

Проект DepoAI:

депозитарий реальных прикладных задач для масштабирования
модульных практико-ориентированных курсов AI/ML/DS

Воронцов Константин Вячеславович

д.ф.-м.н., профессор РАН, руководитель лаборатории
машинного обучения и семантического анализа



**Институт перспективных исследований проблем
искусственного интеллекта и интеллектуальных систем
МГУ им. М.В. Ломоносова**

Проблемы и вызовы

- Нехватка кадров для обеспечения массового внедрения технологий ИИ
- Дефицит инновационных (в т.ч. опережающих) разработок в области ИИ
- Растущая потребность экономики: 15,5 тыс. новых специалистов ежегодно

Цель: масштабировать процесс подготовки специалистов в РФ

Задача: создать экосистему для преподавателей и исследователей

- централизация обмена датасетами и образовательными материалами
- свободный обмен лучшими практиками и методиками преподавания
- устранение барьеров входа для молодых преподавателей

Рабочие гипотезы

- Наиболее эффективное обучение в ИИ — по реальным прикладным кейсам
- За время обучения будущий специалист должен решить около 20 реальных **учебных и исследовательских задач** из различных областей экономики
- Каждая задача должна иметь социально-экономическую значимость и (желательно) конкретного российского заказчика или бенефициара
- Проблема — практически нет преподавателей, имеющих опыт решения или преподавания сопоставимого числа разнообразных реальных задач
- **Необходима «касса взаимопомощи», «общак»
— национальный депозитарий учебных и исследовательских задач AI/ML/DS, вокруг которого сложится экосистема**

Интересы сторон

- **бизнес:**
привлечение научного сообщества к решению промышленных задач
PR, HR, развитие отраслевых профсообществ совместно с университетами
- **система образования:**
внедрение модульных практико-ориентированных курсов по AI/ML/DS
- **государство:**
обеспечение технологического суверенитета,
создание трудового ресурса для цифровизации экономики
- **общество:**
ответ на вопрос «зачем нужна цифровизация и ИИ»;
единая точка входа в индустрии через задачи AI/ML/DS

Стандартизация описания задач (кейсов)

1. Научно-популярное введение, история успеха, мотивация
2. Образовательный контент (лекции, видео, слайды)
3. Источники научной и отраслевой информации
4. Ссылки на связанные задачи
5. **Открытый набор данных**
6. **Открытый код базовых решений и метрик качества**
7. Активный контекст на конкурсной площадке
8. Методики и системы автоматизированного контроля знаний
9. Методические материалы, обмен опытом преподавания
10. Открытые исследовательские задачи для НИР/НИОКР

Две основные категории задач

1. Учебные

- *закрытые проблемы, открытые модели*
- задачи, которые уже кем-то решены (не обязательно хорошо)
- открытый доступ, общественное достояние

2. Исследовательские

- *открытые проблемы, закрытые модели*
- задачи, пока не имеющие готовых отраслевых решений на рынке
- задачи, пока не решаемые с достаточным уровнем качества
- НИР / ОКР от компаний
- с разграничением прав доступа

Аналоги, и чего в них не хватает

- <https://archive.ics.uci.edu/ml> — UC Irvine Machine Learning Repository
— есть данные, **нет контекстов, сообщества и образовательного контента**
- <https://ai-russia.ru/library>
— есть истории успеха, **нет сообщества, данных и образовательного контента**
- <https://www.kaggle.com>
— есть данные, контексты, сообщество, **нет образовательного контента**
- <https://ods.ai>
— есть данные, контексты, сообщество, курсы, **нет унификации и охвата**
- большинство университетских курсов
— есть образовательный контент, **нет открытости, охвата, унификации**
- биржи данных
— есть коммерческие данные, **нет образовательного контента**

Этапы внедрения

- Старт отбора датасетов (не прекращается на всём жизненном цикле проекта)
- Старт формирования сообщества авторов кейсов
- Унификация кейсов
 - отработка методологии, разработка шаблонов
 - доработка кейсов до уровня стандарта
- Прототипирование и разработка платформы депозитария
- Создание консорциума:
 - от бизнеса: альянс в сфере ИИ, ...
 - от образования: МГУ, Иннополис, РАНХиГС, МФТИ, ...
- Формирование и развитие экосистемы

Эффекты от внедрения инфраструктурного проекта DeroAI

- Интеграция бизнеса в образование через практические задачи ИИ
- Унификация взаимодействия компаний и вузов опосредованно через DeroAI
- Сертификация специалистов на платформе через решение практических кейсов
- Кратный рост числа преподавателей, образовательных программ, выпускников
- Поддержка проведения хакатонов, контестов, олимпиад
- Систематическая передача опыта от сильных к слабым

Спасибо за внимание!



Воронцов Константин Вячеславович

д.ф.-м.н., профессор РАН,
руководитель лаборатории
машинного обучения и семантического анализа
Института Искусственного Интеллекта МГУ

k.vorontsov@iai.msu.ru