

**Европейская экономическая комиссия****Комитет по внутреннему транспорту****Всемирный форум для согласования правил
в области транспортных средств**

Сто девяносто вторая сессия

Женева, 5–8 марта 2024 года

Пункт 2.3 предварительной повестки дня

Координация и организация работы:**Интеллектуальные транспортные системы****и координация деятельности, связанной****с автоматизированными транспортными средствами****Предложение по проекту резолюции с руководством
по искусственному интеллекту в контексте дорожных
транспортных средств****Записка секретариата***

Воспроизведенный ниже текст основан на неофициальном документе WP.29-191-06, который был рассмотрен Всемирным форумом для согласования правил в области транспортных средств (WP.29) на его сессии в ноябре 2023 года. В этом документе предложен проект резолюции с руководством по искусственному интеллекту (ИИ) в контексте дорожных транспортных средств. Он включает определения, касающиеся конкретных особенностей систем на основе ИИ, используемых в автомобильной продукции, обзор примеров использования в автомобилях, представленных промышленностью, и обзор взаимодействия принципов, связанных с ИИ, с основой нового метода оценки/испытаний для автоматизированного вождения. Он подготовлен по просьбе WP.29, изложенной на его сессии в ноябре 2023 года (см. ECE/TRANS/WP.29/1175, пункт 25). Он передан WP.29 для рассмотрения и возможного принятия. WP.29 решил, что в этот документ могут быть внесены поправки на основе неофициального документа, подготовленного GRVA на ее сессии в январе 2024 года.

* В соответствии с программой работы Комитета по внутреннему транспорту на 2024 год, изложенной в предлагаемом бюджете по программам на 2024 год (A/78/6 (разд. 20), таблица 20.5), Всемирный форум будет разрабатывать, согласовывать и обновлять правила ООН в целях улучшения характеристик транспортных средств. Настоящий документ представлен в соответствии с этим мандатом.



Преамбула

Договаривающиеся стороны соглашений 1958 и 1998 годов, участвующие в деятельности Рабочей группы по автоматизированным/автономным и подключенным транспортным средствам,

признав значительную степень проникновения некоторых видов искусственного интеллекта (ИИ) в колесные транспортные средства, подпадающие под действие соглашений, относящихся к ведению Всемирного форума для согласования правил в области транспортных средств (WP.29),

приняв во внимание, что в настоящее время отрасль может использовать инструменты машинного обучения для поддержки усилий по разработке и/или тестированию программного обеспечения перед развертыванием,

обсудив основополагающие технические аспекты некоторых систем машинного обучения, используемых в автомобильной продукции, которые широкая публика называет искусственным интеллектом, и рассмотрев соответствующие определения,

напоминая о принятии рекомендаций по единообразным предписаниям, касающимся кибербезопасности и обновления программного обеспечения,

проведя оценку важности надлежащих жизненных циклов ИИ для обеспечения совместимости с существующими режимами сертификации,

признав, что использование таких технологий в автомобильной промышленности все еще находится в стадии разработки,

согласовали нижеследующие рекомендации по использованию алгоритмов на базе ИИ в своей автомобильной продукции:

Обновление программного обеспечения

1. Настоящее руководство распространяется на сертификационные требования и требования о соответствии производства. Отрасль не должна выпускать обновления программного обеспечения, которые будут существенно изменять уже сертифицированные функции, в соответствии с рекомендациями по единообразным предписаниям, касающимся кибербезопасности и обновления программного обеспечения, без возобновления соответствующей процедуры сертификации.

2. По завершении обучения системы на базе ИИ, заложенной в программное обеспечение, рекомендуется производить ее валидацию уполномоченными сторонами и/или при помощи процессов сертификации и оценку с точки зрения безопасности, защиты и экологических характеристик, а также других соответствующих требований. Несертифицированные системы, включающие ИИ, не должны влиять на сертифицированные системы таким образом, чтобы это вредило сертификации. По завершении этого процесса валидированное программное обеспечение может внедряться на транспортных средствах того или иного типа.

Данные, подлежащие использованию для разработки системы на базе ИИ

3. Предполагается полное соблюдение правил защиты данных и конфиденциальности, а также других юридических требований. Настоящий документ не наносит ущерба действующим на конкретном рынке законодательным и нормативным актам, касающимся порядка сбора и использования персональных данных. Там, где такие нормы существуют, они способствуют повышению общей безопасности системы ИИ, устанавливая стандарты безопасности управления персональными данными.

Приложение 1

Упрощенные определения в контексте правил в области транспортных средств — Конкретные особенности систем на базе ИИ в автомобильной продукции

1. Далее под термином «системы на базе ИИ» подразумеваются такие коннекционистские системы ИИ, как нейронные сети, которые обучаются с помощью алгоритмов и данных в рамках машинного обучения. Эти системы демонстрируют качественно новые свойства, открывающие новые возможности и обуславливающие новые вызовы.
2. Системы на базе ИИ, используемые в автомобильной продукции, позволяют найти компромисс между различными желаемыми характеристиками модели; речь идет о треке и ветхости модели, ее сложности, надежности проверяемости предсказуемости, переобучении и т. д. и обеспечении при этом определенного уровня защиты и безопасности. Системы на базе ИИ должны обеспечивать возможность обновления системы.
3. Вероятно, потребуются дополнительные периодические оценки для выяснения того, насколько адекватно положения об обновлениях программного обеспечения (в рекомендациях по единообразным предписаниям, касающимся кибербезопасности и обновления программного обеспечения) позволяют учитывать обновления систем на базе ИИ.
4. Системы на базе ИИ могут содействовать повышению надежности транспортных средств, оказывая дополнительное благоприятное воздействие на безопасность дорожного движения, например позволяя системам АВ предсказывать непредсказуемые в настоящее время виды поведения других участников дорожного движения (речь идет, к примеру о потенциальном предотвращении столкновений).
5. Использование ИИ и алгоритмов машинного обучения в функциях официального утверждения типа пока ограничено. И, хотя уже существуют оптимально отлаженные процессы тестирования обычного программного обеспечения до и в ходе внедрения автомобильной продукции, эти процессы могут оказаться недостаточными для программного обеспечения на базе ИИ. Программное обеспечение — независимо от того, создано ли оно с помощью машинного обучения или нет — должно быть протестировано перед внедрением, с тем чтобы оно соответствовало всем надлежащим законам, нормам, рекомендациям и стратегиям. Это относится также к процессу обновления. Вместе с тем необходимо оценить, насколько эффективно существующие нормативные положения могут удовлетворять конкретные потребности в тестировании и обновлении программного обеспечения на базе ИИ и гарантировать его надежное функционирование.
6. Приведенные ниже термины в значительной мере позаимствованы из определений, находящихся на рассмотрении Международной организации по стандартизации (см. ISO/IEC 22989), Комитета по вопросам использования ИИ в наземных транспортных средствах ОИАТ и Организации экономического сотрудничества и развития (ОЭСР). Список терминов не является исчерпывающим, представленные определения упрощены и поэтому могут не подходить для целей регулирования.
7. Обычно для обеспечения надежного функционирования сертифицированной системы проводится всестороннее тестирование программ «прозрачного»/«серого»/«черного» ящика.

Примечание: некоторые определения, приведенные ниже, были позаимствованы в (либо разработаны под влиянием) ОЭСР¹, ОИАТ² или ИСО³.

- **Агент** — это сущность, которая воспринимает свое окружение, автономно предпринимает действия для достижения целей и может улучшать свою производительность благодаря обучению либо использовать полученные знания.
- **Жизненный цикл ИИ** состоит из этапа проектирования и разработки системы на базе ИИ, включая, в частности, сбор, отбор и обработку данных, а также выбор модели и процесса обучения; этапа валидации; этапа внедрения; и этапа мониторинга. Жизненный цикл завершается, когда система на базе ИИ перестает функционировать.
- **Искусственный интеллект (ИИ)³** — это совокупность методов или автоматизированных сущностей, которые сообща строят, оптимизируют и применяют модель таким образом, чтобы система могла — исходя из заданного набора предварительно определенных задач — вычислять прогнозы, рекомендации или решения.
- **Система ИИ¹** — это машинная система, которая способна воздействовать на окружающую среду, выдавая выходные данные (прогнозы, рекомендации или решения) для заданного набора целей. Она использует машинное обучение и/или данные, полученные от человека для восприятия реальной и/или виртуальной среды; абстрагирования этих восприятий в контексте моделей посредством автоматизированного анализа (например, с помощью машинного обучения) либо вручную; и применения типовых заключений с целью формулирования вариантов результатов. Системы ИИ предназначены для работы с различными уровнями автономности.
- **Предвзятость** — это систематическое различие в отношении (включая категоризацию/наблюдение) к определенным объектам (например, людям или группам) по сравнению с другими.
- **«Черный» ящик** — это система/программа, в случае которой детальная архитектура и подробности обработки неизвестны.
- **Тестирование методом «черного»/«серого»/«прозрачного» ящика** — это способы тестирования системы/программы, в случае которой детальная архитектура и подробности обработки неизвестны/частично известны/известны.
- **Коннекционистские ИИ (кИИ)** системы обычно состоят из множества узлов, называемых нейронами, которые связаны друг с другом по определенным схемам, зависящим от конкретной модели ИИ. Примерами кИИ-систем являются нейронные сети и машины опорных векторов. Во многих прикладных сферах кИИ-системы являются более мощными по сравнению с сИИ-системами, например в компьютерном зрении. В большинстве случаев параметры кИИ-систем не могут быть заданы непосредственно разработчиком. Вместо этого для обучения таких систем используются алгоритмы машинного обучения в сочетании с соответствующими данными. Качество результирующей кИИ-системы особо зависит от качества и количества обучающих данных. В отличие от сИИ-систем кИИ-системы в большинстве случаев не относятся к числу легко интерпретируемых и не являются формально верифицируемыми.
- **Коннекционистское программное обеспечение** обычно создается с помощью процесса, называемого традиционным программированием. Программист вручную кодирует правила с использованием языка программирования.
- **Аннотирование данных** — это процесс присоединения к данным набора описательной информации без какого-либо изменения этих данных.
- **Выборка данных** — это статистический процесс отбора подмножества данных, в котором должны быть представлены закономерности и тенденции, аналогичные тем, которые наблюдаются в более крупном анализируемом наборе данных.

- **Набор данных** — это совокупность данных, имеющих общий формат и содержание, соответствующее конкретной цели.
- **Глубокое обучение** — это процесс, при котором нейронные сети используют большое количество слоев обработки для извлечения из данных все более высокоуровневых характеристик.
- **Объяснимость** означает свойство системы на базе ИИ отражать важные показатели, влияющие на результаты работы системы, в понятном для человека виде.
- **Беспристрастность/Матрица непристрастности** — это способ описания предвзятости.
- **«Серый» ящик** — это система/программа, в случае которой детальная архитектура и подробности обработки известны частично.
- **Человеческий надзор** — это заложенное в систему на базе ИИ свойство, призванное гарантировать, что встроенные операционные ограничения не могут быть отменены самой системой и реагируют на действия человека-оператора и что окончательный контроль осуществляют физические лица, на которые возложена функция человеческого надзора.
- **Машинное обучение** — это совокупность основанных на данных вычислительных методов, обеспечивающих способность системы к обучению без явных инструкций, с тем чтобы поведение модели отражало закономерности в данных или опыте.
- **Модель машинного обучения** — это компьютерная математическая модель, которая генерирует вывод или предсказание на основе входных данных.
- **Модель** — это физическое, математическое или иное логическое представление системы, объекта, явления, процесса или данных.
- **Дрейф модели** — это термин из области машинного обучения. Он относится к тому явлению, что с течением времени точность прогнозирования моделей машинного обучения может ухудшаться. Это может быть обусловлено, например, тем, что допущения или зависимости переменных, которые были актуальны на момент создания и обучения моделей, со временем изменились. Устранить дрейф модели позволяют такие меры, как переобучение или настройка моделей.
- **Ветхость модели** определяется как устаревание, если обученная модель не содержит актуальных данных и/или не отвечает современным требованиям. Устаревшие модели могут негативно сказываться на качестве обеспечиваемого интеллектуальным программным обеспечением прогнозирования.
- **Онлайновое обучение** означает постепенное обучение новой версии системы на базе ИИ в процессе эксплуатации на борту серийных транспортных средств для достижения поставленных целей. Те части системы, которые подвергаются онлайновому обучению, должны быть отделены от системы исполнительного привода транспортного средства, с тем чтобы новая версия системы на базе ИИ не оказывала воздействия на эти системы до прохождения испытания на соответствие надлежащим правилам безопасности.
- **Предсказуемость** — это свойство системы на базе ИИ, которое позволяет заинтересованным сторонам делать достоверные предположения о результатах работы.
- **Обучение с подкреплением** — это способ машинного обучения, позволяющий агенту обучаться действиям на основе закономерностей в данных или опыте, оптимизируя количественную функцию накопленного за это время вознаграждения.
- **Надежность** — это свойство соответствия ожидаемому поведению и результатам.

- **Отказоустойчивость** — это способность системы к быстрому восстановлению работоспособности после инцидента.
- **Робастность** — это способность системы сохранять свой уровень производительности при самых различных обстоятельствах, включая ее способность справляться с естественными и злонамеренными активностями в пространстве входных данных систем.
- **Безопасность по умолчанию** — это свойство системы, обеспечиваемое проактивной разработкой и упреждающей деятельностью на протяжении жизненного цикла, позволяющее свести риски к приемлемому уровню за счет системных мер.
- **Обучение с частичным привлечением учителя** — это сочетание способов обучения с учителем и без учителя. При нем используется небольшой объем размеченных данных и большой объем неразмеченных данных, что обеспечивает преимущества обоих способов обучения, позволяя при этом избегать проблем, связанных с поиском большого объема размеченных данных.
- **Обучение с учителем** — это способ машинного обучения, при котором во время обучения используются размеченные данные.
- **Символьный ИИ (сИИ)** связан с эксплицитным кодированием знаний с использованием символов. Примером такой системы является дерево решений. Интерпретация и формальная верификация сИИ-систем в целом возможны и значительно проще по сравнению с коннекционистскими системами ИИ.
- **Обучение** — это процесс настройки параметров модели машинного обучения.
- **Обучающие данные** — это подмножество образцов входных данных, используемых для настройки модели машинного обучения.
- **Транспарентность организации** — это свойство организации документировать ее деятельность и решения и доносить их до соответствующих заинтересованных сторон в полной, доступной, ясной и понятной форме.
- **Транспарентность системы** — это свойство системы доносить информацию до заинтересованных сторон.
- **Достоверность** — это способность оправдывать ожидания заинтересованных сторон проверяемым способом.
- **Обучение без учителя** — это способ машинного обучения, при котором во время обучения используются неразмеченные данные.
- **Валидация** производится для того, чтобы убедиться в практичности использования программного обеспечения и его способности отвечать потребностям клиента.
- **Результаты валидации** — это данные, служащие для оценки эффективности конечной модели машинного обучения.
- **Верификация** производится для того, чтобы убедиться в высоком качестве программного обеспечения, его отлаженности, надежности и отсутствии ошибок, не вдаваясь при этом в практичность его использования.
- **«Прозрачный» ящик** — это система/программа, в случае которой детальная архитектура и подробности обработки известны.

Приложение 2

Обзор представленных отраслью вариантов использования ИИ в транспортных средствах (в 2022 году)

Тип ИИ	Тип машинного обучения	Функции, не связанные с безопасностью Не подпадают под действие официального утверждения типа	Функции обеспечения безопасности			
			Функция вождения			Функции, не связанные с вождением
			Восприятие	Планирование	Активация	
Обычное программное обеспечение	Отсутствует		Не подпадает под действие (в контексте настоящего документа)	Не подпадает под действие (в контексте настоящего документа)	Не подпадает под действие (в контексте настоящего документа)	Не подпадает под действие (в контексте настоящего документа)
Символьный ИИ	Отсутствует или любой тип МО	Например, информационно-развлекательная, обработка естественно-языковых текстов	Например, обнаружение присутствия других участников дорожного движения для работы систем САЭТ, АСС Обнаружение объектов дорожной инфраструктуры для работы систем СПВП, СУПП	Например, активация систем СПФС и САЭТ с учетом положения испытуемого транспортного средства и присутствия других участников дорожного движения	В настоящее время неприменимо	Например, распознавание лица водителя для идентификации личности (на условиях обеспечения конфиденциальности), иммобилизатор, реагирующий на содержание алкоголя в выдыхаемом воздухе
Коннекционистский ИИ + Машинное обучение	Обучение с учителем (ОСУ)	Считывание жестикуляции Распознавание голоса	Обнаружение присутствия других участников дорожного движения для работы систем САЭТ, АСС Обнаружение объектов пассивной дорожной инфраструктуры для работы систем СПВП, СУПП	Прогнозирование траектории движения посредством прогнозирования пути перемещения по размеченным данным (например, картам высокого разрешения)	В настоящее время неприменимо	Распознавание направление взгляда/поворота головы водителя для работы системы управления данными. Выявление неисправностей, диагностическое обслуживание
	Обучение без учителя (ОБУ)		Оптимизация процесса разметки данных для таких менее критических с точки зрения безопасности систем, как ИСА Извлечение сценариев из реальных данных в обоснование валидации Генерирование синтетических данных для обучения с учителем/искажения реальных данных	Прогнозирование траектории движения с использованием фильтров Калмана, архитектур на основе KalmanNet или гауссовского процесса либо других архитектур	В настоящее время неприменимо	выявление неисправностей (выявление аномалий без учителя)

Тип ИИ	Тип машинного обучения	Функции, не связанные с безопасностью Не подпадают под действие официального утверждения типа	Функции обеспечения безопасности			
			Функция вождения			Функции, не связанные с вождением
			Восприятие	Планирование	Активация	
	Обучение с частичным привлечением учителя (ОЧУ)		Оптимизация процесса разметки данных для таких менее критических с точки зрения безопасности систем, как ИСА	При разработке алгоритмов управления обучением используется режим «тени»	В настоящее время неприменимо	
	Обучение с подкреплением (ОП)		Некоторые изготовители начинают использовать ОП для функции восприятия, а в будущем оно вполне может использоваться для целей кооперативного восприятия	ОП может использоваться в случае систем удержания по центру полосы движения или АСС ввиду снижения стоимости/доступности данных, необходимых для обучения системы	В настоящее время неприменимо	Диагностическое обслуживание

Примечание: В заштрихованных графах указаны пункты, которые не попадают под действие настоящего документа].

Приложение 3

Воздействие искусственного интеллекта на новые методы оценочных испытаний

Для всеобъемлющего схематичного обзора взаимодействия отдельных компонентов, сценариев и требований относительно безопасности таких весьма сложных систем, как системы, используемые при автоматизированном вождении, делаются ссылки на составной документ ЕЭК ООН «Новый метод оценки/испытаний для автоматизированного вождения (НМОИ)» в его текущем варианте. Системы на базе ИИ также могут разрабатываться аналогичным образом.



Рис. 1: Пример схемы

Примечание секретариата:

На рис. ниже представлен НМОИ, причем серым цветом выделены две дополнительные графы (Транспарентность и Управление данными). Цветной текст свидетельствует о том, каким образом НМОИ может применяться к системам на базе ИИ, содействовать оценке таких ключевых принципов, применимых к ИИ, как прозрачность, масштабируемость, робастность и т. д.