

Generative AI: The Latest Disruptive Technology for Copyright

Pamela Samuelson, Berkeley Law School
AI & the Future of Copyright Politics Conference
Keynote Talk
November 7, 2025

OVERVIEW

- Lessons from prior disruptive technology cases for (c) & their politics
 - Photocopying, home taping, Internet, mass digitization
- How generative AI (genAI) is similar to & different from those disruptive technologies
- How courts have assessed the use of in-(c) works as training data
- Politics of genAI
- Prospects for collective licensing as a compromise solution
- Concluding thoughts

TECHNOLOGY THREATS TO (C) ARE NOT NEW

- “Copyright today is under stress. It is also under attack.
- “The stress arises primarily from the rapid succession of marvelous new machines, miraculous enough to delight Jules Verne, for copying and, in new ways, enjoying protected works. They are for that reason a boon.
- “By the same token, unauthorized copying of print, sound, & cinematographic works has become epidemic; copyright owners’ control has weakened & the domestic & international traffic in piratical copies of films & records has risen to floodtides.”
- “If the source of the stress lies in rapidly changing technologies & if the political force of the attack draws from commercial & noncommercial self-interest, **the assault weapon is the notion that some proof of ‘harm’ is a necessary condition for protecting a class of creative industries”**

WHAT TECHNOLOGIES POSED THREATS?

- Quote is from David Ladd, then Register of Copyrights, in *The Harm of the Concept of Harm in Copyright Law*, J. Cop. Soc'y (1982)
- What was he worried about?
 - Unlicensed photocopying of journal articles by libraries for patrons
 - Video tape recording machines (VTRs) that allowed consumers to copy programs off broadcast TV
 - Home audio taping of sound recordings
- Some said these copies were fair uses because not harming any (c) markets
- Ladd thought that these technologies were causing considerable harm & warned they would have devastating impacts for (c) industries
 - Assumption that if large number of copies are being made, there must be harm to (c) markets
 - Sound familiar? Echoes in the generative AI cases?

COPYING FOR RESEARCH & EDUCATION

- Photocopying for research
 - *Williams & Wilkins v. U.S.*: fair use for NIH librarians to make photocopies of W&W's journal articles for researchers
 - *American Geophysical Union v. Texaco*: unfair for commercial firm's research scientists to copy journal articles because CCC "cured" the market failure by offering licenses
- Photocopying & digitizing works for education purposes
 - *Princeton U Press v. MDS*: unfair for copy shop to prepare coursepacks of copyrighted materials for students to buy
 - *Cambridge U Press v. Patton*: publishers only partly successful in challenge to online postings of book chapters by GSU professors

POLITICS OF PHOTOCOPYING

- During legislative history of 76 Act, stakeholders were at “war”
 - Educators/librarians wanted a **broad exemption** for educational & research uses of works
 - Publishers/author groups wanted **every copy to be licensed**
 - After *W&W* decision, Congress created an exemption for library research uses, but left other educational/research uses to vagaries of **fair use**
 - CONTU study in late 1970s concluded no present need for licensing mandate
 - Yet hope expressed that voluntary private licensing would be enabled
 - CCC became licensing agent for commercial firm photocopying & coursepacks
 - Post-*Texaco* concerns that publishers would sue educators & courts would require them to license photocopying, but this did not occur

HOME TAPING OF TV PROGRAMS & MUSIC

- *Sony v. Universal*: contributory infringement claim vs. Sony for materially contributing to consumers' infringements of movies shown on broadcast TV
 - Goal of lawsuit: outlaw technology & require significant redesign to thwart infringement
 - 5-4 SCT decision that most popular use of VTRs--to make time-shift copies of programs--was fair use, which meant the tech had substantial non-infringing uses, so Universal's challenge failed
- Moral panic when DAT machines & tapes about to enter U.S. market
 - *Cahn v. Sony*: goal was to outlaw DAT machines because most frequent use would be to make permanent copies of recorded music, not space-shifting
 - Audio Home Recording Act was compromise to settle *Cahn*

POLITICS OF HOME TAPING

- While *Sony* was pending, bills in Congress either to exempt home taping or create compulsory licenses, but Congress decided to wait to see what Congress did
- OTA study showed that 4 in 10 Americans made home tapes of music
 - As much as RIAA wanted to ban all private copies, this was implausible
- AHRA created noncommercial copy exemption
 - Also created a levy (invisible to DAT users) to establish fund so copyright owners could be compensated for private uses
 - But computer industry lobbied for exclusion of hard drives, so tech bypassed law (*RIAA v. Diamond Multimedia*)

POLITICS OF INTERNET USES

- 1995 Clinton White Paper: all OSPs are strictly liable for user infringements
- WIPO (c) Treaty Agreed Statement: providing facilities for communication not itself a communication to the public
- “War” between (c) industries & tech/telcos -> 1998 DMCA “grand bargain”
 - Copyright industries got very broad regulations of tampering with technical protection measures & with copyright management information in §§ 1201-1202
 - But they compromised on safe harbors in § 512
 - Plaintiffs in major cases have expected courts to narrow safe harbors because of tech industry’s contempt for copyright
 - 2020 Copyright Office Report on § 512 highly critical of court rulings & tech industry practices
 - But copyright industry efforts to revise safe harbor rules have not succeeded (yet)

INTERNET USES

- Peer-to-peer file sharing technologies resulted in billions of infringements
 - Napster held secondarily liable because of centralized index & search function
 - Grokster held secondarily liable for inducing users' infringements
 - BitTorrent protocol is still widely used by file-sharers, including by Cox users
- *Viacom v. YouTube*: DMCA safe harbor protected YT because it complied with notice & takedown regime (even though YT officials contemptuous)
- *Cox v. Sony* case before the SCT: is it contributory infringement to continue to provide broadband service to subscribers after Cox received 3 or more notices of p2p file sharing from those accounts?
 - Jury found vs. Cox, awarded \$1 billion in statutory damages (@\$100K for each work)
 - Lower court found Cox did not qualify for § 512(a) safe harbor because it did not have & enforce a reasonable repeat infringer policy
 - Should it matter that overwhelming % of such accounts are institutional (e.g., hospitals) & innocent people will lose Internet access if Cox terminates accounts?

MASS DIGITIZATION

- *Authors Guild v. HathiTrust*: digital library of millions of books held fair use
 - Highly transformative because of different purpose than the original
 - Information locating tool: find data on all books that mention this person or event
 - Providing access to print-disabled was non-transformative but was also fair use
- *Authors Guild v. Google*: database of digital books held fair use
 - Highly transformative to index book contents & to enable snippets of contents relevant to users' search queries: public benefit because provides otherwise unavailable info
- But *Hachette v. Internet Archive*: controlled digital lending of books held not fair use
 - Google Books cases involved mostly non-expressive uses of books cf. digital lending involves consumptive uses

POLITICS OF MASS DIGITIZATION

- Soon after AG sued G for infringement, AG & AAP proposed settlement to create new collective license
 - G would have the right to commercialize all books (except those whose owners opted out) through subscriptions
 - 63% of revenues would go to Book Rights Registry for (c) owners, 37% to G
 - But court disapproved settlement so court had to decide fair use issue
 - By the time courts ruled on fair use in 2010s, everyone was using GBS search
- Copyright Office proposed collective license to achieve GBS goal, but overwhelming majority of responses were negative
- Copyright Office & courts have also been very skeptical of “first sale” or “format-shifting” defenses as in *Hachette* case

LESSONS FROM DISRUPTIVE TECH CASES?

- Market effects factor looms large
- Speculation about harms is likely insufficient (*Sony*, *GBS* cases)
- Large # of copies made does not = harm to markets (same + *W&W*)
- Look for evidence of actual harm or meaningful likelihood of future harm
 - Getty Images: Stability's use is unfair because Getty has an existing market for licensing uses of its images as training data
- Market failure may exist if infeasible for market to form
 - *Sony* & *GBS* cases as examples
 - With class action cases, no way for genAI firms to get licenses from all authors
- Circularity problem well recognized
 - Just because P willing to license does not mean D must license (*W&W*, *CUP v Patton*)
 - Just because some have licensed such uses does not mean D must license (*Sega*)

HOW IS GEN-AI SIMILAR & DIFFERENT?

- Photocopying, home taping, & P2P file sharing enabled ordinary people to make consumptive copies of copyrighted expression
 - These copies posed risks of supplanting demand for the original or arguably interfered with an emerging or likely-to-develop licensing market
- Generative AI uses of works as training data is more like the *Google Books* cases by enabling non-expressive uses of the works
 - Authors Guild argued in both cases that a licensing market was feasible, but courts found the claims too speculative to undermine fair use
 - GenAI defendants argue licensing markets infeasible & no license is needed if use is otherwise fair
 - But “pirated books” issue looms in numerous cases
 - Market dilution theory may test boundaries of infringement

GENERATIVE AI COPYRIGHT LAWSUITS

- 56 lawsuits (so far) in the U.S., mostly focused on training data uses
 - Fair use is main defense in these cases
- Most are class action cases in which small # of named plaintiffs claim to represent a class of persons similarly injured by the genAI defendants
 - *Anderson v. Stability* for visual artists
 - Numerous lawsuits to represent book authors, including *Bartz & Kadrey*
- Several non-class-action lawsuits for different types of works:
 - *Getty Images v. Stability AI* for stock photography
 - *Concord Music v. Anthropic* over music lyrics
 - *UMG v. Suno* over recorded music
 - *NY Times v. Microsoft & OpenAI* + other news cases for training on news stories
 - *Disney v. Midjourney* over movie character outputs

COMPETING VIEWS ON GENAI TRAINING

Plaintiffs' theories:

- Purpose: commercial & non-transformative (exact copies made); bad faith (use of pirated books to train models)
- Nature: creative expression
- Amount: entire works, multiple times
- Harm: outputs compete with our works, lost licensing revenues

Defendants' theories:

- + Purpose: highly transformative because of different purpose
- + Nature: we only care about works as data, not for their expressions; most works posted on Internet
- + Amount: reasonable in light of our transformative purposes
- + Harm: outputs do not infringe or supplant demand for originals or DWs; licensing markets impossible

THOMSON-REUTERS v ROSS INTELLIGENCE

- TR owns Westlaw, a computer database of legal materials, who sued Ross for copying 25,000 WL headnotes (short descriptions of key parts of judicial opinions) to train Ross' AI model
- Ross obtained the WL headnotes from a 3d party subscriber to WL
- Ross' AI does not embody or output WL headnotes, but rather passages from relevant judicial decisions to answer users' legal ?s
- Trial judge initially denied motions to decide the case as a matter of law, saying it needs to go to trial, but then changed his mind
- Ruling for vs Ross' fair use defense, but appeal pending in 3d Circuit
 - Will fair use rulings in *Bartz & Kadrey* influence its decision?

BARTZ & KADREY CASES

- Bartz & Kadrey are book authors & lead plaintiffs who claim to represent classes of authors whose books the genAI defendants copied for purposes of training foundation models
 - Bartz sued Anthropic
 - Kadrey sued Meta
- Anthropic asked Judge Alsup to rule on its fair use defense without going to trial (i.e., no need to go to trial because we agree on facts)
- Kadrey & Meta asked Judge Chhabria to rule on fair use issue
- In June, both judges issued decisions on the fair use issues

BARTZ v ANTHROPIC

Judge Alsup analyzed fairness of 3 uses:

1. Uses of book contents as training data
2. Digitizing purchased books for database of books & storing in database
3. Downloading collections of pirated books & storing them in books database

#1 Training data uses = fair

1. “Quintessentially transformative” because of different purpose (+A)
2. Books = creative works chosen for training because of expressiveness (+B)
3. It’s necessary to copy entire works for A’s transformative purpose (+A)
4. Not supplanting demand for books, so no harm to market (+A)
 - Authors are not entitled to control licensing for this use of their works

MORE ON *BARTZ* DECISION

#2 Digitizing used books & storing digital copies in books database = fair

1. Transformative different purpose than Bartz's purpose (+A)
2. Books are creative works used for expression (+B)
3. Entire works were copied, but this was reasonable in light of Anthropic's purpose (+A)
4. Not a substitute for books; A paid for copies of the books, just format-shifted them from print to digital form (neutral)

#3 Pirated books in database = unfair

- All factors vs fair use (7 million pirated books)
- Rather than go to trial on pirated books issue, Anthropic settled lawsuit for \$1.5 billion for registered books with ISBN #s (est. 500K works)

KADREY v META

1. Purpose: highly transformative + commercial (+M)
 - Use of pirated books does not “move the needle” (i.e., neutral)
2. Nature: books are creative works, chosen for expressiveness (+K)
3. Amount: entire works, but reasonable in light of purpose (+M)
4. Market effects (+M):
 - No on K’s lost sales theory (Meta’s output filters prevent output infringement)
 - No on K’s lost licensing revenues (not a market K is entitled to monopolize)
 - K failed to plead or offer evidence of market dilution (i.e., AI will flood the market with competing works, undermining incentives for human authors)
 - Prediction: most training data uses of works will infringe if evidence of market dilution

COMPARING *BARTZ* & *KADREY*

- Common views on the fair use training data issue
 - Highly transformative purpose, which favors fair use
 - Defendants chose to train on books because of their expressiveness
 - Rejected claim that authors are entitled to control training data licensing market
- Different views about the training data defenses
 - Significance of uses of pirated data
 - Alsup in *Bartz*: definitely NOT fair use
 - Chhabria in *Kadrey*: does not undermine fair use defense
 - Market dilution theory:
 - Alsup in *Bartz*: “science fiction”; no harm unless outputs are substantially similar
 - Chhabria in *Kadrey*: genAI outputs are indirect substitutes for authors’ works so it doesn’t matter that the outputs are not substantially similar

POLITICS OF GEN-AI (C)

- Congress has held some hearings about the genAI (c) controversies, but no bills introduced for or vs.
- Copyright Office Notice of Inquiry attracted more than 10,000 comments
 - Register was fired day after she released a “pre-publication” report on training data fair use claims
 - Bottom line: some training data uses may be fair, but some unfair
 - Report introduced “market dilution” theory (AI outputs will flood markets)
 - Register has said she got fired because of the report’s conclusions
- Trump & tech industry aligned in support of fair use

COLLECTIVE LICENSING AS SOLUTION?

- Voluntary collective licensing
 - Copyright Office, Authors Guild, AAP favor this
- Legislation to create compulsory collective licensing regime
 - US copyright law has done this for certain uses of music + cable & satellite retransmissions of broadcast signals
- Class action settlements
 - *Authors Guild v. Google* settlement tried to accomplish this
 - G would pay \$60 per book & get right to commercialize books through subscriptions
 - 63% of proceeds would go to book registry to pay out to registered right holders
- Court order at remedy phase of successful infringement lawsuits
 - Sweeping injunctive powers but scope is limited to litigants in lawsuits

COMPLICATIONS

- From whom would developers get permission?
 - Everybody in the world who has posted content on the Internet?
 - Incomplete ownership information for Internet contents
- Who (if anyone) owns the rights?
 - Publishers/other commercial entities who got assignments?
 - Authors because training data uses were not in contemplation when they contracted with publishers?
 - Does it depend on contract language?
 - *Bartz & Kadrey*: both judges say authors do not have rights to control such uses
- Lack of collecting society infrastructure (except for music)
 - Copyright Clearance Center & other entities are offering licenses, but it's unlikely they have rights to make these deals at scale needed + don't have data
 - Developers need data, which is why they are doing deals with publishers

MORE COMPLICATIONS

- Antitrust exemption to allow copyright owners to collaborate on prices each industry should charge for uses of their (c)'d works?
- Licensing mandate as entry barrier?
 - Incumbent firms would be happy to have the ladder pulled up behind them
- Intellectual property arbitrage?
- Other proposals for collective licensing have failed
 - Post-Google Books settlement initiative by Copyright Office for books database
 - Noncommercial use levy to allow peer-to-peer filesharing
- Congress is dysfunctional, so even if economically sound model for compulsory license could be found, it's unlikely to happen
 - But could such a model develop, given how unclear it is how much value each input contributes to the general purpose genAI models?
 - Transaction costs would likely swamp payouts

WHAT ABOUT NONPROFIT GENAI RESEARCHERS?

- Fair use calculus will be different if nonprofit researchers use in-(c) works as training for generative AI systems
 - Commercial v. noncommercial distinction is significant in fair use law
 - Scholarship, research & teaching are favored uses
 - Especially if train on fact-intensive works which have “thinner” copyrights
 - Unlikely to harm markets for works’ expressions
- Virtually all of the U.S. cases involve high profile “big tech” defendants
 - “Tech lash” may cut against these companies’ defenses
 - None of the genAI defendants has incentive to push for a distinction between for-profit & non-profit developers
 - When courts render fair use rulings, they won’t be thinking about researchers like you & if rulings are vs fair use & broadly written, you may be at risk

WHAT ABOUT USERS?

- By the time Sony was before SCT, 5 million American households owned VTRs
 - Under Universal's theory, every person who used VTRs to make unauthorized copies of TV programs was an infringer
 - Universal planned to seek statutory damages for Sony's contributions to the infringements
- If OpenAI & others infringed copyrights, are users infringers simply by virtue of using the technology to generate outputs?
 - OpenAI has 180 million registered users; others have millions also
 - Secondary liability claims in many of the genAI cases, so plaintiffs must be claiming that regular users are infringers too
 - Many users prompt for copyrighted content (e.g., Ironman at the supermarket), but are these fair uses? If so, secondary liability claims may fail, as in *Sony*

CONCLUDING THOUGHTS

- Gen AI is not the first disruptive technology to attract (c) lawsuits
 - Challenges failed in cases involving VCRs, MP3 players, & RS-DVRs
 - Because these technologies had substantial non-infringing uses
 - Other challenges succeeded
 - Grokster & Napster, makers of p2p file sharing SW, held liable because they induced users to infringe or contributed to user infringements knowing they were doing so
- As a matter of equity, the copyright plaintiffs have a point
 - GenAI developers have spent huge \$\$\$ to develop these systems & expect to make billions, if not trillions, of \$\$\$ of profits
 - Yet, developers have generally not paid authors/(c) owners any \$ even though their works are the raw material on which genAI systems were built