as "the inventive or master mind." *Burrow-Giles,* 111 U.S. at 61. A person who provides text prompts to Midjourney does not "actually form" the generated images and is not the "master mind" behind them. Instead, as explained above, Midjourney begins the image generation process with a field of visual "noise," which is refined based on tokens created from user prompts that relate to Midjourney's training database. The information in the prompt may "influence" generated image, but prompt text does not dictate a specific result. Because of the significant distance between what a user may direct Midjourney to create and the visual material Midjourney actually produces, Midjourney users lack sufficient control over generated images to be treated as the "master mind" behind them.

The fact that Midjourney's specific output cannot be predicted by users makes Midjourney different for copyright purposes than other tools used by artists. Like the photographer in *Burrow-Giles*, when artists use editing or other assistive tools, they select what visual material to modify, choose which tools to use and what changes to make, and take specific steps to control the final image such that it amounts to the artist's "own original mental conception, to which [they] gave visible form."[15] *Burrow-Giles,* 111 U.S. at 60 (explaining that the photographer's creative choices made the photograph "the product of [his] intellectual invention"). Users of Midjourney do not have comparable control over the initial image generated, or any final image. It is therefore understandable that users like Ms. Kash-

15 For this reason, the cases cited by Ms. Kashtanova regarding Photoshop do not alter our conclusion. Both cases involved situations where the artist had made deliberate, intentional edits to an image using Photoshop. In *Etrailer Corp. v. Onyx Enters., Int'l Corp*., the court credited the plaintiff's statement that she used Photoshop to "smooth, crop, saturate, and burn" photographs of trailer accessories. Case No. 4:17-CV-01284-AGF, 2018 U.S. Dist. LEXIS 19916, at *4 (E.D. Mo. Feb. 7, 2018) (rejecting motion to dismiss that photographs were not protected by copyright). And in *Payton v. Defend, Inc.*, the court found a triable issue on copyrightability where the plaintiff used Photoshop to create a shirt design containing a silhouette of an AR-15 rifle based on a preexisting "picture of a model AR-15 Airsoft gun." No. 15-00238 SOM/KSC, 2017 U.S. Dist. LEXIS 208358, at *9 (D. Haw. Dec. 19, 2017).

tanova may take over a year from conception to creation of images matching what the user had in mind because they may need to generate hundreds of intermediate images.

Nor does the Office agree that Ms. Kashtanova's use of textual prompts permits copyright protection of resulting images because the images are the visual representation of "creative, human-authored prompts."[16] *Id.* at 10. Because Midjourney starts with randomly generated noise that evolves into a final image, there is no guarantee that a particular prompt will generate any particular visual output. Instead, prompts function closer to suggestions than orders, similar to the situation of a client who hires an artist to create an image with general directions as to its contents. If Ms. Kashtanova had commissioned a visual artist to produce an image containing "a holographic elderly white woman named Raya," where "[R]aya is having curly hair and she is inside a spaceship," with directions that the image have a similar mood or style to a "Star Trek spaceship," "a hologram," an "octane render," "unreal engine," and be "cinematic" and "hyper detailed," Ms. Kashtanova would not be the author of that image. *See id.* at 8 (text of prompt provided to Midjourney). Absent the legal requirements for the work to qualify as a work made for hire, the author would be the visual artist who received those instructions and determined how best to express them. And if Ms. Kashtanova were to enter those terms into an image search engine, she could not claim the images returned in response to her search were "authored" by her, no matter how similar they were to her artistic vision.

The Office does not question Ms. Kashtanova's contention that she expended significant time and effort working with Midjourney. But that effort does not make her the "author" of Midjourney images under copyright law. Courts have rejected the argument that "sweat of the brow" can be a basis for copyright protection in otherwise unprotectable material. The Office will not consider the amount of time, effort, or expense required to create the work because they have no bearing on whether a work possesses the minimum creative spark required by the Copyright Act and the Constitution.

The Office's determination here is based on the specific facts provided about Ms. Kashtanova's use of Midjourney to create the Work's images. It is possible that other AI offerings that can generate expressive material operate differently than Midjourney does. However, on the administrative record before the Office, Ms. Kashtanova is not the author for copyright purpose of the individual images generated by Midjourney.

3. Images Edited by Ms. Kashtanova

Finally, Ms. Kashtanova suggests that she personally edited some of the images created by Midjourney. Her letter points to two specific images contained in the Work. While the Office accepts the statement that the changes were made directly by Ms. Kashtanova, it cannot definitively conclude that the editing alterations are sufficiently creative to be entitled to copyright.

First, Ms. Kashtanova explains that she "modified the rendering of Zarya's lips and mouth" in an image on page 2 of the Work.

---

16 While Ms. Kashtanova suggests that her text prompts are copyrightable because they are similar to poems, she did not submit them in the application and is not seeking to register the text prompts themselves, either separately or as part of the Work. Accordingly, the Office has not addressed the question of copyrightability of prompts here.

Detail before Photoshop

Detail after Photoshop

The changes to Zarya's mouth, particularly her upper lip, are too minor and imperceptible to supply the necessary creativity for copyright protection. The Office will register works that contain otherwise unprotectable material that has been edited, modified, or otherwise revised by a human author, but only if the new work contains a sufficient amount of original authorship"to itself qualify for copyright protection. Ms. Kashtanova's changes to this image fall short of this standard. *Contra Eden Toys, Inc. v. Florelee Undergarment Co.*, 697 F.2d 27, 34–35 (2d Cir. 1982) (revised drawing of Paddington Bear qualified as a derivative work based on the changed proportions of the character's hat, the elimination of individualized fingers and toes, and the overall smoothing of lines that gave the drawing a "different, cleaner 'look'"). …

Second, Ms. Kashtanova points to an image on page 12 of the Work depicting an old woman with her eyes closed. She describes this work as created "using both the Midjourney service and Photoshop together," with edits in Photoshop made to "show[] aging of the face, smoothing of gradients[,] and modifications of lines and shapes." …

Based on Ms. Kashtanova's description, the Office cannot determine what expression in the image was contributed through her use of Photoshop as opposed to generated by Midjourney. She suggests that Photoshop was used to modify an intermediate image by Midjourney to "show[] aging of the face," but it is unclear whether she manually edited the youthful face in a previous intermediate image, created a composite image using a previously generated image of an older woman, or did something else. To the extent that Ms. Kashtanova made substantive edits to an intermediate image generated by Midjourney, those edits could provide human authorship and would not be excluded from the new registration certificate.

IV . Conclusion

For the reasons explained above, the Office concludes that the registration certificate for *Zarya of the Dawn*, number VAu001480196 was issued based on inaccurate and incomplete information. Had the Office known the information now provided by Ms. Kashtanova, it would have narrowed the claim to exclude material generated by artificial intelligence technology. In light of the new information, the Office will cancel the previous registration pursuant to 37 C.F.R, § 201.7(c)(4) and replace it with a new registration covering the original authorship that Ms. Kash-

tanova contributed to this work, namely, the "text" and the "selection, coordination, and arrangement of text created by the author and artwork generated by artificial intelligence." Because these contributions predominantly contain textual material, they will be reregistered as an unpublished literary work. The new registration will explicitly exclude "artwork generated by artificial intelligence." ...

BARTZ V. ANTHROPIC PBC

787 F.Supp.3d 1007 (2025)

William Alsup, United States District Judge:

Introduction

An artificial intelligence firm downloaded for free millions of copyrighted books in digital form from pirate sites on the internet. The firm also purchased copyrighted books (some overlapping with those acquired from the pirate sites), tore off the bindings, scanned every page, and stored them in digitized, searchable files. All the foregoing was done to amass a central library of "all the books in the world" to retain "forever." From this central library, the AI firm selected various sets and subsets of digitized books to train various large language models under development to power its AI services. Some of these books were written by plaintiff authors, who now sue for copyright infringement. On summary judgment, the issue is the extent to which any of the uses of the works in question qualify as "fair uses" under Section 107 of the Copyright Act.

Statement

Defendant Anthropic PBC is an AI software firm founded by former OpenAI employees in January 2021. Its core offering is an AI software service called Claude. When a user prompts Claude with text, Claude quickly responds with text—mimicking human reading and writing. Claude can do so because Anthropic trained Claude—or rather trained large language models or LLMs underlying various versions of Claude—using books and other texts selected from a central library Anthropic had assembled. Claude was first released publicly in March 2023. Seven successive versions of Claude have been released since. Users may ask Claude some questions for free. Demanding users and corporate clients pay to use Claude, generating over one billion dollars in annual revenue.

Plaintiffs Andrea Bartz, Charles Graeber, and Kirk Wallace Johnson are authors of books that Anthropic copied from pirated and purchased sources. Anthropic assembled these copies into a central library of its own, copied further various sets and subsets of those library copies to include in various "data mixes," and used these mixes to train various LLMs. Anthropic kept the library copies in place as a permanent, general-purpose resource even after deciding it would not use certain copies to train LLMs or would never use them again to do so. All of Anthropic's copying was without plaintiffs' authorization. ...

From the start, Anthropic "had many places from which" it could have purchased books, but it preferred to steal them to avoid "legal/practice/business slog," as cofounder and chief executive officer Dario Amodei put it. So, in January or February 2021, another Anthropic cofounder, Ben Mann, downloaded Books3, an online library of 196,640 books that he knew had been assembled from unauthorized copies of copyrighted books—that is, pirated. Anthropic's next pirated acquisitions involved downloading distributed, reshared copies of other pirate libraries. In June 2021, Mann downloaded in this way at least five million copies of books from Library Genesis, or LibGen, which he knew had been pirated. And, in July

2022, Anthropic likewise downloaded at least two million copies of books from the Pirate Library Mirror, or PiLiMi, which Anthropic knew had been pirated. Although what was downloaded and later duplicated from these sources was sometimes referred to as data or datasets, at bottom they contained full-text ebooks or scans of books saved in individual files in formats like .pdf, .txt, and .epub. For Books3, most filenames identified the book inside. For LibGen and PiLiMi, Anthropic downloaded a separate catalog of bibliographic metadata for each collection, with fields like title, author, and ISBN. Anthropic thereby pirated over seven million copies of books, including copies of at least two works at issue for each Author.

As Anthropic trained successive LLMs, it became convinced that using books was the most cost-effective means to achieve a world-class LLM. During this time, however, Anthropic became "not so gung ho about" training on pirated books "for legal reasons." It kept them anyway.

To find a new way to get books, in February 2024, Anthropic hired the former head of partnerships for Google's book-scanning project, Tom Turvey. He was tasked with obtaining "all the books in the world" while still avoiding as much "legal/practice/business slog" as possible. So, in spring 2024, Turvey sent an email or two to major publishers to inquire into licensing books for training AI. Had Turvey kept up those conversations, he might have reached agreements to license copies for AI training from publishers—just as another major technology company soon did with one major publisher. But Turvey let those conversations wither.

Instead, Turvey and his team emailed major book distributors and retailers about bulk-purchasing their print copies for the AI firm's "research library." Anthropic spent many millions of dollars to purchase millions of print books, often in used condition. Then, its service providers stripped the books from their bindings, cut their pages to size, and scanned the books into digital form—discarding the paper originals. Each print book resulted in a PDF copy containing images of the scanned pages with machine-readable text (including front and back cover scans for softcover books). Anthropic created its own catalog of bibliographic metadata for the books it was acquiring. It acquired copies of millions of books, including of all works at issue for all Authors.

Anthropic may have copied portions of Authors' books on other occasions, too—such as while copying book reviews, academic papers, internet blogposts, or the like for its central library. And, Anthropic's scanning service providers may have copied Authors' print books along the way to delivering the final digital copies to Anthropic. But neither side here specifically raises legal issues implicated by any such copies. Nor will this order.

From all the above sources, Anthropic created a general "research library" or "generalized data area." What was this for? As Turvey said, this was a "way of creating information that would be voluminous and that we would use for research," or otherwise to "inform our—our products." The copies were kept in the original "version of the underlying" book files Anthropic had "obtained or created," that is, pirated or scanned. Anthropic planned to "store everything forever; we might separate out books into categories, but there was no compelling reason to delete a book"—even if not used for training LLMs. Over time, Anthropic invested in building more tools for searching its "general purpose" library and for accessing books or sets of books for further uses.

One further use was training LLMs. As a preliminary step towards training, engineers browsed books and bibliographic metadata to learn what languages the

books were written in, what subjects they concerned, whether they were by famous authors or not, and so on—sometimes by "opening any of the books" and sometimes using software. From the library copies, engineers copied the sets or subsets of books they believed best for training and "iterated" on those selections over time. For instance, two different subsets of print-sourced books were included in "data mixes" for training two different LLMs. Each was just a fraction of all the print-sourced books. Similarly, different sets or "subsets" or "parts of" or "portions" of the collections sourced from Books3, LibGen, and PiLiMi were used to train different LLMs. Anthropic analyzed the consequences of using more books, fewer books, different books. The goal was to improve the "data mix" to improve each LLM and, ultimately, Claude's performance for paying customers.

Over time, Anthropic came to value most highly for its data mixes books like the ones Authors had written, and it valued them because of the creative expressions they contained. Claude's customers wanted Claude to write as accurately and as compellingly as Authors. So, it was best to train the LLMs underlying Claude on works just like the ones Authors had written, with well-curated facts, well-organized analyses, and captivating fictional narratives—above all with "good writing" of the kind "an editor would approve of." Anthropic could have trained its LLMs without using such books or any books at all. That would have required spending more on, say, staff writers to create competing exemplars of good writing, engineers to revise bad exemplars into better ones, energy bills to power more rounds of training and fine-tuning, and so on. Having canonical texts to draw upon helped.

Each work selected for training any given LLM was copied in four main ways—and in fact so many times that Anthropic admits it would be impractical even to estimate.

First, each work selected was copied from the central library to create a working copy for the training set.

Second, each work was cleaned to remove a small amount of lower-valued or repeating text (like headers, footers, or page numbers), with a "cleaned" copy resulting. If the same book appeared twice, or if while looking across the entire provisional training set it became clear there was some other reason to cull a book or category, Anthropic had the capability to delete relevant copy(ies) from the set at this step.

Third, each cleaned copy was translated into a "tokenized" copy. Some words were "stemmed" or "lemmatized" into simpler forms (e.g., "studying" to "study"). And, all characters were grouped into short sequences and translated into corresponding number sequences or "tokens" according to an Anthropic-made dictionary. The resulting tokenized copies were then copied repeatedly during training. By one account, this process involved the iterative, trial-and-error discovery of contingent statistical relationships between each word fragment and all other word fragments both within any work and across trillions of word fragments from other copied books, copied websites, and the like. Other steps in training are not at issue here.

Fourth, each fully trained LLM itself retained "compressed" copies of the works it had trained upon, or so Authors contend and this order takes for granted. In essence, each LLM's mapping of contingent relationships was so complete it mapped or indeed simply "memorized" the works it trained upon almost verbatim. So, if each completed LLM had been asked to recite works it had trained upon, it could have done so. Further steps refining the LLM are not at issue here.

However, that was as far as the training copies propagated towards the outside world. When each LLM was put into a public-facing version of Claude, it was complemented by other software that filtered user inputs to the LLM and filtered outputs from the LLM back to the user.As a result, Authors do not allege that *any* infringing copy of their works was or would ever be provided to users by the Claude service. Yes, Claude could help less capable writers create works as well-written as Authors' and competing in the same categories. But Claude created no exact copy, nor any substantial knock-off. Nothing traceable to Authors' works. Such allegations are simply not part of plaintiffs' amended complaint, nor in our record.

Neither side puts directly at issue any copies of any works that might have been used for the filtering software. Nor will this order.

In sum, the copies of books pirated or purchased-and-destructively-scanned were placed into a central "research library" or "generalized data area," sets or subsets were copied again to create training copies for data mixes, the training copies were successively copied to be cleaned, tokenized, and compressed into any given trained LLM, and once trained an LLM did not output through Claude to the public any further copies. Finally, once Anthropic decided a copy of a pirated or scanned book in the library would not be used for training at all or ever again, Anthropic still retained that work as a "hard resource" for other uses or future uses. At least one work from each Author was present in every phase described above.

* * *

In August 2024, the three individual authors brought this putative class action complaining that Anthropic had infringed its federal copyrights by pirating copies for its library and by reproducing them to train its LLMs. …

Anthropic now moves for summary judgment on fair use only. …

Notably, in its motion, Anthropic argues that pirating initial copies of Authors' books and millions of other books was justified because all those copies were at least reasonably necessary for training LLMs—and yet Anthropic has resisted putting into the record what copies or even sets of copies were in fact used for training LLMs. For example, at oral argument, Anthropic asserted that if a purported fair user had retained pirated copies for uses beyond the fair use, then her piracy would not be excused by the fair use. But when Authors earlier interrogated Anthropic in discovery about what library copies (the original copies "obtained or created" by Anthropic) Anthropic had recopied for further uses, Anthropic responded that providing information about any copies made for uses beyond training commercially released LLMs would be overbroad, and that it could not count up all its copying even for LLMs in any case. We know that Anthropic has more information about what it in fact copied for training LLMs (or not). Anthropic earlier *produced* a spreadsheet that showed the composition of various data mixes used for training various LLMs—yet it clawed back that spreadsheet in April. A discovery dispute regarding that spreadsheet remains pending. But Anthropic did not need a court order to offer up what it possessed in support of its motion. All deficiencies must be held against Anthropic and not the other way around. …

To summarize the analysis that now follows, the use of the books at issue to train Claude and its precursors was exceedingly transformative and was a fair use under Section 107 of the Copyright Act. And, the digitization of the books purchased in print form by Anthropic was also a fair use but not for the same reason as applies to the training copies. Instead, it was a fair use because all Anthropic did was replace the print copies it had purchased for its central library with more

convenient space-saving and searchable digital copies for its central library—without adding new copies, creating new works, or redistributing existing copies. However, Anthropic had no entitlement to use pirated copies for its central library. Creating a permanent, general-purpose library was not itself a fair use excusing Anthropic's piracy.

Analysis

Section 107 of the Copyright Act identifies four factors for determining whether a given use of a copyrighted work is a fair use:

> [T]he fair use of a copyrighted work ... for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright. In determining whether the use made of a work in any particular case is a fair use the factors to be considered shall include —
>
> (1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;
>
> (2) the nature of the copyrighted work;
>
> (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
>
> (4) the effect of the use upon the potential market for or value of the copyrighted work.

These factors presuppose a "use." So, at the threshold, a court must decide whether a "copyrighted [work] has been used in multiple ways," then evaluate each. *Warhol,* 598 U.S. at 533. Uses do not turn on "the subjective intent of the user" but on "an objective inquiry into what use was made, *i.e.,* what the user did with the original work." *Id.* at 544-45. A "use" should be construed narrowly enough to not "swallow" distinguishable infringing uses, much less categories of exclusive rights *in toto. Id.* at 541, 543 n.18, 546-48. Sometimes, the challenged copying involves just one use: In *Perfect 10, Inc. v. Amazon.com, Inc.,* Google visited websites having full-sized images, made only reduced-sized copies, and incorporated those directly into its search engine—the sole use of the thumbnails being as "pointers" to the images themselves. 508 F.3d 1146, 1157, 1160, 1165 (9th Cir. 2007). Sometimes, the copying involves many uses: In the *Google Books* cases, Google borrowed books from libraries, made both full-image and text-only copies, and incorporated different copies into different tools—one use being to reveal information "*about those books,*" another use being to provide the books to print-disabled patrons, and still another being to back up the print books if lost. *Authors Guild v. Google, Inc.,* 804 F.3d 202, 217 (2d Cir. 2015) (quoted); *Authors Guild, Inc. v. HathiTrust,* 755 F.3d 87, 97, 101, 103 (2d Cir. 2014) (other cited uses).

Our parties debate an instructive decision. In *American Geophysical Union v. Texaco Inc.,* Texaco employees used scientific articles in a central library, used copies of them in personal desk libraries, and used selected copies again in the scientific laboratory—the first use paid for, the second infringing, and the third plausibly fair but in fact a rare occurrence. 802 F. Supp. 1, 4-5, 14 (S.D.N.Y. 1992) (Judge Pierre Leval), *aff'd,* 60 F.3d 913, 918-19, 926 (2d Cir. 1994).

Here, our parties contest what use or uses are at issue. Anthropic contends it copied Authors' books only for *one* use: *Only* to train LLMs. By contrast, Authors contend it did so for at least *two* uses: *First* to build a vast, central library of potentially useful content, and *second* to train specific LLMs using shifting sets and sub-

sets of that content—over time selecting the more well-organized and well-expressed works for training. Authors also complain that the print-to-digital format change was itself an infringement not abridged as a fair use. Authors do not allege, however, that any LLM outputs infringing upon their works ever reached users of the public-facing Claude service.

This order addresses each of the four factors in turn, pointing out how each applies to the training copies and to the purchased and pirated library copies. It concludes with an integrated analysis.

1. The Purpose and Character of the Use

For a given use at issue, the first factor addresses "the purpose and character of th[at] use, including whether [it] is of a commercial nature or is for nonprofit educational purposes." 17 U.S.C. § 107(1).

A. The Copies Used to Train Specific LLMs

All agree that one use at issue was training LLMs to receive text inputs and return text outputs. More specifically, Anthropic used copies of Authors' copyrighted works to iteratively map statistical relationships between every text-fragment and every sequence of text-fragments so that a completed LLM could receive new text inputs and return new text outputs as if it were a human reading prompts and writing responses. Authors further argue—and this order takes for granted—that such training entailed "memorizing" works by "compressing" copies of those works into the LLM. The LLMs "memorized A LOT, like A LOT." Regardless, the "purpose and character" of using works to train LLMs was transformative—spectacularly so.

To repeat and be clear: Authors do not allege that any LLM output provided to users infringed upon Authors' works. Our record shows the opposite. Users interacted only with the Claude service, which placed additional software between the user and the underlying LLM to ensure that no infringing output ever reached the users. This was akin to the limits Google imposed on how many snippets of text from any one book could be seen by any one user through its Google Books service, preventing its search tool from devolving into a reading tool. Here, if the outputs seen by users had been infringing, Authors would have a different case. And, if the outputs were ever to become infringing, Authors could bring such a case. But that is not this case.

Instead, Authors challenge only the inputs, not the outputs, of these LLMs. They point to the fully trained LLMs and the Claude service only to shed light on how training itself uses copies of their works and the ways the Claude service could be used to produce still other works that would compete with their works. This order does the same. Authors' arguments that the training use is not transformative are unavailing.

First, Authors argue that using works to train Claude's underlying LLMs was like using works to train any person to read and write, so Authors should be able to exclude Anthropic from this use. But Authors cannot rightly exclude anyone from using their works for training or learning as such. Everyone reads texts, too, then writes new texts. They may need to pay for getting their hands on a text in the first instance. But to make anyone pay specifically for the use of a book each time they read it, each time they recall it from memory, each time they later draw upon it when writing new things in new ways would be unthinkable. For centuries, we have read and re-read books. We have admired, memorized, and internalized their

sweeping themes, their substantive points, and their stylistic solutions to recurring writing problems.

Second, to that last point, Authors further argue that the training was intended to memorize their works' creative elements—not just their works' non-protectable ones. But this is the same argument. Again, Anthropic's LLMs have not reproduced to the public a given work's creative elements, nor even one author's identifiable expressive style (assuming arguendo that these are even copyrightable). Yes, Claude has outputted grammar, composition, and style that the underlying LLM distilled from thousands of works. But if someone were to read all the modern-day classics because of their exceptional expression, memorize them, and then emulate a blend of their best writing, would that violate the Copyright Act? Of course not. Copyright does not extend to "method[s] of operation, concept[s], [or] principle[s]" "illustrated[] or embodied in [a] work." 17 U.S.C. § 102(b); *see, e.g., Nichols v. Universal Pictures Corp.,* 45 F.2d 119, 120-22 (2d Cir. 1930) (Judge Learned Hand) (stage properties and storytelling elements); *Apple Comput., Inc. v. Microsoft Corp.,* 35 F.3d 1435, 1445 (9th Cir. 1994) ("user-friendly" design principles and elements); *Swirsky v. Carey,* 376 F.3d 841, 848 (9th Cir. 2004) (music theory principles and chord progressions).

Third, Authors next argue that computers nonetheless should not be allowed to do what people do.

Authors cite a decision seeming to say as much. But the judge there twice emphasized while discussing "purpose and character" of the use that what was trained was "not generative AI (AI that writes new content itself)." Rather, what was trained—using a proprietary system for finding court opinions in response to a given legal topic—was a competing AI tool for finding court opinions in response to a given legal topic. That was not transformative. *Thomson Reuters Enter. Centre GmbH v. Ross Intell. Inc.,* 765 F. Supp. 3d 382, 398 (D. Del. 2025) (Judge Stephanos Bibas), *appeal docketed,* No. 25-8018 (3d Cir. Apr. 14, 2025).

A better analogue to our facts would be an AI tool trained—using court opinions, and briefs, law review articles, and the like—to receive legal prompts and respond with fresh legal writing. And, on facts much like those, a different court came out the other way. It found fair use. *White v. W. Pub. Corp.,* 29 F. Supp. 3d 396, 400 (S.D.N.Y. 2014) (Judge Jed Rakoff).

The latter use stood sufficiently "orthogonal" to anything that any copyright owner rightly could expect to control. *See Warhol,* 598 U.S. at 538-40, 143 S.Ct. 1258. It could thus be freed up for the copyist to use, "promot[ing] the progress of science and the arts, *without* diminishing the incentive to create." *Id.* at 531, 143 S.Ct. 1258 (emphasis added); *see* U.S. Const. art. I, § 8, cl. 8.

In short, the purpose and character of using copyrighted works to train LLMs to generate new text was quintessentially transformative. Like any reader aspiring to be a writer, Anthropic's LLMs trained upon works not to race ahead and replicate or supplant them—but to turn a hard corner and create something different. If this training process reasonably required making copies within the LLM or otherwise, those copies were engaged in a transformative use.

The first factor *favors* fair use for the training copies.

B. The Copies Used to Build a Central Library

But that is not the only use at issue. Recall that Anthropic purchased millions of print books for its central library and pirated millions of digital books for its central library, too. It used specific sets and subsets of books for training specific LLMs. And, it then retained all the copies in its central library for other uses that

might arise *even after deciding it would not use them to train any LLM (at all or ever again).* Anthropic seems to believe that because some of the works it copied were sometimes used in training LLMs, Anthropic was entitled to take for free all the works in the world and keep them forever with no further accounting. There is no carveout, however, from the Copyright Act for AI companies.

Because the legal issues differ between the library copies Anthropic purchased and pirated, this order takes them in turn.

(i) The Purchased Library Copies Converted from Print to Digital

Anthropic *purchased* millions of print copies to "build a research library." It destroyed each print copy while replacing it with a digital copy for use in its library (not for sharing nor sale outside the company). As to these copies, Authors do not complain that Anthropic failed to pay to acquire a library copy. Authors only complain that Anthropic changed each copy's format from print to digital. On the facts here, that format change itself added no new copies, eased storage and enabled searchability, and was not done for purposes trenching upon the copyright owner's rightful interests—it was transformative.

Anthropic purchased its print copies fair and square. With each purchase came entitlement for Anthropic to "dispose[]" each copy as it saw fit. 17 U.S.C. § 109(a). So, Anthropic was entitled to keep the copies in its central library for all the ordinary uses. Yes, Anthropic changed the format of these library copies from print to digital—giving rise to the issue here.

All agree on the facts of the format change. Anthropic "destructively scan[ned]" the print copies to create the digital ones. Anthropic or its vendors stripped the bindings from the print books, cut the pages to workable dimensions, and scanned those pages—discarding each print copy while creating a digital one in its place. The digital copy was then housed in the "research library" or "generalized data area" in place of the print copy. Authors do not allege and our record does not show that Anthropic provided its converted digital copies of print books to anyone outside Anthropic.

The parties disagree about the legal consequences of the format change. Was scanning the print copies to create digital replacements transformative? Anthropic argues it was because it was reasonably necessary to training LLMs. Authors argue it was a distinguishable step requiring independent justification.

Here, for reasons narrower than Anthropic offers, the mere format change was a fair use.

Storage and searchability are not creative properties of the copyrighted work itself but physical properties of the frame *around* the work or informational properties *about* the work. *See Texaco,* 802 F. Supp. at 14 (physical), *aff'd,* 60 F.3d at 919; *Google,* 804 F.3d at 225 (informational); *Sony Corp. of Am. v. Universal City Studios, Inc.* ("*Sony Betamax*"), 464 U.S. 417, 447 (1984) (rightful interests). In *Texaco,* the court reasoned that if a purchased scientific journal article had been copied "onto microfilm to conserve space, this might [have been] a persuasive transformative use." 802 F. Supp. at 14 (Judge Pierre Leval), *aff'd,* 60 F.3d at 919 (reducing "bulk" "might suffice to tilt the first fair use factor in favor of Texaco if these purposes were dominant"). In *Google Books,* the court reasoned that a print-to-digital change to expose information about the work was transformative. *Google,* 804 F.3d at 225 (Judge Pierre Leval). And, in *Sony Betamax,* the Supreme Court held that making a recording of a television show in order to instead watch it at a later time was copying but did not usurp any rightful interest of the copyright owner. 464 U.S. at 447, 455. Important to the Supreme Court's reasoning

was the expectation that most such copiers would not distribute the permanent copies of the work. Finally, in *A&M Records, Inc. v. Napster, Inc.*, our court of appeals recognized the reasoning just explained, and therefore rejected by contrast a digitization effort that was touted as space-shifting but in fact resulted in the multiplication of copies shared with outsiders through a file-sharing service. 239 F.3d 1004, 1019 (9th Cir. 2001), *aff'g in this part* 114 F. Supp. 2d 896, 912-13, 915-16 (N.D. Cal. 2000) (Judge Marilyn Hall Patel) (citing *Sony Betamax* and *Texaco*).

Here, every purchased print copy was copied in order to save storage space and to enable searchability as a digital copy. The print original was destroyed. One replaced the other. And, there is no evidence that the new, digital copy was shown, shared, or sold outside the company. This use was even more clearly transformative than those in *Texaco, Google,* and *Sony Betamax* (where the number of copies went up by at least one), and, of course, more transformative than those uses rejected in *Napster* (where the number went up by "millions" of copies shared for free with others).

Yes, Anthropic is a commercial outfit. And, this order takes for granted that Anthropic in fact benefited from the print-to-digital format change—or it would not have gone to all the trouble. But the crux of the first fair use factor's concern for "commercial" use is in protecting the copyright owners and their entitlements to exploit their copyright as they see fit (or not). *See, e.g., Harper & Row, Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 562 (1985). That the accused is a commercial entity is indicative, not dispositive. That the accused stands to benefit is likewise indicative. But what matters most is whether the format change exploits anything the Copyright Act reserves to the copyright owner. Anthropic already had purchased permanent library copies (print ones). It did not create new copies to share or sell outside.

Yes, Authors also might have wished to charge Anthropic more for digital than for print copies. And, this order takes for granted that Authors could have succeeded if Anthropic had been barred from the format change. "But the Constitution's language in Clause 8 nowhere suggests that the copyright owner's limited exclusive right should include a right to divide markets or a concomitant right to charge different purchasers different prices for the same book, merely]say to increase or to maximize gain." *See Kirtsaeng v. John Wiley & Sons, Inc.*, 568 U.S. 519, 552 (2013). Nor does the Copyright Act itself. Section 106 sets out exclusive rights that fair uses under Section 107 abridge. Section 106(1) reserves to the copyright owner the right to make reproductions. But on our facts we face the unusual situation where one copy entirely replaced the another. And, Section 106(2) reserves to the copyright owner the right to make derivative works that add or subtract creative material—as occurs in a "translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, [or] condensation" of a book, 17 U.S.C. § 101 (definitions). For some "other modification" of a book to constitute a "derivative work," it must itself "represent an original work of authorship." *Ibid.* But on our facts the format was changed but no content was added or subtracted. *See Mirage Editions, Inc. v. Albuquerque A.R.T. Co.*, 856 F.2d 1341, 1342, 1343-44 (9th Cir. 1988) (yes where elements added to create new decorative ceramic). Section 106(3) further reserves to the copyright owner the right to distribute copies. But again, the replacement copy here was kept in the central library, not distributed. *Cf. Fox News Network, LLC v. TVEyes, Inc.*, 883 F.3d 169, 176-78 (2d Cir. 2018) (enabling searching for "information about the material" can be transformative use, even if some distribu-

tion results); *Lewis Galoob Toys, Inc. v. Nintendo of Am., Inc.*, 964 F.2d 965, 968, 971 (9th Cir. 1992) (using nifty converter to "merely enhance" audiovisual displays emitted from purchased videogame cartridge was fair use of those displays partly because no surplus copies of cartridge or displays were ever created).

As a result, Anthropic's format-change from print library copies to digital library copies was transformative under fair use factor one. Anthropic was entitled to retain a copy of these works in a print format. It retained them instead in a digital format, easing storage and searchability. And, the further copies made therefrom for purposes of training LLMs were themselves transformative for that further reason, as above.

To be clear, this print-to-digital conversion involved a different and narrower form of transformative use than the broader one advanced by Anthropic. Anthropic argues that the central library use was part and parcel of the LLM training use and therefore transformative. This order disagrees. However, this order holds that the mere conversion of a print book to a digital file to save space and enable searchability was transformative for that reason alone. Therefore, the digital copy should be treated just as if the purchased print copy had been placed in the central library.

In sum, the first fair use factor *favors* fair use for the digital library copies converted from purchased print library copies—but these do not excuse the pirated library copies.

(ii) The Pirated Library Copies

Before buying books for its central library, Anthropic downloaded over seven million pirated copies of books, paid nothing, and kept these pirated copies in its library even after deciding it would not use them to train its AI (at all or ever again). Authors argue Anthropic should have paid for these pirated library copies. This order agrees.

The basic problem here was well-stated by Anthropic at oral argument: "You can't just bless yourself by saying I have a research purpose and, therefore, go and take any textbook you want. That would destroy the academic publishing market if that were the case" Of course, the person who purchases the textbook owes no further accounting for keeping the copy. But the person who copies the textbook from a pirate site has infringed already, full stop. This order further rejects Anthropic's assumption that the use of the copies for a central library can be excused as fair use merely because some will eventually be used to train LLMs.

This order doubts that any accused infringer could ever meet its burden of explaining why downloading source copies from pirate sites *that it could have purchased or otherwise accessed lawfully* was itself reasonably necessary to any subsequent fair use. There is no decision holding or requiring that pirating a book that could have been bought at a bookstore was reasonably necessary to writing a book review, conducting research on facts in the book, or creating an LLM. Such piracy of otherwise available copies is inherently, irredeemably infringing even if the pirated copies are immediately used for the transformative use and immediately discarded.

But this order need not decide this case on that rule. Anthropic did not use these copies only for training its LLM. Indeed, it retained pirated copies even after deciding it would not use them or copies from them for training its LLMs ever again. They were acquired and retained, as a central library of all the books in the world.

Building a central library of works to be available for any number of further uses was itself the use for which Anthropic acquired these copies. *One further use* was making further copies for training LLMs. But not every book Anthropic pirated was used to train LLMs. And, every pirated library copy was retained even if it was determined it would not be so used. Pirating copies to build a research library without paying for it, and to retain copies should they prove useful for one thing or another, was its own use—and not a transformative one.

Anthropic's briefing contains other reasons why it believes its pirated library copies are irrelevant to our fair use analysis, notwithstanding its own statements at our oral argument.

First, Anthropic accepts in this posture that it acted in bad faith but argues that its bad faith in pirating copies cannot "somehow short-circuit" the fair use analysis. But its bad faith is not the basis for this decision. Each use of a work must be analyzed objectively. *Warhol,* 598 U.S. at 544-45. The objective analysis here shows the initial copies were pirated to create a central, general-purpose library, as a substitute for paid copies to do the same thing. (Of course, if infringement is found, bad faith would matter for determining willfulness. 17 U.S.C. § 504(c)(2).)

Second, Anthropic argues that its goal to put the copies eventually "to a highly transformative use" requires that each copy and use along the way be justified as having a transformative use, too. But now Anthropic seeks to take the shortcut Anthropic just said cannot be taken. Again, the Supreme Court tasks us with looking past the "subjective *intent* of the user" to the objective *use* made of each copy. *Warhol,* 598 U.S. at 544-45(emphasis added). Put another way, what a copyist *says or thinks or feels* matters only to the extent it shows what a copyist in fact *does* with the work. Indeed, the same copy can be used one way, then another, each with a different result. Here, what Anthropic said about its acquisitions at the time —that they were made to "build[ ] a research library" while avoiding a "huge legal/practice/business slog"—are relevant in this regard. And, Anthropic's actual use of these pirated copies was to create its central library of texts that, like any university or corporate library, stored the works' well-organized facts, analyses, and expressive examples for various contingent uses, one being training.

Third, Anthropic argues that *Texaco*—the case involving copies used in a central library, copies used in desk libraries, and copies used in the laboratory—is inapposite. Anthropic argues that the disputed copies in *Texaco* were never used in the laboratory but instead in personal desk libraries for a use "identical to the original purpose and use" of the central library copies, and so not for a transformative use. By contrast, says Anthropic, here it *did use* copies in the laboratory to train LLMs—a very transformative use. But this is a fast glide over thin ice. Like Texaco, Anthropic possessed copies it did not put into use in the laboratory and it kept those copies in a central library even after its transformative use had been completed. But, unlike Texaco, which bought those copies, Anthropic never paid for the central library copies stolen off the internet. *Texaco* also shows why Anthropic is wrong to suppose that so long as you create an exciting end product, every "back-end step, invisible to the public," is excused (Br. 10).

Notably, this is not a case where source copies were unavailable for separate purchase or loan. *See, e.g., NXIVM Corp. v. Ross Inst.,* 364 F.3d 471, 475-76, 478-79 (2d Cir. 2004) (using selections of training manual—otherwise available only to cult's trainees subject to NDAs—to expose cult in critical review); *Time Inc. v. Bernard Geis Assocs.,* 293 F. Supp. 130, 135-36, 138, 146 (S.D.N.Y. 1968) (Judge Inzer Bass Wyatt) (making charcoal drawings of photographs taken of originals

otherwise not on sale or loan out to illustrate a history book). Nor were the copies made only incidentally and necessarily from pirated copies. *See, e.g., Perfect 10,* 508 F.3d at 1164 n.8 (copies of images that had been pirated by third-party websites were used to index those same websites while indexing the entire web). Here, piracy was the point: To build a central library that one could have paid for, just as Anthropic later did, but without paying for it.

Nor were the initial copies made immediately transformed into a significantly altered form. In *Perfect 10,* images were copied by the search engine in thumbnail form only and deployed immediately into the transformative use of identifying the full-sized images and the pages from which they came. 508 F.3d at 1160, 1165, 1167. And, in *Kelly v. Arriba Software Corp.,* images were copied at full size and then into thumbnails for immediate use in building a search engine, after which the full-sized copies were immediately deleted. 336 F.3d 811, 815 (9th Cir. 2003). Not here. The *full-text* copies of books were downloaded and maintained "forever."

Nor does the initial copying here even resemble the full-text copying in the *Google Books* cases. There, libraries of authorized copies *already* had been assembled, and *all* copies therefrom were made for direct employment in a one-to-one further fair use—whether the transformative use of pointing to the works themselves, the use of providing the works in formats for print-disabled patrons, or the use of insuring against going out of print, getting lost, and becoming otherwise unavailable. *HathiTrust,* 755 F.3d at 97, 101, 103; *Google,* 804 F.3d at 206, 216-18, 228 (further distinguishing search and snippet uses, which "tested the boundaries of fair use"). Not so here concerning the pirated copies. No authorized copies existed from which Anthropic made its first copies. No full-text copy therefrom was put immediately into use training LLMs. Not every copy was even necessary nor used for training LLMs. No initial copy was ever deleted, even if never used or no longer used.[7] The university libraries and Google went to exceedingly great lengths to ensure that all copies were secured against unauthorized uses—both through technical measures and through legal agreements among all participants. Not so here. The library copies lacked internal controls limiting access and use.

Nor do the decisions on intermediate copying require anything less than the analysis applied here. Anthropic argues that our court of appeals in *Sega Enterprises Ltd. v. Accolade, Inc.* looked only at the "ultimate use" and "did not analyze a series of atomized acts of 'infringement' distinct from that overall purpose." To the contrary, the appeals court examined the initial, intermediate, and ultimate copies used by the copyist. *The court explained that the copyist initially purchased commercially available copies of game cartridges* and then made further copies necessarily and "solely in order to discover the functional requirements for compatibility." 977 F.2d 1510, 1522 (9th Cir. 1992). Thus, it reached only one result because on those facts there was only one "overall purpose" for the unauthorized copies. Indeed, the court reaffirmed prior caselaw holding that "intermediate copying of a work may infringe the exclusive rights granted to the copyright owner in Section

7 Training LLMs was not a use where perpetually maintaining a library copy was intrinsic to the proffered fair use (e.g., for a plagiarism-checker service). Nor is this an instance where retaining at least one copy was authorized by contract with the copyright owners (e.g., by agreement to express terms upon submission to a plagiarism-checker service, notwithstanding proposed terms scrawled on a paper prior to submission). *A.V. ex rel. Vanderhye v. iParadigms, LLC,* 562 F.3d 630, 635-36 & n.5, 645 n.8 (4th Cir. 2009), *aff'g in relevant parts* 544 F. Supp. 2d 473, 480 (E.D. Va. 2008) (Judge Claude Hilton). Anthropic mischaracterizes this case.

106 of the Copyright Act regardless of whether the end product of the copying also infringes those rights." *Id.* at 1518-19.

Similarly, in *Sony Computer Entertainment, Inc. v. Connectix Corp.,* our appeals court applied the same law to similarly focused conduct. Another copyist allegedly had purchased an authorized copy and then made further copies solely and necessarily to reverse-engineer compatibility requirements. 203 F.3d 596, 601, 602-03 (9th Cir. 2000).

Both *Sega* and *Sony* avoided imposing an "artificial hurdle" to fair use by generously construing the intermediate copying necessary to the fair use. As one example, *Sega* stated that an engineer should be permitted to reboot her computer while undertaking to reverse-engineer software loaded onto it—even if doing so creates another digital copy of the software and is not strictly necessary to reverse-engineering. *Id.* at 605. But neither *Sega* nor *Sony* fathomed gifting an "artificial head start" to a fair user, either, by treating even the initial copy as an intermediate one. …

Finally, Anthropic argues that even if the initial copies served a different use than the intermediate and ultimate copies, it was not a use for which Anthropic necessarily would have needed to pay Authors for a copy. In theory, argues Anthropic, it could have done as Google did in *Google Books*—find an existing reference library willing to loan its copies for free as source copies. Or, in theory, it could have done as Anthropic did later—go buy used copies without having to pay Authors at all. *See* 17 U.S.C. § 109(a). But Anthropic did not do those things—instead it stole the works for its central library by downloading them from pirated libraries.

In sum, the first factor *points against* fair use for the central library copies made from pirated sources—and no damages from pirating copies could be undone by later paying for copies of the same works.

2. The Nature of the Copyrighted Work

The second fair use factor is "the nature of the copyrighted work." 17 U.S.C. § 107(2). This factor "calls for recognition that some works are closer to the core of intended copyright protection than others, with the consequence that fair use is more difficult to establish when the former works are copied." *Campbell,* 510 U.S. at 586. For one thing, less protection is due published works than unpublished ones. For another, less protection is due "factual works than works of fiction or fantasy." *Harper & Row,* 471 U.S. at 563. But less protection is not no protection. Even the arrangement of otherwise unprotectable facts surpasses the low bar for a protectable original work of authorship.

Here, Anthropic accepts that all of Authors' books—all published, whether non-fiction or fiction—contained expressive elements. And, as set out above, this order accepts Authors' view of the evidence that their works were chosen for their expressive qualities in building a central library and then in training specific LLMs.

The main function of the second factor is to help assess the other factors: to reveal differences between the nature of the works at issue and the nature of their secondary use (above), and to reveal any relation between the amount and substantiality of each work taken and the secondary use (next). *E.g., Campbell,* 510 U.S. at 586, 114 S.Ct. 1164; *Kelly,* 336 F.3d at 820; *Google,* 804 F.3d at 220; *HathiTrust,* 755 F.3d at 98; *Bill Graham Archives v. Dorling Kindersley Ltd.,* 448 F.3d 605, 612-13 (2d Cir. 2006).

The second factor *points against* fair use for all copies alike.

3. The Amount and Substantiality of the Portion Used

The third fair use factor is "the amount and substantiality of the portion" of the copyrighted work used by the accused. 17 U.S.C. § 107(3). The crux of this factor is whether the amount was "reasonable in relation to the purpose of the copying." *Campbell,* 510 U.S. at 586. Thus, the amount of copying is considered first against the work itself, then more importantly against the proposed transformative purpose.

A. The Copies Used to Train Specific LLMs

Copies selected for inclusion in training sets were selected because they were complete and because they contained rich protectible expression, or so this order accepts the record shows for Authors. Was all this copying reasonably necessary to the transformative use?

Yes.

"What matters is not so much 'the amount and substantiality of the portion used' *in making a copy,* but rather the amount and substantiality of *what is thereby made accessible* to a public [in the purported secondary use] for which it may serve as a competing substitute [for the primary use]." *Google,* 804 F.3d at 222. Here, once again, there is no allegation of any traceable connection between the Claude service's outputs and Authors' works. The copying used to train the LLMs underlying Claude was thus especially reasonable.

In response, Authors object primarily that the copying used in training was both extremely extensive and not strictly necessary.

As to extensive copying, it is true that entire works were copied. And, "copying entire works militates against a finding of fair use." *Worldwide Church of God v. Philadelphia Church of God, Inc.,* 227 F.3d 1110, 1118 (9th Cir. 2000). But we just addressed why Authors' argument is misdirected. The copies that count for this factor are those that would merely serve the same use as the work's ordinary one. Authors do not allege such copying. The accused use here of the incremental copies is as orthogonal as can be imagined to the ordinary use of a book.

As to strict necessity, Authors make a stronger point. When a productive use is made possible only by borrowing from a specific work, fair use climbs towards its zenith. When a productive use is possible without that borrowing, fair use falls to its nadir—and the borrowing deserves a particularly compelling justification. *See Warhol,* 598 U.S. at 543 & n.18, 547. Here, it is true that Anthropic could have used some other books or no books at all for training its LLMs—or so this order accepts the record shows for Authors. But Anthropic has presented a compelling explanation for why it was reasonably necessary to use them anyway.

For one thing, all agree Anthropic needed billions of words to train any given LLM. If using only books, Anthropic would have needed millions of books per model. If using a set comprising only a small fraction of books and a larger fraction of other texts, Anthropic still would have needed hundreds of thousands of books. Authors contend that because Anthropic showed it could use such smaller sets of books, it surely could have used no books at all—or at least not *their* books. But Authors forget that "reasonably necessary" does not mean "strictly necessary." Authors do not contest that the volume of text required to train an LLM is monumental. Because using so many works was reasonably necessary, using any one work for *actually training* LLMs was about as reasonable as the next.

For another thing, no output to the public was even alleged to be infringing. So, yes, Authors' works were chosen as the strongest examples of writing. But the compelling benefits of training the LLMs on strong examples were not offset by

revelations to the public of any portion of the works themselves. What was copied was therefore especially reasonable and compelling.

The third factor thus *favors* fair use for the training copies.

B. The Copies Used to Build a Central Library

But again, there was a separate use—a distinction that makes some difference as to whether the amount and substantiality of the copying was "reasonable in relation to the purpose of the copying" for the library copies. *Campbell,* 510 U.S. at 586.

(i) The Purchased Library Copies Converted from Print to Digital

For the print library copies that Anthropic purchased and then converted into digital library copies, Anthropic already enjoyed entitlement to keep the copies in its library. The purpose of the copying was to keep them in its library but with more favorable storage and searchability properties. Copying the entire work was exactly what this purpose required. There was no surplus copying. The source copy was destroyed.

The third fair use factor *favors* fair use for the purchased library copies converted from print to digital.

(ii) The Pirated Library Copies

For the pirated library copies, however, Anthropic lacked any entitlement to hold copies of the books at all. Its purpose, it says, was to train LLMs. But its objective conduct was to seek "all the books in the world" and then retain them even after deciding it would not make further copies from them for training—indicating there were other further uses. Against the purpose of acquiring all the books one could on the chance some might prove useful for training LLMs and maybe other stuff too, almost any unauthorized copying would have been too much. Anthropic copied millions of books *in toto,* Authors' works among them.

The third factor *points against* fair use for the pirated library copies.

4. The Effect of the Use upon the Market for or Value of the Copyrighted Work

The final factor is "the effect of the use upon the potential market for or value of the copyrighted work." 17 U.S.C. § 107(4). This factor points against fair use when a copyist makes copies available that displace demand for copies the copyright owner already makes available or readily could. *Texaco,* 60 F.3d at 926-28 (reproduced copies); *Dr. Seuss Enters., L.P. v. ComicMix LLC,* 983 F.3d 443, 461 (9th Cir. 2020) (derivative copies). "While the first factor considers whether and to what extent an original work and secondary use [*in principle* could] have substitutable purposes, the fourth factor focuses on *actual or potential* market substitution." *Warhol,* 598 U.S. at 536 n.12, 143 S.Ct. 1258 (emphasis added).

A. The Copies Used to Train Specific LLMs

The copies used to train specific LLMs did not and will not displace demand for copies of Authors' works, or not in the way that counts under the Copyright Act.

Again, Authors concede that training LLMs did not result in any exact copies nor even infringing knockoffs of their works being provided to the public. If that were not so, this would be a different case. Authors remain free to bring that case in the future should such facts develop.

Instead, Authors contend generically that training LLMs will result in an explosion of works competing with their works—such as by creating alternative summaries of factual events, alternative examples of compelling writing about fictional events, and so on. This order assumes that is so. But Authors' complaint is

no different than it would be if they complained that training schoolchildren to write well would result in an explosion of competing works. This is not the kind of competitive or creative displacement that concerns the Copyright Act. The Act seeks to advance original works of authorship, not to protect authors against competition.

Authors next contend that training LLMs displaced (or will) an emerging market for licensing their works for the narrow purpose of training LLMs. Anthropic argues that transactional costs would exceed Anthropic's expected benefit from any such bargain, prompting it to cease dealing with any rightsholders or else to cease developing such technology altogether. Our record could support either account—so this order must assume Authors are correct. A market could develop. Even so, such a market for that use is not one the Copyright Act entitles Authors to exploit. ...

The fourth factor thus *favors* fair use for the training copies.

B. The Copies Used to Build a Central Library

(i) The Purchased Library Copies Converted from Print to Digital

For these copies, this order assumes Anthropic's format change from print to digital displaced purchases of new digital copies that Anthropic would have made directly from Authors (had it not been able to purchase print copies in used condition). But for reasons stated under the first factor, such losses did not relate to something the Copyright Act reserves for Authors to exploit. It was a format change.

Authors' next argument, it seems, is that the format change nonetheless exposed it to usurpation of the opportunity to sell rightful copies because Anthropic *might* transmit additional unauthorized digital copies more readily than it could have transmitted additional unauthorized print copies—and that the same would be true for all format converters. But after much discovery, there is no inkling in our record of intent to redistribute library copies once acquired nor of inability to secure that valuable library against outside actors. And, if the internal, central library copies did or do in fact lead to further reproduction or distribution, those further copies remain redressable separately by Authors. The format change did not itself usurp the Authors' rightful entitlements.

This factor is thus *neutral* for the purchased library copies converted from print to digital.

(ii) The Pirated Library Copies.

The copies used to build a central library *and* that were obtained from pirated sources plainly displaced demand for Authors' books—copy for copy. Not every person who merely intends to make a fair use of a work is thereby entitled to a full copy in the meantime, nor even to steal a copy so that achieving this fair use is especially simple or cost-effective. Here, the copies employed in training LLMs were one thing, but the copies acquired to assemble a convenient, general-purpose library of works for various uses for which the company might have of them, if any, was a different use altogether.

Anthropic has almost no rebuttal on these points. *First,* Anthropic argues that "Claude's services do not reduce or usurp the value of Plaintiffs' works through substitution in their traditional markets." But stealing pirated copies of Authors' works plainly did. *Second,* Anthropic argues that it may have been able to purchase some books on the open market (and some other texts), but not other texts it copied. But this case does not concern those other texts it could not have pur-

chased. It could have purchased Authors' books (and many others). In fact it later did. *Finally,* Anthropic argues that the effect on these texts from one book foregone was too small to be considered. But the test requires that we contemplate the likely result were the conduct to be condoned as a fair use—namely to steal a work you could otherwise buy (a book, millions of books) so long as you at least loosely intend to make further copies for a purportedly transformative use (writing a book review with excerpts, training LLMs, etc.), without any accountability. As Anthropic itself suggested, "That would destroy the [entire] publishing market if that were the case."

The fourth factor *points against* fair use for the pirated library copies.

5. Overall Analysis

After the four factors and any others deemed relevant are "explored, the results are weighed together, in light of the purposes of copyright." *Campbell,* 510 U.S. at 578.

The copies used to train specific LLMs were justified as a fair use. Every factor but the nature of the copyrighted work favors this result. The technology at issue was among the most transformative many of us will see in our lifetimes.

The copies used to convert purchased print library copies into digital library copies were justified, too, though for a different fair use. The first factor strongly favors this result, and the third favors it, too. The fourth is neutral. Only the second slightly disfavors it. On balance, as the purchased print copy was destroyed and its digital replacement not redistributed, this was a fair use.

The downloaded pirated copies used to build a central library were not justified by a fair use. Every factor points against fair use. Anthropic employees said copies of works (pirated ones, too) would be retained "forever" for "general purpose" even after Anthropic determined they would never be used for training LLMs. A separate justification was required for each use. None is even offered here except for Anthropic's pocketbook and convenience.

And, as for any copies made from central library copies but not used for training, this order does not grant summary judgment for Anthropic. On this record in this posture, the central library copies were retained even when no longer serving as sources for training copies, "hundreds of engineers" could access them to make copies for other uses, and engineers did make other copies. ... We cannot determine the right answer concerning such copies because the record is too poorly developed as to them. Anthropic is not entitled to an order blessing all copying "that Anthropic has ever made after obtaining the data," to use its words.

Conclusion

With respect to the training copies and the print-to-digital converted copies, this order has drawn all ambiguities and inferences in favor of the opposing side, namely Authors. With respect to the pirated copies, this order has also accepted the Authors' version of the facts. Authors did not move for summary judgment but if they had, then we would have been obligated to accept all reasonable views given the evidence in defendant's favor instead.

This order grants summary judgment for Anthropic that the training use was a fair use. And, it grants that the print-to-digital format change was a fair use for a different reason. But it denies summary judgment for Anthropic that the pirated library copies must be treated as training copies.

We will have a trial on the pirated copies used to create Anthropic's central library and the resulting damages, actual or statutory (including for willfulness). That Anthropic later bought a copy of a book it earlier stole off the internet will

not absolve it of liability for the theft but it may affect the extent of statutory damages. Nothing is foreclosed as to any other copies flowing from library copies for uses other than for training LLMs.

NOTES AND QUESTIONS

1. One way to analyze AI training and deployment is in terms of a "supply chain" from training data to models to systems to outputs. *See generally* Katherine Lee, A. Feder Cooper, and James Grimmelmann, *Talkin' 'Bout AI Generation: Copyright and the Generative-AI Supply Chain*, 72 J. COPYRIGHT SOC'Y U.S.A. 251 (2025). On this view, one could ask whether each of the following is a fair use: (1) copying works to assemble a training dataset, (2) training an AI model, and (3) generating an output using an AI model. Which of these does this opinion discuss?
2. The plaintiffs here alleged, and the court accepted as true for purposes of the decision, that an an AI model stores "compressed" or "memorized" versions of specific works of training data in sufficient detail that it can generate near-verbatim versions of those works as outputs. Do you think this is true? If it is, does it matter?
3. Do you accept the court's distinction between books that Anthropic scanned and destroyed, and books that it downloaded from LibGen and similar "shadow libraries?" How would you apply this distinction to data scraped from the web?
4. Consider the argument that widespread use of LLMs could result in "market dilution" or "indirect substitution":

   > The third way that using copyrighted books to train an LLM might harm the market for those works is by helping to enable the rapid generation of countless works that compete with the originals, even if those works aren't themselves infringing. Assume for this discussion that people can (or will soon be able to) use LLMs to generate massive amounts of text in significantly less time than it would take to write that text, and using a fraction of the creativity. People could thus use LLMs to create books and then sell them, competing with books written by human authors for sales and attention.

   Kadrey v. Meta Platforms, Inc. 788 F. Supp. 3d 1026, 1052 (2025). Is this plausible? If so, what does *Bartz* say about how it would affect the fair use analysis? Do you agree?
5. Suppose that Claude is capable of generating both infringing and non-infringing outputs. Should a court order Anthropic to prevent the generation of infringing ones? How could Anthropic implement such an order? How could a court enforce it?

C. First Sale

VERNOR V. AUTODESK, INC.
621 F.3d 1102 (9th Cir. 2010)

Callahan, Circuit Judge: ...

A. Autodesk's Release 14 software and licensing practices ...

Autodesk makes computer-aided design software used by architects, engineers, and manufacturers. It has more than nine million customers. It first released its AutoCAD software in 1982. It holds registered copyrights in all versions of the software including the discontinued Release 14 version, which is at issue in this case. It provided Release 14 to customers on CD-ROMs.

Since at least 1986, Autodesk has offered AutoCAD to customers pursuant to an accompanying software license agreement ("SLA"), which customers must accept before installing the software. A customer who does not accept the SLA can return the software for a full refund. Autodesk offers SLAs with different terms for commercial, educational institution, and student users. The commercial license, which is the most expensive, imposes the fewest restrictions on users and allows them software upgrades at discounted prices.

The SLA for Release 14 first recites that Autodesk retains title to all copies. Second, it states that the customer has a nonexclusive and nontransferable license to use Release 14. Third, it imposes transfer restrictions, prohibiting customers from renting, leasing, or transferring the software without Autodesk's prior consent and from electronically or physically transferring the software out of the Western Hemisphere. Fourth, it imposes significant use restrictions:

> YOU MAY NOT: (1) modify, translate, reverse-engineer, decompile, or disassemble the Software . . . (3) remove any proprietary notices, labels, or marks from the Software or Documentation; (4) use . . . the Software outside of the Western Hemisphere; (5) utilize any computer software or hardware designed to defeat any hardware copy-protection device, should the software you have licensed be equipped with such protection; or (6) use the Software for commercial or other revenue-generating purposes if the Software has been licensed or labeled for educational use only.

Fifth, the SLA provides for license termination if the user copies the software without authorization or does not comply with the SLA's restrictions. Finally, the SLA provides that if the software is an upgrade of a previous version:

> [Y]ou must destroy the software previously licensed to you, including any copies resident on your hard disk drive . . . within sixty (60) days of the purchase of the license to use the upgrade or update. . . . Autodesk reserves the right to require you to show satisfactory proof that previous copies of the software have been destroyed.

Autodesk takes measures to enforce these license requirements. It assigns a serial number to each copy of AutoCAD and tracks registered licensees. It requires customers to input "activation codes" within one month after installation to continue

using the software.1 The customer obtains the code by providing the product's serial number to Autodesk. Autodesk issues the activation code after confirming that the serial number is authentic, the copy is not registered to a different customer, and the product has not been upgraded. Once a customer has an activation code, he or she may use it to activate the software on additional computers without notifying Autodesk. ...

C. Vernor's eBay business and sales of Release 14

Vernor has sold more than 10,000 items on eBay. In May 2005, he purchased an authentic used copy of Release 14 at a garage sale from an unspecified seller. He never agreed to the SLA's terms, opened a sealed software packet, or installed the Release 14 software. Though he was aware of the SLA's existence, he believed that he was not bound by its terms. He posted the software copy for sale on eBay.

Autodesk filed a Digital Millennium Copyright Act ("DMCA") take-down notice with eBay claiming that Vernor's sale infringed its copyright, and eBay terminated Vernor's auction. [The DMCA notice-and-takedown procedure is discussed *infra*.]

[Vernor then purchased four copies of Release 14 from Cardwell/Thomas & Associates ("CTA"). As the court explained, in an earlier part of the opinion, CTA had "upgraded to the newer, fifteenth version of the AutoCAD program, AutoCAD 2000. It paid $495 per upgrade license, compared to $3,750 for each new license. The SLA for AutoCAD 2000, like the SLA for Release 14, required destruction of copies of previous versions of the software, with proof to be furnished to Autodesk on request. However, rather than destroying its Release 14 copies, CTA sold them to Vernor at an office sale with the handwritten activation codes necessary to use the software."]

[Autodesk submitted more DMCA notices, and Vernor filed a declaratory judgment action against Autodesk, claiming that his resales were protected by first sale.]

III. ...

The exclusive distribution right is limited by the first sale doctrine, an affirmative defense to copyright infringement that allows owners of copies of copyrighted works to resell those copies. ...

This case requires us to decide whether Autodesk sold Release 14 copies to its customers or licensed the copies to its customers. If CTA owned its copies of Release 14, then both its sales to Vernor and Vernor's subsequent sales were non-infringing under the first sale doctrine.[6] However, if Autodesk only licensed CTA to use copies of Release 14, then CTA's and Vernor's sales of those copies are not protected by the first sale doctrine and would therefore infringe Autodesk's exclusive distribution right.

---

1 Prior to using activation codes, Autodesk required users to return one disc of an earlier version of the software to upgrade to a later version. Autodesk has abandoned this return policy, deeming it slow and unworkable.

6 If Autodesk's transfer of Release 14 copies to CTA was a first sale, then CTA's resale of the software in violation of the SLA's terms would be a breach of contract, but would not result in copyright liability. See *United States v. Wise*, 550 F.2d 1180, 1187 (9th Cir. 1977) ("[T]he exclusive right to vend the transferred copy rests with the vendee, who is not restricted by statute from further transfers of that copy, even though in breach of an agreement restricting its sale.").

A. The first sale doctrine

The Supreme Court articulated the first sale doctrine in 1908, holding that a copyright owner's exclusive distribution right is exhausted after the owner's first sale of a particular copy of the copyrighted work. *See Bobbs-Merrill Co. v. Straus*, 210 U.S. 339, 350–51 (1908). In *Bobbs-Merrill*, the plaintiff-copyright owner sold its book with a printed notice announcing that any retailer who sold the book for less than one dollar was responsible for copyright infringement. Plaintiff sought injunctive relief against defendants-booksellers who failed to comply with the price restriction. The Supreme Court rejected the plaintiff's claim, holding that its exclusive distribution right applied only to first sales of copies of the work. The distribution right did not permit plaintiff to dictate that subsequent sales of the work below a particular price were infringing. The Court noted that its decision solely applied to the rights of a copyright owner that distributed its work without a license agreement. *Id.* at 350 ("There is no claim in this case of contract limitation, nor license agreement controlling the subsequent sales of the book.").

Congress codified the first sale doctrine the following year. In its current form, it allows the "owner of a particular copy" of a copyrighted work to sell or dispose of his copy without the copyright owner's authorization. 17 U.S.C. § 109(a). The first sale doctrine does not apply to a person who possesses a copy of the copyrighted work without owning it, such as a licensee. *See id.* § 109(d); *cf. Quality King Distribs., Inc. v. L'Anza Research Int'l Inc.*, 523 U.S. 135, 146–47 (1998) ("[T]he first sale doctrine would not provide a defense to … any non-owner such as a bailee, a licensee, a consignee, or one whose possession of the copy was unlawful.").
…

B. Owners vs. licensees

We turn to our precedents governing whether a transferee of a copy of a copyrighted work is an owner or licensee of that copy. We then apply those precedents to CTA's and Vernor's possession of Release 14 copies.

[In *United States v. Wise*, 550 F.2d 1180 (9th Cir. 1977)], a criminal copyright infringement case, we considered whether copyright owners who transferred copies of their motion pictures pursuant to written distribution agreements had executed first sales. The defendant was found guilty of copyright infringement based on his for-profit sales of motion picture prints. The copyright owners distributed their films to third parties pursuant to written agreements that restricted their use and transfer. On appeal, the defendant argued that the government failed to prove the absence of a first sale for each film. If the copyright owners' initial transfers of the films were first sales, then the defendant's resales were protected by the first sale doctrine and thus were not copyright infringement.

To determine whether a first sale occurred, we considered multiple factors pertaining to each film distribution agreement. Specifically, we considered whether the agreement (a) was labeled a license, (b) provided that the copyright owner retained title to the prints, (c) required the return or destruction of the prints, (d) forbade duplication of prints, or (e) required the transferee to maintain possession of the prints for the agreement's duration. Our use of these several considerations, none dispositive, may be seen in our treatment of each film print.

For example, we reversed the defendant's conviction with respect to *Camelot*. It was unclear whether the Camelot print sold by the defendant had been subject to a first sale. Copyright owner Warner Brothers distributed Camelot prints pursuant to multiple agreements, and the government did not prove the absence of a first sale with respect to each agreement. We noted that, in one agreement, Warner

Brothers had retained title to the prints, required possessor National Broadcasting Company ("NBC") to return the prints if the parties could select a mutual agreeable price, and if not, required NBC's certification that the prints were destroyed. We held that these factors created a license rather than a first sale.

We further noted, however, that Warner Brothers had also furnished another Camelot print to actress Vanessa Redgrave. The print was provided to Redgrave at cost, and her use of the print was subject to several restrictions. She had to retain possession of the print and was not allowed to sell, license, reproduce, or publicly exhibit the print. She had no obligation to return the print to Warner Brothers. We concluded, "While the provision for payment for the cost of the film, standing alone, does not establish a sale, when taken with the rest of the language of the agreement, it reveals a transaction strongly resembling a sale with restrictions on the use of the print." *Id.* There was no evidence of the print's whereabouts, and we held that "[i]n the absence of such proof," the government failed to prove the absence of a first sale with respect to this Redgrave print. Since it was unclear which copy the defendant had obtained and resold, his conviction for sale of *Camelot* had to be reversed. …

IV. …

B. Analysis

We hold today that a software user is a licensee rather than an owner of a copy where the copyright owner (1) specifies that the user is granted a license; (2) significantly restricts the user's ability to transfer the software; and (3) imposes notable use restrictions. Applying our holding to Autodesk's SLA, we conclude that CTA was a licensee rather than an owner of copies of Release 14 and thus was not entitled to invoke the first sale doctrine.

Autodesk retained title to the software and imposed significant transfer restrictions: it stated that the license is nontransferable, the software could not be transferred or leased without Autodesk's written consent, and the software could not be transferred outside the Western Hemisphere. The SLA also imposed use restrictions against the use of the software outside the Western Hemisphere and against modifying, translating, or reverse-engineering the software, removing any proprietary marks from the software or documentation, or defeating any copy protection device. Furthermore, the SLA provided for termination of the license upon the licensee's unauthorized copying or failure to comply with other license restrictions. Thus, because Autodesk reserved title to Release 14 copies and imposed significant transfer and use restrictions, we conclude that its customers are licensees of their copies of Release 14 rather than owners.

CTA was a licensee rather than an "owner of a particular copy" of Release 14, and it was not entitled to resell its Release 14 copies to Vernor under the first sale doctrine. 17 U.S.C. § 109(a). Therefore, Vernor did not receive title to the copies from CTA and accordingly could not pass ownership on to others. Both CTA's and Vernor's sales infringed Autodesk's exclusive right to distribute copies of its work. *Id.*

Because Vernor was not an owner, his customers are also not owners of Release 14 copies. Therefore, when they install Release 14 on their computers, the copies of the software that they make during installation infringe Autodesk's exclusive reproduction right …

3. The Supreme Court's holding in Bobbs-Merrill

Vernor contends that *Bobbs-Merrill* establishes his entitlement to a first sale defense. However, *Bobbs-Merrill* stands only for the proposition that a copyright owner's exclusive distribution right does not allow it to control sales of copies of its work after the first sale. Decided in 1908, *Bobbs-Merrill* did not and could not address the question of whether the right to use software is distinct from the ownership of copies of software. Moreover, the Supreme Court in *Bobbs-Merrill* made explicit that its decision did not address the use of restrictions to create a license. *Id.* ("There is no claim in this case of contract limitation, nor license agreement controlling the subsequent sales of the book.")

4. Economic realities of the transaction

Finally, Vernor contends that "economic realities" demonstrate that Autodesk makes "first sales" to its customers, because Autodesk allows its customers to possess their copies of the software indefinitely and does not require recurring license payments. We held supra that neither of these factors is dispositive. Vernor cites no first sale doctrine case in support of this proposition. ...

UMG Recordings Inc. v. Augusto

628 F.3d 1175 (9th Cir. 2011)

[UMG mailed out promotional CDs to music critics and DJs, some of whom sold or gave away the CDs. The court held that UMG had engaged in a first sale.]

Our conclusion that the recipients acquired ownership of the CDs is based largely on the nature of UMG's distribution. First, the promotional CDs are dispatched to the recipients without any prior arrangement as to those particular copies. The CDs are not numbered, and no attempt is made to keep track of where particular copies are or what use is made of them. ...

We also hold that, because the CDs were unordered merchandise, the recipients were free to dispose of them as they saw fit under the Unordered Merchandise Statute, 39 U.S.C. § 3009, which provides in pertinent part that,

(a) [e]xcept for free samples clearly and conspicuously marked as such, the mailing of unordered merchandise constitutes an unfair method of competition and an unfair trade practice.

(b) Any merchandise mailed in violation of subsection (a) of this section may be treated as a gift by the recipient, who shall have the right to retain, use, discard, or dispose of it in any manner he sees fit without any obligation whatsoever to the sender.

There are additional reasons for concluding that UMG's distribution of the CDs did not involve a consensual licensing operation. Some of the statements on the CDs and UMG's purported method of securing agreement to licenses militate against a conclusion that any licenses were created. The sparest promotional statement, "Promotional Use Only—Not for Sale," does not even purport to create a license. But even the more detailed statement is flawed in the manner in which it purports to secure agreement from the recipient. The more detailed statement provides:

> This CD is the property of the record company and is licensed to the intended recipient for personal use only. Acceptance of this CD shall constitute an agreement to comply with the terms of the license. Resale or transfer of possession is not allowed and may be punishable under federal and state laws.

It is one thing to say … that "acceptance" of the CD constitutes an agreement to a license and its restrictions, but it is quite another to maintain that "acceptance" may be assumed when the recipient makes no response at all. …

Because the record here is devoid of any indication that the recipients agreed to a license, there is no evidence to support a conclusion that licenses were established under the terms of the promotional statement. Accordingly, we conclude that UMG's transfer of possession to the recipients, without meaningful control or even knowledge of the status of the CDs after shipment, accomplished a transfer of title.

CAPITOL RECORDS, LLC V. REDIGI, INC.

934 F. Supp. 2d 640 (S.D.N.Y. 2013),
aff'd, 910 F.3d 649 (2d Cir. 2018)

Sullivan, District Judge:

Capitol Records, LLC, the recording label for such classic vinyls as Frank Sinatra's "Come Fly With Me" and The Beatles' "Yellow Submarine," brings this action against ReDigi Inc., a twenty-first century technology company that touts itself as a "virtual" marketplace for "pre-owned" digital music. What has ensued in a fundamental clash over culture, policy, and copyright law, with Capitol alleging that ReDigi's web-based service amounts to copyright infringement in violation of the Copyright Act. …

I. Background

A. Facts

ReDigi markets itself as "the world's first and only online marketplace for digital used music." Launched on October 13, 2011, ReDigi's website invites users to "sell their legally acquired digital music files, and buy used digital music from others at a fraction of the price currently available on iTunes." Thus, much like used record stores, ReDigi permits its users to recoup value on their unwanted music. Unlike used record stores, however, ReDigi's sales take place entirely in the digital domain.

To sell music on ReDigi's website, a user must first download ReDigi's "Media Manager" to his computer. Once installed, Media Manager analyzes the user's computer to build a list of digital music files eligible for sale. A file is eligible only if it was purchased on iTunes or from another ReDigi user; music downloaded from a CD or other file-sharing website is ineligible for sale. After this validation process, Media Manager continually runs on the user's computer and attached devices to ensure that the user has not retained music that has been sold or uploaded for sale. However, Media Manager cannot detect copies stored in other locations. If a copy is detected, Media Manager prompts the user to delete the file. The file is not deleted automatically or involuntarily, though ReDigi's policy is to suspend the accounts of users who refuse to comply.

After the list is built, a user may upload any of his eligible files to ReDigi's "Cloud Locker," an ethereal moniker for what is, in fact, merely a remote server in Arizona. ReDigi's upload process is a source of contention between the parties. ReDigi asserts that the process involves "migrating" a user's file, packet by packet —"analogous to a train"—from the user's computer to the Cloud Locker so that data does not exist in two places at any one time.

Capitol asserts that, semantics aside, ReDigi's upload process "necessarily involves copying" a file from the user's computer to the Cloud Locker. Regardless, at the end of the process, the digital music file is located in the Cloud Locker and not

on the user's computer. Moreover, Media Manager deletes any additional copies of the file on the user's computer and connected devices.

Once uploaded, a digital music file undergoes a second analysis to verify eligibility. If ReDigi determines that the file has not been tampered with or offered for sale by another user, the file is stored in the Cloud Locker, and the user is given the option of simply storing and streaming the file for personal use or offering it for sale in ReDigi's marketplace. If a user chooses to sell his digital music file, his access to the file is terminated and transferred to the new owner at the time of purchase. Thereafter, the new owner can store the file in the Cloud Locker, stream it, sell it, or download it to her computer and other devices. No money changes hands in these transactions. Instead, users buy music with credits they either purchased from ReDigi or acquired from other sales. ReDigi credits, once acquired, cannot be exchanged for money. Instead, they can only be used to purchase additional music. …

Finally, ReDigi earns a fee for every transaction. ReDigi's website prices digital music files at fifty-nine to seventy-nine cents each. When users purchase a file, with credits, 20% of the sale price is allocated to the seller, 20% goes to an "escrow" fund for the artist, and 60% is retained by ReDigi. …

III. Discussion

Section 106 of the Copyright Act grants "the owner of copyright under this title" certain "exclusive rights," including the right "to reproduce the copyrighted work in copies or phonorecords," "to distribute copies or phonorecords of the copyrighted work to the public by sale or other transfer of ownership," and to publicly perform and display certain copyrighted works. …

A. Infringement of Capitol's Copyrights

To state a claim for copyright infringement, a plaintiff must establish that it owns a valid copyright in the work at issue and that the defendant violated one of the exclusive rights the plaintiff holds in the work. It is undisputed that Capitol owns copyrights in a number of the recordings sold on ReDigi's website. It is also undisputed that Capitol did not approve the reproduction or distribution of its copyrighted recordings on ReDigi's website. Thus, if digital music files are "reproduce[d]" and "distribute[d]" on ReDigi's website within the meaning of the Copyright Act, Capitol's copyrights have been infringed.

1. Reproduction Rights

Courts have consistently held that the unauthorized duplication of digital music files over the Internet infringes a copyright owner's exclusive right to reproduce. However, courts have not previously addressed whether the unauthorized transfer of a digital music file over the Internet—where only one file exists before and after the transfer—constitutes reproduction within the meaning of the Copyright Act. The Court holds that it does.

The Copyright Act provides that a copyright owner has the exclusive right "to reproduce the copyrighted work in … phonorecords." 17 U.S.C. § 106(1). Copyrighted works are defined to include, *inter alia,* "sound recordings," which are "works that result from the fixation of a series of musical, spoken, or other sounds." *Id.* § 101. Such works are distinguished from their material embodiments. These include phonorecords, which are the "*material objects* in which sounds … are fixed by any method now known or later developed, and from which the sounds can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device." *Id.* § 101 (emphasis added). Thus, the plain text of the

Copyright Act makes clear that reproduction occurs when a copyrighted work is fixed in a new *material object.*

Courts that have dealt with infringement on peer-to-peer ("P2P") file-sharing systems provide valuable guidance on the application of this right in the digital domain. For instance, in *London-Sire Records, Inc. v. John Doe 1,* the court addressed whether users of P2P software violated copyright owners' distribution rights. 542 F. Supp. 2d 153, 166 & n.16 (D. Mass. 2008). Citing the "material object" requirement, the court expressly differentiated between the copyrighted work —or digital music file—and the phonorecord—or "appropriate segment of the hard disk" that the file would be embodied in following its transfer. Specifically,

> [w]hen a user on a [P2P] network downloads a song from another user, he receives into his computer a digital sequence representing the sound recording. That sequence is magnetically encoded on a segment of his hard disk (or likewise written on other media). With the right hardware and software, the downloader can use the magnetic sequence to *reproduce* the sound recording. The electronic file (or, perhaps more accurately, the appropriate segment of the hard disk) is therefore a "phonorecord" within the meaning of the statute.

Id. (emphasis added). Accordingly, when a user downloads a digital music file or "digital sequence" to his "hard disk," the file is "reproduce[d]" on a new phonorecord within the meaning of the Copyright Act.

This understanding is, of course, confirmed by the laws of physics. It is simply impossible that the same "material object" can be transferred over the Internet. Thus, logically, the court in *London-Sire* noted that the Internet transfer of a file results in a material object being "created elsewhere at its finish." *Id.* at 173. Because the reproduction right is necessarily implicated when a copyrighted work is embodied in a new material object, and because digital music files must be embodied in a new material object following their transfer over the Internet, the Court determines that the embodiment of a digital music file on a new hard disk is a reproduction within the meaning of the Copyright Act.

This finding holds regardless of whether one or multiple copies of the file exist. *London-Sire,* like all of the P2P cases, obviously concerned multiple copies of one digital music file. But that distinction is immaterial under the plain language of the Copyright Act. Simply put, it is the creation of a *new* material object and not an *additional* material object that defines the reproduction right. ... Thus, the right "to reproduce the copyrighted work in ... phonorecords" is implicated whenever a sound recording is fixed in a new material object, regardless of whether the sound recording remains fixed in the original material object.

Given this finding, the Court concludes that ReDigi's service infringes Capitol's reproduction rights under any description of the technology. ReDigi stresses that it "migrates" a file from a user's computer to its Cloud Locker, so that the same file is transferred to the ReDigi server and no copying occurs. However, even if that were the case, the fact that a file has moved from one material object—the user's computer—to another—the ReDigi server—means that a reproduction has occurred. Similarly, when a ReDigi user downloads a new purchase from the ReDigi website to her computer, yet another reproduction is created. It is beside the point that the original phonorecord no longer exists. It matters only that a new phonorecord has been created.

ReDigi struggles to avoid this conclusion by pointing to *C.M. Paula Co. v. Logan,* a 1973 case from the Northern District of Texas where the defendant used

chemicals to lift images off of greeting cards and place them on plaques for resale. 355 F. Supp. 189, 190 (N.D. Tex. 1973). The court determined that infringement did not occur because "should defendant desire to make one hundred ceramic plaques … , defendant would be required to purchase one hundred separate … prints." *C.M. Paula,* 355 F. Supp. at 191. ReDigi argues that, like the defendant in *C.M. Paula,* its users must purchase a song on iTunes in order to sell a song on ReDigi. Therefore, no "duplication" occurs. ReDigi's argument is unavailing. Ignoring the questionable merits of the court's holding in *C.M. Paula,* ReDigi's service is distinguishable from the process in that case. There, the copyrighted print, or material object, was lifted from the greeting card and transferred in toto to the ceramic tile; no new material object was created. By contrast, ReDigi's service by necessity creates a new material object when a digital music file is either uploaded to or downloaded from the Cloud Locker.

ReDigi also argues that the Court's conclusion would lead to "irrational" outcomes, as it would render illegal any movement of copyrighted files on a hard drive, including relocating files between directories and defragmenting. However, this argument is nothing more than a red herring. As Capitol has conceded, such reproduction is almost certainly protected under other doctrines or defenses, and is not relevant to the instant motion. …

2. Distribution Rights

In addition to the reproduction right, a copyright owner also has the exclusive right "to distribute copies or phonorecords of the copyrighted work to the public by sale or other transfer of ownership." 17 U.S.C. § 106(3). Like the court in *London-Sire,* the Court agrees that "[a]n electronic file transfer is plainly within the sort of transaction that § 106(3) was intended to reach [and] … fit[s] within the definition of 'distribution' of a phonorecord." *London-Sire,* 542 F. Supp. 2d at 173–74. For that reason, "courts have not hesitated to find copyright infringement by distribution in cases of file-sharing or electronic transmission of copyrighted works." *Arista Records LLC v. Greubel,* 453 F. Supp. 2d 961, 968 (N.D. Tex. 2006) (collecting cases). …

There is no dispute that sales occurred on ReDigi's website. Capitol has established that it was able to buy more than one-hundred of its own recordings on ReDigi's webite, and ReDigi itself compiled a list of its completed sales of Capitol's recordings. ReDigi, in fact, does not contest that distribution occurs on its website—it only asserts that the distribution is protected by the fair use and first sale defenses.

Accordingly, the Court concludes that, absent the existence of an affirmative defense, the sale of digital music files on ReDigi's website infringes Capitol's exclusive right of distribution.[6] …

---

6 Capitol argues that ReDigi also violated its distribution rights simply by making Capitol's recordings available for sale to the public, regardless of whether a sale occurred. However, a number of courts, including one in this district, have cast significant doubt on this "make available" theory of distribution. *See, e.g., Elektra Entm't Grp., Inc. v. Barker*, 551 F. Supp. 2d 234, 243 (S.D.N.Y. 2008) ("[T]he support in the case law for the "make available" theory of liability is quite limited."); *London-Sire,* 542 F. Supp. 2d at 169 ("[T]he defendants cannot be liable for violating the plaintiffs' distribution right unless a 'distribution' actually occurred."). In any event, because the Court concludes that actual sales on ReDigi's website infringed Capitol's distribution right, it does not reach this additional theory of liability.

B. Affirmative Defenses

Having concluded that sales on ReDigi's website infringe Capitol's exclusive rights of reproduction and distribution, the Court turns to whether the fair use or first sale defenses excuse that infringement. For the reasons set forth below, the Court determines that they do not.

2. First Sale

The first sale defense, a common law principle recognized in *Bobbs-Merrill Co. v. Straus*, 210 U.S. 339, 350 (1908) and now codified at Section 109(a) of the Copyright Act, provides that:

> Notwithstanding the provisions of section 106(3), the owner of a particular copy or phonorecord lawfully made under this title, or any person authorized by such owner, is entitled, without the authority of the copyright owner, to sell or otherwise dispose of the possession of that copy or phonorecord.

17 U.S.C. § 109. Under the first sale defense, "once the copyright owner places a copyrighted item [here, a phonorecord] in the stream of commerce by selling it, he has exhausted his exclusive statutory right to control its distribution." *Quality King Distribs., Inc. v. L'anza Research Int'l, Inc.*, 523 U.S. 135, 152 (1998); *see Kirtsaeng v. John Wiley & Sons, Inc.*, 133 S. Ct. 1351, 1354–55 (2013). ...

As an initial matter, it should be noted that the fair use [*Ed*: presumably "first sale"] defense is, by its own terms, limited to assertions of the *distribution right.* Because the Court has concluded that ReDigi's service violates Capitol's reproduction right, the first sale defense does not apply to ReDigi's infringement of those rights.

In addition, the first sale doctrine does not protect ReDigi's distribution of Capitol's copyrighted works. This is because, as an unlawful reproduction, a digital music file sold on ReDigi is not "lawfully made under this title." Moreover, the statute protects only distribution by "the owner of a *particular* copy or phonorecord ... of *that* copy or phonorecord." Here, a ReDigi user owns the phonorecord that was created when she purchased and downloaded a song from iTunes to her hard disk. But to sell that song on ReDigi, she must produce a new phonorecord on the ReDigi server. Because it is therefore impossible for the user to sell her "particular" phonorecord on ReDigi, the first sale statute cannot provide a defense. Put another way, the first sale defense is limited to material items, like records, that the copyright owner put into the stream of commerce. Here, ReDigi is not distributing such material items; rather, it is distributing *reproductions* of the copyrighted code embedded in new material objects, namely, the ReDigi server in Arizona and its users' hard drives. The first sale defense does not cover this any more than it covered the sale of cassette recordings of vinyl records in a bygone era.

Rejecting such a conclusion, ReDigi argues that, because "technological change has rendered its literal terms ambiguous, the Copyright Act must be construed in light of [its] basic purpose," namely, to incentivize creative work for the "ultimate[ ] ... cause of promoting broad public availability of literature, music, and the other arts." *Sony Corp. of America v. Universal City Studios, Inc*, 464 U.S. 417, 432 (1984). Thus, ReDigi asserts that refusal to apply the first sale doctrine to its service would grant Capitol "a Court sanctioned extension of rights under the [C]opyright [A]ct ... which is against policy, and should not be endorsed by this Court."

The Court disagrees. ReDigi effectively requests that the Court amend the statute to achieve ReDigi's broader policy goals—goals that happen to advance ReDigi's economic interests. However, ReDigi's argument fails for two reasons. First, while technological change may have rendered Section 109(a) unsatisfactory to many contemporary observers and consumers, it has not rendered it ambiguous. The statute plainly applies to the lawful owner's "particular" phonorecord, a phonorecord that by definition cannot be uploaded and sold on ReDigi's website. Second, amendment of the Copyright Act in line with ReDigi's proposal is a legislative prerogative that courts are unauthorized and ill suited to attempt.

Nor are the policy arguments as straightforward or uncontested as ReDigi suggests. For instance, the United States Copyright Office stated that "the impact of the [first sale] doctrine on copyright owners [is] limited in the off-line world by a number of factors, including geography and the gradual degradation of books and analog works." USCO, Library of Cong., DMCA Section 104 Report at xi (2001). Specifically,

> [p]hysical copies of works degrade with time and use, making used copies less desirable than new ones. Digital information does not degrade, and can be reproduced perfectly on a recipient's computer. The "used" copy is just as desirable as (in fact, is indistinguishable from) a new copy of the same work. Time, space, effort and cost no longer act as barriers to the movement of copies, since digital copies can be transmitted nearly instantaneously anywhere in the world with minimal effort and negligible cost. The need to transport physical copies of works, which acts as a natural brake on the effect of resales on the copyright owner's market, no longer exists in the realm of digital transmissions. The ability of such "used" copies to compete for market share with new copies is thus far greater in the digital world.

Id. at 82–83 (footnotes omitted). Thus, while ReDigi mounts attractive policy arguments, they are not as one-sided as it contends.

Finally, ReDigi feebly argues that the Court's reading of Section 109(a) would in effect exclude digital works from the meaning of the statute. That is not the case. Section 109(a) still protects a lawful owner's sale of her "particular" phonorecord, be it a computer hard disk, iPod, or other memory device onto which the file was originally downloaded. While this limitation clearly presents obstacles to resale that are different from, and perhaps even more onerous than, those involved in the resale of CDs and cassettes, the limitation is hardly absurd—the first sale doctrine was enacted in a world where the ease and speed of data transfer could not have been imagined. There are many reasons, some discussed herein, for why such physical limitations may be desirable. It is left to Congress, and not this Court, to deem them outmoded.

Accordingly, the Court concludes that the first sale defense does not permit sales of digital music files on ReDigi's website. …

D. NFTs and Art

FREE HOLDINGS INC. V. MCCOY

No. 22-CV-881 (JLC) (S.D.N.Y. Mar. 17, 2023)

Cott, Magistrate Judge:

Free Holdings, Inc. has brought this action involving the digital artwork, *Quantum,* and the first non-fungible token ("NFT") ever created. It has sued *Quantum*'s creator, artist Kevin McCoy, and auction house Sotheby's Inc. Free Holdings alleges multiple claims stemming from statements made by defendants in promotional material for an auction of *Quantum* and its digital record. Defendants have each moved to dismiss the amended complaint. For the reasons set forth below, the motions are granted.

I. Background

A. Factual Background …

1. NFTs and Namecoin

As alleged in the amended complaint, an NFT is a unique identifier that "authenticates digital content on the blockchain, which is "a digital public ledger maintained on a decentralized computer system and consisting of records called blocks." "The blockchain's record-keeping and authentication technology serves to provide public certificates of authenticity or proof of ownership of NFTs." "Content linked to NFTs can take the form of digital images of physical objects, music, text, and more."

Namecoin, an early blockchain, is a system of "names"—unique combinations of numbers and letters—that can be claimed and traded by users. *See also* Monolithbrah.eth, *et al., Defining "NFT" in Historical Context* ("*Defining NFT*") (Jun. 27, 2022) https://mirror.xyz/chainleft.eth/MzPWRsesC9mQflxlLo-N29oF4iwCgX3lacrvaG9Kjko. Ownership of every name periodically expires, at which point any user may freely claim it on the Namecoin blockchain by re-registering the expired name. In the community of blockchain users, there is an ongoing debate about the status of names that expire and are then re-registered: namely, whether re-registered names become new NFTs or are the same NFTs that were previously claimed. The debate is summarized in the *Defining NFT* article as follows:

> Namecoin is designed around domain names, but when a domain name gets registered the first time, it is represented by a particular identifier (UTXO design) to allow its trading. The chain of these identifiers (UTXOs) can be considered a special coin, which we will call a "colored coin" in this article. …
>
> When a Namecoin domain expires, the colored coin becomes unspendable. In practice, it can be called "burnt", because when the domain name is re-registered later, a new UTXO chain starts. This would mean that when a domain name expires, the chain of events where that domain was represented breaks. During the re-registration, the Namecoin protocol does not make a distinction between a never-used domain name and a previously expired domain name. This fact may support the argument that the expiration is philosophically burning the original token.

> On the other hand, as mentioned before, Namecoin is designed around names. Names are unique identifiers themselves and they function that way immutably in a cryptographically secured blockchain. The history of the unique name, previous updates and transactions remain on-chain, as is the case with blockchains.
>
> The aforementioned "UTXO chains" are actually formed of individual UTXO stations. Each UTXO has an address and domain names are attached to these UTXOs. When Bob makes a standard domain name transfer to Alice, a 0.01NMC is spent on Bob's side where the domain name is attached, and a new 0.01NMC is created on Alice's side where the domain name is attached. The thing that physically (for lack of a better term) moves from one address to another during a typical transfer is the domain name, not UTXOs. The design choice of a typical domain name transfer supports the argument that the domain name is the unique identifier and the token, and it never burns even after expiration.
>
> This domain-first design can be seen not only in block explorers but also in functions to utilize the domain names. There's a built-in [remote procedure call ("RPC"), a software communication protocol] method since the beginning of Namecoin called name_show which returns the information about the domain name even during the expired period. This function does not return anything for never-registered domains.
>
> In fact, an expired domain can still be functional. When expired, Namecoin Core will still have the name and return its local state with name_show (just indicating the expired status). If the resolver setup accepts it, the domain will still work. Similarly, expired "id/" names in Namecoin still work as logins, which indicates the usefulness of the name_show function.

Consequently, as the article explains, there are three different interpretations of the significance of ownership of a re-registered name on Namecoin:

> Interpretation 1: *The token is the colored coin, and that is the asset.* Names may be unique but they are cryptographically represented by tokens and only that representation matters. Since a re-registered name is assigned to an entirely new colored coin, the provenance from the earlier UTXO chain is broken, therefore *this is a new asset and cannot claim historical value.*
>
> Interpretation 2: *The token is the name, and that is the asset.* Namecoin is designed around names where names are the main unique identifiers. This is supported by the technical design when you do a typical name transfer or assign a value field to the name. These are all visible on-chain. Further, when the name expires, the data of the name remains on the blockchain. If the blockchain is designed around names and their whole history is on-chain, when the name is re-registered you are simply reactivating the old token, therefore *provenance is not broken.*
>
> Interpretation 3: *The token is the colored coin because it is a cryptographic representation of the name.* But the main collectible asset is the name itself and these names have their own provenance because Namecoin is designed around names as unique identifiers. A

> name's history and UTXO assignments are all visible on-chain. This is further proven by the existence of name_show and name_history RPC methods. Therefore when the name is re-registered, *the new colored coin (UTXO chain) can be assumed as a new NFT, but it represents the old decentralized collectible asset which was intact and ownable since its birth.*

Id. (emphasis added). To summarize, as the *Defining NFT* article makes clear, some believe that when a new user re-registers an available name previously claimed by someone else, it has claimed the same NFT claimed by the previous user. Others, who believe that the NFT is the UTXO chain, not the name, believe that when a new user re-registers an expired name, the new UTXO chain it creates signifies a new NFT. A third group ascribes to a combination of the two theories: that the UTXO chain becomes a new NFT that retains the entire history of the name.

2. McCoy's Creation and Preservation of Quantum

In 2014, digital artist Kevin McCoy created *Quantum* (viewable at https://static.mccoyspace.com/gifs/quantum.gif). McCoy was interested in experimenting with the blockchain to store his work because he "thought about how the scarcity mechanism that bitcoin presented could be a way for digital artists to create provenance and ownership systems around digital works." Because of this interest, McCoy created the digital record of *Quantum* on Namecoin on May 2, 2014—"the first NFT." AC ¶¶ 2, 98; *see* Namebrow.se, Namecoin Explorer, https://namebrow.se/name/d41b8540cbacdf1467cdc5d17316dcb672c8b43235fa16cde98e79825b68709a/ [https://perma.cc/N4J6-5Y3U] ("-709a Record Page"). To do this, McCoy entered the "NAME_NEW" and "NAME_FIRSTUPDATE" functions on Namecoin to register the name d41b8540cbacdf1467cdc5d17316dcb672c8b43235fa16cde98e79825b68709a ("-709a") as a digital record of *Quantum*. -709a Record Page. The -709a Record Page provides the following disclaimer:

> An important distinction between property & deed must be made when considering Namecoin NFTs. A UTXO, the thing that transfers ownership [between] holders of public/private key pairs, is a DEED to property but NOT property itself. The property that the Namecoin blockchain was built to cryptographically secure ownership of, provided all recurring fees have been paid to the protocol, is a unique plot of digital space known as a Name. As such, Names, along with the history of Values associated to them, are the NFT property.

-709a Record Page. On May 28, 2021, McCoy "minted another NFT" to record the same moving image of *Quantum,* this time on the Ethereum blockchain.

3. Claims on Namecoin and Communications on Twitter

In January 2015, name -709a expired. On April 5, 2021, a new user entered "NAME_NEW" and "NAME_FIRSTUPDATE" on the -709a name entry. -709a Record Page. On April 30, 2021, a user entered "NAME_UPDATE" and wrote a message that appears under the "Value" column of that entry:

> I assert title to the file at the URL http://static.mccoyspace.com/gifs/quantum.gif with the creator's public announcement of it's publishing at the URL [omitted]. The file whose SHA256 hash is d41b8540cbacdf1467cdc5d17316dcb672c8b43235fa16cde98e79825b68709a is taken to

> be the file in question. Title transfers to whoever controls this blockchain entry.

Free Holdings reports that on Namecoin.org, the command "name_history d41b8540cbacdf1467cdc5d17316dcb672c8b43235fa16cde98e79825b68709a," returns the same message.

As is displayed on the -709a Record Page and reproduced in the chart below, above the April 30, 2021 entry, several more "NAME_UPDATE" entries appear with dates, codes, and messages. -709a Record Page. The entries do not identify authors, but delineate "block" numbers. *Id.* For instance, the block number associated with the "NAME_NEW" operation performed on May 2, 2014 is 174910; the block number associated with the "NAME_FIRSTUPDATE" operation performed on May 3, 2014 is 174923. *Id.* Then, the "NAME_NEW" operation performed on April 5, 2021 appears at block 553180, and subsequent entries appear next to higher block numbers. *Id.* Thus, according to Free Holdings, "all subsequent activity associated with [-709a] is forever linked to that name on the Namecoin blockchain" and "all prior history associated with that name … cannot be deleted so long as the Namecoin blockchain continues to operate." Peltz Decl. ¶ 16.

Beginning in April 2021, Free Holdings alleges that it attempted to contact McCoy on Twitter (@mccoyspace) using the handle @EarlyNFT. On April 6, 2021, @EarlyNFT tweeted: "Would you mind enabling PM for me, @mccoyspace? There's a matter regarding your work 'Quantum' I'd like to discuss with you …" On April 12, 2021, @EarlyNFT replied to one of @mccoyspace's tweets: "Kevin, would you mind allowing me to send you a private message (this option is currently turned off on your end)? This concerns your NFT artwork with Monegraph back in 2014." On April 30, 2021, @EarlyNFT replied to @mccoyspace: "I am in possession of Quantum NFT (the record is found here [link inaccessible by the Court]). Can we discuss this over messaging (currently your messaging is turned off)?" [Further tweets omitted.] McCoy never responded to the Twitter communications.

4. The Auction and Sale of Quantum

In May 2021, the well-known auction house Sotheby's began marketing *Quantum* for an auction entitled *Natively Digital: A Curated NFT Sale,* scheduled for June 2021. AC ¶¶ 8, 52; *see* Sotheby's, Natively Digital: A Curated NFT Sale, https://www.sothebys.com/en/buy/auction/2021/natively-digital-a-curated-nft-sale-2/quantum [https://perma.cc/B3YA-A9HP]. Natively Digital includes the following description below the moving image, *Quantum*:

> Owner: Kevin McCoy Kevin McCoy b. 1967 *Quantum* 2014-21 Non-fungible token ERC-721 Token ID: 0 9 mb Tif + file archive with documentation Originally minted on May 3, 2014 on Namecoin blockchain, and preserved on a token minted on May 28, 2021 by the artist.

Sotheby's marketed *Quantum* as "the most art historically important work in the history of NFTs." AC ¶ 59 (quoting Natively Digital). The Catalogue Note for *Quantum* provides the following description:

> In the long timeline of art, there are few works that serve as genesis blocks to their own chain of history. They are seismic forks in direction; forks that usher in new movements that block by block, mint by mint, usher in new art histories. These works close chapters on the art histories that came before, while anchoring a new flowering of human

> creativity. These prime movers occupy a singular position in art history. They came first. Kevin McCoy's *Quantum* is such a work. Minted on 2nd May 2014 21:27:34, or more precisely Namecoin Block 174923, the NFT era quietly dawned. What a noise it makes today.

AC ¶ 58 (quoting Natively Digital, Catalogue Note).

A "Condition Report" on Natively Digital, authored by nameless TM, explains *Quantum*'s digital history, in relevant part:

> The hash and all the information about [*Quantum*] are stored on the Ethereum blockchain, and therefore cannot be modified. This token was minted May 2021, but the referenced pieces were originally made public in 2014, with a link to the image hosted on the Namecoin network. The token is a very early example of utilizing a blockchain token to identify the provenance of a piece of art. Documents contained within the media bundle hosted on [the InterPlanetary File System] indicate the date and block location that the Namecoin DNS pointer was set to point to mccoyspace.com, and shows the transactions used at the time to create this trail. To avoid domain squatting, Namecoin was designed to include removal of pointers after 36,000 blocks. Accordingly, this specific Namecoin entry was removed from the system after not being renewed, and was effectively burned from the chain.

Natively Digital; AC ¶¶ 64-65.

McCoy appears in two videos on Natively Digital. *See* Rothstein Decl., Ex. C-F. In one video, McCoy describes the "restoration and preservation efforts that [he] made to bring this early work back into the present day." *Id.*, Ex. E, Ex. F ("*Quantum* Token — McCoy Tr.") at 2:4-6. On his computer screen, McCoy shows the viewer how he "moved the original on-chain data from a burned Namecoin token into a modern, industry standard, ERC 721 token, while preserving all of the original on-chain information." *Id.*, Ex. E, *Quantum* Token — McCoy Tr. at 2:8-12; *see also* AC ¶ 66 (quoting same). McCoy explains that the "provenance chain ... begins on Namecoin but now is irrevocably transferred to this more modern and stable specification on the Ethereum blockchain." *Quantum* Token — McCoy Tr. at 3:23-4:2. McCoy displays

> an archive called 'wallet documentation[,]' that contains a set of screenshots taken from the original Namecoin wallet showing the original transaction set, the original on-chain data, and a signed message proving that [he] ha[s] continued possession of the private key for the address that originally sent *Quantum*'s minting transaction. That signed message appears separately in the archive as a text file and the data of which can be used in a contemporary Namecoin wallet to independently verify the message's validity.

Rothstein Decl., Ex. E, *Quantum* Token — McCoy Tr. at 4:3-15. McCoy concludes: "[a]s we enter into this new era of tokenization, I think we will increasingly be faced with issues of cross-chain translation, preservation, and archiving and conservation. In fact, we're already there with *Quantum*." *Quantum* Token — McCoy Tr. at 5:3-8.

Sotheby's held the auction for *Quantum* from June 3-10, 2021. AC ¶¶ 77-78. On June 17, 2021, Free Holdings allegedly spoke to Caroline Moustakis ("Moustakis"), Sotheby's Senior Vice President, and told her "that the description that Sotheby's posted ... was inaccurate and misleading because [the record of *Quan-*

tum on Namecoin] had not been 'burned' or 'removed' from the Namecoin blockchain, but rather was still active and controlled by Free Holdings." AC ¶ 79. "Free Holdings requested that Sotheby's make public statements to correct the record, including making a statement that the [the record of *Quantum* on Namecoin] remains active and in private hands, was not listed for sale as part of Sotheby's auction, and that the item Sotheby's auctioned was in fact an authorized print of the Namecoin-*Quantum* token." *Id.* Free Holdings then emailed Moustakis, memorializing the conversation and requesting that the public record be corrected, but neither Moustakis nor any other Sotheby's representative responded.

On August 23, 2021, McCoy and Sotheby's sold the "re-minted" *Quantum* to Alex Amsel ("Amsel"), who uses the Twitter name @Sillytuna, for a reported sum of $1,472,000. AC ¶ 84. Amsel tweeted: "First ever crypto art nft — Quantum has arrived chez Sillytuna. Take that VISA!" AC ¶ 86.

After the sale, an article in the *New York Times* reported that the "first NFT that Kevin McCoy created ... sold last June at Sotheby's for $1.4 million." A segment on Fox Business ran the banner: "FIRST-EVER NFT 'QUANTUM' SOLD FOR $1.47 MILLION." Free Holdings alleges that it expended $5,311.49 "seeking to have McCoy and Sotheby's issue public statements correcting their false and misleading claims."

On May 6, 2022, nameless tweeted:

> "Last year, nameless prepared a condition report in connection with the Sotheby's auction of Kevin McCoy's Ethereum-based 'Quantum' NFT [(hyperlink to Natively digital)]. McCoy's 'Quantum' NFT was originally created on the Namecoin blockchain. The condition report included a statement that Quantum's 'specific Namecoin entry was removed from the system after not being renewed, and was effectively burned from the chain.' In light of subsequent allegations challenging this statement, nameless has decided to retract its condition report. [N]ameless expresses no opinion about the merits of the allegations.

AC ¶ 88.

B. Procedural History

Free Holdings, a Canadian corporation, originally brought this action against McCoy; Sotheby's; nameless, a corporation based in Colorado; and Alex Amsel, the buyer of the *Quantum* NFT, on February 1, 2022, challenging their characterizations that they were involved with the first NFT, and that the Namecoin record associated with *Quantum* had been "burned" or "removed." Complaint ("Compl."), Dkt. No. 1; *see also* AC ¶¶ 60-70. On March 8, 2022, Free Holdings voluntarily dismissed its claims against Amsel. Dkt. No. 24. On April 6, 2022, the parties consented to my jurisdiction for all purposes pursuant to 28 U.S.C. § 636(c). Dkt. No. 38. On May 9, 2022, Free Holdings filed an amended complaint, in which it asserts five causes of action: (1) unjust enrichment (¶¶ 110-16); (2) slander (¶¶ 117-26); (3) deceptive and unlawful trade practices under New York General Business Law ("G.B.L.") § 349 (¶¶ 127-39); (4) commercial disparagement (¶¶ 140-48); and (5) false or misleading representations under 15 U.S.C. § 1125(a) (¶¶ 149-58). Dkt. No. 47. It further requests that the Court:

> declar[e] and adjudicat[e] that (i) Free Holdings is the rightful owner of the Namecoin-based *Quantum*; (ii) the Namecoin-*Quantum* has not been burned or otherwise removed from the Namecoin blockchain; and (iii) the statements issued by McCoy and Sotheby's in

> connection with their sale of the Ethereum-*Quantum* were false and misleading[;]

and seeks damages and injunctive relief. *Id.* at 24-25. ...

II. DISCUSSION

In their respective motions, defendants contend that Free Holdings has failed to allege an injury sufficient for standing or a claim that is ripe for review, warranting dismissal for lack of subject matter jurisdiction under Rule 12(b)(1). In the alternative, they argue that all of Free Holdings' claims should be dismissed under Rule 12(b)(6) for failure to state a claim. For the following reasons, the Court concludes that Free Holdings has not alleged an injury sufficient for standing to sue. The Court further concludes that even if Free Holdings had standing, it has failed to state a claim upon which relief can be granted. ...

B. The Court Lacks Subject Matter Jurisdiction Over Free Holdings' Claims

The gravamen of Free Holdings' amended complaint is that its alleged proprietary interest in "Namecoin-*Quantum*"—its term for the NFT it claims on Namecoin—was injured by defendants' allegedly false statements that the record of *Quantum* on Namecoin was "burned" or "removed" and their allegedly false representation that the "NFT they sold at auction was the first NFT ever created." AC ¶¶ 119-20, 142-43, 151-52.[*] Defendants counter that those statements and characterizations were not false, and that Free Holdings has failed to allege a concrete injury resulting from them. McCoy Mem. at 2; Sotheby's Mem. at 1, 17-18. Thus, defendants move to dismiss for lack of subject matter jurisdiction.

1. Free Holdings' Claims Are Not Justiciable

The Court first analyzes whether Free Holdings has alleged an injury sufficient for standing under Article III of the U.S. Constitution. As discussed below, Free Holdings has alleged no concrete or particularized injury sufficient for standing to sue in federal court.

a. Article III Standing and Ripeness

The Constitution limits the judicial power of federal courts to decide cases or controversies. *See* U.S. Const. art. III, § 2, cl. 1. Constitutional standing thus is the threshold question in every federal case, determining the power of the court to entertain the suit.

Article III requires a plaintiff first to allege that it suffered an injury in fact," *i.e.*, "an invasion of a legally protected interest which is (a) concrete and particularized and (b) actual or imminent, not conjectural or hypothetical. The remaining requirements for Article III standing are a causal connection between the injury and the conduct complained of and a likelihood, as opposed to mere speculation, that the injury will be redressed by a favorable decision. ...

---

12. Free Holdings's choice of terminology—"Namecoin-*Quantum*" and "Ethereum-*Quantum*"—supports its view that there are two NFTs called *Quantum* and that it has ownership rights to the one on Namecoin. This terminology is not used in any of the exhibits Free Holdings attaches or in the sources to which it cites. Defendants alternatively refer to two Namecoin NFTs—a "2014 NFT" and a "2021 NFT"—to support their position that Free Holdings' proprietary interest is in a new, distinct NFT it created on Namecoin in 2021 when it claimed -709a. At the motion to dismiss stage, the Court, construing the pleadings in Free Holdings' favor, refers to its terminology where it would not cause confusion.

i. Free Holdings Has Failed To Allege A Proprietary Interest In Quantum

Free Holdings claims a proprietary interest in "the blockchain record for the Namecoin-*Quantum*," which, it alleges, "remains active and under its control." The claim to -709a, under Free Holdings' theory of NFT ownership, confers a property interest in *Quantum* itself. Notably, Free Holdings nowhere alleges an interest in the NFT minted on Ethereum that McCoy and Sotheby's sold. Indeed, it alleges that the two "are different NFTs." McCoy disputes Free Holdings' theory of ownership, arguing that a name on Namecoin is distinct from the underlying NFT. Thus, according to McCoy, Free Holdings merely claimed -709a, creating a distinct NFT, which confers no proprietary interest in *Quantum* itself..

Free Holdings does provide an adequate factual basis to support a plausible proprietary interest in -709a—that is, assertions of title displayed on the -709a Record Page and by Twitter user @EarlyNFT. *See* AC ¶¶ 38, 42-43. Free Holdings has thus alleged a proprietary interest in -709a sufficient for standing. However, Free Holdings has not articulated any facts to support its claim to ownership of *Quantum* vis-à-vis its claim to -709a. As McCoy points out, Free Holdings "never alleges that it can or could control" *Quantum*. As factual support, Free Holdings provides only its own alleged statements on Namecoin that it "assert[ed] title to the file at the URL http://static.mccoyspace.com/gifs/quantum.gif" and that "[t]itle transfers to whoever controls this blockchain entry," -709a Record Page, along with an article explaining that the significance of name ownership on Namecoin is open to debate. Even taken together, they are insufficient to give rise to a legally-protected interest in *Quantum*.

ii. Free Holdings Has Failed To Allege Injury-In-Fact

Setting aside conclusory statements that merely recite the elements of the claims in the pleadings, the harm that Free Holdings alleges as a result of defendants' statements in Natively Digital, the challenged publication, consists of three theories: (1) loss of opportunity to profit from Sotheby's and McCoy's $1,472,000 sale; (2) damage to the value of its claimed Namecoin property; and (3) the expense of "$5,311.49 seeking to have McCoy and Sotheby's issue public statements correcting their false and misleading claims."

In support of the first theory, Free Holdings points to the tweet from @EarlyNFT that it allegedly directed at McCoy's Twitter account asking, "Are you interested in the sale of Quantum or not then?" AC ¶ 48. Free Holdings argues that "if McCoy had taken up this offer," its claimed Namecoin NFT "could have been included in McCoy's sale ... at the Sotheby's auction (and corrected related marketing statements)." But the pleadings raise no suggestion that defendants either were legally obligated to include Free Holdings in their auction sale, or that McCoy even responded to (or should have responded to) the cryptic offer made by an anonymous Twitter user. ...

Regarding the second theory—damage to the value of its claimed Namecoin property—as McCoy points out, Free Holdings "does not allege any actual, imminent or attempted sale of its claimed NFT," nor that it "even *tried* to market it." McCoy Mem. at 13-14 (emphasis in original). Free Holdings' vague speculation that "any prospective purchaser of Namecoin-*Quantum* would now have valid reason to doubt its value," is legally insufficient. ...

C. The Amended Complaint Fails To State A Claim

Having concluded that Free Holdings has not alleged injury sufficient for standing, the Court does not need to consider, as a matter of law, whether the amended

complaint states any claims. Nonetheless, both for completeness' sake and because of the novelty of the issues presented here, it will analyze each of Free Holdings' claims for relief to determine if any of them are cognizable. For the reasons set forth below, the Court concludes that none are. Free Holdings has not pleaded any viable cause of action against defendants that would otherwise survive a motion to dismiss, even if it were to establish standing. ...

1. Free Holdings Has Failed To State A Claim Of Unjust Enrichment

As to its unjust enrichment claim, Free Holdings alleges that its Namecoin NFT "has significant value due to its status as the first NFT and such value is being significantly diminished by defendants' false and misleading statements." It argues that defendants were unjustly enriched when they "claimed the identity, titles, and prestige associated with the Namecoin-*Quantum* in order to market, promote, and sell their own, separate asset."

"An unjust enrichment claim is rooted in the equitable principle that a person shall not be allowed to enrich himself unjustly at the expense of another." *Georgia Malone & Co. v. Rieder,* 19 N.Y.3d 511, 516 (2012). "Thus, in order to adequately plead such a claim, the plaintiff must allege that (1) the other party was enriched, (2) at that party's expense, and (3) that it is against equity and good conscience to permit the other party to retain what is sought to be recovered." *Id.* While there is no dispute that defendants were enriched (they made a $1,472,000 sale), Free Holdings does not allege that it has any proprietary interest in the NFT that McCoy and Sotheby's sold. Moreover, Free Holdings has made no specific allegations that defendants' sale of an NFT (it does not even claim to own) was at its expense. ...

2. Free Holdings Has Failed To State Claims Of Slander Of Title Or Commercial Disparagement

Free Holdings next alleges claims of slander of title and commercial disparagement based on its purported title on Namecoin. For a viable slander of title claim, a plaintiff must adequately plead: (1) a communication falsely casting doubt on the validity of its title, (2) reasonably calculated to cause harm (in other words, malice), and (3) resulting in special damages. Likewise, product disparagement consists of: (1) a false statement; (2) published by Defendant to a third party; (3) with malice; and (4) with special damages. For either claim to survive, Free Holdings must therefore plausibly allege falsity, malice, and special damages. ...

i. Free Holdings Has Not Alleged Falsity

The burden is on a plaintiff to allege that the complained of statement is substantially false. Expressions of pure opinion are not actionable.

Free Holdings has failed to meet its burden for several reasons. First, it concedes that the representations it challenges—that *Quantum* was "burned" or "removed" from the Namecoin blockchain and that the NFT sold at auction "was the first NFT ever created,"—are topics of debate. It alleges that "Namecoin-*Quantum* ... has not been 'burned,' and is currently controlled and owned by Free Holdings," but goes on to contradict that allegation in its own pleadings and in the materials it cites. Specifically, Free Holdings admits that the NFT McCoy created on Namecoin "expired," yet it also observes that "defining the legal import of control of digital assets on the Namecoin blockchain is the subject of differing opinions." Further, it cites and attaches to its motion papers an article explaining that "when a Namecoin domain expires, ... it can be called 'burnt.'" *Defining NFT*.

Second, Free Holdings uses terminology that is misleading and fails to show falsity. It alleges that as "Namecoin-*Quantum* is considered the first NFT ever created," AC ¶ 37, Sotheby's falsely marketed "Ethereum-*Quantum* by Kevin McCoy ... as the first NFT ever created." AC ¶ 120; *see also* Pl. Mem. Sotheby's at 10 ("the pleaded truth would ... reveal to potential bidders that the Ethereum-*Quantum* was a copy of the presently-extant Namecoin-*Quantum,* rather than the alleged sole surviving vestige of a prized digital collectible."). Its allegations never quote defendants—or anyone else—stating that "Ethereum-*Quantum*" was the first NFT created. *See, e,g.,* AC ¶¶ 54-74. "Ethereum-*Quantum*" and "Namecoin-*Quantum*" are terms Free Holdings employs in its pleadings, and there is no evidence to suggest that defendants or anyone else ever used them in any public statements or challenged materials. Indeed, the statements Free Holdings cites to in its amended complaint—by defendants and others—characterize *Quantum* as the first NFT ever created, and do not distinguish between an "Ethereum-*Quantum*" and "Namecoin-*Quantum.*" *See* AC ¶ 58 ("Kevin McCoy's *Quantum* is such a work. Minted on 2nd May 2014 21:27:35, or more precisely Namecoin block 174923, the NFT era quietly dawned."); ¶ 86 ("First ever crypto art nft — Quantum has arrived chez Sillytuna"); ¶ 95 ((citing Gopnik) ("That first NFT that Kevin created, which he'd attached to a modest digital abstraction titled 'Quantum,' sold last June at Sotheby's for $1.4 million."); ¶ 97 ("FIRST-EVER NFT 'QUANTUM' SOLD FOR $1.47 MILLION").

Furthermore, Free Holdings nowhere disputes that "McCoy and Sotheby's accurately reported that *Quantum* was originally minted in 2014 on the Namecoin blockchain, and that McCoy 'moved the original on-chain data from a burned Namecoin token into a modern, industry standard, ERC 721 token, while preserving all of the original on-chain information.'" Sotheby's Mem. at 18 (citing Rothstein Decl., Ex. F.). Instead, it argues only that defendants purveyed a false "overall impression" that "Namecoin-*Quantum* did not exist," without citing any authority for its position that the "overall impression" of defendants' statements is to be considered in the falsity analysis. Moreover, defendants did not convey an "overall impression ... that Namecoin-*Quantum* did not exist," because they actually explained *Quantum*'s blockchain history multiple times in *Natively Digital.* Pl. Mem. Sotheby's at 11. For instance, the description explains that the work was "originally minted on May 3, 2014 on Namecoin blockchain, and preserved on a token minted on May 28, 2021 by the artist," *Natively Digital*—a recounting consistent with Free Holdings' view that "Ethereum-*Quantum* is definitively not the world's first NFT." The Condition Report on *Natively Digital* explains that the NFT for auction "was minted May 2021, but the referenced pieces were originally made public in 2014, with a link to the image hosted on the Namecoin network," and in one of his videos on Natively Digital, McCoy explains precisely how the "provenance chain ... begins on Namecoin but now is irrevocably transferred to this more modern and stable specification on the Ethereum blockchain."

Moreover, by Free Holdings' own allegations, it was not false for McCoy to explain that "the nature of *Quantum* lies in the immutable engraving of information to the Namecoin blockchain" and that it was based on his "idea to use blockchain technology to create indelible provenance and ownership of digital images of this kind." In fact, Free Holdings acknowledges that *Quantum* was created by McCoy, and that it was the first NFT. Its position is that it claims title to an "indelible" record of *Quantum* on Namecoin, but, as discussed above, it does not claim title to the record of *Quantum* maintained on Ethereum. Free Holdings has thus failed to

sufficiently allege that any statements made by McCoy or Sotheby's were substantially false. ...

JAMES GRIMMELMANN, YAN JI, AND TYLER KELL

COPYRIGHT VULNERABILITIES IN NFTS

Medium (Mar. 21,2022)
https://medium.com/initc3org/copyright-vulnerabilities-in-nfts-317e02d8ae26

Many NFT and DAOs are designed to provide new or more convenient ways to own and sell creative works. Beeple's *EVERYDAYS: The First 5000 Days* sold at auction for $69 million. Some observers think that the Bored Ape Yacht Club's spectacular rise is due to its permissive copyright approach. Some artists and developers are diving in head-first.

But at the same time, many of these projects have run into copyright trouble due to confusion about how copyright applies to NFTs:

- When the SpiceDAO bought a copy of the lavishly illustrated pitch book that director Alejandro Jodorowsky made for a never-filmed version of *Dune*, some participants hoped that buying the book would allow them to bring Jodorowsky's vision to the screen. But this plan was quickly scrapped when owners of the *Dune* copyrights vetoed the idea.
- "Right-clickers" save JPEG copies of the artwork from popular NFTs. The owners of those NFTs say this is copyright infringement. Only one of the two can be right.
- Quentin Tarantino and Miramax are locked in litigation over the rights to *Pulp Fiction* NFTs.
- Too many NFTs to count use stolen art.

We hope in this blog post to clear up some of the confusion around NFT copyrights, and to help people working in the space understand the challenges of fitting NFTs to the framework of copyright law. Our bottom line is simple: ownership of an NFT can be used to give the owner substantial control over a creative work, but that control is not automatic. Copyright law does not give an NFT owner any rights unless the creator takes affirmative steps to make sure that it does. Our survey of some existing NFT projects and their licenses reveals that very few of them take all of the necessary steps needed to make NFT copyrights behave the way that community members expect. Thinking through the legal issues should be part of the NFT design process, not an afterthought.

On-Chain and Off-Chain Assets

When talking about blockchain assets, it is common to say things like "Alice owns 10 Bitcoin." Most of the time, it is clear what this means: Alice controls the private key to a blockchain address that has been transferred 10 Bitcoin in unspent transaction outputs. Alice, if she wishes, can use that private key to transfer those Bitcoin to another address. (Alice might control the key directly or via a wallet; we will ignore this complication in what follows.) In the words of the current draft Uniform Commercial Code article for blockchain assets, Alice has "the power to avail [herself] of substantially all the benefit from" those Bitcoin. For more complicated assets, like an NFT controlled by a smart contract, this control might take the form of knowing the private key needed to initiate a transaction transferring control to someone else.

This case is easy because Bitcoin and smart contract tokens are on-chain assets. They exist entirely as entries on a blockchain. Matters are more complicated for

off-chain assets, in which an entry on a blockchain is used to represent something that exists somewhere else, like a tungsten cube. The cube itself is a physical object that exists in the physical world. It weights 2,000 pounds, measures 14.545 inches on a side, and sits in the Willowbrook, Illinois warehouse of Midwest Tungsten Service. But the Tungsten Cube NFT is an entry in a smart contract deployed on the Ethereum blockchain using the ERC-1155 standard. If Alice buys the Tungsten cube NFT from TungstenDAO, the physical cube will still be sitting in Willowbrook, Illinois.

Although the Tungsten Cube NFT is not the same as the physical tungsten cube, the two are connected. According to the "Description" provided by TungstenDAO to OpenSea when it created the NFT listing, when Alice acquires the NFT, she is also entitled to "One visit to see/photograph/touch the cube per calendar year." If Alice sends the NFT to a special address that prevents anyone from ever controlling the NFT again (a process called burning), she is entitled to receive physical possession of the cube "via freight truck." If she sells the NFT to Bob, then Bob will be entitled to visit the cube once a year, or to burn the NFT and receive the cube. Some lawyers call this connection "tethering": rights in an off-chain asset (the physical cube) are linked by an invisible tether to an on-chain asset (the NFT). (In theory, at least. Some legal scholars are skeptical about whether tethering actually works.)

Thus, there are actually three different kinds of assets involved in an NFT. Yes, three.

First, there is an NFT itself on the blockchain. It looks like this:

You can view the creation of this NFT online on the Etherscan blockchain explorer here .

Second, there is the physical cube in a warehouse. It looks like this:

:ontrol the physical cube. It looks like this, be-
cau have no physical existence:

If everything works correctly, the legal right is what links the on-chain NFT to the off-chain cube. The current owner of the NFT is able to control the physical cube because they also own the associated legal right.

A Copyright Primer

The following quote is sometimes (humorously but incorrectly) attributed to Albert Einstein:

> You see, wire telegraph is a kind of a very, very long cat. You pull his tail in New York and his head is meowing in Los Angeles. Do you understand this? And radio operates exactly the same way: you send signals here, they receive them there. The only difference is that there is no cat.

A copyright is exactly like a tungsten cube except that there is no cube. A tungsten cube is a specific physical object. It exists in one place. Any other tungsten cube is a different cube. But a creative work like a photograph or a story is intangible. It can exist in exactly one object (called a "copy") like an oil painting on a canvas. Or a work can exist in many copies at once like when a publisher prints thousands of copies of a book, or when a photograph is displayed on millions of computer screens. Or a work can exist in no copies at all, like when one person tells another person a story. The point is that a creative work is not the same as any of the copies that embody it.

Thus, the copyright in a creative work behaves differently than ownership of a physical object, like a tungsten cube. The owner of the cube can move it, sculpt it, or melt it down; if someone else does any of these, they violate the owner's property rights. But the owner of a copyright isn't necessarily the owner of any specific copies. If Alice buys a copy of Bob's novel, Alice owns the physical copy—the paper with ink marks on it—but Bob owns the copyright in the words.

Instead, Bob's copyright consists of a limited set of exclusive rights. Most importantly, Bob can prevent anyone, including Alice, from making more copies of his novel. (That's why it's called a "copy" "right.") If Alice wants to make a movie adaptation of Bob's novel—in copyright terminology, a derivative work—she needs Bob's permission. She can get it in one of two ways. Either she can buy the copyright outright from Bob—a transfer of ownership—or Bob can retain the copyright and give Alice a license to make the movie. The difference is that if Alice becomes the new owner via transfer, she can now decide whether to authorize other uses, like graphic novels and action figures. If Bob merely gives Alice a license, he retains the authority to decide how else to use (or not) the copyright.

NFTs, Copies, and Copyrights

Back to NFTs. It should be apparent now that an NFT can be tethered to a creative work in one of two ways. First, it could be used to control a copy of the work: just like whoever owns the Tungsten Cube NFT is entitled to possession of the tungsten cube, whoever owns the Physical Copy NFT is entitled to possess a specific copy of the work. Second, it could be used to control the copyright in the work: whoever owns the Intangible Copyright NFT is now entitled to decide who gets to make new copies. The two could go together, but they don't have to.

This is where the more ambitious hopes for the SpiceDAO went wrong. The project bought and tokenized ownership of one physical copy of Jodorowsky's

pitch book. The owners of SPICE tokens can (collectively) decide to sell or lend it to others, or put it on public display offline. But they never had, and couldn't tokenize, ownership of the copyright in the underlying creative works. The copyright in *Dune* the novel is still held by Frank Herbert's estate, which licensed film rights to Legendary Entertainment, which produced the 2021 film version; the copyrights in the artwork in the pitch book are held by the original artists and their estates.

Another failure mode for NFT copies is that copyright law has an unintuitive concept of what counts as a copy. We have been talking about obviously distinct physical things, like printed books. But "copies" under U.S. copyright law include any "material objects ... in which a work is fixed by any method now known or later developed, and from which the work can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device." This definition includes hard drives and SSDs, and sometimes even RAM. Every computer that interacts with the work makes a separate "copy" for copyright purposes; even just browsing to a webpage makes a "copy" of the images on it on your computer. Thus, for all practical purposes any NFT that includes digital artwork must include some copyright interest (either a transfer or a license) or the owner of the NFT will become an infringer the moment they attempt to do anything with the artwork.

In particular, it is not sufficient to give an NFT purchaser a copy of the artwork. U.S. copyright law explicitly states that transfers of copyrights and transfers of copies are different:

Transfer of ownership of any material object, including the copy or phonorecord in which the work is first fixed, does not of itself convey any rights in the copyrighted work embodied in the object; nor, in the absence of an agreement, does transfer of ownership of a copyright or of any exclusive rights under a copyright convey property rights in any material object.

If you buy an oil painting from an artist, you don't also receive ownership of the copyright. Yes, you own the original, but the artist retains the copyright, and they can sell prints of it if they like. If you want to buy the copyright too, you'll need to get a separate agreement. The same is true for NFTs. Unless an NFT explicitly gives owners copyright interests as opposed to just access to the artwork, owners should not assume that they have any rights to use the artwork or to stop others from using it.

Some popular NFT projects, including the CryptoPunks, have been released with no explicit copyright terms. This is legally risky for all concerned. An adversary could approach the NFT creator, buy out the copyright to the artwork, and then sue NFT purchasers for infringement if they put the images in their profile pictures, on OpenSea listings, etc. This is not the intent of the creators and purchasers, and we hope that courts would not cooperate in a copyright-based attack like this, but without clarity around the copyright rights of NFT owners, there is a risk that it could happen. (The courts are not known for their nuanced understanding of cutting-edge blockchain technologies and community norms.)

Following the initial CryptoPunks launch, its creator, Larva Labs, later went back and tried to retroactively add a copyright license. Some legal scholars are skeptical whether this works. Even more recently Yuga Labs acquired the rights to the CryptoPunks and announced its intention to grant commercial rights to token owners. While many CryptoPunks owners will welcome this change, changing li-

cense terms after the initial launch and minting is trickier than granting them up front.

Even more blatantly, some NFTs create copyright trouble by using artworks stolen from artists, or famous works that the NFT creators have no connection with and no license from. Copying these works as part of the NFT marketing (e.g. for OpenSea listings) can be copyright infringement. In addition, an NFT creator could be engaged in false advertising by implying that NFT owners will receive rights in these stolen works. Indeed, because copyright infringement is "strict liability," NFT owners who make copies of stolen art could also be liable for infringement, even if they were misled by the NFT creator into thinking that it was properly licensed.

While straight-up scam artists are unlikely to care about infringement, it is unfortunate that many well-meaning projects also seem to believe that minting an NFT of a work somehow automatically brings with it some copyright interest in the work. One particularly tragic example is Andy Williams, who created an NFT of TV video footage depicting his daughter's murder. Parker was apparently advised that creating an NFT would give him enough of a copyright in the footage to have it removed from sites like Facebook and YouTube. But copyright doesn't work that way. The television station that filmed the footage owns the copyright. Parker can't change that by minting an NFT of it.

Another related myth of NFTs is that minting one helps enforce copyright against infringers. The Associated Press's Director of Blockchain, for example, has argued that creating NFTs of some of its photographs would make it easier to make unauthorized users take them down. But copyrights come from copyright law, not from the blockchain. The process to file a copyright lawsuit or a DMCA takedown notice isn't made any easier by having an NFT of the work. To be sure, in a Web3 future where absolutely everything is on the blockchain and nothing is possible unless it is approved by a blockchain transaction, it would be technically impossible to post a photograph without an explicit license from the copyright owner. But first, that world is not the world of today, and second, a world where speech is impossible without advance permission would be profoundly dystopian. It would run completely against the values of freedom and openness that blockchain is supposed to stand for.

Copyright Transfers Are Hard

Actually ensuring that NFT owners have the copyrights they think they do is also a subtler problem than it appears. Consider the following passage from the Bored Ape Yacht Club Terms & Conditions:

> i. You Own the NFT. Each Bored Ape is an NFT on the Ethereum blockchain. When you purchase an NFT, you own the underlying Bored Ape, the Art, completely.

This looks like it tethers ownership of the copyright to ownership of the NFT. Suppose that Woodchuck Labs uses these terms for its WoodChuckers NFTs. When Alice buys a WoodChucker NFT, she acquires the copyright. When she sells the NFT to Bob, he acquires the copyright. In copyright terms, there is a transfer of copyright ownership to Alice when she buys the NFT, and then another one from Alice to Bob when she sells him the NFT. Full ownership of the copyright lets Alice use the artwork—e.g. to display in her Twitter profile inside a hexagon. It also lets her sue for infringement, if she wishes, any right-clickers who download and display the artwork.

Unfortunately, copyright doesn't work this way. The problem is that U.S. copyright law sets a high threshold for what it takes to transfer ownership of a copyright. Section 204(a) of the Copyright Act states:

> A transfer of copyright ownership, other than by operation of law, is not valid unless an instrument of conveyance, or a note or memorandum of the transfer, is *in writing and signed by the owner* of the rights conveyed or such owner's duly authorized agent. (emphasis added)

The writing part isn't a problem; the terms on the website count as a "writing" under federal law. The bigger issue is that the BAYC terms aren't "signed" by the copyright owner, Woodchuck Labs. Without a signature, it is not possible to pass ownership of the copyright to Alice.

In theory, Woodchuck Labs could fix the lack of a signature at this first step by modifying the terms to add a signature line. Under the E-SIGN Act, even a digital signature like a person's name printed in a script typeface can be a "process, attached to or logically associated with a contract or other record and executed or adopted by a person with the intent to sign the record" (emphasis added). Indeed, courts have held that clicking "I agree" to a website's terms when you create an account is enough to show the "intent to sign." (There are some timing issues, like the ones with the CryptoPunks relicensing, but that's an issue for another time.)

Unfortunately, there is a bigger problem. It arises when Alice decides to resell the WoodChucker to Bob. The intent of the BAYC terms is that Bob now owns the copyright and Alice doesn't. But thanks to the signature requirement, that's not what happens. There is no signed transfer of copyright from Alice to Bob. Without a signed transfer, Alice still owns the copyright, not Bob.

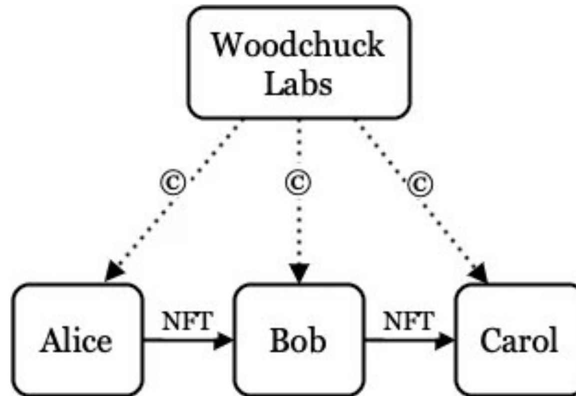
This is where the difference between smart contracts and legal contracts rears its head. Bob might try to argue that Alice has agreed to the BAYC terms, which make him the copyright owner. But Alice hasn't! As far as she's concerned, the BAYC terms are just a bunch of words on a website somewhere. The copyright system wasn't created with digital tokens in mind and doesn't know about them. As far as it's concerned, Alice owns the Woodchucker copyright and hasn't done anything to give up ownership. Legal contracts typically only bind the people who explicitly agree to them.

True, Alice has has invoked an ERC-721 smart contract `transferFrom()` method to transfer the Bored Ape to Bob and applied her cryptographic signature to the smart contract transferring the NFT to Bob. But that method is a smart-contract term, not a legal-contract term. The smart contract doesn't say anything about copyright or link to the BAYC terms. Even if it did, there would be no guarantee that Alice had read or even knew about those terms. She would not have attached her cryptographic signature to a transaction "attached to or logically associated with a [legal] contract ... with the intent to sign" it in a legally binding sense.

Getting from smart contract to legally binding terms is a hard and subtle problem. Adding off-chain assets like tungsten cubes and copyrights into the mix makes it even harder. Changing ownership of these assets requires off-chain effects, but since smart contracts exist on the blockchain, it is entirely possible to interact with them without ever invoking any additional contract terms. If copyright in an NFT-linked artwork is based on a legal contract, users who deal only with the smart contract have a decent argument that nothing in the legal contract applies to them, since they only interacted with the smart contract.

An Alternative: Copyright Licenses

There is another way to structure NFT copyrights that avoids the signed-writing problem. Instead of a transfer of ownership, which passes full copyright ownership to each owner of the NFT, the NFT creator could use a copyright license. The creator holds on to ownership of the copyright, and gives a license directly to each successive NFT owner.



At first glance, this looks more complicated, because now the creator must deal directly with every NFT owner, rather than just with the first owner. But it has a substantial advantage, which is that copyright licenses don't need to be signed the way that copyright transfers do. (Indeed, they don't even need to be in writing, although for any economically serious transactions, writing down the terms is much safer.) Carol and Woodchuck Labs don't need to rely on Alice and Bob to get the signed transfers right. Instead, Woodchuck Labs can simply write its terms so that it directly gives a license to every NFT owner automatically as soon as they acquire the NFT.

There are good precedents for this approach in free and open-source software licensing. The GNU General Public License, for example, says:

> Each time you convey a covered work, the recipient automatically receives a license from the original licensors, to run, modify and propagate that work, subject to this License. You are not responsible for enforcing compliance by third parties with this License.

And the Creative Commons Attribution license says:

> Every recipient of the Licensed Material automatically receives an offer from the Licensor to exercise the Licensed Rights under the terms and conditions of this Public License.

A clear example of this approach in the NFT space is the RTFKT license, which states:

> 1 … Any digital works of authorship or other content made available through the Platform to an owner of a Digital Collectible that is intended as an "Additional Benefit" (as that term is defined in the Digital Collectible Terms) will be identified as such on the Platform or at the time of download. Any such content will be licensed to you for as long as you own the applicable Digital Collectible pursuant to the terms of any license presented at the time of download or, if no such terms are presented, pursuant to the applicable Digital Collectible Terms as Related Content for that particular Digital Collectible. …

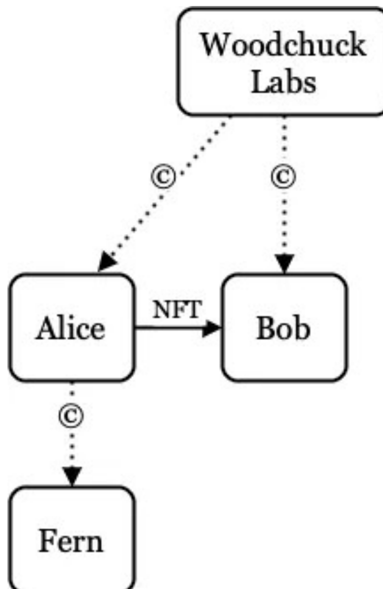
The details are tricky, and this is not meant to be a full legal analysis. Our points are just that NFT creators need to give serious thought to how they structure their terms to ensure that NFT owners actually receive the necessary copyright licenses to NFT-linked artwork, and that copyright licensing is far easier to make work than an outright transfer of ownership.

Derivative Rights

Another difficult issue concerns derivative works—i.e., "a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which a work may be recast, transformed, or adapted." No one can explain why the Bored Apes achieved cultural and economic escape velocity. It will forever be one of the mysteries of the ages. But one factor, at least, is sometimes said to be that the Bored Apes terms allow owners to make more extensive derivative works based on them. While the NFT license allows owners to use the art for their own "personal, non-commercial use" and for projects earning up to $100,000 per year, the BAYC terms allow unrestricted commercial use of the artworks. Specifically, "Yuga Labs LLC grants you an unlimited, worldwide license to use, copy, and display the purchased Art for the purpose of creating derivative works based upon the Art."

The first problem here is that this license grant is inconsistent with the statement, just two paragraphs above in the BAYC terms, that "[w]hen you purchase an NFT, you own the underlying Bored Ape, the Art, completely." If Alice really does "own" the art "completely," then Yuga Labs has nothing left to give and the commercial-use license is superfluous. (This is another sign that the statement that Bored Ape NFT owners "own" the artwork, like many other claims about what users actually own when they "buy" content online, cannot be taken at face value.)

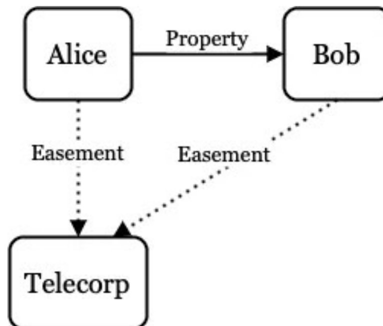
The second problem is that this term does not play nicely with downstream transfers. Consider Alice and Woodchuck Labs again. Suppose that Alice owns WoodChucker number 12345. She allows Fern, a a filmmaker, to create a video series based on WoodChucker 12345. Those videos are derivative works under copyright law and Fern has their own copyright in the videos. Now Alice decides to sell WoodChucker number 12345 to Bob. What should happen to Fern's license?



One simple answer would be that since Alice's copyright license to use WoodChucker 12345 terminates when her ownership of the NFT does, so do any sublicenses she has granted. That would mean that the videos stop being licensed the moment Alice sells to Bob, and Fern becomes a copyright infringer if they continue to show the videos! This is terrible from Fern's perspective, having invested time and money into making the videos. It's also terrible from Alice's perspective, because Fern ought to be very reluctant to pay money to license Alice's rights if Alice can always back out by selling the NFT to Bob. So this solution effectively makes the derivative work rights unmarketable.

Another answer would be that Fern's license continues in full force. Bob has no ability to call backsies once Alice has given Fern a license. This protects Fern, and thereby protects Alice's licensing business. But it creates its own headaches. For example, Bob might grant his own video license to Georg, so now there are two competing WoodChucker 12345 series. Fern will be furious, but what can they do about it? If their lawyers are good, Fern will have insisted that Alice make the video license exclusive, so that Alice can no longer license anyone else to make a video series. But this is a private contract between Alice and Fern. Bob didn't sign it, and isn't bound by it! Bob got his copyright license directly from WoodChucker Labs, without the exclusivity limitation that Alice promised to Fern.

So maybe the licenses ought to travel with the NFT itself. This happens all the time with real estate. If Alice owns a parcel of land and grants Telecorp an "easement" to run a fiber-optic cable under one corner of her land, the easement will still exist after Alice sells the land to Bob. It is said to "bind the owner's successors" or to "run with the land." That is the easement is attached to and limits (or "burdens") the legal right to the land itself. It is not just a personal promise from Alice to Telecorp. When Bob buys the land from Alice, he steps into her shoes. He succeeds not just to her rights in the land (e.g. to build a house or grow crops), but also to her obligations (to allow Telecorp to continue to operate the cable).



Similarly, we could imagine that when Bob buys the NFT from Alice, he steps into her shoes. He succeeds not just to Alice's rights in the Bored Ape 12345 copyright (such as the right to make glossy art prints), but also to any limitations or obligations Alice has taken on (such as the exclusive video license to Fern). Now Bob is not free to to license Georg to make a second video series.

Perhaps this is a good solution. Or perhaps not. If Alice is now free to encumber the artwork copyright in this way, it limits Bob's rights. When he buys the NFT, he buys something less than the full rights Alice bought. Alice has carved up the copyright, and in effect kept a slice for herself. If Bob is in the NFT market, he will have to investigate the entire chain of title of the NFT he is buying to make sure that no Alices before him have quietly given away part of the copyright. This

need for investigation runs contrary to the crypto ethos that as much as possible should be done in public and on-chain. So maybe exclusive licenses entered into by one owner should not run with the NFT.

So far, we have enumerated three different possibilities for what happens when Alice sells the NFT to Bob:

1. Fern's license terminates.
2. Fern's license continues, but Bob can license the same rights to Georg.
3. Fern's license continues, and Bob cannot license the same rights to Georg.

It is possible to imagine a court adopting any of these three outcomes. Indeed, there is no clear consensus as to which of these is the best solution in general. (The three authors of this blog post don't even agree!) Even worse, these don't even exhaust the possibilities. A fourth possibility is that Fern's license to create new derivative works terminates, but that they can continue to use existing derivative works they have already created. This is how copyright law deals with some license terminations. Or if Fern's license continues, what should happen to any royalties that Fern has promised to pay Alice? Should Bob succeed to those too? There are arguments in favor, and arguments against.

Our point is just that these are issues that an NFT license allowing derivative works needs to deal with. Otherwise, NFT owners and their business partners may be unpleasantly surprised by the results. Everyone who does a project based on an NFT that does not answer these questions is putting an immense amount of faith in the courts to get things right if the deal goes sour and the parties end up suing each other. (And blockchain advocates are not generally known for their faith in the courts to get things right.)

We are not saying that there is a best solution for all projects. (This is one reason among many that we are not providing our own proposed license text.) What is right for the Bored Apes may not be right for an NFT project based around musical works, or on a literary work. Instead, we think that NFT creators need to think about these issues, discuss them with their communities, and then communicate clearly how copyright licenses will work in relation to the NFTs.

Conclusion

It is clear that many NFT projects are designed to transfer copyrights along with ownership of the NFT itself. This is a core design goal, on the same level as creating compelling content and making transfers irrevocable. Despite this, many projects seem to have put far less thought into the legal aspects of their designs than into the technical and artistic ones.

We think this is a major mistake. The legal infrastructure on which blockchains run is just as complicated and full of traps for the unwary as the technical infrastructure. While some cryptocurrency and Web3 projects are intended to escape the existing legal system, or to replace it entirely, the same is not true of many creative NFT projects. They are intended to work within the legal system as it currently exists, to allow people to create new and interesting art now, and to commercialize it using real-world contract, property, and copyright law.

Some existing NFT licenses are not fit for purpose. They do not make copyright interests travel along with the NFTs in the way that they intend. If code is law, then these licenses are buggy code. Responsible NFT creators would not launch a project built atop a smart-contract library that had known unpatched vulnerabilities. They should bring the same care to the legal code on which they depend, because otherwise the results could be just as catastrophic. Whatever you think

about NFTs, launching them with broken copyright licenses doesn't do anyone any good.

NOTES AND QUESTIONS

1. Is this a blockchain-specific concern, or will any system to memorialize copyright licenses run into the same difficulties? Is this a copyright-specific concern, or will any attempt to tether off-chain assets to on-chain records run into the same difficulties?
2. For an NFT copyright license that attempts to take the concerns raised in this essay seriously, see *The Token-Bound NFT License* (IC3 and COALA 2022), https://eips.ethereum.org/assets/eip-5218/ic3license/ic3license.pdf.