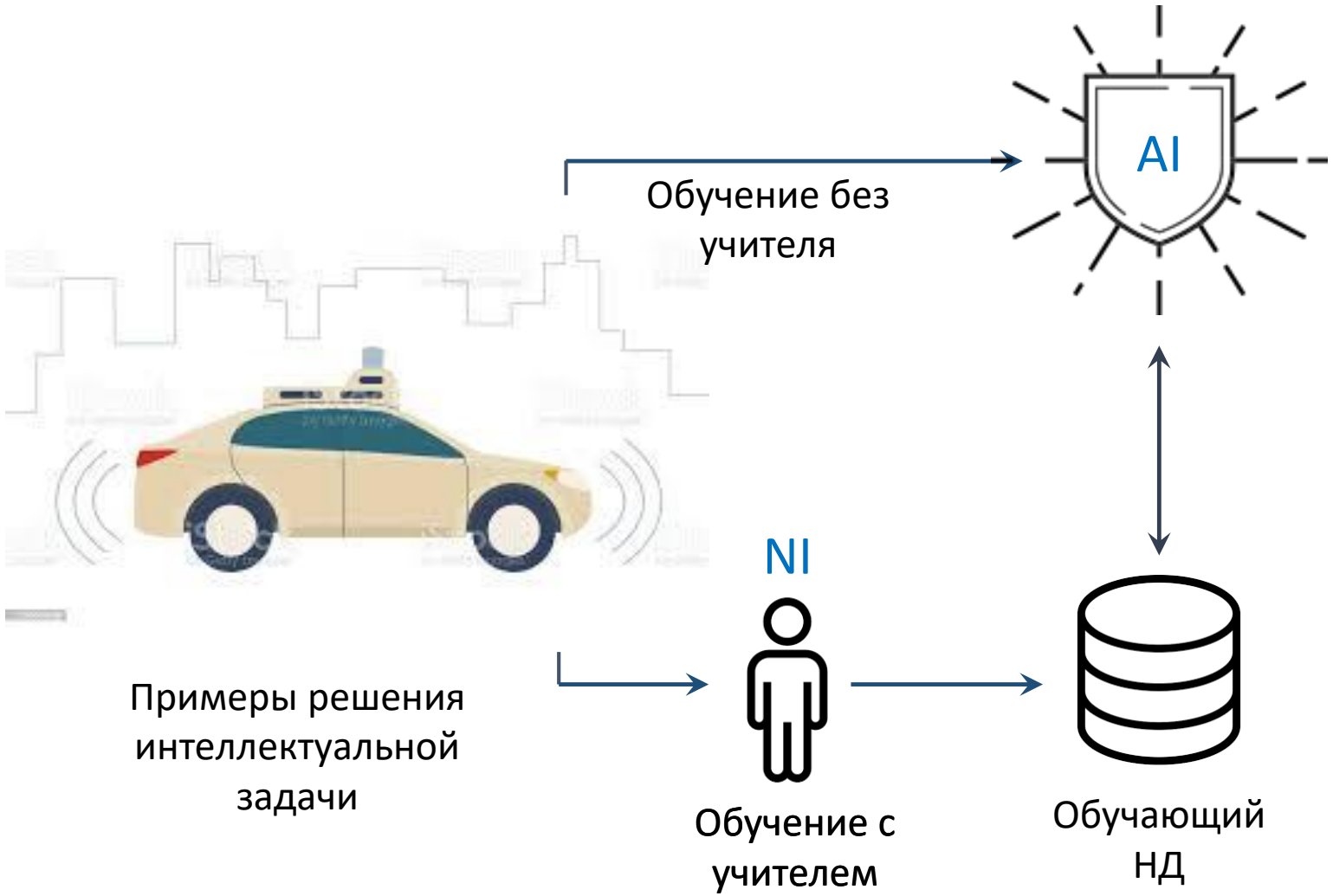
Технический Комитет 164

Особенности стандартизации технологии в области искусственного интеллекта

Заседание подкомитета ПК03 «Искусственный интеллект на транспорте»

ФАУ «РОСДОРНИИ», 01.02.2024

Технологии искусственного интеллекта – технологии обработки данных с использованием методов машинного обучения



Алгоритм системы ИИ принципиально не обладает полной понятностью (объяснимостью) для человека

Плохо предсказуемое поведение системы ИИ в реальных условиях эксплуатации, отсутствие в поведении систем «здравого смысла», подверженность воздействию т.н. «состязательных» атак на исходные данные

Некорректная работа алгоритмов ИИ наблюдается при определённых (плохо предсказуемых):

- условиях эксплуатации (сочетаниях параметров внешней среды и объекта измерения)

- небольших (не значительных с точки зрения здравого смысла человека) неумышленных или умышленных искажениях исходных данных, подаваемых на вход алгоритма ИИ

- характеристиках наборов данных, используемых для дообучения алгоритмов ИИ на стадии их эксплуатации

Актуальность вопроса оценки соответствия

- План мероприятий по совершенствованию законодательства и устранению административных барьеров в целях обеспечения реализации НТИ по направлению «Автонет» → изменения в Технический регламент «О безопасности колесных транспортных средств» ТРТС 018/2011

- Всемирный форум для согласования правил в области транспортных средств WP.29 ЕЭК ООН:
  РГ по автоматизированным/автономным и подключенным ТС (GRVA);

- Международная организация по стандартизации:
  ISO TC 22 Road vehicles/SC 36 Safety and impact testing (ТК «Дорожный транспорт»/ПК «Испытания на безопасность и ударопрочность»)
  ISO TC 204 (WG16 - ITS and C-ITS, CALM)

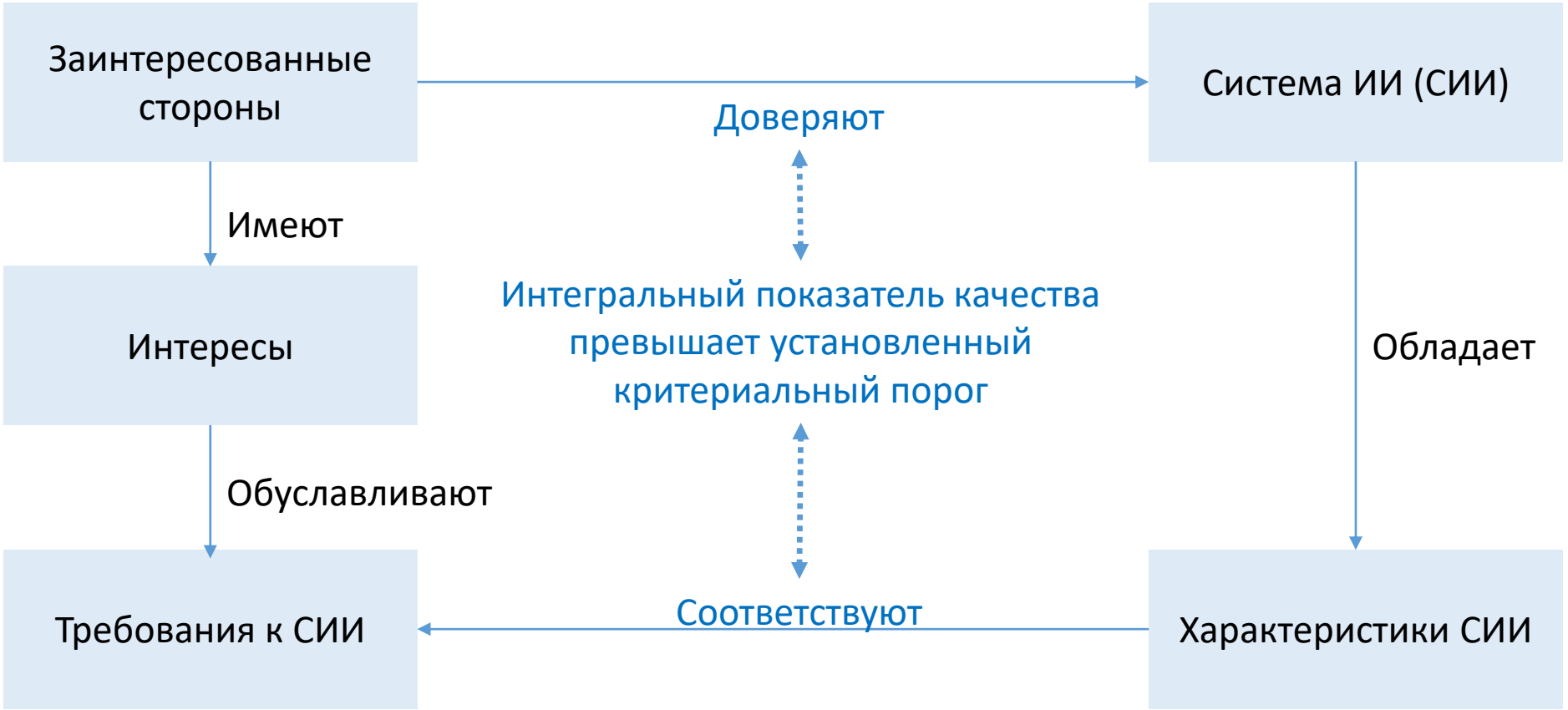
- Международная электротехническая комиссия IEC

- Европейский комитет по стандартизации CEN TC 278 WG16 on co-operative systems:
  Проектные группы PT1601, PT1604

- Европейский институт стандартизации в области телекоммуникаций ETSI TC ITS:
  Специализированная целевая группа – Special task forces: STF 455)

- Институт инженеров в области электротехники и электроники IEEE:
  Рабочая группа WG1609

Соответствие требованиям, качество и доверие к АТС ИИ

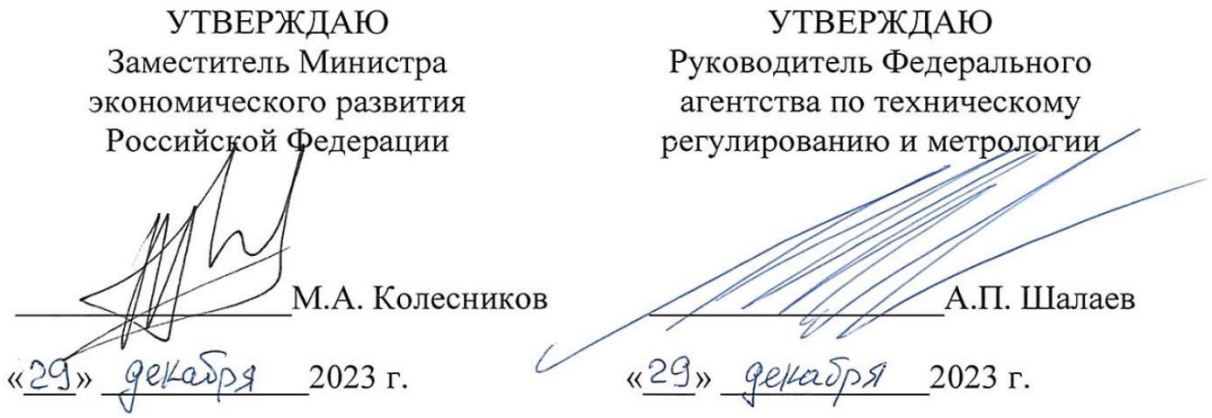


АТС ИИ с гарантированной функциональной корректностью

Для предусмотренных условий эксплуатации могут быть оценены:

- доверительные интервалы и вероятности прогнозирования погрешностей измерений для определенных условий проведения измерений
- предельные интегральные риски, связанные с некорректной работой системы ИИ
- ресурсы, необходимые злоумышленнику для успешного информационного воздействия на измерительную систему с алгоритмами ИИ (опционально, при наличии активного злоумышленника)

Перспективная программа стандартизации по приоритетному направлению «Искусственный интеллект» на 2021-2024 годы



УТВЕРЖДАЮ
Заместитель Министра
экономического развития
Российской Федерации

М.А. Колесников

«29» декабря 2023 г.

УТВЕРЖДАЮ
Руководитель Федерального
агентства по техническому
регулированию и метрологии

А.П. Шалаев

«29» декабря 2023 г.

ПЕРСПЕКТИВНАЯ ПРОГРАММА СТАНДАРТИЗАЦИИ ПО ПРИОРИТЕТНОМУ НАПРАВЛЕНИЮ «ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ» НА 2021–2024 ГОДЫ

Разработана в рамках федерального проекта «Искусственный интеллект» в декабре 2020 и актуализирована в декабре 2023 года.

Включает:

1. Описание принципов стандартизации ИИ;
2. Стандарты общего назначения (как разработанные на основе международных, так и разрабатываемые впервые)
3. Метрологические стандарты, направленные на унификацию способов измерения функциональных характеристик прикладных технологий ИИ в основных отраслях экономики и социальной сферы

Основные цели стандартизации ИИ

1) Обеспечение гарантий функциональной корректности СИИ в реальных условиях эксплуатации, в том числе – при дообучении СИИ в процессе эксплуатации и при автоматизации процессов обработки информации, связанных с заменой человека-оператора
2) Разработка методов и средств оценки и подтверждения безопасности СИИ, в том числе – в отношении третьих лиц (не участвующих непосредственно в эксплуатации систем), включая:
   - обеспечение физической безопасности СИИ для окружающих людей, природной среды и материальных активов (например, в случае беспилотного транспорта)
   - обеспечение специальных требований в области информационной безопасности СИИ
   - оценку уровня социальной приемлемости СИИ, в том числе – этических последствий разработки и применения этих систем
3) Обеспечение терминологического единства
4) Унификация форматов представления данных, необходимых для создания и применения СИИ, обеспечение интероперабельности информационных систем
5) Фиксация вариантов использования и лучших практик создания и применения СИИ при решении различных прикладных задач в отраслях экономики и социальной сферы

Особенности систем ИИ на основе алгоритмов машинного обучения

1. Отсутствие полной интерпретируемости
2. Обязательное использование специальным образом подготовленных наборов данных (НД), содержащих примеры решения конкретных прикладных интеллектуальных задач
3. Возможность дообучения алгоритмов МО в процессе эксплуатации АТС ИИ
4. Универсальность алгоритмов МО и, как следствие, к актуализации вопроса социальной приемлемости применения алгоритмов МО
5. Необходимость сравнения характеристик качества АТС ИИ и функциональных возможностей человека-оператора

Репрезентативные испытания – основной метод оценки соответствия ИИ

Предусмотренные условия эксплуатации АТС ИИ

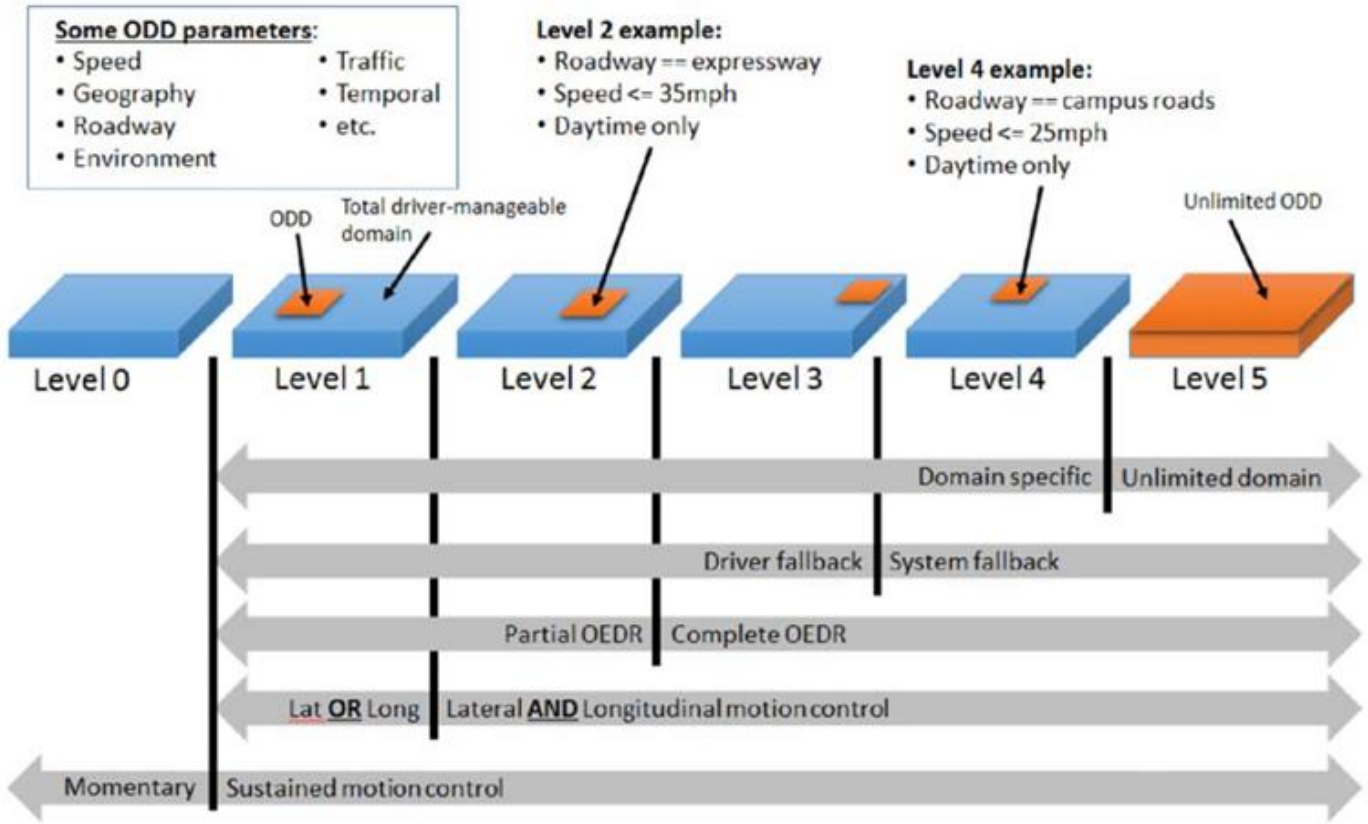
$$E = \{e_1, e_2 \dots e_K\}$$

e_k - существенные факторы эксплуатации АТС ИИ:

- параметры объектов и процессов, моделируемых с использованием методов МО (номенклатура и типоразмеры номеров, подлежащих распознаванию)
- параметры, характеризующие состояние операционной среды АТС ИИ и влияющие на работу алгоритмов МО (характеристики фона, количество одновременно наблюдаемых знаков)
- характеристики сенсоров, используемых для получения данных в АТС ИИ (пространственное и радиометрическое разрешение средств видеонаблюдения)
- условия получения данных (диапазон ракурсов и расстояний до знака, условия освещенности, скорость перемещения АТС относительно знака)
- наличие непреднамеренных искажающих воздействий (характеристики осадков, задымленности, загрязнений апертуры средств наблюдения, процент загрязнения и закрытия информативной части знака мешающими предметами)
- наличие и возможности злоумышленника, реализующего информационные атаки на АТС ИИ («отравление» обучающих НД, реализация состязательных атак на АТС ИИ путем размещения непосредственно на знаке и в поле зрения сенсоров специальных изображений, возможности по нарушению целостности данных от сенсоров АТС ИИ)
- другие необходимые факторы

Предусмотренные условия эксплуатации

(SAE J3016 Surface vehicle recommended practice. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. June, 2018)



$$E_2 = \left\{ \begin{array}{c} e_1(\text{тип автодорог}) = \text{"автострады"}; \\ e_2(\text{скорость движения}) \in \left[0; 35 \frac{\text{миль}}{\text{час}}\right]; \\ e_3(\text{условия освещенности}) = \\ \text{"светлое время суток"} \end{array} \right\}$$

При этом СФЭ определены на различных шкалах, например:

- бинарной – условия ДД/наличие велосипедистов = {есть, нет}
- номинальной – сезонные факторы/погодные условия = {дождь, снег, град}
- шкале отношений – сезонные факторы/температура $\in [-50, 50]$

SAE J3018 Surface vehicle recommended practice. Safety-Relevant Guidance for On-Road Testing of Prototype Automated Driving System (ADS)-Operated Vehicles. December, 2020

1. Тип проезжей части (e_1), включая, но не ограничиваясь:

автомагистрали/автострады с контролируемым доступом (полностью или ограниченно) ($e_1 = e_1^1$);

въезд/выезд на съезды/дорожные развязки ($e_1 = e_1^2$);

автомагистраль (одно- или многополосная) ($e_1 = e_1^3$);

магистральные дороги ($e_1 = e_1^4$);

улицы в жилых массивах ($e_1 = e_1^5$);

подъездная дорожка, автостоянка или сооружение ($e_1 = e_1^6$);

дороги с различным покрытием (асфальт, гравий, бетон и т.д.) ($e_1 = e_1^7$);

железнодорожные переезды ($e_1 = e_1^8$);

трамвайные пути ($e_1 = e_1^9$);

дороги без опознавательных знаков ($e_1 = e_1^{10}$) и др.

Таким образом, для множество всех возможных значений СФЭ, характеризующего тип проезжей части:

$$\{e_1^i\} \supseteq \{e_1^1, e_1^2 \dots e_1^{10}\}.$$

SAE J3018 Surface vehicle recommended practice. Safety-Relevant Guidance for On-Road Testing of Prototype Automated Driving System (ADS)-Operated Vehicles. December, 2020

2. Атрибуты инфраструктуры, включая, но не ограничиваясь:

дорожные условия: $\{e_2^i\} \supseteq \{\text{состояние ремонта, снег, лед}\}$;

наличие/отсутствие определенных средств управления ДД: $\{e_3^i\} \supseteq \{\text{дорожные знаки, сигналы, кольцевые развязки}\}$;

наличие/отсутствие разделительной полосы и/или барьера, разделяющего встречное движение: $\{e_4^i\} = \{\text{есть, нет}\}$;

прочие особенности: $\{e_5^i\} \supseteq \{\text{мосты с одностор. движ., неохр. жд переезды}\}$

3. Условия ДД, включая, но не ограничиваясь:

наличие/отсутствие определенных УДД: $\{e_6^i\} = \{\text{есть, нет}\}$ (для низкоскоростных ТС), $\{e_7^i\} = \{\text{есть, нет}\}$ (для пешеходов), $\{e_8^i\} = \{\text{есть, нет}\}$ (для велосипедистов) и т.д.;

наличие/отсутствие ближнего транспорта (включая относительные отклонения скорости v^9): $\{e_9^i\} = \{v^9, \text{нет}\}$;

наличие/отсутствие средств регулирования ДД/успокоения:

$$\{e_{10}^i\} \supseteq \left\{ \begin{matrix} \text{полосы для ТС с высокой вместимостью (HOV),} \\ \text{реверсивные полосы, полосы с ограничением по времени суток} \end{matrix} \right\};$$

наличие/отсутствие нештатных условий:

$\{e_{11}^i\} \supseteq \{\text{строительные работы, места ДТП, объезды дорог, затопление}\}$;

наличие/отсутствие сложных пересечений, слияний: $\{e_{12}^i\} = \{\text{есть, нет}\}$.

SAE J3018 Surface vehicle recommended practice. Safety-Relevant Guidance for On-Road Testing of Prototype Automated Driving System (ADS)-Operated Vehicles. December, 2020

4. Время суток, включая, но не ограничиваясь:

пробки на дорогах $e_{13} \in [0, 10]$ (количество баллов);

условия освещенности: $\{e_{14}^i\} = \{\text{день}, \text{ночь}\}$, $\{e_{15}^i\} = \{\text{облачно}, \text{ясно}\}$, $\{e_{16}^i\} = \{\text{есть}, \text{нет}\}$ (для уличного освещения);

$\{e_{17}^i\} = \{\text{есть}, \text{нет}\}$ (для бликов или низкого угла наклона солнца)

5. Сезонные факторы, включая, но не ограничиваясь:

погодные условия: $\{e_{18}^i\} \supseteq \{\text{дождь}, \text{снег}, \text{град}\}$;

помехи: $\{e_{19}^i\} \supseteq \{\text{туман}, \text{пыль}, \text{дым}\}$;

температура: $e_{20} \in [-50, 50]$;

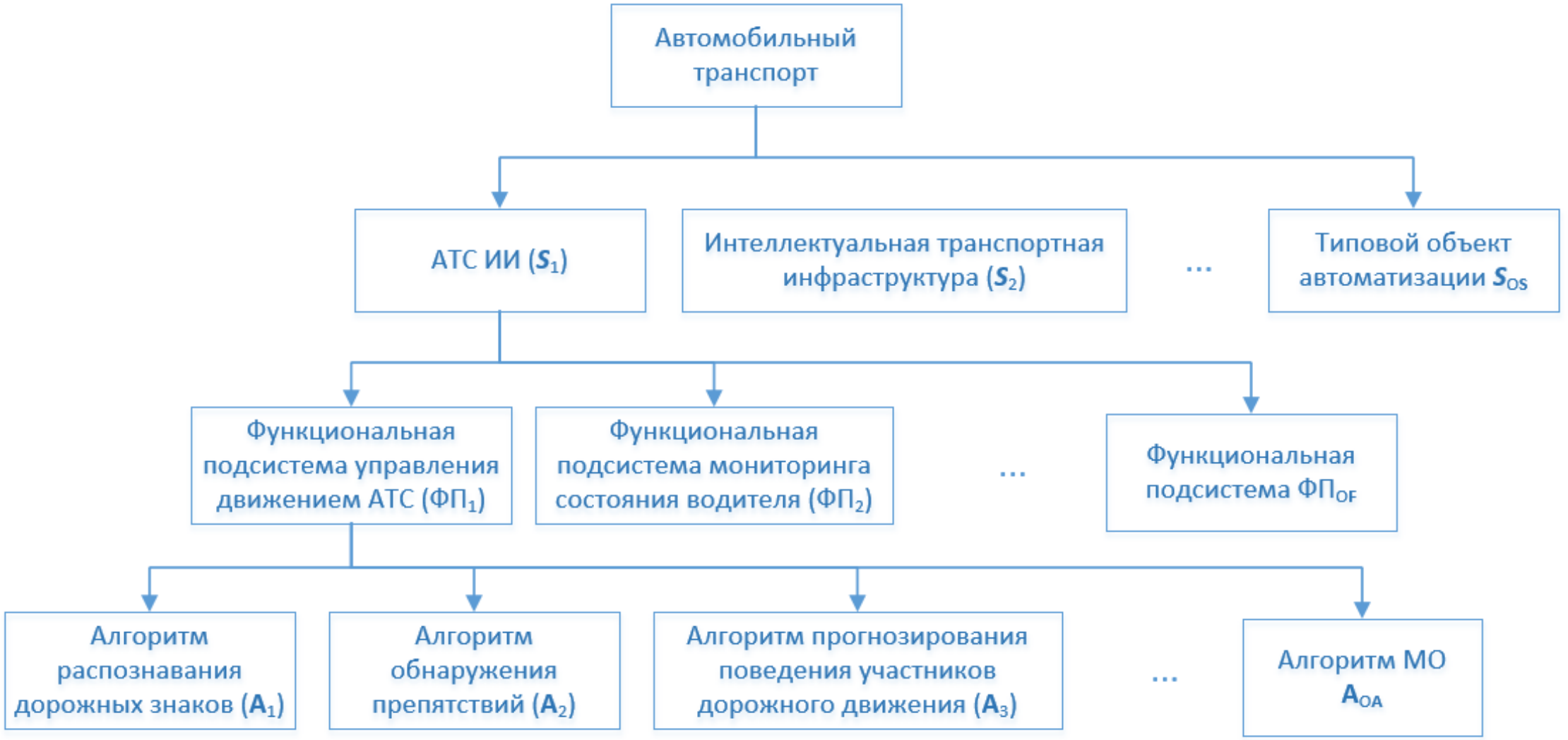
сезонный мусор: $\{e_{21}^i\} \supseteq \{\text{опавшие листья}, \text{сугробы}\}$

6. Местоположение, включая, но не ограничиваясь:

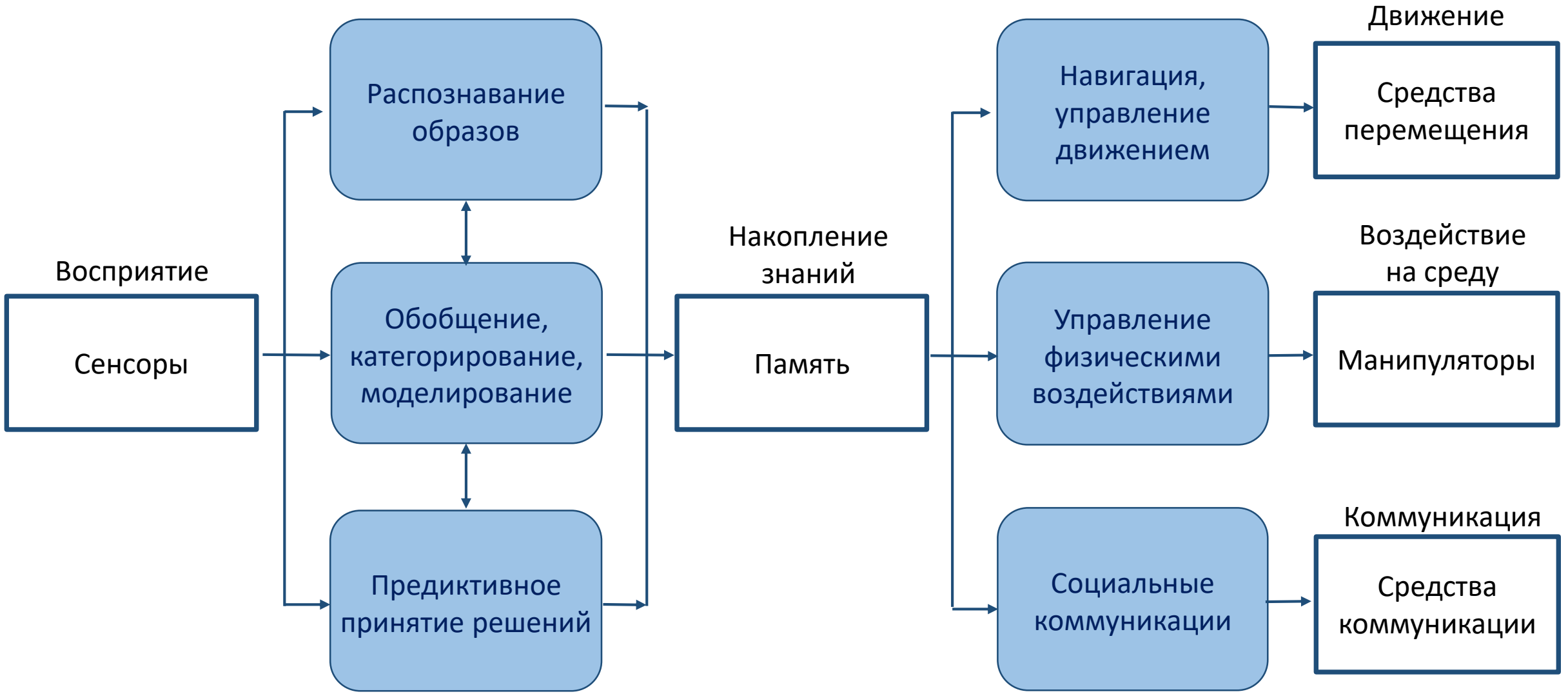
$\{e_{22}^i\} \supseteq \{\text{конкретные маршруты}, \text{дороги}, \text{закрытые кампусные территории}\}$.

Системы искусственного интеллекта на транспорте
Объект стандартизации

Алгоритм ИИ – основной объект оценки соответствия



Универсальные классы задач искусственного интеллекта (по аналогии с естественным интеллектом человека)



Задачи искусственного интеллекта на транспорте

Автономное управление ТС

- Распознавание дорожных знаков и разметки
- Распознавание объектов, препятствий
- Контроль обочины
- Распознавание структуры перекрестков
- Прогнозирование поведения участников дорожного движения

Контроль состояния водителя ТС

- Распознавание психоэмоционального состояния
- Распознавание движений рук, находящихся в руках предметов
- Распознавание состояния средств безопасности (положение ремня безопасности)

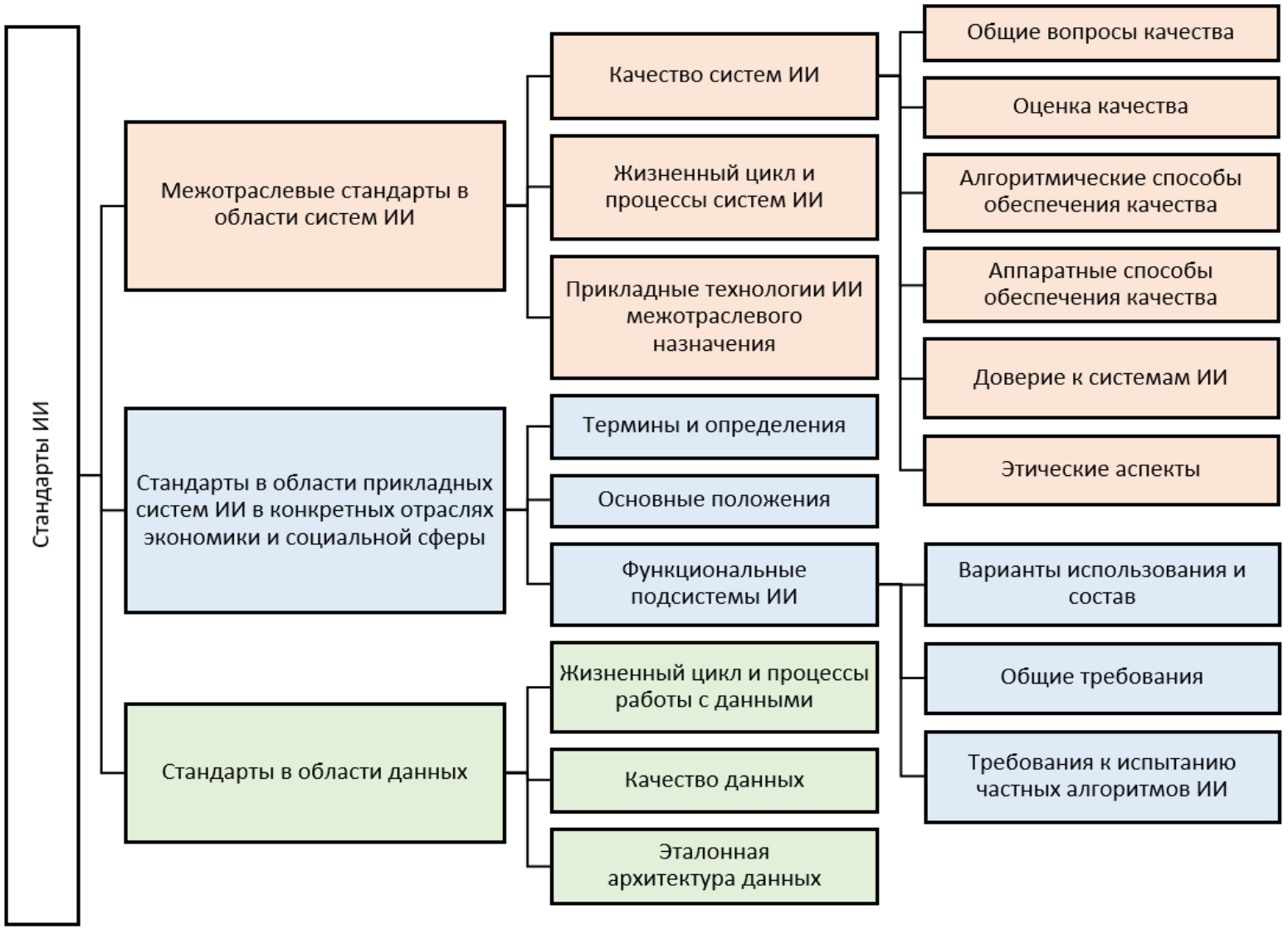
Управление транспортной инфраструктурой

- Моделирование транспортных потоков с использованием предиктивных моделей
- Управление объектами транспортной инфраструктуры

Распределение видов стандартов по уровням иерархии

		Уровень стандарта				
Назначение стандарта		I Комплекс стандартов	II Отрасль	III Типовой объект автоматизации (ТОА)	IV Функцио-нальная подсистема	V Алгоритм ИИ
1	Функциональная корректность	Общие принципы оценки качества и подтверждения соответствия требованиям (ГОСТ Р 59276, 59898...). Показатели репрезентативности обучающих и тестовых НД Общие принципы деклассификации НД, для обучения и тестирования СИИ	Функциональные характеристики и существенные факторы эксплуатации алгоритмов ИИ, применяемых в отраслевых ТОА (сборник). Демонстрационные НД для тестирования алгоритмов отраслевых ТОА	-	-	-
2	Физическая и информационная безопасность, социальная приемлемость			Принципы управления рисками, обусловленными некорректной работой систем ИИ		-
3	Термино-логические	Классификация систем и термины ИИ (ГОСТ Р 59277, 20546…), ПНСТ (553)	Специальные термины ИИ в отрасли (опционально, при небольшом числе терминов не выделяются в отдельный стандарт, а приводятся по месту)	-	-	-
4	Интер-операбельность систем	Эталонные архитектуры систем ИИ, стандарты больших данных	Специальные вопросы сбора, хранения и предоставления доступа к данным в отрасли или на крупном типовом объекте автоматизации		-	-
5	Лучшие практики	-	Лучшие практики и варианты использования ИИ в отрасли	Состав функциональных подсистем и прикладных алгоритмов ИИ в типовом объекте автоматизации	-	-

Комплекс национальных стандартов в области ИИ



Межотраслевые стандарты

- Требования к аппаратным и программно-алгоритмическим средствам, используемым для создания доверенных систем ИИ

Отраслевые стандарты

- Требования к унифицированным процедурам оценки качества прикладных систем ИИ

Данные

- Требования к данным, используемым для создания доверенных систем ИИ

Технический Комитет 164

Спасибо за внимание

Гарбук Сергей Владимирович

Председатель ТК164

www.tc164.ru