ОТЗЫВ

официального оппонента, профессора кафедры
систем информационной безопасности
ФГБОУ ВО «Казанский национальный исследовательский
технический университет им. А.Н. Туполева-КАИ»,
доктора технических наук, профессора
Катасёва Алексея Сергеевича
на диссертационную работу

Сулавко Алексея Евгеньевича

**«Высоконадежная биометрическая аутентификация
на основе защищенного исполнения нейросетевых моделей
и алгоритмов искусственного интеллекта»,**

представленную на соискание ученой степени доктора технических наук
по специальности 2.3.6 – Методы и системы защиты информации,
информационная безопасность

Актуальность темы исследования

В настоящее время человечество вступает в эпоху «технологической сингулярности». Об этом свидетельствуют стремительно ускоряющиеся темпы развития информационных технологий, искусственного интеллекта, средств вычислительной техники и глобальной информатизации общества. Все быстрее растет число веб-сервисов для массовых потребителей. При этом обеспечить безопасность виртуального образа пользователей становится все сложнее, так как при увеличении числа их личных кабинетов возрастает число паролей и криптографических ключей, которые нужно хранить, что, в частности, обостряет проблему обеспечения информационной безопасности. При этом пользователи нуждаются не только в обеспечении высоконадежной аутентификации, но и в защите самих аутентификационных данных от компрометации с учетом наличия «человеческого фактора». Использование современных технологий машинного обучения и глубоких нейронных сетей часто позволяет злоумышленнику обходить даже самые сложные и эффективные алгоритмы сопоставления биометрических данных, генерируя, например, правдоподобное изображение лица или голос человека, личность которого является целью компьютерной атаки.

Диссертационная работа Сулавко А.Е. направлена на решение проблемы повышения надежности многофакторной биометрической аутентификации и защищенности биометрических систем от компьютерных атак на основе технологии защищенного исполнения нейросетевых моделей и алгоритмов искусственного интеллекта. В работе вводится новое понятие защищенного режима для выполнения алгоритмов искусственного интеллекта.

Система искусственного интеллекта, функционирующая в защищенном режиме, является более стабильной при изменяющихся условиях, а также проявляет высокую устойчивость к атакам и воздействиям со стороны нарушителя. Соискатель распространяет данный термин на искусственный интеллект в целом, однако основные результаты, полученные в рамках диссертационной работы, относятся, прежде всего, к биометрическим системам. Защищенный режим выполнения процедуры биометрической аутентификации позволяет повысить ее надежность, обеспечить безопасность биометрической системы, усилить ее сопротивляемость к атакам, защитить от раскрытия конфиденциальности биометрических шаблонов. Кроме того, данный режим дает возможность связать биометрический образ человека с личным паролем или криптографическим ключом, используемым для доступа к данным или информационным системам, избавляя пользователя от необходимости запоминать или записывать пароли. Таким образом, тема диссертации Сулавко А.Е. является актуальной.

Оценка структуры и содержания работы

Диссертация состоит из введения, пяти глав, заключения, списка сокращений, списка литературы и десяти приложений. Основной текст диссертации изложен на 264 страницах, содержит 108 рисунков и 28 таблиц. В приложениях на 78 страницах содержатся вспомогательные материалы, таблицы, результаты экспериментов, исходные коды программ, акты внедрения, патент, а также свидетельства о регистрации программ для ЭВМ и электронных ресурсов. Список литературы включает в себя 362 наименования. Структура и содержание диссертации соответствует цели и поставленным задачам исследования.

Во введении обоснована актуальность темы исследования, дается общая характеристика основных научных положений и результатов работы.

В первой главе выполнен анализ результатов в области биометрической аутентификации и защиты систем и знаний искусственного интеллекта от компьютерных атак. Приведены разновидности угроз и атак, перед которыми уязвимы биометрические системы. Описана концепция защищенного исполнения нейросетевых алгоритмов искусственного интеллекта. Проведен анализ научных публикаций и стандартов (как международных, так и национальных), посвященных тематике защиты искусственного интеллекта от компьютерных атак. Представлены подходы к реализации концепции в задачах биометрической аутентификации. В ходе анализа установлено, что для задач классификации в защищенном режиме архитектуру искусственного интеллекта можно разделить на два блока: блок извлечения признаков и блок классификации образов. Сформулированы требования к этим блокам. В конце главы сформулирована цель и решаемые в работе задачи.

Вторая глава посвящена разработке концепции защищенного исполнения нейросетевых алгоритмов искусственного интеллекта и модели корреляционных нейронов для высоконадежной биометрической аутентификации. Корреляционные нейроны применяются для формирования блока классификации, устойчивого к деструктивным воздействиям. Предложена модель нейросетевого преобразователя «биометрия-код» на основе корреляционных нейронов и алгоритм ее обучения на малых выборках биометрических данных. Модель позволяет защитить биометрический эталон пользователя от компрометации. Предложено отображение для преобразования биометрических признаков в более информативные мета-признаки Байеса-Минковского. По результатам вычислительных экспериментов показана эффективность предложенных моделей нейронов и нейросетевого преобразователя, а также разработанной концепции.

В третьей главе описывается влияние психофизиологического состояния человека на его динамические биометрические признаки. Данный факт отражается в значительном увеличении вероятности ошибок «ложного отказа» (FRR) и «ложного допуска» (FAR), если на этапах обучения биометрической системы и аутентификации пользователь находился в различных состояниях (происходит дрейф данных). В ходе анализа подходов к построению адаптивных моделей искусственного интеллекта, предложена собственная адаптивная модель, представляющая собой искусственную иммунную сеть, настраивающуюся на верификацию образа конкретного пользователя и обучающуюся на малых выборках образов «Свой» (данные пользователя) и «Чужие» (данные, не принадлежащие пользователю). Приведены алгоритмы формирования врожденного и приобретенного иммунитета искусственной иммунной сети, позволяющие производить ее обучение. Исследованы характеристики предложенной адаптивной модели искусственного интеллекта и получены вероятности ошибок, рассчитанные для разных наборов данных клавиатурного почерка. Установлено, что разработанная модель соответствует основным принципам построения искусственной иммунной сети (модель обладает эмерджентностью, памятью, двойной пластичностью, устойчивостью обучения и другими свойствами).

Четвертая глава посвящена разработке методов и алгоритмов высоконадежной многофакторной аутентификации и объединению независимых биометрических образов с обеспечением защиты от их компрометации. Предложен вариант комплексирования биометрических образов в виде алгоритма трехфакторной высоконадежной аутентификации с последовательным предъявлением биометрических данных. В качестве первого фактора используется эхограмма ушного канала. Второй и третий факторы представляют собой голосовой и рукописный пароли (либо рукописная подпись и голосовой пароль). Для анализа данных применяется адаптивная нейро-иммунная модель искусственного интеллекта. Для защиты знаний адаптивной модели ее параметры после обучения

шифруются на ключе, формируемом моделью нейросетевого преобразователя «биометрия-код», основанном на корреляционных нейронах. Для корректировки ошибочных бит ключа, возникающего на выходе модели нейросетевого преобразователя, предложено использовать коды исправления ошибок, что позволяет балансировать показатели FRR и FAR. Достаточно подробно рассмотрен каждый из предложенных к использованию в алгоритме типов биометрических образов. Для каждого типа образов описаны эксперименты по выделению признаков и классификации. Приведены описания экспериментов по аутентификации и их результаты, подробно описаны наборы данных. Отдельным экспериментом выполнена оценка надежности двухфакторной аутентификации по голосовым и рукописным образам. Экспериментальная оценка надежности двухфакторной аутентификации показала следующие результаты: FRR = 0,03 при $FAR < 10^{-10}$. Доказано, что использование трех факторов аутентификации многократно повышает надежность биометрической системы. Проведенные эксперименты дают наглядное представление о повышении устойчивости биометрической системы к атакам подбора с использованием специально подготовленных наборов данных. Экспериментальная оценка надежности трехфакторной аутентификации показала следующие результаты: FRR = 0,12 при $FAR < 10^{-14}$.

В пятой главе описывается разработанная технология автоматического синтеза и обучения нейросетевых моделей доверенного искусственного интеллекта. Указываются ключевые положения проекта разработанного стандарта, основанного на данной технологии, который в значительной степени расширяет уже имеющийся стандарт ГОСТ Р 52633.5-2011. В отличие от базового, новый стандарт может использоваться не только в приложениях биометрии, но и в любых задачах классификации образов. Отмечено, что использование стандарта направлено на повышение уровня защищенности знаний и систем искусственного интеллекта от несанкционированных манипуляций (чтение, интерпретация знаний, анализ логики принятия решений, изменение логики решений искусственного интеллекта, состязательные атаки). Применение стандарта позволяет снизить вероятность успеха состязательных атак, реализуемых злоумышленником, путем наложения шумов на исходный образ «Свой» или создания синтетических примеров образов «Чужой». Приведено описание восьми внедрений результатов работы на предприятиях страны (география внедрений включает не только Омск – родной город соискателя, но и такие города как Москва, Санкт-Петербург, Самара), в том числе, в ИТ-компаниях, двух ВУЗах и государственной поликлинике. Отдельное внимание уделено применению предложенной технологии при разработке линейки программных продуктов AIConstructor.

В заключении сформулированы основные теоретические и практические результаты и итоги работы, приведены выводы в отношении полученных результатов, намечены направления перспективных исследований.

Приложения содержат дополнительные материалы в виде таблиц, актов внедрения и использования результатов диссертационной работы. Отдельно стоит отметить, что в приложениях представлены документы, подтверждающие активное участие соискателя в деятельности технического комитета по стандартизации №164 «Искусственный интеллект», в том числе, в качестве разработчика государственного национального стандарта.

Материалы диссертации отражают содержание и объем проделанной работы. Автореферат диссертации полностью соответствует ее содержанию.

Степень обоснованности научных положений, выводов и рекомендаций, сформулированных в диссертации

При решении поставленных в диссертационной работе задач корректно и обоснованно использовались методы системного анализа, математического и имитационного моделирования, теории вероятностей и математической статистики, спектрального и корреляционного анализа, кодирования и защиты информации, распознавания образов, машинного, в том числе, глубокого обучения, ансамблевые методы, а также аппарат искусственных нейронных сетей.

Основные научные положения, выводы и рекомендации, сформулированные в диссертации, представлены в 80 работах автора, в том числе в 1 монографии, 38 статьях в рецензируемых журналах из перечня ВАК либо в изданиях, индексируемых в RSCI, 21 статье в изданиях, включенных в международные базы Web of Science и Scopus, а также 11 статьях в материалах международных и российских конференций. Кроме того, получен 1 патент и 8 свидетельств о регистрации программ для ЭВМ, что удовлетворяет предъявляемым требованиям.

Содержание диссертации соответствует пунктам 9, 12 и 15 паспорта специальности 2.3.6 – Методы и системы защиты информации, информационная безопасность.

Основные выводы и результаты с оценкой их научной новизны

1. Предложена концепция защищенного исполнения нейросетевых алгоритмов искусственного интеллекта, позволяющая обеспечить устойчивость моделей и алгоритмов искусственного интеллекта к извлечению знаний в задачах классификации образов, отличающаяся преобразованием корреляционных связей между признаками в мета-признаки. Экспериментально установлено, что корреляция между признаками увеличивает количество информации о распознаваемых образах, а один мета-признак содержит в 2-3 раза больше информации, чем содержится суммарно в паре исходных признаков, от которых он порожден, что повышает надежность классификации биометрических данных.

2. Разработаны модель корреляционных нейронов и модель нейросетевого преобразователя «биометрия-код» на их основе, позволяющая связать криптографический ключ или пароль пользователя с его биометрическим образом. Новая модель нейросетевого преобразователя «биометрия-код» отличается тем, что выполняет анализ корреляционных связей между признаками вместо анализа значений самих признаков. Разработан робастной алгоритм автоматического синтеза и обучения этих моделей на малых выборках, что позволяет повысить защищенность биометрических данных от компрометации, длину ключа, связываемого с биометрическими образами субъектов, и устойчивость систем биометрической аутентификации к состязательным атакам.

3. Разработана адаптивная нейро-иммунная модель искусственного интеллекта, основанная на иммунологическом подходе, для распознавания биометрических образов, отличающаяся использованием новой гибкой архитектуры иммунных детекторов, а также применением методов ансамблирования и обучения с подкреплением. В основе детекторов применены ядерные функции. Предложенные нейро-иммунная модель и алгоритмы ее обучения, в отличие от существующих, позволяют снизить влияние концептуального дрейфа и вероятность ошибок биометрической аутентификации, даже если исходная обучающая выборка недостаточно репрезентативна или незначительна в объеме.

4. Разработаны методы и алгоритм высоконадежной многофакторной биометрической аутентификации, отличающиеся использованием новых биометрических параметров, характеризующих внутреннее строение ушного канала, комплексированием признаков различной природы (ухо, голос, рукописный почерк) с учетом их приоритезации, информативности и стабильности, а также совместным использованием нейросетевого преобразователя «биометрия-код» на базе корреляционных нейронов и нейро-иммунной модели. Предложенные методы и алгоритм, в отличие от аналогов, позволяют обеспечить более высокую надежность аутентификации, робастность дрейфующих характеристик, защиту биометрических образов от компрометации, а также снизить показатели FRR и FAR.

5. Разработана технология синтеза нейросетевых моделей доверенного искусственного интеллекта, позволяющая снизить объем тренировочной выборки, повысить надежность и защищенность биометрических систем аутентификации и других приложений искусственного интеллекта. Технология обладает возможностью автоматического обучения нейросетевых моделей искусственного интеллекта с учетом информативности признаков и режимом защищенного исполнения нейросетевых алгоритмов классификации образов. Разработан программный комплекс на основе предложенной технологии, позволяющий создавать системы высоконадежной биометрической аутентификации и другие приложения доверенного искусственного интеллекта.

Достоверность полученных результатов

Достоверность результатов работы подтверждается корректным использованием положений стандартов, методов исследования и грамотной постановкой методик проведения экспериментов. В работе использовались признанные методики статистической обработки данных и методы численного и компьютерного моделирования, все расчеты выполнялись математически строго, что свидетельствует о достоверности результатов. Точность и непротиворечивость результатов подтверждены актами внедрения. Вводимые допущения мотивировались фактами, известными из практики. Предложенные в работе концепция, модели, методы и алгоритмы теоретически обоснованы и не противоречат известным результатам исследований других авторов.

Теоретическая и практическая значимость полученных автором результатов

Теоретическая значимость работы является высокой. Фактически соискатель разработал математический аппарат, позволяющий создавать нейросетевые модели искусственного интеллекта, обладающие повышенной защищенностью от деструктивных воздействий. Предложенный аппарат может применяться в различных приложениях искусственного интеллекта, однако в рамках диссертации рассматривается только задача высоконадежной биометрической аутентификации. Важным положением работы является использование корреляции между признаками в качестве отдельных признаков. Эта идея и авторский подход являются глубоко проработанными, на них опираются многие результаты работы.

Теоретической значимостью обладают следующие результаты:

- концепция защищенного исполнения нейросетевых алгоритмов искусственного интеллекта, основанная на преобразовании корреляционных связей между признаками в мета-признаки;
- модель корреляционных нейронов;
- модель нейросетевого преобразователя «биометрия-код» на основе корреляционных нейронов;
- алгоритм автоматического синтеза и нейросетевого преобразователя «биометрия-код» на базе корреляционных нейронов на малых выборках;
- адаптивная нейро-иммунная модель искусственного интеллекта;
- алгоритмы обучения адаптивной нейро-иммунной модели искусственного интеллекта с учителем и с подкреплением;
- методы и алгоритм высоконадежной многофакторной аутентификации на основе рукописных и голосовых образов, а также акустических образов уха с обеспечением защиты биометрических данных от компрометации;
- технология автоматического синтеза и обучения нейросетевых моделей доверенного искусственного интеллекта на малых выборках.

Практическая значимость работы заключается в том, что на базе предложенной технологии синтеза и обучения нейросетевых моделей искусственного интеллекта под руководством соискателя разработан проект государственного стандарта ГОСТ Р «Искусственный интеллект. Нейросетевые алгоритмы в защищенном исполнении. Автоматическое обучение нейросетевых моделей на малых выборках в задачах классификации». Это первый стандарт, регламентирующий особенности создания и автоматического обучения нейросетевых моделей доверенного искусственного интеллекта, выполнение которых возможно в защищенном режиме. Стандарт прошел экспертизу технических комитетов Росстандарта и включен в программу стандартизации технического комитета «Искусственный интеллект».

Практически значимыми являются также результаты, которые легли в основу линейки программных продуктов AIConstructor. Под руководством соискателя разработан программный комплекс для проведения научных исследований по машинному обучению и корпоративная среда управления жизненным циклом искусственного интеллекта. Также результаты работы внедрены на семи предприятиях, использованы в проектно-конструкторской деятельности и учебном процессе. Отдельно стоит отметить методы и алгоритм трехфакторной биометрической аутентификации по особенностям ушного канала, рукописным и голосовым образам с показателями FRR=0,12 при $FAR<10^{-14}$, которые сами по себе представляют значимость для практики.

Замечания

1. В диссертации дважды сформулирована цель: сначала во введении, а затем в выводах по первой главе. При этом формулировка «целей» несколько отличается. В первом случае целью является «повысить надежность многофакторной биометрической аутентификации на основе **защищенного исполнения нейросетевых моделей доверенного искусственного интеллекта и алгоритмов их автоматического синтеза и обучения на малых выборках биометрических данных**». Во втором случае – «повысить надежность многофакторной биометрической аутентификации на основе **технологии автоматического синтеза и обучения нейросетевых моделей доверенного искусственного интеллекта**». Повторная формулировка цели является избыточной. При этом в заключении отсутствует привычная фраза о достижении цели диссертационного исследования: «все поставленные задачи решены, а цель работы достигнута».

2. Все решаемые задачи, сформулированные для достижения поставленной цели, связаны исключительно с разработкой (разработка концепции, разработка моделей, разработка методов, разработка технологии). При этом нет ни одной задачи, связанной с анализом, исследованием, реализацией, тестированием, апробацией, хотя такие задачи в диссертации фактически решаются.

3. В работе желательно было представить, хотя бы схематично, общую методологию решения поставленной проблемы повышения надежности биометрической аутентификации и защищенности биометрических систем. Схема этой методологии давала бы наглядное комплексное представление о связи используемых в работе принципов, разработанных методов и алгоритмов, а также их реализации и апробации. Без такой схемы предложенные в работе методы, алгоритмы и модели выглядят несколько разрозненно.

4. Во втором приложении шесть из восьми актов являются не «актами внедрения результатов работы», а актами об использовании. Внедрение выполнялось только на уровне учебного процесса двух университетов. В остальных случаях апробация выполнена на уровне использования результатов диссертационного исследования (использование при разработке стандарта, использование в научно-исследовательской, проектно-конструкторской и проектной деятельности, использование в рамках инструментально-лабораторных обследований и тестирования пациентов). Следовательно, заголовок указанного приложения должен иметь название «Акты об использовании и внедрении...».

5. Пятая глава диссертации, в частности, посвящена вопросам апробации. В ней представлены результаты внедрения в учебный процесс, а также использования результатов диссертационного исследования в медицине и при разработке биометрических систем аутентификации и мониторинга пользователей в различных организациях. Однако описание апробации представлено излишне лаконично, что не позволяет в полной мере судить о достигнутых результатах.

6. Некоторые пункты оглавления имеют нумерацию, не соответствующую нумерации в тексте диссертации. Например, «заключение» указано на 261 странице, а фактически начинается на 262 странице.

7. Некоторые рисунки в диссертации плохо читабельные, что затрудняет их понимание и интерпретацию. Так, на рисунке 5.8 (стр. 239) во многих блоках диаграммы практически невозможно разобрать текст. При подготовке диссертации автору следовало бы обратить на это внимание.

8. В диссертации встречаются пропуски и опечатки. Например, в первом предложении заключения «В диссертационной работе на основе выполненного автором исследования решена актуальная научная повышения...» пропущено слово «проблема».

9. В работе имеются терминологические неточности. Так, например, на рисунке 6 автореферата представлены не алгоритмы обучения, а их блок-схемы. В данном случае подрисуночная подпись должна начинаться со слов «Блок-схемы алгоритмов обучения».

Заключение

В целом указанные замечания не снижают высокой научной ценности и практической значимости выполненного исследования. Диссертация Сулавко А.Е., представленная на соискание ученой степени доктора технических наук, обладает научной новизной, теоретической и практической значимостью, является завершенной научно-квалификационной работой, в которой на основании проведенных автором исследований решена актуальная научная проблема повышения надежности многофакторной биометрической аутентификации и защищенности биометрических систем от компьютерных атак. Внедрение и использование полученных в диссертации результатов вносит значительный вклад в развитие методов и средств высоконадежной биометрической аутентификации, а также защиты знаний и систем доверенного искусственного интеллекта.

На основании вышеизложенного считаю, что диссертация Сулавко А.Е. на тему «Высоконадежная биометрическая аутентификация на основе защищенного исполнения нейросетевых моделей и алгоритмов искусственного интеллекта» соответствует требованиям п.9 «Положения о присуждении ученых степеней», предъявляемым к диссертациям на соискание ученой степени доктора технических наук, а ее автор Сулавко Алексей Евгеньевич заслуживает присуждения ученой степени доктора технических наук по специальности 2.3.6 – Методы и системы защиты информации, информационная безопасность.

Официальный оппонент:
профессор кафедры систем информационной
безопасности ФГБОУ ВО «Казанский национальный
исследовательский технический университет
им. А.Н. Туполева-КАИ»,
доктор технических наук, профессор / Катасёв Алексей Сергеевич /

«17» августа 2023 г.

Докторская диссертация защищена по специальности 05.13.18 – Математическое моделирование, численные методы и комплексы программ

Даю согласие на обработку персональных данных.

ФГБОУ ВО «Казанский национальный исследовательский технический университет им. А.Н. Туполева-КАИ»
420111, г. Казань, ул. К. Маркса, д. 10
Телефон: +7 927 408-94-68
E-mail: ASKatasev@kai.ru

Подпись
заверяю. Начальник управления
делопроизводства и контроля