Valentin Malykh

Research Scientist

+7 (903) 966 8133
valentin@maly.hk
val.maly.hk
GitHub: madrugado
Google Scholar: https://goo.gl/b6UPb5

Education

2019 **PhD**, *Study Completed at Moscow Institute of Physics and Technology, Defended at Institute for Systems Programming, Russian Academy of Sciences*, Moscow.
Thesis: Noise Robustness in Various NLP Tasks
Scientific Advisor: Prof. Vladimir Arlazarov

2009 **M.Sc.**, *Moscow Institute of Physics and Technology*, Moscow.
Major: Computer Science

2007 **B.Sc.**, *Moscow Institute of Physics and Technology*, Moscow.
Major: Computer Engineering

Experience

2019–p.t. **Senior Research Scientist**, *Huawei Noah's Ark lab*, Moscow.
Research on dialog systems and related NLP problems.
I am teaching Natural Language Processing course in collaboration with MIPT.

2019–p.t. **Senior Research Scientist**, *Kazan Federal University*, remote.
Research on summarization for Russian language.

2016–2019 **Research Scientist**, *Laboratory of Neural Systems and Deep Learning, Moscow Institute of Physics and Technology*, Dolgoprudny, Moscow Region / remote.
Work on `iPavlov.ai` project. Research on noise robustness in NLP tasks.
Detailed achievements:
- Published 15 papers,
- including ACL Demo Paper on DeepPavlov library,
- and 4 papers in NIPS Proceedings on ConvAI Challenge series.

2018–2019 **Applied Research Scientist**, *VK.com*, Saint-Petersburg.
Work on research problems related to social network, like text classification, summarization, etc. Accepted paper at ECIR'2019 conference. Shared task on headline generation at Dialogue'2019 conference.

2016–2019 **Lecturer**, *Moscow Institute of Physics and Technology*, Dolgoprudny, Moscow Region.
I was teaching Deep Learning in NLP course in 2016-2018. The cumulative audience of that course is estimated to 3000 people. Also I was teaching Neural Networks and Reinforcement Learning modules in Machine Learning course.

2015–2016 **Machine Learning Engineer**, *Yandex*, Moscow.

I was working in Yandex.News on the whole ML stack for ranking and also participated in clusterization development.
Detailed achievements:

- New ranking formula for news clusters.
- Improved news agency ranking.

2014–2015 **Research Engineer**, *Sputnik*, Moscow.

Sputnik is a Russian government-sponsored search engine, which was developed from scratch by a small team of engineers. I was in the Search Quality Department.
Detailed achievements:

- Web pages classifier.
- Malicious documents ranking.

2012–2014 **Research Engineer**, *Cognitive Technologies*, Moscow.

Cognitive Technologies is a company with long history of work in computer vision domain. It created second most common OCR solution in Russia. Nowadays company's main interest is self-driving cars.
Detailed achievements:

- Computer vision & control for a robot car.
- Research at Astarta project – high-load document classifier.

Miscellaneous

2017 **Certified Instructor**, *NVIDIA Deep Learning Institute*.

In addition to being a Certified Instructor, I have authored whole NLP Workshop in NVIDIA DLI and presented it at GTC EU in 2017.

2017&2018 **Organizer**, *ConvAI Challenge series*.

ConvAI Challenge is devoted to creation of human-level conversational intelligence. I was co-organizing in both scientific and technical areas.

Publications

- Burtsev, M., Logacheva, V., Malykh, V. ... & Bengio, Y. The first conversational intelligence challenge. In The NIPS'17 Competition: Building Intelligent Systems (pp. 25-46), 2018.
- M. Burtsev et al. DeepPavlov: Open-Source Library for Dialogue Systems. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics: Demo Track, 2018.
- Alekseev, A., *Malykh, V.*, et al. AspeRa: Aspect-based rating prediction model. Proceedings of 41st European Conference on Information Retrieval, 2019.
- Gavrilov, D., Kalaidin, P., *Malykh, V.* Self-attentinve model for headline generation. Proceedings of 41st European Conference on Information Retrieval, 2019.
- V. Malykh. Robust to Noise Models in Natural Language Processing Tasks. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, 2019.
- I. Shenbin et al. RecVAE: a New Variational Autoencoder for Top-N. Recommendations with Implicit Feedback. Proceedings of the 13th International Conference on Web Search and Data Mining.
- T. Shavrina et al. RussianSuperGLUE: A Russian Language Understanding Evaluation Benchmark. Proceedings of The 2020 Conference on Empirical Methods in Natural Language Processing, 2020.
- *V. Malykh* et al. SumTitles: a Summarization Dataset with Low Extractiveness. Proceedings of The 28th International Conference on Computational Linguistics, 2020.

- E. Tutubalina, I. Alimova, Z. Miftahutdinov, A. Sakhovskiy, *V. Malykh*, S. Nikolenko, The Russian Drug Reaction Corpus and neural models for drug reactions and effectiveness detection in user reviews, Bioinformatics, 15 January 2021.

Languages

English	Fluent	*IELTS Academic 7.0*
Russian	Native speaker	