Об информационной безопасности – в новых статьях журнала СПбПУ

«Как [redacted] «Проблемы информационной безопасности. Компьютерные системы».

Новые подходы к противодействию угрозам нарушения конфиденциальности в федеративном обучении исследуют в своей статье д-р. техн. наук, профессор [redacted] СПбПУ Дарья Лаврова и студент Института компьютерных наук и кибербезопасности СПбПУ Павел Безбородов. Авторы берут за основу методы оптимизации, позволяющие преобразовывать веса локальных нейросетевых моделей и создавать новые веса для передачи на узел совместного градиентного спуска, что способно предотвратить перехват злоумышленником весов локальной модели. Проведено 64800 экспериментов, которые подтвердили эффективность разработанного подхода. Учёным удалось достичь ограничения точности глобальной модели до 8 % при дальнейшем увеличении точности локальной модели в федеративном обучении.

В другой статье выпуска рассматривается проблема преодоления кризиса доверия к современным системам искусственного интеллекта. Нейросети работают по принципу «черного ящика», когда видны действия на входе и выходе, но непонятны процессы, происходящие внутри. Авторы – д-р техн. наук, профессор Военно-космической академии имени А. Ф. Можайского Денис Бирюков и канд. техн. наук, доцент СПбПУ Александр Супрун – сделали вывод о том, что объяснение и интерпретация – не просто термины, а фундаментальные процессы, связывающие рациональное познание с практикой машинного обучения. Фокусируясь на роли объяснения (раскрытия причинно-следственных связей) и интерпретации (адаптации смысла для аудитории) в контексте машинного обучения, учёные говорят о взаимосвязи этих процессов и необходимости междисциплинарного диалога для выработки общих взглядов на вопрос «объяснимости» и «интерпретируемости» применительно к моделям машинного обучения и решениям, формируемым ими.

Ознакомиться с этими и другими статьями журнала «Проблемы информационной безопасности. [redacted] можно, оформив подписку.