



О национальной системе оценки соответствия технологий искусственного интеллекта требованиям в области функциональной корректности и безопасности

Форум «Технологии и безопасность»

«Искусственный интеллект: вызовы и возможности»

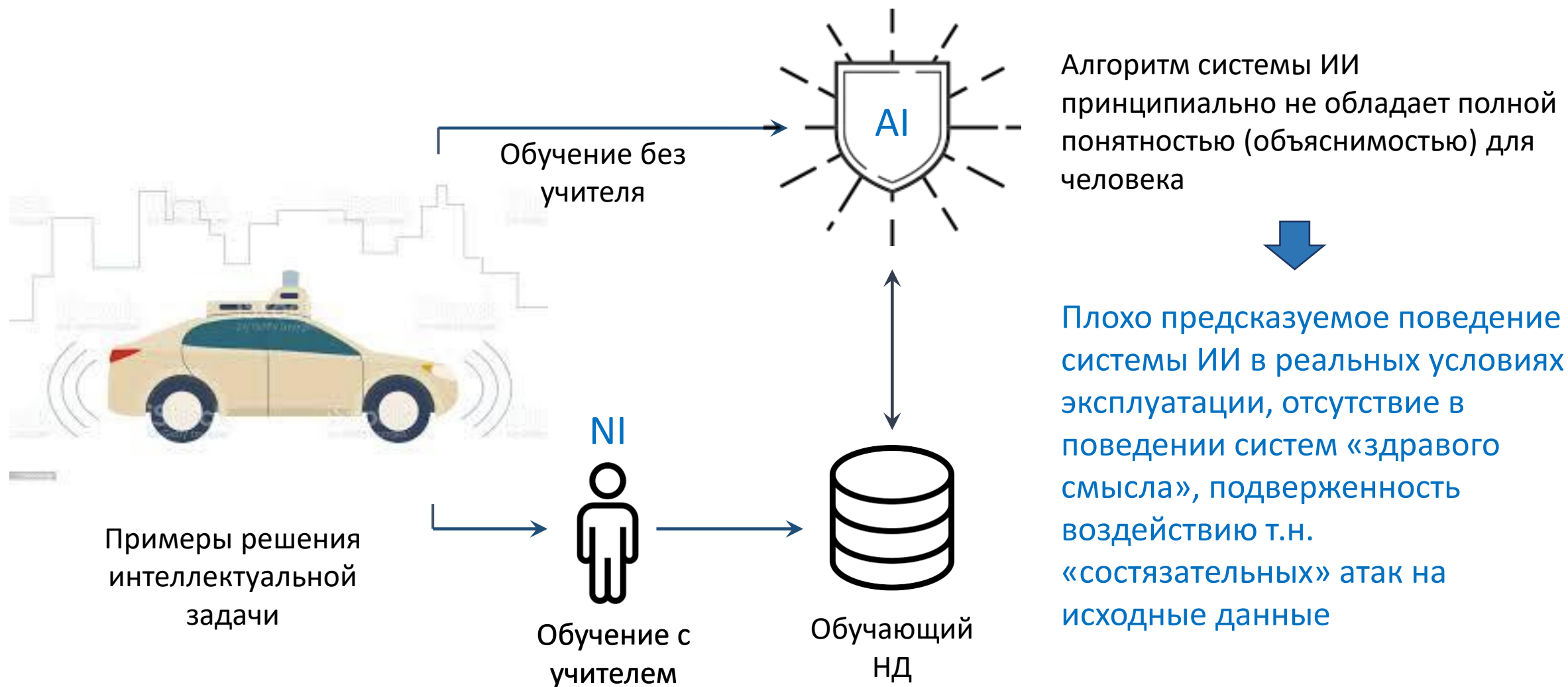
13 февраля 2024



План доклада

1. Особенности технологий искусственного интеллекта, реализованных с использованием методов машинного обучения
2. Структура национальной системы оценки соответствия в области ИИ
3. Комплекс национальных стандартов в области ИИ
4. Система добровольной сертификации ИИ «Интеллометрика»
5. Отраслевые эффекты оценки соответствия в области ИИ

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



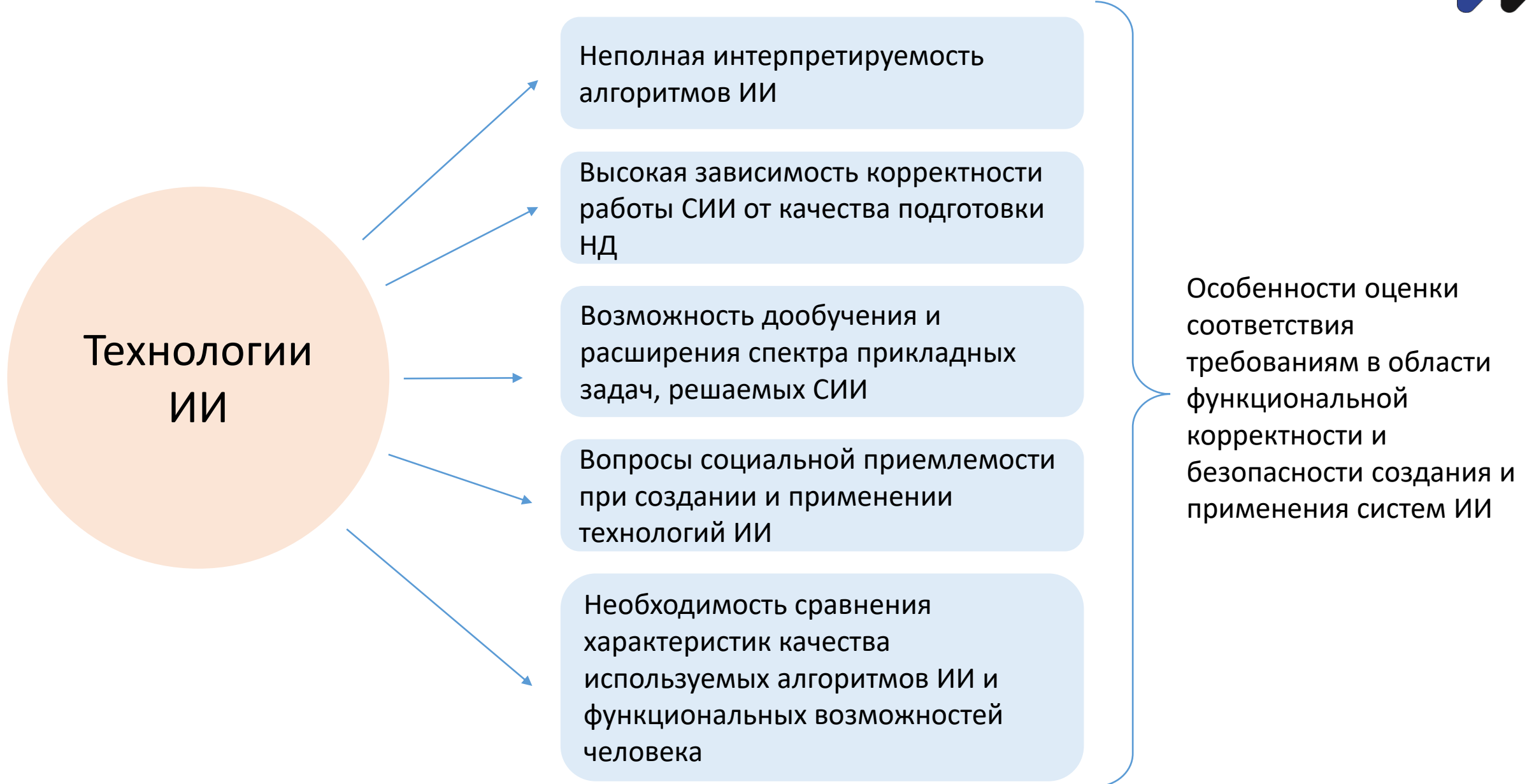
Некорректная работа алгоритмов ИИ наблюдается при определённых (сложно предсказуемых):



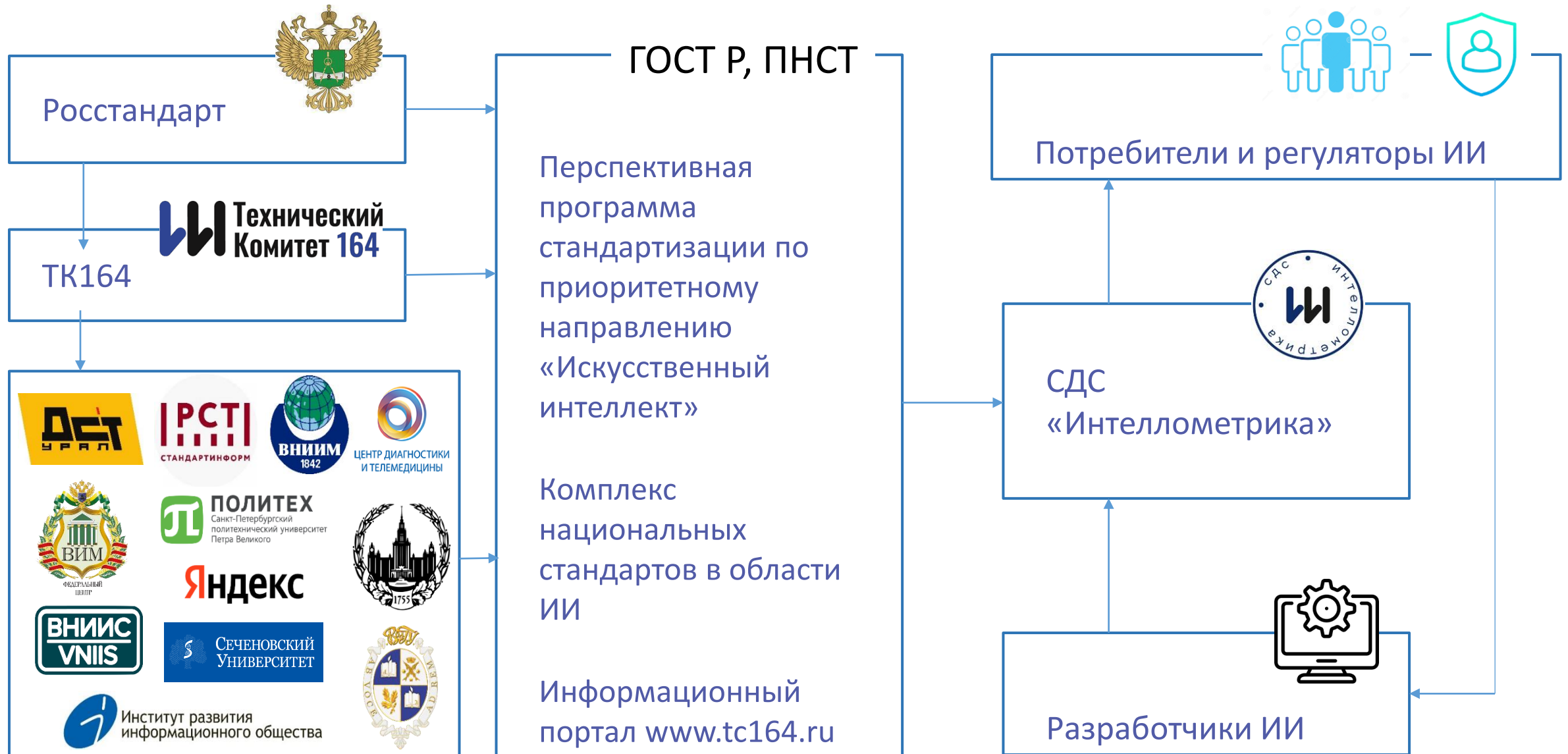
- условиях эксплуатации (сочетаниях параметров внешней среды и объекта измерения)
- небольших (не значительных с точки зрения здравого смысла человека) неумышленных или умышленных искажениях исходных данных, подаваемых на вход алгоритма ИИ
- характеристиках наборов данных, используемых для дообучения алгоритмов ИИ на стадии их эксплуатации



Особенности технологий ИИ



Национальная система оценки соответствия в области искусственного интеллекта





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

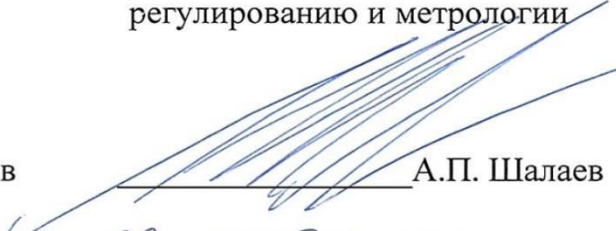
УТВЕРЖДАЮ
Заместитель Министра
экономического развития
Российской Федерации



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:

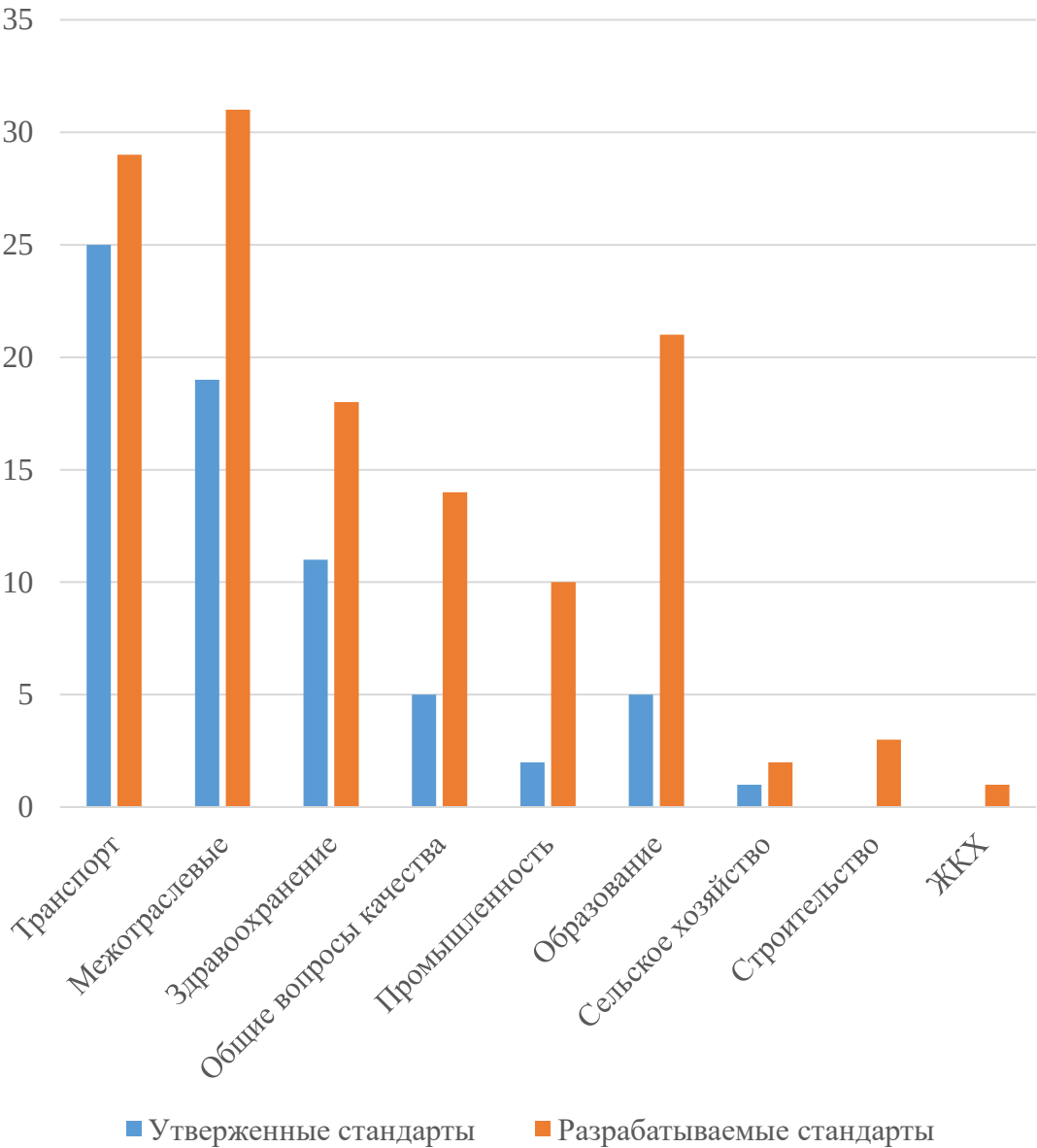
1. Описание принципов стандартизации ИИ;
2. Стандарты общего назначения (как разработанные на основе международных, так и разрабатываемые впервые)
3. Метрологические стандарты, направленные на унификацию способов измерения функциональных характеристик прикладных технологий ИИ в основных отраслях экономики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

Комплекс национальных стандартов, разработанный в рамках ТК 164 «Искусственный интеллект»



*количество утвержденных стандартов/количество разрабатываемых стандартов

Комплекс национальных стандартов в области ИИ (2020-2022)



Фонд стандартов ТК 164 на начало реализации ФП ИИ
7 документов

	План	Факт
2021 г.	0	21
2022 г.	32	40
За 2 года	32	61



Действующие стандарты ТК 164 доступны по ссылке: <https://www.rst.gov.ru/portal/gost//home/standarts/aistandarts>

Стандарты ИИ, утвержденные в 2023 году



1. ИИ в растениеводстве. Варианты использования для управления процессами
2. ИИ в животноводстве. Варианты использования для управления процессами
3. ИИ в переработке сельхозпродукции и производстве пищевой продукции. Варианты использования для управления процессами
4. Системы ИИ на водном транспорте. Варианты использования
5. Технологии ИИ в образовании. Функциональная подсистема управления успеваемостью обучающихся по программам бакалавриата и специалитета. Общие положения и методика испытаний
6. Технологии ИИ в образовании. Функциональная подсистема управления успеваемостью обучающихся по программам среднего профессионального образования. Общие положения и методика испытаний
7. Технологии ИИ в образовании. Функциональная подсистема формирования контингента абитуриентов по программам бакалавриата и специалитета. Общие положения и методика испытаний
8. Технологии ИИ в образовании. Функциональная подсистема управления успеваемостью обучающихся по программам подготовки научных и научно-педагогических кадров в аспирантуре. Общие положения и методика испытаний
9. Технологии ИИ в образовании. Функциональная подсистема управления успеваемостью обучающихся по программам ДПО. Общие положения и методика испытаний
10. Технологии ИИ в образовании. Функциональная подсистема организации участия в конкурсных мероприятиях с целью финансирования научной деятельностью. Общие положения
11. Технологии ИИ в образовании. Функциональная подсистема организации и проведения научных мероприятий. Общие положения
12. Технологии ИИ в образовании. Применение ИИ в научно-исследовательской деятельности. Варианты использования
13. ИТ. ИИ. Структура жизненного цикла данных (ISO/IEC 8183:2023, MOD)
14. ИИ. Структура описания систем ИИ, использующих машинное обучение (ISO/IEC 23053:2022, IDT)
15. ИИ. Методология оценки среды разработки ПО для глубокого обучения (ITU-T F.748.12(2021), NEQ)
16. ИИ. Управляемость автоматизированных систем ИИ (ISO 82000, IDT)
17. Приложения и сервисы для интеллектуального производства на основе машинного зрения. Требования (ITU F.748.16(2022), NEQ)
18. ИТ. Стратегическое управление ИТ. Последствие влияния стратегического управления при использовании ИИ организациями (ISO 38507:2022 MOD)
19. ИИ. Системы операционной аналитики потоков пространственно-временных данных на основе ИИ. Основные положения
20. ИИ. Системы операционной аналитики потоков пространственно-временных данных на основе ИИ. Термины и определения
21. ИИ. Смещённость в системах ИИ и при принятии решений с помощью ИИ (ISO 24027:2021 MOD)
22. Программная инженерия. Требования и оценка качества систем и ПО (SQaRE). Модель качества для систем ИИ (ISO 25059:2023, MOD)
23. Системная и программная инженерия. Требования и оценка качества систем и ПО (SQaRE). Руководство по оценке качества систем ИИ (ISO 25058, MOD)
24. ИИ. Обзор этических и общественных аспектов (ISO 24368:2022 MOD)
25. ИИ. Функциональная безопасность и системы ИИ (ISO 5469 MOD)
26. ИИ. Оценка эффективности моделей и алгоритмов МО в задаче классификации (ISO 4213:2022 MOD)
27. Системы поддержки принятия врачебных решений с использованием ИИ. Методы клинических испытаний
28. Образовательные цифровые платформы (тренажёры) с использованием ИИ для получения практических знаний в области клинических дисциплин. Основные положения
29. Системы ИИ в лучевой диагностике. Основные положения
30. Системы ИИ на автомобильном транспорте. Системы управления интеллектуальной транспортной инфраструктурой. Общие требования
31. Системы ИИ на автомобильном транспорте. Системы управления интеллектуальной транспортной инфраструктурой. Требования к испытанию алгоритмов прогнозирования характеристик транспортного потока
32. Системы ИИ на автомобильном транспорте. Системы управления интеллектуальной транспортной инфраструктурой. Требования к испытанию алгоритмов прогнозирования дорожных условий
33. Системы ИИ на автомобильном транспорте. Системы управления интеллектуальной транспортной инфраструктурой. Требования к испытанию алгоритмов распознавания автомобильных номеров
34. Системы ИИ на автомобильном транспорте. Системы технического диагностирования транспортного средства. Общие требования
35. Системы ИИ на автомобильном транспорте. Системы управления движением транспортного средства. Требования к структуре и архитектуре V2X-взаимодействия

Распределение видов стандартов ИИ по уровням иерархии

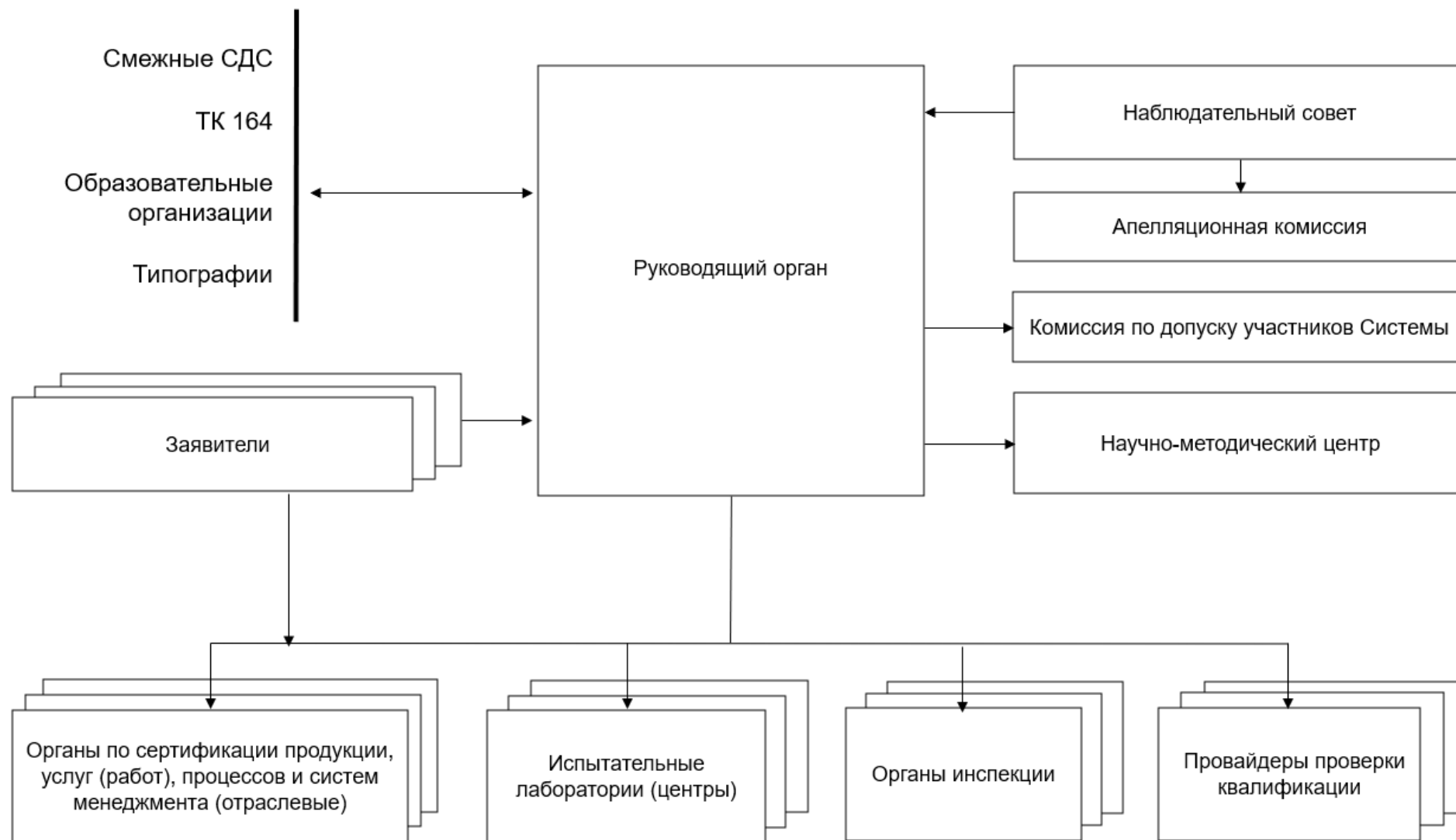


Назначение стандарта		Уровень стандарта				
		I Комплекс стандартов в целом	II Отрасль	III Типовой объект автоматизации (ТОА)	IV Функциональная подсистема	V Алгоритм ИИ
1	Функциональная корректность	Общие принципы оценки качества и подтверждения соответствия требованиям (ГОСТ Р 59276, 59898...). Показатели 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	Терминологические	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	Интероперабельность 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	-	-

Система сертификации технологий ИИ «Интеллометрика»



Зарегистрирована Росстандартом в едином реестре систем добровольной сертификации 26.12.2023
свидетельство № РОСС RU.B2915.04ВШЭ0



Подходы к формированию требований в отношении продукции, содержащей ИИ



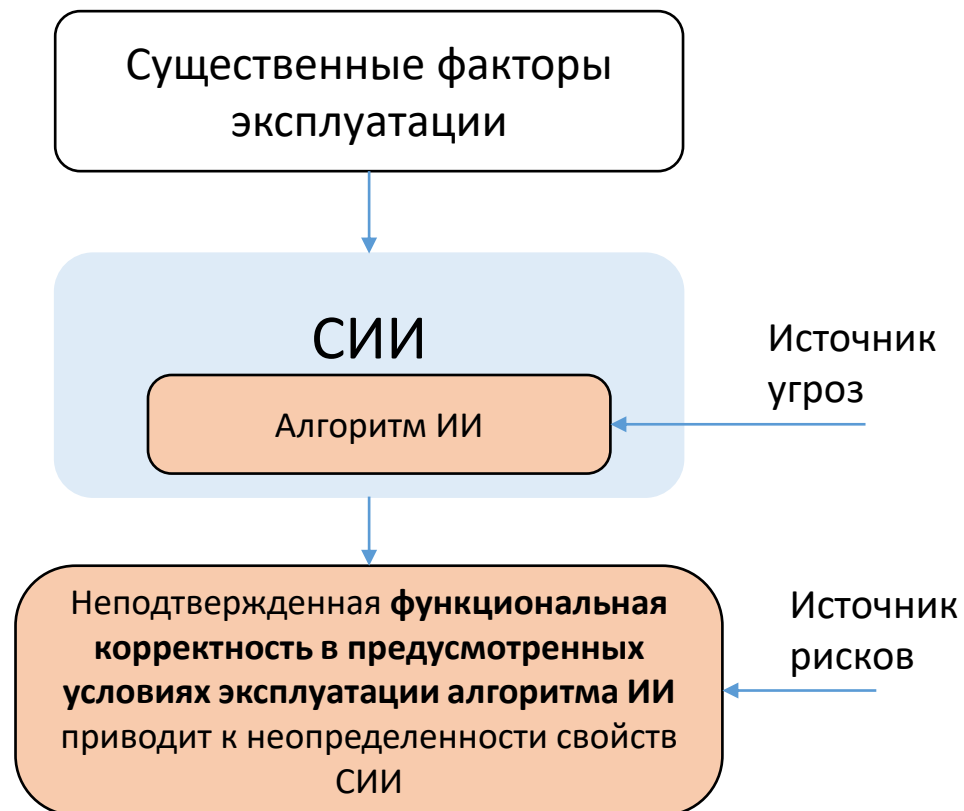
1

ИИ рассматривается как конструкционная особенность продукции, в результате чего разработка отдельных требований к системам ИИ не предусматривается

2

к используемым в продукции технологиям ИИ (программному обеспечению, алгоритмам, аппаратно-программным решениям) предъявляются специальные требования, учитывающие особенности реализации методов машинного обучения.

Технологии ИИ являются как самостоятельным объектом автономных испытаний, так и объектом комплексных испытаний в составе продукции в целом



Иерархическая модель объекта стандартизации



Взаимное признание сертификатов как принцип взаимодействия с отраслевыми системами оценки соответствия



Репрезентативность испытаний СЗИ с ИИ





ВИДЫ СУЩЕСТВЕННЫХ ФАКТОРОВ ЭКСПЛУАТАЦИИ

Существенные факторы эксплуатации

параметры, характеризующие состояние окружающей среды, объектов, процессов и сценариев, влияющие на качество решения прикладной задачи алгоритмом искусственного интеллекта

Контролируемые управляемые факторы

включённые в методику испытания (как независимые переменные)

Контролируемые неуправляемые факторы

факторы, постоянные в течение испытаний, являющиеся частью условий испытания

Неконтролируемые факторы

случайные и независимые от испытания факторы

Незначительные факторы

факторы, эффект от которых не будет учитываться

Примеры СФЭ ИИ для задачи распознавания

параметры объектов и процессов, моделируемых с использованием методов МО

параметры, характеризующие состояние операционной среды ИИ и влияющие на работу алгоритмов МО

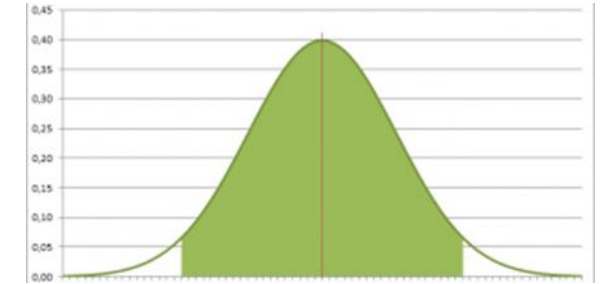
характеристики сенсоров, используемых для получения данных в ИИ

условия получения данных и другие необходимые факторы

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)



Аспекты безопасности автоматизированных систем с ИИ



Техническая надежность

- Аккредитованные лаборатории по проведению испытаний на устойчивость к климатическим, механическим и иным воздействиям
- КГВС «Мороз-7», «Климат-8» и др.

Информационная безопасность

- Система сертификации на соответствие требованиям в области защиты информации
- Национальные стандарты и руководящие документы ФСБ и ФСТЭК России

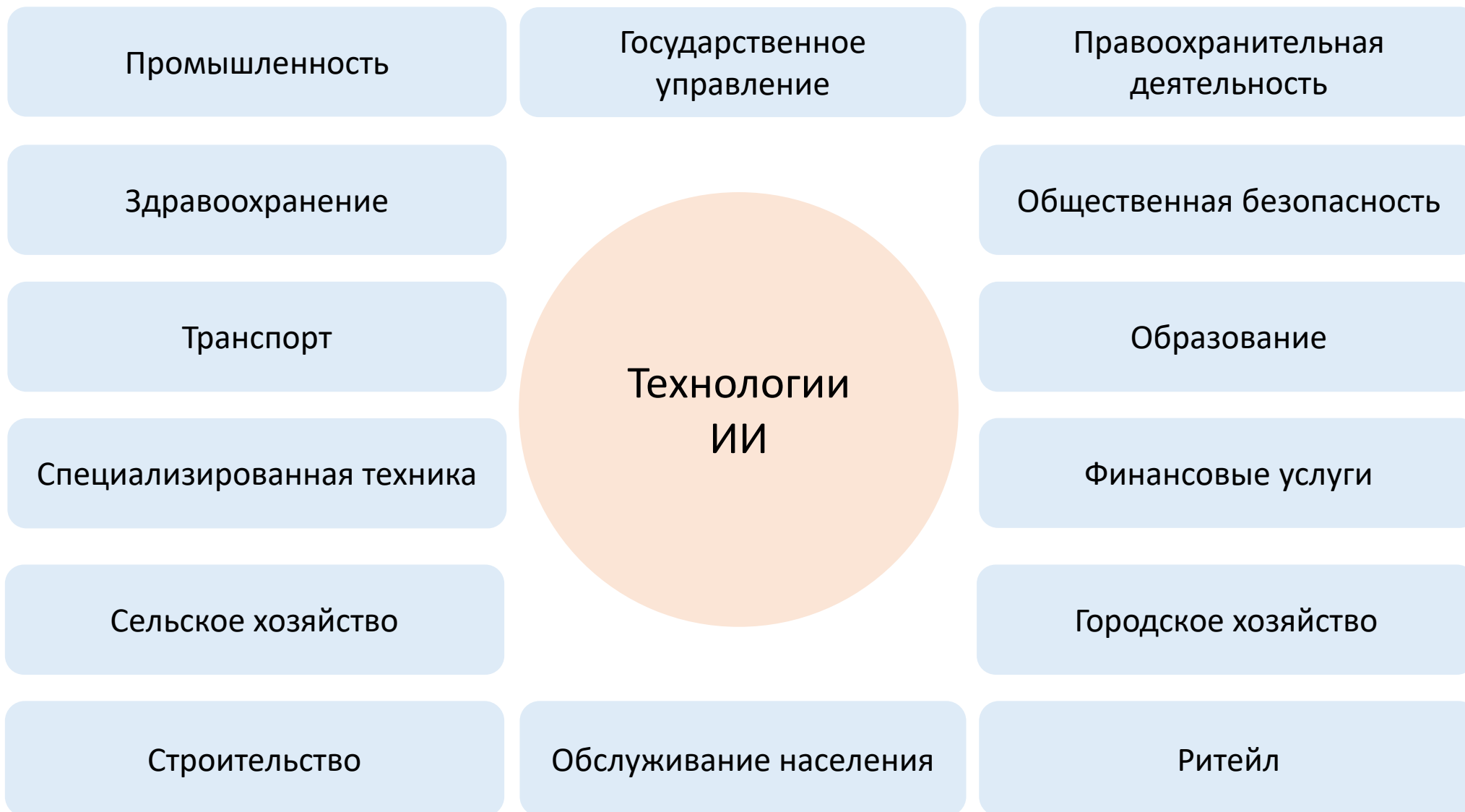
Импортонезависимость

- Реестр отечественного ПО, Минцифры России
- Реестр радиоэлектронной продукции российского происхождения, Минпромторг России

Функциональная корректность

- Комплекс национальных стандартов в области ИИ
- Система добровольной сертификации ИИ «Интеллометрия»

Сферы применения технологий ИИ



Отраслевые РГ по применению и стандартизации технологий ИИ



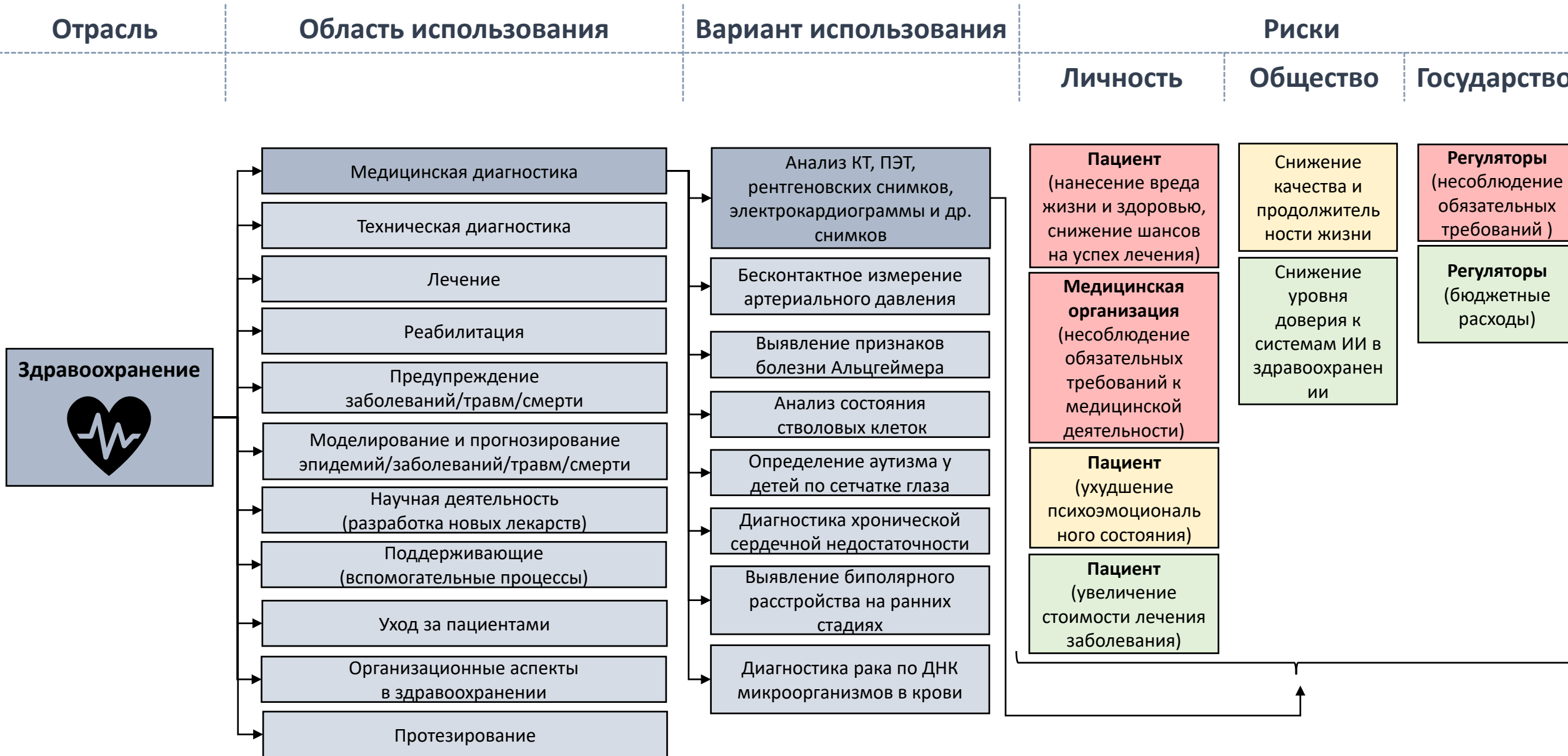
№	Область	Ведущая организация	Смежные ТК
1	Здравоохранение	НПКЦ диагностики и телемедицины Депздрава Москвы	ТК 011 «Медицинские приборы, аппараты и оборудование» ТК 436 «Управление качеством медицинских изделий» ТК 468 «Информатизация здоровья»
2	Образование	НИУ ВШЭ, ВолГУ	ТК 461 «Информационно-коммуникационные технологии в образовании»
3	Гражданская авиация	Союз авиапроизводителей России, ГосНИИАС	ТК 323 «Авиационная техника» ТК 363 «Радионавигация»
4	Специализированная техника	Ассоциация Росспецмаш	ТК 267 «Строительно-дорожные машины и оборудование» ТК 284 «Тракторы и машины сельскохозяйственные»
5	Дорожно-транспортный комплекс	РОСДОРНИИ, НАМИ	ТК 056 «Дорожный транспорт» ТК 057 «Интеллектуальные транспортные системы» ТК 278 «Безопасность дорожного движения» ТК 418 «Дорожное хозяйство» ТК 315 «Автомобильный и городской электрический транспорт»
6	Городской электрический транспорт	Горэлектротранс	
7	Водный транспорт	Отраслевой центр Маринет	ТК 032 «Водный транспорт»
8	Беспилотные летательные аппараты	НИЦ «Институт имени Н.Е.Жуковского»	ТК 323 «Авиационная техника»
9	Промышленность	ФЦПРИИ, СТАНКИН	ТК 070 «Станки»
10	Средства измерений и неразрушающий контроль	ВНИИМ им. Менделеева	ТК 053 «Основные нормы и правила по обеспечению единства измерений» ТК 371 «Неразрушающий контроль»
11	Ж/д транспорт	НИИАС	ТК 045 «Железнодорожный транспорт»
12	Ритейл	НП «РУССОФТ»	
13	Сельское хозяйство	Фед. научный агроинженерный центр ВИМ	ТК 284 «Тракторы и машины сельскохозяйственные»
14	Технические средства охраны	НИЦ «Охрана» Росгвардии	ТК 234 «Системы тревожной сигнализации и противокриминальной защиты»
15	Следственная деятельность	Московская академия СКР	ТК 134 «Судебная экспертиза»

Риски, обусловленные деградацией функциональных характеристик ИИ



Вид угроз, обусловленных нарушением функциональной корректности СИИ	Категория заинтересованной стороны	
	Лица, непосредственно участвующие в создании и применении СИИ (акторы ИИ)	Третьи лица
1 Угрозы жизни и здоровью людей, экологические угрозы	1.1 Потребители, разработчики и поставщики (собственная безопасность, дополнительные требования гос. регуляторов)	1.2 Общество в целом и регуляторы (безопасность общества и окружающей среды)
2 Угрозы информационной безопасности в отношении заинтересованных сторон	Нет	2.2 Общество в целом и государственные регуляторы (защита персональных данных, предотвращение деструктивных информационно-психологических воздействий)
3 Нарушение этических и других норм «мягкого» права	Нет	3.2 Общество в целом (социальная приемлемость создания и применения СИИ)
4 Неопределенные потребительские свойства, не влияющие непосредственно на безопасность жизни и здоровья людей, экологическую безопасность	4.1 Потребители (функциональные характеристики, определяющие возможность применения СИИ по назначению), разработчики и поставщики (характеристики конкурентоспособности СИИ)	Нет

Риски применения СИИ в здравоохранении



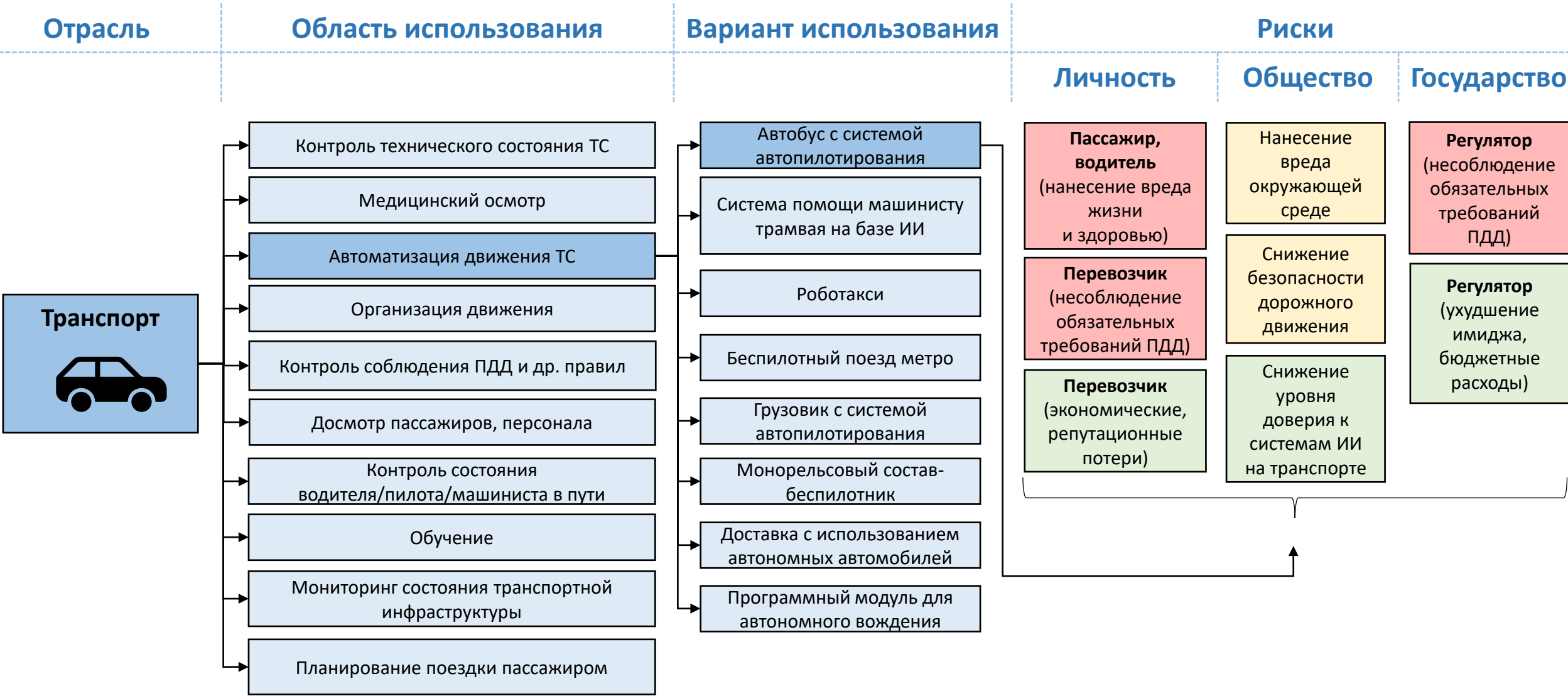


Риски применения СИИ в промышленности





Риски применения СИИ на транспорте



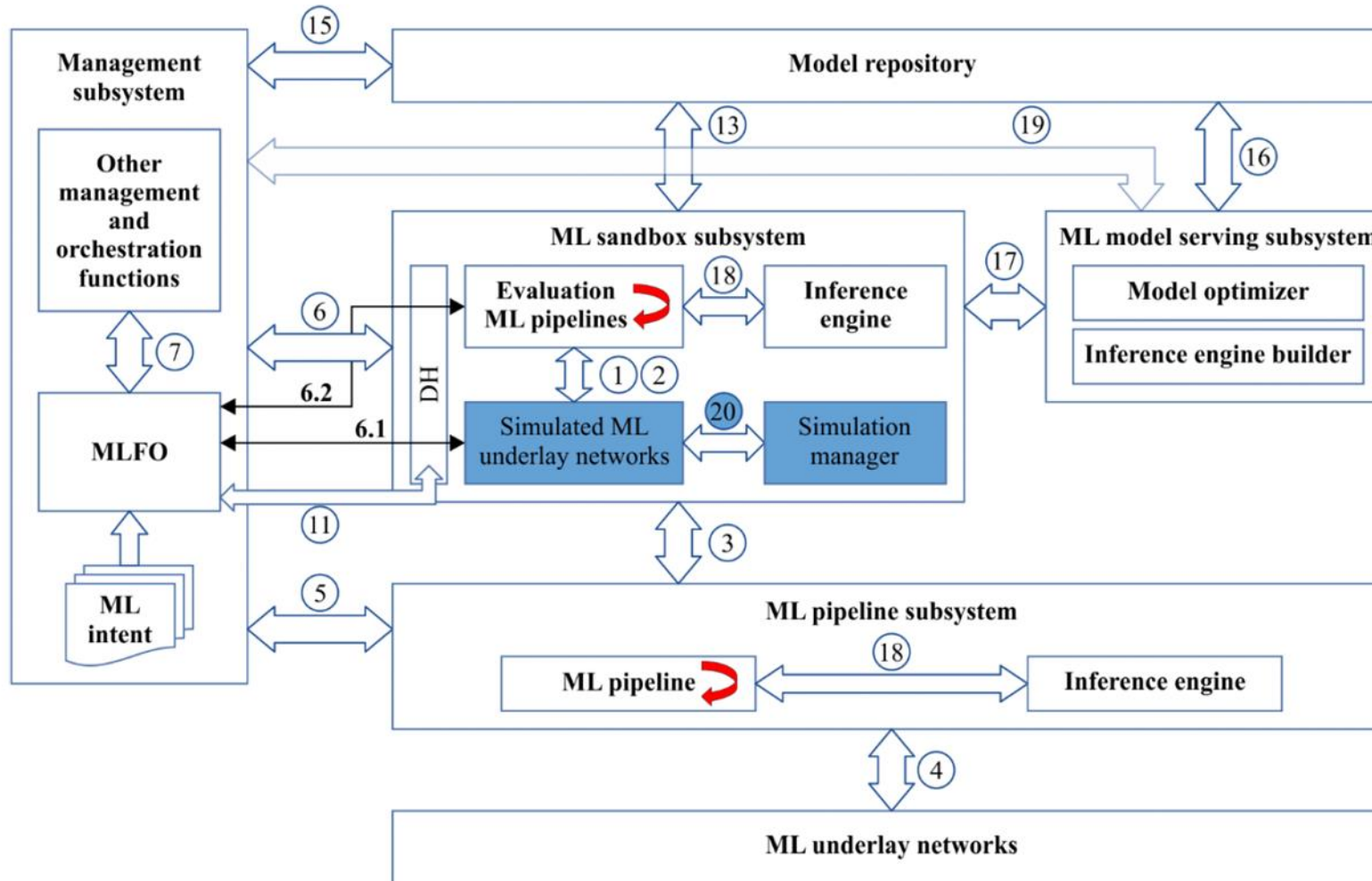
Межгосударственная система оценки соответствия ИИ (Interstate Conformity Assessment System)



РУС®СОФТ

RISING

(Russian-Indian
Special Interest
Group)



Y.3181(22)

Поручения по вопросу создания системы оценки соответствия в области искусственного интеллекта



- Перечень поручений Президента Российской Федерации по итогам конференции «Путешествие в мир искусственного интеллекта» Пр-172 от 29.01.23 (п.3.а):
Правительству Российской Федерации при участии ассоциации «Альянс в сфере искусственного интеллекта» обеспечить внедрение хозяйствующими субъектами унифицированной системы оценки отечественных решений, в которых используются технологии ИИ
Срок – 01.10.23
- План-график представления материалов по исполнению поручений... (п.29):
то же, отв. – Минцифры России
Срок – 11.04.24



ВЫСШАЯ ШКОЛА ЭКОНОМИКИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ



Спасибо за внимание

Гарбук Сергей Владимирович

garbuk@list.ru

www.tc164.ru