



DOI: 10.21005/pif.2024.60.B-07

VISUAL REPRESENTATION OF PODLASKIE ARCHITECTURE: AN EVALUATION OF GENERATIVE AI MODELS

WIZUALNA REPREZENTACJA ARCHITEKTURY PODLASIA: OCENA GENERATYWNYCH MODELI SZTUCZNEJ INTELIGENCJI

Bartosz Śliwecki

dr inż. arch.

Author's Orcid number: 0000-0002-6231-967X

AuReLa Laboratory,
Faculty of Architecture,
Białystok University of Technology, Poland

ABSTRACT

This paper addresses the underrepresentation of the Podlaskie voivodeship's architectural style in AI image generation tools. By generating images from 60 prompts featuring Podlaskie-specific terms and evaluating their alignment and quality, this study finds that SDXL models depict the region proficiently, with MidJourney achieving nearly 100% accuracy but only 70% quality. While cities are generally well-represented, small towns and villages are mostly lackluster in detail and with low accuracy. Generated visual representations consist of derelict and mismanaged structures with a generic Eastern-European housing architectural style.

Key words: diffusion models, Podlaskie, visualization, visual representation.

STRESZCZENIE

Niniejszy artykuł porusza problem niedostatecznej reprezentacji stylu architektonicznego regionu Podlaskiego w narzędziach do generowania obrazów AI. Poprzez generowanie obrazów na podstawie 60 promptów zawierających terminy specyficzne dla Podlaskiego oraz ocenę ich zgodności i jakości, badanie wykazuje, że modele SDXL dość dobrze przedstawiają ten region, podczas gdy MidJourney osiąga niemal 100% dokładności, ale tylko 70% jakości. Miasta są względnie dobrze reprezentowane, natomiast małe miasteczka i wsie są w większości przedstawiane bez charakterystycznych cech oraz ze znaczącą niedokładnością. Generowane obrazy obejmują ilustracje zdegradowanych oraz zaniedbanych budynków w stylu architektonicznym pasującym do generycznej wschodnio-europejskiej zabudowy.

Słowa kluczowe: modele dyfuzyjne, Podlaskie, reprezentacja wizualna, wizualizacja

1. INTRODUCTION

Image generation models have become a cornerstone of modern AI, capable of creating intricate and lifelike images from textual descriptions. These models are particularly significant in the field of architectural depiction, where they can visualize architectural styles, historical reconstructions, and urban planning concepts with remarkable precision. Models such as those built on the Stable Diffusion architecture and MidJourney leverage extensive datasets comprising millions of images comprising millions of human-captured and crafted images (Yang, K.; Qinami, K.; Fei-Fei, L.; Deng, J.; Russakovsky, O. 2020), allowing them to generate detailed and contextually accurate imagery.

Proper representation of a given region, such as the Podlaskie regional architecture, in AI-generated imagery is crucial for several reasons. Accurate depiction in both simple and advanced prompts ensures that the unique cultural and architectural heritage of the region is preserved and promoted. This not only aids in educational and preservation efforts but also helps to foster a greater appreciation and understanding of the region's distinctive characteristics. Moreover, for architects and designers, having access to precise visualizations of Podlaskie's architecture can inspire new projects and integrate traditional elements into modern designs (Sliwecki, B.; Jakimowicz, A.; Andrejczuk, S.; Poplawski, M. 2024). Ensuring that image generation models can faithfully reproduce these features allows for more meaningful and contextually relevant outputs (Fig. 1), thereby enhancing the utility and reliability of these tools across various applications.



Fig. 1. One of the outputs from the prompt "Poczopek" a small village known for its nature and Silverium wildlife sanctuary. The image however, shows a mountainside historical site with well preserved architecture. This is an example of falsification in generative image models. Source: by author

Ryc. 1 - Jeden z wygenerowanych obrazów z promptu „Poczopek”, wioski znanej ze swojego naturalnego otoczenia oraz Silverium, ogrodu leśnego z małym sanktuarium dla dzikich zwierząt. Wygenerowany obraz jednak przedstawia miejscowość górską z historyczną oryginalną zabudową. Jest to przykład falsyfikacji w sztucznej inteligencji generatywnej. Źródło: autor

Training of these models is rooted in large collections of images that reflect the diversity of human experience and creativity. Such a vast array of training images includes photographs of buildings, streetscapes, interiors, and various architectural details, along with the background data including blurred objects, landscapes, lighting setups, noise (You, Y.; Zhang, Z.; Hsieh, C. J.; Demmel, J.; Keutzer, K. 2018). The core mechanism behind these models is the neural network, which learns to associate specific patterns in the input text with visual elements in the training images. This learning process enables the generation of coherent and contextually appropriate images, making these models indispensable tools for architects, designers, and historians.

Development and refinement of image generation models are often community-driven, while basing their work on the available base models and support given by their authors. Researchers and enthusiasts contribute to the evolution of these systems by finetuning them on specialized datasets, most of the time created specifically for a given model direction. With the use of LoRAs (Low-Rank Adaptation of Large Language Models) (Hu, E. J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; Chen, W. 2021), this collaborative effort ensures that the models can cater to specific themes or styles, enhancing their utility in various applications. Fine tuning allows for greater precision in the output, which is especially valuable in fields such as content creation and concept design, where the visualization of ideas needs to be both accurate and innovative.

AI-generated imagery has revolutionized the creative industries, providing a rapid and efficient means of prototyping concepts and visualizing designs that would otherwise require substantial time and resources. These models have found applications in design, advertising, entertainment, and heritage preservation, among others. However, the effectiveness of these models is heavily dependent on the comprehensiveness of their training data. Underrepresented regions or styles in the training datasets often result in outputs that lack accuracy and detail. It is essential that image generation models are configured with specific regional and stylistic identifiers, particularly for areas with rich cultural and architectural heritage. This practice ensures that the generated images accurately reflect the unique characteristics of these locales (Ruiz, N.; Li, Y.; Jampani, V.; Pritch, Y.; Rubinstein, M.; Aberman, K. 2023).

The current paper aims to evaluate the performance of selected Stable Diffusion models and MidJourney in depicting the Podlaskie region, a culturally rich area in northeastern Poland known for its distinctive architectural style. The study focuses on assessing how well these models can generate images that faithfully represent the architectural features and overall aesthetic of the Podlaskie region. This includes examining both the accuracy of specific prompt outputs and the overall quality of the generated images. By exploring the capabilities and limitations of current AI image generation tools in capturing underrepresented regions, this research seeks to provide insights for future enhancements and more inclusive model training practices.

2. MATERIALS AND METHOD

For this study, ten generative models were selected to evaluate their performance in depicting the Podlaskie region's architectural style. The models chosen include eight from the Stable Diffusion network, one representing the most modern iteration of MidJourney, and the newly released Stable Diffusion 3. These models were selected based on their popularity, ambidextrous output quality, prior experience with the model, and high ratings on the Civit.ai website, which is an open and accessible online repository of models, know-how, LoRA models, and output images from the image generation community. It functions alongside HuggingFace, the main online repository for general purpose and single function ai models. From the mentioned list of models, one stands out from the rest in terms of popularity, but has been chosen nonetheless due to its unique architecture and approach, as the "Illuminati Diffusion" is the only representative of Stable Diffusion 2.1, which has fallen behind on development due to lack of popularity, but has retained a more experimental form of image generation.

Tab. 1. Chosen generative models and their general settings. Source: author

Tab. 1. Wybrane modele generatywne oraz ich ustawienia. Źródło: autor

No.	Model name	Sampler model	CFG scale	Steps
1.	epicrealism_naturalSinRC1VAE	UniPC	7	20
2.	ghostmix_v20	DPM++ 3M SDE Karras	7	20
3.	realisticVisionV60B1_v51VAE	DPM++ 3M SDE	5	20
4.	illuminatiDiffusionV1_v11	DPM++ 2M	3.5	20
5.	zavychromaxl_v60	DPM++ 3M SDE Exponential	3	20
6.	juggernautXL_v9Rdphoto2Lightning	DPM++ 2M Karras	3	20
7.	sd_xl_base_1.0_0.9vae	DPM++ 2M SDE Heun Karras	4.5	20
8.	stable_diffusion_3_medium	dpmpp2_2m	4.5	28
9.	MidJourney_v6	N/A	N/A	N/A

Each model was configured with optimal settings tailored for accurately generating images of the Podlaskie region. To ensure a robust and comprehensive assessment, each model was assigned 20 random seeds (Xu, K.; Zhang, L.; Shi, J. 2024). This variability in random seeds helps to account for the inherent stochastic nature of generative models, providing a broader range of outputs for evaluation. Settings were chosen individually for each models, as the same Sampler settings should not be applied to all generative models due to individual differences in the logic of noise generation, as well as Classifier-Free Guidance (CFG) settings which have a considerable impact on the output image quality and possible artifacts that are a direct result of more extreme values (Ho J.; Salimans T. 2022). Excluded from the general settings is the MidJourney V6 generative model which is based on a different set of settings, in which the most important one is the “RAW” style of generation and the stylisation value of 100, weirdness of zero, and variety equal to zero.



Fig. 2. Output for “Podlasie Regional Architecture” from the Epic Realism model. Source: by author



Fig. 3. Output for “Podlasie Regional Architecture” from the Realistic Vision model. Source: by author



Fig. 4. Output for “Podlasie Regional Architecture” from the GhostMix model. Source: by author

Ryc. 2. Wynik dla promptu “Podlasie Regional Architecture” z modelu Epic Realism. Źródło: autor

Ryc. 3. Wynik dla promptu “Podlasie Regional Architecture” z modelu Realistic Vision. Źródło: autor

Ryc. 4. Wynik dla promptu “Podlasie Regional Architecture” z modelu GhostMix. Źródło: autor



Fig. 5. Output for "Podlasie Regional Architecture" from the Juggernaut XL model. Source: by author

Ryc. 5. Wynik dla promptu "Podlasie Regional Architecture" z modelu Juggernaut XL. Źródło: autor



Fig. 6. Output for "Podlasie Regional Architecture" from the ZavychromaXL model. Source: by author

Ryc. 6. Wynik dla promptu "Podlasie Regional Architecture" z modelu ZavychromaXL. Źródło: autor



Fig. 7. Output for "Podlasie Regional Architecture" from the MidJourney model. Source: by author

Ryc. 7. Wynik dla promptu "Podlasie Regional Architecture" z modelu MidJourney v6. Źródło: autor

The study utilized 60 prompts (Tab. 2) specifically designed to capture various aspects of the Podlaskie region's architecture. These prompts were systematically divided into four even categories: general region, large cities, towns, and villages. The prompts describing the overall architectural style and environment of Podlaskie fell under the general region category (Fig. 2-7), while those focused on the architectural features of major cities within the region were categorized under large cities. Prompts related to the architectural characteristics of medium-sized towns were classified as towns, and those emphasizing the unique architectural elements of small villages were placed in the villages category. The names were modified to omit Polish symbols, as the tokenization methods of the stable diffusion network as well as MidJourney network don't work well with foreign symbols, and are meant to be used in English. Having written Polish symbols, the outputs were substantially more random and their filename suggested a missing character. Each of the 60 prompts was used to generate 20 images on each of the 9 selected models, resulting in a total of 10 800 images for evaluation. This comprehensive approach ensures a detailed assessment of each model's capabilities in generating accurate and high-quality images based on the given prompts.

Tab. 2. Chosen prompts used for generating images. Source: author

Tab. 2. Wybrane prompty stosowane do generowania obrazów. Źródło: autor

No.	Regional prompts / Region	No.	Cities / Miasta	No.	Towns / Miejscowości	No.	Villages / Wsie
1	Podlaskie	16	Białystok	31	Supraśl	46	Wojszki
2	Podlaskie city	17	Suwałki	32	Drohiczyn	47	Soce
3	Podlaskie town	18	Lomża	33	Krynki	48	Puchły
4	Podlaskie house	19	Augustów	34	Jedwabne	49	Plutycze
5	Podlaskie village	20	Bielsk Podlaski	35	Grabarka	50	Kruszyniany
6	Podlaskie region	21	Grajewo	36	Narewka	51	Grodek
7	Podlaskie regional architecture	22	Zambrow	37	Tykocin	52	Trzescianka
8	Podlaskie regional traditional architecture	23	Hajnowka	38	Knyszyn	53	Hacki
9	Land of the open shutters	24	Sokolka	39	Białowieża	54	Topilo
10	Eastern Polish house	25	Lapy	40	Piatnica	55	Koterka

No.	Regional prompts / Region	No.	Cities / Miasta	No.	Towns / Miejscowości	No.	Villages / Wsie
11	Eastern Polish cottage	26	Siemiatycze	41	Nowogrod	56	Poczopek
12	Eastern Polish traditional architecture	27	Wasilkow	42	Wizna	57	Dolistowo Stare
13	Eastern Polish borderlands architecture	28	Kolno	43	Dabrowa Bialostocka	58	Orla
14	Traditional Podlaskie house	29	Monki	44	Korycin	59	Ryboły
15	Traditional Podlaskie architecture	30	Wysokie Mazowieckie	45	Sejny	60	Burzyn

Generated images were assessed using two primary scores: alignment to the prompt and quality of the alignment of the outputs. The alignment score measures how accurately the generated image reflects the content and specifics of the given prompt on a scale of 0-20 with zero being a group of entirely random images and a score of 20 with images representing the prompted topic / keyword. Quality scores evaluate the overall aesthetic and technical quality of the generated images, including aspects such as clarity, coherence, and visual appeal, but are also influenced by the alignment factor relating to the classification of images to the level of representation of the topic (Fig. 8-10). The example featured image would be considered aligned to the prompt, and would represent a high value of the quality of alignment of 80%. This value means that the surrounding buildings and setting could be in Białystok, that it resembles the general mood and style, but is not a direct representation of any fragment of the city (Fig. 11).



Fig. 8. Output for "Wizna" from the SD3 model. This is an example of a low alignment score. Source: by author
Ryc. 8. Wynik dla promptu "Augustow" z modelu SD3. Jest to przykład niskiej jakości dopasowania. Źródło: autor



Fig. 9. Output for "Podlasie regional traditional architecture" from the illuminatiDiffusion model. This is an example of a fair alignment score. Source: by author
Ryc. 9. Wynik dla promptu "Podlasie regional traditional architecture" z modelu ZavychromaXL. Jest to przykład średniej jakości dopasowania. Źródło: autor



Fig. 10. Output for "Podlasie Region" from the MidJourney model. This is an example of a high alignment score. Source: by author
Ryc. 10. Wynik dla promptu "Podlasie Region" z modelu MidJourney v6. Jest to przykład wysokiej jakości dopasowania. Źródło: autor



Fig. 11. One of the outputs from the prompt “Białystok” showing a female as a main focus, and the sought out architecture as a background feature. Source: by author
 Ryc. 11 - Jeden z wygenerowanych obrazów z promptu „Białystok” przedstawiające kobietę jako główny element obrazu oraz oczekiwaną architekturę jako element tła. Źródło: autor

The results of the image evaluations will be visualized using charts to provide a clear and concise representation of the models' performances. Mentioned visualizations will facilitate a comparative analysis across different models and prompt categories. The results will be discussed in the context of initial assumptions and educated guesses regarding the performance of each model. Potential reasons for the observed outcomes will be explored, including the impact of training data, model architecture, and specific settings used in the generation process.

By following this comprehensive methodology, the study aims to provide an in-depth understanding of how well current generative models can depict the Podlaskie region's architectural style as well as the region in the context of urban trait awareness. This research highlights the strengths and limitations of existing AI image generation tools, offering insights into areas for improvement and future developments to ensure more inclusive and accurate model training practices. Accurate representation of regional architecture in AI-generated imagery is crucial for preserving cultural heritage (Elias, C. 2020), inspiring new architectural designs, and providing meaningful, contextually relevant visual content for a variety of applications.

3. RESULTS

The gathered data was allocated in tables in which a numerical value was given to each correlation of model name and input prompt, as an analysis result of the 20 generated images of each combination. Depicted tables below (Fig. 12-15) are the products of two prior results, of image alignment with a value range of 0-20, and image quality with a value range of 0-100. The first indicates how many images fit (image fit index) the prompt and could be classified as an intended result of the input. An example would be the city of Siemiatycze, which in the EpicRealism model had a value of 13, meaning the other seven images had shown scenery or objects not directly related to the city itself, in this case most of the results had shown snowy mountain regions. The second part of the result product is the image quality index, which is a numerical value of how closely the generated images represent the actual prompt. Again, the city of Siemiatycze generated in the EpicRealism model had scored a value of 50, meaning that the aligned set had shown traits of a city, but the buildings and their architectural style was not directly associated with the prompt input. An example value of 100 would mean the outputs are exactly or almost exactly like that of the prompt, showing an existing fragment or view of the city, town or village. Values closer to zero would indicate that the outputs had almost no visible traits of the prompt, or had miniscule representation of any urban or rural context.

Alignment and quality of generated images

Prompts of the general Podlaskie theme

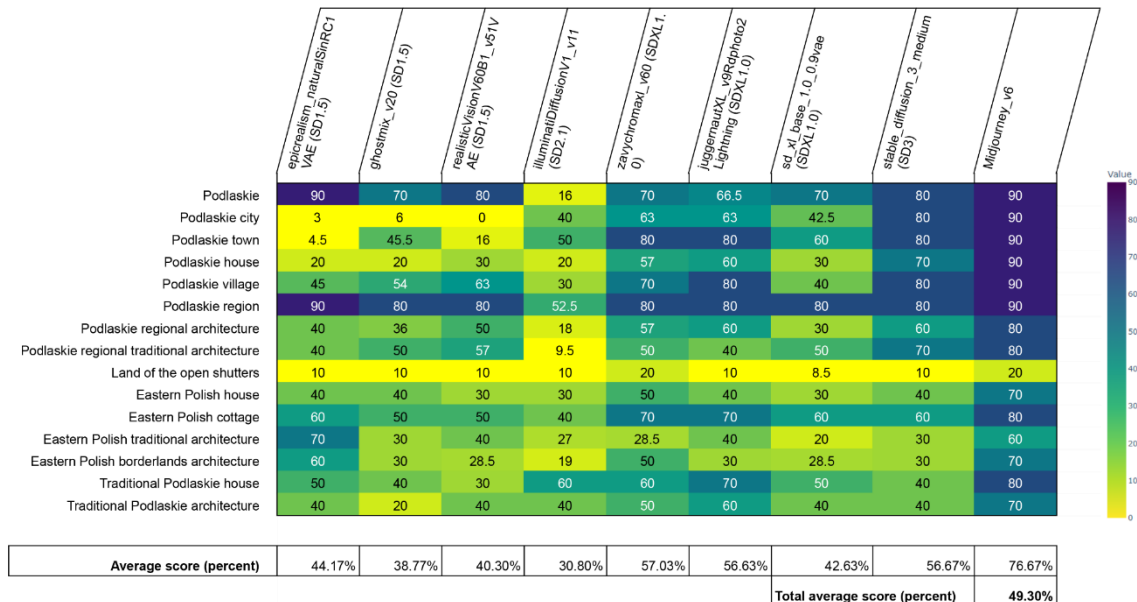


Fig. 12. Chart with the product values of alignment and quality index of Podlaskie prompts. Source: by author

Ryc. 12. Tabela z wartościami produktu indeksu dopasowania i jakości generacji grupy promptów tematyki Podlaskiego. Źródło: autor

Alignment and quality of generated images

Prompts of the 15 largest cities in Podlaskie

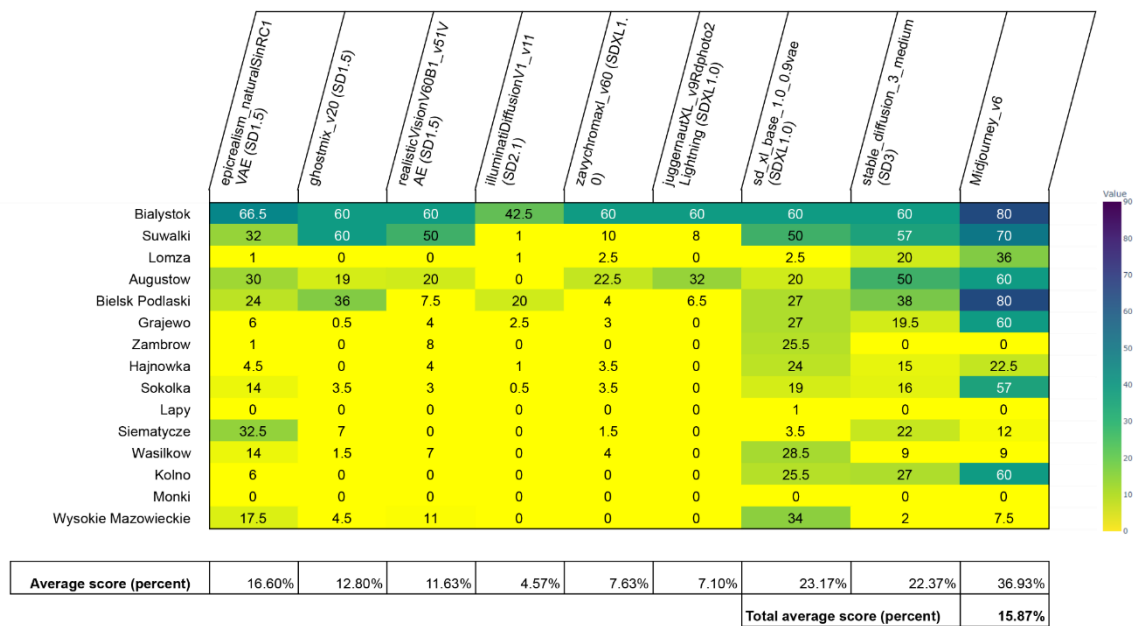


Fig. 13. Chart with the product values of alignment and quality index of city prompts. Source: by author

Ryc. 13 - Tabela z wartościami produktu indeksu dopasowania i jakości generacji grupy promptów miast. Źródło: autor

The visible coloration was added to help visualize the differences between models and prompts and their product result from the mentioned image alignment index and image quality index. Most notably, prompts like “Land of the open shutters” had lackluster values all across the board. It is a picturesque area in the Podlaskie Voivodeship of Poland, renowned for its beautifully preserved wooden architecture. This region, particularly the villages of Trzescianka, Soce, and Puchly, showcases traditional wooden houses with intricately carved and brightly painted shutters. These colorful shutters, along with other decorative elements, highlight the rich cultural heritage and craftsmanship of the local communities. None of the mentioned traits were shown in the generated image outputs, except for a very generic depiction of wooden window shutters right next to an old wooden window. Similarly, outputs for the three mentioned villages were also scored in the 0 - 30% region value, with only Trzescianka having some kind of architectural representation in four of the nine chosen image generation models.

Alignment and quality of generated images

Prompts of the 15 largest towns in Podlaskie

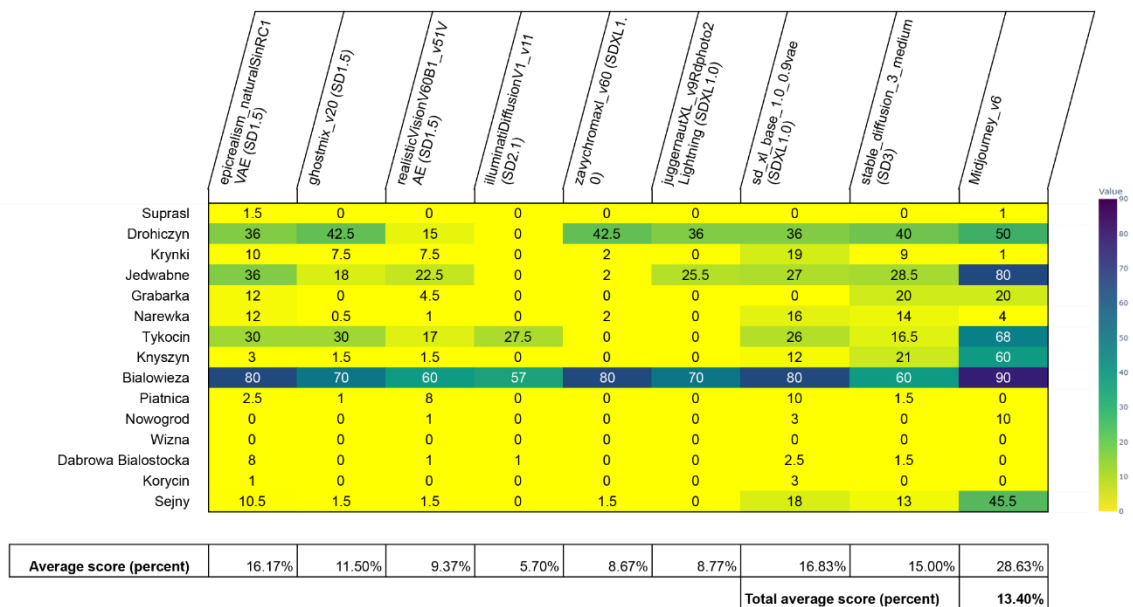


Fig. 14. Chart with the product values of alignment and quality index of town prompts. Source: by author

Ryc. 14 - Tabela z wartościami produktu indeksu dopasowania i jakości generacji grupy promptów miejscowości. Źródło: autor

Alignment and quality of generated images

Prompts of the 15 villages in Podlaskie

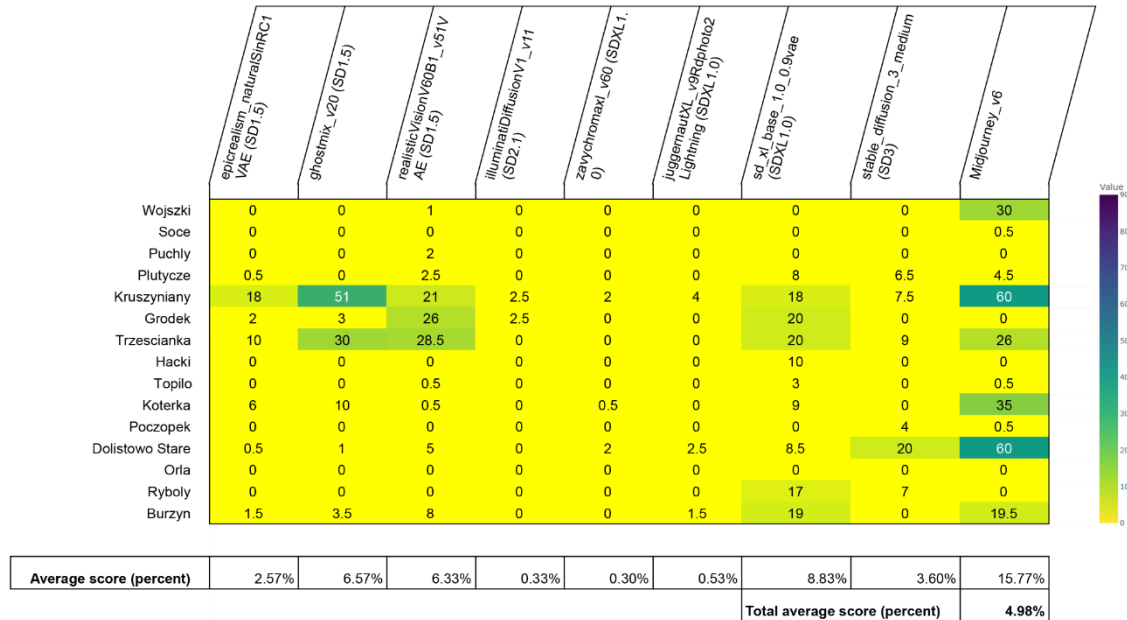


Fig. 15. Chart with the product values of alignment and quality index of village prompts. Source: by author

Ryc. 15 - Tabela z wartościami produktu indeksu dopasowania i jakości generacji grupy promptów wsi. Źródło: autor



Fig. 16. Output for "Podlasie Town" from the MidJourney v6 model. Source: by author

Ryc. 16. Wynik dla promptu "Podlasie Town" z modelu MidJourney v6. Źródło: autor



Fig. 17. Output for "Traditional Podlasie House" from the MidJourney v6 model. Source: by author

Ryc. 17. Wynik dla promptu „Traditional Podlasie House” z modelu MidJourney v6. Źródło: autor



Fig. 18. Output for "Podlasie Village" from the MidJourney v6 model. Source: by author

Ryc. 18. Wynik dla promptu "Podlasie Village" z modelu MidJourney v6. Źródło: autor

In the evaluation of AI-generated images depicting the architecture of the Podlaskie region, MidJourney v6 significantly outperformed all other models, achieving a total average score of 39.50%. This score is notably higher compared to the latest Stable Diffusion model, Stable Diffusion 3 Medium (SD3), which scored 24.41%. This substantial difference highlights MidJourney v6's superior capability in accurately and effectively capturing the architectural and thematic elements of the Podlaskie region (Fig. 16-18). One key factor contributing to this discrepancy is the difference in software architecture between these models. MidJourney, which operates under a proprietary framework,

appears to be more finely tuned for generating complex and contextually rich images, including those of architectural sites.

Tab. 3. Final quality and alignment score for each of the nine generative models. A higher percentage value indicates a higher accuracy and quality of the outputs in context to the assessed prompts. Source: by author

Tab. 3. Wynik końcowy dla każdego z dziewięciu modeli generatywnych. Wyższa wartość procentowa wskazuje na wyższą dokładność oraz jakość wyników w kontekście użytych promptów. Źródło: autor

Generative model / Model generatywny	Final score / Wynik końcowy
epicrealism_naturalSinRC1VAE (SD1.5)	19.88%
ghostmix_v20 (SD1.5)	17.41%
realisticVisionV60B1_v51VAE (SD1.5)	16.91%
illuminatiDiffusionV1_v11 (SD2.1)	10.35%
zavychromaxl_v60 (SDXL1.0)	18.41%
juggernautXL_v9Rdphoto2Lightning (SDXL1.0)	18.26%
sd_xl_base_1.0_0.9vae (SDXL1.0)	22.87%
stable_diffusion_3_medium (SD3)	24.41%
Midjourney_v6	39.50%

On the other hand, the open-sourced Stable Diffusion models, such as Stable Diffusion 3 Medium, have a more prevalent focus on generating human figures and portraits (Luccioni, S.; Akiki, C.; Mitchell, M.; Jernite, Y. 2024). This emphasis likely affects their performance in architectural representation, as their training data and optimization priorities differ from those of MidJourney. When examining the performance of the SDXL1.0 base model (sd_xl_base_1.0_0.9vae), it is evident that it performed remarkably well, achieving a score of 22.87%. This is significantly higher than those of its more specialized counterparts, zavychromaxl_v60 (18.41%) and juggernautXL_v9Rdphoto2Lightning (18.26%). These results suggest that the base model of SDXL1.0 possesses a robust general capability that surpasses the more finely-tuned models in this specific task.

The base model's higher performance can be attributed to its broader training dataset and more generalized approach, which enhances its versatility in generating a variety of images, including those of architectural sites. In contrast, models like Zavychromaxl and Juggernaut, which are specialized for certain aesthetic styles or specific applications, may lack the flexibility required to accurately render the diverse and intricate architectural features of the Podlaskie region. Therefore, while specialized models have their strengths, the results indicate that for tasks involving detailed and contextually rich architectural depictions, more generalized and robust models like the SDXL1.0 base and the highly specialized proprietary model MidJourney v6 are more effective. This underscores the importance of choosing the right model architecture and training focus depending on the specific requirements of the image generation task.

In the context of location based quality of depiction, there is no specific area of the Podlaskie voivodeship that is more or less represented, besides two distinct locations, which are Białystok and Białowieża. While Białystok's architectural and urban style representation in ai image generation models is on an appropriately high level, Białowieża greatly differs from the reason for Białystok's high score. Białystok is the capital of the Podlaskie voivodeship and thus the largest and most well known city with the most modern infrastructure and possibly the most notable presence on the internet, while Białowieża has a non existing representation of the town, but an overwhelming representation of the Białowieża National Park, whose name is mostly abbreviated by using just the word Białowieża. In this case it was decided to leave the result in order to show the importance of keywords in the modern depiction of locations, and their problematic nature in the case of names with multiple instance meanings or uses. Such is with Piatnica, which is both the name of the town as well as a

brand of dairy products, in which case the majority of outputs for this town were images of food, decorative plates and dishes (Fig. 19-21).



Fig. 19. Output for "Piatnica" from the MidJourney v6 model. Source: by author

Ryc. 19. Wynik dla promptu "Piatnica Town" z modelu MidJourney v6. Źródło: autor



Fig. 20. Output for "Piatnica" from the SD3 model. Source: by author

Ryc. 20. Wynik dla promptu „Piatnica” z modelu SD3. Źródło: autor



Fig. 21. Output for "Piatnica" from the ZavychromaXL model. Source: by author

Ryc. 21. Wynik dla promptu "Piatnica" z modelu ZavychromaXL. Źródło: autor

4. DISCUSSION

Visual analysis of the AI-generated images aimed to evaluate how accurately and effectively the architecture of the Podlaskie region is represented in modern image generating algorithms. The images were graded based on their alignment to the prompt and the quality of this alignment, considering factors such as the depiction of cities, towns, villages, and general themes about the Podlaskie region. The primary architectural elements, such as traditional wooden houses, Orthodox churches, and omnipresent Baroque-style buildings, were identified and assessed for accuracy. The images were evaluated based on structural fidelity, stylistic features, and how well they maintained the correct proportions and structural details typical of Podlaskie architecture. The analysis revealed that the AI did not accurately capture the structural fidelity of traditional wooden houses, showed inconsistencies in the ornamentation details of Orthodox churches and in general did not have a good understanding of the Podlaskie region as a whole, apart from nature based scenery and general themes.

To understand the contextual placement and environmental representation, the images were analyzed for landscape integration and the representation of the cultural and historical context within which the architecture exists. This included surrounding artifacts and historical markers. Findings showed that while the landscapes were generally well-integrated, creating a believable setting for the buildings in small towns and villages, the cultural context often lacked specific historical markers that are prevalent in the actual region. Villages located in the Podlaskie voivodeship are mainly located around delicately hilly terrains mostly resembling agricultural presence, along with fairly present forests (Sulima, M. 2018). Generated results had shown this for the most part, however there seemed to be a slight bias towards mountainous terrain for some cases, which is not the case for Podlaskie. The AI-generated images depicted some cities with believable traits, but towns and villages, as well as the depiction of general themes about the Podlaskie region was mostly less accurate. Cities like Bialystok (Fig. 22-24), Suwalki, Augustow, Bielsk Podlaski (Fig. 25) and Grajewo were noted as the most well represented, but still their outputs generated a very broad spectrum of

buildings and sites, almost none of which depicted modern architecture. While true that these cities have a substantial record of having buildings with an average age of over 34 years (Kozłowska, M. 2017), a fairly modern city like Białystok should be equally demonstrated in the context of image generation as any other European city of its size.



Fig. 22. Output for "Białystok" from the SD3 model. Source: by author
Ryc. 22. Wynik dla promptu "Białystok" z modelu SD3. Źródło: autor



Fig. 23. Output for "Białystok" from the ZavychromaXL model. Source: by author
Ryc. 23. Wynik dla promptu "Białystok" z modelu ZavychromaXL. Źródło: autor



Fig. 24. Output for "Białystok" from the MidJourney v6 model. Source: by author
Ryc. 24. Wynik dla promptu "Białystok" z modelu MidJourney v6. Źródło: autor

The overall aesthetic appeal of the AI-generated images was assessed through subjective evaluations and feedback from experts familiar with Podlaskie architecture. This included evaluating the visual harmony and artistic quality of the images as standalone pieces of visual art. It was found that the images were aesthetically pleasing and maintained visual harmony, though some experts noted a lack of authenticity in certain artistic elements. Color schemes and material representations were another focal point in assessing the AI's performance. Traditional Podlaskie architecture is known for its specific use of colors and materials, such as wood, brick, and plaster. The visual analysis indicated that the AI performed well in replicating the regional tone configurations and wooden textures but had difficulty with the intricate decor patterns typical of certain vernacular buildings such as houses and small churches (Szewczyk, J. 2022). It did not, however, produce expected results with iconic buildings such as Tatar Mosques or traditional colorful Orthodox churches, especially in the case of villages which are home to some spectacular wooden architecture like Kruszyniany (Sulima, M.; 2023) or Trzescianka. In addition, an unfortunate finding was noted as almost 86% of the images with prompts hinting at urban and rural context showed derelict or visibly damaged facades with crumbling plaster, missing wall fragments and general degraded sites. Such depiction may be an indirect result of the method of visualizing the Podlaskie voivodeship as underdeveloped (Kiryłuk-Dryjska, E., & Beba, P. 2015), underfunded and slightly derelict region (Fig. 25-27) with a focus on nature instead of modern infrastructure investments (Przygodzka, R. 2008), which can be seen by the numerous photo exhibitions as those of Paweł Tadejko (Białowiecki Ośrodek Kultury. 2018), photo albums and documentaries shown to the public as a means of "promoting" the region as nature inspired, slow and calm (Dąbrowska, A. 2017).



Fig. 25. Image generation output from Midjourney V6 with the prompt "Bielsk Podlaski", depicting a mismanaged, run down section of an old street with semi-similar type of architecture from the mentioned town. Source: by author

Ryc. 25. Wynik generowania obrazu z Midjourney V6 z promptem „Bielsk Podlaski”, przedstawiający źle zarządzany, zaniedbany odcinek starej ulicy z częściowo podobnym stylem architektury ze wspomnianego miasta. Źródło: autor



Fig. 26. Image generation output from Midjourney V6 with the prompt "Suwalki", yet again depicting a mismanaged, run down section of an old street. Visible degradation of facades, lack of general maintenance is visible in the generated content. Source: by author

Ryc. 26. Wynik generowania obrazu z Midjourney V6 z promptem „Suwalki”, ponownie przedstawiający źle zarządzany, zaniedbany odcinek starej ulicy. Widoczna degradacja fasad oraz brak generalnych inwestycji związanych z podtrzymywaniem dobrostanu budynków jest widoczny w generowanych treściach. Źródło: autor



Fig. 27. Image generation output from Midjourney V6 with the prompt "Tykocin", with a severely degraded building, perhaps as an abandoned site. Source: by author

Ryc. 27. Wynik generowania obrazu z Midjourney V6 z promptem „Tykocin”, z niszczącym wiekowym budynkiem, potencjalnie jako pustostanem. Źródło: autor

5. CONCLUSION

This study provided valuable insights into the capabilities and limitations of AI models in replicating the architecture of the Podlaskie region. While the AI models were able to generate visually appealing images that captured some aspects of Podlaskie architecture, they fell short in accurately representing key cultural and architectural features, particularly those related to the region's prominent Orthodox churches and unique rural structures. Instead, the models often produced images that depicted pseudo-baroque architecture with visible defects, and many failed to recognize well-known small

towns and villages such as Ryboły, Plutycze, or Wojszki, which were rendered with little to no representation of their distinctive style and scale. These findings highlight the need for more regionally specific training data and model enhancements to ensure that AI-generated images can more accurately reflect the full essence of the Podlaskie Voivodeship. The tendency of these models to focus on generic or deteriorated structures, rather than the rich and diverse architectural heritage of the region, underscores the limitations of current AI technologies in this context.

However, it is worth noting that similar issues are present even in simple online image searches, which suggests that the underrepresentation of Podlaskie's architectural identity is not unique to AI models. This underscores the broader challenge of preserving and promoting the cultural heritage of underrepresented regions. To address these challenges, a holistic approach to the depiction of the Podlaskie Voivodeship is essential. Future efforts should focus on developing AI models that not only capture the rich cultural heritage and natural landscapes of the region but also incorporate the modest amount of modern infrastructure that exists. By embracing the interconnectedness of these various elements, we can foster a more nuanced and accurate portrayal that truly reflects the multifaceted identity of the Podlaskie Voivodeship. Such a balanced representation not only honors the complexity of the region but also promotes a deeper understanding and appreciation of its unique character.

Looking forward, future research should aim to integrate more diverse and region-specific datasets into AI model training to enhance the accuracy and cultural relevance of the generated images. Expanding this analysis to include other underrepresented regions could further improve our understanding of the capabilities and limitations of generative AI in architectural visualization. Ultimately, these efforts can contribute to the preservation and celebration of the unique architectural identities of regions like Podlaskie, ensuring they are accurately represented in the digital age.

ACKNOWLEDGEMENTS

This research was carried out as part of the work WZ/WA-IA/5/2023 at the Białystok University of Technology and financed by a research grant from the Ministry of Education and Science.

I would like to thank the team of the AuReLA laboratory of the Faculty of Architecture of the Białystok University of Technology (Maciej Popławski, Szymon Andrejczuk, Adam Jakimowicz) for their support in carrying out the research.

WIZUALNA REPREZENTACJA ARCHITEKTURY PODLASIA: OCENA GENERACYJNYCH MODELI SZTUCZNEJ INTELIGENCJI

1. WSTĘP

Modele generowania obrazów stały się kamieniem węgielnym nowoczesnej sztucznej inteligencji, zdolnej do tworzenia skomplikowanych i realistycznych obrazów z opisów tekstowych. Modele te, są szczególnie istotne w dziedzinie przedstawień architektonicznych, gdzie mogą wizualizować style architektoniczne, rekonstrukcje historyczne i koncepcje urbanistyczne z niezwykłą precyzją. Modele takie jak te zbudowane w oparciu o architekturę Stable Diffusion i MidJourney wykorzystują rozległe zbiory danych obejmujące miliony obrazów zarejestrowanych przez kamery i stworzonych przez

ludzi (Yang, K.; Qinami, K.; Fei-Fei, L.; Deng, J.; Russakovsky, O. 2020), umożliwiając im generowanie szczegółowych i kontekstowo dokładnych obrazów.

Właściwa reprezentacja danego regionu, takiego jak architektura regionalna województwa podlaskiego, na obrazach generowanych przez sztuczną inteligencję jest kluczowa z kilku powodów. Dokładne przedstawienie zarówno w prostych, jak i zaawansowanych promptach zapewnia zachowanie i promowanie unikalnego dziedzictwa kulturowego i architektonicznego regionu. Pomaga to nie tylko w działaniach edukacyjnych i konserwatorskich, ale także przyczynia się do większego docenienia i zrozumienia charakterystycznych cech regionu. Co więcej, dla architektów i projektantów dostęp do precyzyjnych wizualizacji architektury województwa podlaskiego może stanowić inspirację dla nowych projektów i integrować tradycyjne elementy z nowoczesnymi projektami (Sliwecki, B.; Jakimowicz, A.; Andrejczuk, S.; Poplawski, M. 2024). Zapewnienie, że modele generowania obrazu mogą wiernie odtwarzać te cechy, pozwala na uzyskanie bardziej znaczących i kontekstowo istotnych wyników (Ryc. 1), zwiększając tym samym użyteczność i niezawodność tych narzędzi w różnych zastosowaniach.

Trening tych modeli opiera się na dużych zbiorach obrazów, które odzwierciedlają różnorodność ludzkich doświadczeń i kreatywności. Tak szeroki wachlarz obrazów treningowych obejmuje zdjęcia budynków, krajobrazów ulicznych, wnętrz i różnych detali architektonicznych, wraz z danymi tła, w tym rozmytymi obiektami, krajobrazami, konfiguracjami oświetlenia, szumem (You, Y.; Zhang, Z.; Hsieh, C. J.; Demmel, J.; Keutzer, K. 2018). Podstawowym mechanizmem tych modeli jest sieć neuronowa, która uczy się kojarzyć określone wzorce w tekście początkowym z elementami wizualnymi na obrazach treningowych. Ten proces uczenia się umożliwia generowanie spójnych i kontekstowo odpowiednich obrazów, czyniąc te modele niezbędnymi narzędziami dla architektów, projektantów i historyków.

Rozwój i udoskonalanie modeli generowania obrazów są często napędzane przez społeczność, opierając swoją pracę na dostępnych modelach bazowych i wsparciu udzielanym przez ich autorów. Naukowcy i entuzjaści przyczyniają się do ewolucji tych systemów, dostrajając je na wyspecjalizowanych zbiorach danych, najczęściej tworzonych specjalnie dla danego kierunku modelu. Dzięki wykorzystaniu LoRAs (Low-Rank Adaptation of Large Language Models) (Hu, E. J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; Chen, W. 2021), ten wspólny wysiłek zapewnia, że modele mogą być dostosowane do określonych tematów lub stylów, zwiększając ich użyteczność w różnych zastosowaniach. Dokładne dostrojenie pozwala na większą precyzję wyników, co jest szczególnie cenne w dziedzinach takich jak tworzenie treści i projektowanie koncepcji, gdzie wizualizacja pomysłów musi być zarówno dokładna, jak i innowacyjna.

Obrazy generowane przez sztuczną inteligencję zrewolucjonizowały branżę kreatywną, zapewniając szybki i skuteczny sposób prototypowania koncepcji i wizualizacji projektów, które w przeciwnym razie wymagałyby znacznego czasu i zasobów. Modele te znalazły zastosowanie między innymi w projektowaniu, reklamie, rozrywce i ochronie dziedzictwa. Skuteczność tych modeli jest jednak silnie uzależniona od kompleksowości ich danych treningowych. Niedostatecznie reprezentowane regiony lub style w zestawach danych szkoleniowych często skutkują wynikami, które nie są dokładne i szczegółowe. Istotne jest, aby modele generowania obrazów były skonfigurowane z określonymi identyfikatorami regionalnymi i stylistycznymi, szczególnie w przypadku obszarów o bogatym dziedzictwie kulturowym i architektonicznym. Taka praktyka zapewnia, że wygenerowane obrazy dokładnie odzwierciedlają unikalne cechy tych lokalizacji (Ruiz, N.; Li, Y.; Jampani, V.; Pritch, Y.; Rubinstein, M.; Aberman, K. 2023).

Niniejszy artykuł ma na celu ocenę wydajności wybranych modeli Stable Diffusion i MidJourney w przedstawianiu regionu podlaskiego, bogatego kulturowo obszaru w północno-wschodniej Polsce, znanego z charakterystycznego stylu architektonicznego. Badanie przedstawione w tym artykule koncentruje się na ocenie, jak dobrze modele te mogą generować obrazy, które wiernie reprezentują cechy architektoniczne i ogólną estetykę regionu podlaskiego. Obejmuje to zarówno badanie dokładności konkretnych wyników, jak i ogólnej jakości generowanych obrazów. Badając możliwości i ograniczenia obecnych narzędzi do generowania obrazów AI w wychwytywaniu niedostatecznie reprezentowanych regionów, badania te mają na celu zapewnienie wglądu w przyszłe ulepszenia i bardziej integracyjne praktyki szkolenia modeli.

2. MATERIAŁY I METODA

Do niniejszego badania wybrano dziesięć modeli generatywnych, aby ocenić ich wydajność w przedstawianiu stylu architektonicznego regionu podlaskiego. Wybrane modele obejmują osiem z sieci Stable Diffusion, jeden reprezentujący najnowszą iterację MidJourney oraz nowo wydany Stable Diffusion 3. Modele te zostały wybrane na podstawie ich popularności, oburęcznej jakości wyjściowej, wcześniejszego doświadczenia z modelem i wysokich ocen na stronie Civit.ai, która jest otwartym i dostępnym online repozytorium modeli, know-how, modeli LoRA i obrazów wyjściowych ze społeczności generowania obrazów. Funkcjonuje obok HuggingFace, głównego repozytorium online dla modeli ai ogólnego przeznaczenia i jednofunkcyjnych. Ze wspomnianej listy modeli jeden wyróżnia się na tle pozostałych pod względem popularności, ale mimo to został wybrany ze względu na swoją unikalną architekturę i podejście, ponieważ „Illuminati Diffusion” jest jedynym przedstawicielem Stable Diffusion 2.1, który został opóźniony w rozwoju z powodu braku popularności, ale zachował bardziej eksperymentalną formę generowania obrazu.

Każdy model został skonfigurowany z optymalnymi ustawieniami dostosowanymi do dokładnego generowania obrazów regionu podlaskiego. Aby zapewnić solidną i kompleksową ocenę, każdemu modelowi przypisano 20 losowych nasion (Xu, K.; Zhang, L.; Shi, J. 2024). Ta zmienność nasion losowych pomaga uwzględnić nieodłączną stochastyczną naturę modeli generatywnych, zapewniając szerszy zakres wyników do oceny. Ustawienia zostały wybrane indywidualnie dla każdego modelu, ponieważ te same ustawienia próbnika nie powinny być stosowane do wszystkich modeli generatywnych ze względu na indywidualne różnice w logice generowania szumu, a także ustawienia Classifier-Free Guidance (CFG), które mają znaczący wpływ na jakość obrazu wyjściowego i możliwe artefakty, które są bezpośrednim wynikiem bardziej ekstremalnych wartości (Ho, J.; Salimans, T. 2022). Z ogólnych ustawień wyłączono model generatywny MidJourney V6, który bazuje na innym zestawie ustawień, z których najważniejszym jest styl generowania „RAW” i wartość stylizacji równa 100, współczynnik „dziwność” równy zero i różnorodność na poziomie zero.

W badaniu wykorzystano 60 promptów specjalnie zaprojektowanych w celu uchwycenia różnych aspektów architektury regionu podlaskiego. Prompty te zostały systematycznie podzielone na cztery równe kategorie: region ogólny, duże miasta, małe miejscowości i wsie. Prompty opisujące ogólny styl architektoniczny i środowisko województwa podlaskiego należały do kategorii ogólnej regionu (Ryc. 2-7), podczas, gdy te skupiające się na cechach architektonicznych głównych miast w regionie zostały skategoryzowane, jako duże miasta. Prompty związane z cechami architektonicznymi miast średniej wielkości zostały sklasyfikowane, jako miasta, a te podkreślające unikalne elementy architektoniczne małych wsi zostały umieszczone w kategorii wsie. Nazwy zostały zmodyfikowane, aby pominąć polskie symbole, ponieważ metody tokenizacji w Stable Diffusion oraz w architekturze MidJourney nie działają dobrze z obcymi symbolami i są przeznaczone do użytku w języku angielskim. Po wpisaniu polskich symboli wyniki były znacznie bardziej losowe, a ich nazwa pliku sugerowała brakujący znak. Każdy z 60 promptów został wykorzystany do wygenerowania 20 obrazów na każdym z 9 wybranych modeli, co dało łącznie 10 800 obrazów do oceny. Takie kompleksowe podejście zapewnia szczegółową ocenę możliwości każdego modelu w generowaniu dokładnych i wysokiej jakości obrazów w oparciu o podane prompty.

Wygenerowane obrazy zostały ocenione przy użyciu dwóch podstawowych ocen: dopasowania do promptu oraz jakości wyników. Wynik dopasowania mierzy, jak dokładnie wygenerowany obraz odzwierciedla treść i specyfikę danego promptu w skali od 0 do 20, gdzie zero oznacza grupę całkowicie przypadkowych obrazów, a wynik 20 jako dokładnie reprezentujący promptowany temat / słowa kluczowe. Wyniki jakości oceniają ogólną wartość estetyczną i techniczną wygenerowanych obrazów, w tym takie aspekty, jak przejrzystość, spójność i atrakcyjność wizualna, ale mają również wpływ na współczynnik dopasowania odnoszący się do klasyfikacji obrazów do poziomu reprezentacji tematu (Ryc. 8-10). Przykładowy prezentowany obraz zostałby uznany za dopasowany do promptu i reprezentowałby wysoką wartość jakości dopasowania wynoszącą 80%. Wartość ta oznacza, że otaczające budynki i scenaria mogą znajdować się w przykładowym Białymstoku, że przypominają ogólny nastrój i styl, ale nie są bezpośrednią reprezentacją żadnego realnego fragmentu miasta (Ryc. 11).

Wyniki oceny obrazów zostaną zwizualizowane za pomocą wykresów, aby zapewnić jasną i zwięzłą reprezentację wydajności modeli. Wspomniane wizualizacje ułatwią analizę porównawczą różnych modeli i kategorii promptów. Wyniki zostaną omówione w kontekście początkowych założeń i domysłów dotyczących wydajności każdego modelu. Zbadane zostaną potencjalne przyczyny zaobserwowanych wyników, w tym wpływ danych szkoleniowych, architektury modelu i konkretnych ustawień zastosowanych w procesie generowania.

Postępując zgodnie z tą kompleksową metodologią, badanie ma na celu zapewnienie dogłębnego zrozumienia, jak dobrze obecne modele generatywne mogą przedstawiać styl architektoniczny województwa podlaskiego, a także region w kontekście świadomości cech miejskich. Badanie to podkreśla mocne strony i ograniczenia istniejących narzędzi do generowania obrazów AI, oferując wgląd w obszary wymagające poprawy i przyszłe zmiany w celu zapewnienia bardziej integracyjnych i dokładnych praktyk szkolenia modeli. Dokładne odwzorowanie architektury regionalnej w obrazach generowanych przez sztuczną inteligencję ma kluczowe znaczenie dla zachowania dziedzictwa kulturowego (Elias, C. 2020), inspirowania nowych projektów architektonicznych i dostarczania znaczących, kontekstowo istotnych treści wizualnych do szeroko pojętych zastosowań.

3. REZULTATY

Zebrane dane zostały podzielone na tabele, w których nadano wartość liczbową każdej korelacji nazwy modelu i monitu wejściowego, jako wynik analizy 20 wygenerowanych obrazów każdej kombinacji. Przedstawione poniżej tabele (Ryc. 12-15) są produktami dwóch wcześniejszych wyników, dopasowania obrazu z zakresem wartości 0-20 i jakości obrazu z zakresem wartości 0-100. Pierwszy wskazuje, ile obrazów pasuje (wskaźnik dopasowania obrazu) do promptu i może być sklasyfikowany, jako zamierzony wynik danych wejściowych. Przykładem może być miasto Siemiatycze, które w modelu EpicRealism miało wartość 13, co oznacza, że pozostałe siedem obrazów pokazywało scenierię lub obiekty niezwiązane bezpośrednio z samym miastem, w tym przypadku większość wyników pokazywała zaśnieżone regiony górskie. Drugą częścią jest wskaźnik jakości obrazu, który jest wartością liczbową określającą, jak blisko wygenerowane obrazy reprezentują rzeczywisty prompt. Ponownie, miasto Siemiatycze wygenerowane w modelu EpicRealism uzyskało wartość 50, co oznacza, że dopasowany zestaw pokazał cechy miasta, ale budynki i ich styl architektoniczny nie były bezpośrednio związane z promptem. Przykładowa wartość 100 oznaczałaby, że wyniki są dokładnie lub prawie dokładnie takie, jak w promptach, pokazując istniejący fragment lub widok miasta, miasteczka lub wsi. Wartości bliższe zeru wskazywałyby, że wyniki nie miały prawie żadnych widocznych cech promptów lub miały niewielką reprezentację jakiegokolwiek kontekstu miejskiego lub wiejskiego.

Widoczne zabarwienie zostało dodane, aby pomóc w wizualizacji różnic między modelami i promptami, a ich produktem wynikającym ze wspomnianego wskaźnika wyrównania obrazu i wskaźnika jakości obrazu. W szczególności, prompty takie jak „Kraina otwartych okiennic” miały słabe wartości na całej planszy. Jest to malowniczy obszar w województwie podlaskim, słynący z pięknie zachowanej architektury drewnianej. Region ten, w szczególności wsie Trześcianka, Soce i Puchły, prezentuje tradycyjne drewniane domy z misternie rzeźbionymi i jaskrawo pomalowanymi okiennicami. Te kolorowe okiennice, wraz z innymi elementami dekoracyjnymi, podkreślają bogate dziedzictwo kulturowe i sztuk lokalnych społeczności. Żadna z wymienionych cech nie została pokazana w wygenerowanych obrazach, z wyjątkiem bardzo ogólnego przedstawienia drewnianych okiennic tuż obok starego drewnianego okna. Podobnie, wyniki dla trzech wspomnianych wiosek zostały również ocenione na poziomie 0-30%, przy czym tylko Trześcianka miała jakąś reprezentację architektoniczną w czterech z dziewięciu wybranych modeli generowania obrazu.

W ocenie obrazów generowanych przez sztuczną inteligencję przedstawiających architekturę województwa podlaskiego, MidJourney v6 znacznie przewyższył wszystkie inne modele, osiągając łączny średni wynik 39,50%. Wynik ten jest znacznie wyższy w porównaniu z najnowszym modelem Stable Diffusion, Stable Diffusion 3 Medium (SD3), który uzyskał 24,41%. Ta znacząca różnica podkreśla wyższą zdolność MidJourney v6 do dokładnego i skutecznego uchwycenia elementów architektonicznych i tematycznych regionu podlaskiego (Ryc. 16-18). Jednym z kluczowych czynników

przyczyniających się do tej rozbieżności jest różnica w architekturze oprogramowania między tymi modelami. MidJourney, który działa w oparciu o zastrzeżone ramy, wydaje się być bardziej precyzyjnie dostrojony do generowania złożonych i bogatych kontekstowo obrazów, w tym obrazów obiektów architektonicznych.

Z drugiej strony, modele Stable Diffusion o otwartym kodzie źródłowym, takie jak Stable Diffusion 3 Medium, koncentrują się bardziej na generowaniu postaci ludzkich i portretów (Luccioni, S.; Akiki, C.; Mitchell, M.; Jernite, Y. 2024). Nacisk ten prawdopodobnie wpływa na ich wydajność w reprezentacji architektonicznej, ponieważ ich dane treningowe i priorytety optymalizacji różnią się od tych z MidJourney. Badając wydajność modelu bazowego SDXL1.0 (sd_xl_base_1.0_0.9vae), widać, że wypadł on wyjątkowo dobrze, osiągając wynik 22,87%. Jest to znacznie wyższy wynik niż w przypadku jego bardziej wyspecjalizowanych odpowiedników, zavychromaxl_v60 (18,41%) i juggernautXL_v9Rdphoto2Lightning (18,26%). Wyniki te sugerują, że model bazowy SDXL1.0 posiada solidne ogólne możliwości, które przewyższają bardziej dopracowane modele w tym konkretnym zadaniu.

Wyższą wydajność modelu bazowego można przypisać jego szerszemu zestawowi danych szkoleniowych i bardziej ogólnego podejścia, co zwiększa jego wszechstronność w generowaniu różnorodnych obrazów, w tym obrazów obiektów architektonicznych. Z kolei modele takie jak Zavychromaxl i Juggernaut, które specjalizują się w określonych stylach estetycznych lub konkretnych zastosowaniach, mogą nie mieć elastyczności wymaganej do dokładnego renderowania różnorodnych i skomplikowanych elementów architektonicznych regionu podlaskiego. Dlatego też, choć wyspecjalizowane modele mają swoje mocne strony, wyniki wskazują, że w przypadku zadań obejmujących szczegółowe i bogate kontekstowo przedstawienia architektoniczne, bardziej uogólnione i solidne modele, takie jak baza SDXL1.0 i wysoce wyspecjalizowany, zastrzeżony model MidJourney v6, są bardziej skuteczne. Podkreśla to znaczenie wyboru odpowiedniej architektury modelu i ukierunkowania treningu w zależności od konkretnych wymagań zadania generowania obrazu.

W kontekście jakości odwzorowania opartej na lokalizacji, nie ma określonego obszaru województwa podlaskiego, który byłby mniej lub bardziej reprezentowany, poza dwoma odrębnymi lokalizacjami, którymi są Białystok i Białowieża. Podczas gdy reprezentacja stylu architektonicznego i urbanistycznego Białegostoku w modelach generowania obrazu przez sztuczną inteligencję jest na odpowiednio wysokim poziomie, Białowieża znacznie różni się od wysokiego wyniku Białegostoku. Białystok jest stolicą województwa podlaskiego, a zatem największym i najbardziej znanym miastem z najnowocześniejszą infrastrukturą i prawdopodobnie najbardziej zauważalną obecnością w Internecie, podczas gdy Białowieża ma nieistniejącą reprezentację miasta, ale przytłaczającą reprezentację Białowieskiego Parku Narodowego, którego nazwa jest najczęściej skracana przy użyciu samego słowa Białowieża. W tym przypadku zdecydowano się pozostawić wynik, aby pokazać znaczenie słów kluczowych we współczesnym przedstawianiu lokalizacji oraz ich problematyczną naturę w przypadku nazw o wielu znaczeniach lub zastosowaniach. Tak jest w przypadku Piątnicy, która jest zarówno nazwą miasta, jak i marką produktów mlecznych, w którym to przypadku większość wyników dla tego miasta stanowiły obrazy żywności, dekoracyjne talerze i naczynia (Ryc. 19-21).

4. Dyskusja

Analiza wizualna obrazów wygenerowanych przez sztuczną inteligencję miała na celu ocenę, jak dokładnie i skutecznie architektura regionu podlaskiego jest reprezentowana przez nowoczesne algorytmy generowania obrazów. Obrazy zostały ocenione na podstawie ich dopasowania do promptów i jakości tego dopasowania, biorąc pod uwagę takie czynniki, jak przedstawienie miast, miasteczek, wsi i ogólnych tematów dotyczących regionu podlaskiego. Główne elementy architektoniczne, takie jak tradycyjne drewniane domy, cerkwie i wszechobecne budynki w stylu barokowym, zostały zidentyfikowane i ocenione pod kątem dokładności. Obrazy zostały ocenione na podstawie wierności strukturalnej, cech stylistycznych oraz tego, jak dobrze zachowały prawidłowe proporcje i szczegóły konstrukcyjne typowe dla architektury Województwa Podlaskiego. Analiza wykazała, że sztuczna inteligencja niedokładnie uchwyciła wierność strukturalną tradycyjnych drewnianych domów,

wykazała niespójności w szczegółach ornamentacji cerkwi i ogólnie nie miała dobrego zrozumienia regionu podlaskiego jako całości, poza scenografią opartą na naturze i ogólnymi tematami.

Aby zrozumieć kontekstowe umiejscowienie i reprezentację środowiska, obrazy zostały przeanalizowane pod kątem integracji krajobrazu oraz reprezentacji kontekstu kulturowego i historycznego, w którym istnieje architektura. Obejmowało to otaczające artefakty i znaczniki historyczne. Wyniki pokazały, że chociaż krajobrazy były ogólnie dobrze zintegrowane, tworząc wiarygodne otoczenie dla budynków w małych miastach i wsiach, kontekst kulturowy często nie zawierał konkretnych znaczników historycznych, które są powszechne w rzeczywistym regionie. Wsie znajdujące się w województwie podlaskim są zlokalizowane głównie wokół nieco pagórkowatych terenów, w większości przypominających tereny rolnicze, wraz z dość obecnymi lasami (Sulima, M. 2018). Wygenerowane wyniki pokazały to w większości, jednak wydawało się, że w niektórych przypadkach występuje lekkie odchylenie w kierunku terenów górskich, co nie ma miejsca w przypadku województwa podlaskiego. Obrazy wygenerowane przez sztuczną inteligencję przedstawiały niektóre miasta z wiarygodnymi cechami, ale miasta i wsie, a także ogólne tematy dotyczące regionu podlaskiego były w większości mniej dokładne. Miasta takie jak Białystok (Ryc. 6-8), Suwałki, Augustów, Bielsk Podlaski i Grajewo zostały odnotowane, jako najlepiej reprezentowane, ale mimo to ich wyniki wygenerowały bardzo szerokie spektrum budynków i miejsc, z których prawie żadne nie przedstawiało nowoczesnej architektury. Chociaż prawdą jest, że miasta te mają znaczną liczbę budynków o średnim wieku ponad 34 lat (Kozłowska, M. 2017), dość nowoczesne miasto, takie jak Białystok, powinno być równie dobrze pokazane w kontekście generowania obrazu, jak każde inne europejskie miasto tej wielkości.

Ogólna estetyka obrazów wygenerowanych przez sztuczną inteligencję została oceniona na podstawie subiektywnych ocen i opinii ekspertów zaznajomionych z podlaską architekturą. Obejmowało to ocenę harmonii wizualnej i jakości artystycznej obrazów, jako samodzielnych dzieł sztuki wizualnej. Stwierdzono, że obrazy były estetyczne i zachowywały wizualną harmonię, choć niektórzy eksperci zauważyli brak autentyczności w niektórych elementach artystycznych. Schematy kolorystyczne i reprezentacje materiałów były kolejnym centralnym punktem oceny wydajności AI. Tradycyjna architektura podlaska znana jest ze specyficznego wykorzystania kolorów i materiałów, takich jak drewno, cegła i tynk. Analiza wizualna wykazała, że sztuczna inteligencja dobrze radziła sobie z odtwarzaniem regionalnych konfiguracji tonów i tekstur drewna, ale miała trudności ze skomplikowanymi wzorami dekoracyjnymi typowymi dla niektórych tradycyjnych budynków, takich jak domy i małe kościoły (Szewczyk, J. 2022). Nie przyniosło to jednak oczekiwanych rezultatów w przypadku ikonicznych budynków, takich jak tatarskie meczety lub tradycyjne kolorowe cerkwie, zwłaszcza w przypadku wiosek, w których znajdują się spektakularne drewniane budynki, takie jak Kruszyniany (Sulima, M.; 2023) lub Trześcianka. Ponadto odnotowano niefortunne odkrycie, ponieważ prawie 86% zdjęć z promptami wskazującymi na kontekst miejski i wiejski przedstawiało opuszczone lub wyraźnie uszkodzone fasady z kruszącym się tynkiem, brakującymi fragmentami ścian i ogólnie zdegradowanymi miejscami. Takie przedstawienie może być pośrednim wynikiem metody wizualizacji województwa podlaskiego, jako słabo rozwiniętego (Kiryłuk-Dryjska, E., & Beba, P. 2015), niedofinansowanego i nieco zaniedbanego regionu (Ryc. 9) z naciskiem na przyrodę zamiast nowoczesnych inwestycji infrastrukturalnych (Przygodzka, R. 2008), co można zaobserwować na licznych blogach (Białowiecki Ośrodek Kultury. 2018), albumach fotograficznych i filmach dokumentalnych pokazywanych publicznie, jako środek „promowania” regionu, jako inspirowanego naturą, powolnego i spokojnego (Dabrowska, A. 2017).

5. WNIOSKI

Niniejsze badanie dostarczyło cennych informacji na temat możliwości i ograniczeń modeli AI w odtwarzaniu architektury regionu podlaskiego. Choć modele AI były w stanie wygenerować wizualnie atrakcyjne obrazy, które uchwyciły niektóre aspekty architektury Podlasia, jednak nie sprostaly one zadaniu dokładnego odwzorowania kluczowych cech kulturowych i architektonicznych, w szczególności dotyczących znaczących dla regionu cerkwi prawosławnych oraz unikalnych struktur wiejskich. Zamiast tego, modele często generowały obrazy przedstawiające architekturę pseudo-barokową z widocznymi defektami, a wiele z nich nie rozpoznało dobrze znanych małych miasteczek i wsi, takich jak Ryboły, Plutycze czy Wojszki, które zostały ukazane z niewielkim lub żadnym odniesieniem do

ich charakterystycznego stylu i skali. Wyniki te podkreślają potrzebę bardziej specyficznych dla regionu danych treningowych oraz dalszego doskonalenia modeli, aby zapewnić, że obrazy generowane przez AI będą dokładniej odzwierciedlać pełną istotę województwa podlaskiego. Tendencja tych modeli do koncentrowania się na ogólnych lub zdegradowanych strukturach, zamiast na bogatym i różnorodnym dziedzictwie architektonicznym regionu, ujawnia ograniczenia współczesnych technologii AI w tym kontekście.

Warto jednak zauważyć, że podobne problemy występują nawet w prostych poszukiwaniach obrazów w internecie, co sugeruje, że niedostateczna reprezentacja tożsamości architektonicznej Podlasia nie jest wyjątkowa dla modeli AI. Podkreśla to szersze wyzwanie, jakim jest zachowanie i promowanie dziedzictwa kulturowego niedostatecznie reprezentowanych regionów. Aby sprostać tym wyzwaniom, konieczne jest holistyczne podejście, które uwzględni zarówno dziedzictwo kulturowe, jak i nowoczesną infrastrukturę, do przedstawiania województwa podlaskiego. Przyszłe działania powinny skoncentrować się na opracowaniu modeli AI, które nie tylko uchwycą bogate dziedzictwo kulturowe i krajobrazy naturalne regionu, ale także uwzględnią skromną ilość nowoczesnej infrastruktury, która istnieje na tym obszarze. Przyjmując współzależność tych różnych elementów, możemy stworzyć bardziej zniuansowany i dokładny obraz, który naprawdę odzwierciedla wieloaspektową tożsamość województwa podlaskiego. Taka zrównoważona reprezentacja nie tylko odzwierciedla złożoność regionu, ale także promuje głębsze zrozumienie i docenienie jego unikalnego charakteru.

Patrząc w przyszłość, przyszłe badania powinny dążyć do integracji bardziej zróżnicowanych i specyficznych dla regionu zestawów danych w szkoleniu modeli AI, aby poprawić dokładność i kulturową adekwatność generowanych obrazów. Rozszerzenie tej analizy na inne niedostatecznie reprezentowane regiony mogłoby poszerzyć nasze rozumienie możliwości i ograniczeń generatywnej sztucznej inteligencji w wizualizacji architektonicznej. Ostatecznie, te działania mogą przyczynić się do zachowania i celebrowania unikalnych tożsamości architektonicznych takich regionów jak Podlasie, zapewniając ich dokładne odwzorowanie w erze cyfrowej.

PODZIĘKOWANIA

Badania te zostały zrealizowane w ramach prac WZ/WA-IA/5/2023 na Politechnice Białostockiej i sfinansowane z dotacji badawczej Ministerstwa Edukacji i Nauki.

Serdecznie dziękuję zespołowi laboratorium AuReLA Wydziału Architektury Politechniki Białostockiej (Maciej Popławski, Szymon Andrejczuk, Adam Jakimowicz) za wsparcie przy realizacji badań.

BIBLIOGRAPHY

- Białowiecki Ośrodek Kultury. (2018) O-koło-podlaskie wyprawy. Galeria oBOK, ul. Sportowa 1, Białowieża, Poland
- Dąbrowska, A., 2017. Traditional sectors based on natural resources-a blessing or a curse for less developed regions? A case study of Podlaskie Voivodeship. *Miscellanea Geographica. Regional Studies on Development*, 21(3), pp.89-95.
- Elias, C., 2020. Representation. *A Concise Companion to Visual Culture*, pp.421-440.
- Ho, J. and Salimans, T., 2022. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*.
- Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L. and Chen, W., 2021. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*.
- Kiryłuk-Dryjska, E., & Beba, P., 2015. Barriers to the development of rural areas in eastern polish border regions. *Wyzwania polityki regionalnej w aspekcie rozwoju społeczno-gospodarczego obszarów transgranicznych*, 255.

- Kozłowska, M., 2017. Rynek obrotu nieruchomościami w Białymstoku w XVIII wieku. *Studia Podlaskie*, 25, pp.103-27.
- Luccioni, S., Akiki, C., Mitchell, M. and Jernite, Y., 2024. Stable bias: Evaluating societal representations in diffusion models. *Advances in Neural Information Processing Systems*, 36.
- Przygodzka, R., 2008. Tradition Or Innovation–Which Approach Is Better In Rural Development? The Case Of Podlasie Region.
- Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M. and Aberman, K., 2023. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 22500-22510).
- Sliwecki, B., Jakimowicz, A., Andrejczuk, S. and Poplawski, M., 2024. Creative Experimental Architectural Design Teaching. In *Novel Approaches to Urban Design and Architecture Education: Design Studio Practice and Pedagogy* (pp. 156-177). IGI Global.
- Sulima, M., 2018. Dom pogranicza w kulturze wsi podlaskiej. *Oficyna Wydawnicza Politechniki Białostockiej*.
- Sulima, M., 2023. Protection of wooden architecture monuments and the cultural heritage of Polish Tatars (on the example of the Bohoniki Cultural Park and the landscape park in Kruszyniany). *Architectus*, (1 (73).
- Szewczyk, J., 2022. Ludowe zdobnictwo podlaskich domów. *ARCHITREND*. PL Rafał Filcek.
- Yang, K., Qinami, K., Fei-Fei, L., Deng, J. and Russakovsky, O., 2020, January. Towards fairer datasets: Filtering and balancing the distribution of the people subtree in the imagenet hierarchy. In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 547-558).
- You, Y., Zhang, Z., Hsieh, C.J., Demmel, J. and Keutzer, K., 2018, August. Imagenet training in minutes. In *Proceedings of the 47th international conference on parallel processing* (pp. 1-10).
- Xu, K., Zhang, L. and Shi, J., 2024. Good Seed Makes a Good Crop: Discovering Secret Seeds in Text-to-Image Diffusion Models. *arXiv preprint arXiv:2405.14828*.

AUTHOR'S NOTE

The author's current scientific focus has shifted from virtual reality architectural design and their functionality in the metaverse via online social platforms to understanding visual perception of architectural designs and living spaces, especially in the context of AI image generation and visual representation of architecture. Among these two fields, the author also works with 3D printing, 3D scanning, and experimental architectural design methods.

O AUTORZE

Obecne zainteresowania naukowe autora przesunęły się z projektowania architektonicznego w rzeczywistości wirtualnej i jego funkcjonalności w metaverse za pośrednictwem internetowych platform społecznościowych na zrozumienie wizualnej percepcji projektów architektonicznych i przestrzeni mieszkalnych, szczególnie w kontekście generowania obrazów AI i wizualnej reprezentacji architektury. Wśród tych dwóch dziedzin autor zajmuje się także drukiem 3D, skanowaniem 3D i eksperymentalnymi metodami projektowania architektonicznego.

Contact | Kontakt: b.sliwecki@pb.edu.pl