

NEW YORK UNIVERSITY
JOURNAL OF INTELLECTUAL PROPERTY
AND ENTERTAINMENT LAW

VOLUME 4

WINTER 2014

NUMBER 1

MEDIA LAW AND COPYRIGHT IMPLICATIONS OF
AUTOMATED JOURNALISM

LIN WEEKS*

New computer algorithms are capable of taking data-heavy input, such as stock prices, and outputting paragraphs of prose that read like traditional journalism stories. Such programs have recently been adopted by top media outlets, including Bloomberg and the Associated Press, for generating stories in some contexts. Given that such technology is a departure from traditional print media reporting, it is unclear whether its implications are adequately addressed by standing jurisprudence in the realms of mass media and copyright law. In general, those areas of law for which courts have developed strong abstract or conceptual frameworks are more easily applied to new technologies.

* J.D. Candidate, New York University School of Law, 2015; B.A. Economics, University of Wisconsin-Madison, 2012. Special thanks to Professor David McCraw, Assistant General Counsel, *New York Times*, and Leah Rosenbaum, NYU JIPEL Senior Notes Editor 2013–2014, for their thoughtful editing and feedback.

INTRODUCTION	68
I. INTRODUCTION TO THE TECHNOLOGY	70
A. <i>Definitions and Terminology</i>	71
B. <i>Database Journalism</i>	71
C. <i>The Business of Automated Journalism</i>	72
II. AUTOMATED JOURNALISM AND THE FIRST AMENDMENT	74
A. <i>A Three-Dimensional Problem:</i> <i>Theories of Protection, Manner of Restraint, Type of Output</i>	74
B. <i>Media Law</i>	77
1. <i>Adjudicated Issues in Media Law:</i> <i>Implications of Brown for Automated Journalism</i>	77
2. <i>Open Questions in Media Law: Defamation</i>	79
i. How should actual malice or negligence be determined in defamation cases arising from animated journalism articles?	80
ii. How will Section 230 be applied to the new technology?.....	84
C. <i>Copyright</i>	85
1. <i>Adjudicated Issues in Copyright: Protection for Algorithms</i>	86
2. <i>Open Questions in Copyright: Forms, Fair Use, and Authorship</i>	87
i. What protection should be given to the spreadsheet used to create automated journalism stories?.....	87
ii. Who will courts favor in an authorship dispute?	91
CONCLUSION.....	93

INTRODUCTION

The following paragraphs appeared in a story on Forbes.com on May 22, 2014:

Workday is expected to book a wider loss than a year ago when it reports first-quarter earnings on Tuesday, May 27, 2014. Analysts are expecting a loss of 28 cents per share, down from a loss of 20 cents per share a year ago.

The consensus estimate is down from three months ago when it was a loss of 26 cents, but is unchanged over the past month. For the fiscal year, analysts are projecting a loss of \$1.15 per share. Revenue is projected to be 66% above the year-earlier total of \$91.6 million at

\$152.4 million for the quarter. For the year, revenue is expected to come in at \$735.4 million.¹

In a critique of the write-up, one might note a lack of conclusions about the information presented. This could be intentional: Forbes may believe investors should draw their own conclusions from the simple, informational story. From a stylistic standpoint, the prose is choppy, but technically sound; it is unclear whether the writer has spoken to the “analysts” referenced in the story or just noted their opinions online or in print. A more perceptive reader might notice the striking similarity to another story published an hour later on the same day:

Aeropostale is expected to book a wider loss than a year ago when it reports first-quarter earnings on Thursday, May 22, 2014. Analysts are expecting a loss of 72 cents per share, down from a loss of 16 cents per share a year ago.

The consensus estimate remains unchanged over the past month, but it has decreased from three months ago when it was a loss of 17 cents. For the fiscal year, analysts are expecting a loss of \$1.75 per share. A year after being \$452.3 million, analysts expect revenue to fall 9% year-over-year to \$409.9 million for the quarter. For the year, revenue is projected to come in at \$1.94 billion.²

Now what looked like curt but useful prose looks like formulaic, if not lazy, writing. Questions of self-plagiarism arise; it appears that the article is written from a template. A close examination reveals that the phrasing of the second paragraph of each piece is slightly different, and that the syntax of the second and third sentences of that paragraph is flipped in a similar fashion. At the very least, one could feel reasonably comfortable in the assumption that the two articles were either authored by the same person, or that the latter piece was written by a person free to take liberties with the first writer’s work. But neither article was written by a person at all; both are the product of a computer program.

New software allows computer programs to translate data-heavy content, such as box scores, stock prices, housing starts, and weather reports, into prose that

¹ Narrative Science, *Loss Expected to Widen for Workday*, FORBES (May 22, 2014, 9:00 AM), <http://www.forbes.com/sites/narrativescience/2014/05/22/loss-expected-to-widen-for-workday>.

² Narrative Science, *Loss Expected to Widen for Aeropostale*, FORBES (May 22, 2014, 10:00 AM), <http://www.forbes.com/sites/narrativescience/2014/05/21/loss-expected-to-widen-for-aeropostale>.

reads much like traditional news stories. In the above example, Forbes used software produced by a company called Narrative Science to automatically generate blog posts for its website. These articles are generated and derived from information about the stock market,³ but they include statistic-driven vocabulary that refers directly to the opinions of analysts. Other ventures into automated journalism are ongoing — the Knight Lab at Northwestern University is “working at advancing news media innovation through exploration, experimentation” in the digital realm.⁴

The new technology has several interesting legal implications, specifically in the realms of copyright and media law. Part I of this note further introduces the technology underlying automated journalism, and explores its development, usage, and business applications. Part II examines both the traditional mass media law and copyright-related problems created by the usage of automated journalism programs, including problems affecting input, output, and the algorithm itself. These could include bad input or programming that leads to a falsehood in an automated story and potentially exposes the publisher to liability for defamation. In another scenario, competition between media entities might lead to copyright disputes over the algorithms or output of automated journalism stories.

A theme is present throughout the note: Computer-generated journalism is just one type of information that will be disseminated with increasing frequency as similar technologies are adapted to various ends. The popularity of algorithmic reporting will require courts to more fully and definitively articulate a set of first principles for free speech lest they work case-by-case or see a fractal splintering of decisions in the lower courts. One effect of the relative clarity of copyright’s theoretical underpinnings, in comparison with the more open questions surrounding the First Amendment, will be a more straightforward translation of existing jurisprudence to the new questions presented by automated journalism technology.

I

INTRODUCTION TO THE TECHNOLOGY

The weather report and a listing of stock prices and changes were once pillars of major American newspapers. The practice was rooted in tradition, but it was also deeply practical — people care about things that affect their lives. The

³ Narrative Science, *About*, FORBES, <http://blogs.forbes.com/narrativescience/profile> (July 10, 2014).

⁴ *About Us*, KNIGHT LAB, <http://knightlab.northwestern.edu/about> (last visited July 10, 2014).

stock market is a proxy for retirement savings; the weather may affect commutes and plans for the day. Those two sections are low-level journalism, easy to investigate and simple to report. But the newspaper is far from the only way to get such information.

Consider the stock ticker and the digital thermometer. Each takes a raw data input and processes it into output that humans can quickly and intuitively understand. News consumers take for granted the accuracy of this output, despite the lack of human moderation in the form of fact checkers or editors standing by to ensure the accuracy of a temperature or stock price. Similar output occur at stoplights, when GPS systems give directions, and even through Artificial Intelligence bots capable of sustaining a facsimile of conversation through a chat program.

Returning to journalism, what happens when computers are enabled to provide information on topics like the weather and financial sector? More interestingly, what happens when computers are asked to tackle topics more complex than a simple report of a temperature or price? Traditional print news reporting is a highly developed field, with many conventions that have been developed to serve readers. But what if computers could be taught to “write” reports on multi-faceted subjects much the same way humans do? This technology is being developed and, in some instances, already being used.

A. Definitions and Terminology

This note will use the term “automated journalism” to refer to the process by which computer algorithms turn data-rich input into prose that reads like a traditional write-up. “Algorithm” is a general term that refers broadly to the category of computer programs that transform input into different sets of output. Input will be at times referred to as “data” or “clean data,” terminology that allows a distinction between the numbers and information that go into a story and the spreadsheet formatting itself. “Output,” “reports,” and “stories” are used interchangeably throughout.

B. Database Journalism

In a way, automated journalism is one of the most logical outgrowths of database journalism, although the two may seem at odds. Data journalism is a subversion of the normal prose structure of news stories, which was catalyzed by the realization that some of what traditional journalists do is better stored as

numbers in a spreadsheet rather than in prose form.⁵ For instance, if a journalist were to track down every phone call made during a certain time period from certain state offices through public records requests, society might benefit more from the production of an organized spreadsheet identifying caller, recipient, time of day, etc. rather than a simple repurposing of that data into one story and one interpretation. Of course, data journalism still leaves room for the reporter to write his story — the point is simply that by making the clean data available for mining by others, perhaps more patterns or narratives will emerge.

Automated journalism operates on a different part of the process, by attempting to use the systems inherent in data journalism to identify the most relevant story. It also thrives on the exploitation of large data-rich caches of information already available in several areas of public interest.⁶ In some cases, the information needs to be organized or cleaned up. In other cases, it is already in usable form. But the upshot is that automated journalism programs are a systematic, rather than human-driven, way of turning data collections into a format most news consumers are comfortable with: a prose translation of the underlying information into words, sentences, paragraphs, and articles.

C. The Business of Automated Journalism

Narrative Science, one example of the practicing leaders in the field, was buoyed when the *New York Times* published an article in September 2011 introducing the technology to a large cross section of the public.⁷ This led to a bout of coverage in mass-media publications with headlines such as “This Article Was Not Written By a Computer,”⁸ “Can the Computers at Narrative Science Replace

⁵ See, e.g., Nate Silver, *What the Fox Knows*, FIVETHIRTYEIGHT (Mar. 17, 2014, 5:38 AM), <http://fivethirtyeight.com/features/what-the-fox-knows/> (“Data journalists, meanwhile, can organize information by running descriptive statistics on it, by placing it into a relational database or by building a data visualization from it.”).

⁶ See, e.g., Michael Sallah, Debbie Cenziper & Steven Rich, *Left With Nothing*, WASHINGTON POST (Sept. 8, 2013), <http://www.washingtonpost.com/sf/investigative/2013/09/08/left-with-nothing/> (investigation into property tax liens placed on homeowners in the District of Columbia).

⁷ Steve Lohr, *In Case You Wondered, a Real Human Wrote This Column*, NEW YORK TIMES (Sept. 10, 2011), http://www.nytimes.com/2011/09/11/business/computer-generated-articles-are-gaining-traction.html?pagewanted=all&_r=0.

⁸ Rachel Arndt, *This Article Was Not Written By a Computer*, FAST COMPANY (Nov. 8, 2011, 5:37 PM), <http://www.fastcoexist.com/1678779/this-article-was-not-written-by-a-computer>.

Paid Writers,”⁹ and, perhaps most optimistically, “Can an Algorithm Write a Better News Story than a Human Reporter?”¹⁰ Reports have indicated that the company raised \$6 million and \$11.5 million in two highly publicized rounds of venture capital funding.¹¹

To understand the value investors see in Narrative Science, and to understand the legal implications, it is instructive to break down in general terms how the program creates a story. For an individual client, such as a media outlet — although, uses for the technology have been imagined in the medical, financial, and tech industries, as well as a host of others — the company tailors an algorithm to its needs, based on the expected output. For instance, the vocabulary utilized in the output story is tweaked.¹² The outcome of a baseball game and a football game are both dependent on the number of points scored by each of two competing teams, but any human reader would blanch at a description of a 21-7 contest that stated that the Packers beat the Vikings by 14 runs. Specializations also exist for various fields; the algorithm that interprets housing starts is different from one that deals with polling numbers. But in every case, the key is data-rich input. Automated journalism software is adept at interpreting a large set of data, be that barometric pressure over time or TV ratings, running that data through an algorithm, and releasing a story about that data using traditional grammar, vocabulary, and syntax.

The business implications of such technology are deep and still developing. Bloomberg, Forbes, and the Big 10 Network all have some form of automated journalism integrated into their regular news output.¹³ In June of 2014, the Associated Press announced that it would soon follow suit.¹⁴ Though it remains to be seen whether the lowered cost and increased speed of reporting such technology provides can compensate for the lack of human voice and interpretation, it is likely

⁹ Joe Fassler, *Can the Computers at Narrative Science Replace Paid Writers?*, THE ATLANTIC (Apr. 12, 2012, 8:03 AM), <http://www.theatlantic.com/entertainment/archive/2012/04/can-the-computers-at-narrative-science-replace-paid-writers/255631>.

¹⁰ Steven Levy, *Can an Algorithm Write a Better News Story Than a Human Reporter?*, WIRED (Apr. 24, 2012, 4:46 PM), http://www.wired.com/2012/04/can-an-algorithm-write-a-better-news-story-than-a-human-reporter/?utm_source=twitter&.

¹¹ John Pletz, *Narrative Science Gets \$11.5 Million to Write Next Chapter*, CRAIN'S CHICAGO BUSINESS (Sept. 10, 2013), <http://www.chicagobusiness.com/article/20130910/BLOGS/11/130919987/narrative-science-gets-11-5-million-to-write-next-chapter>.

¹² Lohr, *supra* note 7 (“The data also determines vocabulary selection. A lopsided score may well be termed a ‘rout’ rather than a ‘win.’”).

¹³ Lohr, *supra* note 7.

¹⁴ Paul Colford, *A Leap Forward in Quarterly Earnings Stories*, ASSOCIATED PRESS (June 30, 2014), <http://blog.ap.org/2014/06/30/a-leap-forward-in-quarterly-earnings-stories>.

— inevitable, in fact — that future uses of this technology will not be confined to journalism, however broadly defined. Other such uses are outside the scope of this note, which is confined to automated journalism “reporting,” while focusing specifically on the mass media implications of such work.

II

AUTOMATED JOURNALISM AND THE FIRST AMENDMENT

A. *A Three-Dimensional Problem:*

Theories of Protection, Manner of Restraint, Type of Output

Does algorithmic output fall within the realm of speech protected by the First Amendment? Courts have only begun to flesh out the answer to this question in the multitude of circumstances in which it might arise. But the normative answer likely depends on one’s preferred theory of First Amendment protection as well as the type of protection being contemplated. Once an abstract framework is in place for these concepts, it becomes easier to make sense of ramifications for the various points along the potential spectrum of machine-generated output.

As a category, algorithmic output may be considered more or less valuable — that is, more or less worth protecting — depending on the lens through which one views the First Amendment. The four traditional justifications for the protection of free speech¹⁵ may lead proponents to different baseline calibrations for evaluation.

Adherents to the theory of the “marketplace of ideas,” following Justice Holmes’ famous articulation that “the best test of truth is the power of the thought to get itself accepted in the competition of the market,”¹⁶ may welcome any new voice, human or otherwise. Likewise, as long as programmers stand behind it, those that see individual self-fulfillment as the main function of speech might appreciate algorithmic output as such.¹⁷ Proponents of the self-governance theory

¹⁵ Namely, “the need to protect the truth-seeking function of the marketplace of ideas; the facilitation of democratic self-actualization; the pragmatic value of providing a social safety valve; and the safeguarding of individual liberty or autonomy.” Steven G. Gey, *The First Amendment and the Dissemination of Socially Worthless Untruths*, 36 FLA. ST. U. L. REV. 1, 6 (2008); see also Thomas I. Emerson, *Toward A General Theory of the First Amendment*, 72 YALE L.J. 877, 902 (1963).

¹⁶ *Abrams v. United States*, 250 U.S. 616, 630 (1919) (Holmes, J., dissenting).

¹⁷ Or maybe not — as Emerson construes the self-fulfillment theory,

Man is distinguished from other animals principally by the qualities of his mind. He has powers to reason and to feel in ways that are unique in degree if not in kind. He has the capacity to think in abstract terms, to use language, to communicate his thoughts and

would probably view machine speech more skeptically, as the role of algorithmic output in achieving Justice Brandeis' construction of the goal of the state ("to make men free to develop their faculties; and that in its government the deliberative forces should prevail over the arbitrary")¹⁸ certainly depends on how you define "deliberative." And there would seem to be almost no room for algorithmic output as speech for those who envision free speech's main social benefit as providing a forum for potentially dangerous actors to let off steam in a manner less harmful than engaging in physical action.¹⁹

The application of media and copyright law to automated journalism raises several important First Amendment questions within this framework. In the media law realm, the Supreme Court's recent decision in *Brown v. Entertainment Merchants Association* sounds in the area of content-based governmental restriction, but some of the most interesting as-yet unanswered questions arise in tort. For instance, when will false information disseminated by an algorithm be considered defamation?

The First Amendment questions surrounding copyright law go the other direction. When does Congress's goal of "promot[ing] the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries"²⁰ supersede the free usage of those creations? The first-principles formulation of copyright law is more well settled than that of free speech — the moral rights doctrine having been basically rejected in favor of a pecuniary rights regime (excepting the limited protection of VARA)²¹ — but open questions still exist as to authorship.

emotions, to build a culture. He has powers of imagination, insight and feeling. It is through development of these powers that man finds his meaning and his place in the world.

Thomas I. Emerson, *Toward A General Theory of the First Amendment*, 72 YALE L.J. 877, 879 (1963).

¹⁸ *Whitney v. California*, 274 U.S. 357, 375 (1927) (Brandeis, J., concurring), *overruled by* *Brandenburg v. Ohio*, 395 U.S. 444 (1969).

¹⁹ This may be a function of journalism, rather than of automated journalism, however. See John J. Watkins Charles, Gertz and the Common Law of Defamation: Of Fault, Nonmedia Defendants, and Conditional Privileges, 15 TEX. TECH L. REV. 823, 850 (1984) ("[A]s Professor Nimmer has noted, the self-fulfillment and safety valve aspects of the first amendment have little relevance to the press.").

²⁰ U.S. CONST. art. I, § 8, cl. 8.

²¹ Visual Artists Rights Act, 17 U.S.C. § 106A, 1990; See *Carter v. Helmsley-Spear, Inc.*, 861 F. Supp. 303, 313 (S.D.N.Y. 1994) ("In passing VARA, Congress for the first time provided for protection of artists' "moral rights" under the Copyright Act.").

The theoretical lines are familiar, but technology has presented a new palette with which to color. Courts and scholars have discussed the spectrum of possible algorithmic output. This discussion warrants a brief overview, as it helps contextualize the specific type of output that is this note's focus. From a positive standpoint, it is clear that in some cases the product of an algorithm or software program triggers the same protections as a piece of political writing produced by traditional means.²² However, it is equally clear that in other cases it does not. For example, no one argues that the "opinion" expressed by an automatic door that opens in response to motion, nor a sign on subway platforms indicating when the next train will arrive should receive special protection of any sort.

Indeed, these examples are definitively outside protection. Much output produced by computer algorithms does not meet the threshold of the "ideas" and "social messages" that the Court considers sufficient for First Amendment purposes. For illustration, in a set of debating articles published in the *University of Pennsylvania Law Review* in 2013, Stuart Minor Benjamin and Tim Wu both concede that there are plenty of permutations of "speech" produced by a machine, computer or algorithm that neither receive nor deserve constitutional treatment, regardless of where the line is drawn for First Amendment protection.²³ In his article, Wu notes the importance of such line drawing stating, "Too little protection would disservice speakers who have evolved beyond the printed pamphlet. Too much protection would threaten to constitutionalize many areas of commerce and private concern without promoting the values of the First Amendment."²⁴

The articles conceive of "machine speech" as a category, on one end of which lie videogames, on the other, automatic doors, car alarms, and the like. Wu, Benjamin and other scholars have dived eagerly into borderline cases. Arguments have been made for and against the speech value of GPS directions, search engine results, and Facebook "likes."²⁵ Necessarily, such discussions revolve around multiple axes. For instance, Wu distinguishes "speech" from "communication," investigates questions of personhood, and further explores the traditional exclusions and inclusions that the Supreme Court has defined for the category of speech — explicit exclusions including incitement, false statements of fact,

²² *Brown v. Entm't Merchants Ass'n*, — U.S. —, 131 S. Ct. 2729, 2733 (2011) (stating that video-games deserve First Amendment protection).

²³ See Tim Wu, *Machine Speech*, 161 U. PA. L. REV. 1495, 1496 (2013); Stuart Minor Benjamin, *Algorithms and Speech*, 161 U. PA. L. REV. 1445, 1450 (2013).

²⁴ Wu, *supra* note 23, at 1498.

²⁵ See *Bland v. Roberts*, 730 F.3d 368, 380 (4th Cir. 2013).

obscenity, and child pornography will receive no First Amendment protection regardless of their vessel.

Wu and Benjamin reach opposing conclusions about where the line should be drawn on First Amendment protection for machine speech. Benjamin would exclude “algorithmic outputs that do not reflect human decision making,”²⁶ whereas Wu advocates for an extended application of the functionality doctrine that he detects in First Amendment jurisprudence.²⁷ Regardless, it is clear that any attempt at line drawing requires an accounting for at least three dimensions of the problem. Not only must the specific character of algorithmic output in question be identified, but, equally important, courts and scholars attempting to fit new technologies into existing First Amendment schema must contemplate the type of restriction at issue and establish first principles as well.

B. Media Law

1. Adjudicated Issues in Media Law: Implications of Brown for Automated Journalism

In *Brown v. Entertainment Merchants Association*, the Supreme Court spoke volumes about the First Amendment value of algorithmic speech by its omission of any acknowledgement of the category.²⁸ In that case, the constitutionality of California Assembly Bill 1179, which prohibited the sale of some violent video games to minors, was in question.²⁹ Specifically, the Bill banned the sale of those games for which a “reasonable person, considering the game as a whole, would find appeals to a deviant and morbid interest of minors.”³⁰

The statute was therefore a content-based restriction; one question addressed by the court was whether the videogame medium was truly speech. The Court’s opinion, written by Justice Scalia, dispenses with this query immediately: “California correctly acknowledges that video games qualify for First Amendment protection.”³¹ The idea that the output might be distinguishable from the code that created the game is not touched upon. Rather, Scalia is blunt in stating the Court’s view that “[l]ike the protected books, plays, and movies that preceded them, video games communicate ideas — and even social messages — through many familiar

²⁶ Benjamin, *supra* note 23, at 1450.

²⁷ Wu, *supra* note 24, at 1479.

²⁸ *Brown*, — U.S. at —, 131 S. Ct. at 2733.

²⁹ *Id.* at 2732.

³⁰ *Id.*

³¹ *Id.* at 2733.

literary devices ... that suffices to confer First Amendment Protection.”³² In fact, Scalia writes, “whatever the challenges of applying the Constitution to ever-advancing technology, ‘the basic principals of freedom of speech and the press, like the First Amendment’s command, do not vary,’ when a new and different medium for communication appears.”³³

Because the Court holds that the conveyance of “ideas” and “social messages” is sufficient for First Amendment protection, it appears that the relatively narrow question of whether stories created by automated journalism programs will be treated as speech can be answered in the affirmative.³⁴ Though Scalia moves on from the speech categorization question without much discussion (as the point was uncontroverted by the parties in dispute), the Court’s justification for the holding is worth parsing further — what happens if an algorithmic output lacks the requisite “ideas” or “social messages?”

Rather than grounding its argument in any of the traditional theories of protection, the Court avoids explicitly endorsing any framework in favor of oblique references (intended or not) to the marketplace of ideas³⁵ and safety valve theories.³⁶ The Court also touches on a justification from history in stating, “[f]or better or worse, our society has long regarded many depictions of killing and maiming as suitable features of popular entertainment, including entertainment that is widely available to minors.”³⁷

The overarching result of *Brown* is clarity as to the speech value of machine output of the highest cognitive level — specifically, output expressing ideas and social messages along the lines of what humans express. This clearly includes automated journalism. Accordingly stories produced using automated journalism technology will trigger strict scrutiny for content-based regulation and intermediate scrutiny for content-neutral regulation.

³² *Id.*

³³ *Id.* (citing *Burstyn, Inc. v. Wilson*, 343 U.S. 495, 503 (1952)).

³⁴ To elaborate, it is hard to imagine a court acknowledging that “ideas” and “social messages” are inherent in videogames yet lacking in a news story.

³⁵ The marketplace of ideas is the theory that “government has no power to decree [esthetic and moral judgments about art and literature], even with the mandate or approval of a majority.” *Brown*, — U.S., at —, 131 S. Ct. at 2733 (quoting *United States v. Playboy Entm’t Group*, 529 U.S. 803, 818 (2000)).

³⁶ The safety valve theory states that “the obscenity exception to the First Amendment does not cover whatever a legislature finds shocking, but only depictions of “sexual conduct.” *Brown*, — U.S., at —, 131 S. Ct. at 2734.

³⁷ *Id.* at 2745.

However, the court's reluctance to explicitly endorse one or several rationales for accepting video games as speech leaves open the issue of the speech value of more borderline cases: Do videogames that do not express an idea warrant categorization as speech? What about an automated journalistic output with incorrect input — seven paragraphs of gibberish about the market's expectations for the National League MVP's third quarter earnings? In declining to address the foundation from which it categorized video games as speech beyond repeated reference to the jurisprudential tradition of recognizing new technologies as they arise, the Court missed a valuable chance to pick the low hanging fruit of the machine speech conversation. As discussed below, issues concerning authorship and personhood are much thornier and harder to reach.

2. *Open Questions in Media Law: Defamation*

Given the several sources of human control inherent in a piece of automated journalism — e.g. the input may be recorded falsely by an overworked newsroom employee, the algorithm may contain flaws that lead to inconsistent output — there is real potential for automated pieces to occasionally contain inaccuracies or falsehoods. This means that the potential exists for disgruntled subjects to commence legal action against outlets that use automated journalism technologies. However, in a defamation suit, the analysis might be different than it would be for a piece authored by a human. For instance, imagine that a Forbes employee accidentally entered data from 2009 into the algorithm that created the story about Aeropostale quoted above. The story is a prediction, and would still register as such, but it would be premised on false information.

A prima facie defamation claim requires that the defendant publish a false, defamatory statement of fact concerning the plaintiff, with some level of fault with respect to the falsity of the statement.³⁸ The implications of some parts of this definition are independent of the machine status of the author. However, some aspects of such a claim, especially the level of fault, raise interesting implications for the best practices of a media organization attempting to regularly publish stories produced by an algorithm.

The truth or falsity of the statement is one aspect of defamation that will not have to be re-examined in light of a machine author. This is because courts have tended to make tests of the meaning of the words in question dependent on the understanding of outside parties, rather than on the intent of the writer. Some courts ask, “How would a reasonable actor interpret the allegedly defamatory

³⁸ RESTATEMENT (SECOND) OF TORTS § 558 (1977).

statement?”³⁹ Others afford the statement a meaning that would be given by a reasonable person of ordinary intelligence.⁴⁰ In either case, the court adopts an external, reader-centric viewpoint; the intent or status of the writer will not be in question at this point.

United States courts generally require that a statement’s topic contain a degree of moral opprobrium in order to be defamatory.⁴¹ This, too, is a requirement that hinges on the perception of readers, not on the intent or actions of the defendant. Still, one can imagine a situation in which the relevant moral opinion of a “substantial and respectable minority,”⁴² of the community is influenced differently by a statement written by a human versus that of an algorithm. But until courts decide on an appropriate application of scienter, it will be difficult to predict the way this standard will swing.

i. How should actual malice or negligence be determined in defamation cases arising from animated journalism articles?

The level of fault required in a successful defamation suit depends on the type of plaintiff claiming to have been defamed. For public figures and public officials, “actual malice” is required for defamation to be found.⁴³ (For non-public figures, mere negligence is the standard; see the discussion below.) Actual malice is defined as false information published “with knowledge that the information was false,” or with “reckless disregard for whether it was false or not.” This is a

³⁹ See, e.g., *Masson v. New Yorker Magazine, Inc.*, 501 U.S. 496, 515 (1991) (“[W]e can think of no method by which courts or juries would draw the line between cleaning up and other changes, except by reference to the meaning a statement conveys to a reasonable reader.”).

⁴⁰ See, e.g., *Romaine v. Kallinger*, 537 A.2d 284, 288 (N.J. 1988) (“In making this determination [on whether the statement at issue is reasonably susceptible of a defamatory meaning], the court must evaluate the language in question according to the fair and natural meaning which will be given it by reasonable persons of ordinary intelligence.” (internal quotation marks omitted)).

⁴¹ For instance, false information about a party’s address would probably not be defamation, but falsely descriptions of a party abusing his child probably would be. See, e.g., *Moss v. Camp Pemigewassett, Inc.*, 312 F.3d 503, 507 (1st Cir. 2002) (defining a defamatory statement as one that “tends to lower the plaintiff in the esteem of any substantial and respectable group of people” and finding that false accusations of “inappropriate contact” with young campers meet the definition) (citation omitted).

⁴² See *Jews For Jesus, Inc. v. Rapp*, 997 So. 2d 1098, 1100 (Fla. 2008) (quoting Restatement (Second) of Torts § 559 (1972)).

⁴³ See, e.g., *New York Times Co. v. Sullivan*, 376 U.S. 254, 279–80 (1964); *Curtis Pub. Co. v. Butts*, 388 U.S. 130, 133–34 (1967).

standard based on the state of mind of the publishing party — unlike the standards for moral opprobrium and falsity, the actual malice standard asks the court to take into consideration the mindset of the speaker and not simply the perception of subjects or recipient parties.

This has not traditionally been a problem, nor would it be for a mass media enterprise in possession of an automated journalism program today. The court would simply be able to impute the level of fault of a publishing entity, through its various copy editors and fact checkers, in much the same way it does in other organizational contexts. But that set up is certainly the easy case in today's media landscape, as the Internet has decreased the cost of publishing to almost nothing. This has enabled individuals or small enterprises without traditional editing systems to reach larger audiences than ever before.

For an illustration of the problem of assigning a mental state for works produced via automated journalism, recall the crossed input example above, with some tweaks to avoid complications posed by group or organizational defamation. Imagine a computer programmer with a passing interest in politics. On his personal computer, he registers a domain name for a website onto which he begins to post blog entries. After a while, his workload picks up, and he begins staying later at the office. So as not to abandon his side project, he licenses or creates an algorithm that combines keywords from certain news stories with results from a reputable public opinion poll to create articles juxtaposing candidates' latest public statement with their polling numbers on that topic. After seeing that it has worked correctly the first few days, he lets this program run without supervision. But soon, due to a programming error that confuses the two inputs in a small number of cases, the headline appears about a popular but beleaguered candidate: "Smith tells voters he's accepted campaign bribes, 75% believe 'I love this country.'" Would that publisher's failure to exercise oversight over the automated statement generated by his algorithm rise to the level of actual malice? What if, instead of a public figure, the defamed subject's name and information were pulled randomly from the blogger's Facebook friends? The private figure analysis is even more fraught with difficulty.

Clearly, the algorithm itself cannot be said to have acted with actual malice or even negligence in any situation. One interesting effect of automated journalism is that it removes any possible culpability from the "writer" of the story and places it squarely upon the publisher, whether a media conglomerate or an individual blogger. Precedent indicates courts' hesitance to assign this level of responsibility to a non-writing party.

Taking public figures first, the mere failure of a fact checker to catch an error does not rise to the level of actual malice.⁴⁴ In fact, courts have held that even a publisher's possession of facts that contradict false information contained in a story does not automatically amount to actual malice, either (though, it would violate the negligence requirement for defamation cases in which the plaintiffs are non-public figures or public officials).⁴⁵

However, there are instances in which editorial oversights are egregious enough to rise to the level of actual malice, particularly in cases where there is evidence of some suspicion that further investigation may be needed to verify information contained in a story.⁴⁶ For instance, the Supreme Court has held that "inherently improbable" information, such that "only a reckless man would put ... in circulation," may lead to a finding of actual malice when a publisher does not follow up with fact checking.⁴⁷ In *Harte-Hanks Communications, Inc. v. Connaughton*, the Court further clarified that "evidence of an intent to avoid the truth" is also sufficient to satisfy the actual malice standard.⁴⁸

For private individuals, however, the only constitutional requirement placed on state defamation statutes is that the plaintiff be required to show negligence.⁴⁹ While some state statutes, such as New York's, heighten the standard by "requir[ing] private figure[s] to show that the media defendant acted in a grossly irresponsible manner regarding its statements about a legitimate public concern," others, like Pennsylvania, require only that a private figure show "mere negligence."⁵⁰ So, while a news outlet's failure to catch a mistaken defamatory statement about a public figure might not lead to a successful defamation claim in

⁴⁴ *Sullivan*, 376 U.S. at 277–78 (finding that "negligence in failing to discover the misstatements ... is constitutionally insufficient to show the recklessness that is required for a finding of actual malice).

⁴⁵ *Great Lakes Capital Partners Ltd. v. Plain Dealer Publ'g Co.*, 8th Dist. Cuyahoga No. CV-599853, 2008-Ohio-6495, ¶ 45.

⁴⁶ 1 RODNEY A. SMOLLA, *LAW OF DEFAMATION* § 3:49 (2d ed.).

⁴⁷ *St. Amant v. Thompson*, 390 U.S. 727, 732 (1968).

⁴⁸ *Harte-Hanks Commc'ns, Inc. v. Connaughton*, 491 U.S. 657, 693 (1989).

⁴⁹ *Gertz v. Robert Welch, Inc.*, 418 U.S. 323, 347 (1974) (holding that "states may define for themselves the appropriate standard of liability" as long as they do not "impose liability without fault" for defamatory injuries to private individuals).

⁵⁰ *Franklin Prescriptions, Inc. v. The New York Times Co.*, 267 F. Supp. 2d 425, 432 (E.D. Pa. 2003).

some jurisdictions,⁵¹ a simple user input error could lead to culpability in others. Though state-by-state analysis is not conducive to generalized discussion, the negligence requirement, at least, seems to translate fairly well from traditional journalism — the requirement is that an editor or fact checker act as a reasonably prudent person would under a corresponding set of circumstances.⁵²

It will also be important for courts to determine whether automated journalism programs act as newsgatherers, or whether their function is more akin to that of a page designer. In other words, is the main function to report a previously unknown story or to take a story that the media entity already owns and simply place it on the page? Criticisms of either view are conceivable — an algorithm, by definition, relies on input that the news outlet must have in its possession. On the other hand, such a program may perform a reporting function in its ability to draw conclusions from a quantity or type of data that would be impractical for human reporters.

Will courts continue to apply the current standards, which afford leeway for poor fact checking by publishers, to defamation cases where the “writer” of an allegedly defamatory story is a computer rather than a human being? If courts are sympathetic to viewing algorithms as newsgatherers, perhaps one way to determine how they will treat such cases is to see how they have dealt with unknown or unreliable writers or sources. In *St. Amant v. Thompson*, the Supreme Court stated, “Professions of good faith will be unlikely to prove persuasive, for example, where a story is fabricated by the defendant, is the product of his imagination, or is based wholly on an unverified anonymous telephone call.”⁵³ Based on this skepticism of “unverified” sources, lower courts have been loath to accept arguments of good faith reliance upon anonymous sources without further verification prior to publication.⁵⁴

⁵¹ See, e.g., *Chapadeau v. Utica Observer-Dispatch, Inc.*, 341 N.E.2d 569, 571 (N.Y. 1975) (stating that failure of newspaper to catch an error does not raise a question as to “grossly irresponsible conduct” so as to preclude summary judgment in its favor).

⁵² See, e.g., *Straw v. Chase Revel, Inc.*, 813 F.2d 356, 359 (11th Cir. 1987) (“The jury was entitled to find that Mr. Smith’s failure to verify the assertions contained in it amounted to a failure to exercise that degree of care exercised under the same or similar circumstances by ordinarily prudent persons...”).

⁵³ *St. Amant*, 390 U.S. 732.

⁵⁴ See, e.g., *Holter v. WLCY T.V., Inc.*, 366 So. 2d 445, 453 (Fla. Dist. Ct. App. 1978) (“[W]here an anonymous tipster conveys the information, one would be hard put not to have serious doubts about the authenticity of the tip.”).

The analogy is imperfect, but one might extrapolate from these examples that courts will ask publishers of automated journalism — which, in a way, involve unnamed sources — to meet a higher level of verification than they would ask for traditional human-written pieces. But if courts view automated journalism programs as simple republishers, the fact checking question would fall to the original data gatherers instead.

ii. How will Section 230 be applied to the new technology?

Section 230 of the Communications Decency Act (“CDA”) limits the liability of online service providers for defamation claims. Congress passed the section as an explicit response by Congress to the ruling in *Stratton Oakmont v. Prodigy*, which held that Prodigy could be liable for statements made on its online bulletin board, even if Prodigy had no knowledge of the information being posted.⁵⁵ In part, the section reads, “No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.”⁵⁶ Interestingly, within § 230, Congress signaled its preference for a marketplace of ideas theory of the First Amendment as well as its distaste for the safety valve theory. As the section states, “It is the policy of the United States ... to preserve the vibrant and competitive free market that presently exists for the Internet and other interactive computer services, unfettered by Federal or State regulation ... [and] to ensure vigorous enforcement of Federal criminal laws to deter and punish trafficking in obscenity, stalking, and harassment by means of computer.”⁵⁷

Section 230 has had several implications for algorithm-generated content since its passage in 1996. For instance, in *Parker v. Google, Inc.*, Google’s status as an interactive computer service immunized it from libel claims stemming from the search engine’s caching and displaying of defamatory content originally created by USENET users.⁵⁸ Courts have also found immunity for website operators whose sites contain defamatory material in posts that they did not author.⁵⁹

⁵⁵ *Stratton Oakmont, Inc. v. Prodigy Servs. Co.*, 1995 WL 323710, at *3 (N.Y. Sup. Ct. May 24, 1995).

⁵⁶ 47 U.S.C.A. § 230.

⁵⁷ *Id.*

⁵⁸ *Parker v. Google, Inc.*, 422 F. Supp. 2d 492, 500–01 (E.D. Pa. 2006) *aff’d*, 242 F. App’x 833 (3d Cir. 2007).

⁵⁹ *Best W. Int’l, Inc. v. Furber*, CV-06-1537-PHX-DGC, 2008 WL 4182827, at *9 (D. Ariz. Sept. 5, 2008).

To envision a more controversial application § 230, imagine an Internet provider who provides users with an algorithm with which they can input their own data. By statutory definition, a provider “means a provider of software ... or enabling tools that ... (A) filter, screen, allow, or disallow content; (B) pick, choose, analyze, or digest content.” The applicability of § 230 to automated journalism will depend on how courts view the technology. If courts are apt to view the algorithm simply as a proxy for a traditional editor, it is unlikely that § 230 will provide shelter from defamation claims. Further, the Ninth Circuit has held that “the CDA does not grant immunity for inducing third parties to express illegal preferences” through the use of form questionnaires.⁶⁰ Only if courts give great deference to Congress’ preference for the marketplace of information rationale as to third-party content hosted on the Internet could such an algorithm be viewed as a simple “tool” used by the “information content providers,” to allow algorithm writers to escape culpability for any potential defamation.

C. Copyright

The oft-cited line from *Feist Publications, Inc. v. Rural Telephone Service Co., Inc.* is the starting point for most discussions of whether certain content is too formulaic or obvious to receive copyright protection: “The *sine qua non* of copyright is originality.”⁶¹ In that case, the Supreme Court found that entries in a phonebook were not protected because the facts therein (phone numbers, names, etc.) were not protected, and the organizational scheme employed (alphabetization by last name) was not original enough to meet the Court’s standard. Justice O’Connor, writing for the Court, uncoupled the concepts of ideas within a work and the work’s expression of those ideas, stating: “A factual compilation is eligible for copyright if it features an original selection or arrangement of facts, but the copyright is limited to the particular selection or arrangement. In no event may copyright extend to the facts themselves.”⁶²

At the highest level of abstraction, automated journalism stories consist of an algorithm, of input (known in the industry as clean data), and of prose output. However, the new technology poses a number of questions related to the organization and usage of clean data input. As for output, major questions on

⁶⁰ Fair Hous. Council of San Fernando Valley v. Roommates.Com, LLC, 521 F.3d 1157, 1165 (9th Cir. 2008) (“Roommate’s own acts – posting the questionnaire and requiring answers to it – are entirely its doing and thus, section 230 of the CDA does not apply to them.”).

⁶¹ *Feist Publ’ns, Inc. v. Rural Tel. Serv. Co., Inc.*, 499 U.S. 340, 345, (1991).

⁶² *Id.* at 350–51.

assignment of authorship loom as the popularity of automated journalism technology grows.

1. Adjudicated Issues in Copyright: Protection for Algorithms

One relatively uncontroversial aspect of the new automated journalism technology is the protection of the algorithm itself. Professor Arthur Miller, in his article on copyright protection for computer programs, explains that just as other forms of expression have been codified into the Copyright Act, computer programs are the most recent candidate for the same reasons that Congress has historically extended intellectual property protection.⁶³ Miller notes that, historically, many new technologies have brought about “fear and concern” that the traditional doctrines and boundaries of protection would not cover them adequately, writing “[t]hese apprehensions were voiced about photography, motion pictures, sound recordings, radio, television, photocopying, and various modes of telecommunication ... As their labors progressed, most members of CONTU⁶⁴ became convinced that computer programs were the latest manifestation of this recurrent phenomenon.”⁶⁵

In support of the 1980 Computer Software Copyright Act, Miller notes, “[c]omputer programs, like other literary works, are expressive. The imagination, originality, and creativity involved in writing a program is comparable to that involved in more time-honored literary works.”⁶⁶ The 1980 Computer Software Copyright Act, now codified as 17 U.S.C. § 117, states in part that the lease, sale, and transfer of rights in a computer program or of its exact copies may be made only with the authorization of the copyright owner.⁶⁷ As for the definition of computer program, § 101 of the Copyright Act reads in relevant part, “A ‘computer program’ is a set of statements or instructions to be used directly or indirectly in a computer in order to bring about a certain result.”⁶⁸ There can be little doubt that the algorithm used to produce an automated journalism story falls under this rubric.

⁶³ Arthur R. Miller, *Copyright Protection for Computer Programs, Databases, and Computer-Generated Works: Is Anything New Since CONTU?*, 106 HARV. L. REV. 977, 982–83 (1993) (articulating three reasons computer programs should be extended the same rights as other forms of expression).

⁶⁴ See, e.g., *id.* at 979 (CONTU is the National Commission on New Technological Uses, established by Congress in 1974).

⁶⁵ *Id.* at 982.

⁶⁶ *Id.* at 983.

⁶⁷ See 17 U.S.C.A. § 117(b).

⁶⁸ 17 U.S.C. § 101.

According to Miller, the fact that automated journalism algorithms are designed to mimic the human process of writing and are thus a form of artificial intelligence, does not change the paradigm.⁶⁹ In fact, the very thing that makes artificial intelligence different from standard computer programs “is more comfortably dealt with under traditional copyright principles than the issues raised by [1993’s] comparatively mundane commercial software.”⁷⁰ As Miller would have it, “these issues were nothing more than the same old wine, and they fit nicely into the old doctrinal bottles.”⁷¹

2. *Open Questions in Copyright: Forms, Fair Use, and Authorship*

i. What protection should be given to the spreadsheet used to create automated journalism stories?

Facts cannot be copyrighted.⁷² However, an interesting question arises when the role of facts for automated journalism is considered in context. In order to create news stories, many automated journalism programs require data sets to be organized in a specific fashion, generally through the use of a spreadsheet program.⁷³ These set-ups, which are used to organize “clean data,” allow systematic sorting and usage of raw input into the eventual prose story. The organization of a typical spreadsheet may be particular to an algorithm, and may in fact be necessary to its function. As such, publishers may wish to protect the organization of their input, the input itself, or both. The law is somewhat in conflict on whether such organizational systems can be copyrighted.

In general, this has meant that creative compilations, using systems of organization less obvious than simple alphabetization, have been copyrightable.⁷⁴ However, some recent decisions in the Courts of Appeals have called into question the precise boundaries for protection on “blank forms.” Since *Lotus Development Corp. v. Paperback Software International*, there has been a tension in the lower

⁶⁹ See Miller, *supra* note 63, at 1036.

⁷⁰ *Id.*

⁷¹ *Id.* at 979.

⁷² *Feist*, 499 U.S. at 350–51.

⁷³ See *Harper & Row Publishers, Inc. v. Nation Enterprises*, 471 U.S. 539, 550 (1985); *accord*, S. Rep. No. 94–473, at 65 (1975) [hereinafter Senate Report].

⁷⁴ See, e.g., Miller, *supra* note 63, at 1039 n.278 (citing instances in which non-alphabetized databases were/may be copyrightable and explaining growing Congressional debate over the issue).

courts between decisions that have granted protection to spreadsheet programs⁷⁵ with those that have denied it to the data-entry inroads used in professions like medicine and dentistry. Decisions denying copyright protection to such systems generally cite to the 1880 Supreme Court case *Baker v. Seldon*,⁷⁶ and to the Copyright Act, which states that no “idea, procedure, process, system, method of operation, concept, principle, or discovery, regardless of the form in which it is described, explained, illustrated or embodied in such work,” may receive protection.⁷⁷

The circuits are split on how to treat so-called “blank forms.” According to the Code of Federal Regulations, “[b]lank forms, such as time cards, graph paper, account books, diaries, bank checks, scorecards, address books, report forms, order forms and the like, which are designed for recording information and do not themselves convey information,” are not subject to copyright.⁷⁸ But, considering the issues of construction inherent in that regulation,⁷⁹ the circuits have split on borderline cases.⁸⁰ The Ninth Circuit, in *Bibbero Sys., Inc. v. Colwell Sys., Inc.* articulated a bright-line rule for blank forms, stating that just because a form contains “possible categories of information” that does not make it any less blank: “[a]ll forms seek only certain information, and, by their selection, convey that the information sought is important. This cannot be what the Copyright Office intended by the statement ‘convey information’ in 37 C.F.R. 202.1(c).”⁸¹ The bright-line rule for the Ninth Circuit is thus defined by what it calls the “text with forms” exception:⁸² Text integrated with blank forms comprises copyrightable

⁷⁵ See *Lotus Dev. Corp. v. Paperback Software Int’l*, 740 F. Supp. 37, 54 (D. Mass. 1990) (holding that some computer programs with literal and nonliteral aspects that are distinguishable from universal aspects of the article may be copyrightable).

⁷⁶ *Baker v. Seldon*, 101 U.S. 99 (1879) (holding that the text of a book describing a special method is copyrightable, but the method itself is not). Cf. *Utopia Provider Sys., Inc. v. Pro-Med Clinical Sys., L.L.C.*, 596 F.3d 1313, 1320 (11th Cir. 2010) (citing *Baker v. Seldon*, 101 U.S. 99 (1879)); 17 U.S.C. § 102).

⁷⁷ 17 U.S.C. § 102(b).

⁷⁸ 37 C.F.R. § 202.1(c).

⁷⁹ Clearly, the list is not meant to be an exclusive rendering of things considered “blank forms.” But, as there is no general definition given, courts have taken it upon themselves to shade in the rest of the picture.

⁸⁰ See *Utopia Provider Sys.*, 596 F.3d at 1320 n.17 (discussing various cases which support the tension in the courts and the split over whether spreadsheets are copyrightable).

⁸¹ *Bibbero Sys., Inc. v. Colwell Sys., Inc.*, 893 F.2d 1104, 1107 (9th Cir. 1990).

⁸² *Id.* at 1106–07 (“Although blank forms are generally not copyrightable, there is a well-established exception where text is integrated with blank forms”).

work; blank forms, even those with unique organizations or carefully chosen categories, do not.

Other circuits, however, have declined to follow *Bibbero* in applying a bright-line rule.⁸³ In *Whelan Assos., Inc. v. Jaslow Dental Laboratory, Inc.*, the Third Circuit noted that like “the majority of courts,” it would find copyrightability for blank forms “if they are sufficiently innovative that their arrangement of information is itself innovative.”⁸⁴ Particularly useful for the purposes of this note is the Second Circuit’s decision on this topic, in which it considered whether a “baseball pitching form,” could be copyrighted.⁸⁵ According to the District Court’s statement of facts, “[t]he form listed various statistics in a tabular format with a legend at the bottom to explain the categories.” The creator of the form, George Kregos, included nine categories, among which were the names of the starting pitchers, the game time, which team was favored to win, as well as each pitcher’s statistics for the current season and his success against the present opponent.⁸⁶

Some newspapers then published Kregos’s form in that precise format — a step less than one might imagine a company like Narrative Science taking.⁸⁷ The Second Circuit held that a decider of fact would not likely find Kregos’ form to lack the creativity *Feist* requires, and that such a conclusion “certainly could not be reached as a matter of law.”⁸⁸ Thus, the nearly identical form that the Associated Press had been circulating could be subject to an infringement claim by Kregos. More generally, the court noted, “all forms need not be denied protection simply

⁸³ See *Utopia Provider Sys.*, 596 F.3d at 1320 n.17 (summarizing many of the decisions weighing in on this issue).

⁸⁴ *Whelan Associates, Inc. v. Jaslow Dental Lab., Inc.*, 797 F.2d 1222, 1243 (3d Cir. 1986).

⁸⁵ *Kregos v. Associated Press*, 937 F.2d 700, 702 (2d Cir. 1991).

⁸⁶ *Kregos v. Associated Press*, 731 F. Supp. 113, 114 (S.D.N.Y. 1990), *aff’d in part, rev’d in part*, 937 F.2d 700 (2d Cir. 1991) (“The first major category lists each pitcher’s statistics for the current season. Under this heading are the sub-categories of wins, losses, and earned run average. The second general category represents the pitcher’s performance during his career against the scheduled opponent. Plaintiff further tailors this category to only include the pitcher’s statistics against this opponent at the particular site where the upcoming game is to be played. This category is then divided into wins, losses, innings pitched, and earned run average. Finally, the last of the main categories lists various statistics for the pitcher’s last three starts. Included within this category are wins, losses, innings pitched, earned run average, and men on base average (‘MBA’)”).

⁸⁷ An automated journalism program could take a similar spreadsheet to Kregos’s form to allow for an algorithm to produce previews upcoming baseball games in much the same fashion, but probably would not publish the underlying spreadsheet.

⁸⁸ *Kregos*, *supra* note 85, at 705.

because many of them fail to display sufficient creativity.”⁸⁹ This is surely welcome news for the media entity that wishes to protect its unique method of data organization for the clean data it feeds into an automated journalism program.

Further good news for such entities, and in further contravention of the Ninth Circuit’s bright-line rule, is the treatment computer spreadsheet programs have received. These programs have generally received copyright protection, resolving some of the tension at the margin of what is and is not considered a “blank form.” The leading cases on this area both involve Lotus Development Corporation, which twice sued to protect the “menu command structure” of its program, Lotus 1-2-3, a main competitor in the market for spreadsheet applications in the late 1980s and early 1990s. In *Lotus Development Corp. v. Paperback Software Intern.*, the Massachusetts District Court found copyright protection for the command elements and menus of Lotus 1-2-3,⁹⁰ which comprise the parts of *any* spreadsheet program that can be considered creative and therefore protectable.

As noted above, it is a matter of some dispute whether the spreadsheet itself should receive copyright protection, but it is relatively clear that, in combination, a spreadsheet with attendant data is indeed copyrightable. One ancillary question to this discussion is whether a second news organization could use this spreadsheet and data combination for its own end by applying the doctrine of fair use. In other words, does the transformation that the clean data and spreadsheet undergoes on its way to becoming an English-language news story rise to the level needed for the fair use limitation to apply to the original author’s exclusive copyright?

The four factors weighed in a decision as to whether a certain use of copyrighted work falls under fair use are: (1) the purpose and character of the use; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (4) the effect of the use upon the potential market for the copyrighted work.⁹¹ Also of note is the list of “purposes” identified in the Copyright Act’s section on fair use, which includes “criticism, comment, news reporting, teaching ... scholarship, or research.”

An analogy can be drawn between the unauthorized use of an unpublished data-spreadsheet set and the facts of the case in *Harper & Row Publishers, Inc. v.*

⁸⁹ Kregos, *supra* note 85, at 709.

⁹⁰ *Lotus Dev. Corp. v. Paperback Software Int’l*, 740 F. Supp. 37, 70 (D. Mass. 1990).

⁹¹ 17 U.S.C.A. § 107.

Nation Enterprises.⁹² That case, in which *The Nation* magazine had scooped *Time* magazine by publishing extensive quotes from an soon-to-be published excerpt of President Gerald Ford's memoirs for which *Time* had paid \$25,000 for the right to publish, was decided for the plaintiffs on the grounds that "The Nation effectively arrogated to itself the right of first publication, an important marketable subsidiary right."⁹³ Fair use doctrine, Justice O'Connor wrote for the court, "has always precluded a use that 'supersedes the use of the original.'" On this view of the role organized data plays in the automated journalism process (that is, as a "marketable subsidiary right"), courts would be unlikely to be sympathetic to a fair use argument.

However, it is possible that courts would see the usage as a technological interchange — the data and its organization could be analogized to a videogame cartridge; the algorithm to a system that can interpret the data therein to form a cognizable image on a television screen. The analogy here would implicate the Ninth Circuit's decision in *Sony Computer Entertainment, Inc. v. Connectix Corp.*⁹⁴ In that case, Connectix corporation had created a "Virtual Game Station," which was intended to perform a very similar function as a Sony PlayStation, but, rather than hooking up to a television, the Virtual Game Station was designed to allow users to play PlayStation games on their computers.⁹⁵ Sony sued for a violation of the copyright it held in the PlayStation's firmware,⁹⁶ which was known as BIOS. The Ninth Circuit found that Connectix's reverse engineering and copying of BIOS for usage in its Virtual Game Station was "modestly" transformative, found fair use, and encouraged Sony to avail itself of the patent system. In sum, if courts view the cumulative input of automated journalism as similar to exclusive, unpublished news material, they will probably be unsympathetic to fair use arguments. However, if the material is seen more as a pathway or reverse engineering of a component piece necessary to the functioning of a system, fair use arguments are more likely to succeed.

ii. Who will courts favor in an authorship dispute?

A final open question in copyright implicated by automated journalism is the treatment of the output of such programs. The major question presented is one of

⁹² *Harper & Row Publishers, Inc.*, 471 U.S. 539.

⁹³ *Id.* at 549.

⁹⁴ *Sony Computer Entm't, Inc. v. Connectix Corp.*, 203 F.3d 596, 598 (9th Cir. 2000).

⁹⁵ *Id.*

⁹⁶ Firmware is software that is embedded in a particular product to allow that product to permanently function as the manufacturer intended.

authorship. Namely, who is the author of the work generated by a computer for the purposes of the initial allocation of copyright?

There are three obvious potential answers to this question, not necessarily exclusive, each of which has been explored to a varying degree in scholarship surrounding this issue. First, the authorship for computer-generated works could be assigned to the programmer or firm who created the algorithm by which the work was generated. Second, it could be assigned to the data entry clerk or data provider, much like the authorship of a photograph attaches to the person standing behind the sight and depressing the shutter.⁹⁷ Or, third, the rights could be assigned to the computer program itself, by finding that the algorithm, in its creative endeavor, has attained the legal personhood necessary to be assigned copyright.

To illustrate how an argument for assigning exclusive rights to a computer program would proceed, Annemarie Bridy's recent article published in the *Stanford Technology Law Review* is instructive.⁹⁸ Essentially, the argument goes that legal personhood is *often* uncoupled from being human — business corporations and government agencies have legal personality in some instances, for example; on the flip side of the coin, slaves “were not legal persons at all under antebellum law.”⁹⁹ Therefore, given the stunning advances made in computer-generated works, the law should be prepared to recognize the true talent (balancing the “creativity of the coder with the creativity of the code”) by awarding some form of copyright to the algorithm itself.¹⁰⁰ An argument for such a drastic change in the copyright system reflects an opinion that “few on either side of the ‘copyfights’ would argue that the system is not broken, and many believe it is irretrievably so.”¹⁰¹

So far, US courts have not agreed. One day automated journalism programs that “write” news stories may be accompanied by automated data collectors, automated newsroom meetings that decide which stories to pursue, automated data input systems, automated editors, and automated publishing suites. But until such a system is in place, the human input necessary for automated journalism to be produced will probably control the copyright.

⁹⁷ C.f. *Burrow-Giles Lithographic Co. v. Sarony*, 111 U.S. 53 (1884).

⁹⁸ Annemarie Bridy, *Coding Creativity: Copyright and the Artificially Intelligent Author*, 2012 STAN. TECH. L. REV. 5 (2012).

⁹⁹ *Id.* at 69, n.157.

¹⁰⁰ Bridy would like to do so by recognizing AI-authored works as works for hire (from computers).

¹⁰¹ Compare Annemarie Bridy, *supra* note 96, at 4 n.189, with Miller, *supra* note 63.

This being the case, the assignation of authorship between the former two categories articulated above — the programmer or the data entrant — will become a very important copyright question as automated journalism gains popularity and wider usage. In the famous case involving a photograph of Oscar Wilde, the Supreme Court found that the photograph in question was copyrightable and that such copyright vested initially in the person taking the picture (as opposed to the manufacturer of the camera or the subject of the photograph).¹⁰² But the Court's opinion anticipated arguments that there was no creativity inherent in new technologies like photography, which held the composer at more of a remove from the process than previous methods like painting.

In predicting how courts will resolve the divide between programmers and those that input data, it is useful to return to the theoretical underpinnings of copyright law. Traditionally, American copyright law and jurisprudence “seeks to vindicate the economic, rather than the personal, rights of authors.”¹⁰³ Where a strict moral conception of copyright might assert that all Polaroid photographs owe dependency to Edwin H. Land,¹⁰⁴ or that the creator of an algorithm has a claim in everything that algorithm output, our current pecuniary conception values the promulgation of technologies into more and newer works. If faced with the decision, therefore, courts will probably prefer the rights of parties who enter data over the claims of algorithm writers in deference to copyright law's abstract framework.

CONCLUSION

In both media law and copyright, the advent of new technology such as automated journalism raises important questions about attribution, mental state, fair use, and more. Some of the questions are easily answered, while others are unlikely to be definitively addressed by the courts for years. However, in identifying these new questions, while courts have been clear in defining a set of first principles and embracing a consistent theoretical structure for copyright law, they have been much less so in the media law realm.

This observation does not necessarily lead to a simple conclusion — one's penchant for judicial principles spanning time, taste, and technology probably

¹⁰² *Burrow-Giles Lithographic Co.*, 111 U.S. 53.

¹⁰³ *Gilliam v. Am. Broad. Companies, Inc.*, 538 F.2d 14, 24 (2d Cir. 1976).

¹⁰⁴ Eric Pace, *Edwin H. Land Is Dead at 81; Inventor of Polaroid Camera*, NEW YORK TIMES (Mar. 2, 1991), <http://www.nytimes.com/1991/03/02/obituaries/edwin-h-land-is-dead-at-81-inventor-of-polaroid-camera.html>.

inform the perceived wisdom of articulating such principles. But as computer technologies rapidly proliferate and concepts like automated journalism arise, courts large and small will have to choose between ruling on correspondent questions case-by-case or picking a conceptual structure to follow.