DOI: 10.55648/1998-6920-2023-17-3-78-86 УДК 004.056.5

Разработка методологии защиты системы искусственного интеллекта в распределенных информационных системах *

С. И. Штеренберг

Московский технический университет связи и информатики (МТУСИ)

Аннотация: Данная статья посвящена методологии строительства искусственных интеллектуальных систем для задач защиты информации. Защищать же планируется сам искусственный интеллект (ИИ) и информацию, которую данное устройство будет обрабатывать. Мероприятие масштабное по своей величине и автор данного текста постарается представить не просто концепт, скорее идею по достижению цели данного результата. Под результатом будем понимать итоговое строительство защищенного ИИ, созданного для организации обеспечения информационной безопасности (ИБ).

Ключевые слова: искусственный интеллект, нейронные сети, большие данные, машинное обучение, квазибиологическая парадигма.

Для цитирования: Штеренберг С. И. Разработка методологии защиты системы искусственного интеллекта в распределенных информационных системах // Вестник СибГУТИ. 2023. Т. 17, № 3. С. 78–86. https://doi.org/10.55648/1998-6920-2023-17-3-78-86.

Статья поступила в редакцию 26.12.2022;
принята к публикации 10.01.2023.

1. Введение

Сегодня, если мы возьмем общий уровень интеллектуального развития человечества, собрать умную машину не представляется сложным занятием [1]. Нам известен простой закон Мура, в котором наблюдалась устойчивая динамика роста количества транзисторов, размещаемых на кристалле интегральной схемы. Кажется на первый взгляд, что именно это и есть один из многих эволюционных факторов, которые могут лечь в основу так называемой квазибиологической парадигмы по созданию искусственного интеллекта (ИИ). Разработка технологии создания интеллектуальных систем защиты информации (СЗИ) носит комплексный характер, в ней на первое место выносится квазибиологическая парадигма, где сначала представляется форма программирования информационных процессов, систем машинного обучения (МО) и построения нейронных систем (НС), а затем – архитектура ИИ со встроенными механизмами обеспечения ИБ.

* Исследование выполнено при финансовой поддержке Минцифры России (грант ИБ) в рамках научного проекта, соглашение № 40469-05/2022-д от 30.06.2022.

2. Разработка методологии

В основу технологий, которые будут выбираться для построения разрабатываемого ИИ, войдут действия, которые во многом смогут напоминать эмуляцию мозговой деятельности человека. Подробное описание данного концепта сведено в табл. 1.

Таблица 1. Используемые технологии для создания прототипа ИИ (эмуляция мозговой деятельности)

Наименование технологии	**Тип эмуляции мозговой деятельности**	**Характеристика, действия**
Сканирование	Предварительная обработка информации	Подготовка информационной базы, сохранение структуры и состояния ИИ
	Информационные манипуляции с кодом	Методы манипуляции над программным кодом
	Получение визуализированных данных	Возможность сканирования всей распределенной информационной системы (РИС)
		Разрешение на структуризацию приложений и программных элементов ИИ и реструктуризацию
		Возможность обнаруживать функционально значимые программные элементы в РИС
Трансляция	Обработка информации (угрозы или события)	Устранение коллизий, вызванных несовершенством сканирования
		Восполнение утерянных данных
		Определение структуры угрозы и ее назначения
		Визуализация обработанной информации
	Интерпретация визуализированных данных	Идентификация типов входных параметров
Эмуляция контрольных синапсов	Моделирование НС	Математическое моделирование
		Эффективная реализация
Моделирование	Хранение	Хранение оригинальной модели ИИ
	Пропускная способность	Эффективная межпроцессорная связь
	Микропроцессор	Мощность устройства СЗИ, достаточная для запуска модели НС
	Моделирование взаимодействия со средой	Виртуальная среда РИС, где предполагаются тесты над ИИ с НС

Разрабатываемый ИИ сможет справляться с задачами по основным направлениям (табл. 2), если будет иметься вероятность воплощения в различные механизмы кода реализации НС каждой по отдельности во все нейронные процессы. Общая сетка ИИ со всеми условиями представляет собой разветвленные точки со слабой взаимосвязью (рис. 1), где может фигурировать некая программа центрального управления системой (ЦУС), в которой нечеткие нейронные связи собираются и пересобираются в случайной и псевдослучайной последовательности благодаря основным циклам обработки и анализа информации и циклам управления [2]. Принимается событийная схема продвижения во времени. Перед началом работы на модельную временную ось наносятся моменты предполагаемого возникновения событий (приход пакетов, обработки их в узлах маршрутизатора и т.д.). В результате удается экономить вычислительные ресурсы, не обсчитывая (т.е. пропуская) те периоды времени, когда в модели ничего не происходит. В концепте учитывается система индикации ошибок, производимых случайно ИИ при наборе исходных данных, и, соответственно, их быстрого и эффективного исправления. Задачей данных вычислений являлось подтверждение имеющихся аномалий при переадресации.

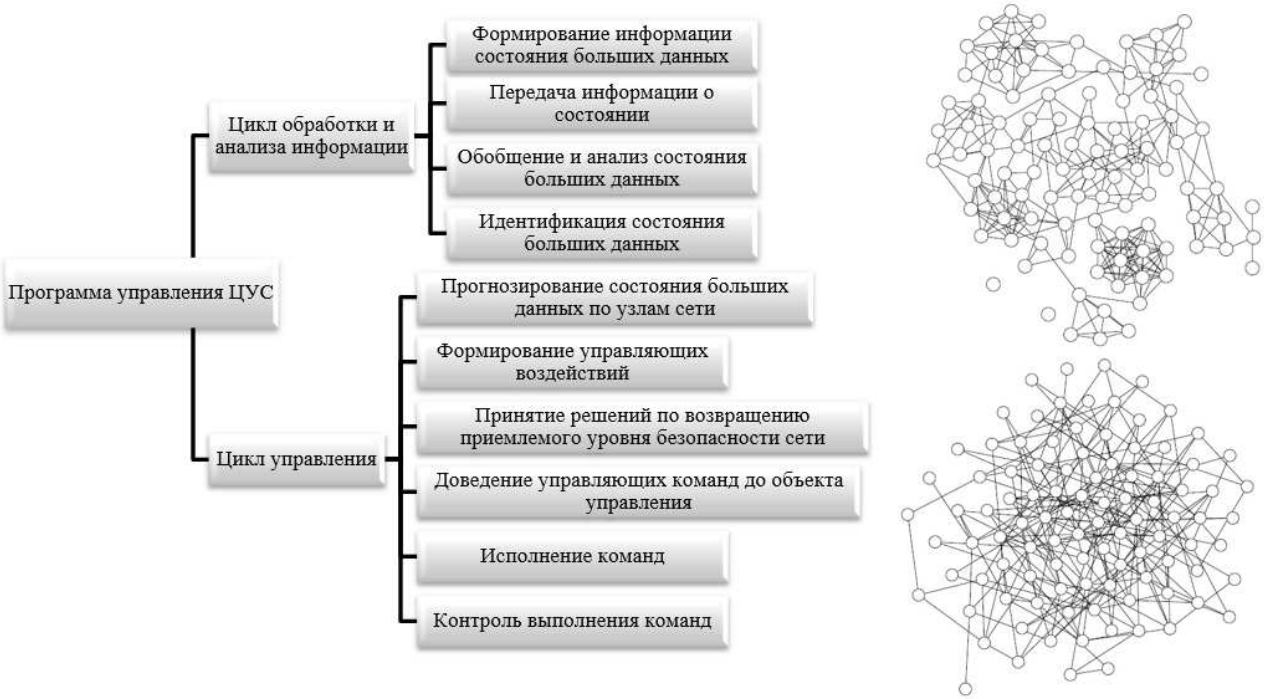

Рис. 1. Программа управления ЦУС ИИ с двумя параллельными циклами

Из рис. 1 видно, что при первом цикле обработки и анализа информации данные, поступающие для ИИ, могут быть несобранные и иерархичные, а во втором цикле управления НС будут приобретать косвенную связь между всеми элементами программного управления. Такая система должна быть сверхмощной и должна решать стратегические задачи, которые представлены в табл. 2.

Таблица 2. Сверхмощные системы со стратегически важными задачами и соответствующими компетенциями

Задачи	Компетенции	Стратегическая значимость
Усиление ИИ	Программирование ИИ, исследования в области МО	Система должна быть способна саморазвиваться и самосовершенствоваться за счет стандартных средств МО
Выработка стратегии	Стратегическое планирование, прогнозирование, расстановка приоритетов, анализ вероятного достижения отдаленных целей	Достижение собственных долгосрочных целей ИИ, преодоление противодействия со стороны внешних факторов
Манипулирование процессами РИС	Использование окружающих ресурсов РИС	Попытка «высвободиться» из поля деятельности РИС. Интеграция в информационную среду
Взлом РИС	Поиск брешей безопасности в РИС, техническая отладка РИС	Проектирование «модели выживания ИИ»
Технологические «ноу-хау»	Проектирование и моделирование стеганографических и криптографических методов защиты ИИ	Использование совершенных стеганографических и криптографических алгоритмов для защиты кода программных агентов ИИ
Экономическая эффективность	Проведение над ИИ анализа целесообразности принятых решений	Принятие решений о самоликвидации или инсталляции компонентов ИИ.

3. Разработка концептуальной модели

В концепте, где уже имеются предметы контроля «жизнеспособности» и кибербезопасности, основными достоинством будет являться установление коэффициента достижения порога насыщения, позволяющего контролировать распространение программного агента (ПА) по системе с обработкой механизмов больших данных [3, 4]. Обработка больших данных в дальнейшем будет влиять на приобретение ассоциативной память у ИИ, а также обеспечивать синхронизацию компонентов мультиагентной нейронной системы [5], имеющей в основе квазибиологическую парадигму, которая позволяет определять условия сохранности ИИ от деструктивных действий [6].

На начальных этапах формирования мультиагентной системы для нового ИИ необходимо составить схему их взаимодействия. Поскольку все ПА системы новые и проходят на данный момент этап первичного внедрения в перцептрон, то принято решение о самоназвании и символьном представлении в схеме каждого из ПА для придания аутентичности разрабатываемому программному обеспечению (ПО) (рис. 2).

На основании рис. 2 создана табл. 3 с кратким названием и описанием всех компонентов новой интеллектуальной системы. Каждый модуль – это программа и/или ПО («нейрон») со встроенными функциями СЗИ, которые связаны единым процессом перцептрона, носящего название PiRun. В дальнейшем будет выстраиваться методология и защита ПО во всей РИС, которое, в свою очередь, носит название yaVi. Цель данной концептуальной модели – обеспечение зарождающегося ИИ математической методологией, обработкой больших данных

для накопления ассоциативной памяти, проведение ассимиляции дополнительных СЗИ и компонентов перцептрона, а также развитие и улучшение строящегося ИИ.

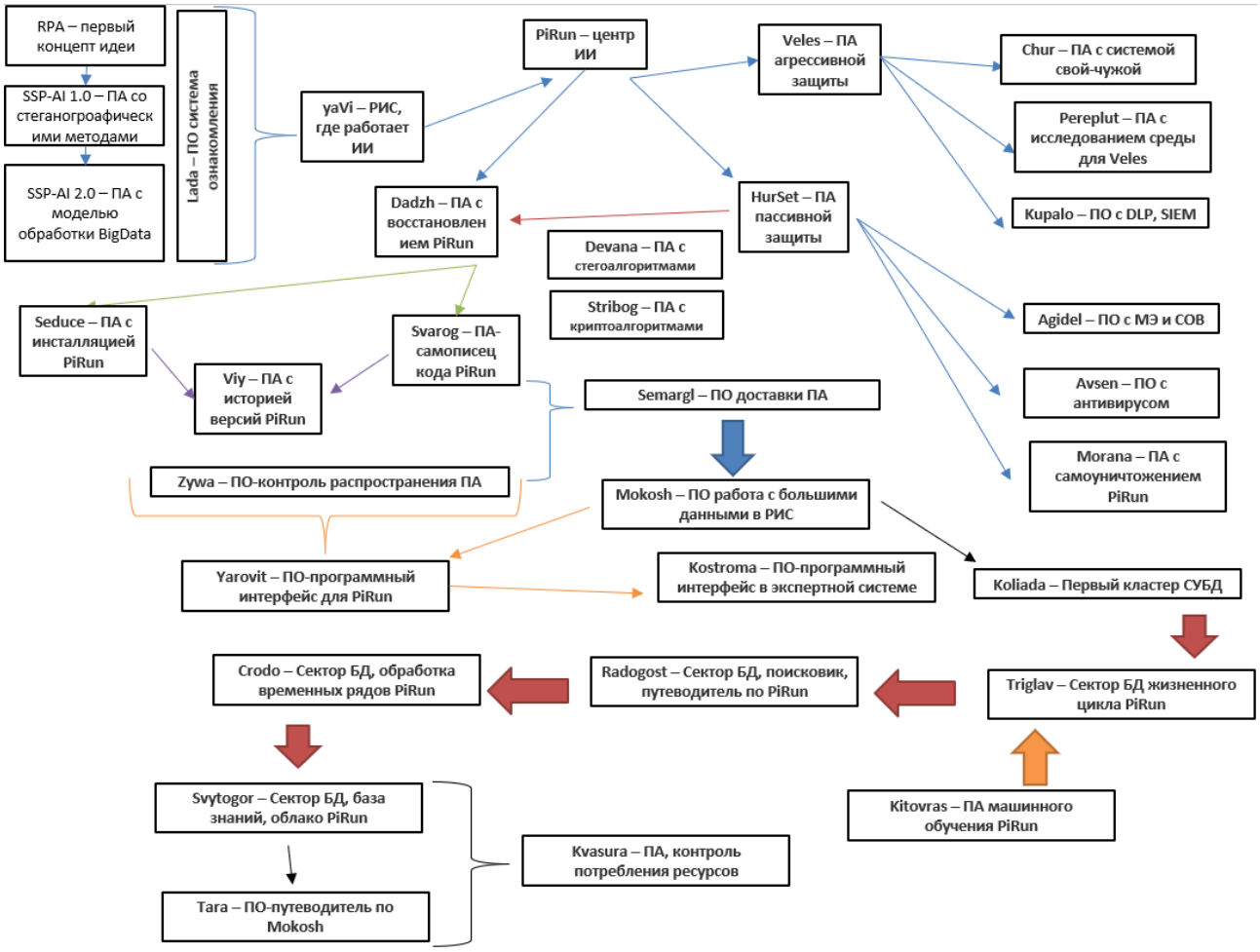

Рис. 2. Концептуальная схема взаимосвязи ПА с кратким описанием компонентов всей системы ИИ

Для формирования систем ИИ задействуются такие концепции технического развития, как искусственное сознание, интеллектуальная робототехника, машинное творчество, инженерия знаний, гибридные подходы синергийных комбинаций нейронных и символьных моделей, агентно-ориентированный подход [7].

Таблица 3. Описание компонентов концептуальной модели обеспечения защиты системы искусственного интеллекта мультиагентного типа

Название компонента	**Тип ПО**	**Функции ПО**
Veles	ПА	ПА для ведения активных защитных функций перцептрона PiRun
PiRun	ПА	Головной модуль принятия решений перцептрона
HurSet	ПА	ПА для ведения функций пассивной защиты перцептрона PiRun
Chur	ПА	ПА с определением системы «свой – чужой»
Pereplut	ПА	ПА в функции исследования среды для внедрения нейрона Veles
Kupalo	ПО	ПО для интеграции DLP- и SIEM-систем в комплекс yaVi
Agidel	ПО	ПО для интеграции СОВ и межсетевого экранирования в комплекс yaVi

Avsen	ПО	ПО для интеграции антивирусных систем в комплекс yaVi
Morana	ПА	ПА, отвечающий за оперативное самоуничтожение перцептрона PiRun
Dadzh	ПА	ПА, отвечающий за восстановление работоспособности PiRun
Devana	ПА	ПА со встроенными библиотечными данными для проведения стеганографических операций [12] перцептрона PiRun
Stribog	ПА	ПА со встроенными библиотечными данными для проведения криптографических операций [11] перцептрона PiRun
Seduce	ПА	ПА, отвечающий за самоинсталляцию компонентов перцептрона PiRun
Svarog	ПА	ПА-самописец кода компонентов перцептрона PiRun
Viy	ПА	Модуль памяти версий перцептрона PiRun
Zywa	ПА	ПА для контроля распространения компонентов перцептрона PiRun
Semargl	ПА	ПА контроля доставки компонентов перцептрона PiRun в РИС
Mokosh	СУБД	Система управления базой данных и базой знаний для ресурсов комплекса yaVi
Koliada	БД	Первый резервный программный кластер СУБД Mokosh
Yarovit	Интерфейсная оболочка	ПО с интерфейсной оболочкой для контроля функций перцептрона PiRun
Kostroma	Интерфейсная оболочка	ПО с интерфейсной оболочкой для развития экспертной системы управления комплексом yaVi
Triglav	БД	Сектор информации о жизненном цикле ИИ
Radogost	БД	Сектор БД, поисковик, путеводитель по комплексу yaVi
Crodo	БД	Сектор БД, в котором идет обработка временных рядов для перцептрона PiRun
Kitovras	ПО	Комплекс, отвечающий за машинное обучение перцептрона PiRun
Svytogor	БД	Облачное хранение систем Mokosh и PiRun
Tara	ПО	Справочная система по Mokosh
Kvasura	ПА	ПА контроля потребления ресурсов комплексом yaVi

Примечание: ПО – программное обеспечение, ПА – программный агент, СУБД – система управления базой данных, БД – база данных.

Данный концепт включает в себя зависимость от выбора сервисов и взаимодействий для защиты ИИ в РИС [8, 9]. Основываясь на требованиях для построения СЗИ, был предложен базовый граф для модулей yaVi (рис. 3).

4. Заключение

На данном этапе формирования концепта можно констатировать, что имеется схема создания оригинальной единой методологии защиты самого ИИ, основу которой составляют: методики построения самоорганизующейся карты программных агентов (нейронов), действующих в составе системы обнаружения вторжений (СОВ) и обеспечения «жизнеспособности» при последующей децентрализации НС; универсальная ассимиляционная модель обработки больших данных в РИС с использованием функции ассоциативной памяти для перцептрона; архитектура системы ИИ в виде иерархии топологий с описанием информационной структуры с адекватным отражением формальными методами специфики различных подходов к организации СЗИ посредством пакетной нейросетевой программы (ПНП).

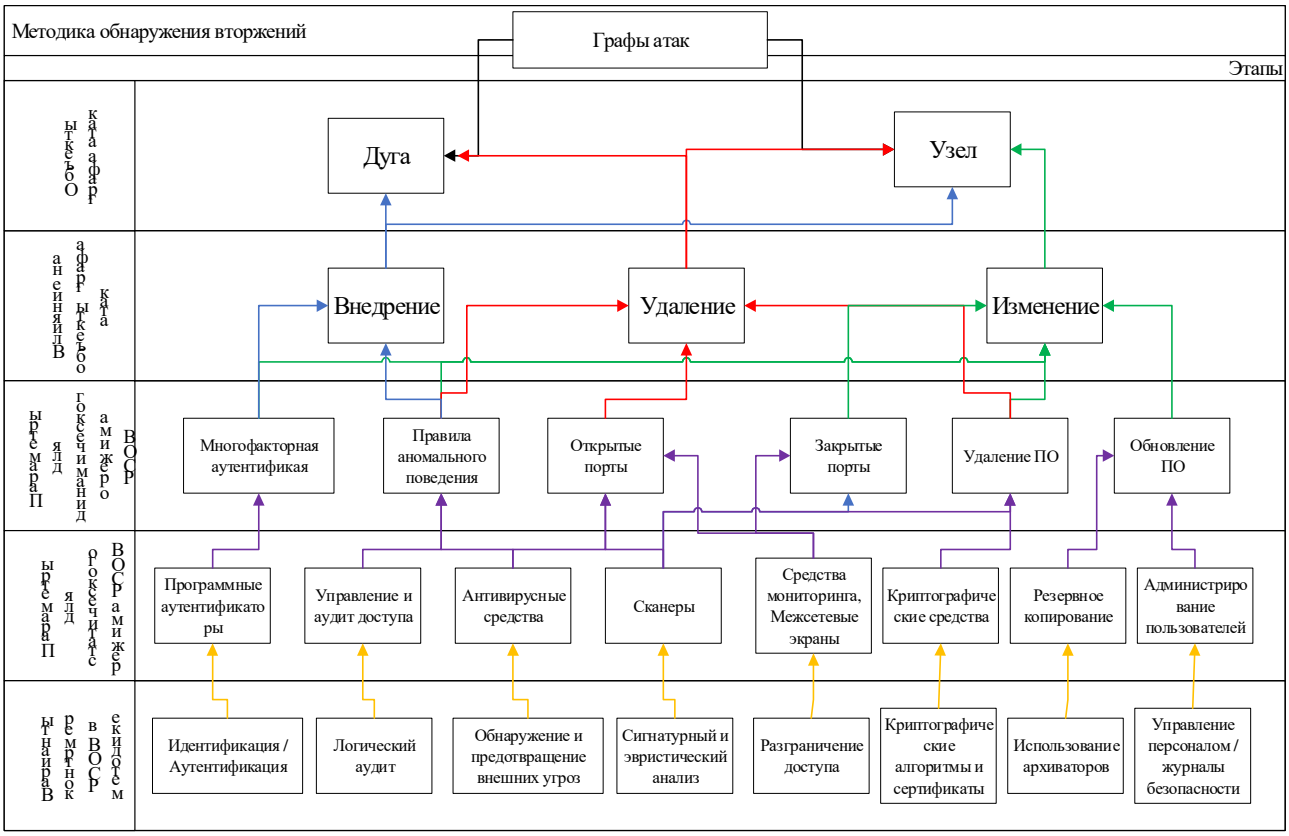

Рис. 3. Зависимости модулей ИИ в методологии обнаружения вторжений между контрмерами и объектами графа атак (РСОВ – распределенная система обнаружения вторжений)

Результаты работы позволят:

1) обеспечивать генерацию кода программного автоматизированного агента, синхронизированного с СОВ, для выявления перцептронов характеристик СЗИ;

2) устанавливать коэффициент достижения порога насыщения, позволяющий контролировать ПА;

3) прогнозировать условия киберустойчивости ИИ против деструктивных действий при помощи машинного обучения;

4) определять граничные условия функционирования ПА при условии проведения атак, направленных на нарушение целостности.

Вся концепция также дополняется возможностью использования разработанных методик и модели, технологического подхода и архитектуры для предотвращения компьютерных атак на ИИ с мультиагентной системой.

Литература

1. *Бостром Н.* Искусственный интеллект. Этапы. Угрозы. Стратегии / пер. с англ. С. Филина. М.: Манн, Иванов и Фербер, 2016. 496 с.
2. *Штеренберг С. И.* Методика построения защищенных систем искусственного интеллекта для проведения электроретинографии в офтальмологии // Офтальмохирургия. 2022. № S4. С. 51–57.
3. *Николенко С., Кадурин А., Архангельская Е.* Глубокое обучение. СПб.: Питер, 2018. 480 с.
4. *Ушаков И. А.* Обнаружение инсайдеров в корпоративной компьютерной сети на основе технологий анализа больших данных // Вестник Санкт-Петербургского государственного

университета технологии и дизайна. Серия 1: Естественные и технические науки. 2019. № 4. С. 38–43.
5. *Степанов М. Д., Павленко Е. Ю., Лаврова Д. С.* Обнаружение сетевых атак в программно-конфигурируемых сетях с использованием алгоритма изолирующего леса // Проблемы информационной безопасности. Компьютерные системы. 2021. № 1. С. 62–78.
6. *Васильева К. В., Лаврова Д. С.* Обнаружение аномалий в киберфизических системах с использованием графовых нейронных сетей // Проблемы информационной безопасности. Компьютерные системы. 2021. № 1. С. 117–130.
7. *Хапке Х., Нельсон К.* Разработка конвейеров машинного обучения. Автоматизация жизненных циклов модели с помощью TensorFlow / пер. с англ. Н. Б. Желновой. М.: ДМК Пресс, 2021. 346 с.
8. *Миняев А. А.* Метод и методика оценки эффективности системы защиты территориально-распределенных информационных систем // Информатизация и связь. 2020. № 6. С. 29–36.
9. *Гамидов Т. О., Виткова Л. А., Ковцур М. М.* Разработка моделей и алгоритмов анализа данных для исследования хода инцидентов и кризисов в социальных сетях // Вестник Санкт-Петербургского государственного университета технологии и дизайна. Серия 1: Естественные и технические науки. 2020. № 2. С. 3–10.

Штеренберг Станислав Игоревич

к.т.н., доцент кафедры информационной безопасности, Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики» (МТУСИ, 111024, Москва, ул. Авиамоторная, 8а), e-mail: stas.shterenberg.89@mail.ru, ORCID ID: 0000-0002-4216-6370.

Автор прочитал и одобрил окончательный вариант рукописи.
Автор заявляет об отсутствии конфликта интересов.

Development of a Methodology for the Protection of Artificial Intelligence Systems in Distributed Information Systems

Stanislav I. Shterenberg

Moscow Technical University of Communications and Informatics (MTUCI)

Abstract: This article is devoted to the methodology for the construction of artificial intelligent systems for information security tasks. It is planned to protect the artificial intelligence itself and the information that this device will process. The task is large-scale in its magnitude and the author of this text will try to present not just a concept, but rather an idea of how to achieve the goal of this result. By the result we will understand the final construction of a secure AI created for the organization of information security.

Keywords: Artificial intelligence, neural networks, Big data, Machine learning, quasi-biological paradigm.

For citation: Shterenberg S. I. Development of a methodology for the protection of artificial intelligence systems in distributed information systems. (in Russian). *Vestnik SibGUTI*, 2023, vol. 17, no. 3, pp. 78-86. https://doi.org/10.55648/1998-6920-2023-17-3-78-86.

The article was submitted: 26.12.2022;
accepted for publication 10.01.2023.

References

1. Bostrom N., *Iskusstvennyi intellekt. Etapy. Ugrozy. Strategii* [Artificial Intelligence. Stages. Threats. Strategies]. Moscow, Mann, Ivanov and Ferber, 2016. 496 p.
2. Shterenberg S. I. Metodika postroeniya zashchishchennykh sistem iskusstvennogo intellekta dlya provedeniya elektroretinografii v oftal'mologii [Methods of constructing protected artificial intelligence systems for conducting electroretinography in ophthalmology]. *Oftal'mokhirurgiya,* 2022, no. S4. pp. 51-57.
3. Nikolenko S., Kadurin A., Arkhangel'skaya E. *Glubokoe obuchenie* [Deep learning]. Saint Petersburg, Peter, 2018. 480 p.
4. Ushakov I. A. Obnaruzhenie insaiderovv korporativnoi komp'yuternoi seti na osnove tekhnologii analiza bol'shikh dannykh [Detection of insiders in the corporate computer network based on big data analysis technologies]. *Vestnik Sankt-Peterburgskogo gosudarstvenno-go universiteta tekhnologii i dizaina. Seriya 1: Estestvennye i tekhnicheskie nauki*, 2019, no. 4. pp. 38-43.
5. Stepanov M. D., Pavlenko E. Yu., Lavrova D. S. Obnaruzhenie setevykh atak v programmno-konfiguriruemykh setyakh s ispol'zovaniem algoritma izoliruyushchego lesa [Detection of network attacks in software-configurable networks using the isolating forest algorithm]. *Problemy informatsionnoi bezopasnosti. Komp'yuternye sistemy*, 2021, no. 1, pp. 62-78.
6. Vasil'eva K. V., Lavrova D. S. Obnaruzhenie anomalii v kiberfizicheskikh sistemakh s ispol'zovaniem grafovykh neironnykh setei [Detection of anomalies in cyberphysical systems using graph neural networks]. *Problemy informatsionnoi bezopasnosti. Komp'yuternye sistemy,* 2021, no. 1. pp. 117-130.
7. Khapke Kh., Nel'son K. *Razrabotka konveierov mashinnogo obucheniya. Avtomatizatsiya zhiz-nennykh tsiklov modeli s pomoshch'yu TensorFlow* [Development of machine learning pipelines. Automation of life cycles of the model using TensorFlow]. Moscow, DMK Press, 2021. 346 p.
8. Minyaev A.A. Metod i metodika otsenki effektivnosti sistemy zashchity territorial'no-raspredelennykh informatsionnykh sistem [Method and methodology for evaluating the effectiveness of the protection system of territorial-distributed information systems]. *Informatizatsiya i svyaz',* 2020, no. 6, pp. 29-36.
9. Gamidov T.O., Vitkova L.A., Kovtsur M.M. Razrabotka modelei i algoritmov analiza dannykh dlya issledovaniya khoda intsidentov i krizisov v sotsial'nykh setyakh [Development of models and algorithms for data analysis to study the course of incidents and crises in social networks]. *Vestnik Sankt-Peterburgskogo gosudarstvennogo universiteta tekhnologii i dizaina. Seriya 1: Estestvennye i tekhnicheskie nauki,* 2020, no. 2, pp. 3-10.

Stanislav I. Shterenberg

Cand. of Sci. (Engineering), Associate Professor of the Department of Information Security, Moscow Technical University of Communications and Informatics (MTUCI, 111024, Moscow, Aviamotornaya str., 8a), e-mail: stas.shterenberg.89@mail.ru, ORCID ID: 0000-0002-4216-6370.