

В нашей работе мы хотим показать, что большая объяснительная сила у современных алгоритмов достигается благодаря подходам, которые уже были описаны в философии науки конца прошлого века.

В качестве примера мы рассматриваем две архитектуры нейронных сетей, которые получили распространение в последние годы и зарекомендовали себя как модели, обладающие объяснительной силой. Мы показываем, что архитектуры этих сетей выстроены в соответствии с положениями двух теорий научного объяснения: DN-модели (дедукто-номологической модели) и SR-модели (модели статистической значимости).

Этот доклад призван показать, что философские подходы к научному объяснению снова становятся актуальными. Сформулированные в 70-80ых теоретические подходы получают сейчас прикладное значение в области Data Science.