

Large Language Models and GPTs

Exercise 1.3

Advanced Deep Learning in Computer Vision

February 2026



In this exercise, you will implement and train a GPT model called **AndersenGPT** to generate fairy tales in the style of Hans Christian Andersen.

Your task is to complete the implementation of a GPT-2 style model, train it on the provided dataset, and analyze how different design choices affect its performance.

Your tasks are as follows:

1. Complete the implementation of the masked attention module (`class MaskedAttention(nn.Module)`) in `gpt.py`. You may use `testimplementation.py` to verify your implementation before proceeding with training.
2. Train your model on the “andersen_fairy_tales” dataset using `train.py` using a pretrained GPT model. You can find the model here.¹ ² Use free Google Colab GPUs or the DTU HPC server, as training on a CPU may be too slow.

¹If you are using Google Colab, you can create a shortcut to the file in your Google Drive and access it from Colab. "File→Add shortcut": <https://drive.google.com/file/d/13AtVjRahDjEgtnsx0fv1sgL-Dmgh/> You do **not** have to download it to your local machine and then upload it to your drive.

²If you are using DTU HPC, you can access the file from the shared directory: `/dtu/datasets1/02501/`

3. Implement the autoregressive generation function (`generate_text`) in `test.py`. Start with a simple greedy strategy. Choose the next token as the token with the highest probability.
4. Implement multinomial sampling to generate more diverse outputs. In order to convert logits to a probability distribution, use a softmax layer with temperature. Comment on the effect of the temperature. Compare the results with the greedy strategy.
5. **OPTIONAL:** Train the model from scratch instead of using the provided pretrained GPT checkpoint. Compare the performance of both approaches.
6. **OPTIONAL:** Train another model with a different context length and analyze its impact on fairy tale generation quality.
7. **OPTIONAL:** Implement beam search as an alternative to greedy and multinomial sampling. Compare the performance and output diversity of the different strategies.
8. Your model, data, process, performance evaluation, and results should be documented and discussed in a PDF (up to 4 pages) to be uploaded on DTU Learn together with exercises 1.1 and 1.2. More details are in the given template.