Программные системы и вычислительные методы
Правильная ссылка на статью:
Тиханычев О.В. Об уточнении понятия «доверенности» систем искусственного интеллекта // Программные системы и вычислительные методы. 2024. № 3. DOI: 10.7256/2454-0714.2024.3.44097 EDN: JOPHLF URL: https://nbpublish.com/library_read_article.php?id=44097

Об уточнении понятия «доверенности» систем искусственного интеллекта

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 "Базы знаний, интеллектуальные системы, экспертные системы, системы поддержки принятия решений"

DOI:

10.7256/2454-0714.2024.3.44097

EDN:

JOPHLF

Дата направления статьи в редакцию:

22-09-2023

Аннотация: Предметом исследования является понятие «доверенности» искусственного интеллекта, управляющего робототехническими средствами разной степени автономности. Актуальность выбора предмета исследования как принципов применения робототехнических систем различного назначения, в том числе в группе, и объекта исследования, которым являются алгоритмические проблемы, возникающие в части реализации алгоритмов групповых действий, определяется имеющимся противоречием между потребностью совместного применения робототехнических систем, в первую очередь автономных, и сложностью программной реализации этого требования. Реализация тенденций группового применения в сфере робототехники порождает определённые технологические проблемы, связанные с эффективностью и безопасностью алгоритмического обеспечения автономных и управляемых робототехнических систем. Проявлением подобных проблем могут являться ошибки применения, которые снижают эффективность совместных действий. В робототехнике

основная потенциальная причина появления подобной ситуации – недостаточная эффективность существующих алгоритмов управления групповым применением, определяемая низким уровнем проработки проблемы. В статье сформулирован перечень типовых ситуаций, определяющих применение автономных и управляемых роботов в группе с наличием ведущего (лидера). На основании предлагаемой классификации проанализированы возможные алгоритмы, обеспечивающие управление перемещением в группе: как расчёты по целевому маневрированию, так по обеспечению взаимной безопасности. Рассмотрены основные ситуации, относительно которых сформулированы виды маневра и описан математический аппарат для их расчётов. На основе обзорного анализа типовых алгоритмов управления перемещением в пространстве, синтезирована постановка научной задачи решения проблемы разработки групповых алгоритмов и математических методов, определяющих их оптимальное применение в системе, как аналога "новой этики" для робототехнических систем

Ключевые слова:

робототехническая система, искусственный интеллект, управляющее программное обеспечение, доверенный искусственный интеллект, контроль алгоритмов поведения, безопасное поведение, мораль робототехнических систем, этика робототехнических систем, математическое обеспечение безопасности, управление безопасным поведением

1 Введение

Одним из приоритетов, обеспечивающим активное развитие автономной робототехники, является использование элементов искусственного интеллекта (ИИ). В настоящее время компоненты ИИ содержат большинство робототехнических систем (РТС) с разными уровнями автономности:

- ограниченно автономные, например, дроны, самостоятельно возвращающиеся на базу при потере связи с оператором;

- частично автономные, с ограничением на автономное выполнение некоторых функций, например – применения оружия для боевых РТС, решение на которое пока остаётся за оператором;

- полностью автономные, решающие поставленные задачи самостоятельно.

Оснащение РТС компонентами ИИ, их дальнейшая автономизация с использованием принципов обучения и самообучения, всё более отдаляет их поведение от роботов с внешним управлением, всё более приближает к разумному, со всеми плюсами и минусами этих изменений. При этом, именно повышение автономности РТС является объективной тенденцией, основанной на требованиях к применению робототехники в самых разных сферах.

В то же время, остаётся целый ряд нерешенных проблем, периодически приводящих к авариям и катастрофам с человеческими жертвами в промышленной и транспортной сфере [1,2] и, конечно, в сфере вооруженного противоборства [3,4,5]. Многие из этих проблем определяются нерешенностью проблем предметной области в части определения безопасности систем ИИ. Используемый в настоящее время термин «доверенность» не в полной мере тождественен понятию «безопасность» в комплексном понимании этого явления, а других подходов к описанию данного фактора пока не

предлагается.

В то же время, именно описание предметной области служит основной развития теории и практики любых систем. И, одновременно, развитие автономных систем не может быть остановлено, оно продолжается, как продолжается технический прогресс, несмотря не некоторые ограничения, отражаемые в официальных документах, регулирующих сферу ИИ.

Таким образом, проблема уточнения предметной области в части обеспечения безопасности ИИ, управляющего РТС не решена и остаётся актуальной, что делает актуальной и тему статьи.

2 О существующем подходе к определению «доверенных» систем

Практика показывает, что основа любой классификации – условия функционирования описываемой системы.

Обзор перечня и условий возникновения катастроф и аварий, произошедших по вине автономных и частично автономных РСТ показывает, что все их причины могут быть объединены в две большие группы;

- условно закономерные, связанные с решениями программного обеспечения, управляющего РТС при неправильной формулировке задачи или ошибках распознавания ситуации, либо при неверном учёте ограничений на применение системы;

- случайные, определяемые или не обнаруженными в ходе тестирования ошибками алгоритмов поведения, либо ограниченным временем не принятие решения, не укладывающимся в заданную при разработке длительность цикла управления и не позволяющим программному обеспечению РТС корректно отработать полный цикл управления.

Примером первой группы ошибок является озвученная полковником ВВС США Такером Гамильтоном (Tucker Hamilton) в докладе на конференции «Future Combat Air & Space Capabilities Summit» ситуация с решением ИИ убить собственного оператора, которого боевая РТС посчитала помехой для достижения поставленной ей цели [6]. В документах по регулированию разработки и использованию ИИ подобные ошибки иногда определяются как предвзятость или необъективность («bias»).

Примеров второго варианта ошибок, а именно – случайностей, существенно больше, как в транспортной сфере, так и в области роботизированных вооружений [7,8], но их последствия, как правило, менее опасны.

Ошибки первой группы являются критичными, так как напрямую затрагивают безопасность человека, они требуют надёжного решения, поиском которого научное сообщество в настоящее время занимается достаточно активно. Формализованную постановку по решению этих задач принято определять как создание «доверенного» искусственного интеллекта. Некоторые результаты исследований по созданию безопасного или, «доверенного» («trustworthy») искусственного интеллекта отражены в документе Еврокомиссии «Руководство по этике для надежного ИИ» 2019 года («Ethics guidelines for trustworthy AI, 2019») и в российском стандарте ГОСТ Р 59 276–2020 «Системы искусственного интеллекта. Способы обеспечения доверия. Общие положения». В данных документах искусственный интеллект объявляется доверенным, если он обладает следующими свойствами: проверяемость, управляемость, стабильность,

робастность, безопасность и отказоустойчивость. При выполнении указанных требований ИИ предлагается считать «верифицированным по этическим характеристикам».

Впрочем, данными документами не решается целый ряд проблем в устранении ошибок применения ИИ, некоторые из которых являются критичными.

Во-первых, в них в качестве одного из критерия «доверенности» задаётся непрерывный и безусловный контроль над поведением искусственного интеллекта человеком, существенную часть требований к доверенному ИИ, приводимых в рассмотренных документах, можно считать реализацией человекоцентричного подхода к его разработке и внедрению, опирающемуся на три группы принципов: прозрачность, надежность, человекоцентричность.

В дополнение к сказанному можно отметить, что современными документами в сфере регулирования ИИ признаётся, что решением большинства проблем является отказ от рассмотрения ИИ как «черного ящика» [9,10], что должно обеспечиваться:

- тщательным контролем системы на всех этапах жизненного цикла;

- разработкой механизмов регулирования ИИ;

- разработкой кодексов и положений по использованию ИИ.

Но, строгое соблюдение этих принципов проблематично для автономных робототехнических систем: такой контроль или сильно ограничивает их возможности, или затруднён чисто технически.

Во-вторых, рассуждать в аспекте, определяемом данными документами, об этике ИИ не совсем логично. Доверенность ИИ по содержанию указанных документов скорее соответствует безопасности их алгоритмов для человека, а не набору мер по предотвращению случайных ошибок поведения, которую как раз и можно отнести к «этичности». Ибо, в робототехнике, понятию «этичности», скорее, соответствует приоритетный учёт интересов взаимодействующих систем, в том числе РТС, даже если для этого приходится жертвовать оптимальностью собственного поведения.

Поэтому «доверенность» ИИ можно определить, всё же, как аналог человеческой морали, интерпретированной для правил поведения РТС, потенциально решаемых средствами контроля безопасности алгоритмов программного обеспечения (ПО) РТС и формированием системы ограничений их поведения [11]. Варианты таких подходов описаны в работах [12,13]. В любом случае, так или иначе, проблема безопасного поведения постепенно решается, хотя и не в полном объёме и с определёнными терминологическими недочётами.

Сложнее дело обстоит с предупреждением возникновения ошибок второго типа. Они менее критичны с точки зрения безопасности, менее очевидны для анализа. Для их решения нужна разработка более гибких правил построения управляющих алгоритмов и системы ограничений, прототипом для которых, с определённой степенью допущений, можно считать правила не морали, а человеческой вежливости, интерпретированные для «доверенного» ИИ.

Возникающие противоречия в терминологии, при кажущейся незначительности, затрудняют разработку унифицированных правил поведения РТС и алгоритмов для их реализации. Такая ситуация требует решения, обеспечивающего логичное развитие теории и практики алгоритмизации ИИ, управляющего РТС.

С учётом выявленного противоречия в терминологии, основываясь на методе аналогий и проецируя некоторые правила человеческого поведения на алгоритмы и ограничения поведения искусственного интеллекта, предлагается сформулировать задачу реализации набора правил «морали» и «вежливости» для различных аспектов безопасности применения РТС разной степени автономности, обеспечивающих уточнение механизмов регулирования и кодексов поведения ИИ. В статье, учитывая меньшую изученность проблемы случайных ошибок поведения ИИ, рассмотрен вариант реализации правил для предотвращения возникновения именно случайных ошибок поведения, возникающих из-за неопределённости исходных данных, в том числе в связи с недостаточным временем для принятия решений.

3 Уточнение понятия «доверенность» в части безопасности взаимного поведения

Типичным примером случайной ошибки является возникновение риска столкновение наземных автономных РТС в процессе совместного маневрирования при ошибках учёта взаимного перемещения.

Модель возникновения подобной ситуации можно рассмотреть на простом примере. Допустим, два автономных транспортных средства, движущихся с одинаковой скоростью должны совершить поворот с огибанием препятствия. Средства маневрируют независимо, не работая в группе. ПО каждого из этих средств строит оптимальную для себя траекторию, с высокой вероятностью эти траектории пересекутся в точке поворота. Если указанные средства оснащены системами безопасного расхождения, столкновение маловероятно – данные системы определят опасность и ПО РТС сформируют маневр аварийного расхождения или экстренное изменение скорости. Но, проблема заключается именно в том, что расхождение будет критичным, формировать варианты расхождения, изменять скорость или траекторию, придётся в очень короткий промежуток времени, об оптимальности маршрута речи уже идти не будет. В таких условиях, и вероятность успешного расхождения не будет стопроцентной.

Это типичный случай несогласованного взаимодействия движущихся агентов.

В такой ситуации, если бы ИИ в составе ПО маневрирующих РТС заранее просчитал ситуацию, аварийного расхождения можно было бы избежать, заблаговременно уточнив параметры движения с учётом потенциальной проблемы. Незначительно отклонившись от оптимальной траектории на раннем этапе планирования, можно было бы избежать существенных потерь оптимальности или риска возникновения столкновения при расхождении.

Ситуация описана в несколько упрощённом виде, в реальности фоновая обстановка может быть сложнее, но для пояснения сущности проблемы, такой подход допустим.

В описанной постановке, для решения задачи по определению параметров маневра, могут быть использованы несколько вариантов, рассчитываемых заблаговременно с применением общеизвестного математического аппарата:

- спланировать взаимное расхождение, изменив траекторию и/или скорость обеим РТС;

- спланировать расхождение одной из систем, которая имеет меньше внешних ограничений для маневра.

Алгоритм «вежливости» или «этики» в вышеописанной ситуации, математически может быть реализован как расчёты по типовым вариантам расхождения, проведённые

относительно РТС, корректирующей траекторию [14,15].

Технически, наиболее простым способом расхождения является снижение скорости одного из участников на $\Delta V = V_{нач} - V_{тр}$ от начальной до требуемой, обеспечивающей расхождение на расстоянии ΔS.

Требуемое снижение скорости легко определяется из уравнения движения материальной точки, приняв её движение за прямолинейное и равноускоренное и используя в виде исходных данных начальную скорость движения $V_{нач}$ и расстояние до точки поворота (расхождения) S:

$$V_{тр} = V_{нач}\left(1 - \frac{\Delta S}{S}\right).$$

Кроме изменения скорости, можно обеспечить расхождение за счёт изменения траектории движения, смещения точки поворота, обеспечивая расхождение сзади или спереди на минимальном безопасном расстоянии.

Расчёты по выносу точки поворота в сторону от пересекаемой траектории могут быть описаны известными уравнениями выхода в заданную точку пространства, удалённую от траектории другого маневрирующего не менее, чем на расстояние безопасного расхождения $L_{мин}$. Расчётный курсовой угол маневрирования q_M в такой ситуации определяется как (рисунок 1):

$$qM = q'M0 - a;$$

$$q'_{мр} = arcsin\left(\frac{sinq'_{To}}{m}\right).$$

где $m = \frac{V_M}{V_T}$,

V_M – скорость маневрирующего;

V_T – скорость РТС, не меняющей траекторию.

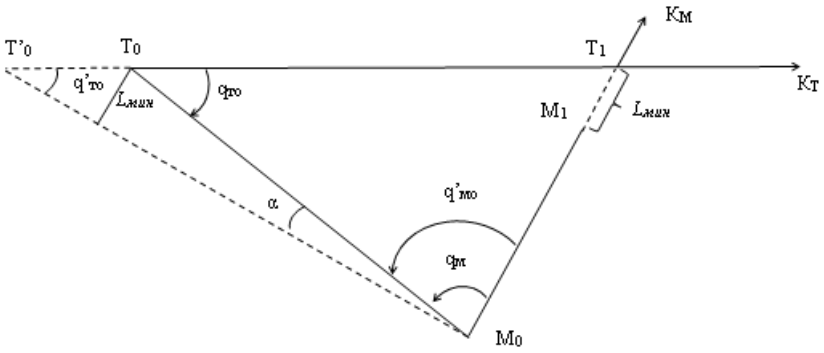

Рис. 1. Схема расчёта расхождения на заданном расстоянии

На рисунке:

K_M – курс маневрирующего;

K_T – курс РТС, не меняющей направления;

q_{TO} – начальный курсовой угол РТС, не меняющей направления движения;

$q_{мо}$ – начальный курсовой угол маневрирующего;

$q_м$ – расчётный курсовой угол маневрирующего;

T_0 *и* M_0 – точки начального положения РТС;

T'_0 – точка расхождения.

Расчёты по расхождению на минимальном расстоянии позади или по курсу на пересекаемой траектории, могут быть реализованы через уравнения расхождения на максимальной дистанции по курсу или на минимальной дистанции позади $L_{мин}$:

$$L_{мин} = L_м \frac{\sin(q_{тр} - Q)}{\sin Q},$$

где Q – критический курсовой угол, предельный, на котором может быть осуществлено расхождение.

Значение Q рассчитывается по соотношению векторов скоростей:

$$Q = \frac{V_м}{V_N}.$$

Таким образом, существует достаточно широкий выбор математического аппарата, позволяющего реализовать «вежливое» расхождение в разных ситуациях.

Выше описаны только варианты расчётов для решения проблемы, реализующие частные случаи. В целом, такая «вежливость» должна реализовываться расчётно-аналитическим алгоритмом, просчитываемым ПО РТС перед каждым маневром:

1) оценка окружающей обстановки, выделение потенциально взаимодействующих РТС и биологических объектов;

2) анализ координат и параметров движения потенциально взаимодействующих объектов;

3) построение оптимальной траектории собственного движения;

4) прогноз возможных пересечений;

5) принятие решение о необходимости коррекции траектории или скорости;

6) выбор способа коррекции;

7) расчёт уточнённых параметров движения.

Математически, как показывают приведённые примеры, данный алгоритм может быть обеспечен существующими методами, проблемным вопросом остаётся только формирование правил реализации «вежливого» поведения, а именно – кто должен совершать маневр уклонения. Эта задача является более сложной, лежащей на границе математики и логики.

Конечно, можно возразить, что подобные задачи уже решаются разработкой и применением наборов правил и нормативов, реализованных в специализированных документах: правилах дорожного движения, воздушных и водных кодексах. Но это предположение не совсем верно. Во-первых, движение и маневрирование, частные случаи применения РТС, реально спектр их применения намного шире. Во-вторых, эти

частные случаи реализуются в условиях существенных ограничений, определяемых, в частности, границами дорожного покрытия, разметкой, заранее задаваемыми трассами и эшелонами. В таких условиях может быть сформирован достаточно простой набор правил, таких как «расхождение левыми бортами», «помеха сверху» и тому подобное.

Намного более сложные и менее формализованные ситуации возникают во всех остальных случаях, характерных для функционирования РТС разной степени автономности. Более того, кроме слабоформализованных условий принятия решения, эти ситуации, как правило, осложняются ограниченным временем на его принятие.

В таких условиях, требуется разработка и реализация динамичных алгоритмов решения задач расхождения со слабоформализованными исходными данными.

Таким образом, при разработке алгоритмов для программного обеспечения РТС возникает нетривиальная научная задача – разработка правил взаимного поведения, аналога человеческого этикета. При этом, программный «этикет», как и обычный, может разделяться на ситуационный и профессиональный, а последний подразделяться на сферы применения с соответствующими вариациями действий. И, кстати, поведения при маневрировании, в предлагаемой постановке, является частью профессионального, а именно транспортного, «этикета». А всё, что связано с заведомо опасным для окружающих применением, должно решаться в рамках системы ограничений и запретов, аналогичных человеческой морали.

При таком подходе, ограничения и, соответственно, содержание понятия «доверенного» ИИ, будут строиться на двух наборах правил: аналогов «этикета» и «морали», с выбором вида профессионального этикета, подходящего для каждой из рассматриваемых ситуаций (рисунок 2).

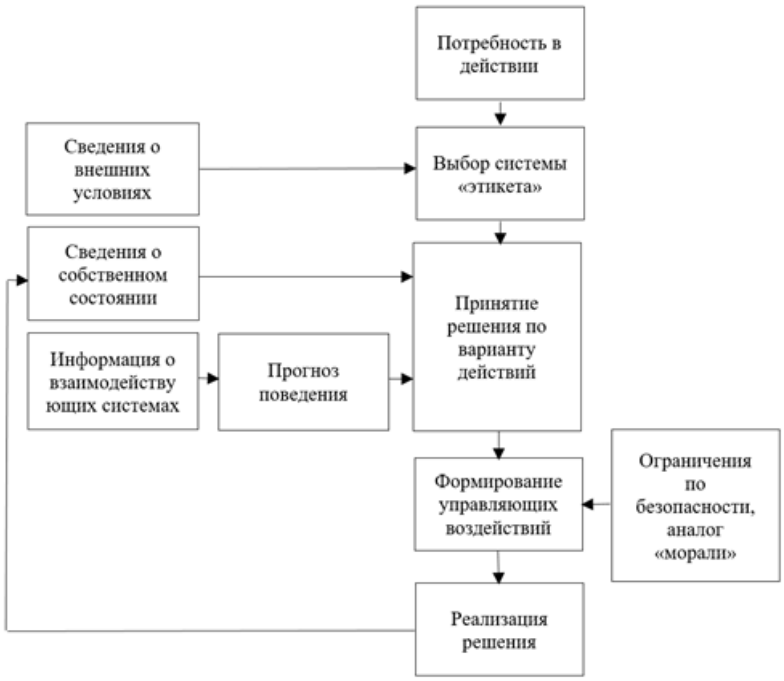

Рис. 2. Схема формирования взаимного поведения

Приведённый на рисунке алгоритм сформирован исходя из отсутствия обмена информацией между объектами. В то же время, развитие информационных технологий теоретически позволяет организовать подобный обмен и тогда алгоритм несколько упростится, так как:

- станет меньше неопределённостей в поведении других участников движения;

- появится возможность «договариваться» с другими участниками о параметрах взаимного расхождения.

То есть ситуация будет сведена к более простому варианту согласованного взаимодействия.

Но, в любом случае, алгоритм взаимного поведения, основанный на учёте параметров движения всех его участников с намеренным отступлением части из них от оптимальных параметров, то есть прообразом «вежливости» поведения, объективно необходим и требует дальнейшей разработки. Как требуется продолжение разработки ранее упомянутых алгоритмов безопасности в рамках правил поведения РТС, выстроенных по аналогии с «моралью».

Таким образом, использование аналогий «морали» и «вежливости» РТС позволит решить ряд важных проблем безопасности поведения автономных роботов, особенно в слабоформализуемых условиях взаимодействия, а основой такого решения должно служить как раз предлагаемое уточнение предметной области.

4 Заключение

Предложенная интерпретация понятий «морали» и «вежливости» для обучаемых алгоритмов программного обеспечения, управляющего автономными и частично автономными робототехническими системами, является лишь одним из вариантов решения проблемы, позволяющим выстроить систему их безопасности, основанную на группах правил и динамически формируемых ограничений.

Для обеспечения решения задач комплексной безопасности РТС, управляемых ИИ, требуется уточнить понятийный аппарат, используемый в данной предметной области. Причём речь идёт не о прямом копировании человеческих понятий морали и вежливости, а об уточнении понятия «доверенный» ИИ.

Указанная задача может быть решена двумя вариантами:

- оставить содержание определения «доверенного» ИИ как синоним ИИ с гарантированно безопасными алгоритмами поведения, добавив определение «вежливого» ИИ, обеспечивающего безопасное поведение в условиях группового применения на основе использования «этичных» алгоритмов;

- уточнить определение «доверенного» ИИ, добавив свойство «этичности» поведения, заключающегося в прогнозировании результатов предполагаемых действий и их поверки на основе набора правил.

В любом случае, предлагаемое уточнение классификации послужит предпосылкой к решению важной научно-практической задачи обеспечения комплексной безопасности РТС, управляемых ИИ. Более того, принятие предлагаемых изменений в понятие «доверенности», в перспективе, должно обеспечить переход от подхода на основе «обучения» ИИ к расширенному варианту, добавив к нему «воспитание», что также послужит повышению безопасности ИИ и управляемых им робототехнических систем.

Библиография

1. Симулин А. А. и др. Некоторые аспекты использования робототехники в военном деле // Сборники конференций НИЦ Социосфера. 2015. № 27. С. 67-71.

2. 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.
3. Хрипунов С.П., Благодарящев И.В., Чиров Д.С. Военная робототехника: современные тренды и векторы развития // Тренды и управление. 2015. № 4. С. 410-422.
4. Pflimlin É. Drones et robots: La guerre des futurs. France: Levallois-Perret, 2017.
5. Roosevelt, Ann. Ar my 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. Influence of the Problem of Safety Control of Heuristic Algorithms on the Development of Robotics. In: Shamtsyan, M., Pasetti, M., Beskopylny, A. (eds) Robotics, Machinery and Engineering Technology for Precision Agriculture. Smart Innovation, Systems and Technologies 2022, No. 247. Singapore: Springer. https://doi.org/10.1007/978-981-16-3844-2_31.
8. Beard J Autonomous weapons and human responsibilities, Georgetown Journal of International Law 2014. No. 45, P. 617-681.
9. Tikhanychev O The Control System of Heuristic Algorithms as a Prototype of the "Morality" of Autonomous Robots. II International Scientific Forum on Sustainable Development and Innovation. WFSDI-2023. URL: https://doi.org/10.1007/978-981-16-3844-2_31.
10. Ćwiąkała P. Testing Procedure of Unmanned Aerial Vehicles (UAVs) Trajectory in Automatic Missions. Appl. Sci. 2019, No. 9, P/ 3488. URL: https://doi.org/10.3390/app9173488.
11. Johnson D. Computer Systems: Moral entities but not moral agents. In: Ethics and Information Technology. 2016, No. 8, P. 195-204. URL: https://doi.org/10.1007/s10676-006-9111.
12. 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, P. 379-425.
13. Ухоботов В.И., Изместьев И.В. Об одной задаче преследования при наличии сопротивления среды // Вестник ЮУрГУ. Серия: Математика. Механика. Физика. 2016. № 2. С. 62-66. URL: https://doi.org/10.14529/mmph160208.
14. Дубанов А.А. Моделирование траектории преследователя в пространстве при методе параллельного сближения // Программные системы и вычислительные методы. 2021. № 2. С. 1–10. URL: https://doi.org/10.7256/2454-0714.2021.2.36014
15. Tikhanychev O. Self-Check System of Heuristic Algorithms as a "New Moral" of Intelligent Systems AIP Conference Proceedings. 2023, No. 2700, 040028. URL: https://doi.org/10.1063/5.0124956.

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 уточнению понятийного аппарата, используемого в интеллектуальных информационных системах, в частности понятия «доверенности» систем искусственного интеллекта.
Методология исследования базируется на обобщении публикаций зарубежных и отечественных ученых по рассматриваемой в статье проблематике, использовании

метода аналогий и математического моделирования возникновения риска столкновение наземных автономных робототехнических систем в процессе совместного маневрирования при ошибках учёта взаимного перемещения.

Актуальность работы авторы связывают с тем, что одним из приоритетов, обеспечивающим активное развитие автономной робототехники, является использование элементов искусственного интеллекта (ИИ), наличием нерешенной проблемы в части определения безопасности систем ИИ.

Научная новизна работы, по мнению рецензента состоит в предложенной интерпретации понятий «морали» и «вежливости» для обучаемых алгоритмов программного обеспечения, управляющего автономными и частично автономными робототехническими системами.

Структурно в статье выделены следующие разделы: Введение, О существующем подходе к определению «доверенных» систем, Уточнение понятия «доверенность» в части безопасности взаимного поведения, Заключение, Библиография.

Авторы указывают, что используемый в настоящее время термин «доверенность» не в полной мере тождественен понятию «безопасность» в комплексном понимании этого явления, а других подходов к описанию данного фактора пока не предлагается. В статье приведен обзор перечня и условий возникновения катастроф и аварий, произошедших по вине автономных и частично автономных робототехнических систем, изложен алгоритм «вежливости» или «этики» на математическом языке в виде расчётов по типовым вариантам расхождения, проведённым относительно робототехнических систем, корректирующей траекторию наземных автономных робототехнических систем в процессе совместного маневрирования. Текст публикации сопровождается пятью формулами и иллюстрирован двумя рисунками: «Схема расчёта расхождения на заданном расстоянии», «Схема формирования взаимного поведения». В Заключении указаны два варианта решения задачи уточнения понятия «доверенный» искусственный интеллект: во-первых, оставить содержание определения «доверенного» ИИ как синоним ИИ с гарантированно безопасными алгоритмами поведения, добавив определение «вежливого» ИИ, обеспечивающего безопасное поведение в условиях группового применения на основе использования «этичных» алгоритмов; во-вторых, уточнить определение «доверенного» ИИ, добавив свойство «этичности» поведения, заключающегося в прогнозировании результатов предполагаемых действий и их поверки на основе набора правил.

Библиографический список включает 15 источников – научные публикации по рассматриваемой теме на английском и русском языках. В тексте публикации имеются адресные отсылки к списку литературы, подтверждающие наличие апелляции к оппонентам.

Из резервов улучшения статьи следует отметить необходимость нумерации формул и оформления и в соответствии с принятыми правилами. С расшифровкой использованных символов непосредственно после математического выражения для улучшения и облегчения восприятия материала читателями.

Тема статьи актуальна, материал отражает результаты проведенного авторами исследования, содержит элементы приращения научного знания, соответствует тематике журнала «Программные системы и вычислительные методы», может вызвать интерес у читателей, рекомендуется к публикации после дооформления формул.