Программные системы и вычислительные методы
Правильная ссылка на статью:
Тиханычев О.В. О разделении ответственности за ошибки эксплуатации робототехнических систем // Программные системы и вычислительные методы. 2025. № 3. DOI: 10.7256/2454-0714.2025.3.72826 EDN: OMKHOA URL: https://nbpublish.com/library_read_article.php?id=72826

О разделении ответственности за ошибки эксплуатации робототехнических систем

Тиханычев Олег Васильевич

ORCID: 0000-0003-4759-2931

кандидат технических наук

заместитель начальника отдела управления перспективных разработок, ГК "Техносерв"

111395, Россия, г. Москва, ул. Юности, 13

to.technoserv@gmail.com

Статья из рубрики "Показатели качества и повышение надежности программных систем"

DOI:

10.7256/2454-0714.2025.3.72826

EDN:

OMKHOA

Дата направления статьи в редакцию:

23-12-2024

Аннотация: Актуальность выбора предмета исследования как основы обеспечения безопасности применения робототехнических систем различного назначения, в первую очередь – использующих для управления искусственный интеллект, и объекта исследования, которым являются проблемы разделения ответственности за разработку и эксплуатацию робототехнических систем, определяется имеющимся противоречием между потребностью автономного применения робототехнических систем и сложностью программной реализации этого требования. В то же время, в робототехнике, довольно часто, именно ошибки алгоритмов управления служат источником большинства проблем. На основании анализа нормативных документов, регламентирующих разработку средств искусственного интеллекта, проанализированы возможные проблемы обеспечения безопасности применения автономных робототехнических систем. Синтезирован вывод, что в существующем состоянии данные документы не обеспечивают решения проблемы безопасности систем искусственного интеллекта. В качестве методологической основы исследования был избран системный подход. Использование системного подхода,

методов декомпозиции и сравнительного анализа, дало возможность рассматривать в комплексе проблемы разделения зон ответственности разработчиков и эксплуатантов автономных и частично автономных роботов, реализующих принципы управления на основе искусственного интеллекта. Источниковую базу исследования составили научные статьи, нормативные и законодательные документы, находящиеся в открытом доступе. Сделан вывод, что существующие подходы к обучению и самообучению систем искусственного интеллекта, управляющих автономными роботами “размывают” границы ответственности участников процесса, что, в теории, может приводить к возникновению критических ситуаций при эксплуатации. С учётом этого, на основе анализа типового процесса разработки и применения, предложено уточнить распределение ответственности, а также добавить в процесс новых участников: дополнить его специалистами, направленно занимающимися безопасностью и непредвзятостью искусственного интеллекта (AI Alignment), а таже обеспечить групповой подход в разработке алгоритмов искусственного интелелкта и машинного обучения, обеспечивающий снижение фактора субъективности. Применение синтезированых в статье принципов разделения ответственности, теоретически, обеспечит повышение безопасности робототехнических систем, построенных на основе использования искусственного интеллекта

Ключевые слова:

робототехнические системы, автономные роботы, искусственный интеллект, ошибки при разработке, ошибки эксплуатации, контроль безопасности алгоритмов, ответственность разработчика, качество тестирования, ответственность эксплуатанта, разделение ответственности

Введение

В последнее время в специальной литературе всё чаще появляются упоминания о проблемах в обеспечении безопасности при эксплуатации робототехнических систем (РТС) различного назначения, особенно в автономном исполнении. Даже несмотря на то, что любая автономность является довольно относительным понятием: автономная система просто по гибким алгоритмам выполняет задания, поставленные человеком, реализуя большую или меньшую степень самостоятельности.

Как показывает практика, доля автономных систем постоянно растёт во всех областях: в промышленности, в транспорте, в военной сфере, что является объективной тенденцией, важность и влияние которой невозможно не учитывать [1,2,3].

Разумеется, при выполнении этих заданий могут возникать ошибки. Часть из них будет критическими, приводящими к материальным потерям или даже гибели людей. И тогда возникает вопрос ответственности за совершения данных ошибок.

Существующая система разделения ответственности между разработчиками и эксплуатантами сложных технических систем построена на следующих принципах:

- разработчик создаёт систему, функционирующую в рамках заданных заранее условий, разрабатывает и описывает правила её эксплуатации, проводит, с привлечением заказчика, приёмо-сдаточные испытания и гарантирует, что в заданных границах система будет функционировать успешно;

- пользователь до начала эксплуатации обязан изучить систему, подтвердить свои знания и, в дальнейшем, эксплуатировать её в рамках требований, описанных в документации, с соблюдением общепринятых нормативных правил;

- техническая служба пользователя или заказчика, служба поддержки, должны обеспечивать своевременный и полный контроль технических параметров системы, установленных в эксплуатационной документации, своевременно проводить требуемое обслуживание;

- в случае аварий и катастроф привлекается внешний арбитраж, создаётся комиссия, определяющая причины происшествия и его виновников.

Эта система долго и успешно работает в различных областях: на производстве, на транспорте и в других.

Актуальность поднятой в статье проблемы в том, что для РТС, особенно автономных, такая система может давать сбой. Данный вывод основан на том, что вся используемая в настоящее время система распределения ответственности за эксплуатацию сложных технических систем основана на статичности их структуры и правил поведения. Эта структура разрабатывается в ходе создания систем и поддерживается неизменной в ходе эксплуатации. Изменения структуры производятся достаточно редко и по строго определённому алгоритму, с фиксацией доработок в конструкторской и эксплуатационной документации, с доведением изменений до всех участников процесса.

Данный подход не предусматривает, что в ситуации применения РТС, особенно автономных, может проводиться самостоятельная модификация некоторых компонентов системы в ходе эксплуатации. А последнее является нормой для автономных систем, использующих алгоритмы управления на основе обучения и самообучения.

Таким образом, проблема распределения ответственности при разработке и эксплуатации автономных РТС является своевременной и актуальной.

1 Анализ проблемы разделения ответственности

Рассмотреть проблему предлагается на основе анализа типовых ошибок, возникающих при эксплуатации РТС.

Учитывая типовую структуру РТС, ошибки её эксплуатации могут порождаться двумя основным группами причин [4]:

- поломками в электромеханической части, служащей, преимущественно, в качестве исполнительных механизмов РТС;

- сбоями и ошибками алгоритмов программного обеспечения, обеспечивающего управление РСТ.

С ответственностью за проблемы первого рода относительно понятно – они делятся между разработчиком и эксплуатантами РТС в соответствии с их обязанностями, прописанными в рабочей конструкторской и эксплуатационной документации (ЭД). Этот процесс отработан многолетней практикой и корректно описан в нормативно-технической документации. Разумеется, может возникнуть ситуация, когда изменения в собственную конструкцию, потенциально приводящие к нештатной ситуации, попытается внести автономный робот. Но это скорее исключение, причём относящее к сфере алгоритмических, а не технических, проблем.

Намного сложнее вопрос разделения ответственности решается в части ошибок программного обеспечения. Особенно для автономных РТС, преимущественно реализующих обучаемые, то есть модифицируемые алгоритмы.

2 Варианты решения проблемы

Несмотря на внешнее впечатление самостоятельной работы алгоритмов автономных РТС, как технических, так и программных «ботов», на самом деле за любым действием алгоритма стоит человек. Вопрос лишь в количестве циклов преобразования между сформированным человеком начальным, и реализуемым РТС конечным алгоритмом. В разработке и уточнении алгоритмов автономных РТС участвует целая группа специалистов: аналитики, алгоритмисты, программисты, тестировщики, специалисты по управлению знаниями, эксплуатанты в части как определения границ постановки задач, так и дообучения в ходе применения. Каждый из указанных специалистов выполняя свою часть совместной работы, задаёт исходные данные для последующих участников цикла разработки и эксплуатации. Теоретически, каждый из участников процесса может совершать ошибки разного рода: как создавая их, так и пропуская, и закладывая основу для ошибок, создаваемых следующими участниками цикла разработки. В итоге формируется вложенная многоуровневая система, потенциально содержащая набор взаимоувязанных частных алгоритмов, работающих с той или иной степенью надёжности [5,6,7].

В случае реализации в РТС принципа самообучения, задача становится ещё сложнее, так как алгоритмы превращаются в эвристические, а система дополняется ещё одним уровнем вложенности и, соответственно, увеличивает количество степеней свободы. В части разработки управляющего программного обеспечения (ПО) РТС это могут быть следующие уровни:

- уровень формирования алгоритмов;

- уровень программного описания;

- уровень тестирования;

- уровень обучения системы;

- уровень постановки задач при применении.

Как показывает анализ, в настоящее время ситуация относительно ответственности на каждом этапе разработки робототехнических систем различной сложности складывается весьма неопределённым образом (таблица 1).

Таблица 1 - Существующий подход к распределению ответственности за ошибки РТС

Тип РТС	Ответственность участников процесса				
	заказчик	разработчик технической части	разработчики программного обеспечения	тестировщики	учи (Mas Leane
Дистанционно управляемая	В рамках сформулированных технических требований	За штатный функционал в рамках технического задания на разработку	За штатный функционал в рамках требований ТЗ	В границах проверок, проведённых по утверждённых заказчиком	

		(ТЗ)		методик	
С частичной автономностью	В рамках сформулированных технических требований	За штатный функционал в рамках ТЗ	За штатный функционал в рамках требований ТЗ	В границах проверок, проведённых по утверждённых заказчиком методик	За об поведе ча допус гра автоно
Автономная	В рамках сформулированных технических требований	За штатный функционал в рамках ТЗ	В настоящее время не определено	В настоящее время не определено	Н опред

Ситуация осложняется тем, что деятельность части специалистов, функционал которых приведён в таблице 1, вообще не регламентируются существующими правовыми и нормативно-техническими документами и они официально выполняют свои обязанности на должностях с другим функционалом.

Разумеется, подобная ситуация сохраняться в текущем состоянии не может, она мешает развитию РТС, особенно автономных и частично автономных, использующих для управления искусственный интеллект (AI). Исходя из этого, требуется корректировка нормативно-правовой документации, в том числе – в части уточнения ответственности всех участников процесса.

На первый взгляд, как основа для формирования изменений, напрашивается аналогия с человеческим воспитанием: вклад родителей за счёт передачи генетики, ответственность за воспитание. Но, в данном случае, это не совсем корректная аналогия. Аналог генетической наследственности, это, скорее, техническая составляющая РТС. А вот программная часть, с точки разделения на вклад разработчика, инженера по обучению (ML-инженера) и эксплуатанта, продолжающего обучение и закрепление навыков – это, скорее, деление на безусловные и условные рефлексы. Причём вырабатываемых как целенаправленно, при ML, так и относительно ненаправленно, при самообучении (дообучении) алгоритмов в процессе применения РТС.

Приняв такое предположение, можно попытаться сформулировать соответствующее ему перспективное распределение ответственности участников процесса разработки. Дополнительно, учитывая, что основные нерегулируемые угрозы потенциально исходят от автономных РТС, управляемых ИИ, может потребоваться оценка основных параметров ИИ, влияющих на безопасность управляемых им систем. В первую очередь к ним относят «доверенность» (trustworthy) и «предвзятость» (Fair).

В части оценки параметра «доверенность», в соответствии с положением Еврокомиссии «Руководство по этике для надежного ИИ» 2019 года (“Ethics guidelines for trustworthy AI, 2019”), ИИ считается безопасным, если он обладает следующими качествами [8,9]:

- проверяемость (verifiability);

- управляемость (controllability);

- стабильность (stability);

- робастность (robustness);

- безопасность (security);

- отказоустойчивость (fault).

В России безопасность систем, управляемых искусственным интеллектом регулируется положениями «Национальной стратегии развития искусственного интеллекта на период до 2030 года», утверждённой указом Президента РФ от 10 октября 2019 № 490 «О развитии искусственного интеллекта в Российской Федерации», а также документом «Кодекс этики в сфере искусственного интеллекта». Последний можно считать образцом мягкого регулирования, это, скорее, даже не кодекс ИИ, а набор базовых правил для его разработчиков. В остальном, положения указанных документов во многом похожи на соответствующие документы зарубежных государств.

Во всех указанных документах предусмотрено, что алгоритмы, определяющие доверенность ИИ, закладываются на этапе разработки программного обеспечения (ПО) РТС.

В части непредвзятости ИИ, которая, по сути, коррелируется с некоторыми качествами доверенности, базовым положением считается то, что искусственный интеллект ассоциирован с программами, поэтому считается, что он всегда будет занимать справедливую позицию, свободную от предвзятостей. Как показывает практика, это заблуждение: алгоритмы ИИ «знают» что-либо только потому, что они обучены на данных, которые создаются и отбираются людьми. Поскольку все люди по своей природе субъективны, это неизбежно влияет на результаты работы алгоритмов. Качество алгоритмов ИИ определяется данными, на которых они обучены. Это подтверждает известное эмпирическое правило, касающееся любых компьютерных систем: «Мусор на входе – мусор на выходе». Если система часто переобучается, например, с использованием новых данных из соцсетей, она будет уязвима к предвзятости или злонамеренным влияниям.

Примером последнего может служить попытка создания нейросети Delphi, которая должна была стать неким «этическим компасом». Основная цель разработки Delphi – избавить ее от всевозможных предвзятостей и сделать беспристрастной (дескриптивная моральная оценка запросов). Результат эксперимента получился двояким, разработанная модель столкнулась с высокими оценками от профессионального сообщества, но критикой со стороны обычных пользователей.

В то же время, специалисты рассматривают предвзятость ИИ как одну из основных опасностей в мире, где компьютерные программы могут принимать самостоятельные решения. Значительная часть ведущихся в настоящее время исследований направлена на минимизацию и устранение риска предвзятости. Так появилось целое направление оценки непредвзятости ИИ (AI Alignment). Основная цель этой области — гарантировать, что ИИ будет безопасным, предсказуемым и действовать в соответствии с человеческими ценностями [10,11].

Учитывая вышеизложенное, уровни и содержание ответственности разработчиков ИИ могут быть определены на основе типичного алгоритма процесса разработки, включающего, в части разработки ПО для РТС и программных ботов:

- поисковые и предпроектные исследования;

- описание функциональной и информационной моделей управляемой системы и

процессов её функционирования;

- описание базовых алгоритмов функционирования;

- написание программного кода;

- первичное обучение систем;

- тестирование;

- контроль поведения и дообучение в ходе эксплуатации.

Возможный вариант распределения ответственности при использовании подобной модели разработки, основанный на уточнении существующих подходов с учётом особенностей использования компонентов ИИ, приведен в таблице 2. Отметим, что данные в таблице формировались с преимущественной оценкой безопасности применения ИИ с разными уровнями автономности, от управляемых человеком, до полностью автономных).

Таблица 2 - Предложения по уточнению распределения ответственности

Тип РТС	Участник процесса	Зона ответственности
Дистанционно управляемая	Разработчик технических средств	Исправность системы, включая предохранительные механизмы и каналов управления
	Тестировщик	Проверка выполнения требований технического задания на разработку системы
	Оператор РТС	Выполнение заранее определённых правил и границ применения системы
С частичной автономностью	Разработчик технических средств	Исправность системы, включая предохранительные механизмы и каналов управления
	Алгоритмист	Разработка корректных алгоритмов управления, обеспечивающих надежное выполнение требований технического задания
	Разработчик программного обеспечения	Разработка управляющего ПО, обеспечивающего точную реализацию заданных алгоритмов
	Тестировщик	Проверка выполнения требований технического задания на разработку системы
	Оператор РТС	Выполнение заранее

		определённых правил и границ применения системы
Автономная с жесткими алгоритмами	Разработчик технических средств	Исправность системы, включая предохранительные механизмы и каналов управления
	Алгоритмист	Разработка корректных алгоритмов управления, обеспечивающих надежное выполнение требований технического задания
	Разработчик программного обеспечения	Разработка управляющего ПО, обеспечивающего точную реализацию заданных алгоритмов
	Тестировщик	Проверка выполнения требований технического задания на разработку системы
	Оператор РТС	Постановка задач РТС, укладывающихся в заданные в технической документации границы применения
Автономная обучаемая	Разработчик технических средств	Исправность системы, включая предохранительные механизмы и каналов управления, в том числе, при работе в нерасчётных режимах
	Алгоритмист	Разработка корректных алгоритмов управления, обеспечивающих надежное выполнение требований технического задания, с учётом принципов их возможной модификации в ходе обучения
	Разработчик программного обеспечения	Разработка управляющего ПО, обеспечивающего точную реализацию заданных алгоритмов, как базовых, так и настраиваемых
	ML-инженер	Формирование новых алгоритмов (правил) поведения, учитывающих обязательные ограничения на определённые действия
	Тестировщик	Проверка выполнения требований технического задания на разработку

		системы, а также ограничений по безопасности для вновь формируемых алгоритмов поведения
	Оператор РТС	Постановка задач РТС, укладывающихся в заданные в технической документации границы применения, формирование дополнительных ограничений для разрабатываемых алгоритмов применения
Автономная самообучаемая	Разработчик технических средств	Исправность системы, включая предохранительные механизмы и каналов управления, в том числе, при работе в нерасчётных режимах
	Алгоритмист	Разработка корректных алгоритмов управления, обеспечивающих надежное выполнение требований технического задания, с учётом принципов их модификации в ходе обучения
	Разработчик программного обеспечения	Разработка управляющего ПО, обеспечивающего точную реализацию заданных базовых алгоритмов с учётом возможных границ их модификации
	ML-инженер	Формирование начальных алгоритмов поведения и алгоритмов самообучения, учитывающих формирование обязательных ограничений по безопасности
	Тестировщик	Проверка выполнения требований технического задания на разработку системы, а также ограничений по безопасности для вновь формируемых алгоритмов поведения
	Эксплуатант, оператор (постановщик задач) РТС	Постановка задач РТС, укладывающихся в заданные в технической документации границы применения, формирование дополнительных ограничений

		по результатам выполнения задач
	Регуляторы нормативной сферы ИИ	Разработка нормативов и правил поведения РТС, управляемых ИИ, регламентов разработки

Приняв содержание таблицы 2 за вариант распределения ответственности всех участников процесса, можно с некоторой точностью определить ответственных за критичные ошибки РТС, а также сформулировать меры по их предотвращению [12,13].

Детализация предложений, описанных в таблице 2, может быть осуществлена выполнением следующего перечня мер по обеспечению решения проблемы безопасности РТС, в том числе, управляемых ИИ:

- перечень участников процесса необходимо дополнить специалистами, направленно занимающимися безопасностью и непредвзятостью ИИ: так называемым AI Alignment, работающими на стыке областей алгоритмизации и машинного обучения. Отметим, что у крупнейших разработчиков современных систем ИИ в структуру уже включены собственные отделы, занимающиеся AI Alignment: в OpenAI,это Superalignment Team, в Anthropic, команда AI Safety and Alignment, в DeepMind, это AI Safety and Alignment Team, а в Google Research – AI Safety and Alignment Organization;

- обеспечить групповой подход в разработке алгоритмов AI и ML, обеспечивающий, при независимой организации работ, снижение фактора субъективности;

- уточнить подходы к тестированию ИИ, расширив их рамки и полноту;

- обеспечить создание «внешних», в том числе программно-механических ограничителей на деятельность потенциально опасных систем, управляемых ИИ.

Данные дополнения не противоречат предложениям, изложенным в таблице 2, а дополняют их. Для практической реализации указанных предложений потребуется корректировка нормативно-технической документации, регламентирующей данную предметную область [14].

3 Некоторые выводы

Таким образом, анализ сформулированной в статье проблемы распределения ответственности за возникновение критичных ошибок в ходе эксплуатации РТС показывает, что при существующем состоянии нормативно-технической и правовой документации, эта проблема решена быть не может. В то же время, учитывая рост критичных ситуаций, возникающих при применении РТС [15,16], сформулированная проблема является своевременной и актуальной. В рамках разрешения данной ситуации, используемые в настоящее время принципы обучения AI и контроля его «доверенности», на выполнении которых основано содержание современной нормативно-технической документации, предлагается дополнить понятием «воспитание», формируемым на основе синтезированных в статье принципов разделения ответственности, а перечень участников процесса дополнить специалистами по безопасности ИИ. В более отдалённой перспективе вполне вероятно появление специалистов по раннему выявлению и исправлению ошибок поведения ИИ, своего рода «роботопсихологов», с формированием соответствующей отрасли науки, например «психология искусственного интеллекта». Впрочем, это отдалённая перспектива, а в настоящее время необходимо оперативно

решить вопрос именно с ответственной организацией процесса первичного и последующего обучения ИИ, с внесением изменений в нормативную и правовую документацию.

Предложенный в статье вариант уточнённого распределения ответственности участников процесса разработки может служить обобщённой базой для решения проблемы безопасности ИИ и управляемых им РТС.

Заключение

В существующей нормативно-технической документации, определяющей разработку ИИ, сформирован в том или ином виде перечень специалистов, участвующих в разработке и применении компонентов ИИ и задекларированы в общем виде принципы обеспечения безопасности ИИ, которые рекомендованы для соблюдения.

В статье впервые сформулирована проблема ответственности за поведение РТС, управляемых ИИ, а также предложены общие принципы разделения ответственности за критичные ошибки ИИ. В качестве дальнейшего направления исследований можно определить дальнейшую детализацию границ ответственности и их закрепление в нормативно-технической документации.

Библиография

1. Чиров Д.С., Новак К.В. Перспективные направления развития робототехнических комплексов специального назначения // Вопросы безопасности. 2018. № 2. С. 50-59. DOI: 10.25136/2409-7543.2018.2.22737 URL: https://e-notabene.ru/nb/article_22737.html
2. John W. Tammen NATO Basic Concept of Warfare: Looking Ahead – The Changing Nature of Warfare // NATO Review. 2021 URL: https://www.nato.int/docu/review/ru/articles/2021/07/09/bazovaya-kontseptsiya-boevyh-dejstvij-nato-v-perspektive-menyayushchijsya-harakter-vojny/index.html.
3. Хрипунов С.П., Чиров Д.С., Благодарящев И.В. Военная робототехника: современные тренды и векторы развития // Тренды и управление. 2015. № 4. С. 410-422. URL: https://e-notabene.ru/tumag/article_67141.html
4. Pflimlin É Drones et robots: La guerre des futurs. France: Levallois-Perret. 2017.
5. Roosevelt, Ann. Army Directs Cuts, Adjustments, To FCS. Defense Daily. 2017.
6. Hamilton T How AI will Alter Multi-Domain Warfare // Future Combat Air & Space Capabilities Summit. 2023. No.4. URL: https://www.aerosociety.com/events-calendar/raes-future-combat-air-and-space-capabilities-summit.
7. Tikhanychev O.V. Exploring Morality and Politeness in the Context of Robotic Systems: A Conceptual Interpretation // BIO Web of Conferences. 2024, No.138. 03020. doi.org/10.1051/bioconf/202413803020.
8. Beard J. Autonomous weapons and human responsibilities // Georgetown Journal of International Law. 2014. No. 45, pp. 617–681.
9. Schuller A. At the Crossroads of Control: The Intersection of Artificial Intelligence in Autonomous Weapon Systems with International Humanitarian Law // Harvard National Security Journal. 2017. No. 8. pp. 379-425.
10. Ухоботов В.И., Изместьев И.В. Об одной задаче преследования при наличии сопротивлении среды // Вестник Южно-Уральского государственного университета. Серия «Математика. Механика. Физика». 2016. № 8(2). С. 62-66. doi.org/10.14529/mmph160208.
11. Tikhanychev O.V. Self-Check System of Heuristic Algorithms as a "New Moral" of Intelligent Systems // AIP Conference Proceedings. 2023. No. 2700. 040028 https://doi.org/10.1063/5.0124956.

12. Ćwiąkała P. Testing Procedure of Unmanned Aerial Vehicles (UAVs) Trajectory in Automatic Missions // Applied Science. 2019. No. 9. pp. 3488. doi.org/10.3390/app9173488.
13. Johnson D Computer Systems: Moral entities but not moral agents // Ethics and Information Technology. 2016. No. 8. pp. 195-204. doi.org/10.1007/s10676-006-9111.
14. Дубанов А.А. Моделирование траектории преследователя в пространстве при методе параллельного сближения // Программные системы и вычислительные методы. 2021. № 2. С. 1-10. doi.org/10.7256/2454-0714.2021.2.36014.
15. Tikhanychev O.V. Development of situational algorithms for the use of robotic systems // E3S Web of Conferences. 2024. No. 531. 02004. doi.org/10.1051/e3sconf/202453102004.
16. Курденкова Е.О., Черепнина М.С., Чистякова А.С., Архипенко К.В. Влияние трансформаций на успешность состязательных атак для классификаторов изображений Clipped BagNet и ResNet. Труды ИСП РАН, том 34. Вып. 6. 2022. С. 101-116. doi.org/10.15514/ISPRAS-2022-34(6)-7.

Результаты процедуры рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.
Со списком рецензентов издательства можно ознакомиться здесь.

Рецензируемая статья посвящена исследованию вопросов разделения ответственности за ошибки эксплуатации робототехнических систем.
Методология выполненной работы базируется обобщении современных публикаций по изучаемой теме и практического опыта эксплуатации робототехнических систем, управляемых искусственным интеллектом.
Актуальность работы определяется тем, в процессе эксплуатации робототехнических систем (РТС) возможны сбои, сопровождающиеся нанесением вреда жизни и здоровью людей, причинением материального ущерба, а существующий подход к распределению ответственности подход не предусматривает, что в ситуации применения РТС может проводиться самостоятельная модификация некоторых компонентов системы в ходе эксплуатации с использованием алгоритмов управления на основе обучения и самообучения.
Научная новизна рецензируемого исследования состоит в предложенных авторами общих принципах разделения ответственности за критичные ошибки при эксплуатации РТС с привлечением искусственного интеллекта. Предложенный в статье вариант распределения ответственности участников процесса разработки РТС может служить обобщённой базой для решения проблемы безопасности применения систем искусственного интеллекта.
Структурно в статье выделены следующие разделы: «Введение», «Анализ проблемы разделения ответственности», «Варианты решения проблемы», «Некоторые выводы», «Заключение» и «Библиография».
В статье отражена существующая система разделения ответственности между разработчиками и эксплуатантами сложных технических систем. Сказано, что ошибки эксплуатации могут порождаться двумя основным группами причин: во-первых, поломками в электромеханической части, служащей, преимущественно, в качестве исполнительных механизмов РТС; в-вторых, сбоями и ошибками алгоритмов программного обеспечения, обеспечивающего управление РСТ. Авторы справедливо считают, что несмотря на внешнее впечатление самостоятельной работы алгоритмов автономных РТС, как технических, так и программных «ботов», на самом деле за любым действием алгоритма стоит человек. Поскольку в разработке и уточнении алгоритмов автономных РТС участвует целая группа специалистов, каждый из которых теоретически

может совершать ошибки разного рода, то распределить ответственность бывает непросто, а при реализации принципа самообучения эта задача становится ещё сложнее. В публикации содержатся предложения по уточнению распределения ответственности между участниками процесса: разработчиками технических средств, тестировщиками, операторами РТС, разработчиками алгоритмов и программного обеспечения, ML-инженерами, эксплуатантами, регуляторами нормативной сферы применения искусственного интеллекта с указанием зон ответственности каждого участника с учетом типа РТС. Используемые в настоящее время принципы обучения искусственного интеллекта и контроля его «доверенности» предлагается дополнить понятием «воспитание», формируемым на основе синтезированных в статье принципов разделения ответственности, а перечень участников процесса дополнить специалистами по безопасности искусственного интеллекта.

Библиографический список включает 16 источников – научные публикации по рассматриваемой теме отечественных и зарубежных авторов на русском и иностранных языках. В тексте публикации имеются адресные отсылки к списку литературы, подтверждающие наличие апелляции к оппонентам.

Рецензируемый материал соответствует направлению журнала «Программные системы и вычислительные методы», отражает результаты проведенного авторского исследования, содержит весьма своевременные разработки, может вызвать интерес у читателей, рекомендуется к опубликованию.