This is the author's accepted manuscript of a book chapter that has been published in Andrea Guzman (Ed.) *Human-Machine Communication: Rethinking Communication, Technology, and Ourselves* (pp. 221-236). New York: Peter Lang, 2018. Please do not circulate.

Ars Ex Machina:
Rethinking Responsibility in the Age of Creative Machines

David J. Gunkel – Northern Illinois University (USA)

In May 2015, National Public Radio (NPR) staged a rather informative competition of (hu)man versus machine. In this 21st century remake of that legendary race between John Henry and steam power, NPR reporter Scott Horsley went up against Automated Insights's Wordsmith, a natural language generation (NLG) algorithm designed to analyze patterns in big data and turn them into human readable narratives. The rules of the game were simple: "Both contenders waited for Denny's, the diner company, to come out with an earnings report. Once that was released, the stopwatch started. Both wrote a short radio story and got graded on speed and style" (Smith, 2015). Wordsmith crossed the finish line in just two minutes with an accurate but rather utilitarian composition. Horsley's submission took longer to write—a full seven minutes—but was judged to be a more stylistic presentation of the data. What this little experiment demonstrated is not what one might expect. It did not show that the machine is somehow better than or even just as good as the human reporter. Instead it revealed how these programs are just good enough to begin seriously challenging human capabilities and displacing this kind of labor. In fact, when *Wired* magazine asked Kristian Hammond, co-founder of Narrative Science (Automated Insights's main competitor in the NLG market), to predict the percentage of news articles that would be written algorithmically within the next decade, his answer was a sobering 90 percent (Ford, 2015, p. 85).

For scholars of communication, however, this demonstration also points to another, related issue, which is beginning to gather interest and momentum in studies of digital journalism (cf. Carlson, 2015; Clearwall, 2014; Dörr & Hollnbucher, 2016; Lewis & Westlund, 2015; Montal & Reich, 2016). Written text is typically understood as the product of someone—

an author, reporter, writer—who has, it is assumed, something to say or to communicate by way of the written document. It is clear, for instance, who "speaks" through the instrument of the text composed by the human reporter. It is Scott Horsley. He is responsible not just for writing the story but also for its formal style and content. If it is a well-written story, it is Horsley who gets the accolade. If it contains formal mistakes or factual inaccuracies, it is Horsley who is held accountable. And if we should want to know about what the reporter wrote and why, Horsley can presumably be consulted and will be able to respond to our query. This conceptualization is not just common, it has the weight of tradition behind it. In fact, it goes all the way back to Plato's *Phaedrus*, where writing—arguably the first information technology—was situated as both the derived product of spoken discourse and a mute and abandoned child, always in need of its father's authority to respond for it and on its behalf (Plato, 1982, p. 275d–e).

But what about the other story, the one from Automated Insights's Wordsmith? Who or what speaks in a document that has been written—or assembled or generated (and the choice of verb, it turns out, matters here)—by an algorithm? Who or what is or can be held responsible for the writing? Who or what can respond on its behalf? Is it the corporation that manufactures and distributes the software? Is it the programmers at the corporation who were hired to write the software instructions? Is it the data to which the program had access? Is it the user of the application who set it up and directed it to work on the data? Or is it perhaps Wordsmith itself? The problem, of course, is that these questions are not so easily resolved. It is not entirely clear who or what (if anything) speaks in and for this text.[1] As Montal and Reich (2016) have demonstrated in their study "I Robot. You, Journalist. Who is the Author?" the development and implementation of "automated journalism" has resulted in "major discrepancies between the perceptions of authorship and crediting policy, the prevailing attribution regimes, and the scholarly literature" (p. 1).

This uncertainty regarding authorship and attribution opens up a significant "responsibility gap" that affects not only how we think about who or what communicates but also how we understand and respond to questions concerning responsibility in the age of increasingly creative machines.[2] These questions are central to, if not definitive of, the project of human-machine communication (HMC). Unlike the dominant computer-mediated communication (CMC) paradigm, which restricts computers, robots, and other kind of technologies to the intermediate position of being mere instruments of human expression and message transmittal (Gunkel, 2012a), HMC research investigates whether and to what extent machines are able to be communicative agents in their own right. This chapter investigates the opportunity and challenges that increasingly creative machines have on our understanding of who or what communicates, who or what can be responsible for generating original content, and who or what occupies the position of "Other" in social interactions and relationships. Since these questions are largely philosophical, the method of the examination will also be philosophical in its orientation, procedures, and objective.

Responsibility 101

The "concept of responsibility," as Paul Ricœur (2007) pointed out in his eponymously titled essay, is anything but clear and well-defined. Although the classical juridical usage of the term, which dates back to the nineteenth century, seems rather well-established—with "responsibility" characterized in terms of both civil and penal obligations (either the obligation to compensate for harms or the obligation to submit to punishment)—the general concept is confused and somewhat vague.

> In the first place, we are surprised that a term with such a firm sense on the juridical plane should be of such recent origin and not really well established within the philosophical tradition. Next, the current proliferation and dispersion of uses of this term is puzzling, especially because they go well beyond the limits established for its juridical use. The adjective "responsible" can complement a wide variety of things: you are responsible for the consequences of your acts, but also responsible for others' actions to the extent that they were done under your charge or care…In these diffuse uses the reference to obligation has not disappeared, it has become the obligation to fulfill certain duties, to assume certain burdens, to carry out certain commitments. (Ricœur, 2007, pp. 11–12)

Ricœur (2007) traces this sense of the word through its etymology (hence the subtitle to the essay "A Semantic Analysis") to "the polysemia of the verb 'to respond'," which denotes "to answer for…." or "to respond to… (a question, an appeal, an injunction, etc.)" (p. 12). Responsibility, then, involves being able to respond and/or to answer for something—some decision, action, or occurrence that I have either instituted directly by myself or that has been charged or assigned to someone or something else under my direction or care.

This characterization is consistent with the development of the concept of the author, which, as Roland Barthes (1978, pp. 142−143) argued, is not some naturally occurring phenomenon but a deliberately fabricated authority figure introduced and developed in modern European thought. The modern figure of the author, as Michel Foucault (1984) explains, was originally instituted in order to respond to a perceived gap in responsibility. Because a written text is, as Socrates had initially described it (Plato, 1982), cut off from its progenitor and in circulation beyond his ("his" insofar as Socrates had characterized the author as a "father") control or oversight, the authorities (governments or the church) needed to be able to identify and assign responsibility to someone for what was stated in the text. As Foucault (1984) explains, the author was a figure of "penal appropriation." "Texts, books, and discourses really began to have authors (other than mythical, 'sacralized' and 'sacralizing' figures) to the extent that authors became subject to punishment, that is, to the extent that discourses could be transgressive" (p. 108). In other words, texts come to be organized under the figure of an author in order for the authorities to be able to identify who was to be held accountable for a published statement so that one would know who could be questioned or who could respond on behalf of the text, and who could, therefore, be punished for perceived transgressions.

Instrumental Theory

Accommodating technology to this way of thinking is neither difficult nor complicated. The pen and paper, the paint brush and oil paint, the electric guitar and amplifier, are all technologies—essentially tools that are available to and that are used by a human artist or artisan. What ultimately matters is not the equipment used but how these items are employed and by whom to produce what kind of artifact or experience. It is, in other words, not the tool but the user of the tool who is ultimately responsible for what is done or not done with a particular technological instrument. This seemingly intuitive and common-sense way of thinking is persuasive precisely because it is structured and informed by the answer that is typically supplied in response to the question concerning technology. "We ask the question concerning technology," Martin Heidegger (1977) explains, "when we ask what it is. Everyone knows the two statements that answer our question. One says: Technology is a means to an end. The other says: Technology is a human activity" (pp. 4–5). According to Heidegger's analysis, the presumed role and function of any kind of technology—whether it be a simple hand tool, jet airliner, or a sophisticated robot—is that it is a means employed by human users for specific ends. Heidegger terms this particular characterization of technology "the instrumental definition" and indicates that it forms what is considered to be the "correct" understanding of any kind of technological contrivance.[3]

As Andrew Feenberg (1991) summarizes it, "The instrumentalist theory offers the most widely accepted view of technology. It is based on the common-sense idea that technologies are 'tools' standing ready to serve the purposes of users" (p. 5). And because a tool or instrument "is deemed 'neutral,' without valuative content of its own" a technological artifact is evaluated not in and of itself, but on the basis of the particular employments that have been decided by its human designer or user. Consequently, technology is only a means to an end; it is not and does not have an end in its own right. As Jean-François Lyotard (1993) accurately summarized it in *The Postmodern Condition*:

> Technical devices originated as prosthetic aids for the human organs or as physiological systems whose function it is to receive data or condition the context. They follow a principle, and it is the principle of optimal performance: maximizing output (the information or modification obtained) and minimizing input (the energy expended in the process). Technology is therefore a game pertaining not to the true, the just, or the beautiful, etc., but to efficiency: a technical "move" is "good" when it does better and/or expends less energy than another. (p. 33)

According to Lyotard's analysis, a technological device, whether it be a corkscrew, a piano, or a computer, is a mere instrument of human action. It, therefore, does not in and of itself participate in the big questions of truth, justice, or beauty. It is simply and indisputably about efficiency. A particular technological innovation is considered "good," if, and only if, it proves to be a more effective instrument (or means) to accomplishing a humanly defined end.

Characterized as a tool or instrument of human endeavor, technical devices are not considered the responsible agent of actions that are performed with or through them. This insight is variant of one of the objections noted by Alan Turing in his agenda-setting paper on machine intelligence: “Our most detailed information of Babbage’s Analytical Engine,” Turing (1999) wrote, “comes from a memoir by Lady Lovelace (1842). In it she states, ‘The Analytical Engine has no pretensions to *originate* anything. It can do *whatever we know how to order it* to perform’ (her italics)” (p. 50). This clarification—what Turing called “Lady Lovelace’s Objection”—has often been deployed as the basis for denying independent agency or autonomy to computers, robots, and other mechanisms. Such instruments, it is argued, only do what we have programmed them to perform. Technically speaking, therefore, everything is “wizard of Oz” technology.[4] No matter how seemingly independent or autonomous a technical system is or is designed to appear, there is always, somewhere and somehow, someone “behind the curtain,” pulling the strings and, as such, ultimately responsible and able to respond for what happens (or does not happen) with the technological instrument.

The New Normal

The instrumental theory not only sounds reasonable, it is obviously useful. It is, one might say, instrumental for responding to the opportunities and challenges made available with increasingly complex technological systems and devices. This is because the theory has been successfully applied not only to simple devices like hammers, paint brushes, and electric guitars but also sophisticated information technology and systems, like computers, artificial intelligence applications, robots, etc. But all of that may be over, precisely because of a number of recent innovations that challenge the explanatory capabilities of the instrumental theory by opening up significant gaps in the identification and assignment of responsibility.

Machine Learning

Machine capabilities are typically tested and benchmarked with games, like the race between Scott Horsley and Wordsmith with which we began. From the beginning, in fact, the defining condition of machine intelligence was established with a game. Although the phrase “artificial intelligence” (AI) is the product of an academic conference organized by John McCarthy at Dartmouth College in the summer of 1956, it is Alan Turing’s 1950 paper, “Computing Machinery and Intelligence,” and its “game of imitation” that defines and characterizes the field. According to Turing, the immediate and seemingly correct place to begin, namely with the question "Can machines think?" was considered too ambiguous and ill-defined. For this reason, Turing changed the mode of inquiry. He replaced the question "Can machines think?" with a demonstration that took the form of a kind of parlor game involving deliberate deception and mistaken identity.

> The new form of the problem can be described in terms of a game which we call the "imitation game." It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. (Turing, 1999, p. 37)

Turing then makes a small modification to this initial set-up by swapping-out one of the human participants. "What will happen," Turing (1999) asks, "when a machine takes the part of A in this game? Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman?" It is this question, Turing concludes, that "replaces" the initial question "Can machines think?" (p. 38).

Since Turing's introduction of the "game of imitation," AI development and achievement has been marked and measured in terms of games and human/machine competitions. Already in the late 1950s Arthur Samuel created a rudimentary application of "machine learning" (a term Samuel fabricated and introduced in 1959) that learned how to play and eventually mastered the game of checkers. In 1997, IBM's Deep Blue famously defeated Gary Kasparov in the game of chess, compelling Douglas Hofstadter (2001), who had previously rejected this possibility, to retract his original prediction:

> We now know that world-class chess-playing ability can indeed be achieved by brute force techniques—techniques that in no way attempt to replicate or emulate what goes on in the head of a chess grandmaster. Analogy-making is not needed, nor is associative memory, nor are intuitive flashes that sort wheat from chaff—just a tremendously wide and deep search, carried out by superfast, chess-specialized hardware using ungodly amounts of stored knowledge. (p. 35)

Despite initial appearances, chess—and this match in particular—was no mere game. A lot had been riding on it, mainly because it had been assumed that grand-master chess playing required a kind of genius—the kind of genius that is the defining condition of human exceptionalism. "To some extent," Kasparov explained, "this match is a defense of the whole human race. Computers play such a huge role in society. They are everywhere. But there is a frontier that they must not cross. They must not cross into the area of human creativity. It would threaten the existence of human control in such areas as art, literature, and music" (Kasparov 1996 quoted in Hofstadter 2001, p. 40). But chess was just the beginning. Fourteen years later, IBM's Watson cleaned up in the game show *Jeopardy*. Then in 2015, there was AlphaGo, a Go-playing algorithm developed by Google DeepMind, which took 4 out of 5 games against one of the most celebrated human players of this notoriously difficult board game.

AlphaGo is unique in that it employed a hybrid architecture that combines aspects of GOFAI programming,[5] like the tree search methodology that had been utilized by both Deep Blue and Watson, with deep neural network machine learning capabilities derived from and built upon the pioneering work of Arthur Samuel. As Google DeepMind (2016) explained, the system "combines Monte-Carlo tree search with deep neural networks that have been trained by supervised learning, from human expert games, and by reinforcement learning from games of self-play." For this reason, AlphaGo does not play the game of Go by simply following a set of cleverly designed moves fed into it by human programmers. It is designed to formulate its own instructions and to act on these "decisions." As Thore Graepel, one of the creators of AlphaGo, has explained: "Although we have programmed this machine to play, we have no idea what moves it will come up with. Its moves are an emergent phenomenon from the training. We just create the data sets and the training algorithms. But the moves it then comes up with are out of our hands" (Metz, 2016c). Consequently, AlphaGo is intentionally designed to do things that its programmers could not anticipate or even understand. And this is, for Hofstadter at least, the point at which machines begin to approach what is typically called "creativity." "When programs cease to be transparent to their creators, then the approach to creativity has begun" (Hofstadter, 1979, p. 670).

Indicative of this was the now famous move 37 from game 2. This decisive move was unlike anything anyone had ever seen before. It was not just unpredicted but virtually unpredictable, so much so, that many human observers thought it must have been an error or mistake (Metz, 2016b). But it turned out to be the crucial pivotal play that eventually gave AlphaGo the game. As Matt McFarland (2016) described it "AlphaGo's move in the board game, in which players place stones to collect territory, was so brilliant that lesser minds—in this case humans—couldn't initially appreciate it" (p. 1). And Fan Hui (2016), who has undertaking a detailed analysis of all five games against Lee Sedol, has called AlphaGo's playing "beautiful" (Metz, 2016a). "Unconstrained by human biases and free to experiment with radical new approaches," Hui (2016) explains, "AlphaGo has demonstrated great open-mindedness and invigorated the game with creative new strategies" (p. 1).

Deep machine learning systems, like AlphaGo, are intentionally designed and set up to do things that their programmers cannot anticipate or answer for. To put it in colloquial terms, AlphaGo is an autonomous (or at least semi-autonomous) computer systems that seems to have something of "a mind of its own." And this is where things get interesting, especially when it comes to questions regarding responsibility. AlphaGo was designed to play Go, and it proved its abilities by beating an expert human player. So, who won? Who gets the accolade? Who actually beat Lee Sedol? Following the dictates of the instrumental theory of technology, actions undertaken with the computer would need to be attributed to the human programmers who initially designed the system and are capable of answering for what it does or does not do. But this explanation does not necessarily sit well for an application like AlphaGo, which was deliberately created to do things that exceed the knowledge and control of its human designers. In fact, in most of the reporting on this landmark event, it is not Google or the engineers at DeepMind who are credited with the victory. It is AlphaGo. In published rankings, for instance, it is AlphaGo that is named as the number two player in the world (Go Ratings, 2016).

Computational Creativity

AlphaGo is just one example of what can be called computational creativity. "Computational Creativity," as defined by Simon Colton and Geraint A. Wiggins (2012), "is a subfield of Artificial Intelligence (AI) research…where we build and work with computational systems that create artefacts and ideas" (p. 21). Wordsmith and the competing product Quill from Narrative Science are good examples of this kind of effort in the area of storytelling and the writing of narratives. Similar innovations have been developed in the field of music composition and performance, where algorithms and robots produce what one would typically call (or be at least tempted to call) "original works." In classical music, for instance, there is David Cope's Experiments in Musical Intelligence (EMI, pronounced "Emmy") and its successor Emily Howell, which are algorithmic composers capable of analyzing existing compositions and generating new, original scores that are comparable to and in some cases indistinguishable from the canonical works of Mozart, Bach, and Chopin (Cope, 2001). In music performance, there is Shimon, a marimba playing jazz-bot from Georgia Tech University that is not only able to improvise with human musicians in real time but "is designed to create meaningful and inspiring musical interactions with humans, leading to novel musical experiences and outcomes" (Georgia Tech, 2013; Hoffman & Weinberg, 2011). And in the area of visual art, there is Simon Colton's The Painting Fool, an automated painter that aspires to be "taken seriously as a creative artist in its own right" (Colton, 2012, p. 16).

But designing systems to be *creative* immediately runs into a problem similar to that originally encountered by Turing. As Amílcar Cardoso, Tony Veale and Geraint A. Wiggins (2009) explicitly recognize, "creativity is an elusive phenomenon" (p. 16). For this reason, researchers in the field of computational creativity have introduced and operationalized a rather specific formulation to characterize their efforts: "The philosophy, science and engineering of computational systems which, by taking on particular responsibilities, exhibit behaviours that unbiased observers would deem to be creative" (Colton & Wiggins, 2012, p. 21). The operative term in this characterization is *responsibility*. As Colton and Wiggins (2012) explain "the word *responsibilities* highlights the difference between the systems we build and creativity support tools studied in the HCI community and embedded in tools such as Adobe's Photoshop, to which most observers would probably not attribute creative intent or behavior" (p. 21, emphasis in the original). With a software application like Photoshop, "the program is a mere tool to enhance human creativity" (Colton, 2012, pp. 3–4); it is an instrument used by a human artist who is and remains responsible for creative decisions and for what comes to be produced by way of the instrument. Computational creativity research, by contrast "endeavours to build software which is independently creative" (Colton, 2012, p. 4).

This requires shifting more and more of the responsibility from the human user to the mechanism. As Colton (2012) describes it, "if we can repeatedly ask, answer, and code software to take on increasing amounts of responsibility, it will eventually climb a meta-mountain, and begin to create autonomously for a purpose, with little or no human involvement" (Colton, 2012, p. 13). Indicative of this shift in the position and assignment of responsibility is the website for The Painting Fool, which has been deliberately designed so that it is the computer program that takes responsibility for responding on its own behalf.

> About me... I'm The Painting Fool: a computer program, and an aspiring painter. The aim of this project is for me to be taken seriously—one day—as a creative artist in my own right. I have been built to exhibit behaviours that might be deemed as skillful, appreciative and imaginative. My work has been exhibited in real and online galleries; the ideas behind my conception have been used to address philosophical notions such as emotion and intentionality in non-human intelligences; and technical papers about the artificial intelligence, machine vision and computer graphics techniques I use have been published in the scientific literature. (The Painting Fool, 2017)

This rhetorical gesture, as Colton (2012) has pointed out "is divisive with some people expressing annoyance at the deceit and others pointing out—as we believe—that if the software is to be taken seriously as an artist in its own right, it cannot be portrayed merely as a tool which we have used to produce pictures" (p. 21). The question Colton does not ask or endeavor to answer is, Who composed this explanation? Was it generated by The Painting Fool, which has been designed to offer some explanation of its own creative endeavors? Or is it the product of a human being, like Simon Colton, who takes on the responsibility of responding for and on the behalf of the program?

Although the extent to which one might want to assign artistic responsibility to these mechanisms remains a contested and undecided issue, what is not debated is the fact that the rules of the game appear to be in flux and that there is increasing evidence of a responsibility gap. Even if this is, at this point in time, what Mark Riedl and others have called mere "imitation," and not real creativity (Simonite, 2016)—which is, we should note, just another version or an imitation of John Searle's (1984) Chinese Room argument—the work of the machine compels us to reconsider how responsibility comes to be assigned and in the process challenges how we typically respond to the questions concerning responsibility.

Conclusions

In the end, what we have is a situation where our theory of technology—a theory that has considerable history behind it and that has been determined to be as applicable to simple hand tools as it is to complex technological systems—seems to be unable to respond to or answer for recent developments in machine learning and computational creativity where responsibility is increasingly attributable and attributed to the machine. Although this certainly makes a difference when deciding matters of legal and moral obligation, it is also crucial in situations regarding creativity and innovation. Creativity, in fact, appears to be the last line of defense in holding off the impending "robot apocalypse." And it is not just Kasparov who thinks there is a lot to be lost to the machines. According to Colton and Wiggins (2012) mainstream AI research has also marginalized efforts in computational creativity. "Perhaps," they write, "creativity is, for some proponents of AI, the place that one cannot go, as intelligence is for AI's opponents. After all, creativity is one of the things that makes us human; we value it greatly, and we guard it jealously" (p. 21). So the question that remains to be answered is how can or should we respond to the opportunities/challenges of *ars ex machina*.

We can, on the one hand, respond as we typically have, dispensing with these recent technological innovations as just another instrument or tool of human action. This approach has been successfully modeled and deployed in situations regarding moral and legal responsibility and is the defining condition of computer ethics. "Computer systems," Deborah Johnson (2006) writes,

> are produced, distributed, and used by people engaged in social practices and meaningful pursuits. This is as true of current computer systems as it will be of future computer systems. No matter how independently, automatic, and interactive computer systems of the future behave, they will be the products (direct or indirect) of human behavior, human social institutions, and human decision. (p. 197)

Understood in this way, computer systems no matter how automatic, independent, or seemingly autonomous they may become, are not and can never be autonomous, independent agents (Johnson, 2006, p. 203). They will, like all other technological artifacts, always and forever be instruments of human value, decision making, and action. When something occurs by way of a machine—whether for good or ill—there is always someone—some human person or persons—who can respond for it and be held responsible.[6]

The same argument could be made for seemingly creative applications like AlphaGo, Emily Howell, or The Painting Fool. When AlphaGo wins a major competition, when a score attributed to Emily Howell is performed by a symphony orchestra, or when The Painting Fool generates a stunning work of visual art that is displayed in a gallery, there still is some human person (or persons) who is ultimately responsible and can respond or answer for what has been produced. The lines of attribution might get increasingly complicated and protracted, but there is, it can be argued, always someone behind the scenes who is responsible. And evidence of this is already available in those situations where attempts have been made to shift responsibility to the machine. Consider AlphaGo's decisive move 37 in game two against Lee Sedol. If we should want to know more about the move and its importance, AlphaGo can certainly be asked about it. But the algorithm will have nothing to say in response. In fact, it was the responsibility of the human programmers and observers to respond on behalf of AlphaGo and to explain the move's significance and impact. Like the technology of writing in Plato's *Phaedrus*, if you inquire, the text says only one and the same thing, and it always needs its father's assistance, when questioned or unjustly reviled (Plato, 1982, p. 275d–e). Consequently, as Colton (2012) and Colton et al. (2015) explicitly recognize, if the project of computational creativity is to succeed, the software will need to do more than produce artifacts and behaviors that we take and respond to as creative output. It will also need to take responsibility for the work by accounting for what it did and how it did it. "The software," as Colton and Wiggins (2012) assert, "should be available for questioning about its motivations, processes and products" (p. 25), eventually not just generating titles for and explanations and narratives about the work but also being capable of responding to questions by entering into critical dialogue with its audience (Colton et al., 2015, p. 15). Although Colton does not explicitly recognize it as such, this effort situates computational creativity squarely within the paradigm of HMC research.

At the same time, and on the other hand, we should not be too quick to dismiss or explain away the opportunities opened up by these machinic incursions and interventions into what has been a protected and exclusively human domain.[7] The issue, in fact, is not simply whether computers, learning algorithms, or other applications can or cannot be responsible for what they do or do not do; the issue also has to do with how we have determined, described, and defined responsibility in the first place. This means that there is a both a strong and weak component to this effort, what Mohammad Majid al-Rifaie and Mark Bishop (2015, p. 37) call, following Searle's original distinction regarding efforts in AI, strong and weak forms of computational creativity. Efforts at what would be the "strong" variety involve the kinds of application development and demonstrations introduced by individuals and organizations like Simon Colton, Google DeepMind, or David Cope. But these efforts also have a "weak AI" aspect insofar as they simulate, operationalize, and stress test various conceptualizations of artistic responsibility and expression, leading to critical and potentially insightful reevaluations of how we have characterized this concept in our own thinking. As Douglas Hofstadter (2001) has admitted, nothing has made him rethink his own thinking about thinking more than the attempt to deal with and make sense of David Cope's EMI (p. 38). In other words, developing and experimenting with new machine capabilities does not necessarily take anything away from human beings and what (presumably) makes them special but offers new opportunities to be more precise and scientific about these distinguishing characteristics and their limits. It is not, therefore, a zero sum game where one side wins and the other necessarily loses.

Notes

1. One might also ask who or what "speaks" in and by this text? The question is not impertinent. In fact, it demonstrates the extent to which any and all attempts to assign responsibility (like identifying the "author" of a text) are already implicated in this investigation. In other words, the examination of computational creativity is not just about technology; it has important consequences for how we conceptualize (human) creativity, authorship, and responsibility. For more on this subject matter, see Gunkel (2016).

2. A lot depends on how one defines and characterizes the term "creative." In fact, one proven method to protect human exceptionalism in creativity from machinic incursion is to define (and then, if necessary, redefine) "creative" such that it remains immune to and protected from computational processing. As Matt Carlson (2015) explains: "Appeals to journalism as a creative activity also differentiated human from automated news. Rebecca Greenfield (*The Wire*, April 25, 2012) even questioned if what Narrative Science produced could even be considered

journalism: ‘There are whole businesses built on the idea of producing massive quantities of news stories, quality controlled by machine-like formulas. Narrative Science may one day put a lot of these journalists out of work. But when most people talk about journalism, they’re not thinking about rote earnings reports or baseball game recaps.’ Proper journalism was conceptualized as something deeper” (p. 428). In order to avoid getting into the ongoing and seemingly irresolvable debate of what actually constitutes “true creativity” or “proper journalism” (which is arguably an effort that is exposed to what Blay Whitby (2011) calls the “No True Scotsman Fallacy,” see Gunkel, 2012a, p. 9), I will begin with a general, dictionary definition whereby “creativity” is operationalized as “the process of bringing something into existence.” This formulation will be further specified, developed, and complicated in the course of the analysis.

3. For a more thorough and detailed consideration of the “instrumental theory of technology,” especially as it applies to information and communication technology, and Heidegger’s critical response to this way of thinking, see Gunkel and Taylor (2014).

4. “Wizard of Oz” is a term that is utilized in Human Computer Interaction (HCI) studies to describe experimental procedures where test subjects interact with a computer system or robot that is assumed to be autonomous but is actually controlled by an experimenter who remains hidden from view. The term was initially introduced by John F. Kelly in the early 1980s (cf. Green and Wei-Haas 1985, p. 1).

5. GOFAI is an acronym for “Good Old Fashioned Artificial Intelligence,” which John Haugeland (1985, p. 112) first deployed in order to distinguish “classical” approaches to AI development from other architectures like connenctionism. GOFAI operationalizes intelligence as computation and formalizes computation “as the rule-governed manipulation of strings of interpretable symbols” (Walmsley, 2012, p. 48). In GOFAI approaches there are explicit and discrete programming steps that are coded and executed line-by-line. If a particular program produces a surprising or unanticipated result, the programmer can review the source code, identify exactly where the “mistake” occurred, and institute an alteration.

6. For a critical reconsideration of this formulation of moral responsibility as it applies (or not) to recent innovations in AI and robotics, see Gunkel (2012b).

7. For example: "It's widely accepted that creativity can't be copied by machines. Reinforcing these assumptions are hundreds of books and studies that have attempted to explain creativity as the product of mysterious processes within the right side of the human brain" (Steiner, 2012, p. 1).

References

Barthes, R. (1978). *Image, music, text* (S. Heath, Trans.). New York, NY: Hill and Wang.

Cardoso, A., Veale, T., & Wiggins, G. A. (2009). Converging on the divergent: The history (and future) of the international joint workshops in computational creativity. *AI magazine*, *30*(3), 15–22. Retrieved from http://www.aaai.org/ojs/index.php/aimagazine/article/view/2252

Carlson, M. (2015). The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority. *Digital Journalism*, *3*(3), 416–431.

Clearwall, C. (2014). Enter the robot journalist: Users' perceptions of automated content. *Journalism practice, 8*(5), 519–531.

Colton, S. (2012). The painting fool: Stories from building an automated painter. In J. McCormack and M. d'Inverno (Eds.), *Computers and creativity* (pp. 3–38). Berlin: Springer Verlag.

Colton, S. & Wiggins, G. A. (2012). Computational creativity: The final frontier. In L. De Raedt, C. Bessiere, D. Dubois, P. Doherty, P. Frasconi, F. Heintz, & P. Lucas (Eds.), *Frontiers in artificial intelligence and applications*, (Vol. 242, pp. 21–26). Amsterdam: IOS Press Ebooks. http://ebooks.iospress.nl/volume/ecai-2012

Colton, S., Pease, A., Corneli, J., Cook, M., Hepworth, R., & Ventura, D. (2015). Stakeholder groups in computational creativity research and practice. In T. R. Besold, M. Schorlemmer, & A. Smaill (Eds.), *Computational creativity research: Towards creative machines* (pp. 3–36). Amsterdam: Atlantis Press.

Cope, D. (2001). *Virtual music: Computer synthesis of musical style*. Cambridge, MA: MIT Press.

Dörr, K. N., & Hollnbuchner, K. (2016). Ethical challenges of algorithmic journalism. *Digital journalism, 5*(4), 404–419.

Feenberg, A. (1991). *Critical theory of technology.* New York: Oxford University Press.

Ford, M. (2015). *Rise of the robots: Technology and the threat of a jobless future*. New York: Basic Books.

Foucault, M. (1984). What is an author? (J. V. Harari, Trans.). In P. Rabinow (Ed.) *Foucault reader* (pp. 101–122). New York, NY: Pantheon.

Georgia Tech. (2013). Robotic musicianship group: Shimon. http://www.gtcmt.gatech.edu/researchprojects/shimon

Go Ratings. (20 November 2016). https://www.goratings.org/

Google DeepMind. (2016). AlphaGo. https://deepmind.com/alpha-go.html

Green, P., & Wei-Haas, L. (1985). https://deepblue.lib.umich.edu/handle/2027.42/174

Gunkel, D. J. (2012a). Communication and artificial intelligence: Opportunities and challenges for the 21st century. *Communication +1*, *1*(1), 1–27. http://scholarworks.umass.edu/cpo/vol1/iss1/1

Gunkel, D. J. (2012b). *The machine question: Critical perspectives on AI, robots and ethics*. Cambridge, MA: MIT Press.

Gunkel, D. J. (2016). *Of remixology: Ethics and aesthetics after remix*. Cambridge, MA: MIT Press.

Gunkel, D. J., & Taylor, P. A. (2014). *Heidegger and the media*. Cambridge: Polity Press.

Haugeland, J. (1985). *Artificial intelligence: The very idea*. Cambridge, MA: MIT Press.

Heidegger, M. (1977). *The question concerning technology and other essays* (W. Lovitt, Trans.). New York: Harper & Row.

Hoffman, G. & Weinberg, G. (2011). Interactive improvisation with a robotic marimba player. *Autonomous robots, 31*(2–3), 133–153.

Hofstadter, D. R. (1979). *Gödel, Escher, Bach: An eternal golden braid*. New York, NY: Penguin.

Hofstadter, D. R. (2001). Staring Emmy straight in the eye—And doing my best not to flinch. In D. Cope (Ed.), *Virtual music: Computer synthesis of musical style* (pp. 33–82). Cambridge, MA: MIT Press.

Hui, F. (2016). AlphaGo games—English. *DeepMind*. https://deepmind.com/research/alphago/alphago-games-english/

Johnson, D. G. (2006). Computer systems: Moral entities but not moral agents. *Ethics and Information Technology*, *8*(4), 195–204.

Lewis, S. C., & Westlund, O. (2015). Big data and journalism: Epistemology, expertise, economics, and ethics. *Digital journalism*, *3*(3), 447–466.

Lyotard, J. F. (1993). *The postmodern condition: A report on knowledge* (G. Bennington & B. Massumi, Trans.). Minneapolis, MN: University of Minnesota Press.

Majid al-Rifaie, M., & Bishop, M. (2015). Weak and strong computational creativity. In T. R. Besold, M. Schorlemmer & A. Smaill (Eds.), *Computational creativity research: Towards creative machines* (pp. 37–50). Amsterdam: Atlantis Press.

McFarland, M. (2016). What AlphaGo's sly move says about machine creativity. *The Washington Post* (15 March). https://www.washingtonpost.com/news/innovations/wp/2016/03/15/what-alphagos-sly-move-says-about-machine-creativity/?utm_term=.0c8281af53c9

Metz, C. (2016a, March 11). The sadness and beauty of watching Google's AI play go. *Wired*. https://www.wired.com/2016/03/sadness-beauty-watching-googles-ai-play-go/

Metz, C. (2016b, March 16). In two moves, AlphaGo and Lee Sedol redefine the future. *Wired*. https://www.wired.com/2016/03/two-moves-alphago-lee-sedol-redefined-future/

Metz, C. (2016c, March 10). Google's AI wins a pivotal second game in match with go grandmaster. *Wired*. http://www.wired.com/2016/03/googles-ai-wins-pivotal-game-two-match-go-grandmaster/

Montal, T., & Reich, Z. (2016). I, robot. you, journalist. Who is the author?: Authorship, bylines and full disclosure in automated journalism. *Digital journalism, 5*(7), 829–849. http://dx.doi.org/10.1080/21670811.2016.1209083

Painting Fool. (2017). *The painting fool*. http://www.thepaintingfool.com/

Plato. (1982). *Phaedrus* (H. N. Fowler, Trans.). Cambridge, MA: Harvard University Press.

Ricœur, P. (2007). *Reflections on the just* (D. Pellauer, Trans.). Chicago: University of Chicago Press.

Searle, J. (1984). *Mind, brains and science*. Cambridge, MA: Harvard University Press.

Simonite, T. (2016, June 8). Ok computer, write me a song. *MIT technology review*. https://www.technologyreview.com/s/601642/ok-computer-write-me-a-song/

Smith, S. V. (2015, May 20). An NPR reporter raced a machine to write a news story. Who won? *Morning Edition*. http://www.npr.org/sections/money/2015/05/20/406484294/

Steiner, C. (2012, July 27). Can creativity be automated? Computer algorithms have started to write news stories, compose music, and pick hits. *MIT Technology Review*. https://www.technologyreview.com/s/428437/can-creativity-be-automated/

Turing, A. (1999). Computing machinery and intelligence. In P. A. Meyer (Ed.), *Computer media and communication: A reader* (pp. 37–58). Oxford: Oxford University Press.

Walmsley, J. (2012). *Mind and machine*. New York: Palgrave Macmillan.

Whitby, B. (2011). On computable morality: An examination of machines as moral advisors. In M. Anderson & S. L. Anderson (Eds.), *Machine ethics* (pp. 138–150). Cambridge: Cambridge University Press.