THE LAW OF USING THIRD PARTY CONTENT AND DATA FOR TRAINING ALGORITHMS AND GENERATIVE AI

Ian Ballon, JD, LLM, CIPP/US
Co-Chair, Global IP & Technology Practice Group
Greenberg Traurig LLP
(650) 289-7881   (310) 586-6575   (202) 331-3138
Ballon@GTLaw.com
Threads, Facebook, LinkedIn, X (formerly Twitter): Ian Ballon
www.IanBallon.net

E-Commerce & Internet Law

Ian C. Ballon

E-Commerce & Internet Law: Treatise with Forms - 2d Edition

2022 UPDATES OUT SOON!

The revised and updated edition of this comprehensive work will provide you with a complete legal authority on e-commerce and Internet law, covering business-to-business and business-to-customer issues, regulatory issues, and emerging trends. It includes practice tips and forms and its unique organization facilitates finding quick answers to your questions. This valuable resource on Internet and e-commerce issues contains nearly 10,000 detailed footnotes, plus references to hundreds of unpublished court decisions, many of which are not available anywhere else.

BUY NOW

E-COMMERCE & INTERNET LAW: TREATISE WITH FORMS 2D - 2022
Download File

Use of third party data / Web scraping and AI

- Using content and data sets to train algorithms for ML/AI/Generative AI
    - Your own content or data
    - Content or data freely available for use
    - Content or data licensed for training
    - Third party content or data that may be accessible but not freely available
        - The owner may claim proprietary rights (under IP or other laws)
        - Third party data may be incomplete (due to privacy opt-out laws))
- Ethical Issues
- Potential regulation in the U.S. and E.U.
    - Compare: Japan
- Will AI put independent artists out of business? AI is only as good as the test set data used to train the algorithms?

THE FUTURE OF COPYRIGHT: AI ART AND MUSIC AND COPYRIGHT FAIR USE

TECH   04/26/2023

The Real Problem With Fake Drake

"Heart on My Sleeve" shows that generative artificial intelligence can crea
songs. But the real issue is how fast it will make them.

BY ROBERT LEVINE

Ian Ballon - Internet Law

BA Who is Ian ballon?

Ian Ballon is an American attorney, author, and speaker who specializes in internet, mobile, and intellectual property law. He is a partner at Greenberg Traurig LLP, a global law firm with over 2200 lawyers in 40+ offices around the world, where he co-chairs the firm's Global Intellectual Property & Technology Practice Group and its Blockchain & Digital Assets Group.

Ian Ballon is widely recognized as an expert in his field and has been honored with numerous awards and recognitions. He has authored several books, including "The Complete CAN-SPAM Act Handbook," "E-Commerce and Internet Law: Treatise with Forms 2nd edition," and "The Blockchain and Cryptocurrency

7

Ian Ballon is an American attorney, author, and speaker who specializes in internet, mobile, and intellectual property law. He is a partner at Greenberg Traurig LLP, a global law firm with over 2200 lawyers in 40+ offices around the world, where he co-chairs the firm's Global Intellectual Property & Technology Practice Group and its Blockchain & Digital Assets Group.

Ian Ballon is widely recognized as an expert in his field and has been honored with numerous awards and recognitions. He has authored several books, including "The Complete CAN-SPAM Act Handbook," "E-Commerce and Internet Law: Treatise with Forms 2nd edition," and "The Blockchain and Cryptocurrency Regulation 2021." He is also a frequent speaker and panelist at legal and technology conferences, including the International

Ian Ballon
Dec 9, 2022 · 🌐

👍❤️ You, Patricia Gardner and 65 others   22 comments

👍 Like   💬 Comment   ➤ Share

AI -- Copyrightability and Patent Law

- Machines can't obtain patents

- Machines can't create works
  - Copyright Office position

  - *Naruto v. Slater*, 888 F.3d 418, 426 (9th Cir. 2018) (holding that "animals other than humans . . . lack statutory standing to sue under the Copyright Act.")

- Can the output of generative AI result in liability? (*i.e.,* can "works" created by machines be infringing or a fair use?)
  - Look at the algorithm and the content or data used to train it
  - How many photos/songs/other creative works are used to train the algorithm
  - Does the algorithm replicate a specific creator's style?
  - What if the algorithm is so good that it independently creates a work that appears to be infringing?

ARTIFICIAL INTELLIGENCE, SCREEN SCRAPING AND DATA PORTABILITY

AI/ Screen Scraping/ Data Portability

- Contract/TOU/PP restrictions
  - *Meta Platforms, Inc. v. BrandTotal Ltd.*, _ F. Supp. 3d _, 2022 WL 1990225 (N.D. Cal. 2022) (automated access violated TOU)
- Copyright protection (statutory damages and potentially attorneys' fees if a work is timely registered)
  - Facts vs creative expression
    - *Feist Publications, Inc. v. Rural Telephone Service Co.*, 499 U.S. 340, 350 (1991)
  - Protection for compilations if originality in the selection, arrangement or organization of a database (but thin protection)
  - Data mining as a transformative fair use: *Author's Guild, Inc. v. HathiTrust,* 755 F.3d 87 (2d Cir. 2014)
  - VHT, Inc. v. Zillow Group, Inc., 918 F.3d 723 (9th Cir. 2019) (search function not a fair use)
- Common law claims, such as misappropriation to the extent not preempted by 17 U.S.C. § 301
  - *International News Service v. Associated Press,* 248 U.S. 215 (1918)
  - *National Basketball Ass'n v. Motorola, Inc.,* 105 F.3d 841 (2d Cir. 1997)
- Interference with contract or prospective economic advantage
- Unfair competition
- Trespass and Conversion
  - trespass to chattels may be based on unauthorized access (plus damage)
    - *Intel Corp. v. Hamidi,* 30 Cal. 4th 1342, 1 Cal. Rptr. 3d 32 (2003)
  - conversion usually requires a showing of dispossession or at least substantial interference
- Computer Fraud and Abuse Act – Federal anti-trespass computer crimes statute
  - Must establish $5,000 in damages to sue
  - Exceeding authorized access may not be based on use (vs. access) restrictions: *Van Buren v. United States*, 141 S. Ct. 1648 (2021)
  - *hiQ Labs, Inc. v. LinkedIn Corp.*, 31 F.4th 1180 (9th Cir. 2022) (affirming an injunction prohibiting LinkedIn from blocking hiQ's access, copying or use of public profiles on LinkedIn's website (information which LinkedIn members had designated as public) or blocking or putting in place technical or legal mechanisms to block hiQ's access to these public profiles, in response to LinkedIn's C&D letter)
- Anti-circumvention provisions of the DMCA, 17 U.S.C. §§ 1201 *et seq.*
- Removing, altering or falsifying copyright management information (CMI) - 17 U.S.C. § 1202
- California BOT Law - Cal. Bus. & Prof. Code §§ 17940 *et seq.* prohibits the undisclosed use of bots to communicate or interact with a person in California online, with the intent to mislead the other person about the artificial identity of the bot, to incentivize a purchase or sale of goods or services in a commercial transaction or to influence a vote in an election

- ## Direct Liability
  - If you directly scrape or otherwise copy third party data you could be held liable under the theories noted on the prior slide
- ## Secondary Liability
  - Secondary liability may arise if you pay a third party to access the data or acquire data that has been obtained in breach of an agreement or violation of law
  - Secondary liability theories could be used to seek to impose individual liability, regardless of the corporate form
  - Secondary liability exists under IP laws and to a lesser extent under other laws but may be harder to establish absent strong documentary evidence (emails, text messages, slack), especially if scraping is done offshore
    - Contibutory copyright liability
    - Vicarious copyright liability
    - Inducing copyright liability
    - Secondary liability under the anti-circumvention provisions of the Digital Millennium Copyright Act
    - No secondary liability for breach of contract (but potentially interference with contract)
    - Potential direct liability for unfair competition
    - In extreme cases, fraud

COPYRIGHT
FAIR USE AND THE *WARHOL* CASE

Copyright Fair Use

- Multipart balancing test available when a work is used "for purposes such as criticism, comment, news reporting, teaching . . . Scholarship or research"
  - Courts must consider:
    - The purpose and character of the use, including whether it is of a commercial nature or is for nonprofit educational purposes;
      - Commercial
      - Transformative
    - The nature of the work (creative works are closer to the core of intended copyright protection than informational or functional works)
    - The amount and substantiality of the portion used in related to the copyrighted work as a whole
    - The effect of the use upon the potential market for or value of the copyrighted work
  - Courts may consider other criteria
  - VCR recordings
    - *Sony Corp. of America v. Universal City Studios, Inc.*, 464 U.S. 417 (1984)
  - For security research: *Apple Inc. v. Corellium, LLC,* 510 F. Supp. 3d 1269, 1285-92 (S.D. Fla. 2020)
  - Data mining/ Google books
    - *Author's Guild, Inc. v. HathiTrust,* 755 F.3d 87 (2d Cir. 2014)
    - *Author's Guild, Inc. v. Google Inc.,* 804 F.3d 202 (2d Cir. 2015), *cert. denied,* 136 S. Ct. 1658 (2016)
  - Use in connection with criticism
    - *Katz v. Google, Inc.,* 802 F.3d 1178 (11th Cir. 2015)
- *Google LLC v. Oracle America, Inc.*, 141 S. Ct. 1183 (2021) (6-2) (Breyer)
  - Google's reimplementation of 37 of 166 of Java SE application programming interfaces (APIs) in the Android mobile operating system was a fair use
  - Declined to address software copyrightability but provided some guidance
- *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith,* 143 S. Ct. 1258 (2023)
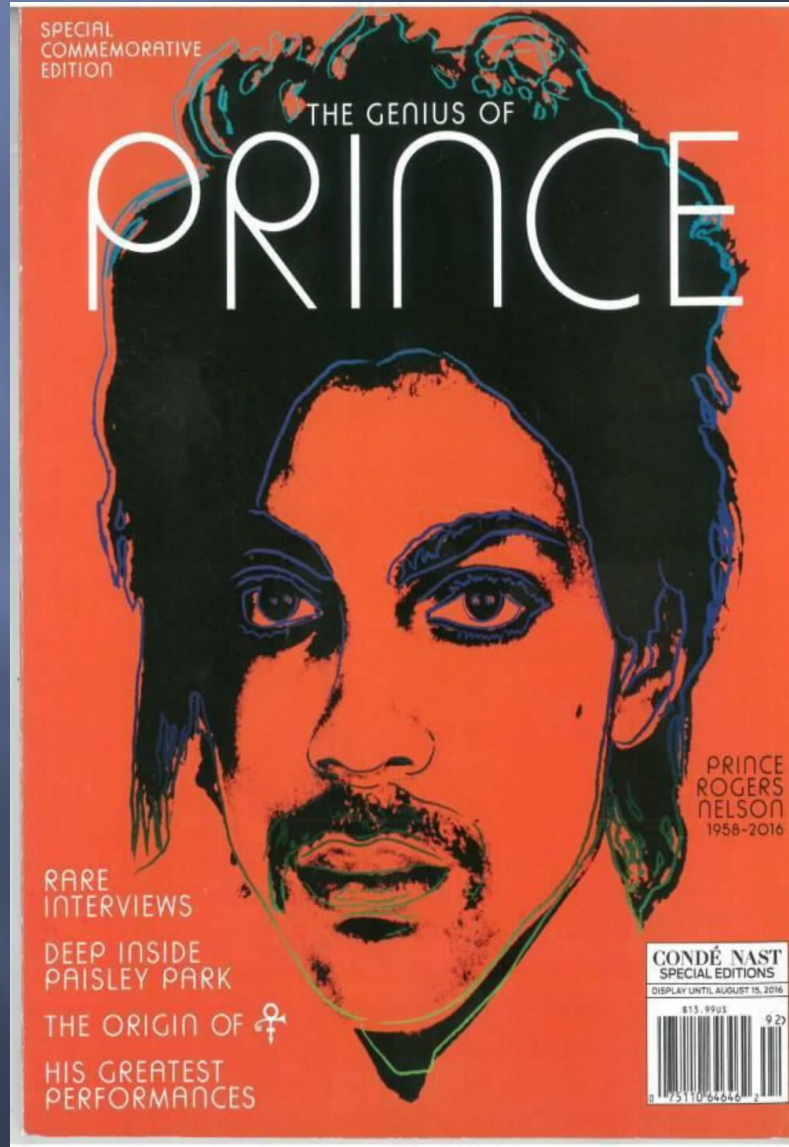
Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith, 143 S. Ct. 1258 (2023) (7-2) (Sotomayor)

- The purpose and character of Warhol's use of Goldsmith's photo *in commercially licensing Orange Prince to Conde Nast* was not a fair use
  - The court only addressed the first factor – not whether the use was fair overall
    - The central question is whether a use merely supersedes the original creation (supplanting the original) or adds something new, with a further purpose or different character (purpose & character judged by an objective inquiry)
    - NEW: As most copying has some further purpose and many secondary uses add something new, the first factor asks whether *and to what extent* the secondary use has a purpose or character different from the original. The larger the difference, the more likely the use is fair.
    - Transformativeness is a matter of degree – to preserve the copyright holder's right to prepare derivative works the degree of transformation must go beyond that required to qualify as a derivative work
    - Stated differently, if an original work and secondary use share the same or highly similar purposes, and the secondary use is commercial, the first factor is likely to weigh against fair use absent some other justification for copying
    - The purpose the court focused on was use of the image to illustrate a magazine article, not the painting itself. Even assuming that Warhol's purpose was to portray Prince as iconic, that difference was not significant enough for purposes of using one work or the other to illustrate a magazine article
      - Likewise Warhol's purpose of commenting on the dehumanizing nature of celebrity was not substantial enough as it was not focused specifically on the Goldstein photo that was used (as opposed to any image of Prince) (analogy to parody)
      - Because the use was commercial, a more substantial justification was required
  - The majority went to great lengths to limit its holding to the facts of the case – competitive commercial licensing, emphasizing that other uses of the Goldstein photo for Orange Prince (such as to display in a museum) could be fair
  - Nevertheless, the decision seems to import the fourth factor – impact on the market – as relevant to the first factor, much in the same way that Justice Breyer in *Google* found transformativeness to be relevant to all four factors.
  - The creative nature of the works – and their competitive use for magazine cover licensing – greatly impacted the decision
  - But if an Andy Warhol painting is not fair use, what is?
    - The decision seems to elevate visual impression over other aspects of whether a secondary use has a further purpose or different character than the original, which is "a matter of degree" (*see* Kagan dissent)
    - The degree of difference must be weighed against other considerations, like whether the use is commercial
    - New expression, meaning or message may be relevant, but is not, without more, dispositive
- Gorsuch (joined by Jackson) concurred (examine the purpose of the particular use challenged, not the artistic purpose of the underlying use)
- Kagan (joined by Chief Justice Roberts) dissented (sharp departure from *Campbell* and *Google*; this opinion will stifle creativity because a license is not always available)

COPYRIGHT DAMAGES AND THE BENEFITS OF TIMELY REGISTRATION

- Copyright owners may elect actual or statutory damages at any time prior to a jury verdict

  - The amount of damages is determined by the jury if a jury trial is selected

- Statutory damages (1 award per work infringed):

  - Usual range: $750-$30,000

  - Increased to $150,000 if plaintiff proves willfulness

  - Decreased to $200 if the defendant proves innocence

- Actual Damages:

  - Actual damages suffered as a result of the infringement and, to the extent not duplicative,

  - Defendant's wrongful profits attributable to the infringement

    - May include indirect (or noninfringing) profits attributable to the infringement.

- Timely Registration:

  - Statutory damages and attorneys fees are not recoverable if a plaintiff failed to timely register its work (but actual damages and injunctive relief may be available)

    - A registration certificate is deemed sufficient even if it contains inaccurate information unless (a) the inaccurate information was included on the application with knowledge that it was inaccurate, and (b) the inaccuracy, if known, would have caused the Registrar of Copyrights to refuse registration. *Unicolors, Inc. v. H&M Hennes & Mauritz, LP*, 142 S. Ct. 941 (2022)

- Timing – Damages for 3 years prior to filing suit

  - *Petrella v. Metro-Goldwyn-Mayer, Inc*., 572 U.S. 663 (2014).

  - Except where the discovery rule applies: *Starz Entertainment, LLC v. MGM Domestic Television Distribution, LLC,* 39 F.4th 1236 (9th Cir. 2022); *Nealy v. Warner Chappell Music, Inc*., _ F.4th _, 2023 WL 2230267 (11th Cir. Feb. 27, 2023)

- Attorneys' fees:

  - Reasonable attorneys' fees, where a copyright has been timely registered, may be awarded to the prevailing party as part of the costs of a case; the decision to award fees is in the sound discretion of the court

  - *Fogerty v. Fantasy, Inc*., 510 U.S. 517, 543 n.19 (1994)

    - Frivolousness
    - Motivation
    - Objective unreasonableness (both in the factual and legal components of a case)
    - The need in particular circumstances to advance considerations of compensation and deterrence

  - *Kirtsaeng v. John Wiley & Sons, Inc*., 136 S. Ct. 1979 (2016)

    - A court should give substantial weight to the objective reasonableness of the losing party's position (while an important factor it is not controlling)
    - A district court may not award fees to a prevailing plaintiff as a matter or course
    - A district court may not treat prevailing plaintiffs and prevailing defendants differently (both should be encouraged to litigate meritorious claims or defenses)
    - A court must look at the totality of the circumstances of a case

Practical Rules of Thumb

- Legal analysis. Ask:
  - What was copied?
  - How was it accessed?
  - How was it used?
  - How long will it be retained?
- Fair Use. Ask:
  - How much was copied?
  - Is the material factual/ functional or artistic/ highly creative?
  - What is it being used for (to train competitive algorithms? For a commercial purpose? For research or scholarship?)
  - Was an intermediate copy made?
    - If so, how long will it be retained?
- Practical business considerations

THE LAW OF USING THIRD PARTY CONTENT AND DATA FOR TRAINING ALGORITHMS AND GENERATIVE AI

Ian Ballon, JD, LLM, CIPP/US
Co-Chair, Global IP & Technology Practice Group
Greenberg Traurig LLP
(650) 289-7881   (310) 586-6575   (202) 331-3138
Ballon@GTLaw.com
Threads, Facebook, LinkedIn, X (formerly Twitter): Ian Ballon
www.IanBallon.net