НОРМАТИВНО-ТЕХНИЧЕСКИЕ ВОПРОСЫ РАЗРАБОТКИ БЕЗОПАСНЫХ АВТОМАТИЗИРОВАННЫХ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ

Гаврилов Д.А.[1]

Цель работы: *рассмотрение нормативно-технических вопросов эффективного создания, функционирования и эксплуатации безопасных, надёжных и эффективных систем, основанных на искусственном интеллекте.*

Метод исследования: *Рассмотрены возможности применения концептуально-логического моделирования эргасистем и инвариантных архитектур рационального моделирования на основе проблемно-ориентированного варианта комплексного «информационно-кибернетически-дидактического» подхода с помощью информационно-математической структуры автоматизированной оптико-электронной системы наземно-космического мониторинга.*

Результаты: *Представлена концептуально-логическая модель системы нормативно-технического регулирования систем, основанных на технологиях искусственного интеллекта, и инвариантная архитектура рациональной модели системы искусственного интеллекта, разработана методика решения задачи функционирования автоматизированной оптико-электронной системы наземно-космического мониторинга.*

Ключевые слова: *технологии искусственного интеллекта, нормативно-техническое регулирование искусственного интеллекта, двухуровневое эргасистема, методологические принципы, концептуально-логическая модель.*

DOI: 10.21681/2311-3456-2020-06-63-71

Введение

В первой части статьи представлены [1] основные подходы к построению эффективных автоматизированных оптико–электронных систем наземно–космического мониторинга (АОЭС НКМ), обеспечивающей защищенную переработку визуальной информации в условиях информационного соперничества. Настоящая работа посвящена рассмотрению нормативно-технических вопросов эффективного создания, функционирования и эксплуатации безопасных, надёжных и эффективных систем, основанных на искусственном интеллекте.

В настоящее время искусственный интеллект развивается в направлении решения практических задач, позволяющих приблизить его возможности к возможностям человека [2]. Наибольшее внимание привлекают автоматизированные информационные системы поддержки и принятия решений в реальном времени или близком к реальному, системы динамического планирования, средства хранения, извлечения, анализа и моделирования знаний. Специалист, использующий автоматизированную информационную систему для решения текущих задач, может достигать по результатам возможностей экспертов в данной области знаний, что позволяет резко повысить квалификацию рядовых специалистов за счет аккумуляции знаний в системе, в том числе знаний экспертов высшей квалификации. Автоматическая обработка визуальной информации является одним из наиболее важных направлений в области развития искусственного интеллекта [3]. В современном мире непрерывно возникают разнообразные задачи анализа множества данных. Широкое разнообразие областей применения требует повышения качества обработки изображений, что, в свою очередь, ведет к необходимости разработки новых технологий, методов и алгоритмов обработки [4]. Задачи обработки информации все чаще требуют решения в режиме реального времени, поэтому все более востребованными становятся системы, использующие машинное зрение в качестве основного источника информации. Техническое зрение, сфера применения которого непрерывно расширяется, приобретает в настоящее время большое значение во многих областях деятельности человека и считается одной из самых перспективных и востребованных цифровых компьютерных технологий. Изображение является формой наиболее полного представления информации, которую, как правило, невозможно ничем заменить [5]. Компьютерное зрение может рассматриваться как составная часть технологий в области искусственного интеллекта. В свою очередь, распознавание образов является одной из важнейших задач искусственного интеллекта, целью которого является копирование и имитация интеллектуальной де-

1 Гаврилов Дмитрий Александрович, кандидат технических наук, научный сотрудник отдела научно-технической подготовки производства, АО «Институт точной механики и вычислительной техники им. С.А. Лебедева РАН», заведующий лабораторией цифровых систем специального назначения, Московский физико-технический институт (национальный исследовательский университет), г. Москва, Россия.
E-mail: gavrilov.da@mipt.ru

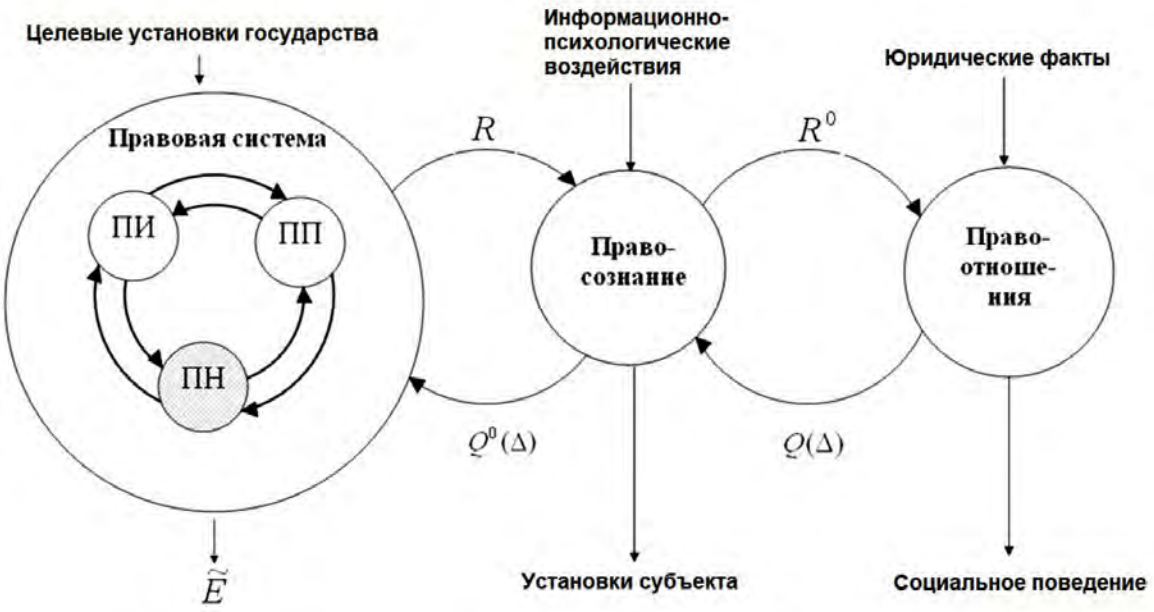

Рис. 1. Концептуально-логическая модель системы правового регулирования

ятельности человека. В отличие от человеческого мозга работа технологий искусственного интеллекта предполагает математический выбор и распознавание шаблонов из массивов обучающих данных.

Основными научно-техническими направлениями развития технологий искусственного интеллекта на сегодняшний день являются следующие: технологии виртуальной реальности [6], робототехнические системы и экспертно-аналитические системы, а также технологии интеллектуального поиска, анализа и синтеза различных видов информации.

Нейросетевые технологии [7] находят применение в системах анализа и предсказания событий. Способности алгоритмов по прогнозированию поведения значительно превосходят возможности человека. На достаточно высоком уровне определяется потребность в экспертных автоматизированных системах. В экспертных системах база знаний представляет собой формализованные эмпирические знания высококвалифицированных специалистов в какой-либо узкой предметной области. Автоматизированные системы предназначены для решения задач в диалоговом режиме со специалистами (конечными пользователями), от которых не требуется знания программирования.

Можно выделить следующие основные классы задач, решаемых автоматизированными информационными системами [8]: диагностика, прогнозирование, идентификация [9], управление, проектирование (конфигурирование), мониторинг.

Наиболее широко встречающиеся области деятельности, требующие использования экспертных систем: медицина, вычислительная техника, военное дело, микроэлектроника, радиоэлектроника, юриспруденция, экономика, экология, управление технологическими процессами, геология (поиск полезных ископаемых).

Концептуально-логическая модель нормативно-технического регулирования систем, основанных на технологиях искусственного интеллекта

Концептуально-логическая модель нормативно-технического регулирования систем, основанных на технологиях искусственного интеллекта может быть представлена в виде множества взаимосвязанных на разных уровнях комплексов и компонентов, обладающих свойствами целостности. По целевому назначению система нормативно-технического регулирования искусственного интеллекта является сложной открытой неравновесной информационно-кибернетической системой, обеспечивающей нормативное регулирование и характеризующейся высокой степенью динамичности, неустойчивости и неопределенности. Система правового регулирования базируется на следующих основных подсистемах: правовая система (ПИ), правосознание, правоотношения, и включает двухуровневый внутренний информационно-кибернетический контур правового нормативного и индивидуального регулирования и три внешних информационно кибернетических контура [10] (Рис. 1).

При этом информационно-кибернетическая цепочка внутреннего регулирования выглядит следующим образом: «правовая система – правовые предписания *R* (предписывающая правовая информация) – правосознание – осознанные (включая индивидуальные) правовые предписания R^0 (правоприменительная правовая информация) – правоотношения – осведомляющая (статистическая и др.) правовая информация $Q(\Delta)$ о качестве Δ соблюдения правовых норм и принципов (правовая реализация) – правосознание – логически обработанная осведомляющая правовая информация $Q^0(\Delta)$ – правовая система» [10]. Основными внешними входными воздействиями и соответствующими внутренними откликами или результатами в данной подси-

стеме будут следующие: целевые установки государства и интегральная оценка эффективности правового регулирования, информационно-психологические воздействия, юридические факты и правовое поведение.

Исследование подобных многосвязных систем в настоящее время осуществляется на основе концептуально-логического моделирования с применением инвариантных архитектур рационального моделирования. В качестве модели архитектуры системы на основе искусственного интеллекта может быть использована инвариантная функциональная структура эргасистемы [11], представленная виде комплекса функциональных подсистем: измерения (P_1), наблюдения (P_2), идентификации (P_3), принятия решений (P_4), координации (P_5), информационного обмена (P_6) и информационной защиты (P_7), необходимой при функционировании в условиях информационного соперничества и обеспечивающих необходимую защищенность переработки информации. На объект управления (P_0) в момент времени t поступают различные входные воздействия: функциональные $R(t)$, внешние целевые $X(t)$ и внешние координирующие $X^0(t)$, на которые формируются соответствующие отклики [10].

В общем случае техническая система, основанная на искусственном интеллекте, представляет собой автоматизированную систему, обеспечивающую распределение информационных и управляющих функций между оператором и ЭВМ, а также анализ и выдачу измерительной информации в удобном для оператора виде. Таким образом, основным результатом работы системы является подготовка информации с высокой степенью достоверности для эффективного принятия дальнейших решений оператором, при этом полученная информация служит рекомендацией, а основная роль в принятии решений принадлежит человеку. В экстренных случаях возможен переход системы в режим автоматизации, в котором обеспечивается выполнение полного цикла операций, включая принятие решений по дальнейшему использованию полученной аналитической информации, без участия оператора.

Построение эффективной автоматизированной системы мониторинга представляется возможным на основе применения проблемно-ориентированного варианта комплексного «ИКД»-подхода (рис. 2).

Комплексный системный «ИКД» подход с акцентированием внимания на его *информационном*, *кибернетическом* и *дидактическом* аспектах, состоящего в интеграции методологии информационного подхода (при котором объект рассматривается как целенаправленная информационная система), методологии кибернетического подхода (при котором объект рассматривается как система управления на уровне информационных процессов и алгоритмов функционирования информационной базы) с методологией дидактического подхода (при котором объект рассматривается как система, способная к самообучению) в составе методологии системного подхода (при котором объект рассматривается как сложноорганизованная многоуровневая и многоаспектная система). Основные подходы к построению эффективной автоматизированной оптико-электронной системы наземно-космического мониторинга, обеспечивающей защищенную переработку

Рис. 2. Проблемно-ориентированный вариант комплексного «ИКД»-подхода

DOI:10.21681/2311-3456-2020-06-63-71

визуальной информации в условиях информационного соперничества представлены в [1].

Основными организационно-техническими требованиями (ОТТ) информационного математического обеспечения (ИМО) АОЭС НКМ, влияющими на эффективность ее функционирования, являются *точность*, характеризующая качество дешифрирования визуальной информации, и *оперативность*, характеризующая обеспечение необходимого быстродействия своевременно в соответствии с поставленными целями и задачами.

Дополнительными ОТТ к ИМО АОЭС НКМ являются *имитостойкость*, характеризующая способность не допускать навязывания дезинформации в условиях информационного противоборства, *устойчивость*, характеризующая способность сохранять состояние равновесия в условиях дестабилизирующих воздействий, *живучесть*, характеризующая способность выполнять установленный минимальный объем функций при подавляющих внешних воздействиях, *добротность,* характеризующая возможность функционирования в условиях отказов [12].

Показатели информационно-целевой эффективности [11] могут быть охарактеризованы следующим образом:

Информационная точность:

$$J_{1ц} = \frac{Э_ц}{\tau},$$

где

$$Э_ц = max|I_z(M,T)| = max \sum_{m=1}^{M} \left| ln \left[T(O_m)/_T \right] \right|;$$

$I_z(M,T)$ – количество получаемой от подсистемы наблюдения осведомляющей содержательной информации; O_m – оператор преобразования тезауруса Т, соответствующий единичному информационному массиву (ИМ) $m \in M$; τ – средний интервал времени переработки осведомляющей информации от одного объекта управления.

Информационная добротность

$$J_{Зц} = \frac{I}{[I_S + I_z[T]]} = = [Э_Ц + I_0 + I_V + + I_Z(T)] / [I_V + I_Z(T)] = = 1 + [Э_Ц + I_0] / [I_V + I_Z(T)].$$

где I – общее количество информации, которое хранится и циркулирует в эргасистеме (узле); I_0 – количество информации, хранимой в информационной базе эргасистемы (узла)

Информационная оперативность:

$$J_{2ц} = \frac{I}{Э_ц},$$

Информационно-технологическая эффективность [10] характеризуется следующими показателями:

Информационная устойчивость:

$$J_{1Т} = \frac{Э_Т}{[Э_Т + I_S(\theta)]}, J_{1Т} \in (0,1),$$

где $Э_Т = H(M_1) - \sum_i p(m_{0i}) H(M_1|m_{0i})$ – мера (оценка) технологического эффекта, получаемого от данного информационного узла в результате выполне-

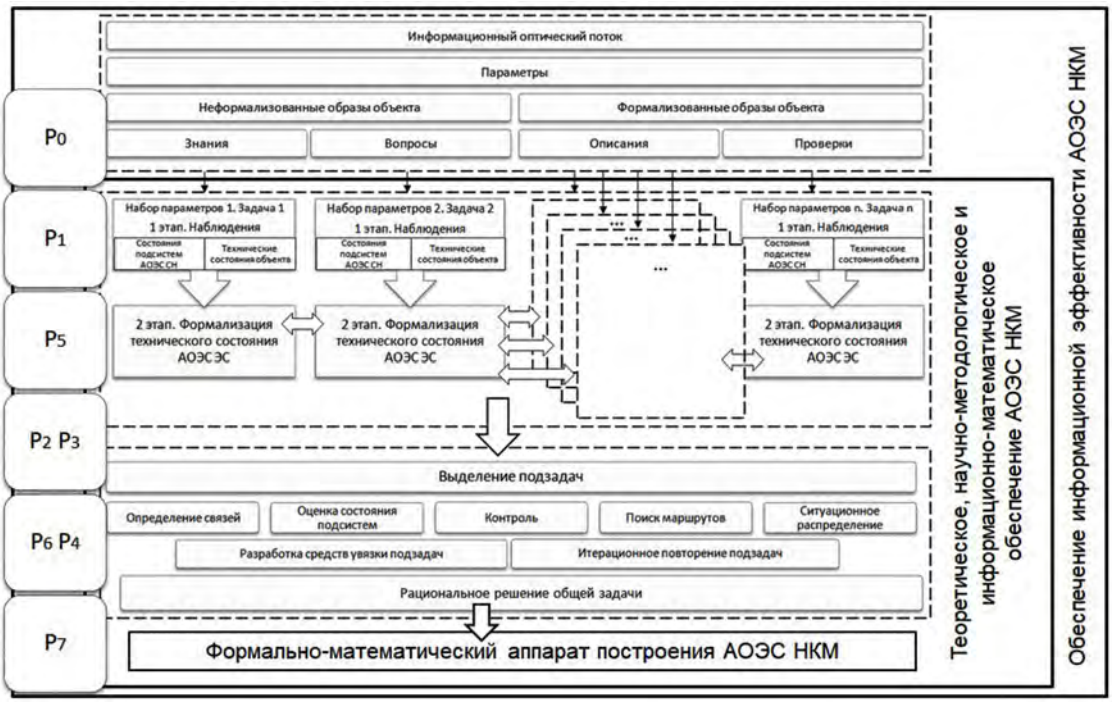

Рис. 3. Информационно-математическая структура АОЭС НКМ

ния процесса переработки информационных массивов (ИМ) m_{0i}, $i = \overline{1, M_0}$;
$H(M_1) = \sum_j p(m_{1j}) \ln p(m_{1j})$ – статистическая энтропия множества переработанных ИМ;
$H(M_1|m_{oi}) = \sum_j p(m_{1i}|m_{oi}) \ln p(m_{1i}|m_{oi})$ – частная (при m_{0i}) условная энтропия множеств переработанных ИМ;
$I_S(\theta)$ – количество используемой структурной информации, содержащейся в информационном узле, определяющее затраты на преобразование содержательной информации.

Информационная живучесть:

$$J_{2\mathrm{T}} = \frac{\mathrm{Э_T}}{N},$$

где N – сумарная производительность всех функциональных компонент.

Информационная имитостойкость:

$$J_{3\mathrm{T}} = \frac{S_{\mathrm{ед}}}{S_{\mathrm{общ}}},$$

где $S_{\mathrm{ед}}$ – единичное подпространство параметров;

$S_{\mathrm{общ}}$ – общее число возможных состояний системы.

Информационно-математическая структура АОЭС НКМ

В общем случае входной информационный поток, поступающий в АОЭС АКМ и подлежащий переработке, содержит множество параметров, характеризующих свойства объекта. Данные параметры, как правило, включают в себя формализованные и неформализованные образы объекта. Формализованные образы объекта представляют собой множество формализованных описаний объекта, отражающих семантические связи между его смысловыми элементами, множество проверок, реализуемых при решении задачи анализа. Неформализованные образы объекта содержат множество знаний об объекте, которыми система располагает и может пополнять в процессе работы, и множество вопросов, формулируемых при принятии решения. Информационно-математическая структура АОЭС НКМ представлена на рис. 3.

Для решения каждой задачи переработки входной визуальной информации используется определенный набор параметров, при этом различные комбинации параметров могут применяться для постановки и решения самых разнообразных задач. Реализация процесса, как правило, проходит через два последовательных этапа. На первом этапе осуществляется наблюдение за состояниями объекта и соответствующими им функциональным состояниям системы. На втором этапе происходит формализация технического состояния системы, в результате переработки формируется информационная модель решения задачи.

Методика решения задачи функционирования АОЭС НКМ

Для решения сложной задачи функционирования АОЭС НКМ разработана методика, использующая новую эффективную технологию переработки визуальной информации в АОЭС НКМ в режиме реального времени с требуемым качеством. Разработка технологии создания программно-аппаратных средств переработки визуальной информации, основанной на информационно-математическом обеспечении многоуровневой АОЭС НКМ, осуществляется путем последовательного выполнения совокупности шагов-этапов до получения результата. Технологическая последовательность шагов обеспечивается упорядоченной совокупностью необходимых моделей, методов и алгоритмов.

Шаг 1. Ввод исходных данных для переработки с учетом заданной ситуационной модели функционирования АОЭС НКМ.

Шаг 2. Оценка движения камеры и стабилизации видеоизображения. Восстановление кадров визуального информационного потока, формирование стабилизированного изображения Φ_{c}.

Шаг 3. Формирование информационной базы АОЭС НКМ, обучение нейросетей для решения задач распознавания визуальной информации. Процесс подготовки информационной базы включает разметку изображений с помощью разработанного универсального метода, учитывающего основные принципы построения обучающих выборок для обучения нейросетевых алгоритмов по комплексным сценариям [13].

Шаг 4. Дешифрирование стабилизированного изображения Φ_{c}. Разработка и модификация эффективных алгоритмов оперативной переработки визуальной информации с помощью формализованных процедур детектирования, локализации и классификации объектов на аэрокосмических изображениях [14, 15].

Шаг 5. Функциональное диагностирование АОЭС НКМ, оценка соответствия заданного качества.

Шаг 6. Выдача результата в виде упорядоченной информационной последовательности описательного характера необходимой для принятия дальнейших решений, а также составления прогнозов и оперативного планирования тактики поведения в условиях информационного соперничества.

Таким образом, получена методология разработки и логической организации модульно-алгоритмического обеспечения эффективной АОЭС НКМ в условиях информационного противоборства. Разработанная объектно-ориентированная технология переработки визуальной информации включает функционально достаточный комплекс методов для решения основных частных задач АОЭС НКМ и многоуровневых решений, основанных как на классических подходах к обработке изображений, так и с применением технологий искусственного интеллекта.

Проблемы внедрения технологий искусственного интеллекта

Внедрение любых новых технологий, как правило, влечет за собой множество вопросов. Высокие темпы технологического прогресса в области искусственного интеллекта создают новые проблемы в защите от негативных последствий неправомерного использова-

DOI:10.21681/2311-3456-2020-06-63-71

ния технологий[2]. Одним из препятствий для широкого внедрения технологий искусственного интеллекта в настоящее время является отсутствие нормативной базы, определяющей единые форматы представления данных на всех стадиях жизненного цикла, принципы отражения специфических угроз информационной безопасности, а также регламентирующей актуальные вопросы создания и эксплуатации безопасных, надёжных и эффективных систем применения искусственного интеллекта. Проблема практически полного отсутствия нормативного правового и технического регулирования условий и особенностей разработки, запуска в работу, функционирования, деятельности, интеграции в другие системы и контроля применения технологий искусственного интеллекта в настоящее время является общемировой. Многообразие областей применения определяет различные формы нормативно-правого регулирования систем искусственного интеллекта.

В разделе I Программы «Цифровая экономика Российской Федерации», утвержденной Распоряжением Правительства Российской Федерации от 28.07.2017 № 1632-р1, технологии искусственного интеллекта (наряду с нейро-технологиями) обозначены как одна из позиций в перечне «сквозных цифровых технологий, которые входят в рамки настоящей Программы»[3]. Стратегия развития отрасли информационных технологий в Российской Федерации на 2014–2020 гг. и на перспективу до 2025 г. прогнозирует, что интеллектуальные системы станут неотъемлемой частью повседневной жизни уже к 2020 году, и предусматривает развитие программных технологий поддержки принятия решений в реальном времени с элементами искусственного интеллекта, а также задействование систем искусственного интеллекта при «анализе больших массивов данных и извлечении знаний, включая новые методы и алгоритмы для сбора, хранения и интеллектуального анализа больших объемов данных (включая вычислительную лингвистику)», а также задействование систем искусственного интеллекта при распознавании образов и интеллектуальном поиске[4].

Выделяют пять критически важных областей для выявления возникающих рисков, связанных с искусственным интеллектом, в том числе доступность ПО, безопасность, подконтрольность, ответственность и этика[5]. Степень влияния данных факторов на общество схематично может быть представлена в виде перевернутой пирамиды, вверху которой указана этика, оказывающая на общество наиболее широкое воздействие, а внизу доступность ПО, влияние которого может быть самым сильным (Рис. 4)

Рис. 4. Критически важные области, связанные с искусственным интеллектом

Программное обеспечение является ключевым элементом систем искусственного интеллекта. Наличие открытых кодов ПО создает риски непреднамеренного неправильного применения технологий. При отсутствии ПО в открытом доступе появляется риск возникновения вредоносного искусственного интеллекта, созданного группой специалистов, допущенных к разработкам, также существует риск использования ИИ преступниками и террористами.

Для обеспечения безопасности необходимы превентивные меры, позволяющие снизить риски непреднамеренных последствий. Для безопасного внедрения необходима организация обязательного тестирования технических систем, основанных на искусственном интеллекте, в условиях приближенных к реальности.

Подконтрольность подразумевает прозрачность и возможность проверки принятия искусственным интеллектом решений. Таким образом, должна обеспечиваться возможность объяснения причин, по которым принято то или иное решение. Прозрачность принятия решений даст возможность обеспечения объективности результатов независимо от личных характеристик потребителя и с учетом того, что обучение алгоритмов проводится с использованием сгенерированных человеком данных.

Система искусственного интеллекта не может нести самостоятельную ответственность за окончательное решение. В то же время решения искусственного интеллекта, получаемые в результате его деятельности, могут приводить к действиям, наносящим ущерб пользователям. Кроме того, с распространением искусственного

2 Взлет искусственного интеллекта: будущие перспективы и возникающие риски [Electronic resource]. URL: https://allianz.ru/ru/stuff/Взлет искусственного интеллекта.pdf (accessed: 16.12.2020).

3 Распоряжение Правительства РФ от 28.07.2017 № 1632-р «Об утверждении программы «Цифровая экономика Российской Федерации» [Electronic resource]. URL: http://government.ru/docs/28653/ (accessed: 29.04.2019).

4 Распоряжение Правительства РФ от 01.11.2013 № 2036-р «Об утверждении Стратегии развития от- расли информационных технологий в Российской Федерации на 2014–2020 годы и на перспективу до 2025 года» // Собрание законодательства РФ. –2013. –№ 46. – С. 5954

5 Распоряжение Правительства РФ от 28.07.2017 № 1632-р «Об утверждении программы «Цифровая экономика Российской Федерации» [Electronic resource]. URL: http://government.ru/docs/28653/ (accessed: 29.04.2019).

интеллекта большое значение приобретают этические проблемы, связанные, как правило, с субъективностью этических принципов.

По итогам конференции разработчиков и исследователей в сфере ИИ Beneficial AI, прошедшей в январе 2017 года в Асиломаре, США были приняты общие принципы искусственного интеллекта[6], авторами которых являются эксперты из Института будущего жизни (Future of Life Institute). Данный кодекс был принят в результат дискуссий ведущих разработчиков в сфере искусственного интеллекта и описывает основные правила изучения и внедрения ИИ, в том числе систем автономного вооружения и интеллектуальных систем способных к самовоспроизведению. Кодекс содержит три основные части: исследования, этические вопросы и долгосрочная перспектива, каждая из которых описывает принципы развития ИИ с максимальной пользой для общества.

Этические вопросы касаются таких проблем как безопасность, открытость сбоев в системе, открытость системе правосудия, ответственность, синхронизация ценностей, общечеловеческие ценности, защита личных данных, свобода и конфиденциальность, совместная выгода, подконтрольность, устойчивость системы, участие в гонке вооружений. Системы ИИ должны быть безопасны и защищены на протяжении всего срока эксплуатации, а в ситуациях, где это целесообразно, штатная работа ИИ должна быть легко верифицируема. Если система ИИ причиняет вред, должна быть возможность выяснить причину. Любое участие автономной системы ИИ в принятии судебного решения должно быть удовлетворительным образом обосновано и доступно для проверки компетентным органами. Системы ИИ с высокой степенью автономности должны быть разработаны таким образом, чтобы их цели и поведение были согласованы с человеческими ценностями на всем протяжении работы. Устройство и функционирование систем ИИ должно быть согласовано с идеалами человеческого достоинства, прав, свобод и культурного разнообразия.

В качестве проблем долгосрочной перспективы рассматривается важность рисков и опасностей недооценки возможностей, рекурсивное автообучение, использование ИИ для всеобщего блага. Применение технологий ИИ может повлечь коренные изменения в истории жизни на Земле, и его разработка и управление должны осуществляться при наличии соответствующих ресурсов и в условиях особой тщательности. Потенциальные риски, связанные с системами ИИ, особенно опасность катастроф или угроза существованию жизни в целом, должны купироваться действиями по планированию и смягчению рисков, соразмерными возможному масштабу воздействия. Системы ИИ, разработанные для улучшения эффективности собственных алгоритмов и самовоспроизведения, ведущего к быстрому изменению качества и количества, должны быть объектом применения мер жесткого регулирования и контроля.

6 Принципы работы с ИИ, разработанные на Асиломарской конференции [Electronic resource]. 2017. URL: https://futureoflife.org/ai-principles-russian/ (accessed: 15.12.2020)

Стандартизация технологий искусственного интеллекта

В июле 2019 года образован технический комитет по стандартизации ТК 164 «Искусственный интеллект», обеспечивающий проведение работ по стандартизации в области технологий искусственного интеллекта. В настоящее время выделен ряд проблем сдерживающих повсеместное внедрение искусственного интеллекта, связанных вопросами нормативно-технического регулирования.

Одной из проблем, возникающих перед разработчиками системы ИИ является необходимость предварительной обработки полученных исходных данных, предназначенных для анализа с помощью систем ИИ. Как правило, разные заказчики используют различные способы получения и хранения данных. В тоже время, сбор данных предполагает получение максимально выверенной исходной информации и является одним из наиболее ответственных этапов в работе с информацией, поскольку от цели сбора и методов последующей обработки полностью зависит конечный результат работы всей информационной системы.

Основным требованием, предъявляемым разработчиками к качеству исходных данных является их унификация в рамках конкретной отрасли применения систем ИИ. Для сбора данных необходимо определить технические средства, позволяющие осуществлять сбор быстро и высококачественно, поддерживающие операции ввода информации и представления данных в электронной форме. Кроме того, необходимо определить нормы и требования к форматам сохранения, предоставления и обмена исходными данными. можно выделить следующие основные требования по стандартизации разработок систем ИИ: стандартизация входных данных для систем ИИ, определение правил сбора, хранения и передачи данных для систем распознавания образов, аудио- и текстовой информации и других возможных входных данных.

Заключение

В основе безопасности технологий искусственного интеллекта возможность внешней проверки данных, четкое выполнение распоряжений человека, а также обязательное включение в программу искусственного интеллекта правил морального и этического характера.

Разработаны подходы к решению задачи эффективного создания, функционирования и эксплуатации безопасных, надёжных и эффективных систем, основанных на искусственном интеллекте.

Формирование моделей нормативно-технического регулирования систем, основанных на технологиях искусственного интеллекта, условий и особенностей разработки, запуска в работу, функционирования, деятельности, интеграции в другие системы и контроля применения технологий искусственного интеллекта может осуществляться на основе многоуровневого концептуально-логического моделирования эргасистем и информационных нормативно-технических отношений.

DOI:10.21681/2311-3456-2020-06-63-71

Литература

1. Гаврилов Д.А., Ловцов Д.А. Автоматизированная оптико–электронная система наземно–космического мониторинга для систем безопасности реального времени // Вопросы кибербезопасности. – 2020. № 5. С. 41–47. DOI: 10.21681/2311-3456-2020-05-41-47
2. Степанов О.А. Правовое регулирование отношений в сфере безопасного функционирования и развития систем искусственного интеллекта: доктринальные аспекты // Правовая информатика. 2019. № 1. С. 56–63. DOI: 10.21681/1994-1404-2019-01-53-63
3. Садыхов Р.Х., Дудкин А.А. Обработка изображений и идентификация объектов в системах технического зрения // Искусственный интеллект. 2006. № 3. С. 694–703.
4. Тропченко А.Ю., Тропченко А.А. Методы вторичной обработки и распознавания изображений. Учебное пособие. СПб: Университет ИТМО. 2015. 215 с.
5. Анисимов Б. В., Курганов В. Д., Злобин В.К. Распознавание и цифровая обработка изображений. М: Высшая школа, 1983. 295 с.
6. Хаустова Е.Ю., Ельцов Д.А., Ершов Д.П. Развитие систем искусственного интеллекта // VIII Международной научно-практической конференции «Научное сообщество студентов XXI столетия. Технические науки». Новосибирск, 2013. С. 67-70.
7. Гаврилов Д.А. Нейросетевой алгоритм автоматического обнаружения и сопровождения объекта интереса в видеосигнале // 16 Национальная конференция по искусственному интеллекту (24–27 сентября 2018 г., г. Москва, Россия). Труды конференции. В 2-х томах. Т 2. М: РКП. 2018. № 8. С. 188–195.
8. Шелухин О.И., Полковников М.В. Классификация зашифрованного трафика мобильных приложений методом машинного обучения // Вопросы кибербезопасности. 2018. т. 28, № 4. С. 21–28. DOI: 10.21681/2311-3456-2018-04-21-28
9. Кругликов С.В., Дмитриев В.А., Степанян А.Б., Максимович Е.П. Политика управления доступом в системе защиты информации высокопроизводительной системы обработки геолого-геофизических данных // Вопросы кибербезопасности. 2018. № 3. С. 22–28. DOI: 10.21681/2311-3456-2018-03-22-28
10. Ловцов Д.А. Основы технологии эффективного двухуровневого правового регулирования информационных отношений в инфосфере // Правовая информатика. 2018. № 2. С. 4–14. DOI: 10.21681/1994-1404-2018-02-4-14
11. Ловцов Д.А. Информационная теория эргасистем: Тезаурус. М: Наука. 2005. 248 с.
12. Скобцов В.Ю., Кругликов С.В., Ким Д.С., Новоселова Н.А., Архипов В.И., Кульбак Л.И., Николаеня Е.Д., Лапицкая Н.В., Вакульчик Е.Н, Саксонов Р. Анализ показателей надежности, живучести и телеметрии бортовой аппаратуры малых космических аппаратов // Вопросы кибербезопасности. 2018. т. 28, № 4. С. 54–69. DOI: 10.21681/2311-3456-2018-04-54-69
13. Гаврилов Д.А., Щелкунов Н.Н. Программное обеспечение разметки крупноформатных аэрокосмических изображений и подготовки обучающих выборок // Научное приборостроение. 2020. т. 30, № 2. С. 67–75.
14. Местецкий Л.М., Гаврилов Д.А., Семенов А.Б. Метод разметки изображений самолетов на аэрокосмических снимках на основе непрерывных морфологических моделей // Программирование. 2019. № 6. С. 3–12.
15. Пунь А.Б., Гаврилов Д.А., Щелкунов Н.Н., Фортунатов А.А. Алгоритм адаптивной бинаризации объектов в видеопоследовательности в режиме реального времени // Успехи современной радиоэлектроники. 2018. № 8. С. 40–48.

REGULATORY ISSUES TO DEVELOP SECURE AUTOMATED SMART SYSTEMS

Gavrilov D.A[7].

Purpose of the Article. *The purpose of the article is to address the regulatory and technical issues of effective creation, operation and operation of safe, reliable and effective systems based on artificial intelligence.*

The research method. *Opportunities for conceptual and logical modeling of ergasystems and invariant architectures of rational modeling based on the problem-oriented version of the integrated "information-cybernetic-didactic" approach using the information and mathematical structure of the automated optical-electronic system of ground-space monitoring are considered.*

Results. *Presented conceptual and logical model of the system of regulatory and technical regulation of systems based on artificial intelligence technologies, and the invariant architecture of the rational model of the artificial intelligence system, developed a method of solving the problem of the operation of the automated optical-electronic system of ground-space monitoring.*

Keywords: *artificial intelligence technologies, regulation of artificial intelligence, two-tier ergasystem, methodological principles, conceptual and logical model.*

7 Dmitry Gavrilov, Ph.D., Researcher, Department of Scientific and Technical Preparation of Production, JSC "Institute of Precision Mechanics and Computer Science named after S.A. Lebedev RAS ", Head of the Laboratory for Special Purpose Digital Systems, Moscow Institute of Physics and Technology (National Research University), Moscow, Russia. E-mail: gavrilov.da@mipt.ru

References

1. Gavrilov D.A., Lovtcov D.A. Avtomatizirovannaia optiko–e`lektronnaia sistema nazemno–kosmicheskogo monitoringa dlia sistem bezopasnosti real`nogo vremeni // Voprosy` kiberbezopasnosti. – 2020. № 5. S. 41–47. DOI: 10.21681/2311-3456-2020-05-41-47
2. Stepanov O.A. Pravovoe regulirovanie otnoshenii` v sfere bezopasnogo funktcionirovaniia i razvitiia sistem iskusstvennogo intellekta: doktrinal`ny`e aspekty` // Pravovaia informatika. 2019. № 1. S. 56–63. DOI: 10.21681/1994-1404-2019-01-53-63
3. Sady`hov R.KH., Dudkin A.A. Obrabotka izobrazhenii` i identifikatciia ob``ektov v sistemakh tekhnicheskogo zreniia // Iskusstvenny`i` intellekt. 2006. № 3. S. 694–703.
4. Tropchenko A.Iu., Tropchenko A.A. Metody` vtorichnoi` obrabotki i raspoznavaniia izobrazhenii`. Uchebnoe posobie. SPb: Universitet ITMO. 2015. 215 s.
5. Anisimov B. V., Kurganov V. D., Zlobin V.K. Raspoznavanie i tcifrovaia obrabotka izobrazhenii`. M: Vy`sshaia shkola, 1983. 295 s.
6. Haustova E.Iu., El`tcov D.A., Ershov D.P. Razvitie sistem iskusstvennogo intellekta // VIII Mezhdunarodnoi` nauchno-prakticheskoi` konferentcii «Nauchnoe soobshchestvo studentov XXI stoletiia. Tekhnicheskie nauki». Novosibirsk, 2013. S. 67-70.
7. Gavrilov D.A. Nei`rosetevoi` algoritm avtomaticheskogo obnaruzheniia i soprovozhdeniia ob``ekta interesa v videosignale // 16 Natcional`naia konferentciia po iskusstvennomu intellektu (24–27 sentiabria 2018 g., g. Moskva, Rossiia). Trudy` konferentcii. V 2-kh tomakh. T 2. M: RKP. 2018. № 8. S. 188–195.
8. Sheluhin O.I., Polkovnikov M.V. Classifikatciia zashifrovannogo trafika mobil`ny`kh prilozhenii` metodom mashinnogo obucheniia // Voprosy` kiberbezopasnosti. 2018. t. 28, № 4. S. 21–28. DOI: 10.21681/2311-3456-2018-04-21-28
9. Kruglikov S.V., Dmitriev V.A., Stepanian A.B., Maksimovich E.P. Politika upravleniia dostupom v sisteme zashchity` informatcii vy`sokoproizvoditel`noi` sistemy` obrabotki geologo-geofizicheskikh danny`kh // Voprosy` kiberbezopasnosti. 2018. № 3. S. 22–28. DOI: 10.21681/2311-3456-2018-03-22-28
10. Lovtcov D.A. Osnovy` tekhnologii e`ffektivnogo dvuhurovnevogo pravovogo regulirovaniia informatcionny`kh otnoshenii` v infosfere // Pravovaia informatika. 2018. № 2. S. 4–14. DOI: 10.21681/1994-1404-2018-02-4-14
11. Lovtcov D.A. Informatcionnaia teoriia e`rgasistem: Tezaurus. M: Nauka. 2005. 248 s.
12. Skobtcov V.Iu., Kruglikov S.V., Kim D.S., Novoselova N.A., Arhipov V.I., Kul`bak L.I., Nicolaenia E.D., Lapitckaia N.V., Vakul`chik E.N, Saksonov R. Analiz pokazatelei` nadezhnosti, zhivuchesti i telemetrii bortovoi` apparatury` maly`kh kosmicheskikh apparatov // Voprosy` kiberbezopasnosti. 2018. t. 28, № 4. S. 54–69. DOI: 10.21681/2311-3456-2018-04-54-69
13. Gavrilov D.A., Shchelkunov N.N. Programmnoe obespechenie razmetki krupnoformatny`kh ae`rokosmicheskikh izobrazhenii` i podgotovki obuchaiushchikh vy`borok // Nauchnoe priborostroenie. 2020. t. 30, № 2. S. 67–75.
14. Mestetckii` L.M., Gavrilov D.A., Semenov A.B. Metod razmetki izobrazhenii` samoletov na ae`rokosmicheskikh snimkakh na osnove neprery`vny`kh morfologicheskikh modelei` // Programmirovanie. 2019. № 6. S. 3–12.
15. Pun` A.B., Gavrilov D.A., Shchelkunov N.N., Fortunatov A.A. Algoritm adaptivnoi` binarizatcii ob``ektov v videoposledovatel`nosti v rezhime real`nogo vremeni // Uspehi sovremennoi` radioe`lektroniki. 2018. № 8. S. 40–48.