Международная научно-практическая конференция. Развитие теории и совершенствование практики использования специальных знаний в условиях цифровизации. МГЮА. Май 2020.

ПЕРСОНАЛЬНЫЕ ДАННЫЕ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: СУДЕБНО-ЭКСПЕРТНЫЙ АСПЕКТ

Нестеров А.В.

PERSONAL DATA AND ARTIFICIAL INTELLIGENCE: FORENSIC ASPECT

Nesterov A.V.

Аннотация. Проблема. Законодательство в области преступлений в сфере компьютерной информации страдает отсутствием современной терминологии, правовых критериев и конкретизации используемых понятий, что стало существенно проявляться с усилением внимания к обороту персональных данных. Внедрение смарт-систем, называемых системами искусственного интеллекта, дает возможность операторам персональных данных осуществлять на основе систем больших банных, накапливаемых в облачных хранилищах центров обработки данных, осуществлять продуцирование профилей персон с нарушением некоторых норм законодательства. Цель статьи. Исследовать вопросы, связанные с использованием юридического инструмента судебной экспертизы, при выявлении неправомерного доступа к компьютерной информации. Результаты. Показано, что систематизация данных о персоне позволит сформулировать критерии для демаркации персональных данных от анонимизированных персональных данных, а также определить логические связи персональных данных с псевдонимизированными персональными данными. Выводы. Пока информационное законодательство не будет упорядочено, судьи будут назначать судебные экспертизы с вопросами о наличие логических связей у обезличенных персональных данных с персональными данными конкретного субъекта персональных данных, а экспертам придется разбираться с критериями и создавать методики

исследования таких связей с учетом используемого программного обеспечения оператора персональных данных. Разработка, апробация и утверждение таких методик ставит достаточно сложную задачу перед судебными экспертами.

Abstract. Problem. Legislation in the field of computer information crimes suffers from a lack of modern terminology, legal criteria and specification of the concepts used, which has become significantly evident with the increased attention to the turnover of personal data. The introduction of smart systems, called artificial intelligence systems, makes it possible for personal data operators to carry out the production of profiles of persons in violation of certain legal norms on the basis of large bath systems accumulated in cloud storage centers of data processing centers. Purpose of article. Investigate issues related to the use of a legal tool for forensic examination when detecting illegal access to computer information. Results. It is shown that the systematization of personal data will allow us to formulate criteria for the demarcation of personal data from anonymized personal data, as well as to determine the logical connections of personal data with pseudonymized personal data. Conclusions. Until the information legislation is regulated, judges will appoint forensic examinations with questions about the existence of logical links between depersonalized personal data and the personal data of a particular personal data subject, and experts will have to deal with the criteria and create methods for investigating such links, taking into account the software used by the personal data operator. The development, testing and approval of such methods poses a rather difficult task for forensic experts.

Ключевые слова. Цифровизация, законодательство, критерии, персональные, данные, судебная, экспертиза, искусственный, интеллект, приватность.

Keywords. Digitalization, legislation, criteria, personal, data, forensic, expertise, artificial intelligence, privacy.

Введение. Появление смарт-программ, с помощью которых осуществляется автоматизация рутинных умственных задач, в частности, распознавание отпечатка пальца, речи, лица и т.п., привело к их использованию в анализе иных персональных данных и данных о частной жизни-деятельности.

Персоны самостоятельно выкладывают в своих персональных страницах, блогах, сайтах, социальных сетях данные, отображающие их свойства, а также ситуации из частной жизни-деятельности. Персоны добровольно и/или вынужденно посещают учреждения и/или бизнес структуры, в которых также собираются персональные данные.

Все эти данные могут накапливаться в виде больших данных (мультимодальных данных) в хранилищах большого объема, в том числе в удаленных облачных хранилищах, в центрах обработки данных.

Цель работы. Учитывая то, что появились негативные явления с персональными данными, они приобрели существенное значение, т.к. их стали использовать мошенники в социальной инженерии. Всем известен фишинг, спаминг, навязчивая реклама, которая может привести на вредоносный сайт, и т.д. По оценке Всемирного экономического форума (ВЭФ), прошедшего в 2020 г., предполагаемая стоимость ущерба, причиненного хакерами, вредоносными программами и нарушениями данных, по прогнозам, достигнет $ 6 трлн. к 2021 году.

На вопрос: опасны ли системы собирающие персональные данные, многие люди говорят, что им нечего бояться, т.к. они не совершают преступлений. Однако объемы хранилищ данных практически не имеют границ, а системы искусственного интеллекта достигли такого уровня, что они практически принимают автоматические решения по поводу жизни-деятельностной ситуации конкретного человека. Например, уже сейчас банковские автоматические системы выносят решение о выдаче кредита, аналогичные юридические системы решают судьбу заключенного и т.д. Поэтому представляется целесообразным рассмотреть вопросы, связанные

с использованием юридического инструмента судебной экспертизы при расследовании преступлений в сфере компьютерной информации.

Материал и обсуждение. К сожалению, социальные сети, банки и иные операторы персональных данных мало уделяют внимания безопасности данных, что приводит к «сливу» сведений при их хранении и «перехвату» сообщений при их передаче в виде файловых данных. Дальше наступает очередь социальной инженерии, т.к. именно, люди, а не машины, часто переходят на вредоносную ссылку или непосредственно взаимодействуют с вредоносными субъектами.

На сайте ВЭФ опубликованы основные тенденции кибербезопасности [The Global Risks Report 2020 (дата размещения: 15 January 2020) URL: https://www.weforum.org/reports/the-global-risks-report-2020 (дата обращения: 03.03.2020)], в частности:

«ИИ по-прежнему используется в качестве посредника для совершения преступлений. Он также будет использоваться для ускорения мер реагирования в области безопасности. Большинство решений для обеспечения безопасности основаны на механизмах обнаружения, построенных на человеческой логике, но поддержание этого в актуальном состоянии против усложненных угроз и через новые технологии и устройства невозможно вручную. ИИ значительно ускоряет выявление новых угроз и реагирование на них, помогая блокировать атаки до того, как они смогут широко распространиться. Однако киберпреступники также начинают использовать те же методы, чтобы помочь им исследовать сети, находить уязвимости и разрабатывать более уклончивые вредоносные программы».

Многие ученые обращают внимание на то, что системы машинного обучения искусственного интеллекта, как бы их не называли (нейросети с глубинным обучением), на ограниченной выборке, какой бы она не была большой, не могут устранить их существенные недостатки, в частности, определить репрезентативность выборки данных, различить причину и

следствие в корреляции, а также обладают «катастрофическим забыванием». Так как «глубинное обучение» не учитывает не формализуемые свойства, то лучше эти операции называть «тренировкой», а сами такие системы – смарт-системами с тренировкой. Кроме того, эти системы работают на основе эвристического «черного ящика» на базе способа проб и ошибок, поэтому кроме отрицательной обратной связи они могут запускать и положительную обратную связь. А это приводит к негативным результатам.

Например, «Программа предупреждения преступности в Калифорнии предлагала отправлять больше полицейских в черные кварталы, основываясь на уровне преступности (количестве зафиксированных преступлений). А чем больше полицейских машин в области видимости, тем чаще жители сообщают о преступлениях (просто есть кому сообщить). В итоге преступность только возрастает — значит, надо отправить еще больше полицейских, и т. д.» [Этические вопросы искусственного интеллекта. (дата размещения: 30 августа 2018) URL: https://habr.com/ru/company/ kaspersky/blog/421791/ (дата обращения: 03.03.2020)].

Смарт-системами с тренировкой пока могут автоматизировать только рутинные умственные операции, а также в сложных и жизненно важных ситуациях должны представлять собой автоматизированные, а не автоматические системы.

Нет сомнения, что они будут использоваться для выявления преступников и правонарушителей не только в действительном, но и виртуальном мире, а также как автоматизированные ассистенты для лиц, принимающих решения, и наконец, в исследовательской деятельности, где требуется поиск неочевидных паттернов в больших массивах данных.

Любые достижения технического прогресса – это только инструменты, которые могут использоваться, как в позитивных, так и негативных целях. Хакерские инструменты применяют не только

физические лица, но и юридические лица, а также органы государственной власти. При этом особенностью персональных данных является то, что во многих случаях для операторов персональных данных важны не конкретные персональные данные, а статистически обработанные данные, т.к. именно они представляют наибольший интерес для принятия решений и оказания информационного воздействия на субъектов персональных данных. Однако, часто персонализированная реклама опаздывает, а таргетированная реклама, направленная на «целевую аудиторию», оказывается нерепрезентативной.

Рекомендации нейросети могут быть ошибочными в юридической деятельности, т.к. они не учитывают неформальное восприятие контрагента. Математическая модель нейросети при принятии решения фактически представляет собой самонастраивающийся фильтр, который просеивает входные данные, но этот фильтр формируется на основе критериев, которые подстраиваются под входной поток. Поэтому злоумышленник может направлять в этот поток объекты, которые очень похожи на правильные объекты, что может привести к деградации границы, разделяющей поддельные данные от подлинных. Такие действия еще называют взлом нейросети. Например, можно направить в самообучающийся спам-фильтр массу якобы чистых писем, которые научат фильтр пропускать спам, считающийся как чистые письма. О рисках использования нейросетей говорится в публикации [1].

Однако из этого не следует делать вывод, что нейросети нельзя использовать для выявления злоумышленников, наоборот, они широко используются для этого, но доверять математическим моделям принимать юридически значимые решения нельзя, т.к. квалифицированный человек всегда будет лучше оценивать сложную ситуацию, чем человекоподобная модель, т.к. в ней нужно думать, а не быстро решать. Поэтому правильно делается вывод в публикации о преждевременной ориентации на замену следователей и судей искусственным интеллектом [2].

Работа с персональными данными относится к сложным ситуациям, т.к. персональные данные не выделены в данных о персоне и не имеют четких критериев их «определения». В ФЗ РФ «О персональных данных» используемое слово «определение» метафорично выражено через отнесение (принадлежность) персональных данных конкретному субъекту персональных данных. Размытость критериев, тавтологичность, метафоричность и противоречивость, как определений понятий, так и норм этого закона, позволяет операторам персональных данных устанавливать связи обезличенных персональных данных с персональными данными и манипулировать ими в своих целях.

В связи с этим, необходимо использовать для юридических целей конструкцию (термин – дефиниция), а не (понятие – определение), создавать терминологические словари, на их основе тезаурусы, и далее информационные онтологии [3] юридических предметных областей знания. Это связано с тем, что осуществить цифровизацию законодательства и юридической деятельности, как их сквозной автоматизации, без наличия такой онтологии будет невозможно. Автоматизация отдельных частей юридической деятельности приведет к тому, что они не будут стыковаться между собой также, как и гос

ударственные услуги. Но важнее то, что без упорядочивания законодательства его математическая модель будет страдать уязвимостями, которыми могут воспользоваться злоумышленники.

В публикации [4] отмечено, что тот, кто обладает Artificial Intelligence of Things (AIoT) - (ИИ + большие данные + интернет вещей), на порядок сильнее и могущественнее того, у кого этого нет. Использование полицией, прокуратурой и судебной властью AIoT делает одну из сторон судебного разбирательства на порядок более мощной и вооруженной, чем другая. В публикации [5] считается, что искусственный интеллект могут использовать эксперты, хотя автор не раскрывает, как и для чего.

Для того, чтобы создавать цифровые системы необходимо упорядочить терминологию в информационной области, в частности в области персональных данных. Будем рассматривать слово «персона» для обозначения конституционной конструкции (человека и гражданина), поэтому персона может обладать свойствами человека (телесными и/или психическими) и/или субъекта общества, в частности, в виде физического лица (лицевые свойства), личности (личностные свойства) и иными общественными свойствами. Таким образом, можно говорить о личных, лицевых и/или личностных свойствах персоны, которые могут быть отображены в виде персональных данных.

Так как персональные данные входят в данные о персоне, которые кроме них еще содержат операционные данные, характеризующие персону, и/или иные данные, связанные с персоной, то необходимо более четко установить демаркацию между персональными данными и иными данными о персоне. Это важно, т.к. персональные данные могут быть, как открытыми, так и ограниченными и/или закрытыми.

Если персональные данные конкретной персоны хранятся на ее персональном электронном устройстве, то их содержимое может представлять собой приватную и/или секретную информацию [6]. Если же эта персона взаимодействует с онлайн сайтом и/или сервисом оператором персональных данных, то она добровольно должна дать согласие на доступ, обработку и/или использование своей конфиденциальной информации в персональных данных. Если в обработку входят операции глубокой обработки, анализа и/или синтеза конфиденциальной информации, в частности, с использованием нейросети с тренировкой, то оператор персональных данных должен предварительно поставить в известность персону для дачи согласия на это. Отметим, анонимизированные персональные данные не требуют согласия персон на их использование.

В связи с этим, необходимо уточнить операцию по об обезличенных данных, а также рассмотреть вопрос о соотношении баз данных

анонимизированных персональных данных как больших данных. При этом такие данные должны использоваться не только для научных и/или личных, но и экономических целей. Все это можно реализовать только при условии, когда законодатели четко определят критерии персональных и анонимизированных персональных данных.

В публикации [7] предложена спорная классификация персональных данных и высказано мнение, что «использование термина "цифровой след" или иных подобных понятий, характерных скорее для профессионального сленга или узкоспециализированных источников». Попробуем с этим не согласиться, т.к. криминалистика более точно характеризует данные о персоне в электронной среде. В электронной среде персона добровольно оставляет свои данные в двоичном формате как следы-отпечатки в определенный момент времени. Ее действия в этой среде в определенный интервал времени также оставляют цепочку следов в виде «теневых» данных от «электронного света». Учитывая наличие распределенной электронной среды, персона оставляет «зеркальные» следы-данные. Такие следы как криминалистически значимые следы могут быть востребованы при расследовании инцидентов в электронной среде.

Систематизация данных о персоне позволит сформулировать критерии для демаркации персональных данных от анонимизированных персональных данных, а также определить логические связи персональных данных с псевдонимизированными персональными данными.

Выводы. Пока информационное законодательство не будет упорядочено, судьи будут назначать судебные экспертизы с вопросами о наличие логических связей у обезличенных персональных данных с персональными данными конкретного субъекта персональных данных, а экспертам придется разбираться с критериями и создавать методики исследования таких связей с учетом используемого программного обеспечения оператора персональных данных. Разработка, апробация и

утверждение таких методик ставит достаточно сложную задачу перед судебными экспертами.

Библиографический список:

1. Бахтеев Д.В. Риски и этико-правовые модели использования систем искусственного интеллекта // Юридические исследования. 2019. № 11. С. 1 - 11.

2. Афанасьев А. Ю. Искусственный интеллект или интеллект субъектов выявления, раскрытия и расследования преступлений: что победит? // Библиотека криминалиста. 2018. № 3(38). С. 28-34.

3. Nesterov, A.V. On the Unification of the Conceptual Model of the Meta-Ontology // Scientific and Technical Information Processing, 2019, Vol. 46, No. 1, pp. 34–37.

4. Овчинский Владимир, Бинецкий Алексей. Судья с искусственным интеллектом (дата размещения: 13.02.2019) URL: http://zavtra.ru/blogs/sud_ya_s_iskusstvennim_intellektom (дата обращения: 02.03.2020).

5. Шамовка А. Искусственный интеллект — будущее криминалистических расследований. (дата размещения: 08 июля 2019). URL: https://www.anti-malware.ru/analytics/Technology_Analysis/ artificial-intelligence-is-computer-forensicfuture?utm_ source=google&amp%3Butm_medium=email&amp%3Butm_campaign=amdelivery (дата обращения: 02.03.2020).

6. Нестеров А.В. Место персональных данных в сведениях, отображающих персону: судебно-экспертный аспект // Актуальные проблемы административного, финансового и информационного права в России и за рубежом. 25.01.2019. М.: РУДН, С. 54-61.

7. Савельев А.И. На пути к концепции регулирования данных в условиях цифровой экономики // Закон. 2019. N 4. С. 174-195.

References:

1. Bakhteev D. V. Risks and ethical and legal models of using artificial intelligence systems // Legal research. 2019. № 11. P. 1-11.

2. Afanasiev A. Yu. Artificial intelligence or intelligence of subjects of detection, disclosure and investigation of crimes: what will win? // Criminalist's library. 2018. no. 3(38). Pp. 28-34.

4. Ovchinsky Vladimir, Binetsky Alexey. Judge with artificial intelligence (date of placement: 13.02.2019) URL: http://zavtra.ru/blogs/ sudya_s_iskusstvennym_intellektom (accessed: 02.03.2020).

5. Shamovka A. Artificial intelligence is the future of forensic investigations. (placement date: July 08, 2019). URL: https://www.anti-malware.ru/analytics/Technology_Analysis/artificial-intelligence-is-computerforensicfuture?utm_source=google&amp%3Butm_medium=email&amp%3Butm_campaign=amdelivery (date accessed: 02.03.2020).

6. Nesterov A.V. the Place of personal data in information displaying a person: forensic aspect // Actual problems of administrative, financial and information law in Russia and abroad. 25.01.2019. Moscow: RUDN, P. 54-61.

7. Savelev A. I. On the way to the concept of data regulation in the digital economy // Law. 2019. N 4. Pp. 174-195.

Сведения об авторе: Нестеров Анатолий Васильевич, д.ю.н., профессор, профессор Юридического института Российского университета дружбы народов, г. Москва, ул. Миклухо-Маклая 6, профессор Российской таможенной академии, г. Люберцы, Комсомольский просп. 4. nesterav@yandex.ru .

Information about the author: Anatoly Nesterov, doctor of law, Professor, Professor of the Law Institute of the peoples ' friendship University of Russia, 6 Miklukho-Maklaya str., Moscow, Professor of the Russian customs Academy, 4 Komsomolsky Ave., Lyubertsy. nesterav@yandex.ru.