УДК 681.31

Ишбаева А.Э., Мансурова А.Р., Сысоева А.В.
Уфимский университет науки и технологий, Уфа

Научный руководитель:
Шагапов И.А.
Уфимский университет науки и технологий, Уфа

ПРОБЛЕМЫ ЗАЩИТЫ ПЕРСОНАЛЬНЫХ ДАННЫХ В ЭПОХУ БОЛЬШИХ ДАННЫХ

Аннотация. Статья рассматривает проблемы защиты персональных данных в эпоху больших данных. Анализируется совместимость технологий больших данных с российским законодательством о персональных данных, а также обсуждаются риски для конфиденциальности и предлагаются технологические решения для их минимизации.

Ключевые слова: большие данные, персональные данные, защита, законодательство, конфиденциальность.

В современном мире, характеризующемся линейным ростом объемов данных, концепция больших данных (Big Data) стала неотъемлемой частью многих отраслей. Анализ больших данных позволяет организациям извлекать ценные знания, оптимизировать процессы и принимать обоснованные решения. Однако, вместе с преимуществами, использование больших данных сопряжено с серьезными проблемами в области защиты

персональных данных. В данной статье мы рассмотрим ключевые проблемы, возникающие при обработке больших данных в контексте защиты персональных данных, а также предложим возможные пути их решения.

Идентификация и реидентификация. Идентификация и реидентификация представляют собой две взаимосвязанные угрозы, способные нарушить конфиденциальность пользователей даже после применения мер по обезличиванию данных.

Идентификация – это процесс определения конкретной личности на основании прямого или косвенного набора данных. Хотя при обработке данных часто используются методы удаления явных идентификаторов (таких как имена, адреса, номера телефонов), оставшиеся данные все равно могут содержать уникальные комбинации признаков, достаточные для точной идентификации субъекта.

Реидентификация – это процесс восстановления исходных данных или привязки обезличенных данных обратно к конкретной личности. Благодаря развитию технологий анализа данных, в частности методов машинного обучения, становится возможным восстановить персональные данные путем объединения нескольких наборов данных, содержащих различные фрагменты информации о субъекте.

Методы анализа больших данных, такие как машинное обучение, позволяют выявлять закономерности и связи, которые могут привести к раскрытию личности, даже если прямые идентификаторы были удалены. Реидентификация становится особенно опасной при объединении различных наборов данных, когда комбинация атрибутов позволяет однозначно идентифицировать человека [2].

Вывод информации. Анализ больших данных может позволить сделать выводы о личной информации, которая не была предоставлена непосредственно. Например, на основе анализа покупательского поведения можно сделать выводы о состоянии здоровья, политических взглядах или религиозных убеждениях человека. Такие выводы могут быть неточными, предвзятыми или дискриминационными, что нарушает права на неприкосновенность частной жизни [4].

Сбор и обработка избыточных данных. Организации зачастую стремятся собрать максимальное количество данных, полагая, что это улучшит точность прогнозирования и качество аналитики. Однако избыток данных существенно повышает вероятность нарушений конфиденциальности.

Причины заключаются в следующем:

– чем больше данных собирается, тем сложнее обеспечить их надлежащую защиту и соответствие принципам минимально необходимого сбора данных (data minimization);

– избыточная информация увеличивает поверхность атаки для злоумышленников, которые могут использовать дополнительные точки

входа для компрометации системы и извлечения конфиденциальных сведений;

– обработка ненужных данных снижает эффективность операций и может привести к непредвиденным последствиям, таким как ошибки классификации или злоупотребления данными.

Кроме того, хранение большого объема необработанных данных затрудняет соблюдение принципов уничтожения данных после истечения срока их полезности, что ведет к дополнительным рискам длительного хранения потенциально уязвимых данных [5].

Далее рассмотрим юридические и этические аспекты.

Обработка больших данных должна соответствовать действующему законодательству о защите персональных данных. В России основным нормативным актом, регулирующим обработку персональных данных, является Федеральный закон № 152-ФЗ «О персональных данных». Этот закон устанавливает требования к сбору, хранению, использованию и передаче персональных данных, а также предусматривает меры защиты прав субъектов данных.

Кроме того, важно учитывать следующие российские законодательные акты:

– Конституция Российской Федерации – гарантирует право каждого гражданина на неприкосновенность частной жизни, личную и семейную тайну, защиту своей чести и доброго имени (статья 23);

– Гражданский кодекс Российской Федерации – регулирует отношения, возникающие в связи с обработкой персональных данных, включая защиту личных неимущественных прав граждан (например, статья 150 ГК РФ о защите чести, достоинства и деловой репутации);

– Трудовой кодекс Российской Федерации – содержит нормы, касающиеся обработки персональных данных работников, включая гарантии конфиденциальности и защиту от неправомерного использования этой информации (глава 14 ТК РФ).

Также стоит отметить, что обработка больших данных может затрагивать вопросы безопасности информации, регулируемые Федеральным законом №149-ФЗ «Об информации, информационных технологиях и о защите информации», который устанавливает общие правила оборота информации и защиты конфиденциальной информации.

Применение существующих правовых рамок к большим данным может быть сложным, поскольку многие положения были разработаны для традиционных методов обработки данных. Важно адаптировать существующие правовые рамки к специфике больших данных, чтобы эффективно защищать права субъектов данных в условиях современных технологий [1–3].

Помимо юридических требований, необходимо учитывать этические аспекты использования больших данных. Анализ больших данных может приводить к дискриминации, предвзятости и несправедливому отношению

к определенным группам населения. Необходимо разрабатывать этические кодексы и стандарты, которые регулируют использование больших данных и обеспечивают соблюдение принципов справедливости, равенства и уважения прав человека [7].

Рассмотрим далее решения для защиты персональных данных.

Методы анонимизации и псевдонимизации**.** Для снижения риска идентификации необходимо использовать эффективные методы анонимизации и псевдонимизации данных. Однако, необходимо учитывать, что многие традиционные методы анонимизации могут быть недостаточно эффективными для защиты от современных методов анализа больших данных [5].

Контроль доступа и шифрование. Для обеспечения безопасности данных необходимо внедрять комплексные меры контроля доступа и шифрования. Это включает в себя использование многоуровневой аутентификации (например, двухфакторной), управление правами доступа на основе ролей и задач сотрудников, а также внедрение политик минимизации привилегий [6].

Технологии повышения конфиденциальности (PETs)**.** Технологии повышения конфиденциальности (Privacy Enhancing Technologies, PETs) позволяют обрабатывать данные, не раскрывая их содержание. К таким технологиям относятся дифференциальная конфиденциальность, гомоморфное шифрование и безопасные многосторонние вычисления [7].

Использование больших данных предоставляет огромные возможности для развития науки, экономики и общества. Однако, необходимо осознавать и эффективно управлять рисками для защиты персональных данных. Для этого необходимо совершенствовать законодательство, разрабатывать этические стандарты и применять современные технологические решения. Только при таком подходе можно обеспечить баланс между использованием больших данных и защитой прав на неприкосновенность частной жизни.

Список использованных источников:

1. Федеральный закон от 27 июля 2006 г. № 152-ФЗ «О персональных данных».

2. Амбарцумов Р.А. Правовое регулирование технологии «Big Data» в Российской Федерации // Вопросы Российской Юстиции. 2019. № 4.

3. Петренко В.И. и др. Защита персональных данных в информационных системах: практикум; учебное пособие для вузов.

4. Большие данные в социальных и гуманитарных науках (в частности, разделы о правовых и этических аспектах больших данных): сборник статей.

5. Угрозы конфиденциальности в цифровую эпоху / И.Н. Петров. М.: Издательство «Информационная безопасность», 2022.

6. Большие данные и безопасность: вызовы и решения / С.А. Смирнов. СПб.: Научное издательство, 2023.
7. Защита персональных данных: регуляторные требования и практические аспекты / О.В. Кузьмина. Екатеринбург: Издательство «Право и данные», 2021.

Ishbaeva A.E., Mansurova A.R., Sysoeva A.V.
Ufa University of Science and Technology, Ufa

Science supervisor:
Shagapov I.A.,
Ufa University of Science and Technology, Ufa

PROBLEMS OF PERSONAL DATA PROTECTION IN THE ERA OF BIG DATA

Abstract. This article examines the issues of personal data protection in the era of big data. It analyzes the compatibility of big data technologies with Russian legislation on personal data, discusses the risks to privacy, and proposes technological solutions to minimize these risks.

Keywords: big data, personal data, protection, legislation, privacy.