



МЕЖГОСУДАРСТВЕННЫЙ СТАТИСТИЧЕСКИЙ КОМИТЕТ СОДРУЖЕСТВА НЕЗАВИСИМЫХ ГОСУДАРСТВ



МЕЖДУНАРОДНЫЙ ОПЫТ ИСПОЛЬЗОВАНИЯ AI и Big Data В СТАТИСТИКЕ

AI

Совещание по вопросам
искусственного интеллекта и
больших данных
г. Санкт-Петербург, 8-9 июня 2026 года

Фоминов Алексей Николаевич
Статкомитет СНГ



История искусственного интеллекта (AI) как научного направления



1956

Джон Маккарти предложил термин «искусственный интеллект». Этот момент считается началом истории AI.

1950

Алан Тьюринг в статье «Могут ли машины мыслить?» предложил тест для определения момента, когда машина сравнивается в разумности с человеком.

Эра генеративного AI: появление мультимодальных моделей, развитие генеративных сетей, больших языковых моделей. ИИ активно внедряется в медицину, финансы, статистику, производство и другие сферы.

2020

Прорыв в области глубокого обучения: разработаны методы глубокого обучения, нейронные сети, которые повысили точность в задачах распознавания изображений, речи и текста.

2010

Активное развитие машинного обучения и нейронных сетей благодаря росту вычислительной мощности.

1990

1960

Появились первые экспертные системы и языки программирования.

1970-1980

После завышенных ожиданий от развития AI финансирование исследований сократилось. Наступила – «Зима AI».



ОПРЕДЕЛЕНИЕ и ХАРАКТЕРИСТИКИ AI



1

Определение AI Евростатом - это системы, использующие следующие технологии для сбора и/или использования данных с целью прогнозирования, рекомендации или принятия решений с различной степенью автономности:

- Компьютерное зрение
- Машинное обучение
- Распознавание речи
- Интеллектуальный анализ текста

2

Ключевые характеристики AI в Статистике

- Автоматизация процессов сбора и анализа данных
- Повышение точности и своевременности статистики
- Обработка больших данных и новых источников информации
- Поддержка принятия решений на основе данных

3

Альтернативное определение AI - системы, использующие методы машинного обучения, обработки естественного языка и компьютерного зрения для автоматизации процессов производства статистики, повышения их качества и эффективности, а также для генерации аналитических инсайтов на основе больших данных.





ОСНОВНЫЕ ХАРАКТЕРИСТИКИ BIG DATA



1

Volume (Объём)

Гигантские масштабы данных, превышающие возможности традиционных систем хранения и обработки.

2

Velocity (Скорость)

Высокая скорость генерации, поступления и необходимости обработки данных — часто в режиме, близком к реальному времени.

3

Variety (Разнообразие)

Гетерогенность форматов и источников: структурированные таблицы, текст, изображения, геоданные, логи, видео — часто несовместимые по природе.

4

Veracity (Достоверность)

Важно, чтобы данные были точными, а способ их сбора — корректным.

5

Variability (Изменчивость)

Поток данных нестабилен: на него могут влиять скорость передачи, изменение источников.

6

Value (Ценность)

Анализ данных может принести пользу бизнесу или организации. Сами по себе данные ничего не значат, но на их основе можно сделать глубокие выводы и принимать взвешенные решения





Глобальные тренды и направления применения AI и BIG DATA в официальной статистике



Ключевые направления применения AI и Big Data, которые трансформируют производство статистики

**Автоматизация
обработки данных:
классификация текстов
с помощью LLM**



**Чат-бот для поиска
статистики**

**Улучшение качества
обследований: генерация
синтетических данных
для обучения моделей,
проверка аномалий,
автоматическая
импутация пропусков**



**Использование
альтернативных
источников данных:
данные мобильных
операторов,
спутниковые снимки,
данные ККМ и прочие.**





Ключевые вызовы для внедрения AI в статистике СНГ



ПРИНЦИПЫ КОДЕКСА ЭТИКИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

КЛЮЧЕВЫМИ ПРИНЦИПАМИ КОДЕКСА ЯВЛЯЮТСЯ ЧЕЛОВЕКОЦЕНТРИЧНОСТЬ И ПРАВДЕННОСТЬ, ПРЕДОСТОРОЖНОСТЬ И БЕЗОПАСНОСТЬ, ПОДКОНТРОЛЬНОСТЬ И РИСК-ОРИЕНТИРОВАННОСТЬ, ОТВЕТСТВЕННОСТЬ И НЕДИСКРИМИНАЦИЯ

ОСНОВНЫЕ ПОЛОЖЕНИЯ КОДЕКСА

- 
- 1** Главный приоритет развития технологий ИИ в защите интересов людей, отдельных групп, каждого человека
 - 2** Необходимо осознавать ответственность при создании и использовании ИИ пропорционально масштабу и значимости применения этих технологий
 - 3** Ответственность за последствия применения ИИ всегда лежит на человеке
 - 4** Технологии искусственного интеллекта можно и нужно внедрять там, где это принесет пользу людям
 - 5** Интересы развития технологий искусственного интеллекта выше интересов конкуренции
 - 6** Важна максимальная прозрачность и правдивость в информировании об уровне развития технологий ИИ, о возможностях, о рисках, об успехах и неудачах применения
 - 7** И технологии искусственного интеллекта, и этика искусственного интеллекта находятся вне границ
 - 8** Принципы этики развиваются по мере появления новых знаний, новых вызовов, новых возможностей

РИСКИ СВЯЗАННЫЕ С АТАКАМИ

НАРУШЕНИЕ НОРМАТИВНЫХ
ТРЕБОВАНИЙ

РИСКИ СВЯЗАННЫЕ С ДАННЫМИ

ТЕСТИРОВАНИЕ И ДОВЕРИЕ



Регион СНГ на пути к достижению Целей Устойчивого Развития



Общие тенденции

- Цифровая трансформация как стратегический императив.
- Фокус на большие данные и административные источники.
- Развитие информационных платформ для распространения данных.
- Углубление региональной координации между странами.

Ключевые проблемы применения технологий AI

- Пробелы и расхождение в данных
- Ограниченность ресурсов и институциональные барьеры
- Технологическое отставание и устаревшая ИТ-инфраструктура
- Сложность методологического обеспечения для новых данных

Задачи и стратегические направления дальнейшего развития AI в статистике.

- Системная трансформация
- Нарращивание статистического потенциала по наименее охваченным ЦУР.
- Развитие партнерства между НСС.
- Внедрение конкретных решений и пилотные проекты.



Глобальная конференция SDMX 2025



ЗАДАЧИ

- Обеспечение AI-Readiness данных
- Доверенные данные для AI
- Автоматизация аналитики и отчётности
- Внедрение управляемых AI
- Естественный язык и визуализация
- Развитие компетенций

ПРОБЛЕМЫ

- Галлюцинации и потеря контекста
- Дефицит качественных метаданных
- Семантические разрывы в реализации SDMX
- Высокие затраты и зависимость от облачных LLM
- Устаревшие метрики и измеримость
- Юридические, этические и суверенные риски
- Организационный и кадровый разрыв

НАПРАВЛЕНИЯ ДАЛЬНЕЙШЕГО РАЗВИТИЯ

- Исключение галлюцинаций AI и обеспечение интероперабельности
- Использование управляемых и локальных AI-решений
- Обогащение LLM формализованными знаниями (Семантический WEB)
- Интеграция Big Data с сохранением статистического качества
- Легитимизация AI в официальной статистике
- Превращение статистики в диалог, рост грамотности пользователей
- Ускорение внедрения, формирование единых стандартов качества



Семинар Группы высокого уровня по модернизации официальной статистики



Общие тенденции

- Стратегическая готовность к AI
- Семантическая интероперабельность фундамент для доверенного AI
- Инновационные методы обучения моделей
- Прагматичный подход к внедрению AI

Задачи

- Создание условий для внедрения AI
- Семантическая интероперабельность
- Инновации в методах обучения моделей

Проблемы

- Кадровый дефицит
- Технологическая зависимость и вопросы суверенитета данных
- Дефицит данных и знаний для обучения моделей
- Отсутствие единой системы оценки качества и неравномерная готовность к внедрению





Архитектура Хаба Данных Статкомитета СНГ



ПАК ИАП

Программно-аппаратный комплекс



- начальная точка формирования метаинформации с учетом международных стандартов (GSBPM, GSIM)
- сбор данных от НСС

ЕИАС

Единая информационно-аналитическая система



- аналитическая визуализация (BI, OLAP)
- сбор данных и метаданных международных источников
- электронные публикации

СУЗ-СОД

Система управления знаниями и связанных открытых данных



- библиотека методологических материалов
- глоссарии терминов
- наборы связанных открытых статистических данных (СОСД)
- инструменты гармонизации



ПЛАНЫ ПО ИСПОЛЬЗОВАНИЮ AI и Big Data в Хабе Данных



1

Испытание возможностей AI для сопоставления структурных метаданных Перечня, шаблонов форм, реестра базовых показателей БД ПАК ИАП

2

Анализ полноты формирования метаданных БД ПАК ИАП на основе шаблонов форм вопросников с применением технологии AI

3

Внедрение искусственного интеллекта в ЕИАС для автоматизации процессов и расширения аналитических возможностей, включая создание с применением ИИ интерфейса произвольных запросов в каталоге открытых данных ЕИАС

4

Создание единой системы метаданных Хаба данных Статкомитета СНГ и технологии автоматизированной гармонизации их

BIG DATA



ПЛАНЫ ПО ИСПОЛЬЗОВАНИЮ AI и Big Data в Хабе Данных



5

Развитие подсистемы СУЗ в части применения AI и совершенствования подготовки информационных, в том числе методологических, материалов

7

Апробация формирования предложений для терминологической статьи

9

Апробация применения AI, встроенного в СУЗ, в статистической практике

LLM

LANGUAGE MODELS

Апробацию формирования аннотаций к документам с применением AI

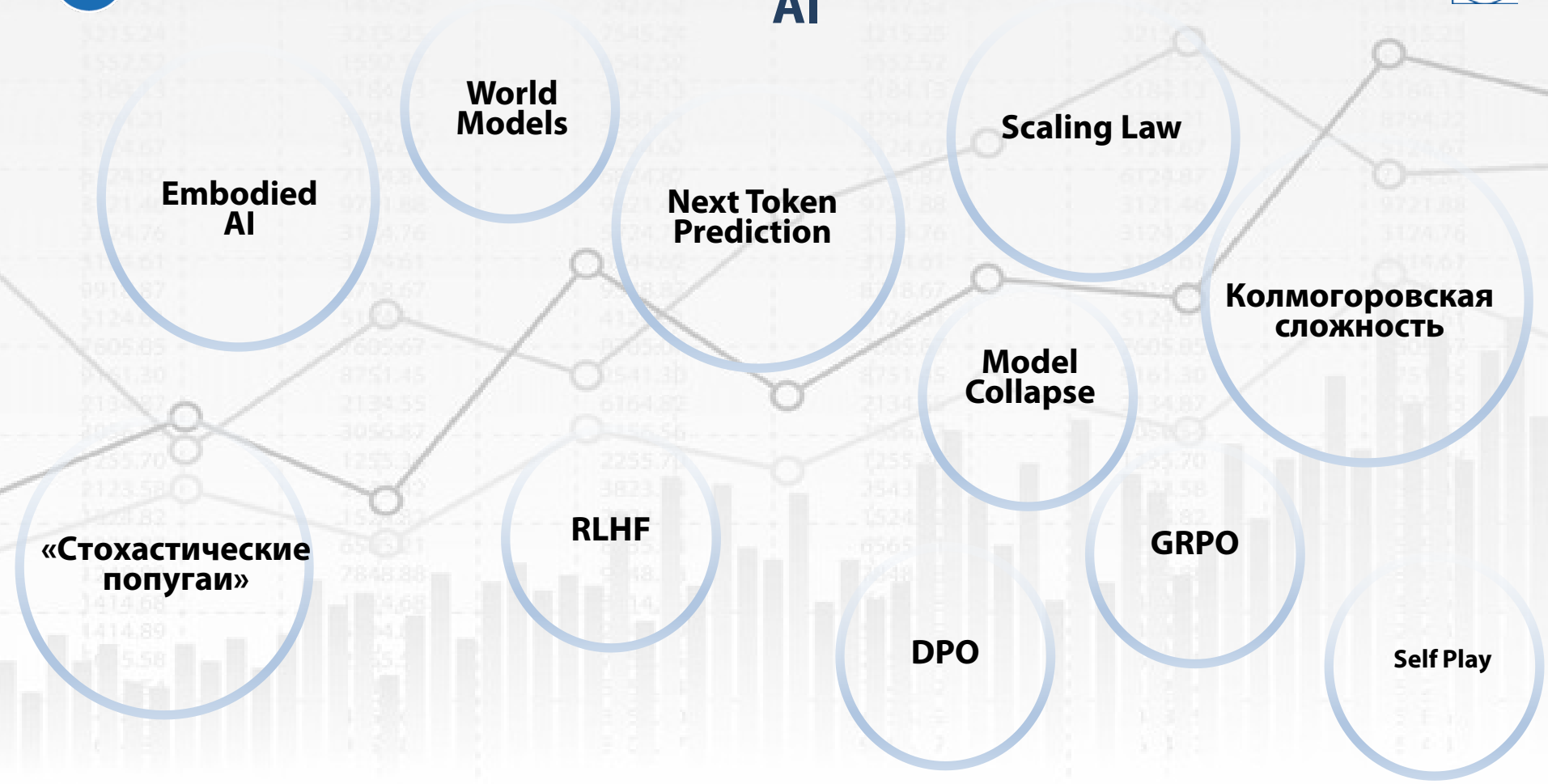
6

Создание виртуального помощника статистика с функциями AI.

8



СПОРЫ о ПРОБЛЕМАХ ЭВОЛЮЦИИ AI





СОЗДАНИЕ СТАТИСТИКИ ГОТОВОЙ ДЛЯ ИСПОЛЬЗОВАНИЯ AI



МЫ ПЫТАЕМСЯ СФОРМИРОВАТЬ ЕДИНУЮ ЭКОСИСТЕМУ СТАТИСТИКИ, В КОТОРОЙ СТАТИСТИЧЕСКИЕ ЗНАНИЯ, ДАННЫЕ, МЕТАДАННЫЕ И AI БУДУТ НЕРАЗРЫВНО СВЯЗАНЫ МЕЖДУ СОБОЙ.

КОГДА ЭТО БУДЕТ ДОСТИГНУТО ПРОИЗОЙДЕТ ЗНАЧИТЕЛЬНОЕ ПОВЫШЕНИЕ КАЧЕСТВА НАШЕЙ АНАЛИТИКИ И ОБЯЗАТЕЛЬНО ВОЗРАСТЕТ УРОВЕНЬ УДОБСТВА РАБОТЫ ПОЛЬЗОВАТЕЛЕЙ НА ПОРТАЛЕ СТАТКОМИТЕТА СНГ ПРИ ПОЛУЧЕНИИ СТАТИСТИЧЕСКИХ ЗНАНИЙ.



СПАСИБО ЗА ВНИМАНИЕ!

