

# Reflectieverslag: Implementatie van een Mamba-model voor Rapgeneratie

Studenten: Nima Agha Ghafar Hamedani, Tommi Lander

Studentnummer: 22085947, 21090874

Datum: 25 maart 2025

## 1. Inleiding

Voor deze opdracht moesten we een Mamba-model from scratch maken. Het model komt uit een nieuw onderzoek, beschreven in het paper "[Mamba: Linear-Time Sequence Modeling with Selective State Spaces](#)". De opdracht die wij hebben verzonnen was het genereren van rapteksten met het mamba model. We mochten tijdens deze opdracht geen gebruik maken van Libraries zoals PyTorch of TensorFlow (uitgezonderd: Numpy, Pandas). In dit verslag vertellen we over de uitdagingen die we zijn tegengekomen, hoe we deze hebben aangepakt en wat we ervan hebben geleerd.

## 2. Problemen en oplossingen

### 2.1 Moeilijke begin: paper, weinig uitleg en structuur vinden

Het grootste probleem aan het begin was dat we weinig informatie hadden om mee te werken. Dit was erg lastig omdat wij geen beeld hadden van dit model. Het paper was best technisch en lastig te begrijpen, vooral omdat het model nog nieuw is. We kregen vanuit de opdracht zelf ook weinig uitleg of richtlijnen. Hierdoor wisten we niet goed hoe we moesten beginnen.

Om toch een start te maken, hebben we drie bronnen gebruikt:

1. We hebben het paper doorgenomen en geprobeerd de structuur eruit te halen.
2. We hebben gekeken naar een eerdere huiswerkopdracht (een notebook) waarin een simpele versie van een model stond.
3. We hebben ChatGPT gevraagd om een Mamba-model te maken in code. Op basis van dit voorbeeld zijn we verder gaan bouwen. We hebben het model aangepast en verbeterd om het te laten werken voor onze opdracht.

Door het combineren van deze drie bronnen zijn we uiteindelijk begonnen met het bouwen van ons eigen model.

## 2.2 Weinig trainingsdata

Een ander probleem was dat we niet veel trainingsdata hadden. We gebruikten een dataset met veel verschillende rapteksten, maar dit was uiteindelijk toch niet genoeg om een woordenschat te bouwen. Het model had daardoor moeite om te leren wat goede zinnen zijn, en begreep de context van woorden niet goed. De gegenereerde tekst was hierdoor vaak niet samenhangend.

We hebben toen geprobeerd om grotere woordenschaten te gebruiken, zoals die van GPT-2, hiermee probeerde we de gegenereerde tekst samenhangend te maken. Dit werkte wel iets beter, maar omdat het niet meer “from scratch” was, hebben we dit uiteindelijk niet meegenomen in onze notebook.

## 2.3 Kwaliteit van de gegenereerde teksten

Het model kon vrij snel tekst genereren, maar de kwaliteit van die tekst was erg slecht. De zinnen kloppen vaak niet qua grammatica of betekenis. Dat komt vooral doordat modellen zoals Mamba heel veel data nodig hebben om echt goede resultaten te geven bij generatie.

Voorbeeld: *money talks money money dough money money money cash chips chips cheese cheddar cheddar cheddar cheese*

We hebben geprobeerd de prestaties iets te verbeteren door te experimenteren met hyperparameters, zoals de lengte van de tekst, de initialisatie van de gewichten en de manier van trainen. Dat hielp een klein beetje, maar het bleef erg lastig om goede rapteksten te genereren

Voorbeeld: *can you down when its sunny and honey got the back control us ayy zoogang we came up learn to afford it with christmas its true and all i done smoked*

## 3. Wat wij hebben geleerd van dit project

Tijdens dit project hebben wij veel geleerd over sequence modellen en hoe lastig het is om zo'n model helemaal zelf te bouwen. Wij begrijpen nu beter hoe de Mamba architectuur werkt, hoe belangrijk de juiste structuur is, en wat er allemaal komt kijken bij tekstgeneratie. Ook hebben wij geleerd dat goede resultaten niet vanzelf komen (je hebt veel data, tijd en experimenten nodig).

## 4. Conclusie

Het bouwen van een Mamba-model voor rapgeneratie was moeilijk, maar wij hebben er veel van geleerd. We hadden weinig informatie, maar door creatief om te gaan met de middelen die we wél hadden (het paper, het huiswerk en ChatGPT), hebben we een werkend model kunnen maken. De kwaliteit van de output is nog niet geweldig, maar het belangrijkste is dat we het model goed hebben begrepen en er wat van geleerd hebben.