Research Article

Analysis of Factors Affecting Common Use of Generative Artificial Intelligence-Based Tools by Machine Learning Methods

Yasin KIRELLİ*

Kütahya Dumlupınar University, Faculty of Applied Sciences, Management Information Systems, Kütahya-Türkiye
* **Corresponding Author**: **Email:** yasin.kirelli@dpu.edu.tr **ORCID:** 0000-0002-3605-8621

Abstract:

Artificial Intelligence is a sub-branch of artificial intelligence used to produce new data or content. These methods can create recent examples in different categorical fields such as natural language processing, image processing, music, and video creation by using models from learning clusters with artificial intelligence (AI) tools in this field. AI tools that can solve real-world problems are also created using different methods apart from generative AI methods. With generative-based artificial intelligence tools, it can facilitate people's work in jobs that require creativity. However, they can offer the opportunity to build advanced models that learn from data with other artificial intelligence methods. In the study, the public dataset has been used. This dataset includes trending artificial intelligence tools, AI methods, and user scores. In this study the working area and user trend of the ai tools in the dataset and the effect of generative AI methods on the development of the tool are discussed. Random Forest and Naive Bayes algorithms from classification methods have been used to measure the impact and estimation. Several AI tools help solve real-life problems. Identifying what type of category is needed for AI tools and method selection are interlinked, and the research provides an overview of this connection.

1. Introduction

Artificial intelligence (AI) is the ability to create applications or tools with computer programs that aim to think like humans, similar to human intelligence. The field of AI was studied in 1956 by a computer scientist at Dartmouth College. They have been based on the identifiability assumption of intelligence and aimed to make the simulation a machine. Thus, it seeks to enable people to solve problems specific to humans and learn from data [1–4]. The primary purpose is to solve the issues that people can learn or solve with their cognitive intelligence with the help of machines or applications. These problems can be in different categories, such as language comprehension, music creation, image recognition, and decision support systems. The human ability in these categories is to be able to meet similar. The primary purpose of artificial intelligence tools is to create advanced systems with cognitive intelligence in many fields,

such as health, education, entertainment, image, video, and the military. Artificial intelligence systems, which can interpret human input with applications such as chatbots and convert it from text to text or from text to picture with natural language processing techniques, serve effectively in many areas today. It has become a personal assistant tool that can help in different areas, such as customer service, search engines, and code generators [5]. As data-based learning systems depend on statistical data, artificial intelligence tools have developed rapidly. Two main reasons have been influential in these tools' widespread use and development. The first of these is the general use of data storage systems and data sets, and the other reason is the increasing computing power of computers. With the deep learning sub-artificial intelligence field, powerful personal assistants such as Siri and Alexa were introduced in 2010. Today, ChatGPT has reached widespread use as an advanced personal

assistant and a language processing tool offered by OpenAI [1].

The deep learning technique is a structure created by combining forms of artificial intelligence with machine learning models [6–8]. Unlike neural networks, which use only one-layer, deep learning structures use multiple layers. Thus, a better estimation model can be made. Deep learning is defined as a multi-layered neural network structure. This neural network transfers information to the last layer in a non-linear form between layers, which is possible with activation functions. In the convolutional form, activation functions take place as parameters between layers in the transfer of information and are used to disrupt linearity. Thus, the deep learning structure will achieve better learning ability. Minimizing the loss value by forward and backward propagation is the main goal [1,9].

The Generative Adversarial Networks (GAN) architecture method is among the artificial intelligence methods that attract great attention today. It is a deep learning method frequently preferred in natural language processing, image, and video processing with artificial intelligence tools created with this method. Creating powerful personal assistant applications such as healthcare or ChatGPT, such as the production of new drugs, is a promising method [10–13]. Responsible use of artificial intelligence is among the discussion topics that come with it. Presenting the text and images produced by AI as an 'original work' is also a research topic [14].

In the study, the artificial intelligence tools dataset in the publicly available Kaggle database is discussed. Today, widely used artificial intelligence tools are listed in this dataset. Extra details, such as usage areas and development methods, are given. It has been tried to analyze the effect of generative artificial intelligence methods or other artificial intelligence methods on the widespread use of the tools in the study, depending on the subject area targeted by an artificial intelligence tool. Random forest, support vector machine, and logistic regression from machine learning techniques have been used for estimation. Evaluation metrics and

analysis are also has been discussed in the following sections.

2. Material and Methods

In this chapter, GAN, which is the most preferred method for the development of ai tools, is explained. The following sections give an exploratory analysis of the data set used. In the last section, the evaluation of the model is given.

2.1. Generative Adversarial Networks Architecture

According to the Generative Adversarial Networks architecture, it is a deep learning method in the convolutional artificial neural network structure. Goodfellow et al. proposed this in 2014 [15,16]. It has been frequently used and preferred as a deep learning method. This method uses real data as input and learns the sample distribution. This produces outputs similar to real data [17]. It is an effective method to create tools such as ChatGPT by giving outcomes above human capabilities [18]. This method, preferred in data creation, is also essential for data imbalance problems. With the development of deep neural network architectures, variants of the GAN architecture have also emerged. It has been proposed as a preferred variant for unsupervised learning (DCGAN), for increased stability (WGAN), and poor sample quality problems (WGAN-GP) [19]. As seen in Figure 1, there is a producer (G) and a separator (D) object. The GAN method is presented in Equation 1. According to the formula, the goal is for Pda to converge to Pg with a gradient.
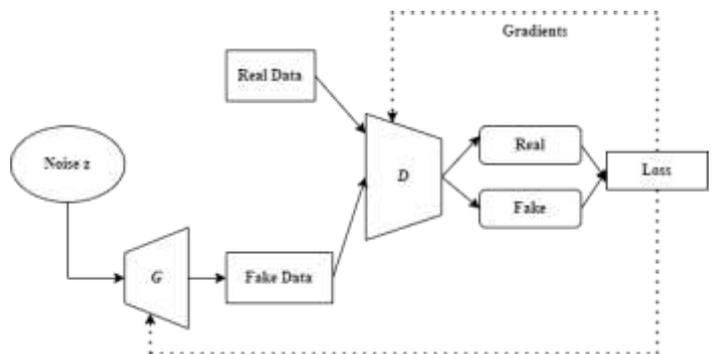
Figure 1. *GAN Architecture*

According to the formula, G Generator, D stands for the discriminator. Pz distribution of the random noise; Pda represents the real sample distribution. D(x) is the discriminant probability value, and G(z) is the sample output.

$$L(G,D) = \min_{G} \max_{D} E_{x \sim P_{da}}[log D(x)] \quad (1)$$
$$+ E_{z \sim P_z}[log(1 - D(G(z)))]$$

According to Equation 1, when Pda = Pg, it is assumed that the generated data distribution is equal to the actual data distribution [19].

2.2. Exploratory Data Analysis (EDA)

The dataset contains publicly available AI tools and information from the Kaggle database. This dataset is currently widely used and is a resource for discovering new tools [20]. The dataset includes the target subject area, user scores and ai methods information of these tools. The data set consists of 1012 rows and sample 5 records are shown in Figure 2.

Figure 2. *Data set sample records*

According to the records in the data set and Figure 3., the most AI tools are in 110 Copywriting, 91 Generative Art and 84 Chat subject areas. According to the results, considering the use of artificial intelligence tools by individual users, personal assistants that assist in writing texts are widely used. It is seen that multimedia solutions are the next most used area. Figure 4 shows the most preferred ai tool categories and the applied methods according to user comments. According to the graph, GAN architecture has been mostly preferred as a method in developing AI tools. In Figure 5, if the AI tools are sorted according to user preferences, it is seen that "LeiaPix," "GPT-3," and "Codeium" have the most votes. All three tools have been developed using the artificial intelligence method. In addition, these tools are in the "LeiaPix" image improvement category, "GPT-3" copywriting, and "Codeium" generative code category, respectively. If it is

considered a personal assistant tool, it attracts attention from users. Figure 6 shows each ai tool in the dataset and a word cloud according to the description field and word frequencies in the dataset. It can be seen that the emphasis is placed on 'generative artificial intelligence' and 'personal assistant' in the word cloud.

2.3 Implementation and Evaluation of the Model

The dataset's 'ai tool development technique' has been chosen as the estimation or dependent variable.
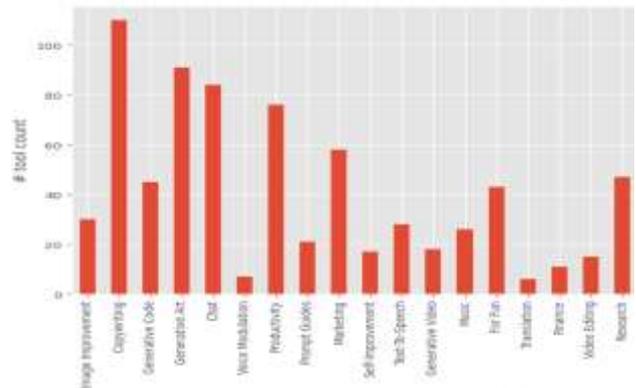
Figure 3. *Category and usage rates*
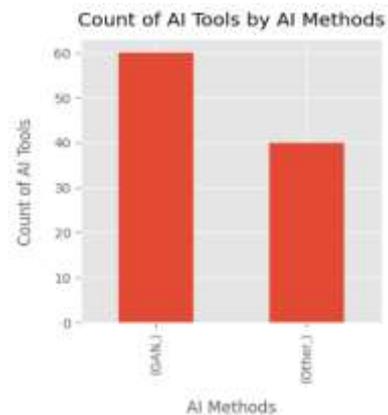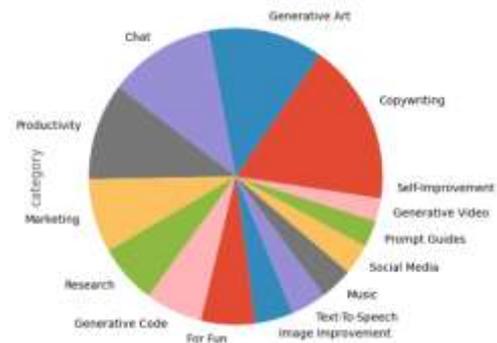
Figure 4. *Category and AI Methods for Top 100 AI Tools*
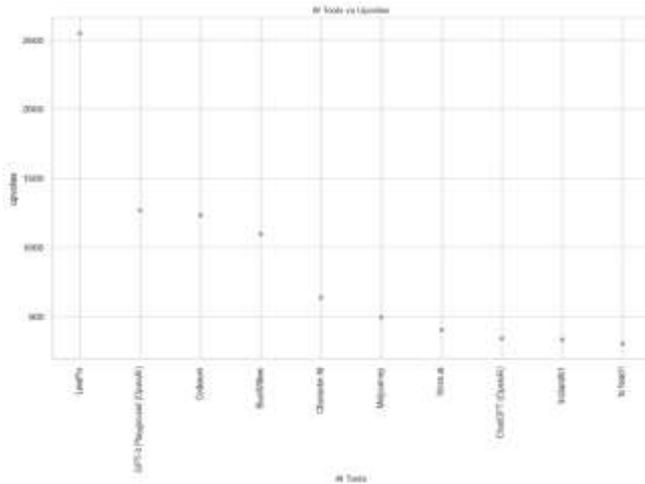
Figure 5. *Trends of AI tools by user ratings*

Figure 6. *AI Tools description field word frequencies*

The dependent variable 'ai tool development technique' consists of two categorical values: 'GAN' or 'Other Methods.' The independent variables are the upvotes of the users and the categorical area addressed by the artificial intelligence tool. While creating the prediction model, Python programming version 3.10 and the Sklearn library have been used. Logistic Regression, Random Forest, and Support Vector Machine machine learning techniques have been used for the prediction model, respectively.

Table 1. *Support Vector Machine Model Evaluation Metrics*

	precision	recall	f1-score
GAN	0.68	0.93	0.79
Other Methods	0.83	0.45	0.59
Accuracy			0.72
Macro Avg	0.76	0.69	0.69
Weighted Avg	0.75	0.72	0.70

Table 2. *Random Forest Model Evaluation Metrics*

	precision	recall	f1-score
GAN	0.97	0.95	0.96
Other Methods	0.95	0.97	0.96
Accuracy			0.96
Macro Avg	0.96	0.96	0.96
Weighted Avg	0.96	0.96	0.96

Table 3. *Logistic Regression Model Evaluation Metrics*

	precision	recall	f1-score
GAN	0.79	0.79	0.79
Other Methods	0.73	0.73	0.73
Accuracy			0.76
Macro Avg	0.76	0.76	0.76
Weighted Avg	0.76	0.76	0.76

The accuracy scores of the evaluation and criteria are presented in Table 1, Table 2, and Table 3. According to the models created, it is seen that the highest accuracy rate in the analysis has been obtained with the Random Forest model at 96%. It is seen that the next closest and with the highest metric ratio is Logistic Regression. While creating the models, hyperparameter tuning has been performed in order to select the most suitable model parameters for the data set. Thus, it can be suitable for the data set, and the estimation rate is increased.

3. Conclusion

Generative artificial intelligence is presented to produce a response, unlike other techniques. It can create outputs similar to human intelligence beyond just generating answers but also go beyond that. AI tools such as ChatGPT are among the popular tools preferred by users in this study as an example of generative intelligence that includes different methods rather than just producing natural language processing output. Due to the increased machine capacities and data processing speeds today, they have become personal assistants in daily life thanks to successful AI tools. According to the study, since artificial intelligence tools created using the generative artificial intelligence technique are preferred as personal assistant services for individual users, using these tools has become more widespread with the users' upvotes. The primary purpose is to use the information provided by generative artificial intelligence tools for ethical and goodwill purposes by their users. Therefore, ethical principles and bias in the effective widespread use of artificial intelligence are another discussion topic. For this reason, it is a different need to confirm the accuracy of the information produced by generative artificial intelligence tools.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.

- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] A. Holzinger, K. Keiblinger, P. Holub, K. Zatloukal, and H. Müller, (2023). AI for life: Trends in artificial intelligence for biotechnology," *N Biotechnol,* 74;16–24 doi: 10.1016/j.nbt.2023.02.001.

[2] N. Eslamirad, F. De Luca, K. Sakari Lylykangas, and S. Ben Yahia, (2023). Data generative machine learning model for the assessment of outdoor thermal and wind comfort in a northern urban environment, *Frontiers of Architectural Research,* doi: 10.1016/j.foar.2022.12.001.

[3] V. Couteaux et al., (2023). Synthetic MR image generation of macrotrabecular-massive hepatocellular carcinoma using generative adversarial networks, *Diagn Interv Imaging,* doi: 10.1016/j.diii.2023.01.003.

[4] H. Woldesellasse and S. Tesfamariam, (2022). Data Augmentation Using conditional Generative Adversarial Network (cGAN): Application for Prediction of Corrosion Pit Depth and Testing Using Neural Network, *Journal of Pipeline Science and Engineering, p. 100091* doi: 10.1016/j.jpse.2022.100091.

[5] S. O'Connor and ChatGPT, (2023). Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse?, *Nurse Education in Practice,* vol. 66. Elsevier Ltd, doi: 10.1016/j.nepr.2022.103537.

[6] T. Ching et al., (2018). Opportunities and obstacles for deep learning in biology and medicine," *J R Soc Interface,* 15;141 doi: 10.1098/RSIF.2017.0387.

[7] D. Dana et al., (2018). Deep Learning in Drug Discovery and Medicine; Scratching the Surface, *Molecules,* 23;9 doi: 10.3390/molecules23092384.

[8] E. Lin, P. H. Kuo, Y. L. Liu, Y. W. Y. Yu, A. C. Yang, and S. J. Tsai, (2018). A Deep Learning Approach for Predicting Antidepressant Response in Major Depression Using Clinical and Genetic Biomarkers. *Front Psychiatry,* 9 doi: 10.3389/FPSYT.2018.00290.

[9] Y. Chauvin and D. E. Rumelhart, Backpropagation : theory, architectures, and applications. Lawrence Erlbaum Associates, 1995. Accessed: Mar. 13, 2023. [Online]. Available: https://www.routledge.com/Backpropagation-Theory-Architectures-and-Applications/Chauvin-Rumelhart/p/book/9780805812596

[10] H. Zhao, H. Li, S. Maurer-Stroh, and L. Cheng, (2018). Synthesizing retinal and neuronal images with generative adversarial nets. *Med Image Anal*, 49;14–26, doi: 10.1016/j.media.2018.07.001.

[11] B. Hu, Y. Tang, E. I.-C. Chang, Y. Fan, M. Lai, and Y. Xu, (2017). Unsupervised Learning for Cell-level Visual Representation in Histopathology Images with Generative *Adversarial Networks.* doi: 10.1109/JBHI.2018.2852639.

[12] M. Mardani et al., (2019). Deep Generative Adversarial Neural Networks for Compressive Sensing MRI. *IEEE Trans Med Imaging*, 38(1);167–179, doi: 10.1109/TMI.2018.2858752.

[13] E. Lin, S. Mukherjee, and S. Kannan, (2020). A deep adversarial variational autoencoder model for dimensionality reduction in single-cell RNA sequencing analysis. *BMC Bioinformatics*, 21(1);64, doi: 10.1186/s12859-020-3401-5.

[14] D. O. Eke, (2023). ChatGPT and the rise of generative AI: Threat to academic integrity?. *Journal of Responsible Technology,* 13;100060, doi: 10.1016/j.jrt.2023.100060.

[15] C. Kruchko, Q. T. Ostrom, H. Gittleman, and J. S. Barnholtz-Sloan, (2018). The CBTRUS story: providing accurate population-based statistics on brain and other central nervous system tumors for everyone. *Neuro Oncol,* 20(3);295–298, doi: 10.1093/neuonc/noy006.

[16] I. J. Goodfellow et al., "Generative Adversarial Networks," Jun. 2014.

[17] R. Ranjbarzadeh, A. Caputo, E. B. Tirkolaee, S. Jafarzadeh Ghoushchi, and M. Bendechache, (2023). Brain tumor segmentation of MRI images: A comprehensive review on the application of artificial intelligence tools," *Computers in Biology and Medicine*, 152 doi: 10.1016/j.compbiomed.2022.106405.

[18] W. M. Lim, A. Gunasekara, J. L. Pallant, J. I. Pallant, and E. Pechenkina, (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. The *International Journal of Management Education*, 21(2);100790, doi: 10.1016/j.ijme.2023.100790.

[19] S. Liu, H. Jiang, Z. Wu, Y. Liu, and K. Zhu, (2022). Machine fault diagnosis with small sample based on variational information constrained generative adversarial network. *Advanced Engineering Informatics,* 54 doi: 10.1016/j.aei.2022.101762.

[20] Cutting-Edge AI Tools: An Up-to-Date Dataset Kaggle"https://www.kaggle.com/datasets/yasirabdaali/740-ai-tools-for-everyone (accessed Mar. 13, 2023).