Spis treści

[Wstęp 4](#_Toc136160180)

[1. Sztuczna inteligencja a twórczość kreatywna 4](#_Toc136160181)

[1.1. Potencjalne możliwości zastosowania sztucznej inteligencji w zadaniach kreatywnych 4](#_Toc136160182)

[1.2. Dotychczasowe osiągnięcia sztucznej inteligencji w zakresie zadań kreatywnych 6](#_Toc136160183)

[2. Sieci neuronowe 11](#_Toc136160184)

[2.1 Klasyczne sieci neuronowe 11](#_Toc136160185)

[2.2 Rekurencyjne sieci neuronowe 16](#_Toc136160186)

[2.2.1 LSTM i GRU 19](#_Toc136160187)

[3. Stworzenie generatora piosenek 22](#_Toc136160188)

[3.1. Budowa modelu 22](#_Toc136160189)

[3.2. Implementacja modelu w aplikacji Flask 22](#_Toc136160190)

[3.3. Wyniki modelu 22](#_Toc136160191)

[4. Dalszy możliwy rozwój AI w dziedzinach kreatywnych 22](#_Toc136160192)

[Bibliografia 24](#_Toc136160193)

[Spis tabel i rysunków 27](#_Toc136160194)

# Wstęp

# Sztuczna inteligencja a twórczość kreatywna

# Potencjalne możliwości zastosowania sztucznej inteligencji w zadaniach kreatywnych

Rozważania na temat potencjalnych możliwości sztucznej inteligencji są obecne od czasów gdy zaawansowanie techniczne nie było nawet blisko poziomu współczesnego. Słynny test Turinga został sformułowany już w 1950 roku przez matematyka Alana Turinga stanowi wyraz ciekawości jak rozwinie się dalej sztuczna inteligencja. Opisanie potencjalnych możliwości sztucznej inteligencji w zakresie zadań kreatywnych jest szczególnie złożone i zróżnicowane ze względu na trudności w zakresie używania jednoznacznych definicji, oceniania tych możliwości obiektywnie i bez względu na dotychczasowe uprzedzenia.

Zgodnie ze stwierdzeniem Ade Lovelace sztuczna inteligencja „nie ma żadnych predyspozycji do zapoczątkowania czegokolwiek. Może zrobić [tylko] wszystko, co wiemy, jak mu zlecić wykonanie”[[1]](#footnote-1). W celu weryfikacji tej tezy postawionej 200 lat temu należy przyjrzeć się procesowi tworzenia kreatywnego stosowanego przez człowieka i zadać pytanie w jakim stopniu współczesne maszyny są w stanie podążać tym procesem, bądź czy istnieje w przyszłości możliwość wpasowania się w ten proces,

Boden (1998) wskazuje na trzy rodzaje kreatywności. Pierwszy rodzaj, combinational creativity, stanowi kombinacje znanych pomysłów. Drugi rodzaj o nazwie exploratory creativity opisuje powstawanie pomysłów poprzez eksplorację ustrukturyzowanych pomysłów, a więc jest to kreowanie nowego dzieła w zakresie danego nurtu. Ostatni rodzaj kreatywności, transformational creativity, polega na powstaniu przekształceń, w celu mogły powstania nowych struktur, które wcześniej nie mogły powstać, a więc są to wydarzenia zapoczątkowujące nowe nurty, wzorce tworzenia. Sztuczna inteligencja odnosi największe sukcesy w zakresie exploratory creativity ze względu na stosunkową łatwość działaniu w zakresie tej kreatywności. Do tego rodzaju kreatywności można zaliczyć narzędzia służące do tworzenia dzieł w danym stylu oparte na wcześniejszym dostarczeniu wystarczającej ilości danych w stylu w jakim algorytm powinien się nauczyć tworzyć. Combinational creativity z kolei pojawia się w systemach tworzących żarty lub odpowiadających w dowcipny sposób. W kwestii transformational creativity AI zdaje się wciąż nie być w stanie w pełni realizować tej idei. Sztuczna inteligencja nie może poszczycić się kreatywnymi wydarzeniami godnymi porównania z powstaniem kubizmu czy bluesa. Braki w kwestii transformational creativity stanowią odzwierciedlenie często podnoszonych argumentów głoszących zdolność AI jedynie do naśladowania kreatywności, a nie prawdziwego tworzenia transformacyjnych dzieł.

Wspomniany test Turinga polegał na przeprowadzeniu rozmowy przez sędziego z człowiekiem oraz maszyną. W sytuacji gdy sędzia nie jest w stanie jednoznacznie odróżnić maszyny od człowieka można mówić o przejściu maszyny przez test Turinga, co oznacza, że maszyna jest w stanie dostatecznie dobrze imitować zachowanie człowieka. Przejście testu Turinga przez maszynę nie stanowi jednak idealnego wskaźnika inteligencji i stopnia rozwoju maszyny. Jest możliwe aby maszyna wyspecjalizowana w jednym aspekcie nie będzie w stanie prowadzić rozmowy w sposób przekonujący na tematy zaproponowane przez ludzkiego sędziego. Taki rodzaj maszyny wyspecjalizowanej w wykonywaniu ograniczonej liczby zadań nazywany jest narrow AI, natomiast maszyna zdolna do wykonania każdego intelektualnego zadania możliwego do wykonania przez człowieka lub zwierzę należy do konceptu artificial general intelligence (AGI). Odpowiednikiem testu Turinga w przypadku tworzenia kreatywnych dzieł sztuki byłoby przeprowadzenie testu polegającego na osądzeniu przez niezależnego sędziego które z dzieł zostało stworzone przez człowieka, a które przez sztuczną inteligencję. W przypadku takiego rodzaju testu możliwe byłoby rzetelne zbadanie faktycznych możliwości danych maszyn w kwestii tworzenia złożonych dzieł sztuki bez obciążenia uprzedzeniami bądź nadmiernym optymizmem dotyczącym zastosowania AI.

# Dotychczasowe osiągnięcia sztucznej inteligencji w zakresie zadań kreatywnych

Sztuka generowana przez sztuczną inteligencję zbiera w ostatnich czasach dużo uwagi. Dostępność generatorów obrazów czy piosenek jest coraz większa, a dzieła generowane przez sztuczną inteligencję otrzymują nawet nagrody na konkursach. Poniżej opisane zostały wybrane przykłady implementacji AI w zakresie tworzenia sztuki.

DALL-E to narzędzie przedstawione w 2021 roku pozwalające tworzyć realistyczne obrazy na podstawie opisu w języku naturalnym. Jedyne kroki niezbędne do wykonania aby wytworzyć taki obraz to dostarczenie krótkiego opisu pożądanego obrazu. Nie jest niezbędna umiejętność pisania kodu, co sprawia że rozwiązanie to jest dostępne również dla osób bez wiedzy technicznej na temat algorytmów stojących za narzędziem. Sam kod na którym zbudowany został DALL-E jest nieznany, ponieważ jego twórca Open AI nie upublicznia go do ogólnego dostępu. Jak wskazuje Gordon (2022) DALL-E nie radzi sobie doskonale z przekształcaniem skomplikowanych opisów w obrazy. Ponadto, problematyczne jest poprawne komponowanie danych obiektów i adekwatne odzwierciedlanie relacji pomiędzy nimi. Przykładowo, gdy użytkownik chce otrzymać obraz czerwonej ciężarówki i zielonego domu jest pewne prawdopodobieństwo otrzymania zielonej ciężarówki i czerwonego domu. Przy dalszym ulepszaniu narzędzia ma ono potencjał stać się nie tylko generatorem obrazów artystycznych, lecz także zgodnie z Adams, Busch, Truhn, Makowski (2023) stanowi obiecujące narzędzie w radiologii.



Rysunek 1. Obraz wygenerowany przez Floris Groesz z Dall-E software “Kermit the frog painted by Munch”

Symbolicznym odzwierciedleniem postępu jaki dokonał się w zakresie generowania przez AI sztuki jest zwycięstwo obrazu wyprodukowanego przez sztuczną inteligencję w konkursie artystycznym Colorado State Fair’s w 2022 roku. Obraz ten został wygenerowany za pomocą Midjourney – programu zmieniającego linię tekstu w grafikę. Zwycięstwo wzbudziło kontrowersję ze względu na fakt, że wygraną przyznawano Jasonowi M. Allenowi, który wygenerował obraz za pomocą AI.



Rysunek 2. Obraz wygenerowany przez Jasona Allena za pomocą AI - “Théâtre D’opéra Spatial”

Możliwość generowania piosenek została zaimplementowana między innymi w Chacie GPT. Ponadto, użytkownik jest w stanie zdefiniować typ muzyczny, klimat, czy inną wybraną cechę pożądanej piosenki. Kwestią dyskusyjną jest jednak jakoś generowanych tekstów. Muzyk Nick Cave swoją opinię o piosence wygenerowanej „w stylu Nicka Cave”:

„ Verse 1: In the depths of the night, I hear a call

A voice that echoes, through the hall

It's a siren's song, that pulls me in

Takes me to a place, where I can't begin

Chorus: I am the sinner, I am the saint

I am the darkness, I am the light

I am the hunter, I am the prey

I am the devil, I am the saviour

Verse 2: I walk the streets, with my head held high

I've got my demons, but I'll make them fly

I'll dance with the devil, and I'll play his game

I'll be the one, to call his name (…)”[[2]](#footnote-2)

wyraził się w następujący sposób: „Z całą miłością i szacunkiem na świecie, ta piosenka to bzdury, groteskowa kpina z tego, co to znaczy być człowiekiem, i cóż, nie bardzo mi się to podoba”.[[3]](#footnote-3) Opinia ta wywodzi się z przekonania Cave’a oraz wielu przedstawicieli świata muzycznego, zgodnie z którym sztuczna inteligencja jest zdolna jedynie odtwórczo i nieudolnie naśladować dzieła ludzkie. Nick Cave postuluje, że ze względu na fakt, że sztuczna inteligencja nie doświadcza własnego cierpienia i radości nie jest w stanie wzbić się na wyższy poziom artystyczny. Teksty piosenek są pisane w wyniku realnych ludzkich przeżyć, co według Cave’a stanowi barierę dla AI. Poziom artystyczny wygenerowanych przez AI piosenek jest sprawą trudną do rozstrzygnięcia obiektywnie i niezwykle trudne jest zweryfikowanie tez postawionych przez Cave’a. Samo narzędzie wciąż jest rozwijane i z pewnością z czasem będzie ulegać ulepszeniu, ze względu na duże zainteresowanie mediów i inwestorów. W momencie powstawania tej pracy, a więc kilka miesięcy po publikacji artykułów z opinią Nicka Cave’a o piosence wygenerowanej przez AI w stylu jego twórczości wynik ponowienia tego zapytania w Chacie GPT daje następujące rezultaty:

„(Verse 1)

In the heart of darkness, where shadows collide,

A solitary figure, he walks with wounded pride.

His voice like thunder, echoing through the night,

With words that pierce the soul, filling it with fright.

(Chorus)

Oh, the devil's in the details, and the truth is hard to find,

In this world of sorrow, where the blind lead the blind.

But he stands tall and fearless, unyielding in his quest,

A troubadour of darkness, he's Nick Cave at his best.

(Verse 2)

In the depths of despair, where despair finds a home,

He weaves tales of longing, of love and of the unknown.

His piano weeps, while his guitar screams and moans,

A symphony of anguish, in a world all his own.

(Chorus)

Oh, the devil's in the details, and the truth is hard to find,

In this world of sorrow, where the blind lead the blind.

But he stands tall and fearless, unyielding in his quest,

A troubadour of darkness, he's Nick Cave at his best.”[[4]](#footnote-4)

Poziom artystyczny powyższego tekstu oczywiście jest ponownie kwestią sporną, lecz zdecydowanie można zauważyć większą złożoność tego tekstu w porównaniu z tekstem cytowanym w artykułach parę miesięcy temu. Z tego względu dalszy rozwój Chata GPT oraz innych generatorów piosenek jest szczególnie ciekawy.

Sukcesy sztuki generowanej przez sztuczną inteligencję oraz łatwa dostępność do niej dostępność powoduje niepewność w środowisku artystycznym. Kwestie związane z przyszłością artystów oraz zapotrzebowanie na ich w porównaniu z generowaną przez AI dużo bardziej kosztowną sztukę pozostają nadal otwarte. Obecna sytuacja stanowi nie tylko wyzwanie dla artystów, lecz w obliczu potrzeby opracowania spójnego podejścia oraz zasad, także dla systemu prawnego i instytucji kulturalnych.

# Sieci neuronowe

Poniższa część pracy stanowi teoretyczne wytłumaczenie mechanizmów funkcjonowania sztucznych sieci neuronowych. Szczególna uwaga została poświęcona na przedstawienie rekurencyjnych sieci neuronowych, które zostały użyte w modelu przygotowanym na potrzeby tej pracy.

# Klasyczne sieci neuronowe

Historia sztucznych sieci neuronowych sięga już 80 lat. W latach 40 Warren McCulloch i Walter Pitts opracowali budowę neuronu, która do dzisiaj jest używana w budowie sztucznych sieci neuronowych. Z kolei pierwsza działająca dwuwarstwowa sieć neuronowa (Percepton) została zbudowana 1958 i służyła do rozpoznawania znaków. Sztuczne sieci neuronowe dzisiaj znajdują zastosowanie w wielu dziedzinach, takich jak rozpoznawanie obrazów, zagadnienia klasyfikacyjne, klastrowanie czy rozpoznawanie wzorców. Rekurencyjne sieci neuronowe, użyte w modelu zamieszczonym w dalszej części pracy, odnajdują zastosowanie także w innych gałęziach NLP takich jak analiza sentymentu, tłumaczenia, rozpoznawanie mowy. Poniżej opisany został mechanizm funkcjonowania sieci neuronowych, z wyszczególnieniem rekurencyjnych sieci neuronowych (RNN) oraz przedstawione zostały argumenty przemawiające za użyciem tej metody przy stworzeniu generatora piosenek.

Zainspirowane działaniem ludzkiego mózgu sztuczne sieci neuronowe są serią algorytmów, które mają za zadanie modelować złożone zależności. Ich budowa składa się z warstw neuronów, które połączone są między sobą synapsami. Neuron odbierając sygnały wejściowe, wykonuje obliczenia, a następnie generuje sygnał wyjściowy. Formalnie ujmując, każdy neuron ma *I* liczbę wejść *xi* i jedno wyjście *y*. Do każdego wejścia xi przypisana jest waga *wi,,* oraz dodatkowo może pojawić się parametr *w0* odpowiadający za bias.Obliczenia odbywające się wewnątrz neuronu składają się na zsumowanie sygnałów wejściowych zmodyfikowanych przez nadanie im wag

*,*

a następnie zastosowanie na otrzymanym wyniku funkcji aktywacji *f(a)* decydującej o ewentualnym przekazaniu informacji dalej.

Obraz zawierający krąg, linia, diagram, design

Opis wygenerowany automatycznie

Rysunek 3. Pojedyńczy neuron. Źródło: MacKay, D., J.C. (2003). Information Theory, Inference, and Learning Algorithms. Cambridge University Press, USA.

Funkcja aktywacji może przybierać różne formy. Popularnymi przykładami takich funkcji są:

1. Liniowa funkcja:

*Obraz zawierający tekst, linia, Wykres, diagram

Opis wygenerowany automatycznie*

Rysunek 4. Liniowa funkcja aktywacji. Źródło: opracowanie własne

II. Sigmoidalna funkcja unipolarna:

*Obraz zawierający linia, Wykres, diagram, zrzut ekranu

Opis wygenerowany automatycznie*

Rysunek 5. Sigmoidalna unipolarna funkcja aktywacji. Źródło: opracowanie własne

III. Tangens hiperboliczny:

*Obraz zawierający tekst, linia, Wykres, diagram

Opis wygenerowany automatycznie*

Rysunek 6. Funkcja aktywacji tangens hiperboliczny. Źródło: opracowanie własne

IV. ReLu:

*Obraz zawierający linia, diagram, Wykres, tekst

Opis wygenerowany automatycznie*

Rysunek 7. Funkcja aktywacji ReLu. Źródło: opracowanie własne

Następnie, analogicznie do biologicznie występujących synaps, synapsa w sztucznej sieci neuronowej może transmitować informację do kolejnego neuronu. Ostatecznie, sieć neuronowa składa się z warstwy wejściowej odpowiedzialnej za zbieranie danych, z pewnej liczby warstw ukrytych odpowiedzialnych za trenowanie sieci oraz warstwy wyjściowej zwracającej wynik.

Dostosowanie się wag oraz innych parametrów odbywa się w trakcie procesu uczenia. Pierwsze podejście do trenowania sieci neuronowej o nazwie propagacja w przód (ang. forward propagation) polega na początkowym losowaniu wag, a następnie procesowaniu danych wejściowych tylko w jednym kierunku. Innym podejściem jest zastosowanie propagacji wstecznej (ang. backpropagation) opierające się na metodzie gradientu. Propagacja wsteczna polega na minimalizacji funkcji kosztu poprzez dostosowanie wag przy użyciu gradientu. Funkcja kosztu *C* jest definiowana przez twórcę sieci neuronowej, może być to przykładowo błąd średniokwadratowy, kategoryczna entropia krzyżowa lub rzadka kategoryczna entropia krzyżowa. Dodatkowo wyznaczany jest współczynnik uczenia się *δ*, który determinuje jak duży wpływ ma gradient, a więc determinuje tempo uczenia. Schemat działania propagacji wstecznej rozpoczyna się od losowego ustalenia wartości wag *w* i biasu *b*. Następnie dla każdego neuronu jest obliczany błąd, a błędy następnie są przenoszone do poprzednich warstw, a więc wartości wyjściowe neuronów są porównywane z wartościami wejściowymi. Następnie wagi oraz bias są zmieniane aby zmniejszyć wartość funkcji kosztu *C*, a proces kończy się gdy funkcja ta zostanie zminimalizowana. Formalnie ujmując, algorytm można zapisać w postaci:

Przy złożonych i starannie wytrenowanych sieciach neuronowych dla większości zagadnień aproksymacyjnych istnieje sieć neuronowa stanowiąca rozwiązanie problemu w dowolnym stopniu dokładności. Wynika to bezpośrednio z twierdzenia Kołmogorowa, mówiącego o tym, że sieci wielowarstwowe są w stanie dowolnie dokładnie aproksymować funkcje rzeczywiste przyjmujące wartości w przedziale [0,1]. Wytrenowanie dobrze działającej sieci neuronowej jednak jest bardziej prawdopodobne przy zastosowaniu dobranych warstw, starannym dobraniu hiper parametrów, dużym i zbilansowanym zbiorze danych oraz długim czasie trenowania wymagającym odpowiednio dużych możliwości obliczeniowych oraz pamięci komputera.

# Rekurencyjne sieci neuronowe

Klasyczne sieci neuronowe opierają się na założeniu, że wejścia i wyjścia są od siebie niezależne. Rekurencyjna sieć neuronowa jest rodzajem sztucznej sieci neuronowej, w której połączenia pomiędzy danymi węzłami mogą tworzyć cykl powstający poprzez wpływanie danych wyjść na kolejne wejścia. Zjawisko to nazywane jest sprzężeniem zwrotnym. Ze względu na swoją specyfikację RNN są często używane przy rozpoznawaniu mowy, nauce struktur gramatycznych czy komponowaniu muzyki. Dokładniejszy opis zasad działa rekurencyjnych sieci neuronowych zostanie przedstawiony poniżej.

Rekurencyjna sieć neuronowa składa się analogicznie do klasycznej sieci neuronowej z neuronów, lecz w przypadku sieci rekurencyjnej dane wyjściowe mogą mieć wpływ na następne dane wejściowe do tych samych węzłów.

Obraz zawierający szkic, diagram, linia, origami

Opis wygenerowany automatycznie

Rysunek 8. a) jednokierunkowa sieć neuronowa b) Rekurencyjna sieć neuronowa. Źródło: MacKay, D., J.C. (2003). Information Theory, Inference, and Learning Algorithms. Cambridge University Press, USA.

Formalnie ujmując, rekurencyjny neuron w danym czasie *t* przyjmuje wejścia *x(t)* oraz swoje wyjście z poprzedniego okresu *y(t –1)*.

Obraz zawierający diagram

Opis wygenerowany automatycznie

*Rysunek 8. Rekurencyjny neuron. Źródło: Géron, A. (2017) Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, O’Reilly Media*

Sieć neuronowa powstała w wyniku połączenia rekurencyjnych neuronów wygląda więc następująco:

Obraz zawierający diagram, wykres

Opis wygenerowany automatycznie

*Rysunek 9. Sieć neuronowa rekurencyjna. Źródło: Géron, A. (2017) Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, O’Reilly Media*

Neurony mają dwa rodzaje wag: wagi wejściowe *wx*oraz wagi wyjściowe *wy*. Stąd wyjście serii rekurencyjnych neuronów można opisać jako:

gdzie: *X(t)* to macierz wejść w czasie *t*, *Y(t – 1)* to macierz wyjść w czasie *t - 1*, *Wx* to macierz wag wejściowych, Wy to macierz waga wyjściowych, *φ (⋅)* to funkcja aktywacji, *b* to bias.

Wyjście neuronu jest funkcją wyjść tego samego węzła w poprzednich momentach w czasie, a więc jest to forma posiadania pamięci na temat tego co do tej pory zostało obliczone. Odpowiada za to stan ukryty (ang. hidden state), który w czasie *t* jest obliczony na podstawie wejścia w czasie *t* oraz stan ukryty w czasie *t - 1* transformowanych przez nieliniową funkcję transformacji f:

*Obraz zawierający diagram, wykres

Opis wygenerowany automatycznie*

*Rysunek 10. Sieć neuronowa rekurencyjna ze stanem ukrytym. Źródło: Géron, A. (2017) Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, O’Reilly Media*

Rekurencyjne sieci neuronowe dobrze radzą sobie z przetwarzaniem danych sekwencyjnych, lecz nie są jedynymi sieciami neuronowymi, które są do tego zdolne. Dla krótkich sekwencji tradycyjne sieci neuronowe również są odpowiednie, natomiast dla dłuższych sekwencji, takich jak przetwarzanie tekstu, mogą być użyte również konwolucyjne sieci neuronowe. Mimo wszystko RNN są częściej używane w tym zagadnieniu, ponieważ radzą sobie najlepiej z przetwarzaniem danych sekwencyjnych i z tego względu również model wykonany w tej pracy opiera się na rekurencyjnej sieci neuronowej.

# 2.2.1 LSTM i GRU

Jednym z problemów związanych z implementacją rekurencujnych sieci neuronowych jest zanik pamięci (ang. short-term memory loss). Polega on na zapominaniu informacji, które były przetworzone dawno, a więc zapominaniu o początkowych wejściach. Jest to spowodowane przez zanikający gradient. Problem ten jest szczególnie istotny w złożonych sieciach neuronowych o wielu warstwach ukrytych, gdzie gradient znacząco zanika podczas propagowania wstecznego. Rozwiązaniem problemu zanikającego gradientu jest zastosowanie LSTM lub GRU.

LSTM (ang. Long Short Term Memory networks) to typ rekurencyjnej sieci neuronowej zdolny do przetwarzania zależności w długim okresie czasu. Zastosowanie LSTM jest uzasadnione, aby sieć neuronowa dalej przykładała odpowiednią wagę do początkowych wejść. Zasada działania LSTM jest zbliżona do tradycyjnych RNN. Różnicą jest proces przekazywania informacji. W LSTM w czasie przekazywania informacji następuje rozróżnienie które informacje powinny być transmitowane dalej. W LSTM istnieją dwa stany: - długotrwały o nazwie *cel state* oraz – krótkotrwały o nazwie *hidden state*, a celem jest zadecydowanie które informacje powinny być przekazane ze stanu krótkotrwałego do stanu długotrwałego. Mechanizm służący do tego opiera się na istnieniu bramek (ang. gates). Bramka wejściowa (ang. input gate) *i* wpływa na to jaki sygnał będzie zapisany w jednostce stanu komórki pamięci . Bramka zapominania *f* (ang. forget gate) formułuje wpływ stanu komórki na stan komórki pamięci , a więc steruje tym które elementy pamięci długoterminowej powinny zostać usunięte. Dodatkowo wyróżnia się również bramkę modulacji wejścia (ang. input modulation gate), często uznawaną jako część bramki wejścia. Ma ona na celu modulację informacji otrzymanej w wejściu. Ostatecznie bramka wyjściowa *o* (ang. output gate) wybiera jakie informacje przepływają ze stanu wewnętrznego do wyjścia pozwalając jednocześnie na zatrzymanie informacji nieistotnych w danym momencie czasu. Schemat działania LSTM składa się więc z bramki wejściowej *i* (oraz jej części bramki *g* lub osobno ujmowanej bramki *g)*, bramki zapominania *f* i bramki wyjściowej o to bramki oraz stanów *c* i *h*

Obraz zawierający tekst, diagram, linia, Plan

Opis wygenerowany automatycznie

*Rysunek 9. Struktura LSTM. Źródło: Géron, A. (2017) Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, O’Reilly Media*

Inne rozwiązanie problemu zanikającego gradientu to użycie GRU (ang. gated recurrent units). W swojej budowie GRU jest podobne do LSTM, lecz posiada mniej parametrów. W GRU oba stany są połączone w jeden stan oraz istnieją jedynie dwie bramki. Bramka aktualizacyjna (ang. update) zajmuje się określeniem informacji przydatnych w przyszłości. Z kolei bramka resetująca (ang. reset) definiuje które informacje nie są istotne i powinny zostać pominięte. GRU zbudowane więc jest z bramek *z* oraz *r*, a także stanu *h.*

Obraz zawierający tekst, diagram, linia, zrzut ekranu

Opis wygenerowany automatycznie

*Rysunek 10. Struktura GRU*. *Źródło: Géron, A. (2017) Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, O’Reilly Media*

Ze względu na mniejszą złożoność GRU jest prostsze pod względem obliczeniowym, a zatem zużywa mniej pamięci podczas trenowania sieci. Ponadto GRU stanowi dobry wybór w przypadku chęci otrzymania szybkich wyników. Jednakże jednocześnie LSTM ma większą możliwość dostosowywać się do danego problemu oraz jest szczególnie polecany do dużych zbiorów danych. Oba rodzaje sieci stanowią rozwiązanie problemu zanikającego gradientu, a wybór pomiędzy nimi zależy od wielkości dostępnej pamięci, wielkości zbioru danych oraz czasu jakim dysponuje twórca sieci neuronowej.

# Stworzenie generatora piosenek

# Budowa modelu

# Implementacja modelu w aplikacji Flask

# Wyniki modelu

# Dalszy możliwy rozwój AI w dziedzinach kreatywnych

Podsumowanie

# Bibliografia

Géron, A. (2017) Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. California: O’Reilly Media.

Boden, M. A., 1998. “Creativity and artificial intelligence”. Artificial Intelligence, Volume 103, Issues 1–2, 347-356

Szaleniec, M, Tadeusiewicz, R. (2015). Leksykon sieci neuronowych. Wrocław: Wydawnictwo Fundacji „Projekt Nauka”. s. 22

[Sparrow](https://www.theguardian.com/profile/jeff-sparrow), J. (2023) “Are AI-generated songs a ‘grotesque mockery’ of humanity or simply an opportunity to make a new kind of music?”. The Guardian. 20 lutego. Dostępny w: <https://www.theguardian.com/commentisfree/2023/jan/20/are-ai-generated-songs-a-grotesque-mockery-of-humanity-or-simply-an-opportunity-to-make-a-new-kind-of-music> [Dostęp 27.04.2023]

Savage, M. (2023) “Nick Cave says ChatGPT's AI attempt to write Nick Cave lyrics 'sucks'”. BBC. 17 stycznia. Dostępny w: <https://www.bbc.com/news/entertainment-arts-64302944> [Dostęp 27.04.2023]

Adams LC., Busch F., Truhn D., Makowski MR, Aerts HJWL, Bressem KK (2023) “What Does DALL-E 2 Know About Radiology?”. J Med Internet Res. 25. e43110 Dostępny w: <https://www.jmir.org/2023/1/e43110> [Dostęp 28.04.2023]

Gordon R., (2022) “AI system makes models like DALL-E 2 more creative”. MIT News. 8 września. Dostępny w: <https://news.mit.edu/2022/ai-system-makes-models-like-dall-e-2-more-creative-0908> [Dostęp 28.04.2023]

Roose, K. (2022) “An A.I.-Generated Picture Won an Art Prize. Artists Aren’t Happy.”. The New York Times. 2 września. Dostępny w:

<https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html> [Dostęp 29.04.2023]

[Clarke](https://www.theguardian.com/profile/laurie-clarke), L. (2022) “When AI can make art – what does it mean for creativity?”. The Guardian. 12 listopada. Dostępny w: <https://www.theguardian.com/technology/2022/nov/12/when-ai-can-make-art-what-does-it-mean-for-creativity-dall-e-midjourney> [Dostęp 29.04.2023]

Sak, H., Senior, A., Beaufays, F. (2014). Long short-term memory recurrent neural network architectures for large scale acoustic modeling. Proceedings of the Annual Conference of the International Speech Communication Association. s. 338-342

Sherstinsky, A. (2020) Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. Physica D: Nonlinear Phenomena. Volume 404, 132306, ISSN 0167-2789

Bouwmeester, H, Dougherty, A; Knyazev, A. V. (2015) Nonsymmetric Preconditioning for Conjugate Gradient and Steepest Descent Methods. Procedia Computer Science. 51, s. 276–285

Hardesty, Larry (2017) Explained: Neural networks, MIT News Office. 14 kwietnia. Dostępny w: <https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414> [Dostęp 10.05.2023]

MacKay, D. (2003). Information Theory, Inference, and Learning Algorithms. USA: Cambridge University Press

Britz, D. (2015) Recurrent Neural Network Tutorial, Part 4 – Implementing a GRU/LSTM RNN with Python and Theano – WildML. 27 października. Dostępny w: <https://web.archive.org/web/20211110112626/http://www.wildml.com/2015/10/recurrent-neural-network-tutorial-part-4-implementing-a-grulstm-rnn-with-python-and-theano/> [Dostęp 27.05.2023]

# Spis tabel i rysunków

[Rysunek 1. Obraz wygenerowany przez Floris Groesz z Dall-E software “Kermit the frog painted by Munch” 7](#_Toc136160259)

[Rysunek 2. Obraz wygenerowany przez Jasona Allena za pomocą AI - “Théâtre D’opéra Spatial” 8](#_Toc136160260)

[Rysunek 3. Pojedyńczy neuron. Źródło: MacKay, D., J.C. (2003). Information Theory, Inference, and Learning Algorithms. Cambridge University Press, USA. 13](#_Toc136160261)

[Rysunek 4. Liniowa funkcja aktywacji. Źródło: opracowanie własne 13](#_Toc136160262)

[Rysunek 5. Sigmoidalna unipolarna funkcja aktywacji. Źródło: opracowanie własne 14](#_Toc136160263)

[Rysunek 6. Funkcja aktywacji tangens hiperboliczny. Źródło: opracowanie własne 14](#_Toc136160264)

[Rysunek 7. Funkcja aktywacji ReLu. Źródło: opracowanie własne 15](#_Toc136160265)

[Rysunek 8. a) jednokierunkowa sieć neuronowa b) Rekurencyjna sieć neuronowa. Źródło: MacKay, D., J.C. (2003). Information Theory, Inference, and Learning Algorithms. Cambridge University Press, USA. 17](#_Toc136160266)

[Rysunek 9. Struktura LSTM. Źródło: Chung, J. (2014) Empirical evaluation of gated recurrent neural networks on sequence modeling. 21](#_Toc136160267)

[Rysunek 10. Struktura GRU. Źródło: Chung, J. (2014) Empirical evaluation of gated recurrent neural networks on sequence modeling. 22](#_Toc136160268)

1. “has no pretensions whatever to originate anything. It can do [only] whatever we know how to order it to perform.” [↑](#footnote-ref-1)
2. Tekst wygenerowany przy użyciu Chatu GPT, dostępny na stronie: https://www.bbc.com/news/entertainment-arts-64302944 [↑](#footnote-ref-2)
3. “With all the love and respect in the world, this song is bullshit, a grotesque mockery of what it is to be human, and, well, I don’t much like it.” [↑](#footnote-ref-3)
4. Tekst wygenerowany przez autora przy użyciu Chatu GPT [↑](#footnote-ref-4)