

Безопасное предобучение глубоких языковых моделей на синтетическом псевдоязыке

И.Ю. Бондаренко (НГУ, Новосибирск)

В данной работе проводится сравнение предварительного обучения нейросети трансформерного типа на текстах естественного языка и на предложениях синтетического псевдоязыка. Искусственные тексты были автоматически сгенерированы по написанным нами правилам в контекстно-свободной грамматике. Результаты дообучения на выполнение заданий проекта RussianSuperGLUE статистически достоверно показали, что модели имеют одинаковые оценки, то есть можно считать, что использование искусственных данных даёт преимущество для “безопасности” искусственного интеллекта за счёт возможности полностью контролировать состав выборки. Также мы можем говорить о том, что на этапе предобучения модели типа RoBERTa достаточно научиться распознавать только синтаксические и морфологические закономерности языка, которые могут быть успешно созданы довольно таким простым способом, как контекстно-свободная грамматика.