

# Ograniczenia współczesnych generatorów obrazów opartych o AI w procesie projektowania architektonicznego



dr inż. arch.

**ADAM GIL**

Politechnika Śląska  
Wydział Architektury

ORCID: 0000-0003-4148-8682

Na rynku dostępne są zarówno narzędzia przeznaczone bezpośrednio do projektowania architektonicznego, jak i narzędzia o charakterze ogólnym, służące do generowania dowolnych obrazów, od fotorealistycznych po malarskie i graficzne. Artykuł analizuje przydatność i ograniczenia tych narzędzi z perspektywy architekta, biorąc pod uwagę stan technologii na koniec 2023 roku.

Rozwój modeli AI do generowania obrazów rozpoczął się w 2014 roku z wprowadzeniem GAN przez Iana Goodfellowa [1]. W 2018 roku DeepMind zaprezentowało ulepszoną wersję, BigGAN. W 2021 roku wprowadzono model BigSleep, łączący BigGAN i CLIP, oraz metodykę VQGAN-CLIP, co znacznie poprawiło jakość. W 2022 i 2023 roku wprowadzono modele Guided Diffusion [2] i Latent Diffusion [3], które przyspieszyły proces generowania.

Wdrożenie ww. narzędzi generatywnych w architekturze nie jest dobrze przebadane w literaturze naukowej, niemniej jest tematem publikacji pozanaukowych w portalach internetowych o tematyce architektonicznej (np. Archdaily). Szczególnie popularne jest prezentowanie możliwości tych narzędzi w serwisach typu YouTube, gdzie odnaleźć można dziesiątki filmów instruktażowych i poradników dotyczących potencjalnego użycia tych narzędzi przez architektów.

## Sytuacja rynkowa w zakresie użycia AI w architekturze

W ciągu ostatnich dwóch lat (2022/2023) zastosowanie AI w projektowaniu architektonicznym przeszło od fazy doświadczalnej do fazy praktycznej. Architekci zainteresowani są głównie generatorami obrazów AI, które umożliwiają eksplorację koncepcji projektowych poprzez konwersję tekstu na obraz (text-to-image). Wywołują one skrajne emocje w środowisku architektonicznym.

W przeciągu tych dwóch lat pojawiła się w internecie znaczna liczba serwisów oferujących takie narzędzia (większość uruchomiono

w 2023 roku, wnioskując z dat rejestracji domen). Aktualizowany na bieżąco ranking (2023) serwisu G2.com [4] wymienia 68 takich narzędzi dostępnych online.

## Sposób dostępu do narzędzi AI

Zdecydowana większość z tych narzędzi AI to narzędzia dostępne wyłącznie online, w postaci płatnych, abonamentowych usług sieciowych. Uzasadnieniem takiego modelu dostępu jest konieczność użycia do generowania tych wariantów dużej mocy obliczeniowej niedostępnej dla stacjonarnych komputerów. Jedynym wyjątkiem jest generator obrazów Stable Diffusion (SD) i jego pochodne, jak SD XL, które mogą być uruchamiane lokalnie. Nie dotyczą go więc ograniczenia finansowe pozostałych.

Podczas przygotowania tego artykułu przebadano 21 z dostępnych obecnie generatorów obrazu (tab. 1.). Pozwoliło to na uogólnienie obserwacji i wyciągnięcie wniosków, które dotyczą całej grupy narzędzi. Niemniej zdecydowana większość testów została przeprowadzona przy użyciu Stable Diffusion. Było to podyktowane zarówno największą kontrolą nad narzędziem (dostępem do parametrów, które są niedostępne w innych narzędziach), jak i kwestią kosztów.

## Metodyka działania z użyciem generatorów obrazów AI

Artykuł ten nie zajmuje się objaśnieniem matematycznej strony generowania obrazów przez ww. narzędzia. Istnieje szeroka literatura specjalistyczna omawiająca tę tematykę [3], [5]. Niniejszy artykuł nie aspiruje także

do roli poradnika obsługi narzędzi AI, skupia się jedynie na tych elementach, które mają bezpośrednie przełożenie na przydatność tych generatorów jako narzędzi w pracy architekta.

Należy podkreślić, że na obecnym etapie narzędzia te nadają się wyłącznie do generowania widoków zbliżonych do zdjęć lub renderingów. Nie nadają się do generowania jakichkolwiek rysunków rzutów czy przekrojów. O ile bowiem istnieje techniczna możliwość wygenerowania obrazu przypominającego taki rzut, o tyle jego zawartość merytoryczna będzie nieakceptowalna, nawet dla bardzo prostych zadań (np. rzut domu jednorodzinnego). Pomiędzy wygenerowanymi obrazami nie ma też związków logicznych (np. nie ma możliwości wygenerowania widoku wcześniej wygenerowanego budynku z przeciwnej strony). W grupie testowanych narzędzi istnieją co prawda takie, które jako bazy do generowanych obrazów używają modelu 3D z programu CAD / BIM, ale nawet one nie radzą sobie z utrzymaniem spójności pomiędzy ujęciami.

## Konstrukcja poleceń (prompt)

Podstawowa metodyka pracy z takimi narzędziami polega na wpisaniu w pole polecenia (prompt) opisu tego, co ma znajdować się na docelowym obrazie (np. dom na zboczu nad jeziorem, o zmiernym). Można się w tym celu posługiwać językiem naturalnym. Polecenie ma z reguły formę opisu charakteryzującego temat (np. house) jego cechy (small, contemporary, cottage house, with sloped roof), kontekst (on a hill, at the lake,

on the hazy morning) i sposób przedstawienia (np. architectural photography, mock-up etc.). Przykłady takich poleceń zostały zaprezentowane w artykule Analiza metod pracy projektantów korzystających z generatorów obrazów opartych na sztucznej inteligencji w kreatywnych fazach projektowania architektonicznego w tym numerze Builder Science [11].

Większość narzędzi posiada funkcjonalności, które – w teorii – powinny pozwolić na lepszą kontrolę nad tworzonym poleceniem. Część z narzędzi oferuje możliwość użycia tzw. negative prompts, czyli opisów tego, co nie powinno znaleźć się w wygenerowanym obrazie (np. old house, weathered facade, noise).

Stable Diffusion XL pozwala zwiększać i zmniejszać wagę poszczególnych pojęć przy użyciu mnożników, np. zapis (rounded building1:2) oznacza zwiększenie ważności pojęcia w nawiasie o 20% w porównaniu do tego samego pojęcia bez mnożnika. Pomimo dużego zakresu możliwości sterowania opisem generowane rezultaty często odbiegają od oczekiwań użytkownika. Podstawowym problemem jest pomijanie przez sieć części opisu, w szczególności gdy wiele pojęć (tzw. tokenów) „walczy” o uwagę sieci. W praktyce skuteczność wszystkich wymienionych metod jest ograniczona.

## Losowość wyników

Wynik jest rezultatem wielu czynników: źródłowego „szumu”, od którego zaczyna się generowanie obrazów, zastosowanej metody próbkowania, liczby kroków, które służą do wygenerowania obrazu, promptu, rozdzielczości obrazu, zastosowanych parametrów oraz sposobu zbudowania i wytrenowania sieci. Prowadzi to do randomizacji rezultatów i powoduje, że prawie zawsze odbiegają one od spodziewanego rezultatu.

Na poziomie promptu powoduje to konieczność szukania pojęć oraz przekształcania polecenia na wiele sposobów, tak aby osiągnąć w końcu spodziewany efekt. Przykładowo aby uzyskać intencjonalnie obłą bryłę budynku, w testach stosowano wymiennie kilkanaście pojęć, takich jak: blob, bubble, egg, sphere, bean, pebble, spheroid wraz z przymiotnikami: spherical, flattened, distorted, organic, blobby, blob-like, egg-shaped etc. Efekty były różne dla różnych narzędzi, np. Stable Diffusion XL najlepiej (tj. najbardziej zgodnie z oczekiwaniami pytającego) reagowało na sformułowanie spherical blob, podczas gdy inne narzędzia zdawały się je rozumieć błędnie.

Współczesne generatory AI nie radzą sobie także z interpretacją przyimków („w”, „nad”, „przed”, „za”) i liczebników. Pojęcia te często są ignorowane (rezultatem polecenia „narysuj 3 kolumny” będzie obraz z ich dowolną liczbą). Podawanie danych liczbowych – jak

odległości, liczba kondygnacji lub wymiary – nie daje więc pożądanych rezultatów. Jest to istotne ograniczenie współczesnych sieci neuronowych [12].

Generowany rezultat jest także wysoce zależny od proporcji kadru. Sieci będą produkować obrazy obiektów, które będą starać się „wypełnić” ten kadr. To znaczy, że jeśli wybrany zostanie kadr pionowy, to sieć będzie tworzyć budynki smukłe, wielokondygnacyjne, nawet jeśli w poleceniu jest wprost wskazany budynek „niski”, „parterowy”, „horyzontalny”. Ponieważ część ww. narzędzi ma ograniczenia w wyborze proporcji obrazu, jest to istotny czynnik ograniczający ich przydatność. Na przykład łatwo dostępny Bing Image Generator (oparty na DALL-e-3) generuje wyłącznie obrazy o proporcjach 1:1, co ma wyraźny wpływ na przydatność z punktu widzenia architekta.

Nie jest tu możliwe wskazanie wszystkich ograniczeń poleceń, ale należy wspomnieć jeszcze o jednym – większość narzędzi przyjmuje fotorealistyczny sposób przedstawienia jako domyślny (wyjątkiem jest DALL-e 3), inne sposoby przedstawiania wymagają zaznaczenia tego w opisie (np. obraz impresjonistyczny, rysunek odręczny, makieta etc.). Często takie doprecyzowania „odwracają uwagę” AI od głównego tematu, w rezultacie prowadząc do ignorowania zapisów dotyczących meritum.

## Efekt motyla

W przypadku szeroko sformułowanego polecenia rozrzut odpowiedzi jest bardzo duży. Jest to oczywista konsekwencja wspomnianych czynników. Ale nawet drobne zmiany promptu prowadzą do wykreowania radykalnie różniących się od siebie obrazów. Liczy się tu, jak wspomniano, kolejność występowania wyrazów (np. wielki biały budynek da inne rezultaty niż biały wielki budynek). Powoduje to, że uzyskanie konwergentnych rezultatów stanowi dużą trudność. Istotny wpływ na losowość wyników mają też inne parametry, takie jak np. CFG scale (classifier-free guidance) w Stable Diffusion i pochodnych, która decyduje o tym, jak bardzo sieć ma starać się przedstawić tokeny wskazane w prompcie kosztem spójności całości obrazu.

Ponieważ generowanie obrazów w trybie text-to-image zaczyna się od obrazu z losowym szumem, sam rodzaj tego szumu ma istotny wpływ na generowaną zawartość. Sam użytkownik ma bardzo ograniczoną możliwość kontrolowania tego szumu. Każdemu wzorcowi szumu przypisywany jest unikatowy identyfikator (tzw. seed), który pozwala po jego wpisaniu (i odtworzeniu pozostałych parametrów) odtworzyć identyczny obraz końcowy.

Użyta metoda próbkowania (sampling method) jest kolejnym z czynników zwiększających losowość wyników. Większość

narzędzi oferuje jedną, maksymalnie kilka metod generowania, ale SD oferuje obecnie 26.

Stable Diffusion ma także możliwość modyfikacji sieci przez zaawansowanych użytkowników (custom checkpoint, LORA i ich pochodne). Sieci takie są wytrenowane tak, aby generowały z wyższą jakością pewien konkretny rodzaj obrazów (np. fotorealistyczne portrety). Część z takich modyfikacji przeznaczona jest także do kreowania architektury (wnętrz, budynków lub np. miast). Tego typu modyfikacje są używane w sieciowych narzędziach dla architektów opartych na Stable Diffusion.

## Baza LAION-5B a tematyka architektoniczna

Reakcja sieci na polecenia, na ogólnym poziomie, zależy od wytrenowania sieci, a w szczególności od zestawu treningowego, jaki został użyty do jej uczenia. Większość narzędzi nie ujawnia, jaka liczba par obraz + opis była używana do szkolenia sieci oraz jakie to były obrazy, ale Stable Diffusion używa ogólnie dostępnego zbioru par obraz + opis „LAION-5B” [6], [7]. Należy przypuszczać, że część tych sieci, które nie podają źródeł, jest również trenowana na tym zbiorze, a pozostałe na analogicznych zbiorach o zbliżonej wielkości. LAION-5B to zbiór 5,85 miliarda obrazów. To zbiór bardzo losowy, uzyskany poprzez przeszukiwanie internetu (web scraping), a następnie przefiltrowanie i skatalogowanie wyników.

Z punktu widzenia architekta istotne jest, że fotografia architektoniczna jest stosunkowo niewielkim podzbiorem zbioru LAION-5B. Większość fotografii budynków znajdujących się w tym zbiorze to fotografia uliczna, reportażowa, zdjęcia archiwalne, a także okładki książek. Oznacza to także, że kluczowe dla uczenia sieci opisy nie zawierają z reguły terminów specjalistycznych, a zbiory obrazów powiązanych z takimi pojęciami są skromne (w skali całej bazy). Ma to bezpośrednie konsekwencje dla łatwości, z jaką tworzy się obrazy (np. łatwiej stworzyć obraz wiarygodnego starego miasta niż obraz budynku w stylu dekonstrukcji).

Podobna sytuacja dotyczy twórców. O ile Le Corbusier, F.L.Wright czy Z. Hadid i ich dzieła są dobrze reprezentowani w tej bazie, o tyle inni architekci, których nazwiska nie przebiły się do tzw. mainstreamu, są reprezentowani bardzo słabo albo wcale. W chwili obecnej nie ma narzędzi, które pozwoliłyby precyzyjnie ustalić, jakie to są liczby, ale ponieważ baza ta jest przeszukiwalna [8], można sprawdzić, jakie rezultaty są zwracane jako najbardziej poprawne oraz jak szybko zbiór odpowiedzi przestaje realnie odpowiadać zapytaniu. Przykładowo wyszukiwanie „Peter Behrens architecture” zwraca jako najlepsze wyniki wnętrza Hoechst AG, Behrensa,

ale obok prezentuje wielokrotnie okładkę książki biograficznej o Behrensie, a ok. 40 pozycji niżej zdjęcie fotela M. van der Rohe, a dalej rysunki Sant'elii i Het Ship de Klerka. W rezultacie polecenia „budynek w stylu Le Corbusiera” i „budynek w stylu Zaha Hadid” mogą być skuteczniejszymi triggerami dla sieci neuronowej trenowanej na LAION-5B niż odpowiednio „budynek modernistyczny” i „architektura organiczna”. Oczywiście użycie nazwisk w poleceniu może budzić wątpliwości etyczne, ale należy zaznaczyć, że nawet użycie konkretnych znanych nazwisk nie zawęży wyników tylko do jednego autora – raczej do grupy autorów, których nazwiska często występują razem w literaturze (np. w opisach stylów architektonicznych).

### Szybkość i liczba wygenerowanych wariantów

Pracę z AI cechuje duża szybkość generowania kolejnych obrazów. Czas wygenerowania pojedynczego obrazu w serwisie internetowym waha się od kilku sekund do kilku minut. W przypadku generowania obrazów lokalnie (używając SD) ten czas zależy od wielu czynników – rozdzielczości obrazu i wydajności karty graficznej – we wspomnianych testach wygenerowanie grupy 100 obrazów (SD XL, Sampling type: UniPC, 20 steps) w rozdzielczości 1200×768 pikseli zajęło ok. 7 s na obraz, na komputerze z GPU Nvidia RTX 4070 TI. W rzeczywistości więc szybkość generowania obrazów nie jest istotnym ograniczeniem w pracy z generatorami.

Łatwość, z jaką uzyskuje się kolejne modyfikacje, oraz konieczność przeformułowywania poleceń w wyniku niezyskiwania satysfakcjonujących wyników powoduje, że w trakcie rozwijania koncepcji powstają ogromne ilości obrazów. Przykładowo w trakcie testów wygenerowano łącznie ok. 1000 obrazów. Przeglądanie, dyskusja i analiza wyników oraz modyfikacja poleceń zdecydowanie zużywają większość czasu przeznaczoną na pracę z AI.

### Użycie obrazów zewnętrznych jako podstawy dla generowanych obrazów

Używanie opisu jako punktu startowego dla tworzenia architektury ma oczywiście ograniczenia, trudno tu o precyzyjne odzwierciedlenie zarówno wizji architekta, jak i rzeczywistych ograniczeń (otoczenia, wymogów prawnych etc.). Stable Diffusion oraz niektóre inne generatory pozwalają na użycie zewnętrznych obrazów jako „punktu startowego” dla obrazów generowanych przez AI. W takiej sytuacji obraz z szumem jest zastępowany przez obraz wczytany przez użytkownika. Zasadniczo powoduje to, że generowany obraz staje się, w pewnym zakresie, podobny do wczytanego obrazu. SD pozwala na różne sposoby interpretacji takich obrazów.



Rys. 1. Przykładowe wyniki pracy Stable Diffusion 1.5 w trybie image-to-image dla polecenia: "a photo of contemporary boxy minimalistic elongated narrow glass; concrete; wood; house cantilevered; long cantilever over a grassy hill slope. pine tree forest in the background, 8K, cinematic, intrinsic details, fog with patches of warm sunlight on the grass" Negative prompt: "columns, fence; thick concrete slabs"; Steps: 20, Sampler: Euler a, CFG scale: 7.5; źródło: g.w.

Podstawy: image-to-image traktuje obraz jako układ barwnych plam, które są następnie reinterpretowane zgodnie z opisem zawartym w poleceniu. Metoda ta działa więc najlepiej z obrazami przypominającymi zdjęcia. Mogą być to dowolne obrazy, więc także np. zdjęcia wykonanej makiety roboczej, rendering wykonany w programie do wizualizacji etc. Ten obraz może być bardziej lub mniej zmodyfikowany przez sieć neuronową. Decydują o tym dwa parametry: wspomniana już CFG scale oraz parametr Denoising strength. Ten drugi pozwala na ustalenie zakresu swobody interpretacji obrazu (wartość 0 daje zupełnie niezmienny obraz, wartość 1 daje obraz, który zupełnie nie przypomina źródłowego). Wartości pomiędzy 0,2 i 0,6 dają obrazy, które w pewnym zakresie przypominają obraz źródłowy. Taka metoda działania pozwala na generowanie wariantów obiektów architektonicznych w zakresie ich kompozycji i rozwiązań materiałowych.

Drugą techniką przydatną w proj. architektonicznym jest sketch-to-image. Wymaga ona użycia rozszerzenia Stable Diffusion o nazwie ControlNet. Jego funkcjonalność Sketch-to-image pozwala na stworzenie wizualizacji, bazując na rysunku kreskowym. W praktyce narzędzie to ma wiele ograniczeń. Szybkie scribe odręczne – z charakterystycznymi dla nich niedoskonałościami i manieryzmami (jak przeciąganie linii, poprawianie, kreskowanie) nie są dobrym źródłem do generowania

obrazów. Wszystkie te niedoskonałości zostaną zinterpretowane przez sieć jako istotne elementy rysunku. Prowadzi to do uzyskania efektów bardzo odległych od zamierzonych. Użycie rysunku kreskowego jako źródła wymaga więc odpowiedniej precyzji rysowania. De facto najlepsze efekty uzyskuje się poprzez użycie czarno-białych, kreskowych widoków z programów CAD, co oczywiście ogranicza zastosowanie tej techniki (lub wymusza zmianę tzw. workflow i rozpoczęcie modelowania na wstępnych etapach projektowania). Niemniej nawet wtedy generowane obrazy bywają w znacznym stopniu ułomne.

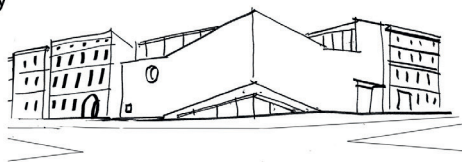
### Modyfikacja istniejących obrazów

Technika image-to-image może być rozszerzona o dodatkowe funkcjonalności zwane inpainting i outpainting. Inpainting pozwala na zaznaczenie fragmentu istniejącego obrazu i zastąpienie go fragmentem wygenerowanym (zgodnie z promptem). Technika outpainting pozwala na rozszerzenie istniejącego obrazu o wygenerowany fragment. Odpowiednik tych technik dostępny jest w wersji beta Adobe Photoshop pod nazwą „generative fill”. Techniki te umożliwiają (w pewnym zakresie) korygowanie istniejących obrazów (wygenerowanych przez AI lub nie). Pozwala to na edycję obrazów (np. wizualizacji) w finalnych fazach procesu projektowania.

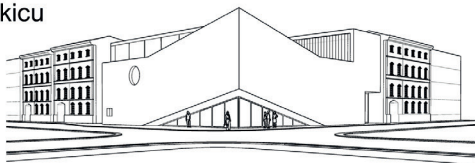
szybki szkic odręczny



skorygowany szkic odręczny



model 3d na podstawie szkicu



Rys. 2. Przykładowe wyniki pracy Stable Diffusion 1.5 w trybie sketch-to-image dla polecenia: "street view of contemporary theater with white sleek facade, glass entrance and glass roof, metal frames. visible intrinsic interior; warm light. located at crossroads in paris old city, architecture, 8K, high-res, soft light. morning haze, lawn close to camera. trees". Negative prompt: "windows"; Steps: 20, Sampler: UniPC sekcja górna – szybki szkic odręczny oraz przykładowe rezultaty interpretacji tego szkicu sekcja środkowa – skorygowany (uproszczony) szkic odręczny oraz przykładowe rezultaty interpretacji tego szkicu sekcja dolna – widok modelu 3d sporządzonego w prog. Sketchup oraz przykładowe rezultaty interpretacji tego widoku; źródło: g.w.

## Wnioski – możliwy zakres i ograniczenia zastosowania generatorów AI w projektowaniu architektonicznym

Przeprowadzone testy i analizy pozwalają na wskazanie potencjalnych obszarów zastosowania generatorów AI. Na obecnym etapie rozwoju narzędzi AI istnieje techniczna możliwość wykorzystania ich w kilku obszarach:

– Inspiracyjnym, na wstępnych etapach projektowania [9], przy użyciu technik text-to-image poprzez tworzenie obrazów, które mogą służyć jako inspiracje lub referencje. Ograniczeniem jest tu obecnie głównie trudność uzyskania wyników, które obejmują wszystkie istotne (wymienione w prompcie) aspekty projektu. Niemniej przy odpowiednio swobodnym podejściu do rezultatów

mogą one wskazać kierunki projektowe, które zupełnie nie były rozważane na tym etapie. Ten sposób podejścia do użycia AI jest najbardziej oczywisty. Jest on często prezentowany w artykułach i we wspomnianych filmach instruktażowych.

– Dyskusji wariantów poprzez reinterpretację szkiców / uproszczonych modeli 3D za pomocą technik sketch-to-image i image-to-image. Obecnie największym ograniczeniem techniki sketch-to-image jest wymagana precyzja szkicu (w przeciwnym razie wyniki będą nieprzydatne). To oczywiście w istotny sposób ogranicza rzeczywistą przydatność tego narzędzia. Żadne znane autorowi publikacje nie opisują przydatności techniki sketch-to-image i image-to-image jako narzędzia do generowania i dyskusji wariantów. Istniejące podejścia do tych technik skupiają się na

„skróceniu drogi” pomiędzy wstępną wizją (szkicem) a „finalnym renderingiem” [10], co okazuje się fałszywą obietnicą ze względu na brak kontroli nad finalnym rezultatem.

– Na późniejszych etapach możliwe jest wykorzystanie techniki image-to-image do generowania wariantów z reprezentacji fotorealistycznych. Przy niskich wartościach współczynnika Denoising strength pozwala to na generowanie np. wariantów elewacji czy umiarkowanych modyfikacji brył. Podstawowym ograniczeniem tutaj jest ponownie zaawansowanie prac, przy którym skorzystanie z tego narzędzia jest możliwe. Z reguły w procesie projektowania tego typu analizy wykonywane są stosunkowo wcześniej (przed powstaniem wizualizacji). Tego typu podejście nie jest reprezentowane w istniejących publikacjach.

– Na etapie finalnym: do optymalizacji przekazu (storytelling), poprzez korygowanie wizualizacji, tak aby odpowiadały one lepiej intencji projektanta. Tego typu podejście jest w ograniczonym zakresie prezentowane w istniejących publikacjach.

Największy, nieodkryty jeszcze, potencjał generatorów AI w architekturze wydaje się kryć w generowaniu wariantów. Wariantowanie to jedna z najbardziej skutecznych metod rozwoju projektu. Niemniej z obserwacji autora jako nauczyciela akademickiego wynika, że wariantowanie wydaje się przeżywać regres wśród młodych projektantów (głównie w wyniku woli użycia narzędzi typu BIM od wstępnych faz projektowania i związanego z tym nakładu pracy).


Sieć neuronowa może w pewnym zakresie zastąpić współpracownika, który prezentuje alternatywną wizję rozwiązania. Pozwala to na porównanie wariantów i w rezultacie pogłębienie zrozumienia projektu.

Użycie AI może także pozwolić do przetrwania tzw. fiksacji projektowej polegającej na powracaniu przez projektanta do ograniczonego zbioru rozwiązań. Istotnym ograniczeniem jest tu pewna „pytkość” propozycji generowanych przez AI – mogą być one wizualnie atrakcyjne, ale jednocześnie pozbawione logiki funkcjonalnej / konstrukcyjnej etc. [11].

Skuteczne użycie tych narzędzi jest obecnie trudne. Problemатyczne są zarówno: potencjalne zmiany w metodyce pracy, przewlekłość eksperymentowania w celu uzyskania przydatnych wyników, bardzo ograniczona kontrola nad rezultatami, konieczność selekcji z ogromnej liczby obrazów oraz ich nadmierna precyzja [11] (na etapie pracy nad ogólnymi założeniami projektu).

## BIBLIOGRAFIA

- [1] Goodfellow I.J. et al., (2014), Generative Adversarial Networks, arXiv:1406.2661 [stat.ML].
- [2] Nichol A. et al., (2021), GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models, arXiv:2112.10741 [cs.CV].
- [3] Rombach R., et al., (2021), High-Resolution Image Synthesis with Latent Diffusion Models, arXiv:2112.10752 [cs.CV].

Tab. 1. Lista narzędzi AI przetestowanych w trakcie przygotowania artykułu				
	Nazwa narzędzia [a-z]	Nazwa serwisu	Opis	Przykładowy rezultat dla takiego samego polecenia
1	Adobe Firefly	<a href="https://firefly.adobe.com/">https://firefly.adobe.com/</a>	Serwis o szerokim zakresie zastosowania w tworzeniu grafiki i mediów. Jego elementem jest generator AI	
2	AI-Architect	<a href="https://ai-architect.net/">https://ai-architect.net/</a>	Generator obrazów AI na podstawie zaimportowanych rysunków (szkiców perspektywicznych lub widoków modelu 3D) przykładowy rezultat na podstawie szkicu	
3	Arko	<a href="https://arko.ai/">https://arko.ai/</a>	Web-based plugin do programów do modelowania architektonicznego (Sketchup, Revit, Rino). Służy do renderowania modelu w oparciu o generator obrazów AI przykładowy rezultat na podstawie modelu 3d	
4	Bing Image Creator (oparty na DALL-e 3)	<a href="http://www.bing.com/images/">www.bing.com/images/</a>	Generator obrazów AI ogólnego zastosowania	
5	Canva	<a href="https://www.canva.com/">https://www.canva.com/</a>	Serwis o szerokim zakresie zastosowania w tworzeniu grafiki i mediów. Jego elementem jest generator AI	
6	DALL-e 2	<a href="https://openai.com/dall-e-2">https://openai.com/dall-e-2</a>	Generator obrazów AI ogólnego zastosowania. Serwis obecnie wyłączony. Obrazy wygenerowano przed wyłączeniem	
7	DALL-e 3	<a href="https://chat.openai.com">https://chat.openai.com</a>	Generator obrazów AI ogólnego zastosowania zintegrowany z usługą Chat GPT 4. Wyłącznie płatny dostęp Nie użyto w badaniu ze względu na ograniczenia licencyjne	
8	Fotor	<a href="https://www.fotor.com/">https://www.fotor.com/</a>	Serwis o szerokim zakresie zastosowania w tworzeniu i korekcie grafiki z użyciem AI. Jego elementem jest generator AI reklamowany również jako narzędzie do generowania architektury	
9	Img 2 Go	<a href="https://www.img2go.com">https://www.img2go.com</a>	Generator obrazów AI ogólnego zastosowania	
10	Imagine	<a href="https://www.imagine.art">https://www.imagine.art</a>	Generator obrazów AI ogólnego zastosowania	

[4] <https://www.g2.com/categories/ai-image-generators> (dostęp: 20.11.2023).  
 [5] Podell D. et. al., (2023) SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis, <https://doi.org/10.48550/arXiv.2307.01952>.  
 [6] Schuhmann Ch. et. al. (2022), LAION-5B: An open large-scale dataset for training next generation image-text models\*, arXiv:2210.08402 [cs.CV].  
 [7] <https://laion.ai/about/> (dostęp: 20.11.2023).  
 [8] <https://haveibeentrained.com> (dostęp: 21.11.2023).  
 [9] Gil A., (2005), Szkic, makieta, model komputerowy jako narzędzia pracy twórczej architekta, Praca doktorska, Wydział Architektury Politechniki Śląska, <http://repolis.bg.polsl.pl/dlibra/publication/edition/4916>.  
 [10] <https://www.youtube.com/watch?v=D2cybt2lf2Y>, How to use AI Rendering in Architecture (dostęp: 21.11.2023).  
 [11] Gil A., (2024), Analiza metod pracy projektantów korzystających z generatorów obrazów opartych na sztucznej inteligencji

w kreatywnych fazach projektowania architektonicznego, w tym numerze Builder Science, DOI: 10.5604/01.3001.0054.3433 [12] Raj Sanjay Shah R. et. al., (2023), Human Behavioral Benchmarking: Numeric Magnitude Comparison Effects in Large Language Models, arXiv:2305.10782. [cs.AI].

**DOI: 10.5604/01.3001.0054.3459**

**PRAWIDŁOWY SPOSÓB CYTOWANIA**  
 Gil Adam, 2024, Ograniczenia współczesnych generatorów obrazów opartych o AI w procesie projektowania architektonicznego, „Builder” 03 (320). DOI: 10.5604/01.3001.0054.3459

#### STRESZCZENIE:

Wraz z wprowadzeniem narzędzi CAD, a potem BIM, architekci zaczęli dostrzegać, że narzędzia, których używają, mają wpływ na ich twórczość. Obecnie architekci zaczynają używać narzędzi AI umożliwiających eksplorację koncepcji projektowych poprzez generowanie obrazów. Te narzędzia, zwane generatorami obrazów AI, generują obrazy na podstawie opisów tekstowych oraz dowolnych wczytanych obrazów, w tym szkiców lub zdjęć. Na rynku dostępne są zarówno narzędzia przeznaczone bezpośrednio do projek-

11	Limewire	<a href="https://limewire.com">https://limewire.com</a>	Serwis o szerokim zakresie zastosowania w tworzeniu i korekcie grafiki z użyciem AI. Jego elementem jest generator AI	
12	Look X	<a href="https://www.lookx.ai/">https://www.lookx.ai/</a>	Generator obrazów AI ogólnego zastosowania	
13	Midjourney	<a href="https://www.midjourney.com/">https://www.midjourney.com/</a>	Generator obrazów AI ogólnego zastosowania. Wyłącznie płatny dostęp Nie użyto w badaniu ze względu na ograniczenia licencyjne	
14	Picsart	<a href="https://picsart.com/">https://picsart.com/</a>	Serwis o szerokim zakresie zastosowania w tworzeniu i korekcie grafiki z użyciem AI. Jego elementem jest generator AI	
15	Prome AI	<a href="https://www.promeai.com/">https://www.promeai.com/</a>	Serwis o szerokim zakresie zastosowania w tworzeniu koncepcji z użyciem AI. Jego elementem jest generator AI reklamowany również jako narzędzie do generowania architektury	
16	Stable Diffusion 1.5	Local host / komputer użytkownika	Generator obrazów AI ogólnego zastosowania	
17	Stable Diffusion XL	Local host / komputer użytkownika	Generator obrazów AI ogólnego zastosowania	
18	Stable Diffusion XL	<a href="https://stablediffusionweb.com/">https://stablediffusionweb.com/</a>	Generator obrazów AI ogólnego zastosowania	
19	Stable Diffusion XL	<a href="https://beta.dreamstudio.ai/">https://beta.dreamstudio.ai/</a>	Generator obrazów AI ogólnego zastosowania	
20	Veras	<a href="https://www.evolvealab.io/veras/">https://www.evolvealab.io/veras/</a>	Web-based plugin do programów do modelowania architektonicznego (Sketchup, Revit, Rhino). Służy do renderowania modelu w oparciu o generator obrazów AI	
21	Wepik	<a href="https://wepik.com/">https://wepik.com/</a>	Generator obrazów AI ogólnego zastosowania	

towania architektonicznego, jak i narzędzia o charakterze ogólnym, służące do generowania dowolnych obrazów, od fotorealistycznych po malarskie i graficzne. Wszystkie te narzędzia są we wstępnych fazach rozwoju. Artykuł analizuje przydatność i ograniczenia tych narzędzi z perspektywy architekta, biorąc pod uwagę stan technologii na koniec 2023 roku.

**SŁOWA KLUCZOWE:**

AI w projektowaniu architektonicznym, AI Image Generators, diffusion models, stable diffusion, fazy kreatywne projektowania

**ABSTRACT:**

**LIMITATIONS OF MODERN AI-BASED IMAGE GENERATORS IN THE ARCHITECTURAL DESIGN PROCESS.** With the introduction of CAD and then BIM tools, architects have begun to recognize that the tools they use have an impact on their creations. Today, architects are beginning to use AI tools that enable them to explore design concepts by generating images. These tools, called AI image generators, generate images based on text descriptions and any loaded images, including sketches or photos. There are tools on the market that are designed

directly for architectural design, as well as general-purpose tools for generating any image, from photorealistic to painterly and graphic images. All of these tools, are in the preliminary stages of development. This article examines the usefulness and limitations of these tools from an architect's perspective, taking into account the current state of technology at the end of 2023.

**KEYWORDS:**

AI in architectural design, AI Image Generators, diffusion models, stable diffusion, creative phases of design