

Summary

This paper proposes a generative language model called AfriKI – *Afrikaanse Kunsmatige Intelligensie* (Afrikaans Artificial Intelligence). Our approach is based on an LSTM architecture trained on a small corpus of contemporary fiction. With the aim of promoting human creativity, we use the model as an authoring tool to explore machine-in-the-loop poetry generation. To our knowledge, this is the first study to attempt creative text generation in Afrikaans.

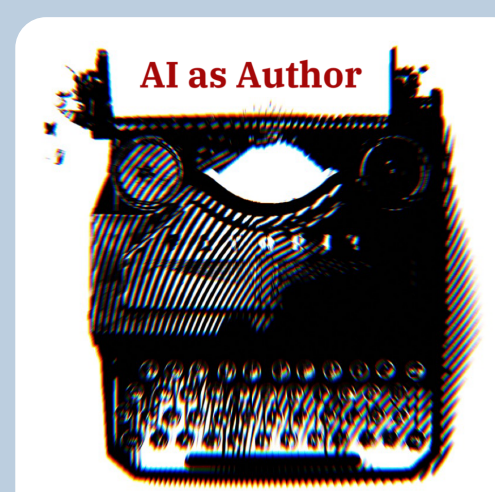
Contributions

- First study in poetry generation in Afrikaans.
- Poetry generation is treated as a co-creative system, an experimental approach that produces high-quality poetic text from very limited data (Afrikaans is a low-resource language).
- Trained on prose, the proposed LSTM network generates individual lines that the human writer selects and vertically arranges into short poems.
- Machine-in-the-loop setting emphasises human creativity, introducing a strong computational component to the creative writing process.
- Results in minimalist free verse poems that contain poetic devices and figurative language.

Machine-in-the-Loop

Machine-in-the-loop frameworks promote human creativity through computational assistance, as opposed to human-in-the-loop pipelines. Applying Clark et al.'s terminology,¹ we employ an iterative interaction structure. Our process consists of a single loop with two stages. First, the model generates a set of unique individual lines. The generated phrases are distinct from the dataset, with hardly any repetition of word order. Second, the first author responds by choosing phrases at will. To create the final artefact, the author arranges the selected lines vertically. Generated text is used without modification (except for some capitalisation and punctuation). The result of our collaborative writing system is a short, compelling work of poetry.

Team



AI as Author is an interdisciplinary research team based in Istanbul that brings NLG into conversation with Literary Studies. We seek practical ways in which literary expertise can strengthen the quality of generated text. <https://authorai.ku.edu.tr>

Model



Frequently occurring words in the corpus

Corpus: AfriKI trained on a lengthy (208,616-word) literary novel titled *Die Biblioteek aan die Einde van die Wêreld* (*The Library at the End of the World*).² This work of new journalism combines fictional techniques with documentary language, and is particularly suitable given its use of rich imagery, figurative language as well as different Afrikaans varieties like *Kaaps* (or Cape Afrikaans) and Standard Afrikaans.

Model Architecture: We obtain best results with the following two-layer LSTM architecture. We use a vanilla LSTM structure.³ We start with 100-dimensional word embeddings with a vocabulary size of 23,317 words, where weights are randomly initialised from a normal distribution with zero mean and standard deviation 0.01. Next, we stack two LSTM layers with 50 units in each layer followed by dropout layers with the rate of 0.2. This is followed by a fully connected layer and a softmax layer. We use the Adam optimiser with a learning rate=0.001, batch size=16, and train for 300 epochs.

References

- [1] E. Clark et al. Creative Writing with a Machine in the Loop: Case Studies on Slogans and Stories. In Proc. ACM IUI, 2018.
- [2] E. van Heerden. *Die Biblioteek aan die Einde van die Wêreld*. NB-Uitgewers, 2019.
- [3] S. Hochreiter and J. Schmidhuber. Long Short-Term Memory. *Neural Computation*, 1997.

Results

Original (Afrikaans)	Translation (English)
<i>Die konstabel se skiereiland</i> Afrika drink onheil in die water. Die landskap kantel sy rug in sigbewaking en vlam. Ons oopgesnyde sake brandtrappe vir die ander state. Hierdie grond word intimidasie.	<i>The constable's peninsula</i> Africa drinks disaster in the water. The landscape tilts its back in surveillance and flame. Our cut-open affairs fire escapes for other states. This soil becomes intimidation.
<i>Gedigte, daar by die brul van 'n brander</i> Hier is die oë katvoet vir die spoelrotse onder uitdrukkings die golwe van gister wat getol en woes en water saam met die son skuim in hul woorde die ingedagte see lig die geure en praat 'n asemhaal	<i>Poetry, there near the roar of a wave</i> Here the eyes are cautious of the sea rocks under expressions the waves of yesterday that whirled and wild and water froth with the sun in their words the introspective sea lifts the scents and utters a breath
<i>Kaapstad</i> Vandag is ons nie net die stad nie maar die vertaler van die son Vanaand se gordyne glinster by skuifvensters in die stadsliggies Die uur van die winde sorg dat dit rondom klink Sy wil die glasvensters deurkosyn eens iets te beskerm Tafelberg maak 'n vraag waarbinne ons 'n duisend name genoem word	<i>Cape Town</i> Today we are not just the city but the translator of the sun Tonight's curtains glitter at sliding windows in the city lights The hour of the winds takes care it sounds around She wants to doorframe the glass windows to protect something Table Mountain creates a question in which we are given a thousand names

Acknowledgements

This paper has been produced benefiting from the 2232 International Fellowship for Outstanding Researchers Program of TÜBİTAK (Project No: 118C285). However, the entire responsibility of the paper belongs to the owner of the paper. The financial support received from TÜBİTAK does not mean that the content of the publication is approved in a scientific sense by TÜBİTAK. We would like to thank Etienne van Heerden for providing his manuscript to be used in this study.