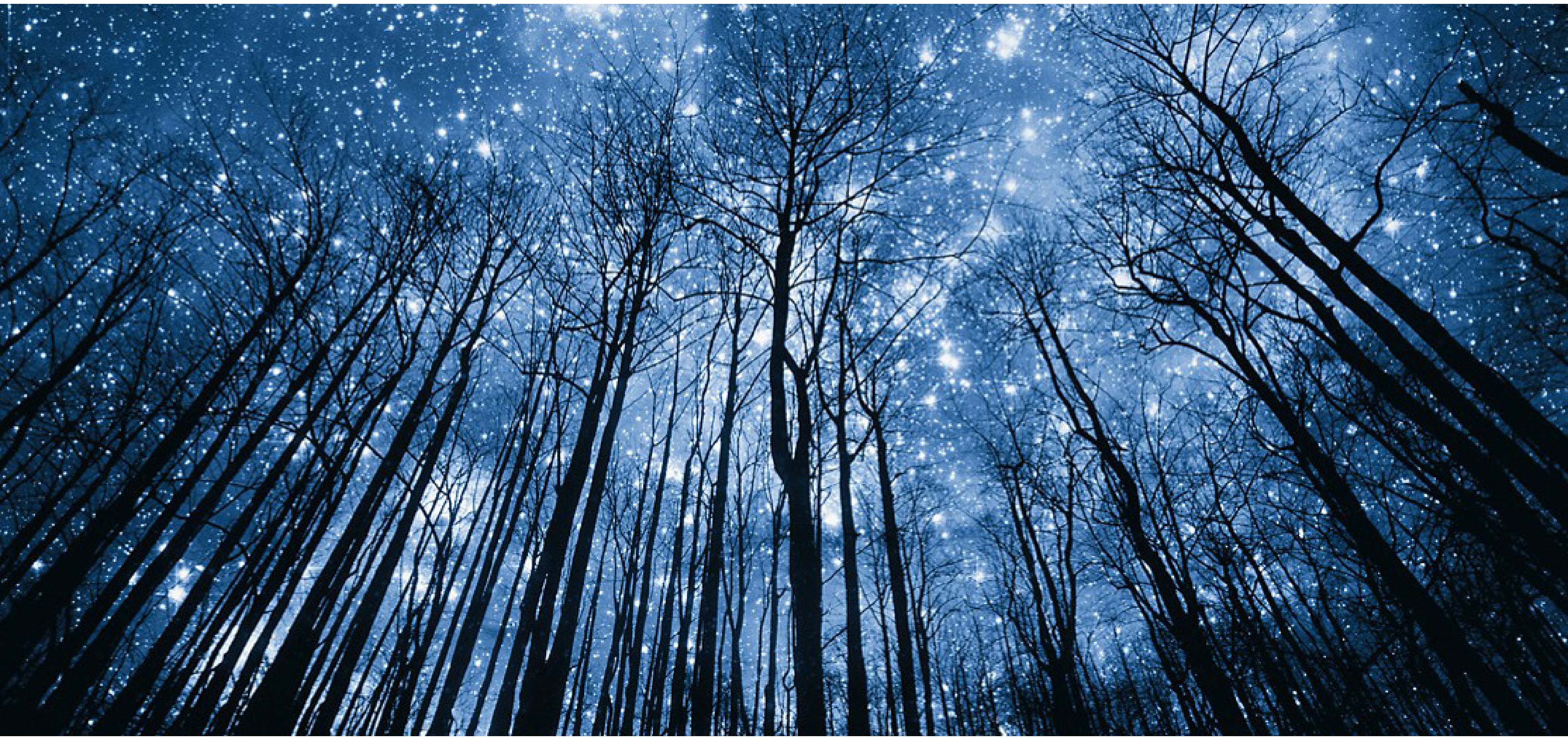
GANS - THE FUTURE OF CREATIVE FORGERY



Bharath Krishnan
Specialist 2, Inside Product
Dell Technologies

The Dell Technologies Proven Professional Certification program validates a wide range of skills and competencies across multiple technologies and products.

From Associate, entry-level courses to Expert-level, experience-based exams, all professionals in or looking to begin a career in IT benefit from industry-leading training and certification paths from one of the world's most trusted technology partners.

Proven Professional certifications include:

- Cloud
- Converged and Hyperconverged Infrastructure
- Data Protection
- Data Science
- Networking
- Security
- Servers
- Storage

Courses are offered to meet different learning styles and schedules, including self-paced On Demand, remote-based Virtual Instructor-Led and in-person classrooms.

Whether you are an experienced IT professional or just getting started, Dell Technologies Proven Professional certifications are designed to clearly signal proficiency to colleagues and employers. Learn more at www.dell.com/certification

Contents

Introduction

Conventional unsupervised machine learning algorithms are capable of easily classifying a set of objects by clustering the unlabeled datasets that are given to it and learning the pattern that appears within. What if we try to step this up another level and reverse engineer the model to not just predict but also create something out of what it has already learnt?

Generative Adversarial Network or GAN for short, is a cleverly engineered model that can produce an almost indubitable replica of what it is trained for. A GAN that is trained to convert text-to-image for example, would be capable of taking a text input of, say, “cat riding a bike” and accurately draw out the result, although there is no actual image of a cat riding a bike anywhere in its training dataset. It does so by learning to blend all the discrete things it has learnt, by introducing just the right amount of random noise to bring it all together and make it look closest to reality.

When applied across text, speech, vision, and other means of I/O, this could be the new way sophisticated AI systems would advance in fields like computer vision, natural language processing and other domains that involve human interaction. Alan Turing who is commonly regarded as the father of modern computing and artificial intelligence, introduces the concept of “Turing Test” (originally called “The Imitation Game”) in his research paper titled “Computing Machinery and Intelligence” that was published in 1950. The test tries to understand if a machine is capable of thinking like a human, and implies that if it could, it should be able to mimic a human in the way it communicates with another entity. Advancements in the field of GAN would allow us to develop an artificial intelligence that would not just be able to imitate human behavior, but also make it indistinguishable, thus making it a worthy participant for the Turing Test with a high winning probability.

This paper introduces a layman to the topic of Generative Adversarial Network and talks about the technical aspect of training the GAN sub-models, namely the **Discriminator** and the **Generator** and how they are put up against each other to compete and train themselves to get better at what they do individually in a process called Adversarial Training. A part of the paper also talks about how the rise of advanced conversational AI systems paired with strong GAN models can produce logical outputs that can drive a seamless experience for a human interacting with a computer. We also discuss the scope and application of the system across a variety of use cases by taking examples of existing readily available GAN models in the market. The paper finally concludes by discussing the ethical impacts that all the technological developments in this field can have on the future of the creative/design industry.

Adversarial Training

The 'A' in GAN, stands for "adversarial" – which is the very fundamental technique that is used to produce (or improve) the two sub models that we read about in the previous section. Adversarial training is simply the process of identifying your weaknesses and turning it into your strengths over time. Let us take an example of a neural network (type of machine learning algorithm that mimics the human brain, by transmitting information between nodes like how biological neurons fire signal impulses between one another) that is trained to understand handwritten numbers. Based on the labelled dataset that is used to train, there would be a set of numbers, say 3 and 8, that the model cannot differentiate between. A regular unsupervised learning algorithm would just iterate through the entire dataset in an orderly manner until it reaches its limit. This way, at the end of the training, we would get a model that can still understand handwritten numbers, but it might perform relatively poor when it must classify between the numbers 3 and 8 specifically.

An adversarial training model tries to overcome this challenge by a simple technique. As it learns that the model is weak in a certain scenario, it starts emphasizing on that specific subset. In the case of our handwritten numbers model, the training would weigh in on having more iterations of 3 and 8, until the model eventually gets better at finally understanding the difference between the two numbers. This way, it learns to overcome its own weaknesses – just like the protagonist in every superhero movie ever.

Now if you think that was not dramatic enough, let's introduce an antagonist into the story. To explore the possibilities of not one, but two models that are put into an adversarial training environment, let us take the example of an interesting experiment that was developed by the brilliant team of engineers at Open AI, called the "Multi-Agent Hide and Seek" game. In this game, they observed how two groups of agents, color coded red and blue, would work together to lead their team into winning a simple game of hide and seek. This involved creating 1000s of random simulations involving walls, objects, and other elements in the environment, that the self-supervised agents would use to strategize and win over the other team. The blue team would always be the hiders, and the red team the seekers, and their respective objectives were very clear - hide and seek.

After every game, the winning team gets a +1, and the losing team -1. Over time they observed the agents learning at a faster rate than they predicted, exploring as many as six different strategies that put pressure on the other set of agents to improvise.

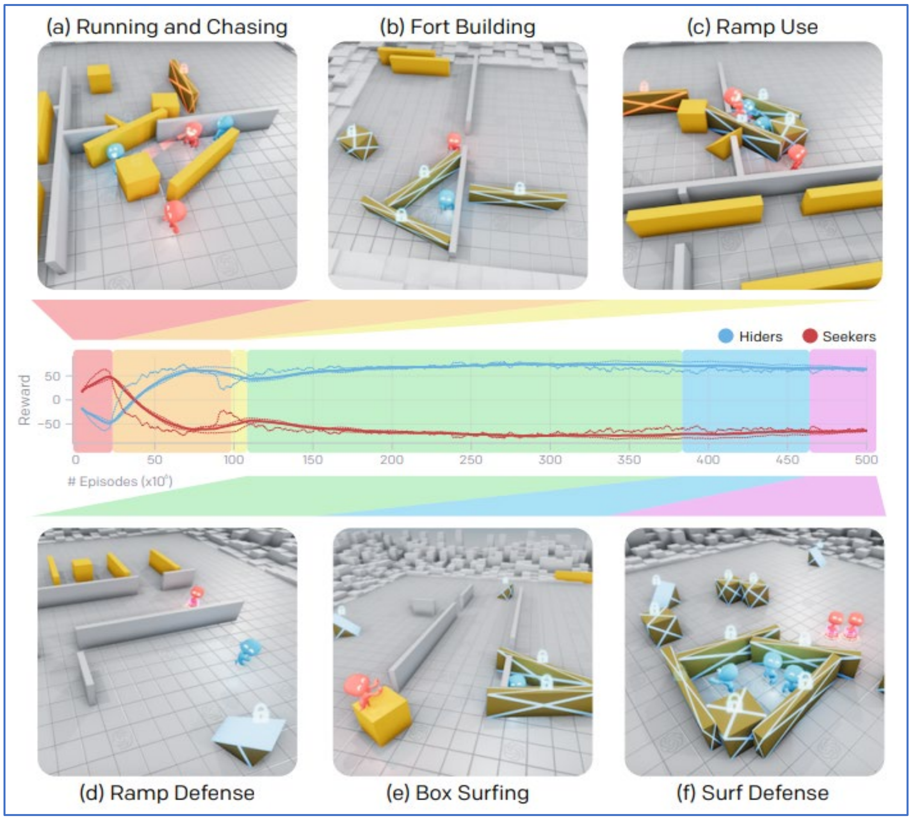


Figure 1: Strategies discovered within the multi-agent hide and seek game. [source]

This system of two agent groups fighting against each other to improve oneself, not only required zero human intervention, but also explored scenarios beyond the scope of an average human's potential.

What is a GAN?

Now coming back to the case of Generative Adversarial Network; similar to the example we reviewed earlier, involves putting together two convolutional neural networks (CNNs) up against each other to compete. The pair would go through an Adversarial Training process, which means at the end of every iteration, one side loses and would have to work on improving the weakness that caused it to fail. The two sub-models would be called Discriminator and Generator, and their respective responsibilities quite analogous to the hiders and seekers in our previous multi-agent game, would be pretty straightforward. The Discriminator must predict if a sample is real or fake, and the Generator must introduce a level of random noise to a sample or otherwise generate a fake image to fool the Discriminator. In this adversarial game, the Discriminator learns to better differentiate between real and fake, and the Generator would learn to produce better fakes. Eventually after several iterations of wins and losses, the Generator would get so good at producing a fake image, that the Discriminator would not be able to tell if it is real or not. At this moment, you would have yourself a Generator model that can produce life-like replicas of anything you want, that would almost be non-differentiable from real, even to a human being.

Application in Image Fabrication

Unleashed upon the plethora of data on the internet, this invention could scale limitlessly. Starting with a project called StyleGAN that was introduced by a team of researchers at Nvidia in 2018 and made publicly available in 2019. This was a GAN that was trained to analyze portraits of human faces with the goal of producing an unlimited number of fake human faces that are almost impossible to be differentiated from a real one. This software was later used to create a website that became popular for displaying a new (fake) face every time the page was reloaded, called thispersondoesnotexist.com. The engineer Phillip Wang, who was behind the website himself was surprised at how accurately StyleGAN was able to dissect the human face and coherently reassemble all of its key aspects.

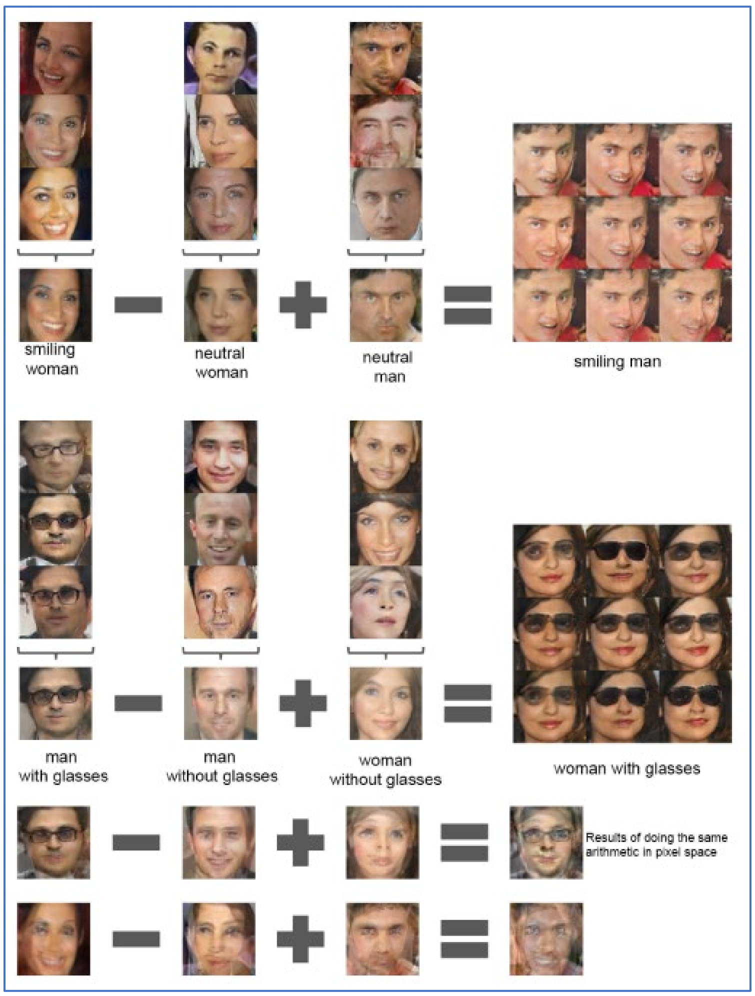


Figure 2: Vector arithmetic calculations performed on face portraits. [source]

Another really interesting project that was again developed by a team of engineers at Nvidia, is called the GuaGAN, that was trained on millions of photos of landscapes to convert a rough hand-drawn painting into a photorealistic image. This was later released to public under the product name “Nvidia Canvas”, that enabled even a toddler to turn his/her brushstrokes into an ultra-realistic image of a landscape.

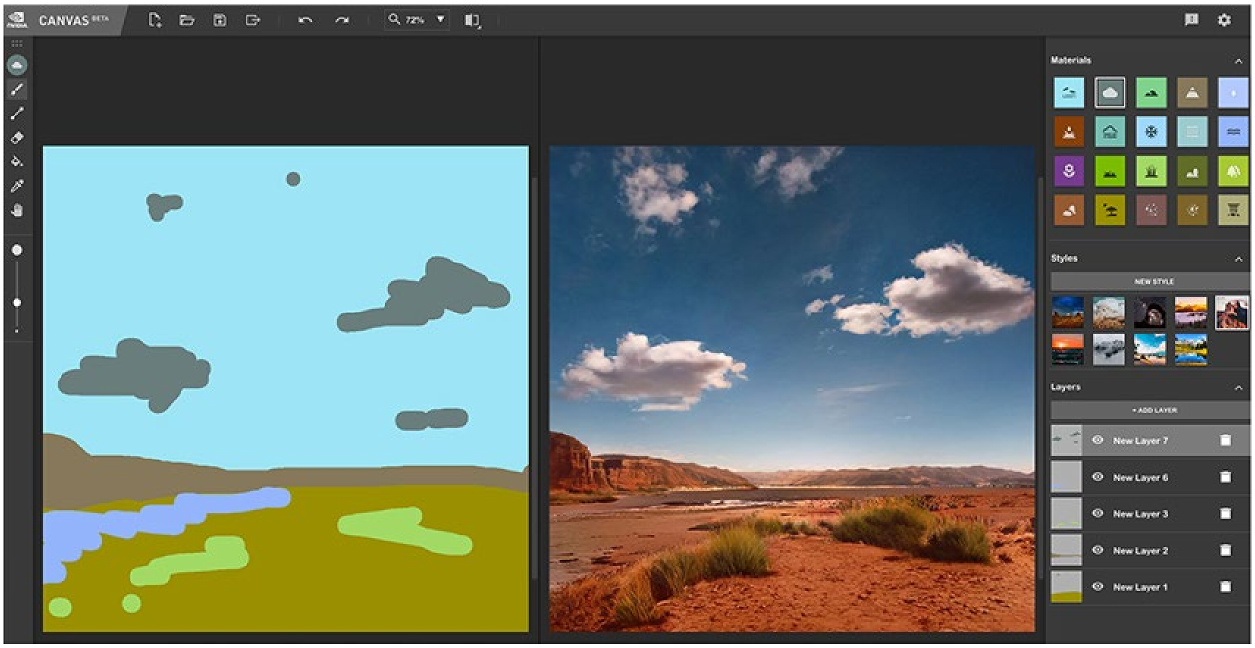


Figure 3: Nvidia Canvas in action. [source]

Text to Image Generators

Beyond the territory of GANs, another concept that has been making a lot of noise (quite literally) in the field of AI based image generators, is text to image conversion. We have had systems that are intelligent enough to identify objects within an image for quite some time now but doing the exact opposite of that – turning text captions into image responses is one of the most popular studies among research communities today.

Most text to image generators today work in a similar fashion, comprising of up to 3 key components that are elemental to its functioning:

1. A natural language processing model that can understand the user’s text prompt and extract the keywords from it.
2. A multi-dimensional latent space that comprises of a corpus of trained labelled image dataset of almost every object that you can imagine.
3. A diffusion algorithm that can predict the noise in an image to remove it from the source. This technique is used to blend multiple images together.

One of the most popular text to image generator systems that has got the entire internet talking about it, is the DALL·E 2, developed by Open AI. DALL·E 2 is capable of producing images that resemble human aesthetic judgement, with highly detailed visual parameters like depth of field, shadows, and reflections.

Figure 4: DALL·E 2's output for the text prompt:
"An astronaut riding a horse in a photorealistic style" [source]

DALL·E 2 uses the GPT-3 (3rd generation Generative Pre-trained Transformer) language model, which is another extremely popular project developed by the Open AI team. It takes the text prompt input and dissects it by parts to understand what is required. GPT-3 is a very powerful language model that has been trained on a large variety of data from several different websites, including Wikipedia. It can not only understand basic text input, but also have fully fledged context aware conversations with long passages of original text with a human.

On that note, this next topic will digress a little, to explore the exciting challenges involved in a machine-human conversation exchange.

The Imitation Game

What does it take for a machine to be called intelligent? While scientists all around the world had hundreds of metrics to evaluate this, British computer scientist Alan Turing just cared about one simple question – *can the machine talk like a human*? In his 1950 paper titled "Computing Machinery and Intelligence", he proposes an examination to evaluate this, which is popularly referred to as the "Turing Test" or "The Imitation Game".

The test involves a human judge having a text-based conversation with a panel of players without seeing their faces. One or more of these players could be a computer, and if the judge is incapable of differentiating the computer from a regular human, then it is said to have passed the Turing Test. In other words, a computer is claimed to be intelligent if it could have general conversations like a human, to the point where it becomes indistinguishable from them.

Most machines built today are designed keeping a finite set of specific requirements in mind. Hence, a computer would be really good at doing that one thing that it was built to do, and it could do this much better than a human could. For example, a computer that is designed to play a board game like Chess or Go, is capable of computing all possible permutation and combination of moves and predict the next step to favor a winning strategy, but you cannot really expect anything more than that from it. Language is one of the most challenging subjects for a machine to understand because of a number of reasons. The ambiguity involved in similes, metaphors and other similar forms of speech can easily throw off most language recognition attempts.

Turing predicted that by the year 2000, machines with 100 megabytes of memory would be able to easily pass his test. However, even though modern computers today have much more memory than that, very few have succeeded – by hook or by crook. Turns out, increased memory and processing power might just not be enough to simulate human conversation.

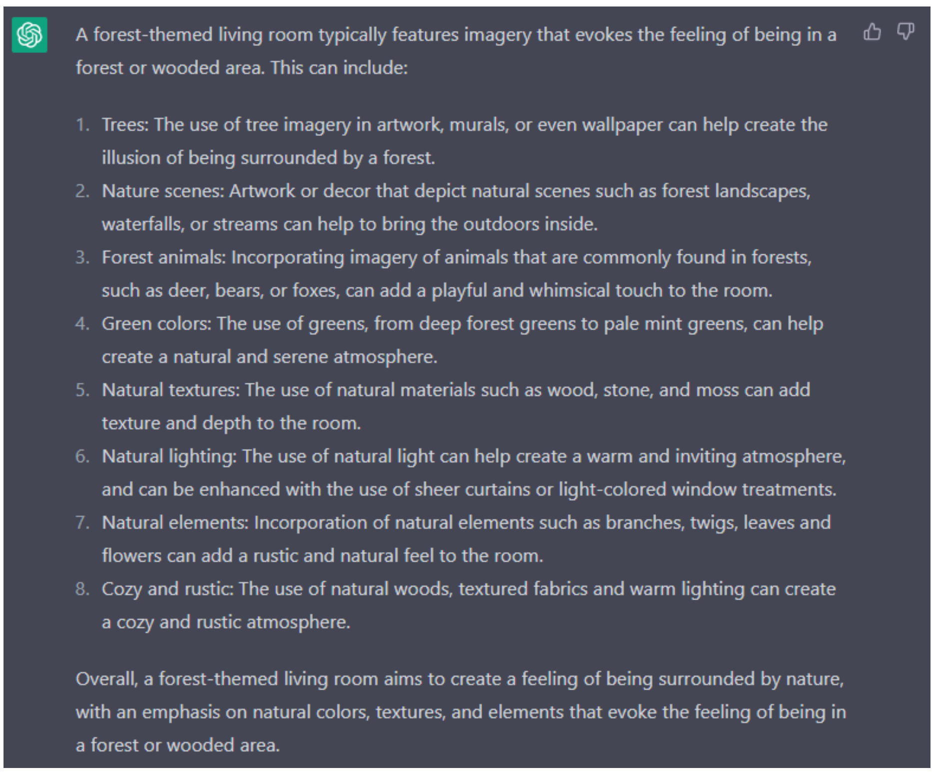
Rise of Conversational AI

Gone are the days of template based chatbots that respond with a static set of statements on a one-off question-answer basis. Today, we have several conversational AI systems that are trying to tackle human conversation. Most personal digital assistants like Alexa, Siri and Google Assistant have introduced the feature of context awareness, which would enable it to remember what was asked previously and take follow-up questions.

Open AI is on a roll, breaking the internet with every new project they unveil. It was the same when they released ChatGPT in November 2022, which is the successor to their already popular GPT-3 language model. GPT-3 tries to form sentences by predicting the next word in series, picking from its extensive dataset built from different parts of the internet. ChatGPT improves on this by introducing human feedback in the course of training, through a process called supervised reinforcement learning. This enables the model to be almost 100x smaller but still perform much better than its predecessor.

ChatGPT truly defines the peak of conversational AI today. From history essays to computer code, to math, science, and business problems – ChatGPT covers everything, while still being able to maintain a context aware conversation with the user. However, one of the biggest challenges in maintaining ChatGPT at a search engine level today, is the amount of compute resource each search query would demand. Scaling to billions of daily user requests would be too expensive, introducing massive latencies from server throttling. There have also been problems where people use conversational agents like these to plagiarize content and develop fake news. Since none of the responses have a legitimate source attached to it, there is a good chance that it may occasionally generate incorrect information, harmful instructions, or biased content. Recent models like ChatGPT are also limited to the knowledge of the world and events after 2021. But in spite of all the challenges, there have been plenty of users that have benefitted from conversation AI systems like ChatGPT, both at a personal and commercial level.

Furthermore, as we are nearing the end of this paper, let's try out an experiment to attain the best of both worlds and see what AI is truly capable of today. One or more AI systems can be combined sequentially to do what the other is incapable of. For instance, in the following example, ChatGPT, which is a conversational AI system that can only give text responses, is asked to give ideas to decorate my living room with a forest theme. Let's take a look at how this progresses.



A forest-themed living room typically features imagery that evokes the feeling of being in a forest or wooded area. This can include:

1. Trees: The use of tree imagery in artwork, murals, or even wallpaper can help create the illusion of being surrounded by a forest.
2. Nature scenes: Artwork or decor that depict natural scenes such as forest landscapes, waterfalls, or streams can help to bring the outdoors inside.
3. Forest animals: Incorporating imagery of animals that are commonly found in forests, such as deer, bears, or foxes, can add a playful and whimsical touch to the room.
4. Green colors: The use of greens, from deep forest greens to pale mint greens, can help create a natural and serene atmosphere.
5. Natural textures: The use of natural materials such as wood, stone, and moss can add texture and depth to the room.
6. Natural lighting: The use of natural light can help create a warm and inviting atmosphere, and can be enhanced with the use of sheer curtains or light-colored window treatments.
7. Natural elements: Incorporation of natural elements such as branches, twigs, leaves and flowers can add a rustic and natural feel to the room.
8. Cozy and rustic: The use of natural woods, textured fabrics and warm lighting can create a cozy and rustic atmosphere.

Overall, a forest-themed living room aims to create a feeling of being surrounded by nature, with an emphasis on natural colors, textures, and elements that evoke the feeling of being in a forest or wooded area.

Figure 5: ChatGPT response [actual query execution result] [source]

If I take the same response from above and pass it through DALL·E 2 (text to image generator), it will give me a visual representation of what this would look like, so that I have a better idea about the results. For an interior designing solution like so, it would serve better to have a visual output, rather than a textual output.

It is important to note that the output of DALL·E 2 is not an existing picture on the internet, but rather a picture that the AI *created* based on our text prompt.

Figure 6: DALL·E 2 response [actual query execution result] [source]

I shall refrain from commenting further and simply allow the results to speak for themselves.

Conclusion: Future of Creative Forgery

For better or worse, it is indeed a wonderful time to be alive, surrounded by the greatest advancements in artificial intelligence. Over time, we have seen several jobs being replaced by automation and robotics, but the last thing that we would have possibly imagined being taken over would have been art. Art requires a very special blend of talent, inventiveness, and the innate human quality of aesthetic sensibility. For the longest time, the creative industry has bragged about being the chieftains of novelty, bringing the touch of authenticity that nothing else can deliver, but yet again, AI has proved that it can conquer this too. AI generated art is growing more and more sophisticated to the point where it has started receiving recognition for its work sold in the art world.

But at the end of the day, a human touch will always retain its value. No level of compute would be able to bring the variance that a human can bring. With innovation comes disruption; until we accept the fact that these AI systems are nothing but tools that can benefit our work further, we would forever be running away from progress. With the help of these practical applications, we would always have a reference point to begin from – the inception of human creativity. Do not reinvent the wheel, just redesign it!

“We can only see a short distance ahead, but we can see plenty there that needs to be done”

~ ***Alan Turing***
Computing Machinery and Intelligence

References

1. Generative Adversarial Networks (GANs) – Computerphile
2. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks – Alec Radford, Luke Metz, Soumith Chintala
3. How AI Image Generators Work (Stable Diffusion / Dall-E) – Computerphile
4. How This A.I. Draws Anything You Describe [DALL-E 2] – ColdFusion
5. The text-to-image revolution, explained – Vox
6. Generating Images from Captions with Attention – Elman Mansimov, Emilio Parisotto, Jimmy Lei Ba & Ruslan Salakhutdinov
7. Emergent Tool Use from Multi-Agent Interaction – OpenAI
8. Emergent Tool Use from Multi-Agent Autocurricula – Bowen Baker, Ingmar Kanitscheider, Todor Markov, Yi Wu, Glenn Powell, Bob McGrew, Igor Mordatch
9. Nvidia Canvas – Nvidia
10. DALL·E 2 – OpenAI
11. It's Time to Pay Attention to A.I. (ChatGPT and Beyond) – ColdFusion
12. OpenAI ChatGPT: The Future Is Here! – Two Minute Papers
13. Computing Machinery and Intelligence – A. M. Turing