

Using artificial intelligence to generate real-time augmented reality content in book publishing

VOLODYMYR YAKYMIV

Department of Applied Mathematics & physics
Faculty of Computer Printing Engineering
Ukrainian Academy of Printing
Pidgolosko St., 19, L'viv, 79020
UKRAINE

LIDIYA PISKOZUB

Department of Media Communications
Faculty of Media Communications & Entrepreneurship
Ukrainian Academy of Printing
Pidgolosko St., 19, L'viv, 79020
UKRAINE

YOSYF PISKOZUB

Department of Applied Mathematics & physics
Faculty of Computer Printing Engineering
Ukrainian Academy of Printing
Pidgolosko St., 19, L'viv, 79020
UKRAINE

Abstract: - Described as augmented and virtual reality as well as artificial intelligence are increasingly being used in various areas of electronic and paper publishing. The main problems with the development of augmented reality applications for book publishing are characterized. An approach is proposed for using artificial intelligence to generate images for display in augmented reality applications for book publications. An algorithm for searching for keywords in the text to use them as input data for artificial intelligence image generation services is also proposed. This study also provides the idea and problems of validation and quality control of generated data.

Keywords: - artificial intelligence, augmented reality, printing, book publishing

Received: May 22, 2022. Revised: March 3, 2023. Accepted: April 2, 2023. Published: April 28, 2023.

1 Introduction

Books are great sources of knowledge and entertainment, but they can sometimes feel outdated compared to the latest technology. But what if technology could help make reading books even more interesting and interactive? Augmented Reality (AR) and Virtual Reality (VR) have the potential to change the way readers experience and interact with books. This is a new direction in the book industry that allows for the creation of new forms of book content and increased engagement with reading among the audience.

The main advantage of using VR and AR in book products is that it makes reading more interactive and engaging for readers, providing new ways to interact with content and opportunities to deepen knowledge and emotional experience. There is a considerable amount of research and work that proves the positive

result of using this technology in education [1,2] or children's books [3].

Most of the existing publications that use AR have additionally developed applications that are necessary for the actual display of augmented reality elements.

This article offers a view on how to change and improve the process of generating content for real-time AR technology for use in print or electronic publications.

2 Problem statement

As virtual or augmented reality technology is increasingly developed and spread, more and more problems arise with the process of writing applications, and developing 3D animations, graphics, and interface elements that need to be developed to use this approach in at least one print or electronic publication. In other words, there is a

problem with the wider spread and use of this technology in printing, print, or electronic media.

There are more and more different tools, libraries, frameworks, or toolkits in the space that are designed to simplify the development and use of this technology for wider distribution and integration into publications. These tools, while simplifying the process, still leave the need for additional development of data and materials for use in the publication. The problem remains in repeating this stage of developing 3D materials and elements for each edition, i.e., constant new content generation.

All this is reflected in the final price of the publication and the time required to develop these materials. In some cases, the development of a large number of materials and 3D scenes may require more time and money than the cost of the publication and printing itself.

2.1 Implementation of artificial intelligence

It is this problem that the use of artificial intelligence (AI) to generate 3D content in real time should help solve.

Despite the significant number of works on the use and development of AR technologies in combination with artificial intelligence, most of them are concerned with improving the integration of this process or improving the result, and most of the works are concerned with the general use of AI to improve the integration of technology [4,5,6] or the use of AI and AR in various fields, such as tourism [7] or medicine [8,9].

There have also been studies that show how imperfect the services that generate content based on the keywords provided are at the moment [10]. However, it is worth noting that these services are at the initial stage of development, and the neural networks and machine learning used for them require time and more materials to improve.

3 Solving the problem

3.1 Overview of AI-based tools

Artificial intelligence is the ability of an engineering system to process, apply, and improve the acquired knowledge and skills [11].

AI is based on algorithms and methods that allow a computer to acquire knowledge, use it to solve problems and improve over time. Artificial intelligence can include the following technologies:

- machine learning;
- deep learning;
- neural networks;
- natural language processing;
- image recognition;

- data processing.

At the moment, there is already a fairly significant amount of software that allows generating content using artificial intelligence. Such software can be divided into the following groups, which differ in the type of input and output data [11]:

- text in GIF (text-to-gif - T2G);
- text to 3D (text-to-3D - T2D)
- text to text (T2T);
- text to NFT (text-to-NFT - T2N);
- text-to-code (T2C);
- text to image (T2I);
- text to audio (text-to-audio - T2S);
- text to video (text-to-video - T2V);
- text to music (T2M);
- text-to-motion (T2Mo);
- audio to text (A2T);
- audio to audio (A2A);
- image-to-text (I2T).

From the above data types, we can conclude that the main type of data input for AI is text. With text, you can provide artificial intelligence with keywords to process to get the desired result.

For this study, examples of tools and algorithms for generating images and video materials that can be used to generate images in augmented reality using text as the main type of data input will be considered.

The following software exists to generate such content [12]:

- to generate images from the text: DALL-E 2, Stable Diffusion, Craiyon, Jasper, Imagen, Midjourney, GauGAN2, Wombo, Wonder, pixray-text2image, neural.love;
- to generate video from the text: runway, Fliki, Synthesia, Meta AI, Google AI, Phenaki;
- to generate audio from text: Play.ht, Murf.AI, RESEMBLE.AI, WELLSAID, descript.
- real-time audio generation AIVA, Jukebox [13,14].

3.2 Using artificial intelligence for content generation

Thus, the key task of artificial intelligence in this case, will be to generate images, video, or audio derived from the text of the book or general data of the publication.

To implement the use of artificial intelligence, it will still be necessary to develop a mobile application, but in this case, it will be rather universal.

The application offers users two options for defining keywords for artificial intelligence generation:

- choose keywords by themselves. The advantages of this method are that users can

choose the right keywords, such as character names or locations, without having to analyze the text by application; the user can choose one keyword that will be easier to process, remember, and display when hovering over the same word (such as a character's name);

- analysis of the page text by application. This method is less reliable because it requires a separate algorithm to select keywords from the text. These words may not have any meaning or be directly related to the story of the publication but may be, for example, references to other materials, which will complicate the content generation process and make it less reliable. However, once this algorithm is improved, and possibly with the use of an artificial intelligence algorithm in the future, the automatic keyword search will become more accurate and will be able to provide a better description of the scene and characters in the story.

The main tasks of this application are:

- entering basic initial data of the publication, such as title and author;
- scanning the page to find keywords for a content generation or selecting keywords by the user;
- sending the obtained keywords along with the title of the publication (for better identification) to the APIs of several services to generate source materials;
- displaying the obtained data above the page or selected text on the page as elements of augmented graphics.

3.3 How the algorithm works to automatically select keywords from the text

Artificial intelligence services cannot properly analyze the entire text of a page or a single paragraph, as such text may not contain any semantic information about what is happening in the selected text, for example, due to the absence of character names. In this case, it will be difficult for artificial intelligence services to generate the correct content that will relate to the selected publication.

To successfully use artificial intelligence services, an algorithm is needed that can independently select keywords from the scanned text.

The algorithm for searching for keywords in the text consists of next steps:

1. Form and separate paragraphs for analysis;
2. Filter out paragraphs that do not contain proper names (names, place names, etc.) or have very little text;

3. Conduct the first analysis to select proper names from individual paragraphs and save the samples in separate arrays if such samples are found;

4. Conduct a second paragraph analysis to search for adjectives in the filtered paragraphs, if they mention proper names, to form descriptions of the characters, location, weather, etc. that were selected in the previous step.

The keywords obtained after the algorithm are ready to be sent to the APIs of artificial intelligence services to generate certain content (for example, images).

3.4 Obtaining images from AI image generation services

To generate images, an AI service needs to support data generation requests using an API interface. Among the listed and existing services, DALL-E 2 and Stable Diffusion systems have such an interface.

To send an HTTP GET or POST request to these services (depending on the service) with a list of selected keywords, you need to receive the generated image. The resulting image is displayed on the application screen in the form of augmented reality above the text of the book.

The application may contain a special cursor, which, when hovered over, will consider the word selected by the cursor as the keyword to be used to send to the artificial intelligence service to obtain the desired result. The disadvantage of this option is that in this way you can choose any word that is not meaningful enough and does not contain the semantic load to get the correct result from the AI service. Therefore, in this case, this option should contain a filter and allow you to select, for example, only your own words that have a semantic load and should ensure that AI generates the correct result.

Practice and examples have shown that using only the title of the work or characters is not enough, and the result itself is incorrect or of inadequate quality. To improve the result, you need to add some words to describe the style of the desired image.

For example, by adding "3D render" to the list of keywords, you can change the style of the resulting images, which will be significantly different and visually better perceived (**Fig. 1, Fig. 2, Fig. 3, Fig. 4, Fig. 5**).

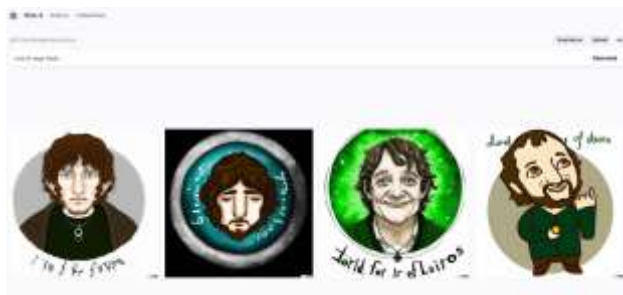


Figure 1. Example of generated images for the query "Lord of the rings Frodo" using the DALL-E 2 system

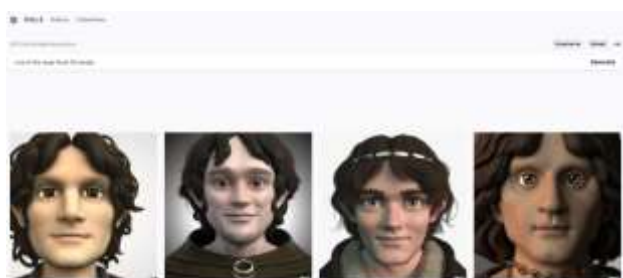


Figure 2. An example of generated images for the query "Lord of the rings Frodo 3D render" using the system DALL-E 2.



Figure 3. Example of generated images by request "white fang wolf near bonfire book illustration" using the Stable Diffusion system



Figure 4. An example of a generated image based on the keyword query "dobby from Harry Potter for kids" using the Stable Diffusion system



Figure 5. An example of a generated image based on the keyword query "beautiful cinderella in ghibli style" to demonstrate the combination of styles using the Stable Diffusion system

4 Conclusions

The proposed model of using artificial intelligence to generate content can significantly simplify the process of developing augmented reality applications for book publications, which makes it possible to reduce the cost of using such applications for a larger number of publications, as well as to use such software as a universal one, without the need to develop separate mobile tools for individual editions. The proposed model for searching for keywords in the text also makes this process automatic and easier for users. Further development of the recommended model should include the introduction of a fully automatic method of generating book content with relevance and accuracy to the content of the

publication, as well as a methodology for quality control and correctness of the generated images.

References:

- [1] Mehta, Soham & Jain, Pratish & Vora, Aayushi & Joshi, Abhijit & Dalvi, Harshal. *Augmented Reality Books: An Immersive Approach to Learning*. 2017.
- [2] Farzeeha, Dayana & Omar, Marlissa & Abdullah, Abdul & Ibrahim, Nor & Mokhtar, Mahani & Zaid, Norasykin & Joe, Ila. *5 Years into Augmented Reality Technology in Education: Research Trends, Bibliometric Study and its Application to Enhance Visualization Skills*. WSEAS TRANSACTIONS ON SYSTEMS AND CONTROL. 16. 253-260. 2021 10.37394/23203.2021.16.21.
- [3] Dong, Chunxia & Si, Zhanjun.. *The Research and Application of Augmented Reality in 3D Interactive Books for Children*. 2018. 10.1007/978-981-10-7629-9_35.
- [4] Sahu, Chandan Kumar & Young, Crystal & Rai, Rahul. *Artificial intelligence (AI) in augmented reality (AR)-assisted manufacturing applications: a review*. International Journal of Production Research. 59. 1-57. 2020. 10.1080/00207543.2020.1859636.
- [5] Zhabokrytskyi, I.V. *Increasing the Efficiency of Creating Augmented Reality Scenes Using Neural Networks*. Elektronnoe modelirovanie. 2023. 44. 69-85. 10.15407/emodel.44.06.069.
- [6] Devagiri, Jeevan & Sidike, Paheding & Niyaz, Quamar & Yang, Xiaoli & Smith, Samantha. *Augmented Reality and Artificial Intelligence in industry: Trends, tools, and future challenges*. Expert Systems with Applications. 2022. 207. 118002. 10.1016/j.eswa.2022.118002.
- [7] Tsepapadakis, Michalis & Gavalas, Damianos & Koutsabasis, Panayiotis. *3D Audio + Augmented Reality + AI Chatbots + IoT: An Immersive Conversational Cultural Guide*. 2022. 10.1007/978-3-031-15546-8_22.
- [8] Gandedkar, Narayan & Wong, Matthew & Darendeliler, M. *Role of Virtual Reality (VR), Augmented Reality (AR) and Artificial Intelligence (AI) in tertiary education and research of orthodontics: An insight*. 2021. Seminars in Orthodontics. 27. 10.1053/j.sodo.2021.05.003.
- [9] Ende, Elizabeth & Ryan, Sean & Crain, Matthew & Makary, Mina. *Artificial Intelligence, Augmented Reality, and Virtual Reality Advances and Applications in Interventional Radiology*. 2023 Diagnostics. 13. 892. 10.3390/diagnostics13050892.
- [10] Leivada, Evelina & Murphy, Elliot & Marcus, Gary. *DALL-E 2 Fails to Reliably Capture Common Syntactic Processes*. 2022 10.48550/arXiv.2210.12889.
- [11] *Information technology - Artificial intelligence - Overview of trustworthiness in artificial intelligence*. International Organization for Standardization and International Electrotechnical Commission. May 2020. ISO/IEC TR 24028:2020
- [12] <https://twitter.com/ethersiim/status/1586824269543112704>
- [13] A neural net that generates music, including rudimentary singing, as raw audio in a variety of genres and artistic styles. <https://openai.com/research/jukebox>.
- [14] AIVA - The AI composing emotional soundtrack music. <https://www.aiva.ai/>

Contribution of individual authors to the creation of a scientific article (ghost-writing policy)

Author Contributions:

Volodymyr Yakymiv carried out the resource search, methodology, research, and algorithmizing. Lidiia Piskozub organized and carried out research, writing, and editing. Yosyf Piskozub supervised and administered the project.