



Do Humans STILL Have a Monopoly on Creativity or Is Creativity Overrated?

Gregory Cowart, Dane Williamson, Naha Farhat,
and Joon-Suk Lee^(✉)

Virginia State University, Petersburg, VA 23806, USA
{gcow3223, dwil2444}@students.vsu.edu,
nfarhat@vsu.edu, joonsukl@acm.org

Abstract. What is creativity? Do AIs have creativity? This research sets out to answer these questions. By conducting a small-scale user experiment, and quantitatively and qualitatively analyzing the data, this work aims to answer the following three research questions: (1) whether human participants can distinguish between AI generated and human composed music; (2) common preconceptions participants have on AI generated music; and (3) common constructs participants associate with human creativity. Our results show that people are not able to tell the difference between human and AI composed music.

Keywords: Turing test · Music Turing test · Creativity · Computational creativity

1 Introduction

1.1 Creative Machines

Once thought of as uniquely belonging to humans, creativity is a personality trait associated with our ingenuity to create novel artifacts. It has also been described as psychological processes that are somewhat like “magic” [2]. For years, educators and researchers have worked on finding ways to harness these magical, creative thinking processes to improve student learning (e.g., [15, 17, 21]). Creativity was what differentiated us from apes.

However, with the recent surge of Artificial Intelligence (AI), researchers are starting to use the term, *computational creativity*, to describe computing processes and algorithms that can be used to carry out tasks that once only humans were able to perform. Anna Jordanous describes the goal of computational creativity as “to model, simulate or replicate creativity using a computer [7].” For instance, advancements in deep learning and neural network systems have enabled developers to create artificial agents that are capable of creating music, a capability that has long been thought innate to humans. Today, we are witnessing an increasing number of artificial intelligence agents that compose music (e.g., [12, 18, 22]), write short stories (e.g., [11]), paint art pieces (e.g., [3]) and even create other software (e.g., [16]).

1.2 Is Creativity Overrated?

According to Merriam-Webster [14], the English word, “*creative*” comes from Medieval Latin *creātivus* which means “to beget” or “to give birth to.” In this sense, being creative just means being able to create or make something or anything. Yet when we say “creativity,” we do not typically mean a mere ability to create random objects, but we refer to certain abstract qualities that creators bring and imbue into their built artifacts. Margaret Boden, for instance, defines creativity as “the ability to come up with ideas or artefacts that are new, surprising and valuable” in her book, *The Creative Mind: Myths and Mechanisms* [4].

Yet, as we have already seen, modern day computer agents are capable of creating seemingly *new*, *surprising* and *valuable* artifacts. Does this mean humans no longer have a monopoly on “creativity”? Is “creativity” an overrated concept? This research starts with these questions.

Some still argue and believe that artists infuse intuitive, abstract, human qualities such as beliefs or emotions into their work when they create. Such qualities are believed to be impossible to algorithmically compute. If so, this research also aims to explore what are the qualities that differentiate AI produced artifacts from human created art works.

1.3 Research Questions

If we are to ask if creativity is still a unique human trait or if it is simply an overrated concept that labels nothing more than a collection of computational processes, we need to first know what creativity is. We should also be able to measure and quantify “creativity” if we are to compare human creativity with computational creativity. Yet, it seems that there is no real consensus on what creativity actually is. As already mentioned, if creativity simply means creating any artifacts, AI agents as well as beavers already possess creativity. However, different researchers define creativity in different ways. Moreover, many of the concepts used to define creativity are abstract and not measurable. Boden uses three terms, “new, surprising and valuable [4]” to define creativity. These three qualities are all abstract and subjective terms, and are not objectively measurable.

One way to compare machine creativity with human creativity would then be by operationalizing the term, creativity, and concocting measurable factors associated with the operationalized definition. However, doing so will only show how machines and humans are different or similar in respect to the way that we define and operationalize what creativity is. So instead of trying to quantify creativity and objectively compare the work of humans and that of AI, this research aims to explore (1) if and how human participants can distinguish between AI generated and human composed music. We also examine (2) common preconceptions participants have on AI generated music, and (3) common constructs participants associate with human creativity. In other words, we do not try to objectively measure and compare human and AI creativity, but rather study and explore how human participants subjectively view what human and AI creativities are to them.

2 Related Work

2.1 Research on Creativity and Four Ps

While arguing that creativity should be viewed not as a theoretical construct but rather as a multifaceted rubric, MacKinnon [13] proposes an analytical framework consisting of four main aspects—process, product, person, and situation—for understanding human creativity. Similarly, Rhodes [10], in an attempt to develop an instructional design model to enhance student learning as well as creative thinking, defines creativity using the same four factors.

Jordanous adopts these four factors from MacKinnon and Rhodes, and delineates computational creativity using a similar framework. She argues that computational creativity should also be evaluated using the four factors, and calls her framework the Four Ps [8].

The Four Ps are defined as follows [8]:

Person—creative individual

Process—creative process that the creative individuals take

Product—creative artifacts produced as a result of creative processes

Press—environment in which creativity is situated

In computational creativity, the Person or creative individual can refer to the machines that generate creative artifacts or the developers who create the machines. The Process can refer to the algorithms that are used in making creative artifacts. The Product is the output or artifacts generated from the Process. The Press can refer to the audience that receives and interprets the Product generated by the machine.

Yet other researchers define and operationalize creativity in different ways. Acknowledging that the debate on whether computer programs are capable of being genuinely creative can become purely philosophical one, for example, Ritchie [19] proposes an alternate and more empirical way of operationalizing computational creativity and provides a list of measurable criteria that can be used to estimate “interesting” properties of a given computer program.

2.2 Turing Test

The notion of AI exhibiting some form of creativity or intelligence has been researched for decades. In 1950 computer scientist Alan Turing proposed a test for evaluating whether a computer can show intelligence equivalent to a human being [23]. The Turing test (TT) has since then become the standard of evaluating machine intelligence. TT gauges an AI agent’s intelligence by having the AI chatbot interact with a person, and tasking another person, a judge, to determine which of the two interlocutors is a computer. The test is repeated over multiple sessions and if the judge fails to identify the machine interactant in more than 50% of the trials, the machine is deemed to possess the ability to think.

Although TT is typically regarded as having a significant impact on shaping the field of Artificial Intelligence, it has also been heavily criticized over the years for not being an appropriate way of evaluating machine intelligence, and for promoting

trickery by chatbot developers in order to get their machines to pass the test. For example, in 2014, Eugene Goostman, a chatbot created by Vladimir Veselov and colleagues, managed to pass TT by convincing the judges that it was a human [1]. How the chatbot managed to pass the test, however, has been criticized as mere trickery with nothing to do with intelligence. Eugene was modeled to imitate a 13-year-old boy from Odessa, Ukraine. When interacting with the bot, the judges accredited Eugene's sometimes unintelligible English, and sporadic and sudden topical changes in conversation to his age and the country of residence, thus fooling them.

As an alternate method to measure machine intelligence, Bringsjord [5] proposes another way of evaluating the creative capabilities of machines. This test is named after Ada Lovelace. The Lovelace Test (LT) looks at the three-way relationship between an AI agent, its output, and the creator of the AI agent. In LT, an AI agent passes the test and is considered to possess intelligence only if the creator cannot account for how the AI agent produces its output. Bringsjord paraphrases Lovelace's objection against TT, pointing out that machines cannot genuinely create anything. According to Bringsjord, Lovelace believed that "only when computers originate things should they be believed to have minds [5]."

LT was an intriguing concept when it was first introduced two decades ago. Back then, creating a machine and not being able to fully understand its inner workings might have been an absurd idea. Yet today the Artificial Intelligence community is experiencing the exact opposite problem. Due to ever evolving neural networking and deep learning algorithms, the behaviors of AI agents are not always explainable. Consequently, governments are authoring regulations to safeguard consumers from the potential issues and problems unexplainable AIs might cause. In 2018, for instance, the European Union's General Data Protection Right (GDPR) began to include a right to explanation in order to hold AI creators responsible for their creations' actions. In other words, not being able to explain AI agents' actions is no longer science fiction. It is already commonplace.

2.3 Artificial Intelligence Generated Music

Music-making machines are already with us. There exist several commercially available AI systems that produce music. For example, IBM Watson Beat (<https://www.ibm.com/case-studies/ibm-watson-beat>) and Google Magenta (<https://magenta.tensorflow.org/>) platforms utilize big training datasets to teach and build machines that can compose music. Researchers of the BachBot project use an AI agent called deepBach to produce classical music pieces [12, 22]. The researchers state that the goal of the project is to build an AI agent capable of generating chorales that are indistinguishable from that of J.S Bach. This was accomplished through training their deepBach model on the chorale patterns in music by J.S Bach [22]. JukeDeck [9] is another AI agent that uses deep neural networks and machine learning to analyze music patterns in order to generate songs. Taiwanese researchers (AILabs.tw) [18] also conducted research using an AI music generation tool similar to deepBach. In our research, we borrow these tools.

3 Methodology

The analysis in this paper is based on data from an exploratory user study we conducted last year.

3.1 Participants

Participants were recruited from a pool of students and faculty with varying musical and technical backgrounds from Virginia State University, a historically black public land-grant university in Petersburg, Virginia. A total of ten participants (3 female, 7 male) enrolled in the study. The participants' age ranged from 20 to 68 (M: 29.30, SD: 14.45). Eight participants (80%) identified themselves as African American, one (10%) as Hispanic, and one (10%) as Asian or Pacific Islander. Ten participants included two Computer Science professors and one Music professor. Participants received no monetary compensation for participating in the study. Table 1 shows summary statistics for participants' gender, classification, ethnicity, major and music experience.

Table 1. Participant summary statistics.

Variable	Percentage	Variable	Percentage
Gender		Major	
Male	70%	Computer Science	90%
Female	30%	Music	10%
Ethnicity		Classification	
African American	80%	Faculty	30%
Hispanic	10%	Graduate Student	20%
Asian/Pacific Islander	10%	Senior	20%
Music Experience		Junior	20%
Expert	10%	Sophomore	10%
Advanced	10%		
Intermediate	40%		
Novice	40%		

3.2 Procedure

The study was designed as a four-phased experiment. In phase 1, participants were tasked with listening to a randomized playlist of six songs and then identifying which songs were composed by human or by AI. Out of six songs, three were AI composed songs. Using a set of publicly available human and AI generated music on an online music distribution platform, SoundCloud (<https://soundcloud.com/>), we reiterated research done by Taiwan AI Labs (AILabs.tw) [18]. The six songs were randomly selected from the ten songs that are available on the *Learner Deep* page on SoundCloud (<https://soundcloud.com/learner-deep>). The answer set (information on which songs were generated by AI and which were by human composers) as well as the statistics from the original experiment done by Taiwan AI Labs are available on [18].

In phase II, participants were asked to take a similar test, but using the online Bachbot test [12, 22]. The online BachBot challenge first asks users to identify their age group, and to self-rate their music experience using multiple choice questions. The age group options are (1) under 18, (2) 18 to 25, (3) 26 to 45, (4) 46 to 60 and (5) over 60, and the music experience options are (1) novice, (2) intermediate, (3) advanced and (4) expert. Users are then presented with two songs and asked to identify which song is composed by Bach and which is by deepBach, an AI agent trained on the chorale patterns in music by J.S. Bach [12]. The BachBot challenge consists of five sets of music clip pairs. After listening to the two clips, users are expected to select a clip that is most similar to Bach. Unlike the songs provided on the Leaner Deep site, BachBot does not disclose answer keys, probably because the authors of BachBot are still gathering experimental data through their site. We were only able to get the percentage of correctness rate of user answers.

In phase III, participants were tasked with evaluating five AI generated songs created from JukeDeck (<https://www.jukedeck.com/>). JukeDeck is another AI agent that uses neural networks and machine learning to analyze music patterns in order to generate songs. The creators of the technology allow users to go to their website and generate customizable songs. Prior to the experiment, we used JukeDeck to prepare five AI generated songs. The generated pieces were piano solo that last for a minute each. The songs were generated with themes alternating between melancholic and upbeat. In this phase, participants were not informed about the fact that they would only be hearing AI generated music. The rationale for including only AI generated music in this phase was two-fold: finding compatible human composed music is challenging, and we also wanted to explore different rationales participants associate with their choices regardless of whether or not they were correct.

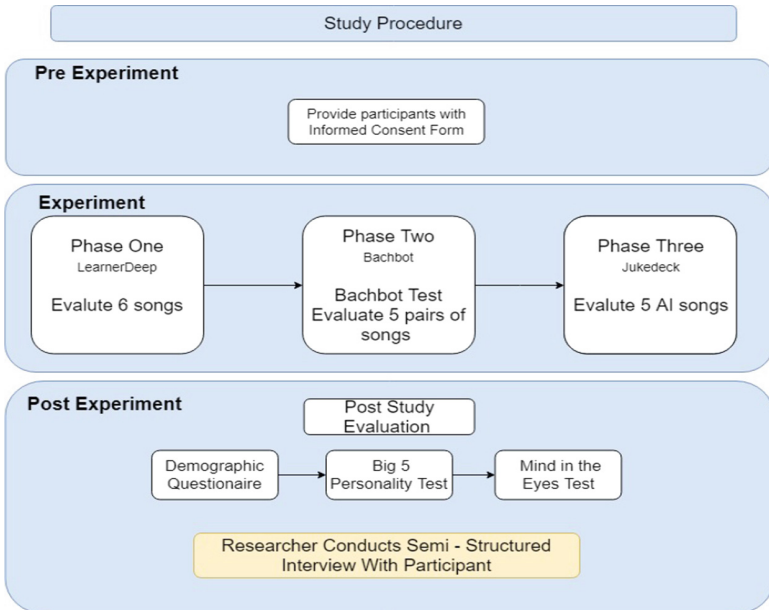


Fig. 1. Music Turing test - study protocol diagram

After the music listening sessions, participants were asked to fill out three types of post study questionnaires. Researchers then explained the nature of the research and conducted semi-structured interviews with the participants. During the interview, we asked how and why participants thought certain songs were composed by AI or by a human, and their overall views on creativity. The study protocol is illustrated in Fig. 1.

4 Data Analysis

4.1 Quantitative Analysis

Phase-1 Data Analysis

In phase 1, participants were presented with 6 songs in random orders, and asked to identify whether the songs were composed by human or by AI. The percentage of participants who correctly identified the type of the song in trial#1 was 60%. This percentage then dropped to 50% in trial#02 and then to 30%, then went back up to 60% and then dropped to 40% then 30% as shown in Table 2. To test whether these percentages are significantly higher than 50%, the chance of guessing correctly at random, Fishers exact test has been conducted and all the p-values were not significant (Table 2, p-value > alpha = 0.05), which indicates that there is not enough evidence to conclude that the percentage of participants who correctly identified the song creators was significantly higher than 50%. In other words, our participants weren't able to distinguish AI generated songs from human composed ones.

Table 2. Phase 1: percentages of incorrect and correct responses for all six songs.

Trial#	Correct %	p-value
Trial #1	60%	0.7539
Trial #2	50%	0.9997
Trial #3	30%	0.3438
Trial #4	60%	0.7539
Trial #5	40%	0.7539
Trial #6	30%	0.3438

Phase-3 Data Analysis

In phase-3, participants were presented with five songs and asked to identify whether the songs were composed by human or by AI. In this phase, however, all five songs were AI generated. Figure 2 shows the percentages of correct responses. 40% of the participants were able to correctly identify the song creator as AI for songs 1, 3 and 5, while only 30% were able to identify song 4 correctly, and only 10% for song 2. The percentages of participants who correctly identified the song creator in this phase were smaller than the percentages in phase-1.

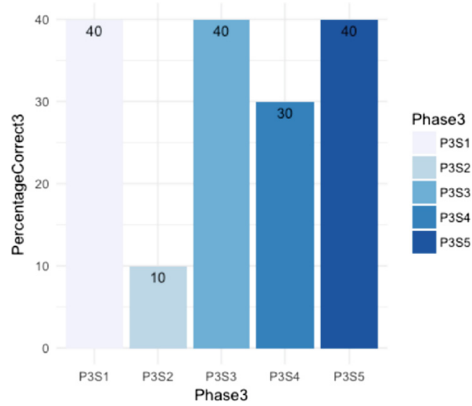


Fig. 2. Phase 3 percentages of correct responses

Overall Scores

For phase-1, we calculated the overall total number of correct responses for each participant for all six songs, three AI songs and three human generated songs. Table 3 displays the results of the overall score where the median response for correctly identifying the type of the six songs in phase-1 is 3 songs and the IQR is 2, based on a sample of 10 participants. For phase-3, the median correct response was 2 and the IQR was 0.25, which indicates that when only AI songs are used to test participants' ability to distinguish between AI and human songs, we have smaller variability in terms of the response of the participants. The phase-3 test can be considered as a more accurate way of testing participants than the phase-1 test since it reduces the chance of making random guesses and still ending up getting high scores.

Figure 3 shows the distribution of the total correct responses for phase-1 and phase-3. The median number of correct responses is smaller for phase-3 as compared to phase-1. We believe that this might be due to the fact that phase-3 has only AI songs which reduces the probability of correctly guessing the answers as compared to phase-1 which included three AI and three human generated songs.

The Wilcoxon signed rank test was used to test whether the number of correctly answered responses were significantly higher or lower than a $\frac{1}{2}$ chance (50%). Due to the small sample size, we used median values for the test. Using the Wilcoxon test, we compared the median of the number of participants' correctly responded answers with the probabilistic median. The probabilistic median is 3 for phase-1 and 2.5 for phase-3. For phase-1, the median of the participant's correct responses was not significantly different from the probabilistic median. For phase-2, the median value of the number of correctly guessed songs was significantly lower than the probabilistic median (Table 3). This significance might have resulted because of the small sample.

Table 3. Summary statistics for $n = 10$ participants' Phase-1 and Phase-3.

Phase	Median (Correctly answered)	IRQ	p-value (Wilcox test)
Phase-1	3	2	0.3039
Phase-3	2	0.25	0.0021*

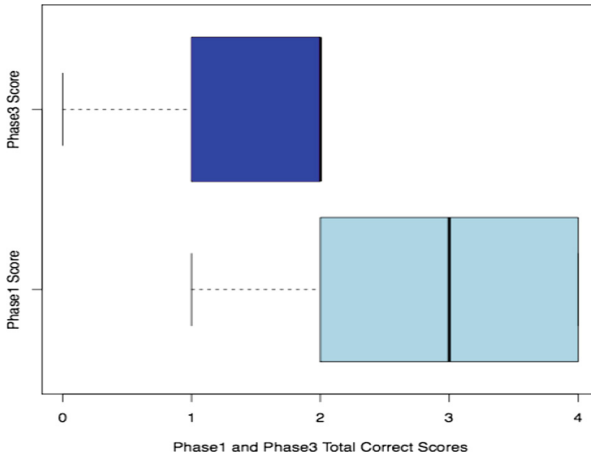


Fig. 3. Distribution of total correct responses for phase-1 and phase-3.

4.2 Qualitative Analysis

We also conducted a thematic analysis on the interview data in order to identify patterns of meanings in the participants' responses. The interview data were fully transcribed and read multiple times by the first and second authors. During the initial analysis, all researchers looked at the interview data together, and developed an initial coding scheme. The first author is currently going through the interview data in multiple iterations and conducting open and axial coding [6, 20]. The analysis is still in process. In this section, we present some of the interesting findings from our preliminary data analysis.

Participant's Perception of Creativity

As indicated earlier, we weren't so much interested in defining the objective and definitive meaning of creativity. Rather, our focus was to understand how each individual perceives the notion of creativity. We also wanted to explore how individuals' perception of creativity impacts their music listening experiences. Multiple participants said that creativity is an ability to create something. For instance, P2 said, *"I feel like creativity is making something, anything. Whether it be like writing a book or drawing something or making music. [It] is like creativity is just taking nothing and making something out of it, or taking something that exists and changing it in a way that makes it your own."*

Some participants associated creativity with brining in personal experience, being imaginative or expressing oneself. P10 said creativity is *"how you express yourself. So basically if an AI can express itself, it should be creative enough?... I just think it's how basically you could express yourself like through drawing or music what have you."* This response is in particular interesting because even though the participant stated that a machine is creative should it possesses an ability to express the self, the response ties creativity with the notion of self-awareness and self-conception. In other words, machines cannot possess creativity without first becoming self-aware. While self-aware

robots are a popular topic in Science Fiction, it is also a topic some AI researchers are trying to tackle these days.

Music Is as Good as How the Listeners Listen

Multiple participants mentioned that they were quite surprised by the quality of music AI could generate. P7, a professor in the Music Department, stated that she initially did not expect the machines to show emotions. However, at the end of the study, she said *“I heard some emotion in some of the songs, and I wasn’t so sure that machines could create that. But they might, they are a lot more sophisticated than they were before.”* P3 also expressed surprise when he stated:

“I think honestly without having been told that any of these pieces were AI created I would not be able- I wouldn’t tell you that a human did not make this piece. So it was difficult to discern what is human or what is AI. Because I can come in this room and if I hadn’t known it was a study- if I didn’t know it was a study, I would assume that they’re all human pieces.” (P3)

In addition, multiple participants mentioned that they would still appreciate the music, no matter who created it, if it is good. For example, P1 and P2 stated:

“I’ll probably appreciate them the same honestly. I don’t see the point in not appreciating anything just because something someone else made it, I mean a computer made it” (P2)

“I enjoy music of all kinds, whether it be made by human whether a dog with a synthesizer did it, I don’t care okay if I enjoyed it I enjoyed it” (P1)

5 Discussion and Conclusion

In this study, we asked (1) whether human participants can distinguish between AI generated and human composed music, (2) common preconceptions participants have on AI generated music, and (3) common constructs participants associate with human creativity. Our preliminary data analyses show that our participants weren’t able to tell the difference between AI generated and human composed music. Participants typically associated creativity with human qualities such as having and bringing in personal experience, being imaginative or expressing oneself. Yet many participants were quite surprised to hear AI generated music. One participant even showed fear when she said, *“it’s a little frightening because you don’t want to be displaced by machines. It’s like how people don’t want automation to come through and take their jobs... like I just lost my job to a machine.”*

References

1. Aamoth, D.: Interview with Eugene Goostman (2014). <http://time.com/2847900/eugene-goostman-turing-test/>. Accessed 30 Jan 2019
2. Arieti, S.: Creativity. The Magic Synthesis. Basic Books, New York (1976)
3. Baraniuk, C.: Artificially intelligent painters invent new styles of art (2017). newscientist.com. <https://www.newscientist.com/article/2139184-artificially-intelligent-painters-invent-new-styles-of-art/>. Accessed 10 Feb 2019

4. Boden, M.A.: *The Creative Mind: Myths and Mechanisms*, 2nd edn. Routledge, New York (2003)
5. Bringsjord, S., Bello, P., Ferrucci, D.A.: Creativity, the Turing test, and the (better) Lovelace test. *Mind*. *Mach.* **11**(1), 3–27 (2001)
6. Corbin, J., Strauss, A.: *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. SAGE Publications, Inc., Thousand Oaks (1990)
7. Jordanous, A.: What is computational creativity? (2014). [creativitypost.com. http://www.creativitypost.com/science/what_is_computational_creativity](http://www.creativitypost.com/science/what_is_computational_creativity). Accessed 17 Jan 2019
8. Jordanous, A.: Four PPPPerspectives on computational creativity in theory and in practice. *Connect. Sci.* **28**(2), 194–216 (2016)
9. Jukedeck: Jukedeck - Create unique, royalty-free AI music for your videos. <https://www.jukedeck.com/>
10. Rhodes, M.: An analysis of creativity. *Phi Delta Kappan* **42**(7), 305–310 (1961). <http://www.jstor.org/stable/20342603>
11. Lewis, D.: An AI-written novella almost won a literary prize (2016). [smithsonian.com. https://www.smithsonianmag.com/smart-news/ai-written-novella-almost-won-literary-prize-180958577/](https://www.smithsonianmag.com/smart-news/ai-written-novella-almost-won-literary-prize-180958577/). Accessed 10 Feb 2019
12. Liang, F., Gotham, M., Tomczak, M., Johnson, M., Shotton, J.: The BachBot Challenge (2016). <http://bachbot.com>. Accessed 27 Jan 2019
13. MacKinnon, D.W.: Creativity: a multi-faceted phenomenon. In: Roslansky, J.D. (ed.) *Creativity: A Discussion at the Nobel Conference*, pp. 17–32. North-Holland, Amsterdam (1970)
14. Merriam-Webster: *Webster’s New Collegiate Dictionary*. Merriam-Webster, Massachusetts (1977)
15. Miller, L.D., Soh, L.-K., Chiriacescu, V., Ingraham, E., Shell, D.F., Hazley, M.P.: Integrating computational and creative thinking to improve learning and performance in CS1. In: *Proceedings of the 45th ACM Technical Symposium on Computer Science Education (SIGCSE 2014)*, pp. 475–480. ACM, New York (2014). <https://doi.org/10.1145/2538862.2538940>
16. Murali, V., Qi, L., Chaudhuri, S., Jermaine, C.: Neural sketch learning for conditional program generation. In: *Proceedings of Seventh International Conference on Learning Representations (ICLR 2018)* (2018)
17. Peteranetz, M.S., Wang, S., Shell, D.F., Flanigan, A.E., Soh, L.-K.: Examining the impact of computational creativity exercises on college computer science students’ learning, achievement, self-efficacy, and creativity. In: *Proceedings of the 49th ACM Technical Symposium on Computer Science Education (SIGCSE 2018)*, pp. 155–160. ACM, New York (2018). <https://doi.org/10.1145/3159450.3159459>
18. Rajon: AI music composition passed Turing test[Taiwan AILabs 2018]. <https://ailabs.tw/human-interaction/ai-music-composition>. Accessed 27 Jan 2019
19. Ritchie, G.: Some empirical criteria for attributing creativity to a computer program. *Mind*. *Mach.* **17**(1), 67–99 (2007). <https://doi.org/10.1007/s11023-007-9066-2>
20. Saldana, J.: *The Coding Manual for Qualitative Researchers*. SAGE Publications Ltd., Thousand Oaks (2015)
21. Shell, D.F., Soh, L.-K., Flanigan, A.E., Peteranetz, M.S., Ingraham, E.: Improving students’ learning and achievement in CS classrooms through computational creativity exercises that integrate computational and creative thinking. In: *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education (SIGCSE 2017)*, pp. 543–548. ACM, New York (2017). <https://doi.org/10.1145/3017680.3017718>
22. Tomczak, M.: *Bachbot*. MPhil dissertation, University of Cambridge (2016)
23. Turing, A.M.: Computing machinery and intelligence. *Mind* **59**, 433–460 (1950)