# Attention Is All I Need

## Oleg Sedukhin[1]

[1]*Novosibirsk, Russia*

E-mail: *sedol.mong@mail.ru*, Telegram: *@olegs1339*

**I'M LOOKING FOR A** research position to solve challenging machine learning problems. Prefer team work, ideally as a member of experienced research group. Prefer remote work or semi-remote from Novosibirsk.

## 1. Education

Master's degree from Faculty of Information Technologies of Novosibirsk State University.

## 2. Work experience

(i) Engineer in Huawei Russian Research Institute (apr. 2022 – now), worked on tabular machine learning and object detection tasks.

(ii) Support and webinar speaker on data science courses in online schools (2020 – 2022).



**Fig. 1**: The author.

## 3. Contributions

Currently I am finishing large-scale tabular machine learning experiments on 100+ datasets to validate a hypothesis that plain baselines (like CatBoost) may be no worse than state-of-the-art tabular autoML systems, if early stopping end ensembling are properly designed. I'm also testing several tweaks to improve gradient boosting performance and studied an intriguing problem of model performance under distributional shift[1]. I plan to publish the results when they are ready.

I have also researched methods for few-shot fine-grained object detection in autoML scenario. For now, this research left unfinished due to changing priorities within the company where I work, but I have a paper draft (link)[2]. The code can be found in this repository.

---

[1] In the modern ML community it is believed that a distributional shift usually occurs either as a result of a change in the ratio of example types, or as a result of a concept shift. However, there are other cases, when the very concept of "data distribution" is ambiguous. For example, the data may be the result of two-stage sampling or a process over time. This may lead to pitfalls during model validation. Simple and clear examples can be given, however, this problem is underexplored in the literature on distributional shift.

[2] I managed to achieve pretty good results with fine-tuning Cascade R-CNN on top of EVA-02 backbone, on a simple benchmark of 10 datasets. However, I soon realized that this task may come in a lot of very different forms, some of which are unexplored in the ML literature and may require com-

## 4. Skills and studying

For 4 years I have been trying to gain a deep understanding of machine learning to find weak spots and limitations of modern algorithms and try to improve them.

(i) I almost fully read **Probabilistic Machine Learning: An Introduction** by Kevin P. Murphy, 2023 (a version with my commentaries at this link).

(ii) I'm half read **Reinforcement Learning Textbook** by Sergey Ivanov, 2022 (a version with my commentaries at this link).

(iii) Since last year, I write short summaries of some of the papers that I read (at this link).

(iv) I have published several longreads on Habr.

In the near future I plan to finish the above books, read more books, gain more systematic understanding of generative models and their mathematics and limitations, and also keep up with modern papers.

**Some skills**: Linux, Python, Numpy, Pytorch, Keras, Pandas, Polars, Scikit-learn, Matplotlib, Catboost, Ray.

---

pletely different approaches. So, my further work consisted of collecting more reliable benchmark to test a wide range of few-shot recognition abilities.