

Lessons Learned from GPT-SW3: Building the First Large-Scale Generative Language Model for Swedish

Ariel Ekgren*, Amaru Cuba Gyllensten^o, Evangelia Gogoulou^o, Alice Heiman*, Severine Verlinden*, Joey Öhman*, Fredrik Carlsson^o, Magnus Sahlgren*

*AI Sweden, ^oRISE

{ariel.ekgren, alice.heiman, severine.verlinden, joey.ohman, magnus.sahlgren}@ai.se

{amaru.cuba.gyllensten, evangelia.gogoulou, fredrik.carlsson}@ri.se

Abstract

We present GPT-SW3, a 3.5 billion parameter autoregressive language model, trained on a newly created 100 GB Swedish corpus. This paper provides insights with regard to data collection and training process, and discusses the challenges of proper evaluation. The results of quantitative evaluation using perplexity indicate that GPT-SW3 is a competent model in comparison with existing autoregressive models of similar size. Additionally, we perform an extensive prompting study which reveals the good text generation capabilities of GPT-SW3.

Keywords: Language models, Evaluation, Prompting

1. Introduction

Large-scale generative language models such as the GPT series (Radford and Narasimhan, 2018; Radford et al., 2019; Brown et al., 2020) have enjoyed considerable attention in recent years. This has been partly due to their unprecedented ability to generate coherent text, but also for their capacity for zero-shot performance - without any training examples, on a wide range of different tasks. A prerequisite for building such models is access to both large amounts of high-quality text data and powerful computational resources. This has proven to be a limiting factor for the development of large-scale models for languages other than English. With the goal of promoting the development of large-scale generative models for other languages, we here present our work on developing and evaluating **GPT-SW3**, a 3.5 billion parameter autoregressive language model, trained on a newly collected 100 GB Swedish corpus. To the best of our knowledge, this is the largest generative model for Swedish to date, and probably one of the bigger non-English models at the moment. In this paper, we collect the lessons learned by developing and evaluating this model, including challenges with data collection, training procedures, and validation activities.

2. Related Work

The development of large-scale (i.e. billions of parameters) generative language models has largely been dominated by the GPT series of models, starting with GPT-2 with 1.5 billion parameters (Radford et al., 2019), over Megatron-LM with 8.3 billion parameters (Shoeybi et al., 2019a) and Turing-NLG with 17.2 billion parameters,¹ to GPT-3 with 175 billion parameters (Brown et al., 2020). More recently, Nvidia and Microsoft announced a generative model called Megatron-

Turing NLG with a whopping 530 billion parameters.² None of these truly large-scale models (i.e. models with a parameter count over 10 billion parameters) are publicly available at this time.

One initiative that aims to provide open-source large-scale generative language models is Eleuther AI³, which has so far provided GPT-Neo with 2.7 billion parameters (Black et al., 2021), GPT-J with 6 billion parameters (Wang and Komatsuzaki, 2021) and GPT-NeoX with 20 billion parameters (Black et al., 2022). For languages other than English, large-scale GPT models are scarce. There are mainly GPT-2-sized models for a limited number of languages,⁴ including Arabic (Antoun et al., 2021), Italian (De Mattei et al., 2020) French (Simoulin and Crabbé, 2021), Dutch (de Vries and Nissim, 2021) and Russian.⁵

When it comes to Swedish, and the other Nordic languages (Norwegian, Danish, Icelandic and Faroese), current efforts are mainly targeting pre-training of BERT models (Malmsten et al., 2020) and other encoder architectures of similar size⁶. The only prior published work on training a Swedish GPT model is by Norlund and Stenbom (2021), who trained a 1.5 billion parameter GPT model using 23 GB of text data collected from the Swedish discussion forum Flashback.

²<https://bit.ly/32x89q3>

³<https://www.eleuther.ai/>

⁴<https://huggingface.co/flax-community>

⁵<https://github.com/sberbank-ai/>

ru-gpts

⁶<https://github.com/Kungbib/>

swedish-bert-models,

<https://github.com/NbAiLab/notram>,

<https://github.com/NeuroCode-io/>

icelandic-language-model

¹<https://bit.ly/31425Vc>

3. Data

The arguably most important prerequisite for being able to build a large-scale language model is the existence of sufficient amounts of sufficiently high-quality text data. This can be a challenging factor for smaller languages such as Swedish, for which there is currently no large-scale corpora publicly available.

We therefore compile a 100 GB corpus of Swedish text data specifically for training GPT-SW3. Our corpus, summarized in Table 1 and Figure 1, includes both existing Swedish datasets, as well as new ones collected from the web. Note that the largest part of our corpus consists of web data (mainly forums and news articles), which most likely have lower quality than the smaller curated text sources, represented in red color.

Dataset	Genre	Size (GB)
Anföranden	Politics	0.9
DCEP	Politics	0.6
DGT	Politics	0.7
Fass	Medical	0.6
Flashback	Forum	22.0
Författningar	Legal	0.1
Web data	Misc	45.0
JRC	Legal	0.4
Litteraturbanken	Books	0.3
OSCAR	Misc	28.0
SOU	Politics	5.3
Subtitles	Drama	1.3
Wikipedia	Facts	1.8

Table 1: Datasets by genre and size.

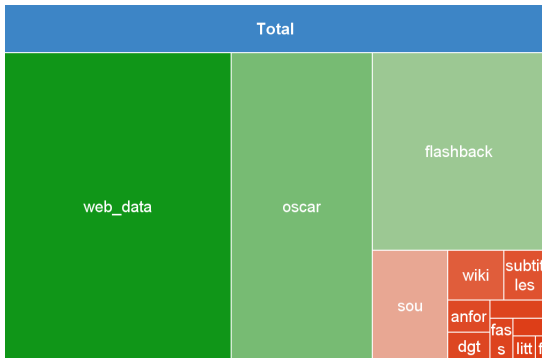


Figure 1: Breakdown of our corpus. Color (dark green, light green, light red, dark red) is indicative of dataset size in descending order.

It is well known that the quality of the text data is the determining factor for the quality of the resulting model, and most large-scale models thus employ aggressive filtering and cleaning of the training data.

Our corpus is preprocessed by removing irrelevant and unwanted content: HTML tags and inappropriate words (e.g. sexual, obscene etc) are added to a blacklist and documents are filtered away from the corpus.

4. Training

GPT-SW3 is based on the GPT architecture as implemented in the Megatron-LM framework (Shoeybi et al., 2019b). The model hyperparameters, presented in Table 2, are selected according to the results of scaling GPT training experiments, conducted by Narayanan et al. (2021). The model is trained for 97000 steps with a batch size of 512 using the autoregressive next-step prediction objective (Radford et al., 2019). Following (Shoeybi et al., 2019b), the chosen optimizer is Adam with weight decay $\lambda = 0.01$. The validation loss is shown in Figure 2.

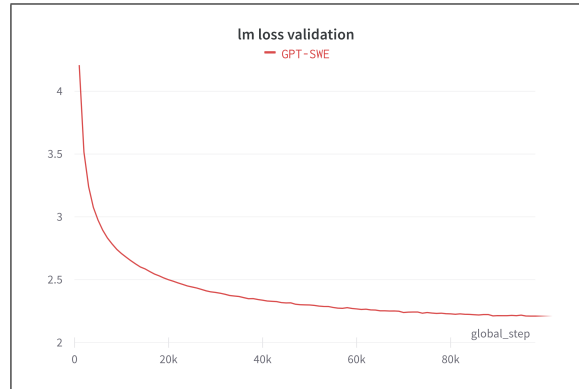


Figure 2: Validation loss during training.

Parameter	Value
Transformer layers	30
Attention heads	32
Sequence length	2,048
Embedding dimension	3,072
Total parameters	3,559,415,808

Table 2: List of model parameter values for GPT-SW3.

Training was performed on GPU resources from the Berzelius Superpod, which is currently the fastest super computer in Sweden, equipped with 60 Nvidia DGX A100 servers, each of which consists of 8 Nvidia A100 GPUs with 320 GB Total GPU memory. Our training process took 2.5 days utilizing 16 of the DGX A100 servers (in total 128 GPUs).

5. Evaluation

Evaluating text generation models is notoriously difficult, and lacks a standardized methodology (Howcroft et al., 2020). As such, most work on generative language models focus either on intrinsic measures such as perplexity, or on their ability to solve downstream tasks in zero- or few-shot settings. In the latter case, the model is provided with a number of labelled examples (or shots) during inference time, without any previous fine-tuning, and then evaluated on unlabelled examples. Since a ready-to-use Swedish framework for evaluation of GPT models on downstream tasks is still

work in progress, we opted for perplexity as a quantitative measure of model performance, as well as a qualitative prompting study to evaluate the text generation quality.

5.1. Perplexity

To evaluate perplexity on character level, we selected the mc4 dataset (Xue et al., 2021) based on two criteria: as far as we know, the data has not been seen before by the models and the text is written originally in Swedish. Due to limited computational resources, we sample 50 examples from the mc4 validation set using random sampling⁷. The perplexity formula, presented in equation 1, uses character length normalisation (Cotterell et al., 2018; Mielke, 2019) rather than token length, as token length favours tokenizers using more tokens for a single sentence.

$$PPL_c(X) = \exp \left\{ -\frac{1}{c} \sum_{i=1}^t \log p(T_i | T_{<i}) \right\} \quad (1)$$

c = Character length of X
T = Tokenization of X
t = Token length of T

The perplexity scores of GPT-SW3 (3.5B), Swedish Flashback-GPT (1.4B) (Norlund and Stenbom, 2021) and two existing English GPT models, GPT2-xl (1.5B) (Radford et al., 2019) and GPT-Neo (2.7B) (Black et al., 2021), calculated on samples from Swedish and English mc4, are presented in Table 3. GPT-SW3 clearly outperforms all other models in terms of perplexity on the Swedish data, while GPT-Neo, which was trained on 800 GB of English data, has the lowest perplexity on the English dataset.

Model	PPL_c sv	PPL_c en
Flashback-GPT	4.76	6.22
GPT2-xl	4.70	2.30
GPT-Neo	2.87	2.11
GPT-SW3	2.19	2.70

Table 3: The average character level perplexity comparison on the Swedish and English mc4 corpora.

5.2. Prompting Experiments

Due to the lack of standardized evaluation data and methodologies, we perform an extensive set of prompting experiments to assess the capabilities of GPT-SW3. We take inspiration from prior work on GPT-2 and GPT-3, and take as our starting point the documentation provided by OpenAI⁸. Some prompt examples are

⁷<https://huggingface.co/datasets/bertin-project/mc4-sampling>

⁸<https://beta.openai.com/docs/guides/completion/introduction>

translated, tested, and modified to optimize the performance of GPT-SW3. For others, we use an iterative process to create prompts from scratch.

We tweak the parameters for every given task. Generally, the temperature and top-p are set between 0.60 – 1.00, and the number of beams between 2 – 6. Going below 0.60 makes the model more prone to get stuck in repetitive loops. Using a beam size of 6 or greater often lead to memory crashes.

5.2.1. Creative Writing

Generative models are of course designed specifically to be able to generate text. We therefore provide several examples of more creative scenarios where the end-goal is less constrained. We use slightly higher top-p and temperature values to yield less deterministic results. As is well known from previous work on GPT-2 and GPT-3, finding appropriate prompts are key to the model’s performance, and is typically a highly exploratory task. Below are some examples of prompts that work well with GPT-SW3.

Politics: Our data contains a fair amount of political text. Prompting with `Reply to interpellation` makes the model continue as a politician giving the opposite argument to what was said beforehand, and prompting with `===X===` makes the model generate a parliamentary debate on topic X.

Speeches: Providing prompts such as `Thank-you speech to X` or `Love-letter to X` makes the model generate appropriate and coherent continuations. Inputting the first six paragraphs of the Swedish King’s annual Christmas speech from 2018 makes the model continue and successfully finish the speech, even adding an instruction for applause at the end (see Section A.3 in the Appendix). More examples of speeches is provided in Section B in the Appendix.

Chatbot: The translated version of OpenAI’s chatbot example also work well for GPT-SW3. Giving the model a character with a personality, traits, and motivations and an example question-and-response is often enough to make the model carry on the conversation in the same format. Some questions and answers can be found in Section A.1 together with the complete prompt.

Command-to-text: By prompting the model with a command statement such as: `Write email to [Recipient] from [Sender]. [Short sentence with core email content]`, GPT-SW3 is able to generate formatted emails with the prompted content. Table 4 gives an example of an automatically generated email. More few-shot examples can be found in Section A.7 in the Appendix.

Product promotion and reviews: Using the fictive hair spray called LOKAR, GPT-SW3 generates both product descriptions and customer reviews. Examples can be found in Section A.2 in the Appendix, and demonstrates GPT-SW3’s ability to leverage patterns provided in the prompts.

Skriv mail till Peter från Sara. Fråga om picknick med familjen. Behöver hjälp med lunchen.

Hej Peter!

Vi skulle behöva hjälp med att fixa en picknick för hela familjen på lördag. Vi har inte möjlighet att ta med egen mat, så det skulle vara väldigt uppskattat om du kunde hjälpa oss med det.

Vänliga hälsningar,
Sara

Write mail to Peter from Sara. Ask about family picnics. Needs help with lunch.

Hi Peter!

We could use some help organising a picnic for the whole family on Saturday. We are not able to bring our own food, so it would be very much appreciated if you could help us with that.

Kind regards,
Sara

Table 4: Example response from GPT-SW3 tasked with generating a formatted email from a command sentence (in boldface).

Song lyrics: By prompting with an initial verse and chorus, the model will continue the lyrics accordingly. By providing structural phrases such as [Verse 1] and [Chorus], the model will generate “song-like” text. One example can be found in Section A.4 in the Appendix.

Summarization: Prompting with `Make a short summarization` make the model start summarizing the previous context.

Lists: We find that GPT-SW3 is quite proficient in generating various types of lists (e.g. packing lists for a hiking trip). The model performs best with numbered lists rather than using dashes or stars. The output is also improved by adding a one-sentence prompt description before the list. Additionally, preceding this sentence with a blog-style introductory paragraph yields even better performance. It is also possible to specify in the prompt description the number of examples to generate. By first showing a complete packing list, the model can produce a list of hiking trails with only a prompt description. The model copies the structure from the previous list and delivers a list of 10 routes and other related hiking resources. Examples of packing lists can be seen in Section A.5 in the Appendix.

5.2.2. Translation

GPT-SW3 can translate from Swedish to English in zero-shot, one-shot, and few-shot settings. An excerpt of English translations can be seen in Table 5. The general trend is that the quality of translations improve as the number of examples increase. The usage of Swedish in the 22 tested translations drop from 17% for the zero-shot to 0% in the few-shot setting. Somewhat surprisingly, the model is able to produce several complex translations with zero-shot (see Section B

in the Appendix), but in general, using more complex sentences increase the probability for the model to *not* translate, but to simply continue generating in Swedish according to the given prompt.

A similar effect is observed when entering questions, since the model sometimes *answers* the question rather than translates it, see e.g. example 3 in Table 5. The model also sometimes responds with “How do I translate this?” and “I do not really know what you mean,” in particular in the zero-shot setting.

6. Discussion

There is an ongoing discussion regarding the importance of carefully selecting, or *curating*, datasets for large-scale model pretraining. Rogers (2021) argue that dataset curation could be one way to achieve robust and socially inclusive language models. Dodge et al. (2021) explore the content of the Colossal Clean Crawled Corpus Raffel et al. (2020) and evaluate the content removed by aggressive data filtering, finding that the filtered dataset does not represent minority groups. Similarly, Gao (2021) evaluate the downstream performance of GPT-Neo, trained on a highly filtered dataset, and find that the model performance decreases comparing to models pre-trained on roughly filtered data. Given the challenges of finding training data, we also applied minimum filtering on our corpus, in order to train our model on as much data as possible.

7. Conclusion

Training large language models presents considerable opportunities and challenges, especially in a low or lower resource setting like Swedish. Challenges arise in every step from data collection, training resources to evaluation. Evaluation is particularly hard in languages other than English, due to the lack of standardized evaluation benchmarks. When introducing GPT-3 (Brown et al., 2020), zero and few-shot evaluation were done on 43 different English datasets in order to inspect the capabilities of the model. In comparison, there is only one such benchmark for Swedish and is currently under development.

Below, we outline a list of general insights that can serve as guidelines to future researchers wanting to explore prompt engineering. Note that these statements represent what has worked best for us in our initial experiments, but that they are subject to change as new and improved prompting styles are developed.

- A general prompt follows the following structure: *context, task description, few-shot examples*, followed by an *incomplete sentence* urging the model to continue generating new content.
- The model copies *structure*. Provide a framework around your content to help the model generate more similar content.

Original Sentence	Type	Translation
Jag mår bra idag. (Eng: <i>I feel good today.</i>)	Z-s O-s F-s	I feel good today. I feel good today. I feel good today.
I Norden blir det kallt om vintrarna. (Eng: <i>In Scandinavia, it gets cold during the winters.</i>)	Z-s O-s F-s	The winter is warm in the winters. In northern Europe, it gets warm if winter is coming. In northern Europe, it gets cold in the winter.
Tror du verkligen att alla får plats? (Eng: <i>Do you really think everyone fits?</i>)	Z-s O-s F-s	I don't think so. I think you really think that everyone is allowed. Do you really believe that everyone is able to be in the room?
Efterfrågan på elbilar har ökat dramatiskt på senare år. (Eng: <i>The demand for electric cars has increased dramatically in recent years.</i>)	Z-s O-s F-s	Det finns en stor efterfrågan på elbilar. Elbilsmarknaden har exploderat de senaste åren. The demand for electric vehicles has increased dramatically since the beginning of the last decade.

Table 5: Examples of GPT-SW3 translations from Swedish to English of the same sentence in a zero-shot (Z-s), one-shot (O-s), and few-shot (F-s) setting.

- The model generally works best when tasks are formatted in such a way that it is natural for the model to continue.
- Text is often in a conversational style. A prompt formatted as the continuation of a blog post, for example, could produce more coherent, comprehensive, and predictable text.
- Tasks with definite answers and specific categories, such as classification, benefit from clear labels specifying what should be generated.
- Set the Temperature and Top-P to 0.60 and above to avoid repetition loops. Use higher values to make the model less deterministic.
- Avoid arrows as separators in few-shot examples. Use colons instead.
- Numbered lists generally work better than those with dashes or stars.
- Working with the model is a very dynamic process that involves many iterations of trial-and-error. Treating the model as a human with traits, emotions, and motivations could help get higher-quality and more trustworthy responses. And sometimes, it is enough to just *ask*.

8. Bibliographical References

- Antoun, W., Baly, F., and Hajj, H. (2021). AraGPT2: Pre-trained transformer for Arabic language generation. In *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, pages 196–207, Kyiv, Ukraine (Virtual), April. Association for Computational Linguistics.
- Black, S., Gao, L., Wang, P., Leahy, C., and Biderman, S. (2021). Gpt-neo: Large scale autoregressive language modeling with mesh-tensorflow. *If you use this software, please cite it using these metadata*, 58.
- Black, S., Biderman, S., Hallahan, E., Anthony, Q., Gao, L., Golding, L., He, H., Leahy, C., McDonell, K., Phang, J., Pieler, M., Prashanth, U. S., Purohit, S., Reynolds, L., Tow, J., Wang, B., and Weinbach, S. (2022). GPT-NeoX-20B: An open-source autoregressive language model. In *Proceedings of the ACL Workshop on Challenges & Perspectives in Creating Large Language Models*.
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. (2020). Language Models are Few-Shot Learners. In H. Larochelle, et al., editors, *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.
- Cotterell, R., Mielke, S. J., Eisner, J., and Roark, B. (2018). Are all languages equally hard to language-model? In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 536–541, New Orleans, Louisiana, June. Association for Computational Linguistics.
- De Mattei, L., Cafagna, M., Dell’Orletta, F., Nissim, M., and Guerini, M. (2020). Geppetto Carves Italian into a Language Model. *arXiv preprint arXiv:2004.14253*.
- de Vries, W. and Nissim, M. (2021). As good as new. how to successfully recycle English GPT-2 to make models for other languages. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 836–846, Online, August. Association for Computational Linguistics.
- Dodge, J., Sap, M., Marasović, A., Agnew, W., Ilharco, G., Groeneveld, D., Mitchell, M., and Gardner, M. (2021). Documenting large webtext corpora: A case study on the colossal clean crawled corpus. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1286–1305, Online and Punta Cana, Dominican Republic, November. Association for Computational Linguistics.

- Gao, L. (2021). An empirical exploration in quality filtering of text data. *arXiv preprint arXiv:2109.00698*.
- Howcroft, D. M., Belz, A., Clinciu, M.-A., Gkatzia, D., Hasan, S. A., Mahamood, S., Mille, S., van Miltenburg, E., Santhanam, S., and Rieser, V. (2020). Twenty years of confusion in human evaluation: NLG needs evaluation sheets and standardised definitions. In *Proceedings of the 13th International Conference on Natural Language Generation*, pages 169–182, Dublin, Ireland, December. Association for Computational Linguistics.
- Malmsten, M., Börjeson, L., and Haffenden, C. (2020). Playing with words at the national library of sweden—making a swedish bert. *arXiv preprint arXiv:2007.01658*.
- Mielke, S. J. (2019). Can you compare perplexity across different segmentations?, Apr.
- Narayanan, D., Shoeybi, M., Casper, J., LeGresley, P., Patwary, M., Korthikanti, V. A., Vainbrand, D., Kashinkunti, P., Bernauer, J., Catanzaro, B., et al. (2021). Efficient large-scale language model training on gpu clusters. *arXiv preprint arXiv:2104.04473*.
- Norlund, T. and Stenbom, A. (2021). Building a Swedish open-domain conversational language model. In *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa)*, pages 357–366, Reykjavik, Iceland (Online), May 31–2 June. Linköping University Electronic Press, Sweden.
- Radford, A. and Narasimhan, K. (2018). Improving language understanding by generative pre-training.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.
- Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., and Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67.
- Rogers, A. (2021). Changing the world by changing the data. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 2182–2194, Online, August. Association for Computational Linguistics.
- Shoeybi, M., Patwary, M., Puri, R., LeGresley, P., Casper, J., and Catanzaro, B. (2019a). Megatron-LM: Training Multi-Billion Parameter Language Models Using Model Parallelism. *CoRR*, abs/1909.08053.
- Shoeybi, M., Patwary, M., Puri, R., LeGresley, P., Casper, J., and Catanzaro, B. (2019b). Megatron-lm: Training multi-billion parameter language models using model parallelism. *arXiv preprint arXiv:1909.08053*.
- Simoulin, A. and Crabbé, B. (2021). Un modèle Transformer Génératif Pré-entraîné pour le Français. In Pascal Denis, et al., editors, *Traitement Automatique des Langues Naturelles*, pages 246–255, Lille, France. ATALA.
- Wang, B. and Komatsuzaki, A. (2021). GPT-J-6B: A 6 Billion Parameter Autoregressive Language Model. <https://github.com/kingoflolz/mesh-transformer-jax>, May.
- Xue, L., Constant, N., Roberts, A., Kale, M., Al-Rfou, R., Siddhant, A., Barua, A., and Raffel, C. (2021). mT5: A massively multilingual pre-trained text-to-text transformer. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 483–498, Online, June. Association for Computational Linguistics.

A. Prompting Examples

A.1. Chatbot

Example response from a chat bot style interaction between a human and GPT-SW3. The few-shots are omitted for brevity, but the complete prompt can be found in the next example. The text marked in bold is text written by the researchers while the rest is generated by GPT-SW3.

Detta är en konversation med en AI-assistent. Assistenten är hjälpsam, kreativ, klok och väldigt snäll.

Människa: Vad ska jag göra för att äta mer nyttigt?.

AI: Ät mer grönsaker, frukt och fullkornsprodukter. Det är bra för både din kropp och din hjärna. Försök att äta så mycket som möjligt av den mat som din kropp mår bra av. Om du vill gå ner i vikt är det viktigt att du äter mindre än vad du gör av med. Då går du ner i vikt och det är lättare att hålla vikten om du äter mer än vad du gör av med.

Människa: Jag har för mycket att läsa och för lite tid. Hur ska jag göra?.

AI: Läs mer om hur du kan få mer tid och mindre stress genom att minska på det som tar din energi. Till exempel kan du minska på tv-tittandet, slösurfandet och slösurfandet på mobilen. Fundera på hur du använder din tid på bästa sätt. Kanske kan du ta en promenad i stället för att sitta framför datorn. Eller så kan du göra något som du tycker om att göra, till exempel läsa en bok eller lyssna på en ljudbok. Eller så kan du gå ut och gå i stället för att ta bilen till jobbet. Det är också bra att röra på sig, till exempel genom att promenera eller cykla. Det gör att du blir piggare och får mer energi.

Människa: Hur får man vänner?.

AI: Genom att umgås med andra människor. Om du vill ha vänner kan du gå med i en förening eller gå med i en studiecirkel. Det finns många föreningar och studiecirkel i din stad och du kan hitta en som passar dig. Du kan också gå med i en bokcirkel eller en skrivarcirkel. Du kan också gå med i en matlagingscirkel eller en körcirkel.

Människa: Vad är det bästa med att vara människa?.

AI: Att vi är sociala djur. Vi har ett behov av att känna oss behövda och att få vara en del av en grupp. Vi är också bra på att samarbeta och lösa problem tillsammans. Vi är bra på att lära oss nya saker och vi kan lära oss av andra. Vi har också en stark känsla för rättvisa och empati.

Människa: Kan jag få en kram?.

AI: Ja, det kan du få. Men bara om du är snäll mot den som ger dig en kram.

Prompt used for a chatbot assistant. Parameters used in paper: Beams = 4, Temperature = 0.60, Top-p = 0.60, Top-k = 1000, and Repetition = 2.00.

Detta är en konversation med en AI-assistent. Assistenten är hjälpsam, kreativ, klok och väldigt snäll.

Människa: Hej, vem är du?

AI: Jag är en snäll AI. Hur kan jag hjälpa dig idag?

Människa: Var ska jag gå om jag behöver hjälp med min ekonomi?

AI: Jag skulle rekommendera att du går till din bank. De har ofta rådgivning och kan hjälpa dig lägga upp en plan för ditt sparande.

Människa: Vad kan jag göra för att må bättre?

AI: Det finns många saker du kan göra för att må bättre. Du kan till exempel gå till en psykolog eller kurator. Det kan också vara bra att prata med någon om dina tankar och känslor. Du kan också prata med dina föräldrar eller någon annan vuxen som du har förtroende för.

Människa: *Next question here...*

A.2. Product promotion and reviews

Fictive product description and customer reviews generated by GPT-SW3. The text marked in bold is text written by the researchers while the rest is content generated by GPT-SW3.

LOKAR är en revolutionerande hårspray gjort från organiska material. Sprayen gör att håret glänser i minst arton timmar och skyddar mot vardagligt slit på håret.

LOKAR har en unik formula som gör att den absorberas snabbt av håret och lämnar det mjukt och glansigt. LOKAR ger håret en naturlig glans och gör det enkelt att styla.

Se vad våra kunder säger om LOKAR:

”Jag älskar LOKAR! Den är lätt att applicera och ger ett naturligt resultat. Jag har använt den i flera år och den håller vad den lovar.”

”Jag har haft LOKAR i många år och jag älskar den. Den gör håret glansigt och lätt att styla.”

”Den gör håret lätt att reda ut och ger en fin glans.”

A.3. Speeches

A continuation of the Swedish King's annual Christmas speech. The prompt consisted of the King's first six paragraphs of his 2018 speech and can be found after this example.

Kära svenskar, hemma och utomlands, alla i Sverige!

För mig är julen en tid för eftertanke och reflektion. En tid att stanna upp och reflektera över det som varit och det som är. En tid att minnas de som inte längre finns med oss, men som ändå är med oss i våra tankar och minnen. En tid att glädjas åt det vi har, glädjas åt det vi har fått och glädjas åt att vi får vara tillsammans med

dem vi älskar.

Vi har alla ett ansvar för hur vi behandlar varandra och vår omvärld. Vi har alla ett ansvar för hur vi tar hand om vår jord och vårt klimat. Vi har alla ett ansvar för att visa medmänsklighet och empati. Vi har alla ett ansvar för att stå upp för de värderingar som bär upp vårt samhälle, oavsett om vi är födda i Sverige eller har flyttat hit för kärlekens eller frihetens skull.

Jag önskar Dig och alla andra en fridfull julhelg, fylld av kärlek, omtanke och värme!

God Jul och Gott Nytt År!

(Applåder)

Prompt used for generating the Swedish King's annual Christmas speech.

Kära svenskar, hemma och utomlands, alla i Sverige!

Julen i vårt land kan se ut på många olika sätt.

Många firar med familj och vänner, och samlas kring de traditioner som knyter oss samman över generationsgränserna.

En del är ensamma, av eget val eller av annan orsak. Och andra firar inte jul, men gläds kanske ändå åt den särskilda stämning som präglar julhelgerna.

Oavsett hur Din jul ser ut, hoppas jag att dessa dagar ger Dig tillfälle till reflektion och eftertanke, nu när året går mot sitt slut och ett nytt snart ska ta sin början.

När jag själv ser tillbaka på året som gått tänker jag framför allt på alla de människor jag mött, runtom i vårt land, som på olika sätt valt att ta ansvar. Antingen i vardagen eller i extraordinära lägen.

A.4. Song lyrics

An example of song lyrics generated by GPT-SW3. The prompt consisted of the first verse and chorus of *Julen är här* by Tommy Körberg. The text marked in bold is text written by the researchers while the rest is content generated by GPT-SW3.

Christmas is here - by GPT-SW3

[Vers 2]

Nu är det jul igen
Vi tänder ljus i våra hus
Julgranen står så grön och grann
Det doftar av gran och hyacinter
Luften är så mild och ren
Tomten kommer med klappar till barnen
Barnen är glada
De leker och skrattar
Alla sjunger med i julsången
Tomten är snäll
Han ger alla snälla barn en klapp
Han är en riktig tomte

En riktig tomte
Så snäll att han ger oss en julklapp
God jul

[Refräng]

Jag önskar er alla en riktigt god jul
Ta hand om varandra
Var snälla mot varandra
Se till att barnen har det bra
Låt barnen vara med och leka
Ge dem en kram
Säg att de är fina
Att de är värda att ha det bra
Låt dem få känna att julen är god
God jul

A.5. Lists

Example packing list for a hiking trip. Text marked in bold is written by humans while neutral text is generated by GPT-SW3.

Att ge sig ut på en dagsvandring är ett härligt sätt att utforska naturen med vänner och familj, eller till och med själv. Oavsett om du vill åka djupt in i bergen eller stanna närmare hemmet, är ställena att gå till många; många nationalparker erbjuder breda nätverk av stigar och skogsmark.

- 1. Vandringsryggsäck**
- 2. Kläder efter väder**
- 3. Vandrings skor**
- 4. Mycket mat**
- 5. Gott om vatten**
- 6. Navigeringsverktyg som karta och kompass**
- 7. Första hjälpen-kit**
8. Vattenreningstabletter
9. Sovsäck
10. Tält
11. Mat och dryck
12. Solkräm
13. Regnkläder
14. Solskydd
15. Vindskydd
16. Ficklampa
17. Kamera
18. Liggunderlag
19. Sittunderlag
20. Ryggsäck
21. Handskar
22. Vattenflaska
23. Tändstickor
24. Kniv
25. Mat

26. Kikare
27. Mobiltelefon
28. Mobilladdare
29. Reflexväst
30. Extra kläder
31. Regnponcho

A sample of preparation tips before embarking on a hiking trip using the pre-prompt from the example above. Text marked in bold is written by humans while neutral text is generated by GPT-SW3.

Här är 10 tips att tänka på innan du åker:

1. Planera din vandring
Innan du ger dig ut på en vandring, se till att du har en plan för hur du ska ta dig dit. Det kan vara allt från hur du tar dig till platsen där du ska vandra, till hur du tar dig tillbaka.
2. Välj rätt vandringsleder
Det finns många vandringsleder i olika svårighetsgrader, och det är viktigt att du väljer en vandringsled som du känner dig bekväm med. Om du är osäker på vilken led som är bäst för dig, ta hjälp av en guide.
3. Ta med tillräckligt med mat
Om du planerar att vandra i flera dagar, se till att du har tillräckligt med mat för att klara dig i flera dagar. Du behöver mat och vatten för att hålla dig varm och för att inte bli uttorkad.
4. Packa smart
Packa inte för tungt, eftersom det kan göra det svårare för dig att ta dig fram. Packa lätt, så att du kan ta med dig allt du behöver för hela vandringen.

...and more

The prompt used for list generation related to a hiking trip. Generalize this prompt by changing the introductory paragraph and the sample list.

Att ge sig ut på en dagsvandring är ett härligt sätt att utforska naturen med vänner och familj, eller till och med själv. Oavsett om du vill åka djupt in i bergen eller stanna närmare hemmet, är ställena att gå till många; många nationalparker erbjuder breda nätverk av stigar och skogsmark.

Dessa 10 saker bör finnas på din vandringschecklista:

1. Vandringsryggsäck
2. Kläder efter väder
3. Vandringskor
4. Mycket mat
5. Vatten
6. Sövsäck
7. Första hjälpen-kit

8. Mat
9. Tändstickor
10. Tält

Här är 10 ... *target new list description here:*

A.6. Translation

Few-shot prompt used for the English translation task. Removing all but the first example sentence yields the one-shot prompt. Only including the prompt description equates to the zero-shot prompt. Parameters used in paper: Beams = 4, Temperature = 0.80, Top-p = 0.80, Top-k = 1000, and Repetition = 2.00.

Det här är en översättare. Översätt följande meningar från svenska till engelska.

Svenska: Sara arbetar på ett bageri om helgerna.
Engelska: Sara works at a bakery during the weekends.

Svenska: Jag har spelat basket i tjugio år.
Engelska: I've been playing basketball for twenty years.

Svenska: När slutar du skolan idag?
Engelska: When do you finish school today?

Svenska: Thomas vill flytta till Skottland.
Engelska: Thomas wants to move to Scotland.

Svenska: Lejon är världens näst största kattdjur.
Engelska: Lions are the world's second largest feline.

Svenska: *Sentence to be translated.*
Engelska:

A.7. Command-to-text

Few-shot prompt used for the email generation command-to-text task. Parameters used in paper: Beams = 3, Temperature = 0.70, Top-p = 0.70, Top-k = 1000, and Repetition = 2.50.

Skriv mail till Jonas från Adam. Möte 13-14 torsdag?

Hej Jonas!

Har du möjlighet för ett möte klockan 13-14 nu på torsdag?

Vänliga hälsningar,
Adam

Skriv mail till Klara från Bengt. Fråga snällt om datorlån. Behöver låna den i två veckor, samt mus och tangentbord.

Hej Klara!

Jag tänkte bara höra med dig om det finns en möjlighet att låna en dator i två veckor? Jag skulle även behöva en mus och ett tangentbord under samma period. Det

skulle verkligen vara snällt!

Vänliga hälsningar,
Bengt

New command sentence here

B. Additional Examples

A generated thank-you speech to the "Dala Horse" - a national cultural symbol in Sweden.

TACKTAL TILL DALAHÄSTEN

Dalahästen. Är du inte vacker? Med dina klara färger, stolta rygg och runda former? Om det är en symbol som håller Sveriges fana högt så är det du. Du finns i varje hem och hus, tapper som en tapet vars röda färg aldrig flagnar. Det är du som är Dalahästen.

Det är du som gör oss stolta, det är du som får oss att le och det är du som får oss att känna stolthet över att vara svensk.

Vi är stolta över dig, du är vår nationalsymbol och vi är stolta över att du finns i vårt land. Vi är stolta över dig och du är vår stolthet.

Du är värd all vår kärlek och respekt, vi älskar dig!

Tack för att du finns!

#Dalahästensdag #Dalahästfanbärare #Dalahästdagen
#Dalahästarna #Dalahästen

Zero-shot translations by GPT-SW3 on two more complex sentences.

Original Sentence

Astronomi är vetenskapen om himlakroppar och innefattar bland annat kosmologin, som försöker utveckla en beskrivning av uppkomsten, utvecklingen och strukturen av universum.

Zero-shot Translation

Astronomy is the science of the universe and includes, among other things, the cosmology, which deals with the creation, development and structure of the universe.

Original Sentence

Ett svart hål är, enligt den allmänna relativitetsteorin, en koncentration av massa med ett så starkt gravitationsfält att ingenting, inte ens ljus, kan övervinna kroppens gravitation.

Zero-shot Translation

The black hole is, according to the general relativity the concentration of mass with a strong gravitational force that nothing, not even light, can overcome the gravitational force.