

Will Algorithms Take the Fairness Out of Fair Use?

Author : Christopher J. Sprigman

Date : January 15, 2018

Dan L. Burk, *Algorithmic Fair Use*, **U. Chi. L. Rev.** (forthcoming), available at [SSRN](#).

As part of an invited symposium, organized by *The University of Chicago Law Review*, on whether artificial intelligence will spell the end of one-size-fits-all laws, Dan Burk has written a terrific essay explaining why he's skeptical that AI or machine learning will lead to better copyright fair use decisions. In the essay, [Algorithmic Fair Use](#), Professor Burk identifies three main bases for his concerns.

First, Professor Burk is skeptical that American fair use law, which is articulated as a relatively open-textured standard (as compared with U.K.-style "fair dealing" provisions that set out a laundry list of tightly specified circumstances in which portions of a copyrighted work may be used without permission), could ever be reproduced with much fidelity as a set of software rules. The resistance of American fair use to ruleification, and therefore to automation, runs deep – not least because the content of any fair use rule depends upon antecedent considerations that are themselves resistant to distillation into rules:

Determining the impact of the unauthorized use of a work on the actual or potential market for the underlying work requires a model of the market and decisions about the data that properly populate that model. The amount of the work used can be mapped to the percentage of lines or words or pixels or bits taken for a given use, but some weight or significance must be accorded that number, whether defined by explicit programming values or by algorithmically learned data patterns. The type of work used and the use to which the protected taking is put require some categorization of works and uses. These and a multitude of other design choices made in advance would determine the allowance or disallowance of uses for protected content; algorithms do not make judgments; they are rather the products of human judgment.

Second, and perhaps more importantly given the direction of technology at the moment, Professor Burk is skeptical of fair use automation through machine learning. Machine learning does not depend on ruleification but would instead seek to determine whether a use was fair by comparing it to patterns that correlate with uses judged to be fair within a large dataset of unauthorized uses. But a machine's ability to produce relevant and reliable results through pattern matching presumes that the machine has been fed a dataset that is representative of the fair use determinations, and the facts underlying them, as they exist out in the world.

Getting the dataset right, Professor Burk argues, is likely to be expensive and difficult. But the problem runs deeper than just cost or the technical difficulties of assembling a reliable dataset. The fundamental conceptual difficulty is that the output of a machine learning algorithm is just a correlation. It isn't a judgment about whether that correlation is meaningful. In an entertaining but important aside, Professor Burk refers to a famous instance where data mining showed a strong correlation between movements in the S&P 500 stock index and the production of butter in Bangladesh. In that case, he notes, "a human decisionmaker is required to designate the trend as spurious rather than meaningful." The same would be true of fair use determinations made by a machine learning algorithm – human intervention would be required to check that the machine's output makes any sense outside the confines of the machine's dataset.

Third, and finally, Professor Burk is wary of proposals to automate fair use because he fears that encoding fair use into the operation of machines will shape human expectations and behavior in ways that are both difficult to predict in

advance or to contest ex post. He outlines this reservation in part by quoting from a video creator's writings describing how Google's Content ID system has shaped creativity on YouTube:

"You could make a video that meets the criteria for fair use, but YouTube could still take it down because of their internal system (Copyright ID) [sic] which analyzes and detects copyrighted material. So I learned to edit my way around that system. Nearly every stylistic decision you see about the channel — the length of the clips, the number of examples, which studios' films we chose, the way narration and clip audio weave together, the reordering and flipping of shots, the remixing of 5.1 audio, the rhythm and pacing of the overall video — all of that was reverse engineered from YouTube's Copyright ID. I spent about a week doing brute force trial-and-error. I would privately upload several different essay clips, then see which got flagged and which didn't. This gave me a rough idea what the system could detect, and I edited the videos to avoid those potholes."¹

Of course, machines are not the only mechanism for shaping behavior. That's what law does too; indeed, that is the very point of having laws. An advantage of the conventional legal system is that when laws and legal reasoning are more readily accessible and comprehensible, they are more easily contested. The inscrutable outputs of trade-secret-protected algorithms or invisible data sets, by contrast, are likely to obscure the ways in which law shapes behavior. In the end, Professor Burk is profoundly pessimistic: "[I]mplementation of algorithmic fair use," he says, "will inevitably, and probably detrimentally, change the nature of fair use."

I am not so sure that we know enough yet to judge whether Professor Burk's intuition is right. It does seem likely that automation will create pressure to "ruleify" fair use, that is, to turn it into a more elaborated version of U.K.-style fair dealing. But what is our normative takeaway if that happens? Is ruleified fair use, where enforcement is done cheaply by machines, necessarily worse than our current fair use standard?

Current American fair use law is more flexible than any set of imaginable fair use rules, yet (in part because of that flexibility) enforcement is expensive and undertaken only in the comparatively rare occasions where a user has both the incentive and means to engage in federal court litigation. Thus, fair use as we know it in the U.S. is flexible, but inaccessible.

Ruleified fair use administered by machines promises to solve the accessibility problem. But will that gain come only at the expense of a bowdlerized set of fair use rules? That depends in part on who would be making the rules that automation demands, and what the process looks like for creating new rules. Would the rule-maker be open to input from users as well as content owners? And would the rule-maker be obliged to periodically revisit the rules to make sure that new exceptions could be added as needed, and exceptions that had proved ill-advised removed?

These are among the important questions that Professor Burk's provocative essay raises, and they should command the attention of the copyright academy in the years to come.

1. Tony Zhou, *Postmortem: Every Frame a Painting*, **Medium** (Dec. 2, 2017), <https://medium.com/@tonyszhou/postmortem-1b338537fab0>.

Cite as: Christopher J. Sprigman, *Will Algorithms Take the Fairness Out of Fair Use?*, JOTWELL (January 15, 2018) (reviewing Dan L. Burk, *Algorithmic Fair Use*, **U. Chi. L. Rev.** (forthcoming), available at SSRN), <https://ip.jotwell.com/will-algorithms-take-fairness-fair-use/>.