

1. What are the main problems of modern NLP and NLU?

Main problems of NLP: sentence boundary disambiguation, POS-tagging, creating vocabularies, understanding the context, semantic analysis, NER, language inequality, discourse parsing, creating datasets and corpus, cross-lingual model, large documents and contests.

Main problems of NLU: meaning interpretation, performative utterance (J. L. Austin), anaphora understanding, pragmatics of language (emotions, intentions, social context of speech act), coreference, polysemy, ambiguity.

2. Which libraries would you pick to use for the following cases and why (all problems should be solved for the Russian)

Sentiment analysis	Dostoevsky, deepPavlov
Multi-label classification	BERT
Dependency parsing	UDPipe, spacy-ru
POS-tagging	Natasha
NER	Natasha, deepPavlov

3. How would you evaluate a classification model, which metrics would you use?

For classification models: Classification Accuracy, Precision, Recall, F1 Score, ROC Curve

For NLP specific: SSER, BLEU, Jaccard Similarity, Perplexity

4. Main pipeline for the text pre-processing.

1. Input data
2. Data cleaning (stop-words, punctuation)
3. Tokenization
4. POS-tagging, NER
5. Normalization and Lemmatizing
6. Count-Vectorization / TF-IDF
7. Word embeddings
8. Analysis

5. Microservices or monoliths? Why.

Monoliths in the start of the project

- concentration of resources on the main tasks required at the start
- saving limited resources at the start
- high development speed and more efficient feedback
- high productivity at the start
- easier to manage
- suitable for small teams
- interservice independence

Microservices after

- *great performance when increasing and diversifying tasks*
- *reliability*
- *profitability*
- *sharing responsibility in large teams and making communication easier*
- *better organization when scaling a project*
- *reduction of demands on the team due to diversification of responsibility*
- *interservice independence*

6. Describe the hardest programming task you've been facing with. It's not necessarily ML task, could be just a programming. Why this task was hard to accomplish? What was your solution for the task? Can you share a github project?

7. Did you work with VCS? Which one?

Yes, with GIT

8. Did you work with Github Actions?

Yes, at starter level

9. How familiar are you with Docker and other orchestration tools?

Always ready for learning

10. What is ed25519 and why is it concerning to be better than ecdsa?

Ed25519 is a public-key signature system

- Faster
- More effective security

11. Do you have any experience in data mining?

Yes, at the entry level. I am currently working on my scientific project about coronavirus in Belarus and want to identify the strategies for disinformation that are used by various media sources. I have finished my work on parsers and formed a dataset already. I have some experience of working with bs4, lxml, selenium libraries for parsing, and a little experience in using nltk, gensim, natasha, pymorphy for nlp tasks. Before I started learning Python, I worked a lot with text analysis in the field of sociology: discourse analysis, content analysis, grounded theory, conversation analysis.