УДК 004.89

ПРАКТИЧЕСКАЯ ОЦЕНКА МОДЕЛЕЙ ИИ С УЧЁТОМ НОРМАТИВНЫХ ТРЕБОВАНИЙ К ЗАЩИТЕ ДАННЫХ

Ортего П., студент направления Industrial Production Engineering Bachelor's degree program, II курс
Миланский технический университет, г. Милан, Италия

В условиях стремительного развития технологий искусственного интеллекта и их массового внедрения в различные сферы человеческой деятельности вопрос соответствия моделей ИИ нормативным требованиям к защите данных становится одним из ключевых аспектов, определяющих как успешность их применения, так и уровень доверия со стороны пользователей и регуляторов [1].

Практическая оценка моделей ИИ с учётом действующих законодательных и нормативных норм представляет собой комплексную задачу, которая требует интеграции технических, юридических и организационных подходов для обеспечения соблюдения требований приватности, безопасности и ответственности при обработке персональных и чувствительных данных.

Современные нормативные акты, такие как Общий регламент по защите данных (GDPR) в Европейском союзе, Закон о конфиденциальности потребителей в Калифорнии (CCPA) и аналогичные документы в других юрисдикциях, устанавливают жёсткие правила по сбору, хранению, обработке и передаче данных [2].

В частности, они требуют минимизации объёмов обрабатываемой информации, прозрачности процедур, возможности контроля и удаления персональных данных, а также реализации технических и организационных мер по предотвращению несанкционированного доступа и утечек. Для моделей ИИ это означает необходимость учитывать данные ограничения на всех этапах жизненного цикла – от сбора и предобработки данных, через обучение и верификацию моделей, до их внедрения и эксплуатации.

Практическая оценка моделей с учётом нормативных требований включает в себя несколько ключевых компонентов. Во-первых, анализ соответствия используемых данных требованиям по анонимизации, минимизации и правам субъектов данных. Это предполагает проверку источников данных, методов их обработки, а также документирование процессов и процедур, обеспечивающих защиту конфиденциальной информации [3].

Во-вторых, проведение тестирования моделей на предмет утечек данных, предвзятости и несправедливости, а также оценка устойчивости к различным видам атак, включая те, что направлены на извлечение личной информации.

Технические инструменты оценки и аудита моделей включают методы статического и динамического анализа, применение дифференциальной приватности, техники обфускации и регуляризации, а также использование специализированных метрик и тестов, направленных на выявление рисков нарушения приватности и безопасности. Важной составляющей является документирование и прозрачность процессов, позволяющая аудиторам и регуляторам проводить независимую проверку соответствия моделей установленным нормам. Современные платформы и фреймворки машинного обучения всё чаще включают встроенные средства контроля и отчётности, упрощающие выполнение этих задач.

Организационные аспекты оценки включают формирование междисциплинарных команд, объединяющих специалистов по машинному обучению, безопасности, праву и этике, что позволяет комплексно подходить к проблемам и минимизировать риски. Внедрение процессов управления рисками, планирования реагирования на инциденты и непрерывного мониторинга состояния моделей способствует поддержанию высокого уровня доверия и соответствия нормативам в долгосрочной перспективе [4].

Особое значение приобретает анализ соответствия моделей ИИ нормативам в контексте их интеграции в критически важные и регулируемые отрасли – здравоохранение, финансы, государственное управление и пр. Здесь требования к безопасности и прозрачности максимально строгие, а последствия нарушений могут быть крайне серьёзными. Практическая оценка в таких условиях требует глубокого понимания специфики применения, использования специализированных протоколов и, зачастую, взаимодействия с контролирующими органами.

В целом, практическая оценка моделей искусственного интеллекта с учётом нормативных требований к защите данных представляет собой фундаментальную задачу для создания ответственных и доверенных ИИ-систем. Она требует системного подхода, включающего технические инновации, юридическую грамотность и организационное управление. Только комплексное соблюдение всех аспектов позволит обеспечить баланс между инновационностью, эффективностью и безопасностью, что является залогом успешного и этически приемлемого внедрения искусственного интеллекта в современное общество [5].

Список литературы:

1. Geyer, R. Differentially Private Federated Learning: A Client Level Perspective / R. Geyer, T. Klein, M. Nabi // arXiv preprint. – 2017. – arXiv:1712.07557.

2. Aono, Y. Privacy-Preserving Deep Learning via Additively Homomorphic Encryption / Y. Aono, T. Hayashi, L. Wang, S. Moriai // IEEE Transactions on Information Forensics and Security. – 2018. – Vol. 13, № 5. – P. 1333–1345. – DOI 10.1109/TIFS.2017.2787987.

3. Phong, L. T. Privacy-Preserving Deep Learning via Weight Transmission / L. T. Phong, Y. Aono, T. Hayashi [et al.] // IEEE Transactions on Information Forensics and Security. – 2018. – Vol. 13, № 11. – P. 3003–3016. – DOI 10.1109/TIFS.2018.2833036.

4. Gilad-Bachrach, R. CryptoNets: Applying Neural Networks to Encrypted Data with High Throughput and Accuracy / R. Gilad-Bachrach, N. Dowlin, K. Laine [et al.] // Proceedings of the 33rd International Conference on Machine Learning. – 2016. – Vol. 48. – P. 201–210.

5. Mohassel, P. SecureML: A System for Scalable Privacy-Preserving Machine Learning / P. Mohassel, Y. Zhang // IEEE Symposium on Security and Privacy. – Piscataway : IEEE, 2017. – P. 19–38. – DOI 10.1109/SP.2017.12.