УДК 004.8+004.056

Л.В. Аршинский, М.С. Жукова, Г.Н. Шурховецкий

Иркутский государственный университет путей сообщения, г. Иркутск, Российская Федерация

ПРОБЛЕМЫ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ В СИСТЕМАХ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА, ИСПОЛЬЗУЮЩИХ МОДЕЛЬ ПРЕДМЕТНОЙ ОБЛАСТИ

Аннотация. В работе исследуются проблемы информационной безопасности интеллектуальных информационных систем, использующих в своём составе модель предметной области: базу знаний для экспертных систем, архитектурный файл для искусственных нейронных сетей и т.п. Вопрос рассматривается с точки зрения защиты команд, данных и коммуникации. Делается вывод, что защищать надо в первую очередь данные в форме модели предметной области как «интеллектуальное ядро» системы. Защита исполнимых файлов вторична, хотя тоже важна. Защита коммуникации может осуществляться традиционными средствами. Для защиты модели предлагается обратить внимание на такой приём, как её фрагментацию с последующим разнесением фрагментов по различным, в том числе географически удалённым хранилищам.

Ключевые слова: искусственный интеллект, информационная безопасность, экспертные системы, искусственные нейронные сети, метод рассечения-разнесения.

L.V. Arshinskiy, M.S. Zhukova and G.N Shurkhovetsky

Irkutsk State Transport University, Irkutsk, Russian Federation

PROBLEMS OF INFORMATION SECURITY IN ARTIFICIAL INTELLIGENCE SYSTEMS USING A DOMAIN MODEL

Abstract. The work examines the problems of information security of intelligent information systems that use a domain model - a knowledge base for expert systems, an architectural file for artificial neural networks, etc. The issue is considered from the point of view of protecting commands, data and communications. It is concluded that it is necessary to protect, first, data in the form of a domain model as the “intellectual core” of the system. Protection of executable files is secondary, although also important. Communication protection can be achieved by traditional means. To protect the model, it is proposed to pay attention to such a technique as its fragmentation with the subsequent distribution of fragments across various, including geographically remote, storage facilities.

Keywords: artificial intelligence, information security, expert systems, artificial neural networks, dissection-placing method.

Введение. Методы и технологии искусственного интеллекта (ИИ) – очевидный тренд современного технологического развития. Зародившись в середине 50-х гг XX века, они прочно заняли своё место в мире. И хотя говорить об их повсеместном распространении пока рано, существует немало направлений, где они используются всё больше. Среди востребованных направлений – вопросно-ответные системы (в первую очередь чат-боты, особенно – основанные на больших лингвистических моделях), автоматизированные системы управления на предприятиях (АСУ и АСУТП), робототехника, медицина, сельское хозяйство, транспорт, государственное и муниципальное управление, искусство, военное дело и многое другое [1-4].

Пройдя через множество этапов как возрастания, так и снижения интереса к ним [5], технологии ИИ созрели до степени признания на законодательном уровне [6, 7].

В [5] ИИ определён как «комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая самообучение и поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека. Комплекс технологических решений включает в себя информационно-коммуникационную инфраструктуру (информационные системы, информационно-телекоммуникационные сети, иные технические

средства обработки информации), программное обеспечение (в котором используются методы машинного обучения), процессы и сервисы по обработке данных и поиску решений.

В литературе, в обществе широко обсуждаются различные риски, которые влечёт ИИ. Это и непредсказуемость, и банальные сбои в условиях, когда под управление ИИ переданы жизненно важные процессы общества, и неправовое вмешательство в частную жизнь граждан, формирование у общества ложных взглядов и представлений, и многое другое [7-9].

В бытовом сознании активно живут представления об угрозе «захвата» и «порабощения» человечества некоей сверхсистемой ИИ, которая преследует свои, не относящиеся к людям цели и интересы (та самая непредсказуемость).

Все перечисленные риски и угрозы так или иначе относятся к влиянию ИИ на людей, хотя некоторые из них носят скорее газетно-публицистический, чем реальный характер [9]. В то же время существует и обратная, причём гораздо более актуальная угроза, – влияние людей на ИИ, – злонамеренные воздействия на системы ИИ для достижения своих, вполне человеческих целей. То есть угроза не людям со стороны ИИ, а угроза людям со стороны других людей через воздействие на системы ИИ.

Классы систем ИИ и их уязвимости. Системы искусственного интеллекта (СИИ) – это широкий класс программных продуктов. Сюда входят и искусственные нейронные сети (ИНС – то, к чему зачастую сводится ИИ в общественном, и не только, сознании), и системы, основанные на знаниях (СОЗ), их наиболее известный подкласс – экспертные системы (ЭС), интеллектуальные информационно-поисковые системы, превращающиеся сегодня в большие лингвистические модели (те же ИНС), и многоагентные системы [10], метаэвристические алгоритмы (алгоритмы направленного поиска), системы моделирования творческих процессов и т.п. [11]. К ним же можно отнести и системы распознавания образов, хотя здесь среди специалистов нет единого мнения.

Рассмотрим потенциальные уязвимости технологий ИИ с точки зрения информационной безопасности (ИБ). Под ИБ понимается состояние защищенности информационных ресурсов (информационной среды) от внутренних и внешних угроз, способных нанести ущерб интересам личности, общества, государства (национальным интересам) [12]. Здесь её сузим до состояния защищённости СИИ и даже более того – программной реализации таких систем.

Всякая (компьютерная) программа состоит из команд и данных. Команды – активная, данные – пассивная составляющие программного обеспечения (ПО). Кроме того, сегодня важное место в функционировании многих программно-информационных систем занимает коммуникация – взаимодействие программ или их компонентов в глобальной сети или сети предприятия. Таким образом разбиваем проблему на три составляющие (рис. 1):

– защита команд;
– защита данных;
– защита коммуникации.

Рис. 1. Защита СИИ

Защита команд. Вторжение в систему команд – пожалуй самая неоднозначная составляющая компьютерной агрессии. С одной стороны, мы можем полностью и целенаправленно

изменить поведение системы. С другой – такой вид атаки требует глубоких знаний поведения ПО, тщательного исследования кодов, уверенности, что код не будет планово обновлён. Атакующему важно вмешаться в команды незаметным для атакуемого образом, чтобы достичь долговременных целей, а не просто разрушить ПО.

Теоретически, вмешаться в команды можно подменив, к примеру, секцию импорта исполняемого файла (список внешних функций или библиотек, которые необходимы программе для выполнения). Но надо быть уверенным, что программа не прекратит своей работы, не найдя необходимого ей ресурса, что также требует тщательного изучения кода.

Примером деструктивного вмешательства в команды являются компьютерные вирусы, но для них программа – главным образом платформа, на которой они достигают своих целей. Да и сами вирусы чаще всего пишутся так, чтобы их существование оставалось не замеченным жертвой максимально долгий срок. Т.е. – целевая функциональность программы не нарушается.

(Примечание. Оставляем в стороне вирусы-вымогатели и другие аналогичные «зловреды», которые быстро обнаруживают себя, ибо имеют совсем иные цели.)

Сложности атакующему добавляет и то, что команды могут сжиматься или шифроваться.

Таким образом представляется, что изменение системы команд – самая краткосрочная из агрессий, которая, скорее всего, будет выявлена достаточно быстро. Даже при возможной доступности команд этот вид атаки в долговременной перспективе выглядит не самым удачным.

Защита данных. Второй составляющей всякого ПО являются данные. В небольших программах они зачастую составляют часть кода и потому вмешательство здесь также довольно проблематично, хотя и возможно. Интереснее и перспективнее такие атаки выглядят, если данные хранятся отдельно от исполняемого файла, как например, при использовании баз данных и других подобных файлов (файлов внешнего хранения данных).

В СИИ принцип внешнего хранения данных широко используется. Например, в ЭС это база знаний (БЗ) – внешний файл или система файлов, в которых хранится (знаниевая) модель предметной области (ПрО). В ИНС во внешних файлах нередко размещаются коэффициенты нейронов сети и другая архитектурная информация. Коэффициенты, в совокупности с архитектурой ИНС, – это тоже модель ПрО. Общая структура СИИ, основанных на моделях ПрО показана на рис. 2.

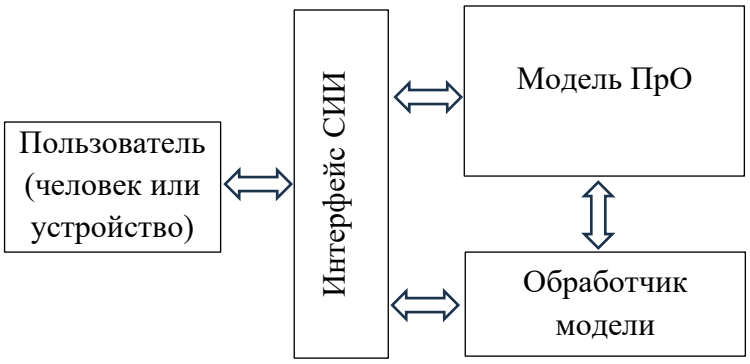

Рис. 2. Общая структура СИИ

Обработчик модели – это универсальный модуль, на работу которого влияет модель; в ЭС это решатель, в ИНС – обработчик связей. Таким образом, вмешательство в БЗ ЭС или коэффициенты ИНС – это вмешательство в модель предметной области, её искажение. Если проводить аналогию с человеческим интеллектом, искажение модели – это разрушение или нарушение представлений человека о мире, разрушение его когнитивных способностей. В СИИ это приведёт к неверному решению задач. Например, неверному принятию решений

(скажем, в АСУТП), неверной классификации или идентификации образа (распознавание образов, ИНС), неверному ответу на вопрос (ЭС). Легко представить возможные последствия, если речь идёт о вмешательстве в АСУТП для критической технологии, или в работу ЭС, если её рекомендации ложатся в основу важных решений.

Интересной темой является не случайное, а целенаправленное искажение модели с целью заставить систему приходить к нужным для атакующего решениям. В ИНС для этого надо учитывать взаимосвязи и веса входов нейронов, знать, что и как обрабатывает сеть. Однако если знать архитектуру атакуемой сети и соответствующая информация хранится во внешнем файле, атакующий может испортить файл или создать у себя аналог ИНС, обучить его деструктивным действиям, после чего подменить соответствующий файл на файл, обученный на аналоге.

В системах, основанных на знаниях, где знания хранятся в явном текстовом виде, можно попытаться переписать фрагменты БЗ. При этом система внешне сохраняет работоспособность и, если не принять специальных мер, пользователь может не сразу заметить изменения.

Для ИНС и систем распознавания образов, кроме того, известны атаки отравлением данных (Data Poisoning, см. напр. [13-15]). Поскольку эти системы настраиваются на обучающих выборках, внесение в эти выборки искажённых данных (отравление) делает зловредным или как минимум бесполезным результат обучения. В [14], к примеру, отмечено, что трёхпроцентное отравление обучающего набора одной из систем снизило её точность на одиннадцать процентов.

В [15] этот метод представлен как метод защиты со стороны авторов от использования их интеллектуальной собственности для обучения СИИ.

Интересный аспект вмешательства в данные интеллектуальных систем состоит в том, что, как отмечается в [16] и ряде других источников, знания (модель ПрО), хранящиеся в интеллектуальной системе, содержат в себе признаки как данных, так и команд. Это называется активностью знаний. Если знания размещаются во внешнем файле, то вмешательство в этот файл аналогично вмешательству в систему команд, что даёт злоумышленнику дополнительные возможности, отсутствующие при работе с традиционным ПО.

Защита коммуникации. Наконец, третий рассматриваемый аспект – защита коммуникации. Одним из подходов к разработке современных ИС является сервис-ориентированные архитектуры, распределённое хранение данных, облачные вычисления и другие подобные технологии. Во всех случаях неотъемлемой частью функционирования ИС является обмен данными между ЭВМ, в том числе географически удалёнными. Обмен данными, распределённое хранение информации возможны и в системах ИИ. Также системы ИИ сами могут быть частью системы хранения данных [17].

Основными целями атаки на коммуникацию рассматривают [18]:

1) нарушение целостности и достоверности передаваемой информации;
2) нарушение конфиденциальности передаваемой информации;
3) нарушение доступности информации системы в целом или отдельных её частей.

Представляется, что для СИИ наиболее актуальным являются первая и третья составляющие. Нарушение конфиденциальности здесь достаточно частный случай, а вот вмешательство в функционирование программы – вполне реальная цель. Указанные цели могут достигаться разными способами [18, 19]. В частности, путём внедрения нерегламентированных возможностей, перехватом телекоммуникационного трафика, изменением параметров СИИ, внедрением ложного доверенного объекта (БЗ ЭС, обучающей выборки, модели ИНС и т.п.), скрытной инсталляции программ удалённого управления и т.д.

Средства коммуникации могут использоваться для блокирования доступа к ключевым файлам СИИ, тайного копирования информации из них (например, для изучения и последующей замены или отравления), наконец, просто для выведения системы из строя.

Серьёзную проблему может создавать т.н. «расширенная постоянная угроза», позволяющая контролировать функционирование СИИ.

Атаки на систему доменных имён могут перенаправить технологический трафик обмена командами и данными и т.п.

Вообще, развитие и усложнение сетевых технологий могут порождать и порождают новые угрозы, которые не всегда можно принять во внимание (угрозы нулевого дня).

Меры противодействия. Полагаем, что данные – самая «интересная» для злоумышленника часть СИИ. Если злоумышленник не является сотрудником соответствующей организации, получить к ним доступ он может через внедрённые в ЭВМ потенциальной жертвы зловредные файлы, а также через компьютерную сеть при обмене технологическими данными внутри распределённой СИИ. Есть ещё методы воздействия на работников соответствующей организации (давление, социальная инженерия и т.п.), но здесь затрагиваются только программно-технические аспекты такого воздействия.

К сожалению, не все данные и не всегда могут быть защищены. Например, если ИНС обучается на основе глобальных данных, размещённых в интернет (те же большие лингвистические модели), это просто не реально. Обсудим только данные, используемые в СИИ.

Рассмотрим два типа нарушителей: внешний и внутренний.

Внешний нарушитель. Первым и уже достаточно традиционным средством защиты от внешнего нарушителя является шифрование. Невозможность разобраться с семантикой модели ПрО, заложенной в СИИ, не позволит атакующему целенаправленно её изменить. Только разрушить. Но эта угроза парируется резервным копированием модели. С момента обнаружения сбоя в работе системы до её восстановления понадобится не так много времени. При этом говорить о долговременном и незаметном для пользователя вмешательстве в работу системы здесь уже не приходится. Тактика разрушения оправдана только, если последствия от атаки наступят быстро и являются фатальными. К примеру – разрушение АСУТП для критических производств, вроде того, как описано в [20].

Если СИИ отделена от глобальной сети, получить к ней на постоянной основе скрытый внешний доступ затруднительно (возможны только разовые акции вроде [20]). Если же она участвует в коммуникации по глобальным сетям, подобные атаки теоретически возможны. Для противодействия им целесообразно использовать защищённые интернет-протоколы, VPN-технологии и другие подобные средства. Также, как минимум, следует привлекать традиционные средства защиты внутренней сети предприятия.

Интересным приёмом защиты внешних данных может быть разделение критической информации и разнесение её по разным местам хранения (хранилищам) [17] (рис. 3).

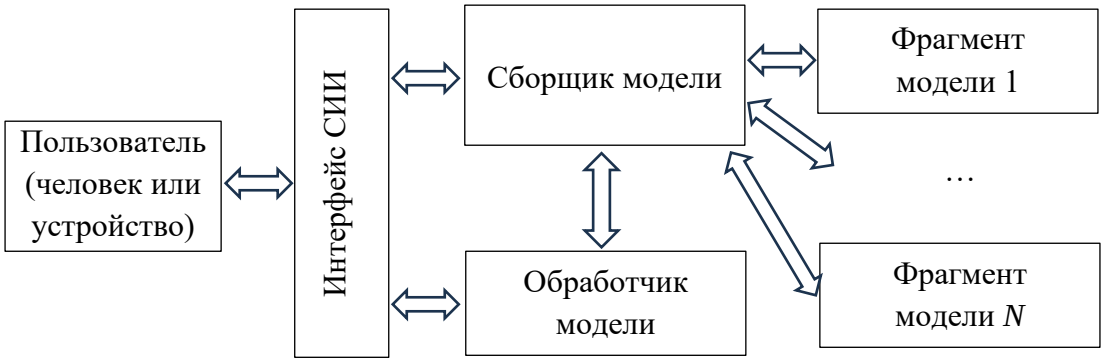

Рис. 3. СИИ с распределённым, в т.ч. географически распределённым хранением модели ПрО

Данные в этом случае перестают быть локальными и монолитными, а злоумышленник, не знающий мест хранения, получает дополнительные, возможно даже непреодолимые для него сложности по их поиску. Задача ещё усложняется, если данные в хранилищах будут зашифрованы.

Ещё более серьёзную степень защиты при раздельном хранении данных обеспечивает метод побитового рассечения-разнесения в варианте алгоритма [21]. Он гарантирует практически полную невозможность извлечения осмысленной информации за приемлемое время даже при отсутствии криптографических средств. Кроме того, эта технология делает необязательным использование закрытых каналов связи. Криптографирование здесь становится непринципиальным (хотя и дополнительно повышает степень защищённости). Расшифровка данных, даже собранных со всех хранилищ (атака «злоумышленник владеет всеми потоками»), здесь ничего не даст. Вскрыть исходный файл данных может только нарушитель внутренний (а это ограниченный круг лиц), знающий порядок сборки, либо внедрённое извне программное обеспечение, получившее доступ к описанию такого порядка при его ненадлежащем хранении.

Внутренний нарушитель. Для защиты от внутренней угрозы первоочередная мера так же традиционна – разграничение доступа. Для критических СИИ – не только логического, но и физического. Ограничение круга допущенных к работе с СИИ и тем более к редактированию модели ПрО должно быть обязательной составляющей безопасности. Здесь нет ничего уникального, характерного только для СИИ. При необходимости может быть использован весь комплекс средств: контроль лиц, получающих физический доступ, оборудование дверей надежными замками, применение видеонаблюдения, защита серверного оборудования, контроль доступа к вычислительной технике и прочее.

Заключение. В результате выполненного анализа можно сделать следующие предварительные выводы:

I. Защищать надо в первую очередь модель ПрО – «интеллектуальное ядро» системы – базу знаний для экспертных систем, архитектурный файл для искусственных нейронных сетей. Для ИНС, а также систем распознавания образов защиты требуют также обучающие выборки.

II. Интересным приёмом защиты интеллектуального ядра является его распределённое хранение на разных, в том числе географически удалённых серверах, местонахождение которых не известно нарушителю. Фрагменты при этом следует шифровать, или использовать иные методы защиты, обеспечивающие сопоставимую стойкость.

III. Эффективным приёмом распределённого хранения является использование для этого метода побитового рассечения-разнесения. В этом случае даже доступ ко всем фрагментам ядра практически не оставляет злоумышленнику шансов получить доступ к оригинальной информации за приемлемое время, даже если фрагменты не зашифрованы и передаются по открытым каналам. Шифрование же ещё более ухудшает его ситуацию.

IV. Защита исполнимых файлов вторична, хотя тоже важна.

V. Защита коммуникации может осуществляться традиционными средствами.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Лапушкин А. Сферы применения систем искусственного интеллекта. – URL: https://maff.io/media/sfery-primeneniya-sistem-iskusstvennogo-intellekta/ (дата обращения: 21.03.2024).

2. Сферы применения искусственного интеллекта: от медицины до сельского хозяйства. – URL: https://gb.ru/blog/sfery-primeneniya-iskusstvennogo-intellekta/ (дата обращения: 21.03.2024).

3. На что способен искусственный интеллект сегодня и каков его потенциал. – URL: https://trends.rbc.ru/trends/industry/cmrm/619766d59a79471862e77e8a (дата обращения: 21.03.2024).

4. 5 применений ИИ, в которых он конкурирует с человеком. – URL: https://habr.com/ru/companies/toshibarus/articles/580930/ (дата обращения: 21.03.2024).

5. Путь искусственного интеллекта от фантастической идеи к научной отрасли. – URL: https://habr.com/ru/companies/cloud4y/articles/469447/ (дата обращения: 21.03.2024).

6. Федеральный закон от 24.04.2020 N 123-ФЗ. «О проведении эксперимента по установлению специального регулирования в целях создания необходимых условий для разработки и внедрения технологий искусственного интеллекта в субъекте Российской Федерации - городе федерального значения Москве и внесении изменений в статьи 6 и 10 Федерального закона «О персональных данных». – URL: http://publication.pravo.gov.ru/Document/View/0001202004240030 (дата обращения: 21.03.2024).

7. Ли Яо. Нормативно-правовое регулирование генеративного искусственного интеллекта в Великобритании, США, Европейском союзе и Китае // Право. Журнал Высшей школы экономики, 2023. – Том 16, № 3. – С. 245-267.

8. Морхат П.М. Риски и угрозы, связанные с применением искусственного интеллекта // Аграрное и земельное право, 2017. – №12(156). – С. 60-65.

9. Артамонов В.А., Артамонова Е.В. Проблемы искусственного интеллекта: мифы и реальность. – URL: https://cyberleninka.ru/article/n/problemy-iskusstvennogo-intellekta-mify-i-realnost/viewer (дата обращения: 21.03.2024).

10. Рассел С., Норвиг П. Искусственный интеллект: современный подход, 2-е изд.: Пер. с англ. – М.: Издательский дом «Вильямс», 2006. – 1408 с.

11. Аршинский Л.В., Жукова М.С. Интеллектуальные информационные системы и технологии: учебное пособие. – Иркутск: ИрГУПС, 2023. – 128 с.

12. Вострецова Е.В. Основы информационной безопасности: учебное пособие для студентов вузов. – Екатеринбург: Изд-во Урал. Ун-та, 2019. – 204 с.

13. Намиот Д.Е. Введение в атаки отравлением на модели машинного обучения // International Journal of Open Information Technologies, 2023. – Т. 11, № 3. – С. 58-66.

14. Bagdasaryan E., Shmatikov V. Blind backdoors in deep learning models // 30th USENIX Security Symposium, 2021. – pp.1505-1521.

15. Апресов С. Что такое «отравление данных». Методы защиты от атак data poisoning. – URL: https://digitalocean.ru/n/podryvnaya-deyatelnost/ (дата обращения: 28.03.2024).

16. Гаскаров Д.В. Интеллектуальные информационные системы. Учеб. для вузов. – М.: Высш. шк., 2003. – 431 с.

17. Жилин В.В., Сафьян О.А. Искусственный интеллект в системах хранения данных // Вестник Донского государственного технического университета, 2020. Т. 20, № 2. – С. 196-200.

18. Бабенко Г.В. Анализ современных угроз безопасности информации, возникающих при сетевом взаимодействии // Вестник АГТУ. Сер.: Управление, вычислительная техника и информатика, 2010. – № 2. – URL: https://cyberleninka.ru/article/n/analiz-sovremennyh-ugroz-bezopasnosti-informatsii-voznikayuschih-pri-setevom-vzaimodeystvii/viewer (дата обращения: 28.03.2024).

19. Вартанян А.А. Угрозы и атаки сетевой безопасности. – URL: https://cyberleninka.ru/article/n/ugrozy-i-ataki-setevoy-bezopasnosti/viewer (дата обращения: 28.03.2024).

20. Ромашкина Н.П., Махукова А.В. Компьютерная вредоносная атака на ядерную программу Ирана // Информационные войны, 2013. – № 4. – С. 40-50.

21. Аршинский Л.В., Шурховецкий Г.Н. Особенности применения метода рассечения-разнесения для безопасного хранения данных во внешних хранилищах // Информационные технологии, 2021. №5. т. 27. С. 259-266. DOI: 10.17587/it.27.259-266

REFERENCES

1. Lapushkin A. Sfery primeneniya sistem iskusstvennogo intellekta [Areas of application of artificial intelligence systems]. – https://maff.io/media/sfery-primeneniya-sistem-iskusstvennogo-intellekta/ (21.03.2024).

2. Sfery primeneniya iskusstvennogo intellekta: ot meditsiny do sel'skogo khozyaystva [Areas of application of artificial intelligence: from medicine to agriculture]. – https://gb.ru/blog/sfery-primeneniya-iskusstvennogo-intellekta/ (21.03.2024).

3. Na chto sposoben iskusstvennyy intellekt segodnya i kakov yego potentsial [What is artificial intelligence capable of today and what is its potential. – https://trends.rbc.ru/trends/industry/cmrm/619766d59a79471862e77e8a (21.03.2024).

4. 5 primeneniy II, v kotorykh on konkuriruyet s chelovekom [5 applications of AI in which it competes with humans]. – https://habr.com/ru/companies/toshibarus/articles/580930/ (21.03.2024).

5. Put' iskusstvennogo intellekta ot fantasticheskoy idei k nauchnoy otrasli [The path of artificial intelligence from a fantastic idea to a scientific industry]. – https://habr.com/ru/companies/cloud4y/articles/469447/ (21.03.2024).

6. Federal'nyy zakon ot 24.04.2020 N 123-FZ. «O provedenii eksperimenta po usta-novleniyu spetsial'nogo regulirovaniya v tselyakh sozdaniya neobkhodimykh usloviy dlya razra-botki i vnedreniya tekhnologiy iskusstvennogo intellekta v sub"yekte Rossiyskoy Federatsii - gorode federal'nogo znacheniya Moskve i vnesenii izmeneniy v stat'i 6 i 10 Federal'nogo zakona «O personal'nykh dannykh» [Federal Law of April 24, 2020 N 123-FZ. “On conducting an experiment to establish special regulation in order to create the necessary conditions for the development and implementation of artificial intelligence technologies in a constituent entity of the Russian Federation - the federal city of Moscow and amending Articles 6 and 10 of the Federal Law “On Personal data."]. – http://publication.pravo.gov.ru/Document/View/0001202004240030 (21.03.2024).

7. Li Yao. *Normativno-pravovoye regulirovaniye generativnogo iskusstvennogo intellekta v Velikobritanii SSHA, Yevropeyskom soyuze i Kitaye* [Legal regulation of generative artificial intelligence in the UK, USA, European Union and China] // *Pravo. Zhurnal Vysshey shkoly ekonomiki* [Law. Journal of Higher School of Economics], 2023, vol. 16, no 3, pp. 245-267.

8. Morhat P.M. *Riski i ugrozy, svyazannyye s primeneniyem iskusstvennogo intellekta* [Risks and threats associated with the use of artificial intelligence] // *Agrarnoye i zemel'noye pravo* [Agricultural and land law], 2017, no 12 (156), pp. 60-65.

9. Artamonov V.A. and Artamonova E.V. *Problemy iskusstvennogo intellekta: mify i real'nost'* [Problems of artificial intelligence: myths and reality]. – https://cyberleninka.ru/article/n/problemy-iskusstvennogo-intellekta-mify-i-realnost/viewer (21.03.2024).

10. Russell S. and Norvig P. *Iskusstvennyy intellekt: sovremennyy podkhod, 2-ye izd.: Per. s angl*. [Artificial intelligence: a modern approach, 2nd ed.: Transl. from English], *M.: Izdatel'skiy dom «Vil'yams»* [Moscow, Williams Publishing House], 2006, 1408 p.

11. Arshinskiy L.V. and Zhukova M.S. *Intellektual'nyye informatsionnyye sistemy i tekhnologii: uchebnoye posobiye* [Intelligent information systems and technologies: textbook], *Irkutsk: IrGUPS* [Irkutsk: Irkutsk State Transport University], 2023, 128 p.

12. Vostretsova E.V. *Osnovy informatsionnoy bezopasnosti: uchebnoye posobiye dlya studentov vuzov* [Fundamentals of information security: a textbook for university students], Yekaterinburg: Izd-vo Ural. Un-ta [Ekaterinburg: Ural Univ. Publishing House], 2019. – 204 p.

13. Namiot D.E. *Vvedeniye v ataki otravleniyem na modeli mashinnogo obucheniya* [Introduction to poisoning attacks on machine learning models] // International Journal of Open Information Technologies, 2023, vol. 11, no 3, pp. 58-66.

14. Bagdasaryan E. and Shmatikov V. Blind backdoors in deep learning models // 30th USENIX Security Symposium, 2021. pp. 1505-1521.

15. Apresov S. *Chto takoye «otravleniye dannykh». Metody zashchity ot atak data poisoning* [What is “data poisoning”. Methods of protection against data poisoning attacks]. – https://digital-ocean.ru/n/podryvnaya-deyatelnost/ (21.03.2024).

16. Gaskarov D.V. *Intellektual'nyye informatsionnyye sistemy. Ucheb. dlya vuzov* [Intelligent information systems. Textbook for universities], *M.: Vyssh. shk* [Moscow, Higher. school], 2003, 431 p.

17. Zhilin V.V. and Safyan O.A. *Iskusstvennyy intellekt v sistemakh khraneniya dannykh* [Artificial intelligence in data storage systems] *// Vestnik Donskogo gosudarstvennogo tekhnicheskogo universiteta* [Bulletin of the Don State Technical University], 2020, vol. 20, no. 2, pp. 196-200.

18. Babenko G.V. *Analiz sovremennykh ugroz bezopasnosti informatsii, voznikayushchikh pri setevom vzaimodeystvii* [Analysis of modern threats to information security that arise during network

interaction] // *Vestnik AGTU. Ser.: Upravleniye, vychislitel'naya tekhnika i informatika* [Vestnik ASTU. Ser.: Management, computer technology and information science], 2010, no. 2. – https://cyberleninka.ru/article/n/analiz-sovremennyh-ugroz-bezopasnosti-informatsii-voznikay-uschih-pri-setevom-vzaimodeystvii/viewer (28.03.2024).

19. Vartanyan A.A. *Ugrozy i ataki setevoy bezopasnosti* [Network security threats and attacks]. – https://cyberleninka.ru/article/n/ugrozy-i-ataki-setevoy-bezopasnosti/viewer (28.03.2024).

20. Romashkina N.P. and Makhukova A.V. *Komp'yuternaya vredonosnaya ataka na yadernuyu programmu Irana* [Computer malicious attack on Iran's nuclear program] // *Informatsionnyye voyny* [Information Wars], 2013, no. 4, pp. 40-50.

21. Arshinskiy L.V. and Shurkhovetsky G.N. *Osobennosti primeneniya metoda rassecheniya-razneseniya dlya bezopasnogo khraneniya dannykh vo vneshnikh khranilishchakh* [Features of using the dissection-diversity method for secure data storage in external storage facilities] // *Informatsionnyye tekhnologii* [Information Technologies], 2021, no. 5. vol. 27. pp. 259-266. DOI: 10.17587/it.27.259-266.

Информация об авторах

Аршинский Леонид Вадимович – д.т.н., доцент, профессор кафедры «Информационные системы и защита информации», Иркутский государственный университет путей сообщения, г. Иркутск, e-mail: larsh@mail.ru.

Жукова Марина Сергеевна – магистрант кафедры «Информационные системы и защита информации», Иркутский государственный университет путей сообщения, г. Иркутск, e-mail: marino_@mail.ru.

Шурховецкий Георгий Николаевич – ассистент кафедры «Информационные системы и защита информации», Иркутский государственный университет путей сообщения, г. Иркутск, e-mail: gshn5@yandex.ru.

Authors

Leonid Vadimovich Arshinskiy – Doctor of Technical Science, professor of department "Information Systems and Information Security", Irkutsk State Transport University, Irkutsk, e-mail: larsh@mail.ru.

Marina Sergeevna Zhukova – master's student of department "Information Systems and Information Security", Irkutsk State Transport University, Irkutsk, e-mail: marino_@mail.ru.

Georgy Nikolaevich Shurkhovetsky – assistant of the department "Information systems and information security", Irkutsk State University of Transport, Irkutsk, e-mail: gshn5@yandex.ru.

Для цитирования

Аршинский Л.В., Жукова М.С., Шурховецкий Г.Н. Проблемы информационной безопасности в системах искусственного интеллекта, использующих модель предметной области // Информационные технологии и математическое моделирование в управлении сложными системами: электрон. науч. журн. 2024. №1 С.36-44. – Режим доступа: http://ismm-irgups.ru/toma/121-2024, свободный. – Загл. с экрана. – Яз. рус., англ. (дата обращения: 17.04.2024)

For citations

L.V. Arshinskiy, M.S. Zhukova and G.N. Shurkhovetsky, Problems of information security in artificial intelligence systems using a domain model // Informacionnye tehnologii i matematicheskoe modelirovanie v upravlenii slozhnymi sistemami: ehlektronnyj nauchnyj zhurnal [Information technology and mathematical modeling in the management of complex systems: electronic scientific journal], 2024. No.1. P.36-44. [Accessed 17.04.2024].