DOI: https://doi.org/10.28925/2519-4135.2024.49

UDC 7.038.5:004.89:004.42

Volodymyr D. TSYSARSKYI,
Postgraduate student,
Borys Grinchenko Kyiv Metropolitan University,
Kyiv, Ukraine,
ORCID: 0009-0003-7659-1118,
e-mail: v.tsysarskyi.asp@kubg.edu.ua

FEATURES, PROBLEMS, AND PROSPECTS OF USING ARTIFICIAL INTELLIGENCE IN THE CONTEMPORARY ART FIELD: WRITING, FINE ARTS, VOICE ACTING

Abstract. This article describes most things related to the rapidly developing Artificial Intelligence which use Neural networks to create all sorts of stunning things. The article covers what exactly Artificial Intelligence is, its short history of development and some techniques it uses. Next comes AI-image generation, which is quite popular right now as it opens possibilities that artist have dreamt for a long time. In the world of fine arts, AI algorithms inspire artists by offering fresh perspectives, styles, and techniques. Here the main advantages of using such AI is described, why it suddenly become popular, roughly explained about its algorithm of work, how users can actually use it and named the most popular tools. Also there described weaknesses of such image generators such as actually low understanding of word-object relationship, it is hard for it to grasp abstract things such as figurative meanings and hard to control what author exactly want to generate. The article also describes usage of Neural Networks in writing field. In the realm of writing, AI-powered tools assist in crafting plots, developing characters, and even producing complete articles or stories.

There it can be used both as a helper tool to find proper wording or improve writing style, as well as writing complete stories or at least helping writers to outline the first draft. It can also present advanced writing techniques in easily editable visual form by using experimental TaleBrush, which can greatly help beginners in writing field but also professionals. Not just that, but creating believable characters using Chat bots or enhancing language translation. AI can revolutionize voice acting by providing realistic and diverse AI-generated voices, opening up new possibilities for character creation. Voice acting is another field that would greatly benefit from using AI, as it allows to basically create entire songs and albums just using Vocaloids, do voiceovers for films and games with VoiceVox and create and edit podcasts with fast-learning AI text to speech synthesizers. In conclusion the future of AI tools were described such as advanced video editing, help with programming and possibility to create an entirely generated and live imaginary world in games with lively NPC which previously was just impossible.

Key words: artificial intelligence, AI models, neural models, fine arts, writing, audio, programming

Introduction. Artificial Intelligence (AI) has made remarkable strides in recent years, permeating various fields and leaving an undeniable mark on contemporary art. This article delves into the features, problems, and prospects associated with the integration of AI in these specific domains of the art world: fine arts, writing, voice acting, programming, game development and more. By examining the advantages and challenges that arise from this fusion, we can gain a comprehensive understanding of how AI is shaping the landscape of contemporary art.

AI brings forth lots of features that improve modern art field. Firstly, it serves as a catalyst for creativity, enabling the

generation of innovative ideas and novel concepts that push the boundaries of traditional art forms. In the realm of writing, AI-powered tools assist in crafting plots, developing characters, and even producing complete articles or stories. Similarly, in the world of fine arts, AI algorithms inspire artists by offering fresh perspectives, styles, and techniques. Additionally, AI can revolutionize voice acting by providing realistic and diverse AI-generated voices, opening up new possibilities for character creation. Same with programming, in which tools like ChatGPT allow programmers to rapidly prototype ideas before actually implementing them, or even just consulting for resolving the problem that is hard to find on the internet or hard to solve in general. For video editing, it brings new features such as automatic slicing and transitioning from one part of the video to another. Not to mention functions like completely changing scenery or adjusting colors on the fly. All of this can be combined for game development with prospects for developing, or at least prototyping, complete games from scratch just with the usage of AI.

Furthermore, AI contributes to the art world by automating and enhancing efficiency in artistic processes. In writing, AI tools streamline tasks such as grammar and spell-checking, content generation, and language translation, allowing artists to focus on their creative vision. In fine arts, AI-assisted tools automate repetitive tasks like colorizing sketches or creating digital prototypes, enabling artists to devote more time to their artistic expression. Voice acting also benefits from AI, as it facilitates voice modulation and sound effects, reducing the need for extensive post-production work.

Problem statement is that the AI and Neural networks, although have a long history by now, is still a fairly new thing. And the goal of this article is to overview this new technologies and determine their strengths, their problems and potential prospects for further development.

Analyses of recent research and publications. This essay can be called one of the first in our country dedicated to this issue. Today, it is possible to find only fragmentary works devoted to this problem, so it belongs to promising aspects of research.

The purpose of this article is to provide overview of the current AI technologies and most popular use cases for end users that anyone can try out.

Results. The integration of AI in the contemporary art field is not without its challenges. Ethical considerations come to the forefront, raising questions about authorship and creativity. Debates ensue in writing regarding the originality and ownership of AI-generated content. In fine arts, concerns arise regarding the attribution and value of AI-generated artworks in the art market. Additionally, voice acting faces the challenge of maintaining the uniqueness and authenticity of human performances when AI-generated voices become indistinguishable from those of real actors.

Despite these challenges, the prospects of AI in the contemporary art field are promising. AI can serve as a collaborative tool, augmenting artists and expanding their creative process. In writing, AI suggests alternative ideas, improves language usage, and aids in research. In fine arts, AI assists in exploring new mediums, materials, and techniques. Voice acting benefits from AI by providing additional vocal effects and expanding the range of emotional expression.

Because of that, more and more companies are implementing AI tools in their workflows to improve efficiency of work and minimize expenses.

In the “Image Generation: A Review” [5] by Mohamed Elasri, Omar Elharrouss, Somaya Al-Maadeed and Hamid Tairi, the authors discuss the challenges associated with generating images from different types of data in the field of computer vision. They highlight the difficulties in manually capturing

images from various perspectives to create objects or products. However, with the advent of deep learning and artificial intelligence, image generation from diverse data has become feasible. The authors claim that significant efforts have been dedicated to developing image generation strategies, leading to notable achievements.

The main contribution of their paper is to provide a comprehensive overview of existing image generation methods. They aim to describe each technique based on the algorithms employed, the type of data used, and the primary objective of the approach. Additionally, the authors discuss each category of image generation by presenting the proposed methodologies. They also present information about existing image generation datasets, offering a comprehensive understanding of the available resources.

Furthermore, the authors address the issue of evaluation metrics suitable for each image generation category. They discuss the metrics in detail and compare the performance of existing solutions to establish the current state-of-the-art. This comparative analysis helps identify the limitations and strengths of different approaches, contributing to a deeper understanding of the subject matter.

Finally, the authors acknowledge the challenges currently faced in the field of image generation. However, their text does not explicitly mention what these challenges are, leaving the reader curious about the specific obstacles that need to be overcome in this domain.

In summary, the authors present a valuable contribution to the field of computer vision by providing an extensive overview of image generation methods. They outline the different techniques, discuss the associated datasets, evaluate the performance of existing solutions, and highlight the current challenges. This analysis offers insights into the state-of-the-art

and paves the way for further advancements in image generation.

But the main goal of this article is to overview modern tools in AI generation with some app examples and alternatives for end-users to try it out.

Artificial Intelligence in general and its short history. The origins of AI have long been debated, with various contenders including Charles Babbage, Alan Turing, McCulloch, Hebb, and Wooldridge in the late 1980s and early 1990s. According to Wooldridge, AI began with Turing's paper on machine intelligence, marking the first wave of AI from the 1950s to the 1980s. This era, known as the Golden Age of AI, introduced symbolic AI, ELIZA, SHRDLU, LISP, and the concept of narrow AI.

However, this wave of AI faced a significant setback with Lighthill's report, as it exposed the flawed assumptions and lack of progress in AI research at the time. Subsequently, knowledge-based AI systems, also known as expert systems, emerged. This period witnessed the development of MYCIN, PROLOG, logic-based AI, Cyc, knowledge graphs, knowledge bases, inference engines, and knowledge engineering. Unfortunately, these logic-based systems failed when confronted with common-sense questions that didn't strictly adhere to logical reasoning.

Robotic systems came into play next, relying on logical AI. However, their capabilities were limited to specific tasks in controlled environments, as the symbolic and logical representation of complex real-world situations proved challenging. In response, Rodney Brooks proposed behavioral AI, advocating for robotic systems situated directly in the environment, with intelligence emerging from interactions rather than preprogrammed rules. Although the idea was sound, its implementation lagged behind.

The advent of agent-based AI systems followed, with the emergence of technologies like Siri, AI assistants, reasoning

under uncertainty, and Deep Blue defeating Gary Kasparov in Chess. This occurred at the turn of the millennium, leading some to believe that AI had reached its pinnacle.

However, with advancements in hardware, new possibilities arose. Artificial neural network (ANN) systems based on biological neural nets gained traction. This technology evolved from perceptrons to connectionism and eventually Deep Learning, culminating in DeepMind's AlphaGo defeating a Korean Go black belt. Presently, self-learning systems have surpassed human capabilities, as demonstrated by DeepMind's mastery of Atari 2600 console games and AlphaGo's Go strategies. These systems possess a self-programming nature that surpasses mere Turing machines.

Wooldridge argues that despite these achievements, we are far from achieving fully adaptive and lifelike AI. The complexities of human intelligence and life cannot be equated to games like Atari or Go. He introduces a section titled "Be careful what you wish for," cautioning that engineering systems, including AI, often fail in unpredictable ways, leading to unforeseen errors and unexpected actions. The paper clip factory scenario serves as a paradigmatic example.

In reflection, all AI revolutions thus far have fallen short of achieving the dream of AGI (Artificial General Intelligence). Wooldridge emphasizes that the measure of AI's progress lies in its ability to attain AGI, not just in exceeding human capacities in specific tasks. The assumptions about the mind, reality, reasoning, thinking, and human nature underlying previous AI systems have proven incorrect. Descartes' and Turing's perspectives on reality and human thought were flawed, and the complexities of life cannot be reduced to clear, logical steps or simplistic world models.

Knowledge cannot be reduced to deterministic, static rules, as evident from the failure of knowledge systems in the second wave of AI. The structure of artificial neural nets should

not be conflated with the intricacies of the human brain. Presently, we lack a comprehensive theory of the mind and an understanding of its mechanisms. The failure of AI technology, therefore, reflects our philosophical limitations and incomplete knowledge of the mind. AI research feels like groping in the dark, as we attempt to replicate something we don't fully comprehend.

Wooldridge concludes by stating that while AGI is not impossible, we are currently far from achieving it, and our current trajectory is not leading us in the right direction. Replicating something we don't fully understand remains a significant challenge. [6]

Artificial Intelligence image generator. In the paper titled "A Review of Image Generation" authored by Mohamed Elasri, Omar Elharrouss, Somaya Al-Maadeed, and Hamid Tairi, the authors discuss the highly challenging task of creating images from various types of data, such as text, scene graphs, and object layouts, within the field of computer technology. Thanks to the advancements in deep learning and artificial intelligence, it has now become possible to generate new images from different types of data. The paper provides a comprehensive overview of existing methods for image generation, presenting a description of each technique based on the adopted algorithms, data types utilized, and main objectives.

Furthermore, the authors delve into each category of image generation, elucidating the proposed approaches. They also present an overview of existing datasets for image generation, and discuss evaluation metrics suitable for each category, allowing for a comparison of performance among different solutions. This comparison serves to inform the state-of-the-art in the field, while identifying the strengths and limitations of each approach.

Finally, the paper addresses the current challenges that exist within this subject, shedding light on the obstacles that

researchers and practitioners face in the pursuit of further advancements in image generation. [5]

This article dives deep into this subject, and is truly a comprehensive overview of most techniques used for image generation. But it is a bit hard to understand for a new readers in this area and doesn’t give examples how users themselves can try this thing out.

“Artificial Intelligence image generator” is an app for creating images using neural networks, which in turn are trained to create these images based on the works of other artists or real photographs. [17]

This "training" is somewhat similar to how humans learn. For example, you can imagine a toddler who is just beginning to learn about the relationship between words/concepts and objects/beings. Artificial intelligence operates on the same principle, but the difference is that it has much more data and trains much faster.

Compared to ordinary artists, such artificial intelligence has the following advantages:

Relative ease of use. In comparison to a long and painstaking art education, learning to use AI is quite simple.

Variability. AI can create up to four variations of an image at one request at a time. If you create them again, another four variations will appear, and so on until you find a good one. Few artists, for example, will agree to create more than a hundred high-quality and painstaking works on one topic, while artificial intelligence doesn't care how many requests you ask it.

Quick creation and correction. Usually, it takes about a few minutes to create an image by artificial intelligence, and often even less. No human artist can physically create such detailed works in such a short period of time. Also, if you don't like something, you can simply clarify your request and fix it in a matter of minutes. This is not even taking into account the many variations of one image. The artist always needs to be paid

extra for corrections, and such corrections often take as much time as the image itself.

Cheapness. Compared to ordering from most artists, where prices start at several tens of c.u. for a single image, buying a subscription for the same price to create thousands of such images seems much more profitable. Not to mention that if you have a gaming computer at your disposal, you can create many such images for free.

The diffusion technique was invented in 2015 by AI researchers at Stanford University. The figure below (fig. 1), taken from the Stanford team's research, illustrates the two phases of the diffusion process using training data in the shape of a spiral [15].

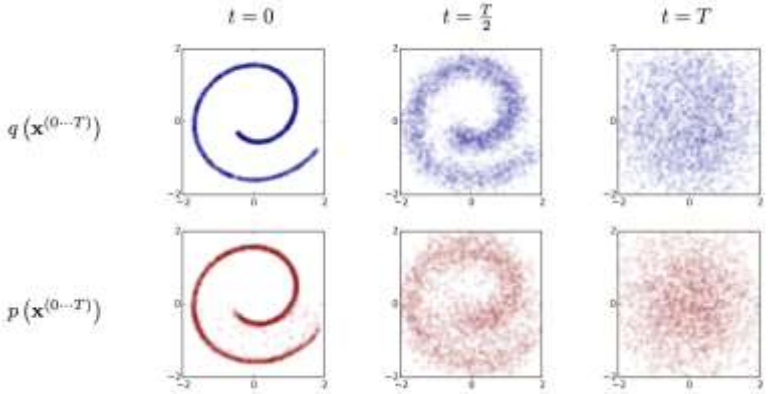

Fig.1 How diffusion based neural networks work

The first phase in diffusion is to take an image (or other data) and progressively add more visual noise to it in a series of steps (This process is depicted in the top row of the diagram). At each step, the AI records how the addition of noise changes the image. By the last step, the image has been "diffused" into essentially random noise.

The second phase is like the first, but in reverse. (This process is depicted in the bottom row of the diagram,

which reads right to left.) Having recorded the steps that turn a certain image into noise, the AI can run those steps backwards. Starting with some random noise, the AI applies the steps in reverse. By removing noise (or "denoising") the data, the AI will produce a copy of the original image.

In the diagram, the reconstructed spiral (in red) has some fuzzy parts in the lower half that the original spiral (in blue) does not. Though the red spiral is plainly a copy of the blue spiral, in computer terms it would be called a *lossy* copy, meaning some details are lost in translation. This is true of numerous digital data formats, including MP3 and JPEG, that also make highly compressed copies of digital data by omitting small details

In short, diffusion is a way for an AI program to figure out how to reconstruct a copy of the training data through denoising. Because this is so, in copyright terms it is no different than an MP3 or JPEG — a way of storing a compressed copy of certain digital data.

The article "Testing Relational Understanding in Text-Guided Image Generation" by Conwell, C., & Ullman, [4] describes that as of right now neural networks still struggle to fully understand the meaning of the objects they are trying analyze. Even on primitives (fig.1) they generate unnecessary noise. Same things of course will happen to much more complex images.

For users it is actually easy to use AI image generator. They just need to type a prompt in words, trying to describe it as precise as they can. Users can choose the style they want to be generated, or even try to ask AI to mimic some other's artists style. The simplest usage is — enter a word prompt, then either choose among four generated variants (fig.2) or regenerate anew altogether.

The word prompt should be as precise or as vague as user desire depending on the results they want to get. In article "The Creativity of Text-to-Image Generation" by Jonas Oppenlaender

[10] author describes that AI image generators can actually generate good artworks if prompt is precise and they were trained on the objects they have to generate. For example, if model is trained on pictures of “beautiful girls” and the author describes exactly a girl of the dreams with precise adjectives, there is higher chances user will get eventually what exactly were imagined.

But if the model were trained on animals and were given a prompt to generate image of human, it will not be able to do it, of course, because it doesn’t even know that such thing actually exist (for this specific trained model, there are plenty of them).

But with something vague, unspecific AI neural network will try its best to guess what author wants, because author may not himself know it and try to conceptualize his ideas.

The worse results comes with abstract or figurative meaning. If human can somehow depict it with its vocabulary, specific grammar slang or with its knowledge about surroundings and traditions, the AI just don’t know this things. It was trained to understand and depict objects “head on” without any hidden sense. Even then, with direct and distinctive words it can fail to understand the meaning, so it’s no surprise that when asking something vague it will probably fail altogether.

Users can then generate variants of the selected images to get the details they want or upscale selected image to desired resolution.

Of course, this is just the simplest method, which uses online services or servers for free or with paid plans. The most popular online paid services is Dall-E and Midjourney. The most popular free one’s are mage.space, playgroundAi, NovelAI, Wombo Dream, etc.

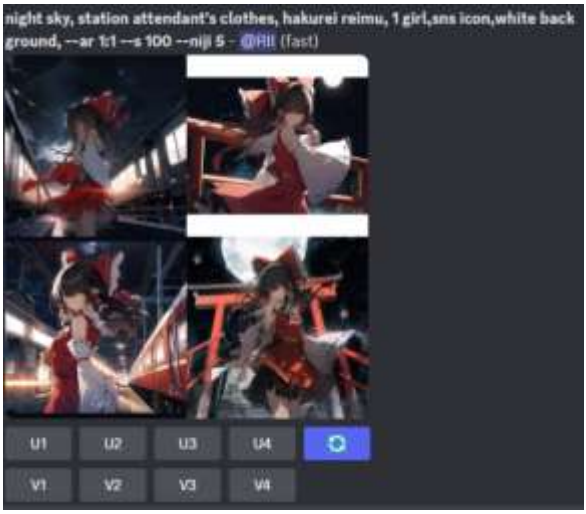

Fig.2. Generated images based on a prompt in Discord server "Niji journey"

Interestingly, the article "Text to image generation: leaving no language behind" by Reviriego, P., & Merino-Gómez, E [12] describes that AI image generators are also sensitive to languages used to enter a word prompt. The best results were achieved using English language as it was most widely used language. With the less known languages, such Spanish, Basque, Latin, the results were much worse. Most probably, with asian languages which uses hieroglyphs in their writing system, which has different meaning depending on the context (like for Kanji in Japanese) AI image generator would really struggle to understand what exactly it should depicture, especially in cases where even humans can be confused.

But users can also install local Stable Diffusion AI image generator for free if they have powerful enough PC or laptop. With this option, they can choose which neuro-model they want to use, download others models if they need specific style (for example, from civitai.com) or they can even train their own neuro-model for generating images in artist's style from their

pictures (recommended amount 20+ images for model training). The most recent thing is ControlNet, which allows users to further specify how should neuro-model generate image based on example. Example can be some simple outline image (fig. 3), human pose from sticks, some instructions on how anatomy works etc. With this tool it is now possible to generate even photorealistic images exactly how user wants with appropriate style and details (Fig.4). Especially hands, which are now drawn correctly.

Nvidia's Canvas app actually lets users do similar things that Control Net does — it turns rough sketch or outline into finished painting using assigned attributes. For example, pick a "sand" brush and draw sand area, pick "water" brush and draw where waterfall and lake should be. All results will be displayed after a short wait on the right part of the app (similar to fig.3, but this time left area is a drawing area). This app does require Nvidia's GPU to work at all though.

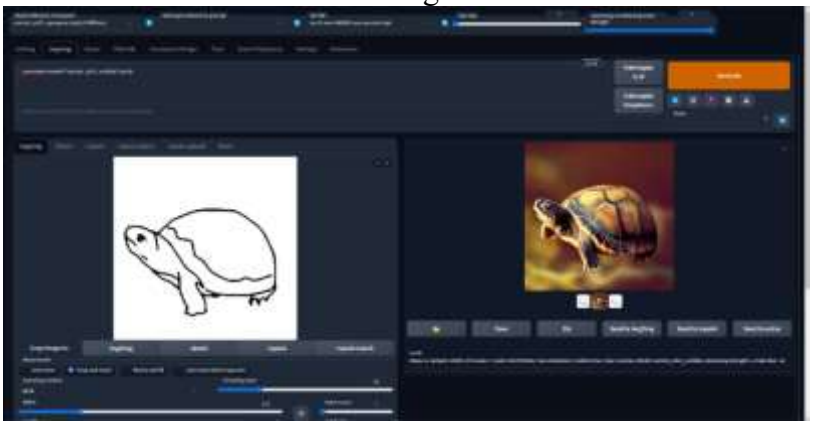

Fig.3. Stable Diffusion in WebUI with ControlNet

Fig.4. ControlNet can detect human pose and generate new image based on that

But can images be generated exactly as authors wants? Right now, AI artists must understand exactly which tech they are using, which settings are best for which model and how exactly type a prompt to get the picture they want. If they want to be even more specific, they had to use Control Net or otherwise neural network will not obey them.

AI storyteller assistant or novel generator. AI neural networks can help humans achieve many things, and image generation is only one example of them. Another logical example would be AI assistance in writing or even generating novels or poems from scratch. Same as with images, scientists takes a large pool a data, literally billions and billions of raw text from fiction and non-fiction, and then train the models from this data. The end result is that trained neuro-model can now understand the meaning of the words (object-semantic relationship), can understand vocabulary and grammar of the languages, can generate sentences with relation to one another and even can imagine (or at least collage) new characters, scenery and the general story for events.

Such thing is called AI-assisted writing or AI-storyteller.

The most prominent examples of such tools are Novel.ai, QuillBot AI, Kobold.ai, ChatGPT and Character.ai.

The most simple use case of such AI is just rephrasing and helping finding the right word (synonyms) for the writers. The interface is quite simple: writer just need to insert text in left column and then AI will generate words replacements suggestions in the right column (fig. 5). If needed, the writer can click on suggested word to get some other suggestions (synonyms) that are better suited for that text. However, such "simple" things could be done even earlier, just with a bit worse quality.

Fig.5. QuillBot AI UI with words suggestions for better readable and consistent writing

The new possibilities, which AI neural networks opens for writers, is actually writing the whole story by itself, which was impossible before. At the very least, it can greatly help writer in shaping the plot for their first draft.

NovelAI is one of such AI-tools, and Kobold.ai is its free local alternative. The writer have to define the genre of this novel, choose neural model best suited for this task, fill some broad details about this story like things AI should remember or avoid, and finally enter the first few sentences (fig. 6).

Fig.6 NovelAI UI for generating and correcting stories

In this program, the text that AI generated will be white, the corrections the author made to the generated text will be pink, and the prompts or sentences author writes further will be blue.

With this tool, writers can experiment with the story, characters, environmental setting etc. It is arguable if the writer can publish such a book without lots of rewriting or corrections, but as for a first draft to help establish what the story is about NovelAi is a great tool. It can even analyze writer's previous books (if uploaded) and adapt its writing style to match the author's writing style more correctly.

Before such advanced AI tool, more primitive widely used AI text generator was voice assistant "Siri" by Apple [14]. Several other articles [18, p. 1] covers technical side of story generation and gives suggestion on how it can be improved.

And another research tool called "TaleBrush" [3] is aimed to further advance tooling for AI writing app by allowing writers to visually modify story events, pacing, character growth, scene depiction etc. by using and editing visual graphs. These tools, which visualize advanced writing concepts in such simple way, could really revolutionize how story is written, or can greatly help writers create better stories much faster.

There is another related research paper named “Deep Learning-Based Short Story Generation foran Image Using the Encoder-Decoder Structure” by Kungbok Min, Minh Dang, and Hyeonjoon Moon. [8] In that article authors describes captures and story generation from image recognition. So it uses similar neuro models that used for AI image generation, but in reverse: image generation recognizes meaning from entered prompt and tries to create an image; but in this case it tries to understand what is actually presented on image and generate captions and short story based on that as closely as possible.

Surprisingly, Novel AI can do both things very similarly: it can help generate text for writers, but it also can generate images based on these stories. It’s like a multi tool for writers: generating both text for novel and illustrations for cover and for placing it in chapters. Truly a dream come true.

In addition, such AI apps now can easily generate poetry that is hard to distinguish from human’s writing. However, it still struggle with understanding the writers’ techniques for writing enjoyable books – sure, it can create a complete story from scratch. But it is a question whether such story will be enjoyable at all or even will be a complete mess with little to no relationship in story structure…

Same thing also goes for ChatGPT is writer wish to use it for helping shaping the story. For example, ChatGPT have limited character memory, which means that after certain character count it will just forget everything it wrote and write something anew without slightest relation (or it can “lucky guess” in best cases).

So, with that in mind, it is impossible to create a consistent story using just ChatGPT. It is a great tool for prototyping ideas, generating/asking some details and getting piece of advice how to write better. Also, as the time will go, it will be even better tool for spelling/grammar checking than existing ones, Grammarly and ProWritingAid just to name a few.

The other interesting concept of such AI models, however, is the Chat-bot, but with pre-defined personality. Surely, programmers now can even integrate ChatGPT into their games for more interactive and believable characters… But they will not sound or behave like the real characters. This is due to the fact that ChatGPT have generic or neutral personality and it is programmed to always reply to the user in polite and respectful tone.

Character.ai was created exactly to solve this problem. It behaves similarly to ChatGPT — user write some text, literally anything they want (because it's AI's problem to recognize it), and get response from AI-model (fig.7).

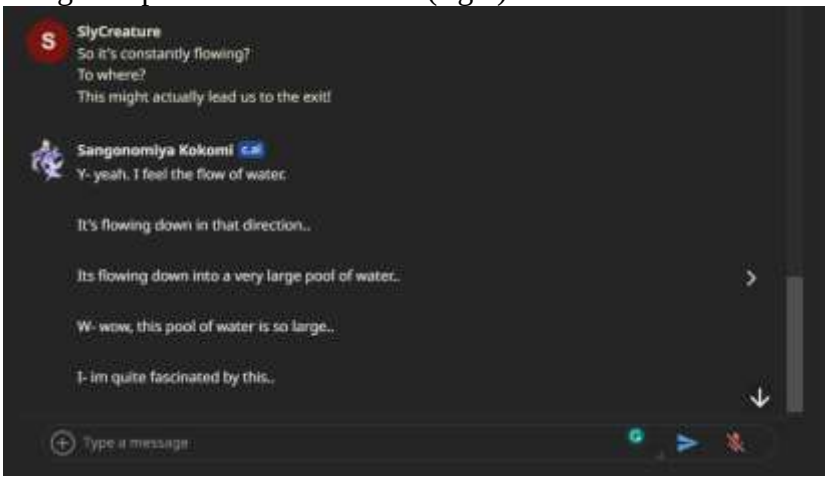

Fig.7. Character.ai UI chat with fiction character Kokomi from Genshin Impact

The biggest difference is that this AI replies with much more emotion, much more character and role-play. It is not just politely neutral, it can be happy, sad, crying, angry, surprised, depressed etc. just like a very good actor or simply as a charismatic person would. The best thing about this app is that the users themselves create these characters, define their character and then play/chat with them. Then this created

character is further defined/corrected by the users who chat with it via feedback. This feedback is a rating from one to five and with commentary or tags chosen by users to further shape this Ai-character. Moreover, users can choose which AI reply they want to select – just scroll right-to-left or press arrow to choose which reply is best suitable for the chat flow.

The article “Story Shaping: Teaching Agents Human-like Behavior with Stories” [11] describes technical side of how it is actually possible to character bot to mimic human thinking.

With such advanced AI neural network and perfectly defined character, it is sometimes really hard to distinguish whether is this AI or the real person you are chatting with. The biggest giveaway is the same as with ChatGPT – forgetfulness, repetition and common patters such as censorship. If after several hours of chatting user ask AI to remember the thing the discussed at the beginning of the chat – It just will not remember that because of the character’s count limitation. Or if the user ask something that is banned by model’s censorship rules (such as extreme violence or NSFW stuff), the AI character will just be confused, react aggressively or do something to avoid replying directly to such prompt. And not just one of the defined characters, but all of the AI characters will behave the same when it comes to censorship, unlike humans who can actually just ignore the rules, even severe ones.

Another honorable mention is DeepL language translator. Its interface is similar to QuillBot and it also allows to choose synonyms or just best variant for this translation. But it behaves more like Novel.Ai, meaning it analyses the semantics of the word and grammar of the languages. Based on this, it creates much meaningful and believable translation that just word-to-word translation, as it tries to mimic the way humans interpret language when they think about something.

AI text to speech generator. One of the most interesting and hardest things is to replicate human voice. The simplest use case for voiceover would be just recording separate characters and play them accordingly to the written text. The problem is – it would sound either robotic or eerie, most commonly both at the same time. When humans speak, they take into consideration how characters form syllables, what intonation they use (some languages are pronunciation sensitive), what emotion they want to convey (voice pacing and volume greatly differs based on that).

The earliest good example of successfully implementing such nuances of human speech was project "Vocaloid" led by Kenmochi Hideki at the Pompeu Fabra University in Barcelona, Spain, in 2000. It became especially popular in Japan with its Vocaloid mascot called Hatsune Miku, which became the singing idol for many audio fans in Japan and in world. It reached popularity comparable to a real human star, with albums and singles releasing to this day and some becoming worldwide hits. Not to mention popular merchandise and games based on this Vocaloid idol.

This project finally brought emotions to otherwise boring standard official-like text to speech generators. However, its distinctive feature that it tries to cover some of its failures in mimicking real human speech with robotic accent.

Another most recent text to speech synthesizer that users can try for free is VoiceVox developed by Hino in 2021. This speech synthesizer actually uses Deep learning neural networks – the same ones that were described earlier for image generation and writing.

Because these models are trained using full recordings of human speech and analyzing every aspect of pronunciation: intonation, pace, emotion, meaning; they can produce speech that closely resembles humans voice.

As a result – nowadays it is possible to actually create a believable character using just an app without voice actor. At least for non-extreme or very specific and distinguishable voice acting.

The app itself is easy to use – just write some text, choose the model you want to use, choose the mood of the model (intonation differs based on that) and fine tune some settings to get the voiceover you need (fig.8).

Fig.8. VoiceVox text to speech synthesizer based on neural networks user interface (currently only in Japanese with Japanese voices)

Another great examples of such text to speech programs are Eleven labs Prime AI for realistic human voice in English and 15.ai for cartoon voices from popular TV shows or games.

From technical side, there are some articles that discuss how exactly such models works [2, 7, 13, 19], how these models can adapt to human characteristics and mood [16], and how models such as VoiceVox can be adapted to English language [9].

But such models requires hours of recordings and tons of data to produce such believable results. There is another technology that uses deep learning neural networks but it analyses data right on the fly without any prerecorded samples.

The earliest widely known demonstration of such technology was presented by Adobe in 2016 and was named Adobe VoCo.

It can analyze literally seconds of human voice and still try to imitate it as best as it can, as it was showcased during the Adobe MAX 2016 Sneak Peeks.

The UI is similar to other text to speech synthesizers, but with some differences: it firstly analyzes the human speech, creates transcription and let user edit text to regenerate audio using sampled voice (fig.9).

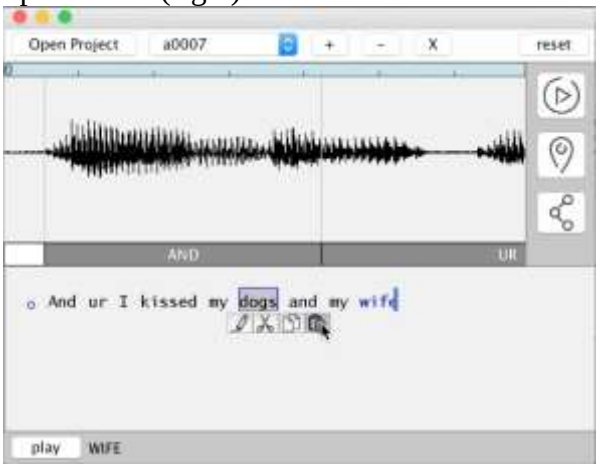

Fig.9. Adobe VoCo UI (tech preview, was not released to public)

Such technology could be used to generate voiceovers for podcasts, film and games based on very little input data like person talking for five to ten minutes.

And it is already implemented in tool for video and podcast editing called “Descript”. In its “Overdub” feature users can choose to voiceover text from pre-trained models, or choose

to overdub their script using their voice from recording. This way it creates model based on user's voice, and then users can either create overdub for text from scratch or edit pre-recorded podcast audio to match their needs.

Conclusions. In conclusion, it is worth saying that AI came a long road from continuous research with failures to the products users can use right now. This article described possible usage of AI for image generation, writing and voiceover, but AI is not limited to just them. It can be used in film industry with example of Adobe incorporating Adobe Firefly into its Premiere for faster video editing or even changing the video itself, like changing seasons from summer to winter. Programmers can already use ChatGPT to help them prototype code faster, and Unity game engine have Machine Learning agents package to help game developers create AI-characters who train themselves and trying to adapt to changing environment. Nvidia even demonstrated a sneak peak of their AI called ACE tailored specifically for NPC actions and dialogues. Some users even manages to create short videos, small 3d scenes or even simple 3d games using mostly AI neural networks as their primal tool.

All the positive point that were described for image generation, like relative ease of use, variability, quick creation/correction and cheapness can be applied to AI models in any creative or technical field. And because of that further development of AI is inevitable, because right now even big corporations, not just researchers, began investing in this field. For artists in any creative field it means that more and more convenient tools will be available for faster, better and cheaper content creation.

References

1. Alabdulkarim, A., Li, S., & Peng, X. (2021). Automatic story generation: Challenges and attempts. arXiv preprint

arXiv:2102.12634. Available at: https://doi.org/10.48550/arXiv.2102.12634 [in English]
2. Arık, S.Ö., Chrzanowski, M., Coates, A., Diamos, G., Gibiansky, A., Kang, Y., ... & Shoeybi, M. (2017, July). Deep voice: Real-time neural text-to-speech. In International conference on machine learning (pp. 195-204). PMLR. Available at: https://doi.org/10.48550/arXiv.1702.07825 [in English]
3. Chung, J.J.Y., Kim, W., Yoo, K.M., Lee, H., Adar, E., Chang, M. (2022, April). TaleBrush: visual sketching of story generation with pretrained language models. In CHI Conference on Human Factors in Computing Systems Extended Abstracts (p. 1-4). Available at: https://doi.org/10.1145/3491101.3519873 [in English]
4. Conwell, C., & Ullman, T. (2022). Testing relational understanding in text-guided image generation. arXiv preprint arXiv:2208.00005. Available at: https://doi.org/10.48550/arXiv.2208.00005 [in English]
5. Elasri, M., Elharrouss, O., Al-Maadeed, S., & Tairi, H. (2022). Image generation: A review. Neural Processing Letters, 54(5), 4609-4646. Available at:: https://doi.org/10.1007/s11063-022-10777-x [in English]
6. Krzanowski, R. (2021). The road to conscious machines: AI through failed ideas. Zagadnienia Filozoficzne w Nauce, (70), 171-181. DOI: 10.13140/RG.2.2.15150.23360 [in English]
7. Lee, J., Choi, H. S., Jeon, C. B., Koo, J., & Lee, K. (2019). Adversarially trained end-to-end Korean singing voice synthesis system. arXiv preprint arXiv:1908.01919. Available at: https://doi.org/10.48550/arXiv.1908.01919 [in English]
8. Min, K., Dang, M., & Moon, H. (2021). Deep learning-based short story generation for an image using the encoder-decoder structure. IEEE Access, 9, 113550-113557. Available at: 10.1109/ACCESS.2021.3104276 [in English]

9. Nakamura, K., Oura, K., Nankaku, Y., Tokuda, K. (2014, May). HMM-based singing voice synthesis and its application to Japanese and English. In 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 265-269). IEEE. Available at: 10.1109/ICASSP.2014.6853599 [in English]
10. Oppenlaender, J. (2022, November). The Creativity of Text-to-Image Generation. In Proceedings of the 25th International Academic Mindtrek Conference (pp. 192-202). Available at: https://doi.org/10.1145/3569219.3569352 [in English]
11. Peng, X., Cui, C., Zhou, W., Jia, R., & Riedl, M. (2023). Story Shaping: Teaching Agents Human-like Behavior with Stories. arXiv preprint arXiv:2301.10107. Available at: https://doi.org/10.48550/arXiv.2301.10107 [in English]
12. Reviriego, P., & Merino-Gómez, E. (2022). Text to Image Generation: Leaving no Language Behind. arXiv preprint arXiv:2208.09333. Available at: https://doi.org/10.48550/arXiv.2208.09333 [in English]
13. Shi, J., Guo, S., Qian, T., Huo, N., Hayashi, T., Wu, Y., ... & Jin, Q. (2022). Muskits: an end-to-end music processing toolkit for singing voice synthesis. arXiv preprint arXiv:2205.04029. Available at: https://doi.org/10.48550/arXiv.2205.04029 [in English]
14. Thorne, S. (2020). Hey Siri, tell me a story: Digital storytelling and AI authorship. Convergence, 26(4), 808-823. Available at: https://doi.org/10.1177/1354856520913866 [in English]
15. Tsysarskyi, V. (2023). Problemy etychnosti vykorystannia robit khudozhnykiv dlia trenuvannia ai-modelei ta problema konkurentsii takykh modelei i khudozhnykiv na komertsiinomu rynku [ethical problems of using artists' works for training AI models and the problem of competition of such artist models on the commercial market]. Mather. Nauk.-

pr.konf. “Etnokulturni tradytsii v obrazotvorchomu mystetstvi ta dyzaini Ukrainy”. Kyiv, 2023. P. 8 [in Ukrainian].
16. Van Rijn, P., Mertes, S., Schiller, D., Dura, P., Siuzdak, H., Harrison, P., ... Jacoby, N. (2022). VoiceMe: Personalized voice generation in TTS. arXiv preprint arXiv:2203.15379. Available at: https://doi.org/10.48550/arXiv.2203.15379 [in English]
17. Wooldridge, M. (2020). The road to conscious machines: The story of AI. Penguin UK. [in English]
18. Yao, L., Peng, N., Weischedel, R., Knight, K., Zhao, D., Yan, R. (2019, July). Plan-and-write: Towards better automatic storytelling. In Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 33, No. 01, pp. 7378-7385). Available at: https://doi.org/10.1609/aaai.v33i [in English]
19. Yi, Y. H., Ai, Y., Ling, Z. H., Dai, L. R. (2019). Singing voice synthesis using deep autoregressive neural networks for acoustic modeling. arXiv preprint arXiv:1906.08977. Available at: https://doi.org/10.48550/arXiv.1906.08977 [in English]

Володимир Дмитрович ЦИСАРСЬКИЙ,
аспірант,
Київський столичний університет імені Бориса Грінченка,
Київ, Україна,
ORCID: 0009-0003-7659-1118,
e-mail: v.tsysarskyi.asp@kubg.edu.ua

ОСОБЛИВОСТІ, ПРОБЛЕМИ ТА ПЕРЕСПЕКТИВИ ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ В СУЧАСНОМУ МИСТЕЦЬКОМУ ПРОСТОРІ: ПИСЬМЕННИЦТВО, ОБРАЗОТВОРЧЕ МИСТЕЦТВО, ОЗВУЧУВАННЯ

Анотація. Ця стаття аналізує багато аспектів, пов'язаних зі стрімким розвитком штучного інтелекту, який використовує нейронні мережі для створення багатьох корисних речей. У статті розповідається про те, що таке штучний інтелект, дана коротка історія його розвитку та деякі методи, які він використовує. Описано AI-генерацію зображень, яка є досить популярною зараз, оскільки відкриває можливості, про які художники мріяли протягом тривалого часу. У світі образотворчого мистецтва AI-алгоритми надихають художників, пропонуючи свіжі перспективи, стилі та техніки. Тут описано основні переваги використання такого AI, чому він раптово став популярним, приблизно пояснено алгоритм його роботи, як користувачі можуть ним користуватися, а також названо найпопулярніші інструменти. Також описані слабкі сторони таких генераторів зображень, такі як фактично низьке розуміння зв'язку «слово-об'єкт», йому важко зрозуміти абстрактні речі, такі як переносні значення, і важко зрозуміти, що саме автор хоче згенерувати. У статті також описується використання нейронних мереж у сфері письма. У сфері письменництва інструменти на основі AI

допомагають створювати сюжети, розробляти персонажів і навіть створювати цілісні статті або оповідання. Вони можуть використовуватися як допоміжний інструмент для пошуку правильних формулювань або покращення стилю письма, а також для написання повноцінних історій або, принаймні, допомоги авторам у створенні перших начерків. За допомогою експериментального TaleBrush можна також представити передові техніки письма у легко редагованій візуальній формі, що може дуже допомогти як початківцям, так і професіоналам у сфері письменництва. І не тільки це, але й створення правдоподібних персонажів за допомогою чат-ботів або покращення мовного перекладу. AI може зробити революцію в озвучуванні, надаючи реалістичні та різноманітні голоси, згенеровані штучним інтелектом, відкриваючи нові можливості для створення персонажів. Озвучування – ще одна сфера, яка значно виграє від використання AI, оскільки дозволяє створювати цілі пісні та альбоми лише за допомогою Vocaloids, озвучувати фільми та ігри за допомогою VoiceVox, а також створювати та редагувати подкасти за допомогою AI-синтезаторів, що швидко навчаються. На завершення було описано майбутнє AI-інструментів, таких як просунутий відеомонтаж, допомога з програмуванням і можливість створювати повністю згенерований і живий уявний світ в іграх з живими NPC, що раніше було просто неможливим.

Ключові слова: штучний інтелект, AI моделі, нейромережі, образотворче мистецтво, письменництво, аудіо, програмування