Neural Predictive Calculator – Process

Anton Mukin 11.11.2025

Overview:

Feed-forward Neural Networks (FNN): Initial experiments and later optimization with Keras-tuner (July 8, 2025; October 3, 2025).

Recurrent Neural Networks (RNN): Research and implementation of SimpleRNN, LSTMs, and GRUs (July 9, 2025).

Transformers: In-depth work, including coding a model from scratch to understand the architecture (July 26, 2025).

Pre-trained Models: Fine-tuning of advanced models like Gemini and Gemma 3 to compare with self-built models (August 18, 2025).

Before entering this project, I already came with some experience working with neural networks, thanks to https://www.kaggle.com/learn

To begin I first started with some research on the topic, which was all collected and presented in the Literature review.

While researching, a working prototype was developed and vaguely evaluated (FNN1.ipynb). This was done in google colab, using publically available free hardware provided by google.

Together, the literature review and the proof-of-concept (FNN1) was submitted and presented as Zwischenprodukt on: 18.06.2025

After the bad experience in google colab, I switched to jupyter lab and used local hardware for the next part of the project: RNNs

With the summer holidays on the rise not many plans were made, on how much and when to work, since it was uncertain if I would even have the time. Still with the literature review done, I already determined which models I wanted to evaluate:

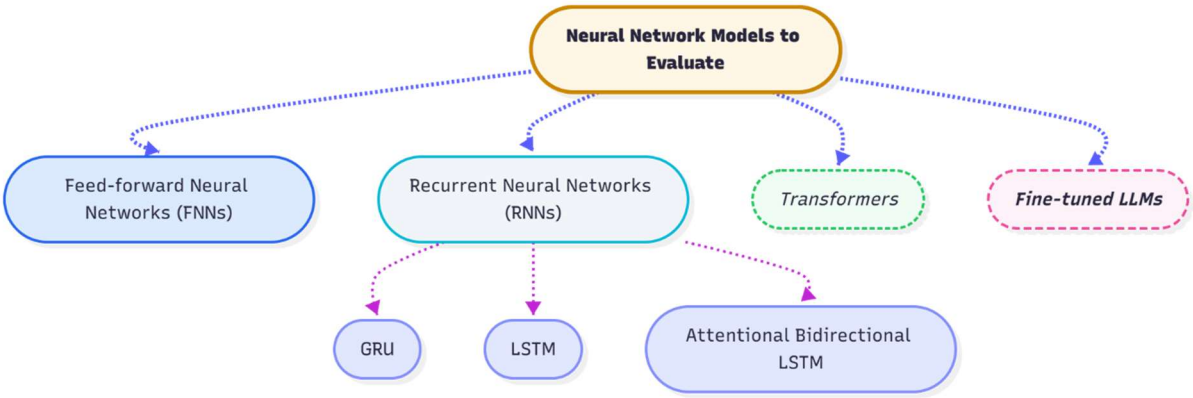


Figure 1: Chart depicting the Neural Network models, panned to evaluate. The diagram was created with mermaidchart.com

Nonetheless, these were the goals for the summer holidays:

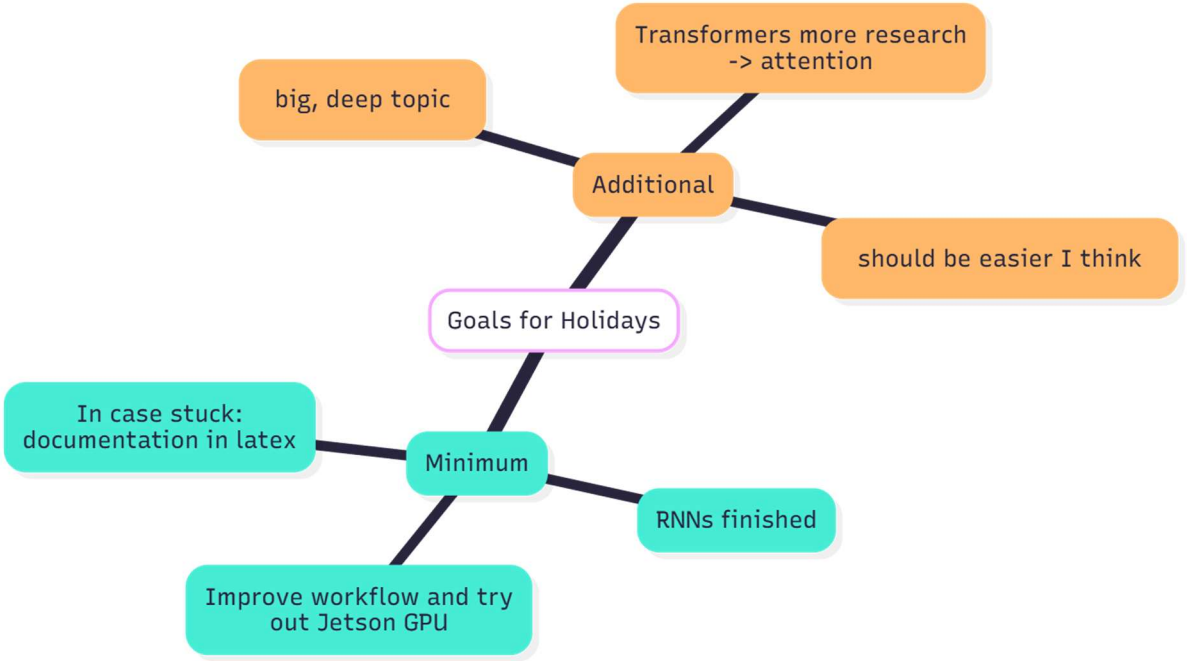


Figure 2: A Mind map with the plans made for the Holidays (8.07.2025). The diagram was made with mermaidchart.com

The RNNs were pretty much finished after the first week, because it was a fun task, the other was more gnarly and took longer.

GRUs and LSTMs were properly researched and implemented. A heatmap was drawn for both models. They weren't used or evaluated any further.

Work shifted to attentional RNNs, specifically the bidirectional LSTM with attention.

Later I also started working on a transformer model.

All of this was done in the first 4 weeks of the summer holidays. I had a lot of spare time to work at home, and I enjoyed it, especially the work on the transformer. It felt really cool to be finally able to understand how they work.

On August 13. A proper plan, as requested by Mr. Schneider, was made:

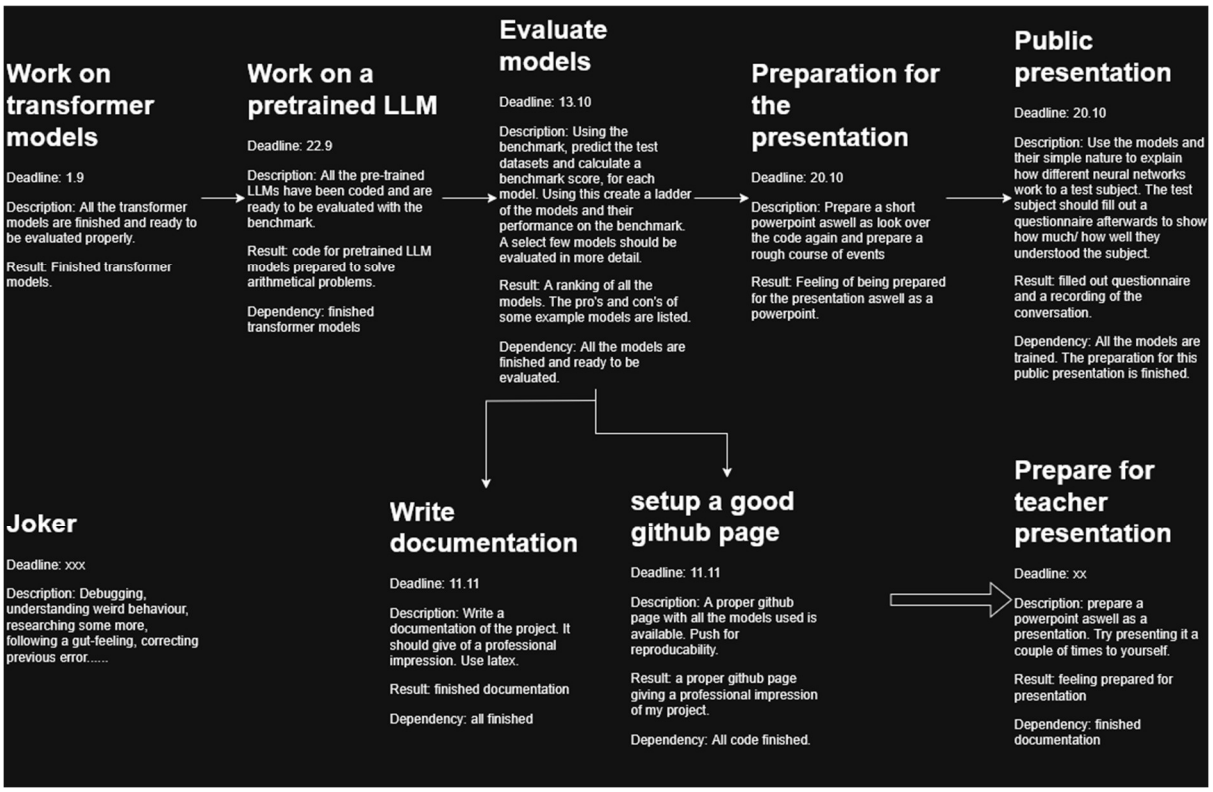


Figure 3: a diagram showing the workloads, I must do in this project. Made myself.

The diagram was followed strictly, with more details, as always in the Arbeitsprozess-pdfs. The first 3 workloads I was about 2 weeks ahead of schedule.

The Guided discussion, here called public presentation, had to be moved, because my discussion partner was on vacation. It was held on 25.10.2025.

The whole situation with the Nvidia GPU also led to a couple of problems, it took me around a week to figure out how to properly use the environment there and stuff. (25.8.2025 – 4.10.2025)

After first trying to work with unsloth, I was pushed towards using jetson-containers on the unsloth discord forum. Jetson containers made my life a lot easier, after identifying the fitting container for my task I was easily able to get TensorFlow running on the gpu as well as Transformers and other lesser known libraries.

Writing the actual code for the notebooks and the actual calculation/evaluation of the two gemma models took up less than half of the total time used for fine-tuning pre-trained Language Models.

This was also in part because I ran into a stop where I had the feeling there must be an error somewhere, but I wasn't able to find any errors. After a week or so of agony, Mr. Schneider suggested I just disprove all my hypothesis for what might have caused the error and that would be enough, so that's what I did.

Approximately around October work on evaluating the models started. It was finished on the 4th of November. It consisted of the excel spread sheet you see in the findings of the findings.

Following this, I reorganized GitHub with Jules – an AI that's supposed to help you with your GitHub repo. I personally didn't find it useful. My project was very difficult for it to grasp.

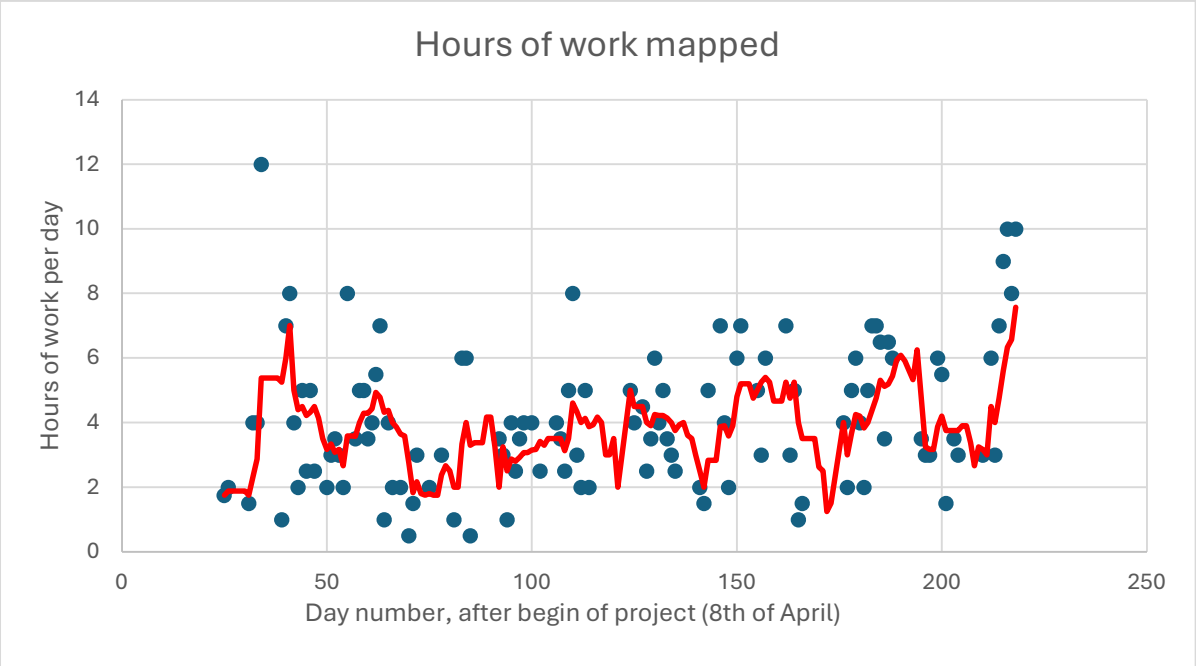
The next step was writing the documentation, consisting of the methodology and findings. This was mainly all done during the autumn holidays.

As previously mentioned, I held the guided discussion with Marton on the 25th and prepared for it around 5 days in advance with a PowerPoint and a questionnaire.

Following this the final phase of the project began, or as I like to call it: "Endsprint" Where I mostly worked on the documentation and on stuff for the documentation. Stuff always kept coming up, like some school rules I hadn't heard of, or a bug in my code.

All in all, I enjoyed working on this project, there were parts I truthfully enjoyed, like coding the transformer0.ipynb, that was my highlight. Apart from that, I ran into a lot of weird errors, where even ai couldn't help me out, which was very annoying. (Think, the accuracy mismatch, or also the Nvidia GPU environment.)

To analyse my work hours, refer to the Graph below.



The total number of Hours is 461.75, not counting the last day (will probably be around 10h).

In reflection, I think I organized myself well. Work hours are spread out more or less evenly. From the beginning a plan was formulated, with clear goals and ideas on what had to be done. Through out the project, the current standing was evaluated time and time again.

Short term goals were made on different scales: from days and weeks to months and holidays. They actually helped me a lot with motivation and reassessment.

Reaffirming the significance of the computed results was done inside the notebooks. It helped avoid overlooking errors, and particularly played a role for example with the Gemma3 notebooks, with the accuracy mismatch. The issue was identified even though an error wasn't even found.

But there are also many more examples when I evaluated models – I mean literally every model trained was technically evaluated even during training.

After every work day I also reflected on my work in the journal, this contained a reflection page, where I wrote lower end stuff like if I was too tired to work with lots of focus that day. Another page in the journal was the "What's next page" where plans for the following day, as well as other plans were noted.

My knowledge expanded a lot after this project. I learned many things. I found it very inspiring my self, when I worked on the FNN2 notebook towards the end of this project. It showed very well and totally convinced me at how quick and clean I was able to code that notebook. In comparison, coding FNN1 at the start of this project took me weeks, and that notebook is weird and probably moldy. I think this proves my performance improvement with Tensorflow and maybe python in general.

Additionally what's funny is. I never used github properly before this project. You can see this in the earliest commits on my repo, I used to think that you're supposed to upload every file manually onto github.com (the website). In hindsight, I obviously see that I was very dumb.

I made a lot of improvements with linux. Right now, I still wouldn't say that I feel comfortable in linux, but still… I'm better than when I started. (For proof, you can see in the Arbeitsjounal that I would sometimes spend lots of time fighting the linux system on my nvidia GPU.) Linux is really cool, I want one of those old thinkpads with linux.

Other Things I've improved at thanks to the project:

- General theory behind different types of neural networks, especially the transformer architecture (obviously).

- Coding with latex took up a very long amount of time, and through out the project I shifted from using overleaf, to using vsc+git
- Jetson containers and docker played a big role in this project, I was able to utilize them for the first time in my life. They're really neat, but also very confusing (still).
- My interest in neural networks also just grew lots

Experts from the discord server helped me setup my environment, on the GPU. Otherwise the rest of the work was done with lots of googling, and AI assistance. https://www.geeksforgeeks.org had the most useful/understandable/helpful tutorials in my opinion. They are cited a lot throughout the documentation.

Other external experts: My mom (statistician) helped me a lot with the p-Value calculation, by pointing me towards one-sample t-tests. And she also reviewed my documentation for grammatical and logical mistakes.

For more details see the Arbeitsprozess itself.