WILLIAM SHAIKESPEARE?: AI AND THE QUEST FOR CREATIVITY

Armaan Kalkat

Background

When attempting to get to the heart of what makes humanity unique, one facet of our intelligence stands out: creativity. In our history, we learn of countless human beings, including the very first of our species, who utilized their intelligence to go beyond what was already possible and ushered in a new era of human ingenuity, from the very first tools to our modern artistic or scientific innovations. It therefore stands to reason that in our attempts to construct a new form of intelligence—Artificial intelligence—that the problem of making AI more creative should stand as one of the primary barriers between simple computation and true intelligence. The question of the possibility of creativity in artificial intelligence programs is not a new one, and was in fact being seriously considered by researchers in the 1990s (e.g. Boden, "Creativity and Artificial Intelligence"), and was much earlier being imagined by pioneers like Alan Turing ("Computing Machinery and Intelligence") and science fiction writers. Much has changed since these early days, including the primary methods by which AI functions, especially when considering the explosion in AI research focused on the creation of neural networks which aim to emulate the human brain's structure. While experiments into AI creativity have not been nearly as successful as other applications of the technology such as analyzing massive amounts of raw data or learning to perform a single task like playing chess, progress is constantly being made in this field, suggesting that it would be unwise to pre-emptively deny its promise. As such, an important question for the future of the field is exactly what the current trends in AI research tell us about the possibilities for cultivating this unique ability. While AI creativity is currently still in its infancy, its trajectory of development and the progress the field has made suggest that as the technology develops, AI will only become more skilled at creating artistic works and combining concepts in a meaningful way like humans.

What is Creativity?

From a Neurocognitive Perspective

The basis for our knowledge of any cognitive process must be a combination of both general observation of the phenomena as well as the use of neuroscience techniques to understand the brain circuitry required to accomplish the task. If AI researchers are to make headway in introducing creativity to artificial intelligence programs, therefore, they must first have a good scientific basis for where creativity comes from and how it arises in the brain in order to be able to create similar systems in a program. However, as psychology professor Arne Dietrich recognizes in his 2004 paper, there has been relatively little research specifically connecting neurocognition to creativity (1011). Furthermore, Margaret Boden, a long-time researcher in creativity, neuroscience and AI, states that even if neuroscientists are able to pinpoint using neuroimaging methods *where* something creative happens in the brain, this doesn't get us any closer to the *how* of what is actually happening in that specific cluster of neurons or broader brain region, a major current issue in neuropsychology ("Creativity as a Neuroscientific Mystery" 5). Despite these limitations, some general conclusions can be made from the research on the neural basis for creativity, and these understandings can be applied in some ways to the potential for creative AI systems.

Firstly, it is important to recognize that creativity, while certainly complex, is viewed from the cognitive perspective as a combination of basic cognitive processes rather than having any mysterious origin (Boden, "Creativity as a Neuroscientific Mystery" 3). Dietrich lists some of these processes: working memory (which allows information to be temporarily stored and, key to creativity, transformed), attention (which allows us to filter out irrelevant information and key in on important features), and temporal integration (allowing us to combine discrete pieces of information and perceive them as one coherent stream) (1013). Interestingly, all of these functions take place in the prefrontal cortex, an area of the brain which has long been associated with higher-order cognitive processes in humans such as planning, knowledge of the self, etc. (Dietrich 1013). The primary function of the prefrontal cortex is to provide a space for the raw information from the senses and memory systems to be combined and transformed, providing a rough outline for creativity as a general cognitive skill: it is the ability to take the already known information about an object or idea and transform it in some way to create something new, like realizing that a stick and a rock can combine to form a tool, merging the properties of these two previously disconnected mental objects and understanding how this combination could be useful.

Of course, creativity in the way people generally conceive of it, in terms of art and much higher order cultural innovations, is not quite so simple, but AI must be able to solve both of these problems—the ability to combine concepts in the first place and apply these new concepts in some form such as words, music, visual art etc.—in order to be said to be fully creative. Dietrich argues that the other primary feature of creativity aside from novelty is "appropriateness," which cannot be explained by the simple combination of concepts within the brain but also must include an additional hierarchical layer of processing which assesses the usefulness or quality of novel combinations based on previous experiences (1015). Finally, the creative insight which is the product of novelty and appropriateness must also rely on communication between various parts of the brain such as the motor system or linguistic systems (depending on the type of creativity) to actually put the creative insight into action or expression through a formulated plan. Combining these insights from psychological researchers, creative AI systems have quite a tall order ahead of them: they must be able to store information, have a "cognitive space" to manipulate said information, have some way to realize that cognitive manipulation or creation outside of their own cognition, and judge their own output. Researchers are, however, currently working on the problem of creating conceptual space wherein an AI can analyze and logically combine concepts (Eppe et al.) and on another track, researchers have created a number of systems to allow AI to create a coherent and structured poem which could, with human help, even fool judges into thinking it came from the mind of a human poet (Köbis & Mossink). While there are, of course, limitations to this kind of research which will be delineated later in the paper, the psychological research has allowed us to slowly but surely penetrate the mysteries of the working of the mind behind creativity, providing an actionable framework for AI researchers to work from.

From a Philosophical Perspective

While psychologists and neuroscientists are now attempting to discover the systematic workings of human creativity in the brain, philosophers since Plato have been interested in defining the concept itself and understanding its value for human society. Berys Gaut, a professor at the University of St. Andrews who studies the philosophy of creativity, states its widely accepted definition as "the capacity to produce things that are original and valuable" (1039). This definition is intended to ensure that works which are wholly derivative from previous works and those which are technically original but entirely worthless (such as a random string of letters) are not deemed creative (Gaut 1039). Gaut raises an issue with this definition, however, in that it does not include the necessary prerequisites of agency (he brings up the example of tectonic plate movements which

produce diamonds as not being creative) and understanding or skill (he contends that accidentally knocking paint onto a canvas cannot be creative), combining these two qualifications into the somewhat vague quality of "flair" (1040-1041). The Harvard philosopher Sean Dorrance Kelly goes even further, stating that the mark of true creativity additionally requires the work to have some level of societal meaningfulness; he contends that the reason Schoenberg is remembered as a great musical genius is because his radical system of composition is interpreted by others as having been valuable in response to the needs of human society at the time and not simply because it is original or artistically valuable. Under this paradigm, the artist is someone who brings to light a new and valuable way of interpreting the world itself (Kelly). As such, he argues that an AI can never reach this standard of creativity because it is not socially embedded; it is instead only formulaically attempting to imitate a human process and its output can therefore never be interpreted as having the same level of meaningfulness (Kelly). Gaut takes a less firm stance, arguing that creativity does take understanding, but leaving the question relatively open whether a computer can or cannot understand (1038).

Although AI may not be socially embedded in the traditional sense, the general principle behind machine learning is to "train" an AI using examples of previous works and then allow the AI to conceptually transform this understanding into a new output (Köbis and Mossink 4), meaning that the AI is, in some sense, using an understanding of previous art to create new art. Even Kelly grants that the artist does not have to be necessarily conscious of the societal value of the work, although he does posit a requirement that the work must not be created by chance, something which he sees as applying to AI art. It can be argued that an AI which uses predictive text methods to create poetry is essentially equivalent to a monkey at a typewriter (to use Kelly's analogy), but this process cannot be said to be truly random or accidental as the AI is taking into account some level of experience when making its decisions rather than only brute force, meaning that AI may not be quite as formulaic as Kelly contends. Taking Gaut and Kelly's arguments together, it is clear that AI creativity must meet a relatively high bar to be accepted by philosophers, but it is also apparent that it is not inherently an impossibility.

Current Areas of Research

Conceptual Blending Studies: How can AI Combine Concepts?

As we have seen, one of the basic foundations for AI creativity is the ability to combine multiple concepts into new iterations based on their shared traits, a phenomenon which is termed combinatorial creativity (Confalonieri and Kutz 481). While this may seem like a simple and automatic process in our own minds, AI researchers Confalonieri and Kutz warn that to implement it in artificial intelligence is "a highly complex, multi-paradigm problem" (479). The basis of this kind of system is the framework of "conceptual blending" which was first proposed by the linguists Fauconnier and Turner in 2003 and entails a system by which some common features of two mental concepts (called the "generic space") are used to combine the two concepts into something new and useful (Confalonieri and Kutz 481). The classic example given by Confalonieri and Kutz is the creation of the concept of a "house-boat" by combining some parameters of the concept of house—such as the fact that it exists stationary on land and houses people—with the parameters of a boat—such as the fact that it moves through water and carries passengers—to create a new kind of dwelling which would could either house people stationary on water or carry them around on water while housing them, or any other number of novel coherent combinations of the concepts of house and boat (500-501). The AI system would therefore need to be able to conceptualize the basic features (the ontology) of "house" and "boat" and rely on some kind of guiding principle to evaluate which features of the two concepts should be kept and which should be forgotten (the houseboat

necessarily requires a removal or weakening of the original necessity of a house being on land to be coherent, for example; Confalonieri and Kutz 501) as well as how useful the output would be (487).

Eppe and colleagues also use a similar framework to posit a potential way conceptual blending could be applied to a creative endeavor such as the composition of music (119). By applying a similar algebraic approach to find the conceptual space which could be shared by two chord progressions and tailoring the output to follow some musicological parameters to avoid dissonance in the notes, the system which Eppe et al. designed was able to combine elements of the phyrigian and perfect cadences (two very old chord progressions) to create the tritone substitution, a progression which was popularized by jazz musicians (119). Detractors may argue that by applying such formulaic rules to something like music creation, the beauty and spontaneous nature of innovation is destroyed, but AI researchers would simply counter that they are applying the little we know about the way creativity takes place in the brain and recreating it in a computational system, meaning that we are simply peering behind the scenes, as it were, which will inevitably remove some of the mystique which our culture has created around creativity through stories of divine inspiration or muses. While there has yet to be an experiment with AI music composition or any other type of creativity which has spontaneously led to a radical paradigm shift vis a vis Kelly's argument, these foundational experiments serve to show that AI can, on its own, apply some process of combinational creativity to create something novel and aesthetically pleasing. Furthermore, researchers like Graeme Ritchie have long been attempting to create better frameworks for assessing the creativity of an AI's output, assessments which he states could potentially be integrated into the AI's own creative process (Ritchie). As a result, the goal of creating an artificially intelligent independent creative agent which can continuously monitor its own output and grow in its abilities like a human artist begins to feel more and more plausible.

A Case Study: AI Poetry

Overview of AI Poetry Generation Methods

Poetry is a somewhat unique art form in that it requires a very broad range of skills, such as the ability to concisely but clearly convey a message, incorporate paralinguistic factors such as rhythm and rhyme, and often an understanding of how to break or bend the rules of language to fit one's goal while still being understandable. As a result, it represents a major challenge to computational creativity researchers, leading to a wealth of research in the creation of AI poetry. In his report "Automatic Generation of Poetry: An Overview," Natural Language Processing (NLP) researcher Hugo Oliveira details a number of broad categories of AI poetry generation methods which have been developed over the years, each having a set of strengths and weaknesses. These categories include Template Based Poetry Generation (which uses poetry templates and inputs words which fit certain parameters), Generate and Test approaches (which generate random word sequences in accordance with a set of constraints and test each one to find the best output), Evolutionary approaches (which utilize the principles of evolution such that the least fit poems are eliminated systematically according to an algorithm) and Case-Based Reasoning approaches (wherein a relevant existing poem is adapted to a user-provided target message; Oliveira 2). Furthermore, Oliveira describes the three primary goals of an AI poetry generation program as having meaningfulness (the poem conveys a conceptual and meaningful message), grammaticality (the poem obeys the grammar rules of the language), and poeticness (the poem contains poetic features such as use of meter, rhyme, form, etc.; 2). Under this paradigm, the ideal poetry generation method takes into account all three of the goals rather than only focusing on obeying grammaticality or sticking to a poetic form (Oliveira 2).

Two systems which fit this requirement but use different methods are discussed by Oliveira: ASPERA and McGonnagall. ASPERA is a system which follows the Case-Based Reasoning

approach (CBR) and asks the user for a description of the intended message and type of poem, allowing the system to select an appropriate meter and form for its output (Oliveira 3-4). ASPERA does not rely on modelling natural language in terms of syntax or lexicon, but instead uses a database of pre-existing verses to draft new lines using the same parts of speech of each word while matching the intended message. These outputted verses are then validated and corrected by a human user, with those validated verses going back into the database, allowing the program to grow and improve over time, albeit using human intervention rather than self-monitoring (Oliveira 4). McGonnagall, on the other hand, is a poetry generation system which uses an evolutionary approach such that a number of potential poems are generated based on the initial information provided to the program, including a meaning for the poem, and then each of these potential poems (called "individuals") are scored in an evaluation phase based on criteria such as adherence to a form, having a certain rhythm of stressed and unstressed syllables, etc. (Oliveira 4). This is then followed by an evolution phase where small random tweaks to the best poems are done (approximating the idea of genetic mutations), with these new individuals then being evaluated again and so on and so forth to eventually produce a poem which is most fit for its purpose (Oliveira 4).

While the output of programs like ASPERA and McGonnagall may approximate human poetry quite well, the question and debate remains as to whether this represents true computational creativity or simple mimicry. Even Oliveira, a proponent of AI poetry generation methods, points out the limitations of only relying on objective measures of the output as an assessment of whether an AI is creative or not, as this may influence researchers to focus more on taking shortcuts which *replicate* creative output rather than really finding methods to *recreate* it ("A Survey" 17-18). Researchers have tried to get around these limitations by introducing new frameworks to encourage a more comprehensive AI creativity such as the FACE framework, which requires the system to be able to frame its own choices in the form of an explanation of the context or logic behind its output, allowing the creator to better understand what process the generation system is using to make its decisions (Oliveira, "A Survey" 17-18). By keeping these limitations and philosophical issues in mind while developing creative AI systems, researchers can position themselves to be able to confidently say that AI systems are truly exhibiting some form of creativity rather than simply going through the formulaic motions. Despite this positive outlook, the day is still clearly far off from when an AI application which is not specifically designed for this purpose would, for example, spontaneously decide to write poetry, a recognition which allows us to reach a more realistic and nuanced understanding of the state of creative computation.

Passing the Turing Test: Can AI Poets Fool Humans?

While not everyone agrees that an ability to fool a human judge is the mark of creativity, it is undeniable that this possibility will play an important role in the potential societal implications of rapidly improving creative AI programs. For example, concerns have already been raised by the creators of such AI text generation programs as GPT-3 that if the tool were to fall in the wrong hands, it could potentially be used to create massive amounts of fake but realistic news (Metz). While machine-generated poetry would not have as much potential to cause political division or civil unrest, the same principles which apply to creating believable prose can apply also to poetry and vice versa, meaning that insights from AI poetry generation research could also play a role in creating other kinds of texts, not to mention the artistic and philosophical implications of acknowledging that a computer may be able to create poetry which humans might even prefer over the works of other humans in the future.

However, as computer scientists Köbis and Mossink state, little empirical research studying whether AI-generated poetry can fool human judges has been done (1). In their study, a neural-network based text generation tool (GPT-2), which is trained on a colossal amount of data from the

internet and can produce a wide variety of text genres, was asked to complete poems given the first two lines of a human-written poem, 10 of which were selected to compete against said human-written poems (4). Human judges were then asked to rate which poem was preferred and were monetarily incentivized to correctly identify which poem was written by an algorithm and which by a human (5). Köbis and Mossink found that when GPT-2 was competing with amateur human writers, participants could not accurately identify which text was algorithmically generated at a level above chance, although they did show a statistically significant slight preference for the human-written poems nonetheless (5). Additionally, they found that participants were overconfident in their abilities to identify GPT-2's poems, rating their confidence levels about their accuracy at 63 out of 100 on average, which is significantly above chance (5-6). Interestingly and contrary to expectations, judges did not prefer human-written poems more when told beforehand the origin of each poem, countering previous research on algorithm aversion wherein humans are hesitant to accept the work of algorithms even if they operate well (5).

In the second part of the study, Köbis and Mossink pitted GPT-2's poems against poems by famous professional poets like Maya Angelou and Hermann Hesse, and they also introduced two new conditions where either humans chose the best of GPT-2's outputs or the poems were randomly selected; this was done in order to understand how important human intervention is in the ability for AI to be perceived similarly to human writers (6-7). With these changes, Köbis and Mossink found that judges still preferred the human-written poems overall, with these preferences being stronger when humans were involved in screening GPT-2's output, and there were again no differences in preference when judges were told which poem was generated by the algorithm (9). In terms of detecting the origins of the poems, judges were able to accurately detect the AI-generated poem when the poem was randomly selected from GPT-2's output, but not when a human selected the best poems to be presented (8-9). Finally, when monetarily incentivized, judges were relatively accurate in their confidence levels in terms of detecting the origin of the poems correctly (9). Taken as a whole, these results suggest that AI text-generation tools, even those not specifically designed for poetry generation (such as GPT-2), can fool humans into thinking its output was written by an amateur human poet or a professional poet, as long as a human is involved in selecting the best of the output, making the algorithm more like a writing tool than a self-sufficient writer. However, the results also suggest that people generally prefer human-written poems, even though they cannot accurately identify them when a human helps screen the output. This makes sense considering the fact that a tool like GPT-2 is basically making calculations concerning what word is likely to be used next, allowing it to approximate other writers' common sentiments rather than expressing any kind of unique conceptual message, making it difficult to expect that its poems would have as much of a direct relatability to the human experience as a human writer's poetry.

Perhaps more alarming is the data implying that people are generally not as accurate as they believe they are at identifying AI-generated text, which could have far-reaching consequences in relation to the fake news debate mentioned earlier. In sum, while these results may not support the conclusion that AI has entirely mastered the creative process of writing poetry, it is clear that if researchers were to find a way to incorporate a more substantive source for the creativity than simple text-prediction (such as the aforementioned conceptual blending frameworks), the foundation for AI to be able to compose language in a way very similar to humans is already in place, and these two aspects together could create a more true-to-life creative AI output. Overall, however, the research being done in AI poetry generation speaks more broadly to the status of AI creativity in general; while AI systems have not yet been able to fully replicate human creative processes, the current trends suggest that as our understanding of our own cognition develops, so too will the ability for AI researchers to implement systems like conceptual blending and working memory into their designs. While it will of course take time to develop the kind of general

intelligence which science fiction writers and researchers dream of, each step in the research process shows that the technologies are becoming more and more sophisticated, lending credence to the idea that we may one day see a radical shift in creative computation similar to those breakthroughs we have seen in terms of the invention of computer chips or AI in the first place.

Conclusion and Future Directions

The idea of artificial intelligence exhibiting such a stereotypically human trait as creativity is understandably frightening for many. Perhaps somewhat assuaging their fears, however, is the understanding that AI creativity is not at the point where human artists could be supplanted by algorithms, for example; human intervention is largely still required to ensure that AI can create works of high quality. However, many researchers are currently looking for innovative new ways to push past these limitations and invent new methods which allow computational systems to combine concepts like humans and even string words together to create meaningful poetry while evaluating their own outputs. While philosophical debates will still as of now rage long and hard about the exact nature of this creativity or whether it even truly fits the definition of the concept at all, the reality is that these technologies are not as fantastical as they may have seemed in the 1950s when artificial intelligence was for the first time being systematically researched. When surveying the current trends in creative computation research, it becomes clear that important conversations must be had within the next few decades so that the possibility of creative AI does not catch us off guard. The innovations currently taking place in the field of computational creativity have direct impacts on how we conceive of ourselves and raise questions about how unique our abilities as a species really are. We may very well be facing an immediate future wherein AI is a useful tool to supplement the output of human artists and a longer-term vision where they truly come into their own in this role. In the meantime, it will suffice to say that this paper was not written by an algorithm.

Armaan Kalkat is a third-year undergraduate student at the University of Florida, where he plans to graduate with a BA in Linguistics and a BS in Psychology with an emphasis in Behavioral and Cognitive Neuroscience. He has always been fascinated by how language use is mediated by cognitive systems, as well as its importance in creating a sense of identity, and hopes to continue his research along these lines.

Works Cited

Boden, Margaret. "Creativity and Artificial Intelligence." *Artificial Intelligence*, vol. 103, nos. 1-2, 1998, pp. 347-356. doi.org/10.1016/S0004-3702(98)00055-1

---. "Creativity as a Neuroscientific Mystery." *Neuroscience of Creativity*, edited by Oshin Vartanian, Adam S. Bristol, and James C. Kaufman, MIT Press, 2013, pp. 3-18.

Confalonieri, Roberto and Oliver Kutz. "Blending Under Deconstruction." *Annals of Mathematics and Artificial Intelligence,* vol. 88, 2020, pp. 479-516. doi.org/10.1007/s10472-019-09654-6

Eppe, Manfred et al. "A Computational Framework for Conceptual Blending." *Artificial Intelligence*, vol. 256, 2018, pp. 105-129. doi.org/10.1016/j.artint.2017.11.005

Gaut, Berys. "The Philosophy of Creativity." *Philosophy Compass*, vol. 5, no. 12, 2010, pp. 1034-1046. doi.org/10.1111/j.1747-9991.2010.00351.x

Kelly, Sean Dorrance. "A Philosopher Argues That an AI Can't Be an Artist." *MIT Technology Review*, 21 Feb. 2019. www.technologyreview.com/2019/02/21/239489/a-philosopher-argues-that-an-ai-can-never-be-an-artist/

Köbis, Nils, and Luca D. Mossink. "Artificial Intelligence Versus Maya Angelou: Experimental Evidence That People Cannot Differentiate AI-generated from Human-written Poetry." *Computers in Human Behavior*, vol. 114, 2021. doi.org/10.1016/j.chb.2020.106553

Metz, Cade. "Meet GPT-3. It Has Learned to Code (and Blog and Argue)." *New York Times,* 24 November 2020, https://nyti.ms/2UV5Llx. Accessed 11 December 2020.

Oliveira, Hugo Gonçalo. "A Survey on Intelligent Poetry Generation: Languages, Features, Techniques, Reutilisation and Evaluation." *10th International Natural Language Generation Conference, Santiago de Compostela, Spain, September 2017*. Association for Computational Linguistics, 2017. doi.org/10.18653/v1/W17-3502

---. "Automatic Generation of Poetry: an Overview." Jan. 2009. *Research Gate,* www.researchgate.net/publication/228610670_Automatic_generation_of_poetry_an_overview.

Ritchie, Graeme. "Assessing Creativity." *AISB Symposium on AI and Creativity in Arts and Science, York, England, March 2001*. Institute for Communicating and Collaborative Systems, Division of Informatics, University of Edinburgh, 2001.

Turing, Alan M. "Computing Machinery and Intelligence." *Mind*, vol. 59, no. 236, 1950, pp. 433-460. doi.org/10.1093/mind/LIX.236.433