

УДК 332.14

АНАЛИЗ РИСКОВ ПРИМЕНЕНИЯ НЕЙРОСЕТЕВЫХ ТЕХНОЛОГИЙ СОТРУДНИКАМИ ПРИ РАБОТЕ С КОНФИДЕНЦИАЛЬНОЙ ИНФОРМАЦИЕЙ

Д. А. ВЕТРОВ

Научный руководитель П. С. ХАРЛАМОВ

Смоленский филиал Российской академии народного хозяйства
и государственной службы при Президенте Российской Федерации
Смоленск, Россия

Использование сотрудниками организаций публичных и сторонних технологий на базе искусственного интеллекта без прямого разрешения отдела информационной безопасности или сотрудников, ответственных за сохранение конфиденциальной информации, на данный момент выделяют в отдельную категорию угроз, называемую «теневого ИИ». В данном случае классические угрозы, связанные с хранением и передачей информации, дополняются угрозами её активной обработки и интерпретации внешними алгоритмами. Бесплатные и публичные нейросети используют пользовательские промты и оценки качества ответов для тренировки будущих моделей, что одновременно нарушает законодательство, связанное с передачей персональных данных, и правила работы организации с конфиденциальной информацией. Количество данных, передаваемых ИИ-агентами, внедряемыми на компьютеры организации, сложно оценить даже в настоящее время, что несёт ещё большие риски, особенно после появления Prompt Injection – атаки, основанной на запросах к таким агентам. В целом, картина возможных рисков и способов их устранения представлена на рис. 1. При правильном применении искусственного интеллекта и оценке опасностей даже в секретных проектах нейросети могут стать незаменимыми.



Рис. 1. Модель рисков безопасности при работе с нейросетевыми моделями