

**КРУГЛЫЙ СТОЛ НА ТЕМУ
«ПРАВО И ЭТИКА
В РАЗРАБОТКЕ
И ПРИМЕНЕНИИ
ИСКУССТВЕННОГО
ИНТЕЛЛЕКТА В
ГОСУДАРСТВЕННОМ
И КОРПОРАТИВНОМ
УПРАВЛЕНИИ»**

**МГУ им. М.В. Ломоносова
19 марта 2024 г. 15:00–18:00**

Технический комитет по стандартизации №164 «Искусственный интеллект» (ТК164)

является «зеркальным» к международному подкомитету **SC 42 “Artificial Intelligence”** объединённого технического комитета **ISO/IEC JTC 1 “Information Technologies”** и выполняет функции постоянно действующего национального органа в ISO.

Методическое руководство работой ТК 164, мониторинг и контроль за его деятельностью осуществляют **Федеральное агентство по техническому регулированию и метрологии (Росстандарт) и ФГБУ «Институт стандартизации».**

Всего в компетенции **ISO/IEC JTC 1** находятся **3472** опубликованных стандарта и **503** документа в стадии разработки.

Подкомитетом **SC42** разработано и опубликовано **25** стандартов, **31** стандарт находится в стадии разработки.

Серии национальных стандартов для отраслей:

Клиническая медицина - 11

ДЗЗ - 7

Транспорт – 8

Образование – 10

**> 40 национальных
отраслевых
стандартов**

TK 164

ПК01 «Искусственный интеллект в здравоохранении»

Базовая организация:
ГБУЗ "НПКЦ ДиТ ДЗМ"

Руководитель:

Васильев Юрий Александрович - к.м.н.,
главный специалист по лучевой и
инструментальной диагностике ДЗМ и МЗ РФ
по ЦФО РФ, директор ГБУЗ «НПКЦ ДиТ ДЗМ»

ПК02 «Данные»

Базовая организация:
МГУ им. М.В. Ломоносова
(Национальный центр цифровой экономики)

Руководитель:

Хохлов Юрий Евгеньевич - к.ф-м.н., доцент,
академик Российской инженерной академии,
Председатель Совета директоров Института
развития информационного общества

ПК03 «Искусственный интеллект в дорожно-транспортном комплексе»

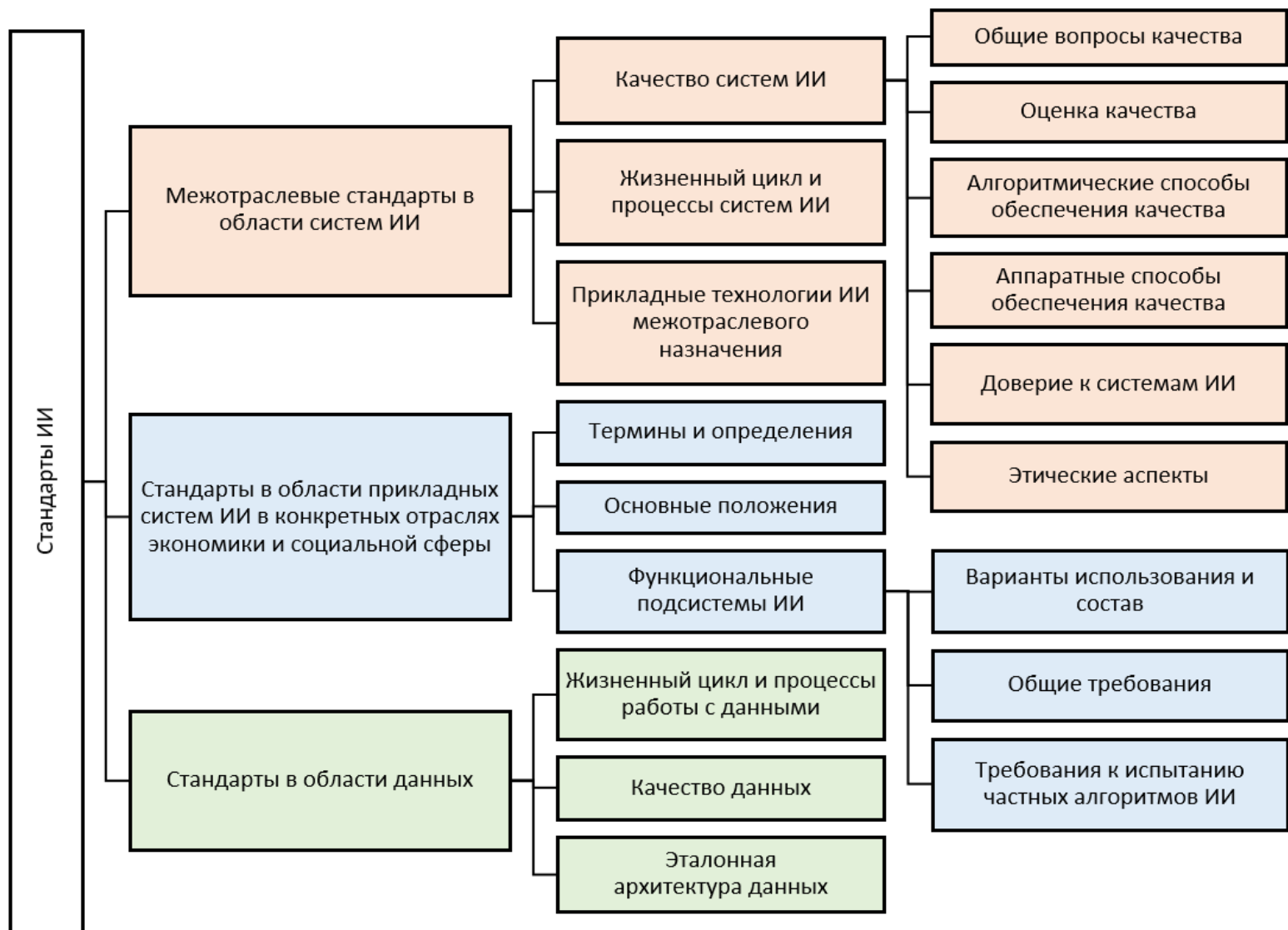
Базовая организация:
ФАУ "РОСДОРНИИ"

Руководитель:

Журавлев Антон Дмитриевич - заместитель
генерального директора ФАУ "РОСДОРНИИ"

ГОСТ Р – 43
ПНСТ - 33

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

ФЕДЕРАЛЬНОЕ АГЕНТСТВО
ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ



ПРЕДВАРИТЕЛЬНЫЙ
НАЦИОНАЛЬНЫЙ
СТАНДАРТ
РОССИЙСКОЙ
ФЕДЕРАЦИИ

ПНСТ
840—
2023
(ISO/IEC TR
24368:2022)

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Обзор этических и общественных аспектов

(ISO/IEC TR 24368:2022, Information technology — Artificial intelligence —
Overview of ethical and societal concerns, MOD)

до 01.01.2027

Содержание

1 Область применения	1
2 Нормативные ссылки	1
3 Термины и определения.	2
4 Описание.	3
4.1 Общие положения	3
4.2 основополагающие источники	4
4.3 Этические рамочные концепции	5
5 Правозащитная практика	6
5.1 Общие положения	6
6 Темы и принципы	7
6.1 Общие положения	7
6.2 Описание ключевых тем и связанных с ними принципов	8
7 Примеры практик для создания и использования этичного и общественно приемлемого ИИ	14
7.1 Согласование внутреннего процесса с принципами ИИ	14
7.2 Рассмотрение рамок этического анализа	16
8 Рассмотрение вопросов создания и использования этичного и общественно приемлемого ИИ	16
8.1 Общие положения	16
8.2 Неполный перечень этических и общественных аспектов для рассмотрения.	17
Приложение А (справочное) Документы, касающиеся принципов в области ИИ	22
Приложение В (справочное) Обзор примеров использования	30
Библиография	39

3.7 справедливость (fairness): Обработка, поведение или результаты, которые основаны на установленных фактах, общественных нормах и убеждениях и не определяются или не зависят от превосходства или несправедливой дискриминации.

Примечания

1 Соображения справедливости в значительной степени контекстуальны и варьируются в зависимости от культуры, поколения, географии и политических взглядов.

2 Справедливость — это не то же самое, что отсутствие предвзятости. Предвзятость не всегда приводит к несправедливости, а несправедливость может быть вызвана факторами, отличными от предвзятости.

ПРЕДВАРИТЕЛЬНЫЙ НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

**Смещенность в системах искусственного интеллекта и при принятии решений с помощью
искусственного интеллекта**

Artificial intelligence. Bias in artificial intelligence systems and artificial intelligence aided decision making

ОКС 35.020

Срок действия с 2024-01-01
до 2027-01-01

Приложение А (справочное) Примеры смещенности

Приложение Б (справочное) Инструменты с открытым исходным кодом

Приложение В (справочное) ISO 26000 – Пример карты соотношений

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

Системы искусственного интеллекта

СПОСОБЫ ОБЕСПЕЧЕНИЯ ДОВЕРИЯ

Общие положения

Artificial intelligence systems. Methods for ensuring trust. General

ОКС 35.240.01

Дата введения 2021-03-01

Предисловие

1 РАЗРАБОТАН Акционерным обществом "Всероссийский научно-исследовательский институт сертификации" (АО "ВНИИС"), Обществом с ограниченной ответственностью "ТВпортал" (ООО "ТВпортал")

ФЕДЕРАЛЬНОЕ АГЕНТСТВО
ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ



НАЦИОНАЛЬНЫЙ
СТАНДАРТ
РОССИЙСКОЙ
ФЕДЕРАЦИИ

ГОСТ Р
59898—
2021

ОЦЕНКА КАЧЕСТВА СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Общие положения

Содержание

1 Область применения	1
2 Нормативные ссылки	1
3 Термины и определения.	2
4 Сокращения	4
5 Модель качества.	4
6 Методология и критерии оценки качества на стадиях жизненного цикла СИИ.	5
7 Общие принципы и порядок оценки качества СИИ	6
7.1 Разработка программы тестирования (испытаний)	6
7.2 Алгоритм оценки качества СИИ	8
8 Представительный набор существенных характеристик и показатели качества СИИ.	11
8.1 Общие положения	11
8.2 Функциональные возможности (functionality) СИИ. Метрики	11
8.3 Уровень производительности (performance efficiency) СИИ. Метрики.	15
8.4 Способность СИИ к взаимодействию (compatibility). Метрики.	15
8.5 Мобильность (portability) СИИ. Метрики	16
8.6 Практичность (usability) СИИ. Метрики	16
8.7 Сопровождаемость (maintainability) СИИ. Метрики.	16
8.8 Надежность (reliability) СИИ. Метрики	17
9 Требования к набору входных и выходных данных для тестирования СИИ	18
Библиография	20

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

Руководство по социальной ответственности

Guidance on social responsibility

ОКС 03.100.01

Дата введения 2013-03-15

Предисловие

1 ПОДГОТОВЛЕН ТК 471 "Социальная ответственность" совместно с Академией труда и социальных отношений на основе выполненного компанией "Эрнст энд Янг" перевода на русский язык международного стандарта, указанного в пункте 4

2.18 социальная ответственность (social responsibility): Ответственность **организации** (2.12) за воздействие ее решений и деятельности на общество и **окружающую среду** (2.6) через прозрачное и **этичное поведение** (2.7), которое:

- содействует **устойчивому развитию** (2.23), включая здоровье и благосостояние общества;
- учитывает ожидания **заинтересованных сторон** (2.20);
- соответствует применяемому законодательству и согласуется с **международными нормами поведения** (2.11);
- интегрировано в деятельность всей **организации** (2.12) и применяется в ее взаимоотношениях.

Примечание 1 - Деятельность включает продукты, услуги и процессы.

Примечание 2 - Взаимоотношения относятся к деятельности организации в рамках сферы ее влияния (2.19).

2.7 этичное поведение (ethical behavior): Поведение, которое соответствует принятым принципам правильного или хорошего поведения в контексте конкретной ситуации, а также согласуется с **международными нормами поведения** (2.11).

2.4 должная предусмотрительность (due diligence): Всеобъемлющий активный процесс выявления фактического и потенциального негативного социального, экологического и экономического влияния деятельности организации в рамках всего жизненного цикла проекта, чтобы избежать или смягчить негативное воздействие.

Проект

СИСТЕМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ЗДРАВООХРАНЕНИИ. ЭТИЧЕСКИЕ АСПЕКТЫ

«Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы» (ГБУЗ «НПКЦ ДиТ ДЗМ»)

нормативные ссылки:

ГОСТ Р 59050 Инновационный менеджмент. Этическая оценка исследований и инноваций. Часть 1.

Компетенции комитета по этике

ГОСТ Р 59051 Инновационный менеджмент. Этическая оценка исследований и инноваций. Часть 2.

Оценка этического воздействия

ГОСТ Р 59276 Системы искусственного интеллекта. Способы обеспечения доверия

ГОСТ Р 59277 Системы искусственного интеллекта. Классификация систем искусственного интеллекта

Если СИИ разрабатываются для последующего их использования медицинскими работниками – специалистами по соответствующим профилям оказания медицинской помощи и процессам организации и оказания медицинской помощи, **их участие в разработке этих систем обязательно.**

ISO/IEC AWI TS 22443

JTC 1/SC 42/WG 3

Information technology — Artificial intelligence — Guidance on addressing societal concerns and ethical considerations

Stage: 20

20.00	1	New project registered in TC/SC work programme	2023-09-06	Current
30.00	1	Committee draft (CD) registered	2024-08-30	
60.60	1	International Standard published	2025-08-01	

ISO/IEC AWI 42105**JTC 1/SC 42/WG 3****Inform. technology — Artificial intelligence — Guidance for human oversight of AI systems**

Stage: 20.00

20.00	1	New project registered in TC/SC work programme	2023-08-21	Current
30.00	1	Committee draft (CD) registered	2024-09-30	
40.00	1	DIS registered.	2025-03-01 limit 2025-08-21	
60.60	1	International Standard published	2026-04-01 limit 2026-08-21	

ISO/IEC DIS 42005 JTC 1/SC 42/WG 1

Information technology — Artificial intelligence — **AI system impact assessment**

Stage: 40.20

40.00	1	DIS registered	2024-07-31	2024-08-09	2023-12-05	Closed
40.20	1	DIS ballot initiated	2024-02-01	2024-02-01	Current	
40.60	1	Close of voting	2024-04-25			
40.99	1	Full report circulated: DIS approved for registration as FDIS				
50.00	1	Final text received or FDIS registered for formal approval				2024-08-15
60.60	1	International Standard published	2024-12-05	limit	2025-08-09	

Риски, обусловленные нарушением функциональной корректности систем 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

Возможности и риски*

усилением возможностей человека за счет вычислительных ресурсов машин

сверхбыстрые вычисления огромной размерности, полезные прогнозы, научно-исследовательские функции, снижение рутинного труда (возможности творческой деятельности), др.



риск потери человеком контроля над управлением сложными социотехническими процессами

передача все больших функций и полномочий интеллектуальным машинам с высоким уровнем автономности. Риски, сопряженные с передачей машинам интеллектуальных функций и делегированием ответственности за принятые решения. Цифровая алгоритмизация среды обитания человека.

Предложения в резолюцию:

1. Развивать взаимодействие между научными площадками для эффективного участия профильных экспертов в **национальной стандартизации в сфере ИИ**. Укреплять **авторитет и позиции России** в области разработки международных стандартов.
2. Актуальные процессы развития сложностных технических систем доказывают необходимость **трансдисциплинарного познания техники и конструирования будущего** в т. ч. при развитии этических регуляторов для разработки и внедрения систем ИИ (в рамках синергетики, социогуманитарных исследований и поиска новых взаимосвязей философии и естественных наук).
2. Укреплять **субъектность человека и государства** при проектировании технических экосистем, в т ч с учётом достижения духовных и социально-значимых целей эволюционного развития человека, общества и планеты.
3. Развивать междисциплинарные проекты в области ИИ для **совершенствования моделей управления, экономического развития и проектирования социальных моделей общества**. Реализация таких проектов должна вестись при ответственной **комплексной оценке воздействия** технологий и систем на всех стадиях жизненного цикла систем ИИ. Обеспечить контроль систем ИИ со стороны человека и субъектность человека при взаимодействии с техническими экосистемами.

Спасибо за внимание!

Игнатъев Андрей

A.ignatev@hse.ru