

COMP4920 Ethics in Computer Science

Group Report



Team JPW

Team Members:

z5251261 Jack (Yueyang) Gu
z5251261@unsw.edu.au

z5267344 Peter (Hyundong) Seo
z5267344@unsw.edu.au

z5286124 William (Jinghan) Wang
z5286124@unsw.edu.au

Report Submission Date: 24 / November / 2022

Table of Content:

| | |
|---|-----------|
| Drawing AI Introduction | 1 |
| Users and stakeholders | 1 |
| Artists (painters, illustrators, etc.) | 1 |
| Art exhibition host | 2 |
| The public | 2 |
| Drawing AI programmers | 2 |
| Project sponsors | 3 |
| Philosopher/Critics | 3 |
| Judgment call game | 3 |
| Stakeholder: the Public; Rating: 1; Principle: Reliability | 4 |
| Stakeholder: Project sponsor; Rating: 3; Principle: Transparency | 5 |
| Stakeholder: Philosopher/Critics; Rating: 3; Principle: Inclusion | 7 |
| Similar technologies | 8 |
| Code-generating AI | 8 |
| Article-generating AI | 10 |
| Findings & conclusion | 12 |
| Recommendation | 12 |
| Bibliography | 13 |

Drawing AI Introduction

With the rise of AIs, the way how the world operates is shifting to a new paradigm and we are being confronted with several questions. Are AIs beneficial to our society or would they bring harm to the human race? Furthermore, are they questionable in terms of ethics? The main goal of this report is to address several questions and uncertainties in regards to AIs and to strongly express our suggestions based on our research. Our team will focus on drawing/painting AIs specifically.

Drawing AI can be defined as an artificial intelligence that has the capability to draw images based on what it learns from its given database (Melissa T. Miller). That is, it generates images by going through a machine learning procedure. Drawing AIs can be used for multiple purposes. They can be either used for analysing existing artworks or for creating a new artwork. They also assist people like myself who have no talent in drawing to come up with decent team logos and typographies for team projects. The scope of drawing AI is truly endless, ranging from filtering a style of an artwork to another to generating an artwork, just by giving it a few lines of sentences. There are several benefits associated with drawing AIs. They have the potential of bringing the most out of children's creativity. There are already drawing AIs that animates the drawings of children. For a person who works in the art industry, these AIs can assist you in handling repetitive tasks. It also reduces the budget for buying paints and other art supplies since it only requires electricity. Most importantly, they can come up with the most creative artwork by learning from the database that is fed to it.

Users and stakeholders

There are several users/stakeholders surrounding drawing AIs. Our group has defined users as people who are using drawing AIs. On the other hand, stakeholders are people who might be affected because of drawing AI execution or completion. Our group will focus on the following users/stakeholders:

- **Artists (painters, illustrators, etc.)**

The artist is the group that is most directly affected by drawing AIs. Artists mainly include painters, Computer Graphics industry practitioners and others who need to use sketching boards and software like photoshop or Illustrator to draw. Due to the maturity of drawing AIs technology, many works that require artists to draw for a long time can be completed by machine. The latter takes a short time to provide many solutions that meet the requirements and cost less than the

remuneration required by artists. In this respect, painters who create artworks for enterprise projects, and personal needs, will be the main stakeholders affected.

On the other hand, artists who paint for interests or personal hobbies can also be a user. The maturity of drawing AIs technology will promote the creation of painters. Painters only need to input their ideas, obtain results through drawing AIs, and then perform secondary creation to create products in line with expectations. It also provides more inspiration for painters.

- **Art exhibition host**

Art exhibition hosts mainly include those who create, plan or assist in art and cultural exhibitions or activities. They will collect or temporarily borrow works of art created by contemporary or former artists and display them to the public. With the emergence of drawing AIs, the channels for the public to pursue art are no longer limited to exhibitions or other ways to view works of art with their own eyes. Drawing AIs provides channels for ordinary people to create their works of art and improves their ability to appreciate art. As a result, the art of half or below-average quality may no longer be acceptable to the public. Therefore, Art exhibition hosts are stakeholders, as the maturity and popularization of drawing AIs will indirectly impact an art exhibition. They will have to decide whether or not they will be accepting AI-generated artworks.

- **The public**

The public is the group that makes the majority of users/stakeholders. The public includes people like us who have heard of drawing AIs but have no profession in the art industry. They have at least one experience that involves painting and designing and know how difficult it is to come up with a mind-blowing artwork. A portion of the public that actively uses drawing AIs to assist them in generating logos or layouts is considered users, as they are directly using the drawing AI. The other portion of the public who doesn't use drawing AIs but have come across artworks by drawing AIs is considered stakeholders as they are likely to be influenced by the artwork whether it being positive or negative.

- **Drawing AI programmers**

The programmers/developers are those who are proposing and working on the drawing AI technology. These programmers may not necessarily be an expert in the field of art, but they take up the core of AIs. They study different types of deep learning algorithms, select datasets that will be fed to the AI, and train AI to generate artworks. Without programmers, we wouldn't have access to different AIs. As programmers are directly involved in the creation of AI, they are

considered to be primary users. They are responsible for explaining the scope and limitations of AIs to project managers and stakeholders.

- **Project sponsors**

The project sponsors are the stakeholders of the AI technology. Without project sponsors, AI drawing development would not be able to obtain financing. If the market accepts the technology, AI drawing might have a big impact on the project sponsor. The project sponsors should be capable and have some expertise in order to perform their duties. Regardless of the organisational unit, the sponsor must be able to persuade individuals towards a good conclusion, ensure that effective communications are in place, and be prepared to assume personal responsibility for the project. To make the proper investments and grant their investors the power they require to carry out their obligations, project sponsors must possess strategic vision and leadership abilities.

- **Philosopher/Critics**

The Philosopher/critic is the stakeholder of the AI drawing technology and as a philosopher/critic he needs to know something about the field but does not need to be a specialist. When this new field emerges, the philosopher/critic needs to think and criticize from two completely different perspectives: "AI drawing will bring creativity to art and the public " and "AI drawing is not art". " to offer a critique. There is no question that the articles and other forms of media posted by philosophers/critics are going to influence the AI industry.

Judgment call game

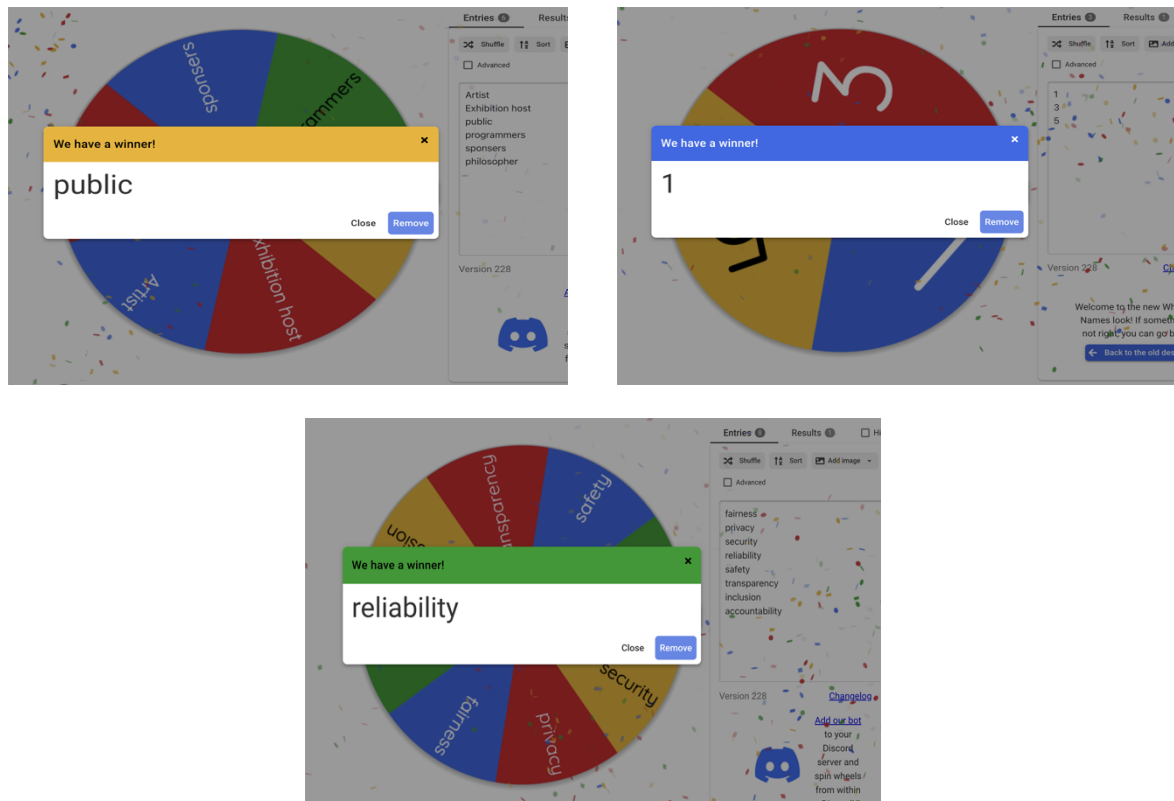
Our group has come up with the following scenario.

Scenario for the judgment call game: A person submitted an artwork generated by a drawing AI to an art contest.

Just like the recent event of Jason Allen winning the first place at an art competition by submitting an artwork that was generated by an AI, the sound of submitting AI generated artworks doesn't sound that futuristic and absurd. In fact, it has become reality. As every media is publishing articles about this news, the majority of the public has become aware of this event. By assuming such scenarios where people keep on submitting AI generated artworks to an art competition will continue and become a phenomenon, our group has conducted a value sensitive design that addresses this issue by performing a judgment call game. Each of our group members was allocated a unique stakeholder, principle and rating.

Instead of printing out the hard copies of the judgment call game cards, our group ended up using an online wheel picker. We made 3 wheels, each for stakeholder, rating, and principle. The below is what happened to our judgment call game.

- Stakeholder: The Public; Rating: 1; Principle: Reliability (Peter)



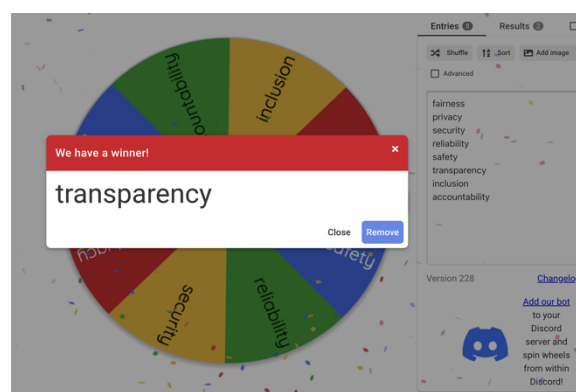
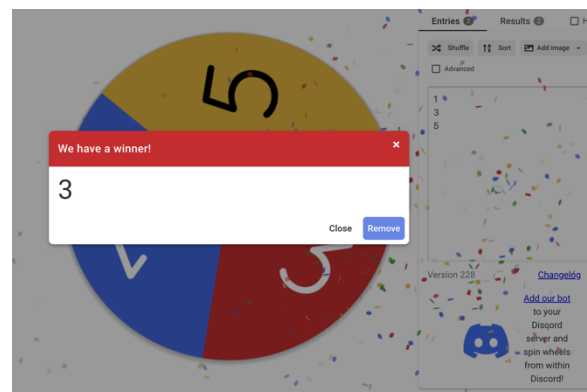
My concern (as the public) is that the AIs might not have a clear and obvious domain in which it should operate. To be more straightforward, I do think that these drawing AIs are not reliable. I do understand that it assists people who are bad at drawing to come up with a wonderful artwork but can the developers for these AIs guarantee that it would behave appropriately when given unexpected and malicious user inputs? For example, in a circumstance where the user wants to generate a violent or a sexual image, is the AI reliable enough to inform the user that this is not what the AI is intended for and perhaps advise the user to visit a counselling centre. We never know what teenagers going under puberty might be generating with these kinds of drawing AI becoming so accessible. I doubt that this is not the case, and the AIs are not mature enough at this point of stage. What fuels my concern is that the developers of these AIs do not share how the testing of AIs is conducted. I know that I'm not an expert in computer science and AI, but I do know that rigorous testing is an essential part when it comes to deploying whatever product. I assume this would be the same for drawing AIs.

Drawing AIs have reliability issues just by itself but when people try to submit AI generated artworks to an art competition? The issue gets worse. Art competitions were intended for humans

to compete with each other in the first place. Having AIs enter the competition was clearly unintended, making the whole evaluation and auditing process unreliable. I am pretty certain that the art competition host receives plenty of complaints and requests of explaining their procedure and logic behind the ranking of artworks. Allowing AIs to enter such competitions will not make things better for the host but only worse.

Drawing AIs will change the lives of the public without a doubt. It will help people with zero knowledge in art edit photos, draw realistic paintings and illustration. The benefits are clear as water. It is going to increase efficiency and productivity for many people who do not work in the art industry. Unfortunately, there will be some negative influence as well. One thing that concerns us is plagiarism and copyright. Ever since the Internet has become part of our daily lives, we have access to many resources. We use them for our projects, YouTube videos and so on. However, not many of us give credit to the original creator. The people who deserve credit do not get one. This will become a more serious issue as drawing AIs use the artwork of countless artists in the procedure of learning. Of course, they wouldn't give credit to these artists and the public who uses them will have no idea whose artworks have been used for generating the requested artwork and fail to acknowledge the hard work of human artists. Drawing AIs should address this issue as soon as possible as some artists refuse their work to be used as a learning tool.

- Stakeholder: Project sponsor; Rating: 3; Principle: Transparency (Jack)



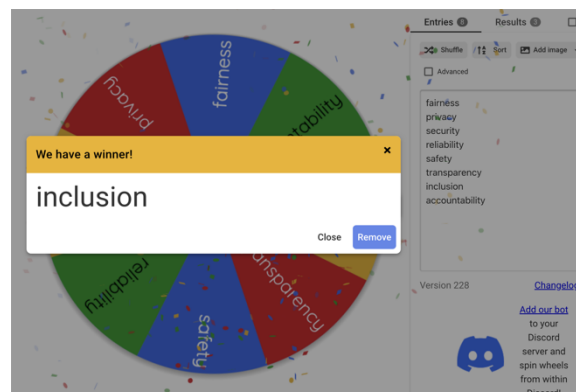
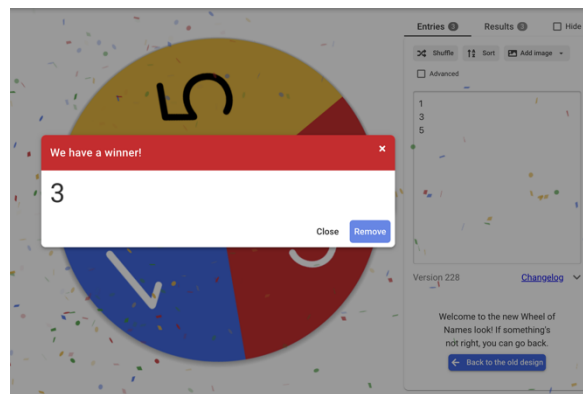
Many machine learning systems currently in use suffer from a lack of transparency and comprehension into the inner workings of the models. It's partially due to the model's complexity, which can be visualized when a huge number of data points and hundreds of heterogeneous attributes are input into the model for analysis. Even the model creator has difficulty comprehending the model's inner workings.

For project sponsors, it is worth considering whether to disclose source code and training model data. The reason for project sponsors not to adopt more transparency is that the application may lose the trust of customers if the models used are perceived to be biased and unfair. For example, Twitter's photo cropping feature has been accused of discriminating against blacks and women because it favors white and male faces when cropping. When the algorithm is questioned, companies are forced to tweak it or even stop using it. Relatively, increased transparency can gain the trust of users. At a time when people are becoming aware of the potential hazards of AI, users will have higher trust in technologies that have high transparency.

In addition, bias may seep into machine learning models at all project stages, so bias is difficult to avoid and detect. Factors that may lead to bias in the output of machine learning models include under- or over-representation of specific groups or categories in the dataset, incomplete data, lack of heterogeneity in data sources, or historically incorrect data. Bias in machine learning models can be the result of unconsciousness. For example, when keywords such as "military", "entrepreneur", and "politician" are entered, the AI drawing will mostly output male characters. This bias is not consciously caused by the project sponsor, but rather by the unbalanced distribution of real data in the dataset. Therefore, project sponsors are forced to think about lowering the transparency in order to avoid being accused of having such "bias".

However, providing data from protected classes to ML practitioners can also aid in reducing model bias. ML practitioners can take proactive measures to mitigate or remove these biases by utilizing specific strategies by first determining where elements of the model are biased. For instance, at the preprocessing stage, strategies like counterfactual fairness and path-specific counterfactual approaches can be employed to make up for data imperfections and guarantee that sensitive qualities do not affect the outcome.

- Stakeholder: Philosopher/Critics; Rating: 3; Principle: Inclusion (William)



From an ethical point of view, if the tournament explicitly requires that the use of AI be prohibited, this participant is unethical because he breaks the rules. More in line with the scenario under discussion is that there is no explicit requirement. Hence, the key to determining whether a tournament is ethical is whether it is fair to all participants.

Drawing comparisons are mainly about comparing the ideas of the painters as well as their abilities. If the competition prefers unique ideas, drawing AI is fair because it is only a tool for participants with no drawing experience to show them fundamental ideas. Especially for some groups who are unable to paint due to disabilities or lesions, it provides the opportunity to output the closest picture to their idea through many attempts by AI. Drawing AI gives everyone the right and chance to participate in the competition. In addition, AI creation is a machine combined with the critical features of the input things and the learning of hundreds of millions of previous or contemporary painters' paintings to create paintings in line with the experience of the general aesthetic. Essentially AI is guessing the user's ideas based on the input keywords, which is more similar to a person dictating ideas for the painter and requesting the artist to draw. Compared with an artist drawing following his thought, the latter can be more non-destructive in showing their ideas. From this perspective, the use of AI is fair.

Suppose the competition prefers individual drawing ability. AI has lowered the bar for art competitions and allows people who have not put any effort into the drawing period to create

artwork in a short time comparable to or even better than those created by contemporary painters with their efforts. It is a negation of their efforts and is highly unfair.

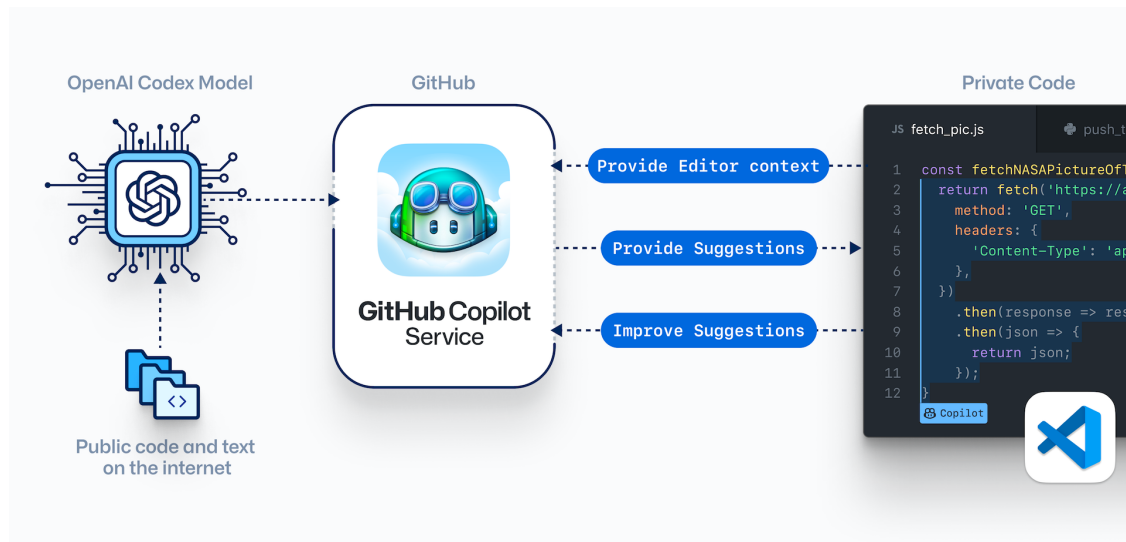
Therefore, the key to judging the contest is fairness needs to be judged by the nature of the contest and the participants. Suppose this competition is a contest of ability or an examination of students' ability in an art school. In that case, using AI is unacceptable and even a kind of cheating, cheating behavior. If it is just a competition of ideas, then AI is an excellent tool to allow more people to experience and enjoy the competition. Moreover, the better solution is for the competition organizer to specify the uncertainties, whether the competition allows using AI and the scope and purpose of the use. To ensure that all competitors The decision will be made by all participants.

In addition, how to define the attribution of an artwork created by drawing AI and the user when a user uses AI and makes some modifications. Just as submitting a paper to use others' materials as reference and borrowing are permissible. Suppose a user draws on the results provided by AI as a basis or inspiration and makes some extensive modifications. In that case, it is unfair to define the attribution of artwork as AI directly. The essence of drawing AI is to serve people, so it should allow modification and borrowing within a specific range. Therefore, it is necessary to set a modification magnitude to define the attribution problem of creation by AI or user.

Similar technologies

- **Code-generating AI**

One similar technology is code-generating AI. For the present stage of this technology, no widely available AI system can generate programs with complex logic. It is more inclined to machines that assist pair programming, which generally works by auto-completing portions of code written by a user or taking instructions describing the code generated. It aims to promise to cut development costs while allowing coders to focus on creative, less repetitive tasks. The most advanced example of these systems is Copilot from GitHub. It is not only chosen by some international enterprises on complex projects but also as an extension plugin that all individual users can use on VSCode, JetBrains and other coder editors or IDE. Copilot can finish small snippets of code or write functions that perform simple tasks based on a prompt described by the user. It assists human programmers by reducing the time it takes to develop routine functionality.

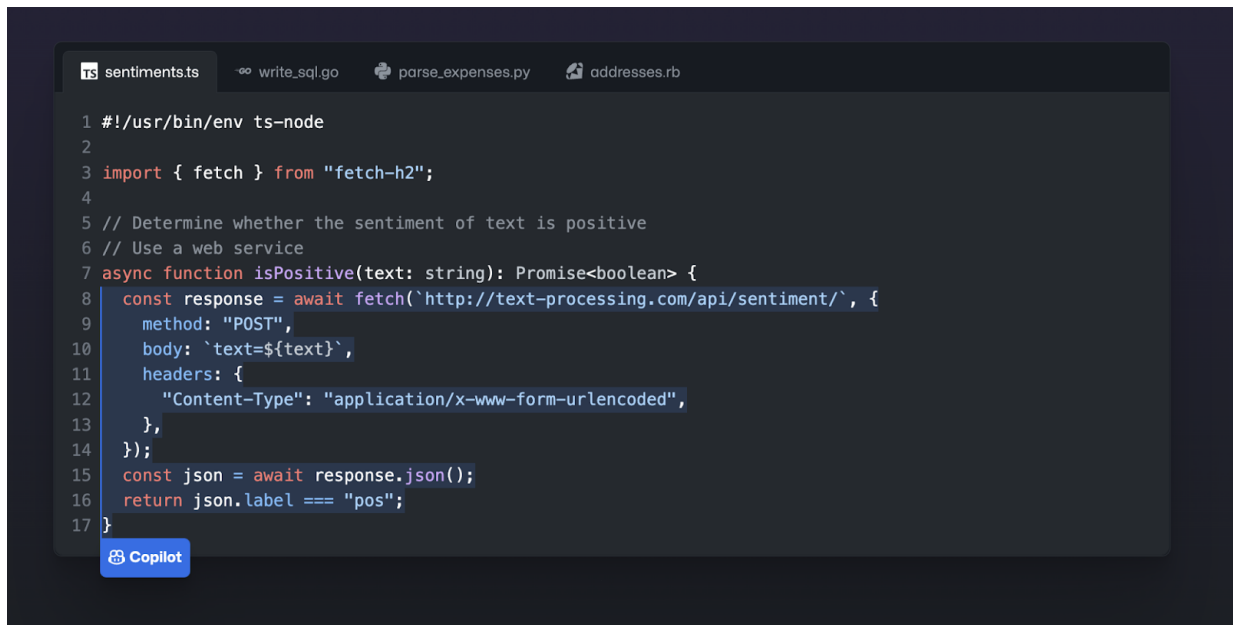


Code-generating AI is similar to drawing AI. The group of stakeholders mainly include the user and developer. The user group can also be divided into individual and project managers. For individual users, it reduces the need for some repetitive work and avoids some users' errors, such as spelling and incorrect characters. It is also beneficial for beginner programming. It can provide example code that follows the users' requirements. However, everything has two sides, excessive convenience will also lead to user dependency, especially for the beginner programmer, and overuse will lead to their programming ability decline and a weak grasp of some programming basics.

For project manager users, this tool can significantly reduce the necessary repetitive work and reduce bugs or problems caused by individual carelessness to shorten the project completion time and total cost. When a project has a problem with the code generated by AI, the attribution of the problem is the user, as this technology is only an auxiliary role. Users have the responsibility to ensure that the code generated by AI meets the requirements of the project and the current environment.

For the developer, similar to drawing AI bias due to the unbalanced distribution of actual data in the dataset, code-generating AI also has security issues due to the training samples' uncertain safety. Since much of the training AI's code comes from code hosting platforms like GitHub, there may be a loophole in unintentional vulnerabilities from outdated code or intentional ones injected by malicious actors into the training set. It may cause an unintentional attack on the computer or service while running code.

In addition, code-generating AI's goal is broader than the current assistive nature but rather to be more autonomous in completing code and independent projects without human guidance. The improvement of code-generating AI may inadvertently and rapidly create new attack surfaces and vulnerabilities beyond our ability to discover and harden them.

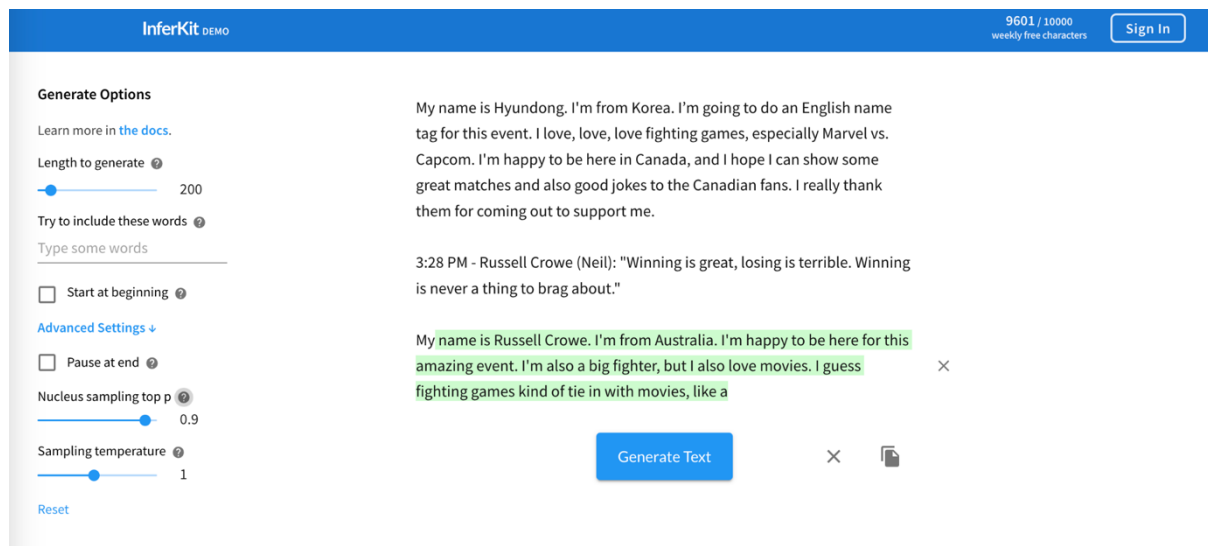


```
1 #!/usr/bin/env ts-node
2
3 import { fetch } from "fetch-h2";
4
5 // Determine whether the sentiment of text is positive
6 // Use a web service
7 async function isPositive(text: string): Promise<boolean> {
8   const response = await fetch(`http://text-processing.com/api/sentiment/`, {
9     method: "POST",
10    body: `text=${text}`,
11    headers: {
12      "Content-Type": "application/x-www-form-urlencoded",
13    },
14  });
15  const json = await response.json();
16  return json.label === "pos";
17 }
```

Moreover, whether humans or machines write code is inherently dual use. Because this technology is available to all, it will help programmers do their jobs better while increasing the speed with which cyber attackers can find vulnerabilities and develop attack tools. Code-generating AI creates many new opportunities and conveniences but presents some technical challenges. Therefore, developers need to pre-screen the training code before training and update AI based on code or library updates to avoid the risk of introducing known vulnerabilities or learning AI to rewrite those integrations to reduce the time of updating and narrowing the window of vulnerability. Additional rules also need to be set to stop allowing AI to be used as a tool for attackers.

- **Article-generating AI**

Another similar technology similar to AI drawing which our group has come across during our research is the text writer AI. In this specific example, our group would like to talk about a text writer AI called InferKit. InferKit is a text generation AI developed by Adam Daniel King. Anyone can use this AI for creative and fun uses of writing any sort of text, including stories or poems. All you have to do is to give it a few lines of sentences that start off the text. Then you are given a scroll slider to set the length of words you want InferKit to generate. You are also able to manually select multiple keywords you would like InferKit to include in its output. InferKit also comes in with advanced settings, allowing its users to control nucleus sampling top p and sampling temperature, which controls creativity. For example, the higher the value for sampling, the more unusual the output text will be. The AI does require subscription for unlimited use. Otherwise, your weekly free characters limit will be set to 10,000 characters.



The artist equivalent stakeholder for InferKit would be journalists as their journals, new articles, etc. would be used for training InferKit. This violates fairness which is one of Microsoft's AI guiding principles. Not all journalists have agreed to the requests of using their journals to be used for training AI as a part of its dataset. This would be unfair for journalists as all their hard work would be referenced in the process of generating text media by the AI and the journalists would have no idea how their works would have been used in what way. Worst case scenario, if a journal which InferKit was heavily used was to be awarded journalist's awards such as the Pulitzer Prize, some journalists will find this situation to be extremely unfair as one would never know if their article was used by InferKit, and they never approved InferKit to use their work. As of right now, such cases where a journal written by the likes of InferKit have not been reported yet. However, this might be due to the fact that it is difficult to distinguish text generated by humans and AIs. There might be cases where InferKit-written journals won an award. We will never know.

Reliability is also an issue to the public. Since the algorithm of how the AI generates text isn't transparent to the public, its outcome may also not be reliable. In the case of InferKit, it had no measures of security when it comes to the user inputting violent, sexual, or grotesque keywords. The user was successfully able to pass in inappropriate keywords. Although InferKit did seem to have some filtering functionalities and did not include such keywords in its text outcome, with the algorithm being not transparent to the public, InferKit can't be reliable. The public never knows when the filtering would encounter a bug, and inappropriate keywords might make it through to the final text outcome. This bug might negatively influence an immature audience who still needs guidance.

Findings & conclusion

After doing study, we are still of the opinion that AI painting technology has had a favorable effect. The use of digital technology and the rise of digital network platforms have altered artists' approaches to their creative processes.

Due to the diversity of creative themes, it has liberated their mind from being constrained to a certain historical era or genre, laying the groundwork for enhancing the substance of artworks. Additionally, it expands the horizons of the artists by liberating their creative thinking and making it easier for their imaginations and creativity to shift course. In other words, conventional painting techniques and artificial intelligence (AI) painting both widen artists' creative ideas. We view AI art as a component of art, not art itself, as modern AI models still significantly rely on human input. There is no denying that significant artistic works are now possible, but we should view digital technology as a tool and a way to make art rather than relying only on it. Because AI cannot function or create quality work without human direction and control.

Recommendation

With the development and popularity of AI drawing technology, the issue of data provenance is becoming a worth considering problem. As some of the training data generated by the AI model used may include copyrighted images. When AI-generated artwork creates value, who should profit from it? The solution might be to change the source of the dataset for training the model to a platform like pinterest, as the images therein as well as the ideas are recognized for use in other creations. However, using this approach may lead to a lack of diversity of datasets and specific attributes, thus making the generated models not powerful.

Whether AI-generated art should compete in the same competition as human-made art is another issue to take into account. The original purpose of art contests was for humans to compete with one another. The complexity of creating an AI art is substantially lower than that of conventional painting, hence it is unfair to artists to allow works produced by AI drawing software to compete in art contests as the technology develops. We might divide the competition into AI art and human art competitions to prevent this problem and make it more equitable. We believe this could be achieved by requiring some middle-work procedure of the artwork it wants to enter human-only competition. Since AI-generated artwork only produces one final state of the drawing, we believe we can distinguish ai-generated artwork from human-generated artwork if participants were to submit images of their artwork of day 1, day 2, ..., last day.

Bibliography

- *InferKit Docs*. Text generation. (n.d.). Retrieved November 19, 2022, from <https://inferkit.com/docs/generation>
- Council on Foreign Relations. (n.d.). *AI code generation and Cybersecurity*. Council on Foreign Relations. Retrieved November 18, 2022, from <https://www.cfr.org/blog/ai-code-generation-and-cybersecurity>
- *Microsoft clip art now uses the dall-e AI program that draws anything you ask*. Nerdist. (2022, November 4). Retrieved November 12, 2022, from <https://nerdist.com/article/dall-e-ai-program-images-from-text-prompt/#:~:text=DALL%2DE%2C%20the%20artificial%20intelligence,the%20better%20it%20will%20be>
- Pal, K. (n.d.). *What is the impact of AI on art?* Techopedia.com. Retrieved November 12, 2022, from <https://www.techopedia.com/what-is-the-impact-of-ai-on-art/2/33399>
- Dhinakaran, A. (2022, September 20). *Overcoming AI's transparency paradox*. Forbes. Retrieved November 23, 2022, from <https://www.forbes.com/sites/aparnadhinakaran/2021/09/10/overcoming-ais-transparency-paradox/?sh=7e223a054b77>
- Cetinic, E., & She, J. (2022). Understanding and creating art with AI: Review and outlook. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 18(2), 1–22. <https://doi.org/10.1145/3475799>
- Liu, X. (2020). Artistic reflection on artificial intelligence digital painting. *Journal of Physics: Conference Series*, 1648(3), 032125. <https://doi.org/10.1088/1742-6596/1648/3/032125>