

Ученые из Германии разработали новый инструмент под названием DataSAIL, который помогает точнее оценивать работу моделей искусственного интеллекта (ИИ). Он автоматически разделяет данные на обучающие и тестовые таким образом, чтобы они максимально отличались друг от друга.

В машинном обучении модели тренируются на большом объеме данных, но перед использованием в реальной жизни должны пройти проверку на тестовом наборе. Если тестовые данные слишком похожи на обучающие, модель может показать хорошие результаты, но провалиться в реальных задачах. DataSAIL помогает этого избежать, делая проверку более честной.

Обычные алгоритмы не умеют качественно разделять данные, из-за чего многие ИИ-системы на деле оказываются переоцененными. Новый инструмент устраняет эту проблему, автоматически создавая два максимально отличающихся набора данных.

DataSAIL работает с любыми типами данных. Пользователю достаточно указать несколько параметров, остальное программа делает сама. Также это первый инструмент, который может работать с так называемыми данными взаимодействий — например, для задач в фармацевтике, где важно понять, как лекарство взаимодействует с разными белками.

