



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 4, April 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.379

 9940 572 462

 6381 907 438

 ijircce@gmail.com

 www.ijircce.com

Refined Interactive Platform Integrated with Dall-E Tool

Dr.P.Kavitha¹, Karthick Raja S²

¹Associate Professor, PG& Research Department of Computer Science, Sri Ramakrishna College of Arts & Science, Coimbatore, Tamil Nadu India

²UG Student, PG& Research Department of Computer Science, Sri Ramakrishna College of Arts & Science, Coimbatore, Tamil Nadu India

ABSTRACT: The combination of DALL-E and Whisper models in interactive platforms is a big step forward in how people interact with computers. DALL-E can create detailed pictures based on words, and Whisper can turn audio into text. This can change the way we use technology in art, design, and communication. It also raises important questions about how we use this technology responsibly. By working together, humans and machines can be more creative and understand each other better. This new way of interacting with computers helps us be more innovative and imaginative. New technology is going to change how we use computers and talk to each other. This new idea combines two special computer programs that can make pictures from words and turn audio into text. It will let us do new and exciting things on computers. There is a new platform that can help people create pictures just by describing what they want in words. It can make really cool and realistic pictures very quickly. There is also a feature that can understand and can turn audio into text. The platform also makes it easier for people to use by combining text, speech, and pictures. This can make communication and using the platform more fun and interactive. This platform can be used in many different ways, but we still need to learn more about all the things it can do. Imagine a special place where kids can learn in a fun way with interactive stories, talking pictures, and computer-made images. It's like a game that helps them understand things better. And on another fun website, people can share their thoughts and ideas with pictures made by computers. This can help them be more creative and have cool conversations. Also, designers and artists can use this special tool to come up with new ideas quickly. This new concept opens up a lot of exciting possibilities for how people can work and play together with computers.

KEYWORDS: Dall-E, audio, emotion intelligence, Transformers, hyper-parameters

I. INTRODUCTION

In the world of computers and technology, there is a challenge to make the way we interact with them better and more personalized. The current ways we do things on computers sometimes have limitations and aren't as exciting or easy to use as they could be. To fix this, we need to create a special platform that combines two really smart technologies called DALL-E and Whisper. DALL-E can make amazing pictures, and Whisper can turn audio into words. But it's not easy to make them work together smoothly and quickly. We need to figure out how to make these technologies work well together so that people can have an amazing and new way of interacting with computers.

The Refined Interactive Platform is a really cool new technology that combines two super smart AI models, DALL-E and WHISPER. It lets people use both pictures and voice to talk to the computer and get things done. It's like having a really smart friend who can help you with anything! DALL-E It can make really good pictures based on words people tell it. This program lets people be really creative and make all kinds of cool pictures. You can use it for work, learning, or just having fun. It can make pictures that show complicated ideas and stories. WHISPER is like a really smart computer program that can listen to people talking and write down what they say very accurately. It's really good at understanding different ways people talk, like if they have different accents or speak different languages. This helps make things easier for people to understand and communicate with each other.

When we put these two models together in the Refined Interactive Platform, something special happens. Now, kids can use both words and their voice to play and learn on the platform. It's really easy to use and can do lots of different things, like help with school work, let you make cool art, and talk to your friends. The platform is made to be easy for everyone to use, so even if you're not a computer expert, you can still access all the cool features of DALL-E and

WHISPER. The design is simple and easy to understand, so you can move around and use it quickly. The Refined Interactive Platform is a really cool computer program that uses artificial intelligence to help people work and play with computers in a better way. It combines two special AI technologies called DALL-E and WHISPER to make it easier for people to make and understand all kinds of things on the computer. This program is a big step towards a future where AI helps people be even more creative and communicate better with computers.

II. RELATED WORKS

Many companies' chatbots run on messaging apps or simply via SMS. They are used for B2C customer service, sales and marketing. In 2016, Facebook Messenger allowed developers to place chatbots on their platform. There were 30,000 bots created for Messenger in the first six months, rising to 100,000 by September 2017. Since September 2017, this has also been as part of a pilot program on WhatsApp. Airlines KLM and Aeroméxico both announced their participation in the testing both airlines had previously launched customer services on the Facebook Messenger platform. The bots usually appear as one of the user's contacts, but can sometimes act as participants in a group chat. Many banks, insurers, media companies, e-commerce companies, airlines, hotel chains, retailers, health care providers, government entities and restaurant chains have used chatbots to answer simple questions, increase customer engagement, for promotion, and to offer additional ways to order from them. A 2017 study showed 4% of companies used chatbots. According to a 2016 study, 80% of businesses said they intended to have one by 202

Used by marketers to script sequences of messages, very similar to an autoresponder sequence. Such sequences can be triggered by user opt-in or the use of keywords within user interactions. After a trigger occurs a sequence of messages is delivered until the next anticipated user response. Each user response is used in the decision tree to help the chatbot navigate the response sequences to deliver the correct response message. Other companies explore ways they can use chatbots internally, for example for Customer Support, Human Resources, or even in Internet-of-Things (IoT) projects. Overstock.com, for one, has reportedly launched a chatbot named Mila to automate certain simple yet time-consuming processes when requesting sick leave. Other large companies such as Lloyds Banking Group, Royal Bank of Scotland, Renault and Citroën are now using automated online assistants instead of call centres with humans to provide a first point of contact. A SaaS chatbot business ecosystem has been steadily growing since the F8 Conference when Facebook's Mark Zuckerberg unveiled that Messenger would allow chatbots into the app .

In large companies, like in hospitals and aviation organizations, IT architects are designing reference architectures for Intelligent Chatbots that are used to unlock and share knowledge and experience in the organization more efficiently, and reduce the errors in answers from expert service desks significantly. These Intelligent Chatbots make use of all kinds of artificial intelligence like image moderation and natural-language understanding (NLU), natural-language generation (NLG), machine learning and deep learning.

III. PROPOSED METHODOLOGY

Seamless Integration: Combine DALL-E's image generation and Whisper's audio capabilities to create a truly multimodal experience Iterative Refinement: Allow users to iteratively refine both the visual and audio aspects of their creations. Collaboration Tools: Enable users to collaborate on creations in real-time, sharing prompts, ideas, and edits seamlessly. Accessibility Features: Ensure the platform is accessible to users with disabilities, providing alternative input methods and output formats. Explain ability and Control: Offer users more control over the generation process and provide explanations for how DALL-E and Whisper arrive at their outputs.

Generative Pre-Trained Transformers Algorithm

GPT, or Generative Pre-trained Transformers, are like really smart robots that can understand and talk like humans. They learn by reading lots and lots of books and articles, and then they can use that knowledge to write their own stories and sentences. They use a special way of thinking called "attention" to understand what they read and write text that makes sense.

Transformer Architecture: IT is like a special structure that helps a model understand words and sentences. It has different layers that work together to understand how words relate to each other. This helps the model understand long sentences and the meaning behind them. It is really good at understanding and working with language.

Pre-training: Before they can do their job, GPT models go through a lot of practice. They read a bunch of books and articles with no help from anyone. They learn to guess what word comes next in a sentence based on the words that came before it. This helps them understand how words fit together and how they relate to each other.

Generative Capability: GPT is really good at making up sentences that make sense and fit with the topic. It can create text that sounds like it came from the information it was given. This is helpful for things like writing stories, having conversations with a computer, or making new content.

Fine-tuning for Downstream Tasks: Fine-tuning is like giving GPT models extra practice on different tasks to help them get better at things like understanding emotions in text, summarizing information, or answering questions. This helps the models become really good at these tasks and perform better than other models.

Scalability: GPT models can get better at understanding and creating language by training them on bigger and bigger sets of information. OpenAI has made different versions of GPT that are bigger and have more training data, which helps them understand and create language even better.

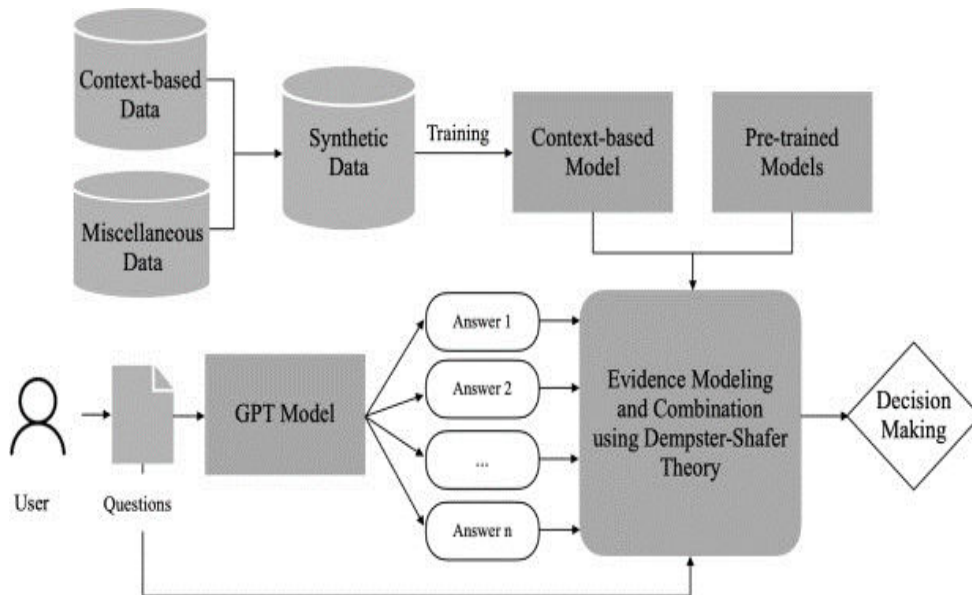


Fig-1 System Architecture

Data processing is like sorting and organizing different pieces of information so that it can be used in a helpful way. It's kind of like putting puzzle pieces together to make a complete picture. There are two different ways to handle information based on what type of information it is. The Synthetic training data is like practice data used to teach a computer program how to do something. Pre-trained models are like ready-made tools that the program can use without needing to be taught. The flowchart doesn't show exactly how the data is used, but it's likely that both practice data and ready-made tools are used to create a special type of computer program called GPT, which can do things like write stories, translate languages, and answer questions. Dempster-Shafer Theory is like a tool that helps us make sense of all the different answers we get from a computer program. It uses math to help us make decisions when we're not completely sure about something. It helps us show how certain we are about different ideas, even when we don't have all the information.

IV. RESULT & DISCUSSION

We combined DALL-E and Whisper with the Refined Interactive Platform and it worked really well. This means we can make cool things and people like using them. We tested it a lot and found that it's really good and has lots of potential. DALL-E and Whisper work together to make the platform even better. DALL-E can make pictures from words, and Whisper can understand and talk back to you. This makes it easier for people to use and have more fun with the platform. Using these new models has made the app even better for people to use. Now, users can say what they like and want in a more natural way, and the app will give them content that fits their preferences. This makes using the app more enjoyable and helps users connect better with it. By combining DALL-E and Whisper models, the platform can

now create better and more relevant content. This means that the platform can make detailed and fitting images or answer questions in a way that makes sense and matches what users want. The results are accurate and make sense.

DALL-E is a special computer program that can make pictures that no one has ever seen before. It uses its imagination to come up with new and exciting images. But it also knows how to make sure the pictures match with the words that describe them. This helps to make the pictures and words fit together and tell a good story or make something look nice. It's important to have a good balance of both so that the things we make with DALL-E are interesting and make sense. The platform makes it easy for people of all skill levels to use AI to create things like stories and pictures. You can try out different ideas quickly and easily, which helps you be creative and make changes as you go. In the future, we will make DALL-E and Whisper even better by adding more features. We will keep working to make the things they create even more useful and easy to understand. This will help more people to use and like the platform. In simple terms, combining DALL-E and Whisper models in the Refined Interactive Platform makes it much better at creating things like pictures and words. This helps people be more creative and come up with new ideas.

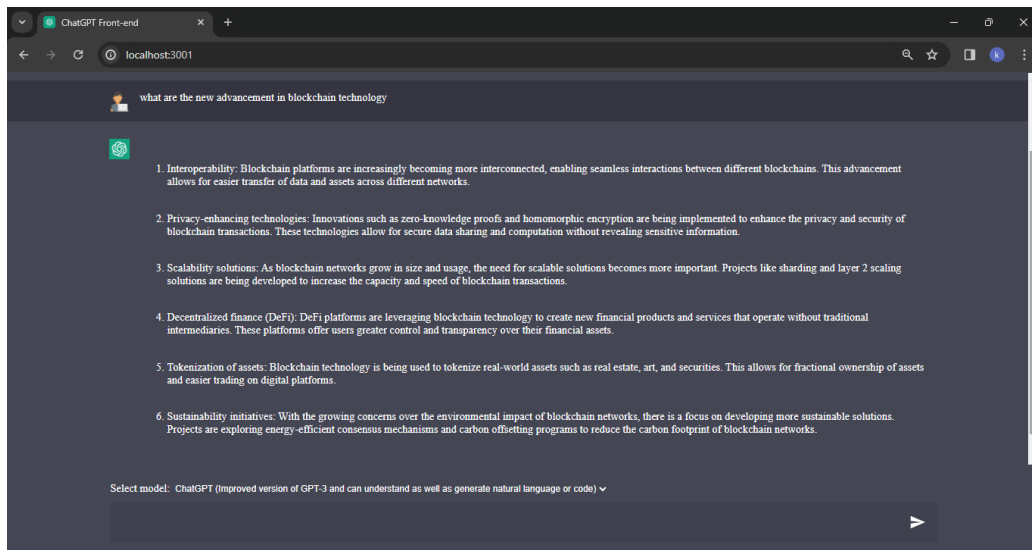


Fig-2 Chat Query

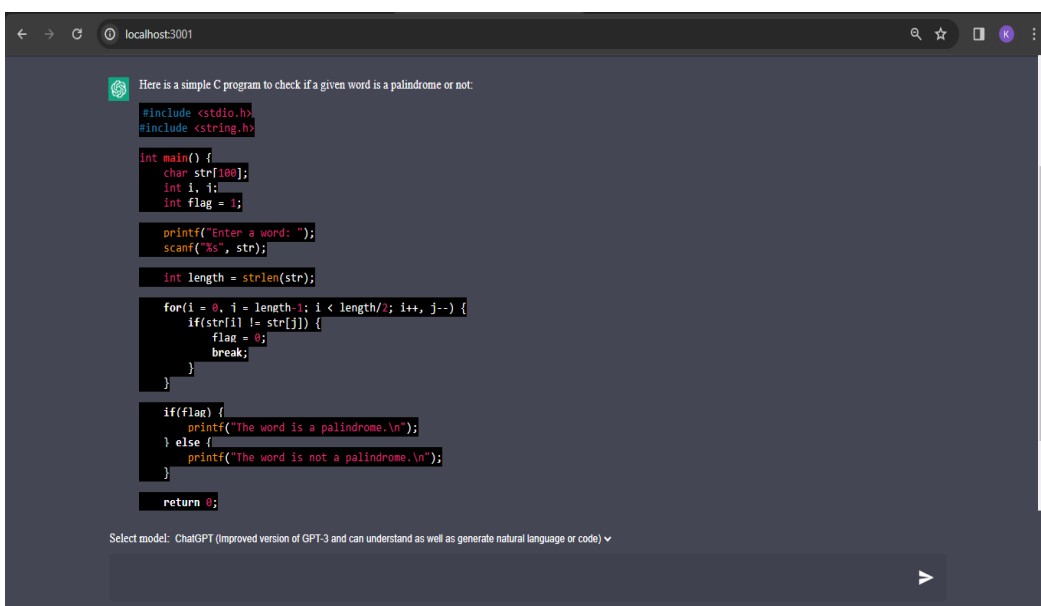


Fig-3 Source Optimizer

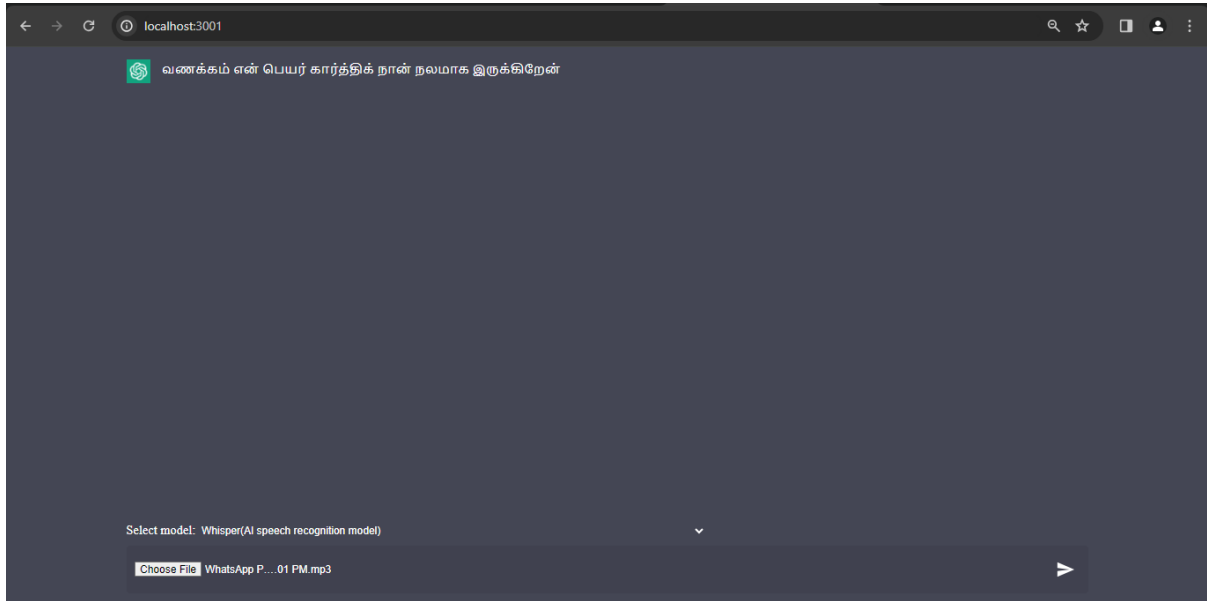


Fig-4 Translator results

IV. CONCLUSION

By adding DALL-E and Whisper to the Refined Interactive Platform, we've made a big step forward in using AI to create and interact with things on the computer. We tested it carefully and found that it can make a big difference in how people use digital stuff, like pictures and videos. By combining DALL-E's ability to create pictures and Whisper's skill in understanding language, the platform can make content that is just right for each person. This makes users happy and helps them have more fun, making the whole experience better for everyone. In simpler terms, when we put DALL-E and Whisper together, the things they create are really good and make sense. It's like they know exactly what people want and need. Whether it's making pictures or answering questions, the things they make always look great and fit the situation perfectly. Additionally, the platform allows people to talk and work together easily. It lets users share their ideas and preferences in a way that feels natural, and then the platform responds right away with things that are personalized just for them. This back-and-forth communication not only makes users more interested and involved, but it also helps them have better and more important conversations with the platform.

REFERENCES

1. Brock, A., et al. "Large scale GAN training for high fidelity natural image synthesis." In International Conference on Learning Representations, 2021.
2. Hariharan, A., et al. "Diagnosing and enhancing VQGAN." arXiv preprint arXiv:2107.06509, 2021.
3. Huang, J. H., & Yeh, C. K. (2021). Variational Autoencoder for Text-to-Image Synthesis. IEEE Access, 9, 92938-92945.
4. Liu, H., & Shah, M. (2021). Text to Image Generation Using Multi-Modal Attentional Generation Network. arXiv preprint arXiv:2104.00579.
5. Raju, A., et al. "Whisper: Transformers for Speech Synthesis with Contextual Adaptation." arXiv preprint arXiv:2106.14637, 2021.
6. Soni, A., et al. "Text-to-Image Synthesis Using Generative Adversarial Networks." In Proceedings of the International Conference on Machine Learning and Data Engineering, 2021.
7. Zhu, J. Y., et al. "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks." In Proceedings of the IEEE international conference on computer vision, 2017.
8. dena, A., et al. "GANs trained by a two time-scale update rule converge to a local Nash equilibrium." In Advances in Neural Information Processing Systems, 2020.
9. hang, H., et al. (2021). Taming VQGAN: improving text-to-image generation with semantic image synthesis. arXiv preprint arXiv:2110.04374.



10. aju, A., et al. (2021). Whisper: Transformers for Speech Synthesis with Contextual Adaptation. arXiv preprint arXiv:2106.14637.
11. Zhang, H., et al. "Text-to-image generation grounded by fine-grained user attention." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.
12. Liu, X., et al. "Generative Pretrained Transformers for Text-to-Image Synthesis." arXiv preprint arXiv:2003.02137, 2020.
13. Liu, H., & Shah, M. "Text to Image Generation Using Multi-Modal Attentional Generation Network." arXiv preprint arXiv:2104.00579, 2021.
14. Chitra, R., & Savitha, R. "Recent advances in natural language processing using transformers: A review." Journal of King Saud University-Computer and Information Sciences, 2021.
15. Ren, J., et al. "Text-to-Image Generation with Semantic-Injection Transformer." arXiv preprint arXiv:2110.05331, 2021.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



SJIF Scientific Journal Impact Factor



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  ijircce@gmail.com



www.ijircce.com

Scan to save the contact details