

## **Od Sztuki do Nauki: Redukując Ograniczenia Generatywnych Modeli Dyfuzyjnych**

Modele generatywne stanowią przełom w sztucznej inteligencji, porywając badaczy, programistów i społeczeństwo swoją bezprecedensową zdolnością do tworzenia treści odzwierciedlających złożoność danych generowanych przez ludzi. Jest to szczególnie widoczne w licznych zastosowaniach kreatywnych, takich jak przekształcanie tekstu na obrazy czy wideo, gdzie modele generatywne tworzą zdumiewające wizualizacje na podstawie krótkich opisów tekstowych. Te imponujące wyniki są możliwe dzięki szczególnemu celowi modelowania generatywnego, którym jest imitacja rozkład danych treningowych. W kontekście generowania obrazów rozkład ten jest określany jako zbiór obrazów treningowych (zazwyczaj kolosalnych rozmiarów – ponad 5 miliardów przykładów). Model uczy się więc, jak tworzyć realistyczne obrazy, układając piksele w taki sposób, by ostateczny wynik był podobny do danych treningowych, a nie kombinacją niepowiązanych losowych wartości.

Oprócz ekscytujących możliwości generowania nowych śmiesznych obrazów kotów, te same metody generatywne mogą być stosowane w programowaniu naukowym. Możliwość generowania nowych instancji podobnych do już zebranych danych stwarza unikalne możliwości w obszarach takich jak odkrywanie leków i nowych materiałów czy fizyka wysokich energii. W tych zastosowaniach, modele generatywne mogą być stosowane jako techniki symulacyjne, za pomocą których można potwierdzać teoretyczne hipotezy lub przyspieszać proces projektowania nowych eksperymentów.

Niemniej jednak, pomimo korzystania z podobnych metod, te dwa scenariusze rządzą się swoimi prawami i są kierowane różnymi celami. W zastosowaniach kreatywnych, głównym celem modelu generatywnego jest stworzenie przyjemnego nowego przykładu podobnego do obrazów stworzonych przez ludzi. W przeciwieństwie do tego, w zastosowaniach naukowych i symulacjach główny nacisk kładziony jest na precyzyjne modelowanie rzeczywistego rozkładu danych. Różnica między tymi dwoma celami jest oczywista. Na przykład, niedoszacowanie częstości występowania czerwonych kotów w generatorze śmiesznych obrazów jest mniej istotnym problemem niż pominięcie fizycznych zjawisk czy subtelnych struktur biochemicznych w naukowych zastosowaniach. Jednocześnie, nietypowe generacje spoza domeny danych treningowych mogą prowadzić do nowego interesującego stylu ilustracji, ale także do fizycznie niemożliwych wniosków, gdy używane są do walidacji eksperymentów w fizyce wysokich energii.

W ramach tego projektu proponujemy zbadanie czego brakuje aktualnym modelom generatywnym do tego by wykorzystać je w obliczeniach naukowych. Naszym głównym celem będzie więc ocena wiarygodności modeli generatywnych opartych na dyfuzji w zadaniu modelowania prawdopodobieństwa danych. Planujemy najpierw przeanalizować ograniczenia najnowszych technik, skupiając się na aspektach takich jak zapamiętywanie przykładów treningowych i ograniczone pokrycie rozkładu danych treningowych. Na podstawie tej analizy opracujemy nowe metody, które umożliwią dodanie brakującej wiedzy do istniejącego modelu. Ponadto planujemy przyjrzeć się bliżej błędnie wygenerowanym przykładom – tzw. halucynacjom. Opracujemy metodę ich identyfikacji i oduczania z już wytrenowanego modelu. Dodatkowo, przeanalizujemy strategię aktualizacji wytrenowanego modelu w oparciu o dodatkowe dane, w taki sposób, by nie kompromitować jednocześnie jego wydajności w kontekście modelowania poprzednich danych.