FIU Law Review

Volume 14
Number 2 *Symposium: Intelligent Entertainment: Shaping Policies on the Algorithmic Generation and Regulation of Creative Works – Florida International University College of Law (Miami, Florida, 2019)*

Article 12

Fall 2020

Can Algorithms Promote Fair Use?

Peter K. Yu
Professor of Law, Professor of Communication, and Director, Center for Law and Intellectual Property, Texas A&M University, peter_yu@msn.com

Online ISSN: 2643-7759

Recommended Citation
Peter K. Yu, *Can Algorithms Promote Fair Use?*, 14 FIU L. Rev. 329 (2020).
DOI: https://dx.doi.org/10.25148/lawrev.14.2.12

CAN ALGORITHMS PROMOTE FAIR USE?

Peter K. Yu*

 Professor of Law, Professor of Communication, and Director, Center for Law and Intellectual Property, Texas A&M University. Earlier versions of this Article were presented at the *FIU Law Review* Symposium on "Intelligent Entertainment: Algorithmic Generation and Regulation of Creative Works" at Florida International University College of Law and the 17th Annual Works-in-Progress Intellectual Property Colloquium at Santa Clara University School of Law. The Author is grateful to the *Review* and Professor Hannibal Travis for their kind invitation to the *FIU Law Review* Symposium. He would also like to thank the symposium participants, in particular Janewa Osei-Tutu, Matthew Sag, Martin Senftleben, and Hannibal Travis, for their valuable comments and suggestions.

I. INTRODUCTION

In the past few years, advances in big data, machine learning,[1] and artificial intelligence ("AI") have generated many questions in the intellectual property field.[2] One question that has attracted growing attention concerns whether algorithms[3] can be better deployed to promote fair use in copyright law. The debate on the feasibility of developing automated fair use systems is not new; it can be traced back to more than a decade ago.[4] Nevertheless, recent technological advances have invited policymakers and commentators to revisit this earlier debate.

As part of the Symposium on "Intelligent Entertainment: Algorithmic Generation and Regulation of Creative Works," this Article examines

1 *See generally* David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. DAVIS L. REV. 653 (2017) (providing an accessible overview of machine learning for lawyers).

2 An obvious question in the copyright area concerns whether creative works generated by intelligent machines are eligible for copyright protection. *See generally* Annemarie Bridy, *Coding Creativity: Copyright and the Artificially Intelligent Author*, 2012 STAN. TECH. L. REV. 5 (2012); Annemarie Bridy, *The Evolution of Authorship: Work Made by Code*, 39 COLUM. J.L. & ARTS 395 (2016); Daniel J. Gervais, *The Machine as Author*, 105 IOWA L. REV. (forthcoming 2020); Jane C. Ginsburg & Luke Ali Budiardjo, *Authors and Machines*, 34 BERKELEY TECH. L.J. 343 (2019); James Grimmelmann, *There's No Such Thing as a Computer-Authored Work—And It's a Good Thing, Too*, 39 COLUM. J.L. & ARTS 403 (2016) [hereinafter Grimmelmann, *There's No Such Thing*]; Daryl Lim, *AI & IP: Innovation & Creativity in an Age of Accelerated Change*, 52 AKRON L. REV. 813, 836–47 (2018); Carys J. Craig & Ian R. Kerr, *The Death of the AI Author* (Osgoode Hall Law Sch. Legal Studies Research Paper Series, 2019), https://ssrn.com/abstract=3374951. For a provocative discussion of the role of robots in copyright's cosmology, see generally James Grimmelmann, *Copyright for Literate Robots*, 101 IOWA L. REV. 657 (2016). For earlier discussions of copyright issues involving computer-generated works, see generally Ralph D. Clifford, *Intellectual Property in the Era of the Creative Computer Program: Will the True Creator Please Stand Up?*, 71 TUL. L. REV. 1675 (1997); Arthur R. Miller, *Copyright Protection for Computer Programs, Databases, and Computer-Generated Works: Is Anything New Since CONTU?*, 106 HARV. L. REV. 977, 1042–72 (1993); Pamela Samuelson, *Allocating Ownership Rights in Computer-Generated Works*, 47 U. PITT. L. REV. 1185 (1986).

3 As the U.S. Public Policy Council of the Association for Computing Machinery explained:

> An algorithm is a self-contained step-by-step set of operations that computers and other "smart" devices carry out to perform calculation, data processing, and automated reasoning tasks. Increasingly, algorithms implement institutional decision-making based on analytics, which involves the discovery, interpretation, and communication of meaningful patterns in data. Especially valuable in areas rich with recorded information, analytics relies on the simultaneous application of statistics, computer programming, and operations research to quantify performance.

U.S. ASS'N FOR COMPUTING MACHINERY: STATEMENT ON ALGORITHMIC TRANSPARENCY AND ACCOUNTABILITY 1 (Jan. 12, 2017), https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf [hereinafter *ACM Statement*].

4 For earlier discussions in this area, see generally Timothy K. Armstrong, *Digital Rights Management and the Process of Fair Use*, 20 HARV. J.L. & TECH. 49 (2006); Dan L. Burk & Julie E. Cohen, *Fair Use Infrastructure for Rights Management Systems*, 15 HARV. J.L. & TECH. 41, 53 (2001); Peter K. Yu, *Anticircumvention and Anti-anticircumvention*, 84 DENV. U.L. REV. 13, 63–73 (2006).

whether algorithms can be better deployed to promote fair use in copyright law. Part II explains why policymakers and commentators have remained skeptical about such deployment. Part III builds the case for greater algorithmic deployment to promote fair use. Part IV concludes by identifying areas to which policymakers and commentators should pay greater attention if automated fair use systems are to be developed. Although this Article draws heavily on U.S. copyright law—due in part to the location of this symposium and in part to the active transplant of the U.S. fair use model abroad[5]—the Article's insights should be applicable to all jurisdictions that have embraced fair use, fair dealing, or hybrid models.[6]

II. SKEPTICISM TOWARD AUTOMATION

Thus far, policymakers and commentators have advanced three major arguments explaining why algorithms cannot be satisfactorily deployed to promote fair use in copyright law. This Part outlines each argument in turn.

A. Backward State of Technology

The first major argument against the satisfactory deployment of algorithms to promote fair use concerns our relatively backward state of technology. As Edward Felten reminded us more than a decade ago: "Fair use is one of the starkest examples of the mismatch between what the law requires and what technology can do. Accurate, technological enforcement of the law of fair use is far beyond today's state of the art and may well remain

---

5 *See* Peter K. Yu, *Fair Use and Its Global Paradigm Evolution*, 2019 U. ILL. L. REV. 111, 129–37 (2019) [hereinafter Yu, *Global Paradigm Evolution*] (documenting a growing trend toward the worldwide adoption of the U.S. fair use model and a slowly emerging paradigm shift in international copyright norms); *see also* Peter K. Yu, *Customizing Fair Use Transplants*, 7 LAWS 1, 3–10 (2018) (discussing the efforts to transplant fair use across the world and the eight different modalities of transplantation that the transplanting jurisdictions have employed). *See generally* JONATHAN BAND & JONATHAN GERAFI, THE FAIR USE/FAIR DEALING HANDBOOK (2013), http://ssrn.com/abstract=2333863 (listing the fair use or fair dealing provisions from around the world).

6 I noted earlier the distinction between fair use and fair dealing as follows:

> Like fair use, . . . fair dealing allows for an unauthorized use of a copyrighted work. Unlike fair use, however, it promotes a closed system of copyright limitations and exceptions. Each fair dealing provision is drafted with a specific purpose, or a set of related purposes. Unless the user's conduct falls within a specified purpose, the use will not be permissible under copyright law.

Yu, *Global Paradigm Evolution*, *supra* note 5, at 126; *see also* Peter K. Yu, *The Quest for a User-Friendly Copyright Regime in Hong Kong*, 32 AM. U. INT'L L. REV. 283, 327 (2016) ("[A] better way to distinguish between fair dealing and fair use is to describe the former as a closed-ended, purpose-based regime and the latter as an open-ended, flexible regime.").

so permanently."[7] Because the current state of technology does not allow us to have a "judge on a chip,"[8] it is very difficult, if not impossible, to satisfactorily deploy algorithms to make automated fair use determinations on a case-by-case basis.

A key part of this technological challenge involves the significant difference between the approaches taken by judges to determine whether copyright law permits a specific use of a copyrighted work and those taken by computer programmers. Under the current copyright system, courts refrain from making ex ante determinations on what uses would be considered fair.[9] Instead, they allow users to test the law's limits. Should conflicts arise and the cases go to courts, judges will make determinations after the fact.[10] By contrast, computer programmers need to know in advance what legal rules and outcomes should be built into automated systems. While they will try their best to translate those rules and outcomes into code and algorithms, they will have considerable difficulty determining ex ante how judges will rule in new situations.[11] Inevitably, such translation will also

7 Edward W. Felten, *A Skeptical View of DRM and Fair Use*, COMM. ACM, Apr. 2003, at 57, 59; *see also* JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM 192 (2019) ("Automated processes have obvious efficiency advantages, but such processes may not align well (or at all) with applicable legal requirements that are couched in shades of gray."); Burk & Cohen, *supra* note 4, at 56 ("At least for now, there is no feasible way to build rights management code that approximates both the individual results of judicial determinations and the overall dynamism of fair use jurisprudence."); Ian R. Kerr et al., *Technical Protection Measures: Tilting at Copyright's Windmill*, 34 OTTAWA L. REV. 7, 31 (2002) ("[T]he technologies employed by [digital rights management systems] are not yet sufficiently sophisticated to mirror the law of copyright because [technological protection measures] themselves remain incapable of distinguishing between infringing and non-infringing uses of digital works."); Mark A. Lemley, *Rationalizing Internet Safe Harbors*, 6 J. ON TELECOMM. & HIGH TECH. L. 101, 110–11 (2007) ("Image-parsing software may someday be able to identify pictures or videos that are similar to individual copyrighted works, but they will never be able to determine whether those pictures are fair uses, or whether they are legitimate copies or displays made under one of the many statutory exceptions . . . .").

8 *See* Felten, *supra* note 7, at 58 ("A [digital rights management system] that gets all fair use judgments right would in effect be a 'judge on a chip' predicting with high accuracy how a real judge would decide a lawsuit challenging a particular use. Clearly, this is infeasible with today's technology."); *see also* Burk & Cohen, *supra* note 4, at 59 ("At present, only human intelligence, reviewing the unique circumstances of a particular use, can determine whether it is likely to be fair.").

9 *See* Dan L. Burk, *Algorithmic Fair Use*, 86 U. CHI. L. REV. 283, 288 (2019) ("[F]air use carries with it the disadvantage of ex ante uncertainty; no one can be entirely certain in advance how a court will weigh the four factors, and hence there is always some apprehension that a use may be found infringing rather than fair."); Burk & Cohen, *supra* note 4, at 61 ("Under the current conception of fair use, the decision whether or not to use a work is made ex ante by the user—if an infringement suit is brought later, the court may or may not validate the user's calculus, but penalties, if any, are imposed after the use has been undertaken.").

10 *See* John S. Erickson & Deirdre K. Mulligan, *The Technical and Legal Dangers of Code-Based Fair Use Enforcement*, 92 PROC. IEEE 985, 992 (2004) ("In the area of copyright law, the evolution of the doctrine of 'fair use' is tightly bound to the practice of after-the-fact adjudication.").

11 As Dan Burk and Julie Cohen observed:

bring up complicated questions concerning the computer programmers' understanding and interpretation of the law.[12]

To be sure, the past decade has seen significant advances in big data, machine learning, and artificial intelligence. One may recall media reports about how IBM Watson prevailed over noted human champions in the quiz show *Jeopardy!*[13] and how Google DeepMind's AlphaGo successfully beat the world's best players in Go, an Asian strategy board game.[14] As amazing as these technological advances have been, they do not automatically translate into automated fair use determinations. Just because Watson has performed well in *Jeopardy!* does not mean that it can perform equally well as a fair use judge. While transferred learning has become increasingly popular,[15] and such learning has allowed "information (or representations) learned on one task to aid learning on another task,"[16] there is no evidence that Watson or AlphaGo can successfully transfer its learning from *Jeopardy!* or Go to intellectual property law.[17]

> We are not optimistic that system designers will be able to anticipate the range of access privileges that may be appropriate for fair uses to be made of a particular work. Neither are we optimistic that system designers will be able to anticipate the types of uses that would be considered fair by a court.

Burk & Cohen, *supra* note 4, at 55.

12 *See* Lisa A. Shay et al., *Confronting Automated Law Enforcement*, *in* ROBOT LAW 235, 257 (Ryan Calo et al. eds., 2016) ("[T]hose who specify and implement the code base of a system will likely make their own interpretations of legal and illegal behavior, perhaps without any legal training."); Maayan Perel & Niva Elkin-Koren, *Black Box Tinkering: Beyond Disclosure in Algorithmic Enforcement*, 69 FLA. L. REV. 181, 189 (2017) ("[T]ranslating legal mandates into code inevitably embodies particular choices as to how the law is interpreted, which may be affected by a variety of extrajudicial considerations, including the conscious and unconscious professional assumptions of program developers, as well as various private business incentives."). *See generally* Lisa A. Shay et al., *Do Robots Dream of Electric Laws? An Experiment in the Law as Algorithm*, *in* ROBOT LAW, *supra*, at 274 (providing an interesting study documenting the variances in an empirical experiment in which three teams of computer programmers were asked to translate a subset of the New York State traffic law into computer code for the purposes of determining traffic violations based on real-world driving data).

13 John Markoff, *Computer Wins on "Jeopardy!": Trivial, It's Not*, N.Y. TIMES, Feb. 16, 2011, at A1.

14 *See* Choe Sang-Hun & John Markoff, *Machine Masters Man in Complex Game of Go*, N.Y. TIMES, Mar. 10, 2016, at A1 (reporting AlphaGo's victory over eighteen-time world Go champion Lee Sedol); Paul Mozur, *In Win for A.I., Google Program Humbles Master of a Mind-Boggling Game*, N.Y. TIMES, May 23, 2017, at B3 (reporting AlphaGo's victory over Ke Jie, the world's then best Go player).

15 For overviews of "transfer learning" in the deep learning context, see generally JOHN D. KELLEHER, DEEP LEARNING 236–37 (2019); Jason Brownlee, *A Gentle Introduction to Transfer Learning for Deep Learning*, MACHINE LEARNING MASTERY (Dec. 20, 2017), https://machinelearningmastery.com/transfer-learning-for-deep-learning/.

16 KELLEHER, *supra* note 15, at 236.

17 *See* Woodrow Barfield, *Towards a Law of Artificial Intelligence*, *in* RESEARCH HANDBOOK ON THE LAW OF ARTIFICIAL INTELLIGENCE 2, 9 (Woodrow Barfield & Ugo Pagallo eds., 2018) ("[W]hile impressive examples of skilled behavior, [IBM, AlphaGo, and other similar victories] are examples of artificial intelligence performing in a narrow domain of expertise; at this time more human-like artificial intelligence remains elusive."). Healthcare is one area in which Watson has not been very successful,

B. Changes in Creative Choices and Practices

The second major argument against the satisfactory deployment of algorithms to promote fair use relates to behavioral changes such deployment will generate. In a recent article, Dan Burk expressed fear that algorithmic fair use would create considerable biases, which in turn would affect authorial choices.[18] As he lamented: "[T]he design values embedded in automated systems become embedded in public behavior and consciousness. Thus, algorithmic fair use carries with it the very real possibility of habituating new media participants to its own biases and so progressively altering the fair use standard it attempts to embody."[19]

Professor Burk's observation is understandable. After all, we have already seen significant behavioral changes following the active deployment of algorithms in technology platforms to facilitate copyright enforcement,[20] such as YouTube's use of the Content ID system (which Part III.B will further discuss).[21] To avoid automatic detection, users have changed the type of content they upload to these platforms.[22] Outside the intellectual property context, we have seen Facebook users modifying behavior to manipulate or circumvent the platform's algorithms.[23] By including hashtags, metadata, or

despite IBM's high ambition. *See* ERIC J. TOPOL, DEEP MEDICINE: HOW ARTIFICIAL INTELLIGENCE CAN MAKE HEALTHCARE HUMAN AGAIN 55 (2019) ("IBM Watson's experience with MD Anderson, one of the country's leading cancer centers, was a debacle noteworthy for many missteps."); Steve Lohr, *Fulfilling Watson's Promise*, N.Y. TIMES, Feb. 29, 2016, at B1 (noting Watson's failure to perform well in the healthcare area).

18 Burk, *supra* note 9.

19 *Id.* at 285.

20 For discussions of algorithmic copyright enforcement, see generally Maayan Perel & Niva Elkin-Koren, *Accountability in Algorithmic Copyright Enforcement*, 19 STAN. TECH. L. REV. 473 (2016) [hereinafter Perel & Elkin-Koren, *Accountability*]; Perel & Elkin-Koren, *supra* note 12.

21 *See How Content ID Works*, YOUTUBE HELP, https://support.google.com/youtube/answer/2797370?hl=en (last visited Dec. 20, 2019) (providing an overview of YouTube's Content ID system). For discussions of the Content ID system, see generally Perel & Elkin-Koren, *Accountability*, *supra* note 20, at 509–16; Matthew Sag, *Internet Safe Harbors and the Transformation of Copyright Law*, 93 NOTRE DAME L. REV. 499, 543–60 (2017).

22 *See* Tony Zhou, *Postmortem: Every Frame a Painting*, MEDIUM (Dec. 3, 2017), https://medium.com/@tonyszhou/postmortem-1b338537fabc, *quoted in* Burk, *supra* note 9, at 303 (explaining how the author and his partner edited around YouTube's Content ID system by making trial-and-error adjustments based on "the length of the clips, the number of examples, which studios' films [they] chose, the way narration and clip audio weave together, the reordering and flipping of shots, the remixing of 5.1 audio, [and] the rhythm and pacing of the overall video").

23 *See* Caleb Garling, *Tricking Facebook's Algorithm*, THE ATLANTIC (Aug. 8, 2014), https://www.theatlantic.com/technology/archive/2014/08/tricking-facebooks-algorithm/375801/ (discussing the experience of tricking Facebook to elevate the author's post); Anjana Susarla, *The New Digital Divide Is Between People Who Opt Out of Algorithms and People Who Don't*, THE CONVERSATION (Apr. 17, 2019), https://theconversation.com/the-new-digital-divide-is-between-people-who-opt-out-of-algorithms-and-people-who-dont-114719 ("A study of Facebook usage found that when

typos, Internet and social media users have also redesigned their expressions to enhance or evade algorithm-driven recognition.[24] As if these examples were not enough, an entire industry has been created to help businesses and individuals "optimize" search results.[25] Thus, if algorithms are deployed to a greater extent to make automated fair use determinations, it will be no surprise to find authors altering their creative choices and practices.

C. Technological Shortcomings

The final major argument against the satisfactory deployment of algorithms to promote fair use pertains to the biases, bugs, and other documented problems now found in automated systems.[26] For instance, ProPublica published an exposé on the racial biases found in COMPAS, the scoring software used by law enforcement and correction personnel to determine risks of recidivism.[27] As the investigatory report stated, "black

---

participants were made aware of Facebook's algorithm for curating news feeds, about 83% of participants modified their behavior to try to take advantage of the algorithm, while around 10% decreased their usage of Facebook."); *see also* Jane Bambauer & Tal Zarsky, *The Algorithm Game*, 94 NOTRE DAME L. REV. 1, 12–14 (2018) (listing avoidance, altered conduct, altered input, and obfuscation among the dominant gaming strategies deployed by users on Internet platforms).

24 As Tarleton Gillespie observed:

> When we use hashtags in our tweets—a user innovation that was embraced later by Twitter—we are not just joining a conversation or hoping to be read by others, we are redesigning our expression so as to be better recognized and distributed by Twitter's search algorithm. Some may work to be noticed by the algorithm: teens have been known to tag their status updates with unrelated brand names, in the hopes that Facebook will privilege those updates in their friends' feeds. Others may work to evade an algorithm: Napster and P2P users sharing infringing copyrighted music were known to slightly misspell the artists' names, so users might find "Britny Speers" recordings but the record industry software would not.

Tarleton Gillespie, *The Relevance of Algorithms*, *in* MEDIA TECHNOLOGIES: ESSAYS ON COMMUNICATION, MATERIALITY, AND SOCIETY 167, 184 (Tarleton Gillespie et al. eds., 2014) (footnote omitted); *see also* Hannah Bloch-Wehba, *Automation in Moderation*, 52 CORNELL INT'L L.J. (forthcoming 2020) (noting the research from the University of Washington that shows how platform users have successfully "fooled" an algorithm-driven anti-trolling system by using misspellings such as "idiiots," "id.iots," or "i.diots" in their comments).

25 *See* Bambauer & Zarsky, *supra* note 23, at 45 ("A cottage industry has developed around gaming the algorithms used by important intermediaries. Google's algorithm is the *raison d'être* for the search engine optimization (SEO) industry and reputation firms.").

26 *See* ANDREW MCAFEE & ERIK BRYNJOLFSSON, MACHINE, PLATFORM, CROWD: HARNESSING OUR DIGITAL FUTURE 53 (2017) (noting the "biases and bugs" in intelligent machines); Burk, *supra* note 9, at 285 (listing "ersatz objectivity, diminished decisional transparency, and design biases" among the potential pitfalls in reliance on algorithmic regulation); Peter K. Yu, *The Algorithmic Divide and Equality in the Age of Artificial Intelligence*, 72 FLA. L. REV. (forthcoming 2020) (discussing algorithmic biases).

27 Jeff Larson et al., *How We Analyzed the COMPAS Recidivism Algorithm*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm. COMPAS stands for "Correctional Offender Management Profiling for Alternative Sanctions." *Id.*

defendants were far more likely than white defendants to be incorrectly judged to be at a higher risk of recidivism, while white defendants were more likely than black defendants to be incorrectly flagged as low risk."[28] In addition, the media provided wide coverage of how Microsoft's twitter bot Tay had quickly become sexist and racist because its "algorithms . . . had [the bot] 'learning' how to respond to others based on what was tweeted at it."[29] Another report showed that Hewlett-Packard's facial recognition technology had failed to properly recognize African Americans because the "[c]ameras on [its] new . . . computers did not track the faces of Black people in some common lighting conditions."[30]

While many of these problems were the result of improperly designed algorithms, properly designed algorithms will face similar problems if they fail to obtain appropriate training data. In computer science jargon, such problems are referred to as "garbage in, garbage out"—that is, the computer will produce faulty outcomes when the inputted data were inaccurate, biased, or otherwise inappropriate.[31] In the copyright context, for example, algorithms that are trained on data involving parodied entertainment will likely provide very different outcomes from those that are trained on data involving textbooks and other educational materials. How well automated fair use systems perform will therefore depend on how well the input data correspond to court decisions and day-to-day fair use practices.

More problematic, because algorithmic outcomes are often fed back into the algorithms as training data, the utilization of machine learning will create self-reinforced feedback loops that amplify the biases found in the initial

---

28 *Id.*

29 Lee Rainie & Janna Anderson, *Code-Dependent: Pros and Cons of the Algorithm Age* 13, PEW RES. CTR. (Feb. 8, 2017), http://www.pewinternet.org/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age.

30 Christian Sandvig et al., *When the Algorithm Itself Is a Racist: Diagnosing Ethical Harm in the Basic Components of Software*, 10 INT'L J. COMM. 4972, 4973 (2016) (citations omitted).

31 *See* Yu, *Global Paradigm Evolution*, *supra* note 5, at 157 (defining "garbage in, garbage out" situation as one "in which incorrect input ends up producing faulty output"). This age-old problem can be traced back to the early days of computing. *See* Amanda Levendowski, *How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem*, 93 WASH. L. REV. 579, 585 n.25 (2018) (noting that Charles Babbage, whom many refer to as the "father of computer," was asked whether the right answers would have come out if the wrong figures had been put into the machine (citing CHARLES BABBAGE, PASSAGES FROM THE LIFE OF A PHILOSOPHER 67 (1864))).

algorithms or training data.[32] Until these biases are corrected—by human intervention, perhaps[33]—the initial biases will be greatly magnified.[34]

As if the algorithmic biases were not disturbing enough, these biases are not easily observable because they are locked inside what commentators have referred to as "black boxes."[35] As Frank Pasquale described, the "workings [in these black boxes] are mysterious; we can observe [their] inputs and

---

32 As Ronald Yu and Gabriele Spina Alì observed:

> [T]here is a strong risk that AI may reiterate and even amplify the biases and flaws in datasets, even when these are unknown to humans. In this sense, AI has a self-reinforcing nature, due to the fact that the machine's outputs will be used as data for future algorithmic operations.

Ronald Yu & Gabriele Spina Alì, *What's Inside the Black Box? AI Challenges for Lawyers and Researchers*, 19 LEGAL INFO. MGMT. 2, 4 (2019); *see also* Sofia Grafanaki, *Autonomy Challenges in the Age of Big Data*, 27 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 803, 827 (2017) (noting that "algorithmic self-reinforcing loops are now present across many spheres of our daily life (e.g., retail contexts, career contexts, credit decisions, insurance, Google search results, news feeds)"); Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54, 69 (2019) ("Bad data . . . can perpetuate inequalities through machine learning, leading to a feedback loop that replicates existing forms of bias, potentially impacting minorities as a result."); *Digital Decisions* 8, CTR. FOR DEMOCRACY & TECH. (Sept. 27, 2018), https://cdt.org/wp-content/uploads/2018/09/Digital-Decisions-Library-Printer-Friendly-as-of-20180927.pdf ("Unreliable or unfair decisions that go unchallenged can contribute to bad feedback loops, which can make algorithms even more likely to marginalize vulnerable populations.").

33 Anthony Casey and Anthony Niblett, for example, noted the continuous role of humans in algorithmic development:

> Algorithmic decision-making does not mean that humans are shut out of the process. Even after the objective has been set, there is much human work to be done. Indeed, humans are involved in all stages of setting up, training, coding, and assessing the merits of the algorithm. If the objectives of the algorithm and the objective of the law are perfectly aligned at the ex ante stage, one must ask: Under what circumstances should a human ignore the algorithm's suggestions and intervene after the algorithm has made the decision?

Anthony J. Casey & Anthony Niblett, *A Framework for the New Personalization of Law*, 86 U. CHI. L. REV. 333, 354 (2019); *see also* Council Regulation 2016/679 art. 22(3), 2016 O.J. (L 119) 1 (requiring a data controller to "implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest [a decision based solely on automated processing, including profiling]"); Aziz Z. Huq, *A Right to a Human Decision*, 105 VA. L. REV. (forthcoming 2020) (discussing whether individuals have a "right to a human decision"); Meg Leta Jones, *The Right to a Human in the Loop: Political Constructions of Computer Automation and Personhood*, 47 SOC. STUD. SCI. 216 (2017) (tracing the historical roots of "the right to a human in the loop" back to rights that protect the dignity of data subjects).

34 *See* Yu, *supra* note 26 ("As time passes, the biases generated through these loops will become much worse than the biases found in the original algorithmic designs or the initial training data.").

35 For book-length treatments of the problems generated by "black box" algorithms, see generally VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR (2017); CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY (2016); FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2015).

outputs, but we cannot tell how one becomes the other."[36] Worse still, these black boxes tend to "disproportionately affect groups that are already disadvantaged by factors such as race, gender and socio-economic background."[37] In the copyright context, "black box" algorithms will likely affect users more than copyright holders or technology platforms, as the two latter groups will have more political clout and will therefore be in better positions to build their preferences into those black boxes, or the processes used to design them.[38]

In sum, there are many arguments against the satisfactory deployment of algorithms to promote fair use in copyright law. If such algorithms are to be deployed to a greater extent, policymakers will need to address most, or all, of these concerns.

III. THE CASE FOR AUTOMATION

Although policymakers and commentators have remained skeptical about the feasibility of developing automated systems to promote fair use in copyright law, several major arguments exist to support greater algorithmic deployment. This Part outlines each argument in turn.

36 PASQUALE, *supra* note 35, at 3; *see also* EUBANKS, *supra* note 35, at 5 ("[T]hat's the thing about being targeted by an algorithm: you get a sense of a pattern in the digital noise, an electronic eye turned toward *you*, but you can't put your finger on exactly what's amiss."); Rainie & Anderson, *supra* note 29, at 19 ("There is a larger problem with the increase of algorithm-based outcomes beyond the risk of error or discrimination—the increasing opacity of decision-making and the growing lack of human accountability." (quoting Marc Rotenberg, Executive Director, Electronic Privacy Information Center)).

37 Kate Crawford & Ryan Calo, *There Is a Blind Spot in AI Research*, NATURE (Oct. 13, 2016), https://www.nature.com/news/there-is-a-blind-spot-in-ai-research-1.20805. As Cathy O'Neil observed:

> [Algorithm-driven weapons of math destruction] tend to punish the poor. This is, in part, because they are engineered to evaluate large numbers of people. They specialize in bulk, and they're cheap. That's part of their appeal. The wealthy, by contrast, often benefit from personal input. A white-shoe law firm or an exclusive prep school will lean far more on recommendations and face-to-face interviews than will a fast-food chain or a cash-strapped urban school district. The privileged, we'll see time and again, are processed more by people, the masses by machines.

O'NEIL, *supra* note 35, at 8; *see also* EUBANKS, *supra* note 35, at 12 (lamenting how "[a]utomated decision-making shatters the social safety net, criminalizes the poor, intensifies discrimination, and compromises our deepest national values"); Rainie & Anderson, *supra* note 29, at 63–65 (surveying views on whether the disadvantaged will lag behind even further in this algorithmic age).

38 *See* COHEN, *supra* note 7, at 193 ("Industry standard-making processes . . . are lengthy, secretive, and notoriously resistant to public interest oversight."); Julie E. Cohen, *DRM and Privacy*, 18 BERKELEY TECH. L.J. 575, 616 (2003) ("[N]ot all standards processes include end user representation, and even in those that do, there is no assurance that end user grievances, once aired, will prospectively shape the standards that are brought to market."). *See generally* MONICA HORTEN, A COPYRIGHT MASQUERADE: HOW CORPORATE LOBBYING THREATENS ONLINE FREEDOMS (2013) (discussing how legislative capture by the copyright industries has undermined online freedom); BRINK LINDSEY & STEVEN TELES, THE CAPTURED ECONOMY: HOW THE POWERFUL ENRICH THEMSELVES, SLOW DOWN GROWTH, AND INCREASE INEQUALITY 64–89 (2017) (discussing capture in the intellectual property area).

A. Gradual Deployment

The first major argument supporting greater algorithmic deployment to promote fair use concerns the need to take incremental steps if automated fair use systems are to be developed. Just because the current state of technology does not support perfect automated fair use determinations does not mean that we should not try to deploy algorithms to a more limited extent. As Microsoft software architects Barbara Fox and Brian LaMacchia declared in the early 2000s:

> [The limitation that no one can mathematically model fair use, as it is understood today,] should not stop us from attempting to identify a useful subset we might approximate in code. That is, we can take a purely pragmatic engineering approach. . . . Focus first on defining and modeling a useful subset of fair use rights in some policy language, then add these expressions to the policy evaluators of [digital rights management] systems.[39]

More recently, Niva Elkin-Koren argued that "the checks that [fair use] intends to create on the rights of authors must . . . be embedded in the design of online systems."[40]

In an article written for the Inaugural Summit on Intellectual Property and Digital Media organized by The Cable Center and the University of Denver Sturm College of Law in May 2006, I also noted the need to distinguish between limitations and exceptions that can be interpreted by machines from those that cannot.[41] As I explained at that time:

---

39 Barbara L. Fox & Brian A. LaMacchia, *Encouraging Recognition of Fair Uses in DRM Systems*, COMM. ACM, Apr. 2003, at 61, 63. Matthew Sag concurred:

> The difficulty of completely automating fair use analysis does not suggest . . . that algorithms have no role to play. Experience, common sense, and recent empirical research suggest that there are some objective characteristics that make a finding of fair use more likely, and there is no reason in principle why matching algorithms could not be fine-tuned to identify common situations associated with a higher probability of fair use.

Sag, *supra* note 21, at 531–32. Likewise, Timothy Armstrong observed:

> The flaw in the conclusion that [digital rights management] cannot accommodate fair use is an unduly hasty inductive leap from the specific (the impossibility of modeling the substance of fair use law in machine-administrable form) to the general (the supposed impossibility of protecting fair use at all in [digital rights management] systems). The foreclosure of one avenue for protecting fair use, however, does not imply that all avenues are likewise foreclosed, but only that design principles other than the creation of a perfect "judge on a chip" must be explored.

Armstrong, *supra* note 4, at 88.

40 Niva Elkin-Koren, *Fair Use by Design*, 64 UCLA L. REV. 1082, 1085 (2017).

41 *See* Yu, *supra* note 4, at 63–73 (discussing the need for such a distinction); *see also* Deirdre Mulligan & Aaron Burstein, *Implementing Copyright Limitations in Rights Expression Languages*, *in*

> The fact that the scope and boundaries of [fair use] are uncertain and that software code at the current state of technology may not be able to capture the full range of exceptions and limitations in the copyright system does not mean that we should not build legitimate uses into the [digital rights management] systems.[42]

Drawing on research in the area of economic, social, and cultural rights, to which intellectual property rights belong,[43] I suggested that automated fair use systems could be developed by first focusing on "minimum essential levels of noninfringing uses" before the coverage is expanded to take advantage of technological improvements and increased technical resources.[44]

B. *Extant Deployment*

The second argument supporting greater algorithmic deployment to promote fair use relates to the fact that such deployment has already taken place in the area of copyright enforcement.[45] Whether we like it or not,

---

DIGITAL RIGHTS MANAGEMENT: ACM WORKSHOP ON DIGITAL RIGHTS MANAGEMENT, DRM 2002, WASHINGTON, DC, USA, NOVEMBER 18, 2002: REVISED PAPERS 137 (Joan Feigenbaum ed., 2002) (discussing ways and challenges to implementing copyright limitations and exceptions in rights expression languages, with a focus on XrML, the eXtensible Rights Markup Language); Fox & LaMacchia, *supra* note 39, at 63 (considering the importance of determining "how to create machine-interpretable expressions that adequately model a set (or subset) of fair use rights").

42 Yu, *supra* note 4, at 63.

43 *See* International Covenant on Economic, Social and Cultural Rights art. 15(1)(c), Dec. 16, 1966, 993 U.N.T.S. 3 (requiring each state party to the Covenant to "recognize the right of everyone . . . [t]o benefit from the protection of the moral and material interests resulting from any scientific, literary or artistic production of which he [or she] is the author."); Comm. on Econ., Soc. & Cultural Rights, *General Comment No. 17: The Right of Everyone to Benefit from the Protection of the Moral and Material Interests Resulting from Any Scientific, Literary or Artistic Production of Which He or She Is the Author (Article 15, Paragraph 1(c), of the Covenant)*, U.N. Doc. E/C.12/GC/17 (Jan. 12, 2006) (providing an authoritative interpretation of Article 15(1)(c) of the International Covenant on Economic, Social and Cultural Rights).

44 Yu, *supra* note 4, at 65–66. As I elaborated:

> Under this proposal, software code would be used to accommodate machine-interpretable noninfringing uses, while the determination of the machine-uninterpretable noninfringing uses would remain in the province of courts. As technology advanced and computer programming became more sophisticated, [digital rights management] systems would be able to accommodate more noninfringing uses. The domain of machine-interpretable noninfringing uses would therefore expand, leaving fewer and fewer copyright matters to courts.

Id. at 65.

45 As Professor Sag observed:

> [D]espite the lack of a de jure obligation to filter under the DMCA [Digital Millennium Copyright Act], many platforms—typically large-scale commercial enterprises are nonetheless implementing

copyright holders will continue to expand such deployment, due in large part to the efficiency and effectiveness provided by automation. A case in point is YouTube. With "more than half a million hours of video" being uploaded on to the platform every day,[46] it is virtually impossible for this streaming platform to manually review each uploaded video file.

Thus far, the Content ID system deployed by YouTube has been the most widely cited example of automated copyright enforcement.[47] Using hashes or digital fingerprints, this system compares files uploaded by Internet users with the reference files provided by copyright holders.[48] If the files match, copyright holders have the choice to "[b]lock a whole video from being viewed," "[m]onetize the video by running ads against it," and "[t]rack the video's viewership statistics."[49] When musical works or sound recordings are involved, the right holders can also mute the video. Although the enforcement provided by the Content ID system has been both underinclusive and overinclusive,[50] the copyright industries and their supportive

> automated copyright enforcement systems. At the present time, platforms using automated copyright enforcement include Scribid, 4shared, Dropbox, YouTube, Facebook, SoundCloud, Twitch, TuneCore, Tumblr, Veoh, and Vimeo. The pressure to adopt automated filtering comes primarily from rightsholders, but these systems also meet some of the business objectives of platforms.

Sag, *supra* note 21, at 538–39; *see also* NICOLAS P. SUZOR, LAWLESS: THE SECRET RULES THAT GOVERN OUR DIGITAL LIVES 72 (2019) ("Automated copyright detection systems have now been built into many other services on the Internet. Facebook has developed its own detection systems, and companies like Audible Magic produce software that has been adopted by many platforms."); Burk, *supra* note 9, at 284 ("In the area of copyright, protection of digitized works is already increasingly mediated by algorithmic enforcement systems that are intended to effectuate the rights of copyright owners while simultaneously limiting the liability of content intermediaries."); Joe Karaganis & Jennifer Urban, *The Rise of the Robo Notice*, COMM. ACM, Sept. 2015, at 28 (expressing concern about the growing use of robo notices to take down potentially infringing copyrighted materials).

46 *See* Sag, *supra* note 21, at 539 ("Collectively, YouTube users now upload more than half a million hours of video and watch hundreds of millions of hours of video every day.").

47 *See How Content ID Works*, *supra* note 21.

48 As Professor Sag explained:

> Content ID begins by taking reference files submitted by a person claiming to represent the copyright owner and converting such files into a hash file or a digital fingerprint. In computer science, a hash function is used to map information of indeterminate size to a long string of letters and digits of fixed size. A "perfect" hash function will generate a unique hash for each unique input. The 128-bit hash for the previous paragraph is *ObllcO463b44082968blf3eedffbOf80*, the hash for the same text with the word "Banana" substituted for "DMCA" is *2863eb5ee4acdb9dO37ea9541ce16b62*. Neither text can be reverse engineered from their hash values, but once the texts are encoded as hash values it is trivial to compare them to see if one is a match for the other. Using hash values to match audio and visual content encoded in differing file formats is no trivial task, but the concepts are similar. Using these hash values, new user content is automatically compared to the reference file as it is uploaded to the site. The system can match audio and/or video; it can detect partial and degraded quality matches as well as perfect high quality copies.

Sag, *supra* note 21, at 545 (footnote omitted).

49 *How Content ID Works*, *supra* note 21.

50 As Nicolas Suzor observed:

policymakers and commentators have slowly embraced it and other similar monitoring and filtering tools. In Europe, for instance, the recently adopted EU Directive on Copyright in the Digital Single Market imposes copyright liability on Internet service providers should they fail to put in place filtering technology that would protect the copyrighted content disseminated online.[51]

C. Technological Improvements

The third major argument supporting greater algorithmic deployment to promote fair use pertains to the new technological advances relating to big data, machine learning, and artificial intelligence. Although the Content ID system has provided a paradigmatic example of automated copyright enforcement, it has yet to realize the full potential generated by these new advances. With the incorporation of big data analytics and machine learning capabilities—and the development of learning algorithms, or so-called "learners"[52]—automated fair use systems will not only function more

---

> Of course, not everybody is happy with Content ID. The system provides no reliable way to resolve disputes about fair use, which upsets both copyright owners and video creators. In modern equivalents to Stephanie Lenz's dancing baby case, the YouTube algorithm will automatically flag music and other copyrighted material that is caught in the background of a video. It will also automatically catch content used in a critique or parody. YouTube's Content ID system cannot tell the difference between someone who copies a few minutes of, say, a professional sporting event to make fun of it and someone who shares parts of a match in a way that might deprive the distributors of revenue. In these cases, YouTube creators have to go through a process to try to convince the copyright owner that their use is fair. Ultimately, the copyright owner makes the decision: if they reject the user's claim, they are redirected through the DMCA process to lodge a formal takedown request. At this point, unless the YouTube user files a counter-notice, they'll get a "strike" against their account. If a user gets three strikes in ninety days, Google will terminate their account.

SUZOR, *supra* note 45, at 72.

51 Article 17(4)(b) of the Directive provides:

> If no authorisation [from the rightholders] is granted, online content-sharing service providers shall be liable for unauthorised acts of communication to the public, including making available to the public, of copyright-protected works and other subject matter, unless the service providers demonstrate that they have . . . made, in accordance with high industry standards of professional diligence, best efforts to ensure the unavailability of specific works and other subject matter for which the rightholders have provided the service providers with the relevant and necessary information . . . .

Directive 2019/790 art. 17(4)(b), 2019 O.J. (L 130) 92. *See generally* Martin Senftleben, *Bermuda Triangle: Licensing, Filtering and Privileging User-Generated Content Under the New Directive on Copyright in the Digital Single Market*, 41 EUR. INTELL. PROP. REV. 480, 482–85 (2019) (highlighting the challenges posed by the new filtering obligation under the EU Directive).

52 *See* PEDRO DOMINGOS, THE MASTER ALGORITHM: HOW THE QUEST FOR THE ULTIMATE LEARNING MACHINE WILL REMAKE OUR WORLD 6 (2015) ("Learning algorithms—also known as learners—are algorithms that make other algorithms. With machine learning, computers write their own programs, so we don't have to.").

efficiently and effectively, but their decisions will also bear stronger resemblances to those made by real-life judges.[53]

Moreover, evidence has already emerged to show that intelligent machines can perform select tasks better than humans. For instance, researchers have documented the advantage of using learning algorithms to diagnose cancer and to perform other tasks in the health area.[54] Commentators have also noted that algorithms "are better and faster than humans at detecting credit card fraud."[55] In addition, the performance of intelligent machines will not be affected by emotion, exhaustion, stress, or other cognitive barriers.[56] These machines can also be tested and therefore

---

53 As Niva Elkin-Koren reminded us:

> Overall, th[e] concerns regarding the limitations of algorithmic fair use overlook recent developments in Artificial Intelligence . . . and machine learning capabilities. AI has already been applied in very sophisticated contexts: physicians use algorithms to guide their diagnoses; banks use them to decide when to approve a loan; security agencies use AI to identify risks; lawyers use them to perform due diligence; and even courts rely on algorithms for sentencing, by scoring the risk of the offender committing future crimes. AI has already been applied for decision-making processes in contexts that are far more complex than fair use, involving critical issues of life and death, health, financial risks, and national security.

Elkin-Koren, *supra* note 40, at 1096–97.

54 As Jonathan Guo and Li Bin observed:

> Esteva et al. trained deep convolutional neural networks (CNN) based on a dataset of 129,450 clinical images to diagnose skin cancer. The results demonstrated that this system is able to classify skin cancer at a comparable level to dermatologists. . . . Liu from Google, Inc. reported a CNN framework to aid the pathological diagnosis of breast cancer metastasis in lymph nodes. The results showed that this system could improve the speed, accuracy, and consistency of diagnosis, as well as reduce the false negative rate to a quarter of the rate experienced by human pathologists.

Jonathan Guo & Li Bin, *The Application of Medical Artificial Intelligence Technology in Rural Areas of Developing Countries*, 2 HEALTH EQUITY 174, 175 (2018); *see also* TOPOL, *supra* note 17, at 117–18 (discussing the impressive progress in algorithmic image processing).

55 *Digital Decisions*, *supra* note 32, at 4.

56 *See* Karni Chagal-Feferkorn, *The Reasonable Algorithm*, 2018 U. ILL. J.L. TECH. & POL'Y 111, 144 ("Unlike humans, algorithms do not have self-interests affecting their judgement, they do not omit any of the decision-making stages or base their decisions on heuristics or biases, and they are not subject to human physical or emotional limitations such as exhaustion, stress or emotionality." (footnotes omitted)); Rebecca Crootof, *"Cyborg Justice" and the Risk of Technological–Legal Lock-in*, 119 COLUM. L. REV. F. 233, 236 (2019) (noting that a "judge's sensitivity to context and penchant for leniency may vary dramatically with whether they are hungry, tired, bored, overworked, overwhelmed, or otherwise distracted"); Lim, *supra* note 2, at 834 ("AI does not suffer from perceptual limitations the way that humans do."); Ozkan Eren & Naci Mocan, *Emotional Judges and Unlucky Juveniles* (Nat'l Bureau of Econ. Research, Working Paper No. 22,611), https://www.nber.org/papers/w22611.pdf (documenting the surprising impact of unexpected outcomes of football games on the type and length of sentences handed down by juvenile court judges); Kurt Kleiner, *Lunchtime Leniency: Judges' Rulings Are Harsher When They Are Hungrier*, SCI. AM. (Sept. 1, 2011), https://www.scientificamerican.com/article/lunchtime-leniency/ ("Judges granted 65 percent of requests they heard at the beginning of the day's session and almost none at the end. Right after a snack break, approvals jumped back to 65 percent again." (citing a study at Ben Gurion University in Israel and Columbia University examining more than 1000 decisions by eight Israeli judges who ruled on convicts' parole requests)).

improved.[57] Should errors be found, the machines "are unlikely to make the same mistake[s] again" once those errors have been corrected.[58] In fact, as Andrew McAfee and Erik Brynjolfsson reminded us, "it is a lot harder to get humans to acknowledge their biases (how many avowed racists or sexists do you know?), let alone the hard work required to overcome them."[59]

When the advantages of automation are extrapolated to the fair use context, these advantages suggest that automated fair use systems may analyze certain fair use factors better than human judges.[60] For illustrative purposes, consider the analysis of the third factor in the U.S. fair use provision, which involves "the amount and substantiality of the portion used in relation to the copyrighted work as a whole."[61] With respect to quantitative analysis,[62] it is not difficult to see how computers could provide quicker and more accurate analysis.[63] In fact, any judge seeking to undertake a quick quantitative analysis will likely rely on computer assistance to count words or compare sizes.

---

57 *See* MCAFEE & BRYNJOLFSSON, *supra* note 26, at 53 ("[M]achine-based systems typically can be tested and improved.").

58 *Id.*

59 *Id.*

60 Dan Burk expressed continuous skepticism toward automated fair use determinations:

> It is perhaps not too farfetched to imagine a programmable exception of the fair dealing laundry list sort—although even for supposedly discrete statutory exceptions, concepts like "educational use" or "news reporting" might be unexpectedly tricky to reduce to computable code. But one can, for example, imagine programming a system to determine, perhaps on the basis of geolocational data and scraped calendaring or advertising data, whether a nondramatic musical work is being performed at an agricultural fair. It is far more difficult to envision how one might program a system to determine whether a given use has a relevant degree of impact on the actual or potential market for the work being used or whether an excerpt from the work is so significant as to constitute the "heart" of an author's creation.

Burk, *supra* note 9, at 292 (footnotes omitted).

61 17 U.S.C. § 107 (2018).

62 *See* Harper & Row, Publishers, Inc. v. Nation Enters., 471 U.S. 539, 583 n.6 (1985) ("The inquiry into the substantiality of appropriation has a quantitative . . . aspect."); *see also* Folsom v. Marsh, 9 F. Cas. 342, 348 (C.C.D. Mass. 1841) (No. 4901) (noting the need to examine "the quantity . . . of the materials used").

63 *See* Elkin-Koren, *supra* note 40, at 1096 ("Some fair use considerations might be relatively easy to automate, such as the amount copied from the original work. For instance, a program could give a higher fair use score based on similarity of less than 10 percent.").

With respect to qualitative analysis,[64] however, the lack of emotion and empathy in machines[65] may suggest their limited ability to determine[66] what courts would consider as the "heart" of the work.[67] On its face, this suggestion is valid.[68] Making value judgment is not the forte of automated systems. In reality, however, there already exists a large trove of data concerning which pages or sentences of a book Kindle users have highlighted the most.[69] Netflix also keeps track of the parts of a movie or TV program that its subscribers have paused or viewed repeatedly.[70] While those highlighted lines or repeat plays may not provide perfect proxies for the heart of the works, they constitute powerful evidence on which parts of the works many users have found important or interesting.

Likewise, intelligent machines can analyze well the fourth factor in the U.S. fair use provision, which involves "the effect of the use upon the potential market for or value of the copyrighted work."[71] This analysis will be even better if the machines can collect additional market information that is currently not in the possession of copyright holders, users, or technology

---

64 *See* Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 587 (1994) ("[The third] factor calls for thought not only about the quantity of the materials used, but about their quality and importance, too.").

65 *See* LEE KAI-FU, AI SUPERPOWERS: CHINA, SILICON VALLEY, AND THE NEW WORLD ORDER 142 (2018) ("Taking the next step to emotionally intelligent robots may require self-awareness, humor, love, empathy, and appreciation for beauty. These are the key hurdles that separate what AI does today—spotting correlations in data and making predictions—and artificial general intelligence."); MCAFEE & BRYNJOLFSSON, *supra* note 26, at 123 ("[T]he ability to work effectively with people's emotional states and social drives will remain a deeply human skill for some time to come."); TOPOL, *supra* note 17, at 290 ("[H]uman empathy is not something machines can truly simulate, despite ongoing efforts to design sociable robots or apps that promote empathy."); Lawrence B. Solum, *Legal Personhood for Artificial Intelligences*, 70 N.C. L. REV. 1231, 1269–71 (1992) (discussing the lack of capacity in artificial intelligence for feelings).

66 *See* Burk, *supra* note 9, at 292 (noting the difficulty in programing an automated fair use system "to determine . . . whether an excerpt from the work is so significant as to constitute the 'heart' of an author's creation").

67 *See Harper & Row*, 471 U.S. at 600 (analyzing whether the defendant magazine "had taken 'the heart of the book'").

68 *See* Burk, *supra* note 9, at 292 ("It is far more difficult to envision how one might program a system to determine . . . whether an excerpt from the work is so significant as to constitute the 'heart' of an author's creation.").

69 *See Viewing Popular Highlights on Kindles*, EBOOK READER BLOG (Feb. 15, 2018), https://blog.the-ebook-reader.com/2018/02/15/viewing-popular-highlights-on-kindles/ ("Popular Highlights show the most highlighted passages that readers have added to Kindle books. . . . Amazon also displays how many times each passage has been highlighted.").

70 *See* Kal Raustiala & Christopher Jon Sprigman, *The Second Digital Disruption: Streaming and the Dawn of Data-Driven Creativity*, 94 N.Y.U. L. REV. 1555, 1587 (2019) ("Some parameters that Netflix tracks include, but are likely not limited to, pause/rewind/fast-forward behavior; day of the week; date of viewing; time of viewing; zip code; preferred devices; completion rate; user ratings; user search behavior; and browsing and scrolling behavior.").

71 17 U.S.C. § 107 (2018).

platforms. Using big data analysis, algorithms can analyze the collected information to predict the user's likely impact on both the actual and potential markets of the copyrighted work. Such collection and analysis will overcome the widely noted problem of having insufficient information about the circumstances surrounding the specific use of a copyrighted work.[72] To be sure, intelligent machines may not be able to make better predictions than trained economists or valuation experts.[73] Nevertheless, they will be able to draw conclusions more quickly than humans, and will thereby facilitate real-time market analysis that will be both costly and time-consuming when conducted manually. Moreover, artificial intelligence has already been widely deployed in the financial area to provide predictive analysis.[74] Such analysis will only improve with technological improvements.

Finally, if deep learning, or the use of neural networks, is involved,[75] the comparison between automated and human performance will become even

---

72 *See* Felten, *supra* note 7, at 58 (identifying the "[l]ack of knowledge about the circumstances" of the use as one of the two key reasons why fair use cannot be built into digital rights management systems). As Professor Elkin-Koren explained:

> [One] concern [regarding the limitations of algorithmic fair use] is that algorithms that analyze fair use will fail to process information that is external to the content itself. For instance, determining the nature of use may require external information and additional analysis of facts. Yet, algorithms could be programmed to extract and analyze data from external sources. For instance, educational use might be determined based on tagging the nature of the user. A program could detect the type of user (e.g., educational institution, governmental agency) based on the domain name (e.g., .edu, .gov) or by checking registration in external databases. Another indication for the nature of use could be the type of tagging selected by the party that uploads the work (educational, commercial, personal/private use). The commercial nature of use might actually be determined by the presence of advertisements, or other means of monetizing the content. External information might also be used to determine "the effect of the use upon the potential market" for the copyrighted work, using the commercial nature of use as a proxy.

Elkin-Koren, *supra* note 40, at 1095–96.

73 *See* Felten, *supra* note 7, at 58 ("[T]he fourth factor in the [fair use] test evaluates the effect of the use on the market for the original work. It requires reasoning about the economics of a particular market, a task even well-trained humans find difficult.").

74 The literature emergent in this area is vast and fast-growing. *See, e.g.*, William Magnuson, *Artificial Financial Intelligence*, 9 HARV. BUS. L. REV. (forthcoming 2020) (discussing the dangers and real-world limitations of deploying artificial intelligence in finance).

75 As a government report on artificial intelligence explained:

> Deep learning uses structures loosely inspired by the human brain, consisting of a set of units (or "neurons"). Each unit combines a set of input values to produce an output value, which in turn is passed on to other neurons downstream. For example, in an image recognition application, a first layer of units might combine the raw data of the image to recognize simple patterns in the image; a second layer of units might combine the results of the first layer to recognize patterns-of-patterns; a third layer might combine the results of the second layer; and so on.

NAT'L SCI. & TECH. COUNCIL, PREPARING FOR THE FUTURE OF ARTIFICIAL INTELLIGENCE 9 (2016). For discussions of deep learning, see generally ETHEM ALPAYDIN, MACHINE LEARNING: THE NEW AI 104–09 (2016); KELLEHER, *supra* note 15; JOHN D. KELLEHER & BRENDAN TIERNEY, DATA SCIENCE 121–30 (2018); THIERRY POIBEAU, MACHINE TRANSLATION 181–95 (2017).

more complicated because automated fair use determinations may not analyze the individual factors the same way a judge or a copyright lawyer would.[76] As Dan Burk observed:

> One can imagine that a neural network or other machine learning system could detect these or other patterns in the data surrounding past cases, matching them to similar patterns in the data surrounding future fair use incidents, situations, and scenarios without formal programming definition of the fair use factors.[77]

While one could argue that a proper fair use analysis must be conducted the same way as how judges would, one cannot help but wonder whether society would find it acceptable to have automated fair use determinations that generate outcomes that have high correlations to the outcomes of judge-made decisions.[78]

D. Scalability

The fourth major argument supporting greater algorithmic deployment to promote fair use regards the scalability of automated fair use systems.[79] As Charles Clark noted in a book chapter that has been widely cited for its title, "the answer to the machine is in the machine."[80] With the creation and

---

76 *See* Elkin-Koren, *supra* note 40, at 1099 ("AI and machine learning would make it difficult for courts to check the rules embedded in the system, since these systems may not explicitly demonstrate the legal specifications of the four factors of fair use.").

77 Burk, *supra* note 9, at 293.

78 *See* Elkin-Koren, *supra* note 40, at 1099 (noting that, with the growing use of artificial intelligence and machine learning, courts may have to "determin[e] acceptable error rates when testing the outcome of such a system compared to determination by the court").

79 *See* AJAY AGRAWAL ET AL., PREDICTION MACHINES: THE SIMPLE ECONOMICS OF ARTIFICIAL INTELLIGENCE 67 (2018) ("One major benefit of prediction machines is that they can scale in a way that humans cannot."); TARLETON GILLESPIE, CUSTODIANS OF THE INTERNET: PLATFORMS, CONTENT MODERATION, AND THE HIDDEN DECISIONS THAT SHAPE SOCIAL MEDIA 97 (2018) ("Artificial intelligence techniques offer . . . to solve the problem of scale."); Tim Wu, *Will Artificial Intelligence Eat the Law? The Rise of Hybrid Social-Ordering Systems*, 119 COLUM. L. REV. 2001, 2002 (2019) ("Compared with the legal system, software has enormous advantages of scale and efficacy of enforcement. It might tirelessly handle billions if not trillions of decisions in the time it takes a human court to decide a single case.").

80 Charles Clark, *The Answer to the Machine Is in the Machine*, *in* THE FUTURE OF COPYRIGHT IN A DIGITAL ENVIRONMENT 139 (P. Bernt Hugenholtz ed., 1996) (capitalization omitted). William Patry pointed out the title of Clark's chapter is largely misnamed because that chapter concluded that "the answer to the machine may turn out to be not only in the machine, but the machine will certainly be an important part of the answer." WILLIAM F. PATRY, HOW TO FIX COPYRIGHT 236–41 (2011).

dissemination of hundreds of exabytes of data and digital content every day,[81] it is almost impossible for technology platforms to not rely on algorithms to determine whether a specific use of a copyrighted work has complied with copyright law.[82] The need for such reliance is indeed why YouTube has deployed the Content ID system to facilitate online copyright enforcement.[83] Fortunately, the greater deployment of algorithms to promote fair use may help address such an ever-growing deluge of content. As Professor Elkin-Koren observed:

> Algorithmic fair use could offer a workable solution for a growing number of circumstances that involve a large volume of content in which the costs of determining fair use on a case-by-case basis, and the risk of mistakenly determining fair use, are simply too high. That is the case, for instance, in educational institutions which make large quantities of teaching materials available for educational purposes using e-reserve systems.[84]

To be sure, a time gap will always exist between the latest judicial decision and the legal rules and outcomes that programmers manage to encode in the algorithms.[85] Nevertheless, if automated fair use systems are constantly upgraded, the time lag between the two may be much more acceptable and less problematic.[86] Moreover, those human actors who

---

81 *See* Jeff Desjardins, *How Much Data Is Generated Each Day?*, WORLD ECON. F. (Apr. 17, 2019), https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/ ("By 2025, it's estimated that 463 exabytes of data will be created each day globally—that's the equivalent of 212,765,957 DVDs per day!"); *see also* Sag, *supra* note 21, at 513 ("In 2016, YouTube users were uploading 400 hours of video content every minute . . . .").

82 *See* Elkin-Koren, *supra* note 40, at 1098 ("The need to address the sheer volume of copyright disputes requires a new approach to fair use that involves rethinking the role of legal oversight in algorithmic adjudication."); Sag, *supra* note 21, at 554 ("With over 400 hours of video being uploaded to YouTube every minute, it is hard to imagine that either rightsholders, or the platform itself, could meaningfully prevent the evisceration of online copyright without relying on automation to some extent.").

83 *See supra* Part III.B (discussing the YouTube's Content ID system).

84 Elkin-Koren, *supra* note 40, at 1100.

85 *See* Burk, *supra* note 9, at 298 ("[O]ne concern that could stem from the dynamic legal nature of fair use is whether automated instantiation of fair use freezes the standard as of the time it was encoded, so that the law and the algorithm diverge.").

86 *See id.* ("The algorithm could of course be updated to learn or incorporate shifts in the legal standard."); Elkin-Koren, *supra* note 40, at 1097 ("Machine learning capabilities could ensure that the system is up to date—because the classifications applied by the algorithm are constantly refined based on new fair use rulings."). This constant upgrade could be compared to the frequent and virtually instantaneous updates found in the thirteen legacy root zone servers containing information about Internet domain names:

> As a past legacy, the database in the A Root Server, which the Internet Corporation for Assigned Names and Numbers ("ICANN") currently controls by virtue of its contract with the U.S.

constantly have to make fair use determinations, such as Internet users and copyright enforcement personnel, do not always keep track of all the latest fair use decisions. If machines can learn those decisions more quickly than the relevant human actors—such as judges, lawyers, and law enforcement personnel—automated fair use systems can still be highly appealing. After all, the limited number of fair use cases suggest that most fair use determinations are made outside the courtroom.[87]

E. Low-Cost Determinations

The fifth major argument supporting greater algorithmic deployment to promote fair use involves the ability of automated fair use systems to provide low-cost determinations to a large number of people who otherwise may not be able to afford copyright lawyers.[88] As Lawrence Lessig put it memorably, "fair use in America simply means the right to hire a lawyer to defend your right to create."[89] With automation, the greater deployment of algorithms will help those users who cannot afford to hire lawyers, or hire them frequently, to explore the boundaries of the law or to provide the support needed to test those boundaries.[90] If automated fair use determinations can have legal effects—even if only on an interim basis—those determinations can enlarge the creative spaces of risk-averse users, some of whom may fear that their creative endeavors will violate current copyright law.[91]

Although automated fair use systems can help users, those systems can also harm users if the algorithms involved fail to capture the full range of limitations and exceptions in copyright law. Indeed, when Dan Burk and Julie Cohen analyzed this topic about two decades ago, one of their key concerns was that automated systems would end up encouraging minimalist

---

> Department of Commerce . . . , is considered authoritative. The other root servers merely copy this root zone file to their servers.

Peter K. Yu, *The Origins of ccTLD Policymaking*, 12 CARDOZO J. INT'L & COMP. L. 387, 390 (2004).

87 *See* Burk & Cohen, *supra* note 4, at 57–58 ("Judicial determinations and negotiated minimum standards are not the only possible measures of current fair use practice; arguably, the more accurate measure of fair use is the daily behavior of ordinary users.").

88 Thanks to Hannibal Travis for pushing me on this point.

89 LAWRENCE LESSIG, FREE CULTURE: HOW BIG MEDIA USES TECHNOLOGY AND THE LAW TO LOCK DOWN CULTURE AND CONTROL CREATIVITY 187 (2004).

90 *See* Burk, *supra* note 9, at 289 ("Automated identification and removal, whether accurate or mistaken, is relatively cheap, whereas legal and institutional engagement is comparatively expensive.").

91 *See id.* at 288 ("Risk averse content users, unable to confidently predict the ultimate decision on their activities, may forgo some socially beneficial uses."); Elkin-Koren, *supra* note 40, at 1100 ("The high cost and high risk involved in fair use implementation prevents users from taking advantage of productive uses that can foster copyright goals, simply because they fear liability.").

interpretations of important safeguards and the establishment of ceilings for these safeguards.[92] As they observed:

> We are . . . skeptical . . . about the ability of negotiated [technical] defaults to capture the full range of social benefit that more flexible legal standards allow. While these defaults sometimes might allow access that would exceed fair use under a judicial determination, the "safe harbor" concept is more likely to tend toward a minimalist view of fair use. We suspect that copyright holders would be willing to concede fair use in only a small fraction of the situations that would constitute fair use—indeed, it was just such insistence upon minimalist guidelines by rights holders that led to the collapse of the [Conference on Fair Use] discussions. Moreover, in the case of the 1976 "safe harbor" guidelines for educational copying, rights holders, content users, and even courts have shown a deplorable tendency to act as though the guidelines defined the outer limits of fair use. To the contrary, such guidelines were intended to delineate fair use minima: a floor rather than a ceiling. We are consequently reluctant to recommend an infrastructure based solely on the design of similar defaults into self-enforcing "lock-out" systems for fear that the "ceiling" effect could be even more pernicious.[93]

Thus, whether algorithmic deployment can benefit users will depend on whether the algorithms involved have been properly designed.

In sum, like the existence of arguments asserting why algorithms cannot be satisfactorily deployed to promote fair use, there are also many arguments supporting such deployment. Whether algorithms should be deployed to a greater extent may ultimately depend on a cost-benefit analysis that weighs the strengths of automated fair use systems against their weaknesses. Such analysis will likely vary, depending on whether the analysis is conducted from the perspective of the copyright holder, the user, the technology platform, or another stakeholder in the copyright system.

---

92 Burk & Cohen, *supra* note 4, at 57; *see also* Elkin-Koren, *supra* note 40, at 1096 ("The main concern is that reducing the four-factor analysis into a simplistic and somewhat rigid set of algorithmic instructions might cause some important aspects of fair use analysis to get lost along the way.").

93 Burk & Cohen, *supra* note 4, at 57.

IV. ENABLING ENVIRONMENT

Given the existence of these strengths, policymakers and commentators will need to pay increased attention to ways that could help develop an environment to support the greater deployment of automated fair use systems.[94] Such increased attention is important for at least three reasons. First, the copyright industries and technology platforms may choose to make greater algorithmic deployment even when they are well aware of the many documented weaknesses of automated copyright enforcement.[95] They would do so either because they disagree with the skeptics or because they see the benefits of automation outweighing its costs. Second, the more supportive this environment is, the greater the benefits of such deployment will become. Such growing support would therefore not only tip the balance of the cost-benefit analysis toward greater automation but would also provide an "enabling environment" to facilitate increased algorithmic deployment to promote fair use in copyright law.[96] Third, preparation is well advised because changes in this area are likely to be fast and sudden, similar to how quickly digital technology disrupted the business models of brick-and-mortar companies two decades ago.[97] While Erik Brynjolfsson and Andrew McAfee used the phrase "[g]radually and then suddenly" to describe the changes that have been ushered in by what they coined "the Second Machine Age,"[98] Lee

94 For those policymakers and commentators who find greater automation undesirable, the establishment of this supporting environment would make thing worse. Nevertheless, this Article takes the view that greater automation is not only beneficial but also inevitable. Even if fair use is not automated to a greater extent, copyright enforcement will be. If we are to retain, or restore, the balance of the copyright system, greater algorithmic deployment will be highly sensible.

95 *See supra* note 45 and accompanying text.

96 *Cf.* Peter K. Yu, *Intellectual Property, Economic Development, and the China Puzzle*, *in* INTELLECTUAL PROPERTY, TRADE AND DEVELOPMENT: STRATEGIES TO OPTIMIZE ECONOMIC DEVELOPMENT IN A TRIPS PLUS ERA 173, 213–16 (Daniel J. Gervais ed., 1st ed. 2007) (discussing the importance of an enabling environment for effective intellectual property protection); Peter K. Yu, *Enforcement, Enforcement, What Enforcement?*, 52 IDEA 239, 265–68 (2012) (criticizing the Anti-Counterfeiting Trade Agreement for its failure to create the enabling environment needed to foster effective enforcement of intellectual property rights).

97 Napster is frequently noted for its disruption to the music industry's traditional business model. *See generally* Raymond Shih Ray Ku, *The Creative Destruction of Copyright: Napster and the New Economics of Digital Technology*, 69 U. CHI. L. REV. 263 (2002) (discussing the disruption caused by Napster and other digital distribution technologies and the creative destruction of copyright). For book-length treatments documenting the adverse impact of digital distribution on the music industry, see generally GREG KOT, RIPPED: HOW THE WIRED GENERATION REVOLUTIONIZED MUSIC (2009); ROBERT LEVINE, FREE RIDE: HOW DIGITAL PARASITES ARE DESTROYING THE CULTURE BUSINESS, AND HOW THE CULTURE BUSINESS CAN FIGHT BACK (2011).

98 ERIK BRYNJOLFSSON & ANDREW MCAFEE, THE SECOND MACHINE AGE: WORK, PROGRESS, AND PROSPERITY IN A TIME OF BRILLIANT TECHNOLOGIES 20 (2014). As the authors observed: "Progress on some of the oldest and toughest challenges associated with computers, robots, and other digital gear

Kai-Fu lamented that "time is one thing that the AI revolution is not inclined to grant us."[99]

To create an environment that will support and enable the greater deployment of algorithms to promote fair use in copyright law, policymakers and commentators should focus attention on two distinct types of needs: the need for preparation for change and the need for support. This Part discusses each need in turn. With respect to the latter, it further offers suggestions on what complementary measures policymakers should introduce.

A. Need for Preparation for Change

1. Legal Practices

The first set of changes for which policymakers and commentators will have to prepare relates to legal practices—or, to be more precise, fair use practices. The current fair use system is based on precision found in ex post decisions. If a user wants to find out whether copyright law permits his or her use of a copyrighted work, that user will have to go to a court. Should the highest court in the country, such as the United States Supreme Court, decide that the use is fair, the legal inquiry will end. Although policymakers, commentators, and industry representatives have criticized fair use decisions for being unclear and unpredictable,[100] the case-by-case analysis provided by courts does indicate, with sufficient clarity and predictability, whether copyright law permits a specific use.

When fair use determinations are made by algorithms, however, precision will have to give way to high probability[101]—a trade-off that would

---

was gradual for a long time. Then in the past few years it became sudden; digital gear started racing ahead, accomplishing tasks it had always been lousy at and displaying skills it was not supposed to acquire anytime soon." *Id.*

99 LEE, *supra* note 65, at 152.

100 As the Australian Law Reform Commission observed:

> The opponents of fair use have pointed to research indicating that the outcome of fair use cases is unpredictable. The outcome of litigation is never completely predictable—if it were, the parties would not have commenced litigation, or would likely have settled. This is also true of recent litigation over the fair dealing exceptions and specific exceptions.

AUSTL. LAW REFORM COMM'N, COPYRIGHT AND THE DIGITAL ECONOMY: FINAL REPORT 115 (2013). *But see* Pamela Samuelson, *Unbundling Fair Uses*, 77 FORDHAM L. REV. 2537, 2542 & n.28 (2009) ("If one analyzes putative fair uses in light of cases previously decided in the same policy cluster, it is generally possible to predict whether a use is likely to be fair or unfair. The only clusters of fair use cases in which it is quite difficult to predict whether uses are likely to be fair is in the educational and research use clusters where judges have tended to take starkly different perspectives on fair use defenses in these settings . . . .").

101 *See* Elkin-Koren, *supra* note 40, at 1099 ("AI systems do not decide fair use, but simply generate a score that reflects the probability of fair use."); *see also* VIKTOR MAYER-SCHÖNBERGER &

make many lawyers uncomfortable. As noted earlier, the current state of technology does not allow us to have a "judge on a chip."[102] As a result, algorithms are incapable of making precise determinations of what the law would or would not permit, unless the use in question is identical, or virtually identical, to the use in a previously adjudicated case. Notwithstanding this shortcoming, algorithms may be able to determine, with high probability, whether the law would permit such a use. Such a determination will become even more accurate as automated fair use systems take better advantage of big data analysis and machine learning capabilities.

To the extent that policymakers and commentators are comfortable with the change in fair use practices from precision to high probability, the deployment of algorithms to promote fair use will receive wider acceptance. By contrast, if they remain uncomfortable with this change, they will likely discourage greater algorithmic deployment. Thus far, there has been little research on whether precision-based fair use analysis will better promote creativity than fair use analysis that has attained high probability.[103] Moreover, the limited number of adjudicated fair use cases suggests that most fair use determinations found in the creative environment are made ex ante outside the courtroom.[104] These decisions are therefore based more on probability than on precision. If so, the development of automated fair use systems that attain high probability may be more appealing than it sounds.

2. Creative Practices

The second set of changes for which policymakers and commentators will have to prepare pertains to creative practices. As Part II.B has noted, the greater deployment of fair use algorithms may cause users to change their creative choices and practices. Such behavioral changes have indeed been a key concern of Professor Burk in his latest article on algorithmic fair use.[105] As he summarized his concern, "[a]ttempting to incorporate fair use into enforcement algorithms threatens to degrade the exception into an

KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK 32–72 (2013) (discussing the trade-offs between exactitude and messiness and between causality and correlation).

102 *See* discussion *supra* Part II.A.

103 The closest research in this area concerns the differences between court decisions and ordinary fair use practices. *See, e.g.*, ASS'N OF INDEP. VIDEO & FILMMAKERS ET AL., DOCUMENTARY FILMMAKERS' STATEMENT OF BEST PRACTICES IN FAIR USE (2005), http://archive.cmsimpact.org/sites/default/files/fair_use_final.pdf (stating the best practices in fair use for documentary filmmakers).

104 *See* discussion *supra* Part III.D.

105 Burk, *supra* note 9.

unrecognizable form. Worse yet, social internalization of a bowdlerized version of fair use deployed in algorithmic format is likely to become the new legal and social norm."[106]

Notwithstanding Professor Burk's concern, to which I am sympathetic, we have seen creative practices changing in response to new laws or judicial decisions, such as in the area of appropriation art[107] and digital sampling.[108] While one could certainly debate whether those changes promote or hurt creativity, changes to creative practices are inevitable whenever legal decisions are made—whether by humans or machines. The key question about automated fair use systems is therefore not whether these systems will make decisions, but whether they will make worse decisions, or make worse decisions more frequently.

Moreover, the ever-growing utilization of algorithm-driven copyright enforcement in technology platforms suggests that creative choices and practices will change regardless of whether automated fair use systems are deployed or not. In fact, such deployment may help offset the excesses found in automated copyright enforcement, thereby providing a better balance to the copyright system and instilling a sense of fairness that induces law-abiding behavior.[109] As Professor Elkin-Koren noted emphatically in her Nimmer Memorial Lecture: "Fair use by design has become a necessity in an era of algorithmic governance. The need to develop such tools is necessary in order to tilt the copyright balance back to its origin in our robo notice environment."[110] Likewise, Dan Burk observed: "[I]t may seem desirable to incorporate context-specific fair use metrics into copyright-policing algorithms, both to protect against automated overdeterrence and to inform users of their compliance with copyright law."[111]

---

106 *Id.* at 306.

107 For discussions of appropriation art, see generally Richard H. Chused, *The Legal Culture of Appropriation Art: The Future of Copyright in the Remix Age*, 17 TUL. J. TECH. & INTELL. PROP. 163 (2014); Marci A. Hamilton, *Appropriation Art and the Imminent Decline in Authorial Control over Copyrighted Works*, 42 J. COPYRIGHT SOC'Y U.S.A. 93 (1994); William M. Landes, *Copyright, Borrowed Images, and Appropriation Art: An Economic Approach*, 9 GEO. MASON L. REV. 1 (2000); Niels Schaumann, *Fair Use and Appropriation Art*, 6 CYBARIS INTELL. PROP. L. REV. 112 (2015); David Tan, *The Transformative Use Doctrine and Fair Dealing in Singapore: Understanding the Purpose and Character of Appropriation Art*, 24 SING. ACAD. L.J. 832 (2012).

108 For discussions of digital sampling, see generally JOANNA TERESA DEMERS, STEAL THIS MUSIC: HOW INTELLECTUAL PROPERTY LAW AFFECTS MUSICAL CREATIVITY 71–110 (2006); KEMBREW MCLEOD & PETER DICOLA, CREATIVE LICENSE: THE LAW AND CULTURE OF DIGITAL SAMPLING (2011).

109 *See* Armstrong, *supra* note 4, at 109 ("Empowering users to exercise their fair use rights without violating the DMCA might . . . increase law-abiding behavior and temper the critical evaluation of the DMCA as a one-sided giveaway to powerful producer cartels." (footnote omitted)).

110 Elkin-Koren, *supra* note 40, at 1100.

111 Burk, *supra* note 9, at 284–85.

Finally, although the use of automated fair use systems will likely affect, or even curtail, select creative practices, one would expect new creative practices to be developed in response to the greater deployment of these systems. Whether those new practices will be better or worse from a creative standpoint is difficult to judge. Moreover, social practices, including creative practices, are constantly defined and redefined through their interactions with technology, and vice versa. Just as automated fair use systems will shape creative practices, the changing creative practices will also shape those systems.[112] In the age of artificial intelligence, in which machines are constantly learning and improving, there is a good chance that the development of new creative practices and the availability of more information for big data analysis would lead to modifications that would allow algorithms to better promote creativity.

B. Need for Support

1. Judicial Support

Apart from the increased preparedness for the two sets of changes mentioned above, policymakers will need to introduce complementary measures to support automated fair use systems if these systems are to be developed. The first type of measures concern judicial support. Even with the greater use of automated systems, courts will remain highly important. As Dan Burk and Julie Cohen observed, the inclusion of "human involvement"

---

112 Tarleton Gillespie noted the entanglement between algorithms and social practices:

> [W]e must consider not [the algorithms'] "effect" on people, but a multidimensional "entanglement" between algorithms put into practice and the social tactics of users who take them up. This relationship is, of course, a moving target, because algorithms change, and the user populations and activities they encounter change as well. Still, this should not imply that there is no relationship. As these algorithms nestle into people's daily lives and mundane information practices, users shape and rearticulate the algorithms they encounter; and algorithms impinge on how people seek information, how they perceive and think about the contours of knowledge, and how they understand themselves in and through public discourse.
>
> It is important that we conceive of this entanglement not as a one-directional influence, but as a recursive loop between the calculations of the algorithm and the "calculations" of people. The algorithm that helps users navigate Flickr's photo archive is built on the archive of photos posted, which means it is designed to apprehend and reflect back the choices made by photographers. What people do and do not photograph is already a kind of calculation, though one that is historical, multivalent, contingent, and sociologically informed. But these were not Flickr's only design impulses; sensitivity to photographic practices had to compete with cost, technical efficiency, legal obligation, and business imperatives. And the population of Flickr users and the types of photos they post changed as the site grew in popularity, was forced to compete with Facebook, introduced tiered pricing, was bought by Yahoo, and so forth.

Gillespie, *supra* note 24, at 183.

will permit the consideration of "a greater level of complexity in the circumstances."[113] In a proposal advanced more than a decade ago, I also advocated the "technology first, courts later" approach that would enable courts (and human actors) to intervene when needed.[114] Although automation enhances efficiency and effectiveness, human intervention can be highly beneficial. To some extent, the debate about the need for such intervention ties to the ongoing debate about whether machines can perform as well as lawyers or judges.[115]

Should we end up choosing to develop automated fair use systems that include judicial intervention at the end of the process, the determinations made by these systems should be viewed as interim, rather than final.[116] In effect, such systems will provide technically driven safe harbors for users until courts step in to make adjustments. Given the potential, and likely constant, modification of these safe harbors, one inevitably will wonder what would happen should a court find out months, or years, later that an earlier automated fair use determination was incorrect.[117]

While it is impossible to go back in time, automated fair use systems can be utilized to facilitate compensation. Consider, for example, a system that has been built into a technology platform that disseminates uploaded

---

113 Burk & Cohen, *supra* note 4, at 75; *see also* Dan L. Burk, *Legal and Technical Standards in Digital Rights Management Technology*, 74 FORDHAM L. REV. 537, 551 (2005) ("[T]echnological controls tend to be relatively blunt instruments for control of digital content, unable to accommodate copyright fair use without the re-introduction of human discretion.").

114 *See* Yu, *supra* note 4, at 73 ("[A] two-step approach—technology first, then courts—seems to be the best compromise we can have today, and it is worth considering developing such a system as we explore the next generation of [digital rights management] systems."). Niva Elkin-Koren outlined a similar approach: "Algorithmic fair use could . . . involve a two-tier review. First, algorithmic screening would be performed and second, for cases which were flagged by the system, but were inconclusive, human review would be conducted." Elkin-Koren, *supra* note 40, at 1098.

115 For discussions in this area, see generally Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L.J. 1147 (2017); Crootof, *supra* note 56; Milan Markovic, *Rise of the Robot Lawyers?*, 61 ARIZ. L. REV. 325 (2019); Andrew C. Michaels, *Artificial Intelligence, Legal Change, and Separation of Powers*, 88 U. CIN. L. REV. 1083 (2020); Frank Pasquale, *A Rule of Persons, Not Machines: The Limits of Legal Automation*, 87 GEO. WASH. L. REV. 1 (2019); Richard M. Re & Alicia Solow-Niederman, *Developing Artificially Intelligent Justice*, 22 STAN. TECH. L. REV. 242 (2019); Dana Remus & Frank Levy, *Can Robots Be Lawyers: Computers, Lawyers, and the Practice of Law*, 30 GEO. J. LEGAL ETHICS 501 (2017); Harry Surden, *Machine Learning and Law*, 89 WASH. L. REV. 87 (2014); Eugene Volokh, *Chief Justice Robots*, 68 DUKE L.J. 1135 (2019); Wu, *supra* note 79. For earlier discussions in this area (collected in Professor Volokh's article), see generally Bruce G. Buchanan & Thomas E. Headrick, *Some Speculation About Artificial Intelligence and Legal Reasoning*, 23 STAN. L. REV. 40 (1970); Anthony D'Amato, *Can/Should Computers Replace Judges?*, 11 GA. L. REV. 1277 (1977); L. Thorne McCarty, *Reflections on "Taxman": An Experiment in Artificial Intelligence and Legal Reasoning*, 90 HARV. L. REV. 837 (1977).

116 *See* Burk, *supra* note 9, at 297 ("Patterns detected by a machine evaluating fair use-related data should not be confused with a legal institutional determination of fair use.").

117 Thanks to Martin Senftleben for asking this important question.

content, similar to YouTube's Content ID system. That system could be easily designed to allow for dissemination if the probability of fair use is over seventy-five percent but prohibit such dissemination if the probability falls below fifty percent. For probability that is in between, the system could further require the revenue stream to be put in escrow for a certain period of time—say, six months or a year.

Should infringement be found later, the court could require the user to reverse the revenue stream based on what the system has documented—including, where applicable, that the technology platform releases the revenue in escrow to the copyright holder. The court could also grant injunctive relief, as the law currently allows.[118] Should the copyright holder prefer injunctive relief over compensation, he or she could file a complaint in court as soon as the alleged infringement has been discovered. Should the copyright holder choose to tolerate such infringement, he or she would still have the option to seek compensation and injunctive relief later, as long as the statute of limitations had not been tolled.[119]

Obviously, the automated fair use system proposed here could be calibrated very differently, depending on legislative or policy preferences. For instance, the seventy-five percent threshold could be easily adjusted upward to eighty or ninety percent or downward to 66.7 percent. Such an adjustment will affect the creative space available to users. In addition, even though the proposal anticipates that only courts will grant injunctive relief, that type of relief could be easily built into the system, similar to how YouTube's Content ID system allows copyright holders to block the use of a specific copyrighted work. The key takeaway of this proposal is not its fine details, but that algorithms can be carefully designed and tailored to accommodate a wide variety of preferences and situations. We should not have a simplistic assumption that automated fair use systems can, or will, make only binary determinations—for example, whether the use is fair or not.

2. Technological Support

The second type of complementary measures relates to technological support. If automated fair use systems are to be developed, the algorithms involved have to be trusted by the different stakeholders in the copyright system.[120] As a result, impartiality has to be built into the algorithms from the

118 *See* 17 U.S.C. § 502 (2018) (providing injunction as a remedy for copyright infringement).

119 *See* 17 U.S.C. § 507(b) ("No civil action shall be maintained under the provisions of this title unless it is commenced within three years after the claim accrued.").

120 As a National Research Council study reminded us:

very beginning. To do so, it would be important to set up a neutral and representative body that would supervise the development of fair use algorithms.[121] These algorithms could not be designed solely by the copyright industries that want to maximize enforcement, those technology platforms that seek to avoid copyright lawsuits and legal liability, or a combination of these two groups of players.[122]

Considering the likely existence of a wide variety of algorithms that could make automated fair use determinations,[123] a process can be further developed to facilitate the certification of different algorithms that are equally capable of making high-quality decisions. Having such a process is desirable because it will facilitate competition over algorithmic quality.[124] Should there

---

> The debate over intellectual property includes almost everyone, from authors and publishers, to consumers (e.g., the reading, listening, and viewing public), to libraries and educational institutions, to governmental and standards bodies. Each of the stakeholders has a variety of concerns . . . that are at times aligned with those of other stakeholders, and at other times opposed. An individual stakeholder may also play multiple roles with various concerns. At different times, a single individual may be an author, reader, consumer, teacher, or shareholder in publishing or entertainment companies; a member of an editorial board; or an officer of a scholarly society that relies on publishing for revenue. The dominant concern will depend on the part played at the moment.

COMM. ON INTELL. PROP. RIGHTS & THE EMERGING INFO. INFRASTRUCTURE, NAT'L RES. COUNCIL, THE DIGITAL DILEMMA: INTELLECTUAL PROPERTY IN THE INFORMATION AGE 51 (2000); *see also* Mark Stefik, *Shifting the Possible: How Trusted Systems and Digital Property Rights Challenge Us to Rethink Digital Publishing*, 12 BERKELEY TECH. L.J. 137, 156 (1997) (identifying among the stakeholders relevant to a proposed Digital Property Trust "publishers, trusted system vendors, financial institutions, lawmakers, librarians, and consumers"); Yu, *supra* note 4, at 31 (noting that stakeholders in the copyright debate, and the debate on digital rights management systems in particular, cannot be nicely divided into binaries).

121 *See* Yu, *supra* note 4, at 68 ("[W]e need to develop a process that brings together copyright holders, technology developers, consumer advocates, civil libertarians and other stakeholders."); *see also* COHEN, *supra* note 7, at 192 ("Mastering the processes by which technical standards are developed . . . requires . . . new public accountability mechanisms.").

122 *See* Armstrong, *supra* note 4, at 121 ("The more technology reflects only one set of interests, . . . the more it departs from the law, which conceptualizes copyright as a balancing of interests, with the ultimate goal of fostering both creative expression and broad public availability of creative works."); *see also* Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 682 (2017) ("A prejudiced decisionmaker could skew the training data or pick proxies for protected classes with the intent of generating discriminatory results.").

123 *See* AGRAWAL ET AL., *supra* note 79, at 189 ("There is often no single right answer to the question of which is the best AI strategy or the best set of AI tools, because AIs involve trade-offs: more speed, less accuracy; more autonomy, less control; more data, less privacy.").

124 As I noted in a recent article:

> Competition is imperative if society is to develop more efficient, more effective, and less biased algorithms. Such competition is particularly needed when algorithmic choices are increasingly difficult, or time consuming, to explain. Indeed, without competition, it would be hard to identify problems within an algorithm or to determine whether that algorithm has provided the best solution in light of the existing technological conditions and constraints.

Yu, *supra* note 26 (footnotes omitted); *see also* Peter K. Yu, *Data Producer's Right and the Protection of Machine-Generated Data*, 93 TUL. L. REV. 859, 927 (2019) (noting that competition law is "a critical area be disagreements between the different algorithms, the "technology first, courts later" approach will allow courts, or the certification body, to step in. Although it would be ideal to eliminate all disagreements, for the sake of clarity and predictability, the preference for competition presumes the existence of disagreements and diversity. Moreover, disagreements over fair use interpretations are not uncommon within our current copyright system. In the U.S. federal system, courts at both the horizontal and vertical levels do not always agree on their interpretations of copyright law.[125]

Once developed, automated fair use systems will have to be constantly audited to ensure that the systems remain neutral and the outcomes consistent with existing copyright law.[126] Such constant auditing reflects the best practices advocated by the technology community. Principle 7 of the *ACM Statement on Algorithmic Transparency and Accountability* declared: "Institutions should use rigorous methods to validate their models and document those methods and results. In particular, they should routinely perform tests to assess and determine whether the model generates discriminatory harm."[127] The *FAT/ML Principles for Accountable Algorithms and a Social Impact Statement for Algorithms* also called for impact assessment "(at least) three times during the design and development

relating to data governance"); Annie Lee, Note, *Algorithmic Auditing and Competition Under the CFAA: The Revocation Paradigm of Interpreting Access and Authorization*, 33 BERKELEY TECH. L.J. 1307, 1310 (2018) ("Online competitors . . . promote fair online practices by providing users with a choice between competitive products . . . .").

125 On disagreements at the horizontal level, compare Cambridge Univ. Press v. Patton, 769 F.3d 1232, 1265 (11th Cir. 2014) (noting that "th[e] reasoning is somewhat circular" when the failure to pay a potential licensing fee is used to disprove fair use), with Am. Geophysical Union v. Texaco Inc., 60 F.3d 913, 930–31 (2d Cir. 1994) ("[I]t is not unsound to conclude that the right to seek payment for a particular use tends to become legally cognizable under the fourth fair use factor when the means for paying for such a use is made easier."). On disagreements at the vertical level, see Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd., 269 F. Supp. 2d 1213 (C.D. Cal. 2003), *aff'd*, 380 F.3d 1154 (9th Cir. 2004), *rev'd*, 545 U.S. 913 (2005).

126 As the Center for Democracy and Technology noted:

> Audits are one method to provide explanations and redress without compromising the intellectual property behind the business model. Designing algorithmic systems that can be easily audited increases accountability and provides a framework to standardize best practices across industries. While explanations can help individuals understand algorithmic decision making, audits are necessary for systemic and long-term detection of unfair outcomes. They also make it possible to fix problems when they arise.

Digital Decisions, *supra* note 32, at 11. *But see* Kroll et al., *supra* note 122, at 660–61 (discussing the limits to auditing in the algorithmic context).

127 *ACM Statement*, *supra* note 3.

process: design stage, pre-launch, and post-launch."[128] As Lorna McGregor, Daragh Murray, and Vivian Ng explained:

> During the design and development stage, impact assessments should evaluate how an algorithm is likely to work, ensure that it functions as intended and identify any problematic processes or assumptions. This provides an opportunity to modify the design of an algorithm at an early stage, to build in . . . compliance—including monitoring mechanisms—from the outset, or to halt development if . . . concerns cannot be addressed. Impact assessments should also be conducted at the deployment stage, in order to monitor effects during operation. . . . [T]his requires that, during design and development, the focus should not only be on testing but steps should also be taken to build in effective oversight and monitoring processes that will be able to identify and respond to [problems] once the algorithm is deployed.[129]

While algorithmic audits have attracted the attention of many commentators, training data increasingly drive the performance of algorithms. As a result, these audits have to cover not only the algorithms themselves but also training data and algorithmic outcomes.[130] When learning algorithms are involved, scrutinizing algorithms alone will unlikely reveal the full extent of any problems that the automated fair use systems may encounter. To the extent privacy concerns are raised when algorithmic outcomes are being disclosed for auditing purposes, the audits could focus instead on representative, anonymized samples of different algorithmic

---

128 *Principles for Accountable Algorithms and a Social Impact Statement for Algorithms*, FAT/ML, https://www.fatml.org/resources/principles-for-accountable-algorithms (last visited July 9, 2019). FAT/ML stands for "Fairness, Accountability, and Transparency in Machine Learning." *Id.*

129 Lorna McGregor et al., *International Human Rights Law as a Framework for Algorithmic Accountability*, 68 INT'L & COMP. L.Q. 309, 330 (2019).

130 *See* O'NEIL, *supra* note 35, at 229 ("We have to learn to interrogate our data collection process, not just our algorithms."); Anupam Chander, *The Racist Algorithm?*, 115 MICH. L. REV. 1023, 1024–25 (2017) ("What we need instead is a transparency of inputs and results, which allows us to see that the algorithm is generating discriminatory impact."); Kartik Hosanagar & Vivian Jair, *We Need Transparency in Algorithms, but Too Much Can Backfire*, HARV. BUS. REV. (July 23, 2018), https://hbr.org/2018/07/we-need-transparency-in-algorithms-but-too-much-can-backfire ("[M]achine learning algorithms—and deep learning algorithms in particular—are usually built on just a few hundred lines of code. The algorithms logic is mostly learned from training data and is rarely reflected in its source code. Which is to say, some of today's best-performing algorithms are often the most opaque."); Kroll et al., *supra* note 122, at 641 ("[W]ithout full transparency—including source code, input data, and the full operating environment of the software—even the disclosure of audit logs showing what a program did while it was running provides no guarantee that the disclosed information actually reflects a computer system's behavior."); *see also id.* at 657–60 (discussing the limits to transparency in the algorithmic context).

outcomes, or of those outcomes that were based on test data provided by auditors or consumer advocacy groups.[131]

3. Legislative Support

The final type of complementary measures pertains to legislative support. If the successful development of automated fair use systems requires the collection and big data analysis of additional records that are currently not in the possession of copyright holders, users, or technology platforms, we will need to introduce legal reforms. We will also need to do so if we are to establish an environment that supports the auditing of automated fair use systems.

As commentators have pointed out, the protection of intellectual property rights has posed significant barriers to both big data analysis and algorithmic audits. Examples range from copyright protection that prevents the mining of text and data that can be used for fair use determinations[132] to

---

131 *See* Yu, *supra* note 26 ("[T]echnology developers could provide a representative, anonymized sample of the different algorithmic outcomes to enable the public to determine for itself the satisfactoriness of algorithm-enhanced technological products and services.").

132 As Amanda Levendowski observed:

> Copyright law causes friction that limits access to training data and restricts who can use certain data. This friction is a significant contributor to biased AI. The friction caused by copyright law encourages AI creators to use biased, low-friction data . . . for training AI systems, like the word2vec toolkit, despite those demonstrable biases. As Google's decision not to freely release the Google News corpus reveals, copyright law can also curtail the implementation of bias mitigation techniques, including interventions like reweighting algorithmic inputs or supplementing datasets with additional data. Copyright law can even preclude potential competitors from converting the customers of dominant AI players. . . .
>
> Good training data is crucial for creating accurate AI systems. The AI system tasked with identifying cats must be able [to] abstract out the right features, or heuristics, of a cat from training data. To do so, the training data must be well-selected by humans—training data infused with implicit bias can result in skewed datasets that fuel both false positives and false negatives. For example, a dataset that features only cats with tortoiseshell markings runs the risk that the AI system will "learn" that a mélange of black, orange, and cream markings [is] a heuristic for identifying a cat and mistakenly identify other creatures, like brindle-colored dogs, as cats. Similarly, a dataset that features only mainstream domestic cats could create an AI system that "learns" that cats have fluffy fur, pointy ears, and long tails and fail to identify cats of outlier breeds, like a Devon Rex, Scottish Fold, or Manx. And, in both examples, all manner of wildcats are excluded from the training data.

Levendowski, *supra* note 31, at 589–92 (footnotes omitted); *see also* Lim, *supra* note 2, at 847–55 (noting the need for expanding fair use to promote artificial intelligence); Benjamin L.W. Sobel, *Artificial Intelligence's Fair Use Crisis*, 41 COLUM. J.L. & ARTS 45, 61–79 (2017) (identifying the potential legal liability for copyright infringement when copyrighted works are used to train intelligent machines). While fair use may need to be expanded to promote artificial intelligence, the non-protection of copyrighted works generated by artificial intelligence may provide an ever-growing trove of helpful training data that reside in the public domain. *See* U.S. COPYRIGHT OFFICE, COMPENDIUM OF U.S. COPYRIGHT OFFICE PRACTICES § 313.2 (3d ed. 2017) ("[The U.S. Copyright] Office will not register works produced by a machine or mere mechanical process that operates randomly or automatically without any creative input trade secret protection that prevents the inspection of copyrighted algorithms and their protected input, training, and feedback data.[133] Outside the intellectual property arena, statutes such as the Computer Fraud and Abuse Act[134] also make it difficult for independent auditors to access algorithms. In addition, privacy protection, such as the protection found in the EU General Data Protection Regulation,[135] could greatly reduce the automated systems' abilities to access and collect personally identifying records that are needed for fair use determinations.

Given these barriers and continuous challenges, policymakers will need to introduce complementary legal reforms if algorithms are to be deployed to

---

or intervention from a human author."); Clark D. Asay, *Independent Creation in a World of AI*, 14 FIU L. REV. 201 (2020) (noting the potentially differing second factor analysis in the fair use determination of machine-generated works). *But see* Grimmelmann, *There's No Such Thing*, *supra* note 2, at 403 ("[N]o one has ever exhibited even one work that could plausibly claim to have a computer for an 'author' in the sense that the Copyright Act uses the term."). For discussions of the importance of exceptions for text and data mining to the copyright systems, see generally Christophe Geiger et al., *Crafting a Text and Data Mining Exception for Machine Learning and Big Data in the Digital Single Market*, *in* INTELLECTUAL PROPERTY AND DIGITAL TRADE IN THE AGE OF ARTIFICIAL INTELLIGENCE AND BIG DATA 95 (Xavier Seuba et al. eds., 2018); Michael W. Carroll, *Copyright and the Progress of Science: Why Text and Data Mining Is Lawful*, 53 U.C. DAVIS L. REV. 893 (2019); Matthew Sag, *The New Legal Landscape for Text Mining and Machine Learning*, 66 J. COPYRIGHT SOC'Y U.S.A. 291 (2019).

133 *See* Burk, *supra* note 9, at 301 (noting the "the explicit or intentional obscurity stemming from trade secrecy and protection of confidential business information—to the extent that algorithms are commissioned or developed by commercial entities, they may attempt to shield proprietary aspects of the technology from misappropriation or competitive copying"). For discussions of the tension between transparency and disclosure on the one hand and trade secret protection on the other, see generally Sonia K. Katyal, *The Paradox of Source Code Secrecy*, 104 CORNELL L. REV. 1183 (2019); Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343 (2018).

134 18 U.S.C. § 1030 (2018). As Sonia Katyal observed:

> [T]he ACLU sued on behalf of four researchers who maintained that [this statute] actually prevented them from scraping data from sites, or from creating fake profiles to investigate whether algorithmic discrimination led some employment and real estate sites to fail to display certain listings on the basis of race or gender. The concern was that the law permitted researchers to be held criminally accountable because the research might involve violating one of the sites' Terms of Service, something that could carry both prison and fines. As one researcher observed, these laws have the perverse effect of "protecting data-driven commercial systems from even the most basic external analysis."

Katyal, *supra* note 32, at 122 (footnotes omitted); *see also* Levendowski, *supra* note 31, at 587–88 ("Some courts have interpreted the [Computer Fraud and Abuse Act] as prohibiting violation of an employer's computer-use policies or a website's Terms of Service, which can chill algorithmic accountability testing, including digital auditing used to uncover racial discrimination." (footnote omitted)); Lee, *supra* note 124, at 1311–38 (discussing how the Computer Fraud and Abuse Act has discouraged algorithmic auditors from exposing questionable business practices and has fostered a hostile market for legitimate competitors).

135 Council Regulation 2016/679, *supra* note 33.

a greater extent to promote fair use in copyright law.[136] Obviously, fair use is only part of the copyright system—and an even smaller part of the overall legal system. When considering what reform to introduce, policymakers will inevitably have to balance the different interests, preferences, and policy objectives to determine how best to facilitate the development of automated fair use systems.

V. CONCLUSION

This Article has documented the case for and against the greater deployment of algorithms to promote fair use in copyright law. If policymakers are eager to develop automated fair use systems, they will need to be prepared for changes in both legal and creative practices that will be brought about by greater algorithmic deployment. They will also need to be willing to introduce complementary measures to build a supportive environment that will enable and facilitate automated fair use determinations. Until we are ready for these changes and until that supportive environment can be built, however, it is understandable why many policymakers and commentators have remained skeptical of the satisfactory deployment of algorithms to promote fair use in copyright law.

---

136 *See* Yu, *supra* note 26 ("To provide support for external audits that do not involve regulatory authorities, adjustments will have to be made to those laws that have posed barriers to external reviews of source code and computer systems, such as the Computer Fraud and Abuse Act and the Digital Millennium Copyright Act." (footnotes omitted)).