

Искусственный интеллект и проверка ответов школьников: может ли нейросеть заменить экспертов ЕГЭ?

Юхина Нонна Юрьевна

Руководитель проектов ООО «ФИЗИКОН»

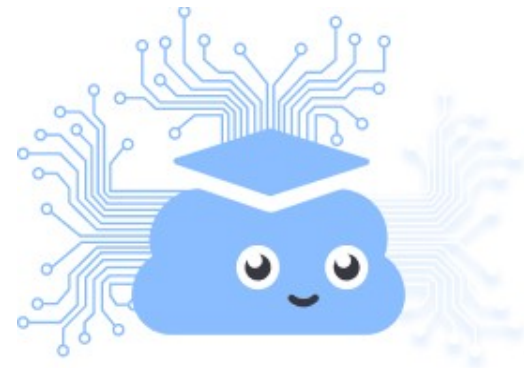
30.01.2024

–

31.01.2024

Применение нейросетей в образовании

- С помощью нейросетей можно решать следующие задачи:
 - **выстраивать индивидуальные траектории** обучения;
 - **персонализировать образование** под индивидуальную скорость обучения каждого обучающегося;
 - **анализировать учебные результаты и прогресс** в обучении;
 - **выявлять и устранять пробелы** в знаниях (адаптивное обучение);
 - **оценивать** правильность выполнения заданий;
- Современные чат-боты уже помогают в организации процесса обучения и отвечают на вопросы по самым разным дисциплинам.



Нейросети в образовании

- Нейросети – это **инструмент** и **помощник**.
- Нейросети могут выполнять озвучку и перевод на другой язык видеозаписей в режиме онлайн, отвечать на общеобразовательные вопросы и выполнять математические вычисления, собирать презентации, генерировать картинки по запросу и другое.
- Использование нейросетей может привести к неконтролируемым процессам в образовании:
 - несанкционированный **сбор персональных данных**;
 - показ **неприемлемого контента**;
 - уменьшение **роли учителя** в образовании;
 - ухудшение **качества критического мышления** у обучаемых.
- Необходимо осознать **потенциал** использования нейросетей в образовании и **разработать стратегии** и **правила** их применения.

Учитель и нейросети

- Нейросети – это **инструмент** и **помощник**.
- **Все знания доступны в Интернете**.
- Учитель перестает быть носителем уникальных знаний.
- Но **обучить, объяснить материал** ученикам и **заинтересовать** учёбой **лучше учителя** нейросети пока **не могут**.
- Ученик будет использовать нейросети для выполнения домашних заданий и в подготовке к контрольным работам.
- С развитием нейросетей становится **важным проверять** не то, что и как написано учеником, а его **понимание** написанного, **умение анализировать** и **дискутировать**.
- Появление нейросетей в образовании **нужно принять** и **использовать** в своей работе.

Онлайн-сервис «Облако знаний»

- **Онлайн-сервис «Облако знаний»** представляет собой систему онлайн-обучения, полностью интегрированную (технологически и методически) с размещённым в ней цифровым контентом.
- **В Федеральном перечне ЭОР.** Решение работает во всех школах Москвы и Московской области.
- Цифровой контент онлайн-сервиса «Облако знаний» можно использовать **тремя способами**:
 - учитель может **назначать** цифровые работы;
 - ученик может **заниматься самостоятельно**;
 - функцию учителя может взять электронный помощник Оз и **подобрать программу обучения.**



Формы деятельности в онлайн-сервисе

- **Фронтальная работа у доски** (интерактивные презентации)
- Автоматизированные **самостоятельные работы** (самостоятельные и контрольные работы)
- **Лабораторные работы** в компьютерном классе и дома (виртуальные практикумы)
- **Подготовка к экзаменам и олимпиадам** (задания повышенной сложности, подготовка к ЕГЭ, ОГЭ, ВПР)
- **Аттестация: контрольные работы** (тематические контрольные работы, задания для подготовки к ГИА)
- **Самостоятельное повторение материала** (опорные конспекты)

Возможности пользователя в роли Учитель

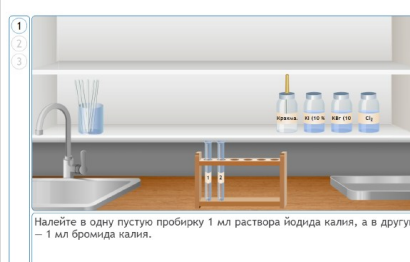
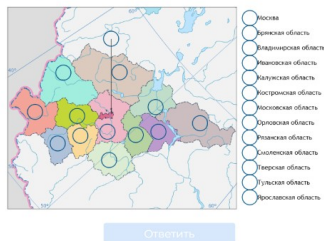
- Учитель имеет **бесплатный доступ** ко всему контенту (и базового, и углубленного уровней) без ограничения времени и может:
 - **просматривать** работы (в том числе решения и ответы);
 - **назначать** работы в своих классах (работы углублённого уровня и ЕГЭ/ОГЭ будут доступны только ученикам, имеющим подписки «Плюс» или «Школа»);
 - **проверять и анализировать** результаты выполнения работ в своих классах.



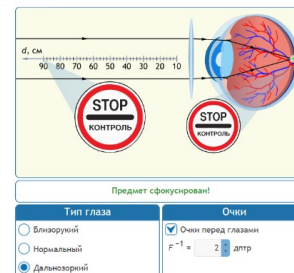
Примеры работ



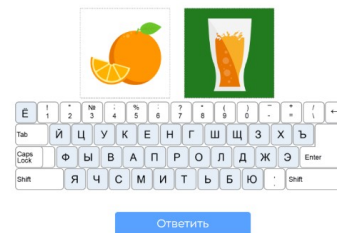
Укажите на карте расположение субъектов Российской Федерации в составе Центрального экономического района.



Налейте в одну пустую пробирку 1 мл раствора йодида калия, а в другую – 1 мл бромида калия.

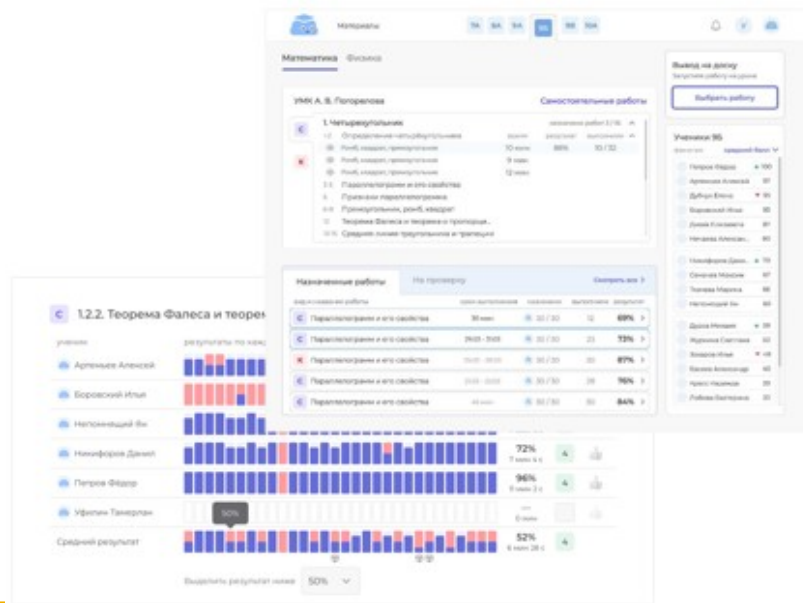


Щёлкните на изображении клавиатуры клавишу, с помощью которой можно к выделенной области добавить ещё одну выделенную область в Adobe Photoshop.



Проверка и анализ результатов учеников

- Учитель **организовывает** обучение учащегося, предлагает ему контент и следит за эффективностью обучения.
- Когда ученик выполнит задание, учитель может при необходимости **проверить** его (например, если это задание с открытым ответом).
- Учитель **анализирует**, как ученик выполнил то или иное задание, и **корректирует** на основе этого траекторию обучения.

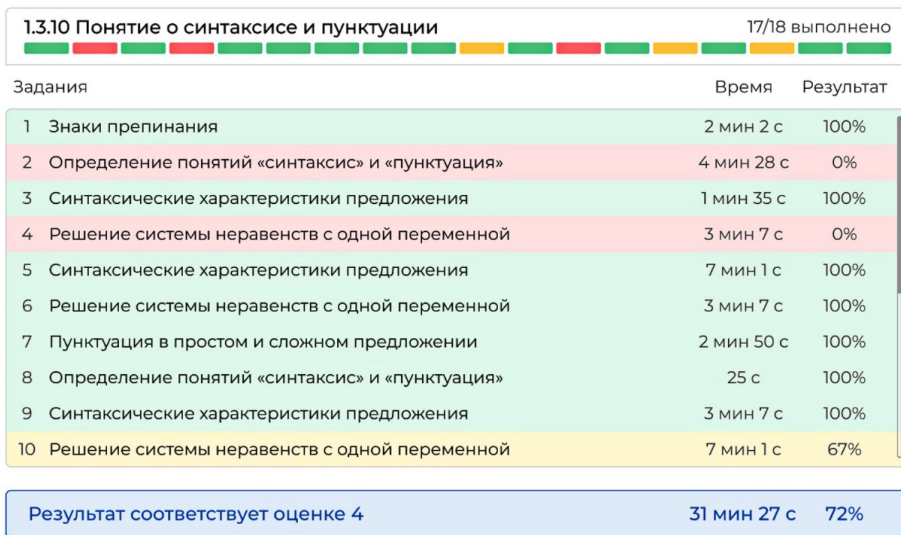


Возможности пользователя в роли Ученик

- Для ученика предусмотрено **два варианта подписки**: бесплатная и «Облако знаний. Плюс».
- Режимы работы ученика:
 - **выполнение назначенных** учителем работ;
 - **самоподготовка**;
 - **подготовка под управлением компьютера** (индивидуальные траектории).
- Подписка «Облако знаний. Плюс» даёт ученику доступ:
 - к работам **углубленного уровня**, а также по подготовке к **ЕГЭ и ОГЭ**;
 - к работам базового уровня **без ограничения времени**.



- Ученику результаты выполнения своей работы:
 - затраченное время;
 - насколько верно он выполнил задания.



Подготовка к ЕГЭ

- Учитель:
 - **назначает** работу ученикам;
 - **проверяет** развернутые ответы и анализирует результаты.
- Ученик:
 - **назначает** работу ученикам;
 - **проверяет** развернутые ответы и анализирует результаты.



Тренировочные наборы
к заданиям ЕГЭ



10 экзаменационных
вариантов КИМ



Иллюстрированные
конспекты



Задания обновляются по
свежим спецификациям
ФИПИ

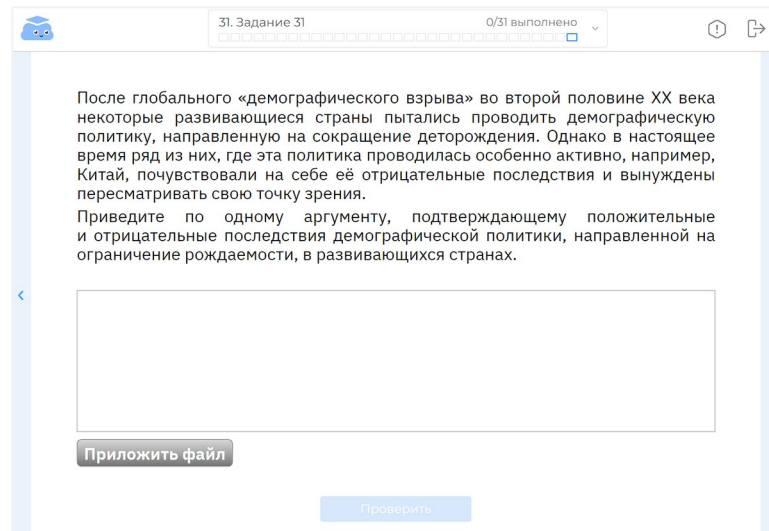
Задания с проверкой развернутых ответов

- Формы ввода ответа в интерактивных заданиях могут быть закрытыми или открытыми. **Открытые формы** подразумевают ввод краткого ответа или ввод развернутого ответа.
- **Ввод развернутого ответа** широко используется в диагностических измерениях, так как позволяет измерить более сложные компетенции, чем краткий или закрытый ответ (ЕГЭ и ОГЭ).
- Развёрнутые ответы в настоящее время проверяются квалифицированными **экспертами**. В ответственных процедурах каждый ответ проверяется 2 экспертами.
- При масштабных тестированиях проверка ответа требует большого количества экспертов и занимает много времени.



Задания с развернутым ответом в ЕГЭ

- В сервисе «Облако знаний» для ответа на задание ученик может **ввести символы в поле ввода**, либо **приложить фотографию** со своим ответом.
- Оценить ответ с помощью **критериев** может только учитель.



31. Задание 31 0/31 выполнено

После глобального «демографического взрыва» во второй половине XX века некоторые развивающиеся страны пытались проводить демографическую политику, направленную на сокращение деторождения. Однако в настоящее время ряд из них, где эта политика проводилась особенно активно, например, Китай, почувствовали на себе её отрицательные последствия и вынуждены пересматривать свою точку зрения.

Приведите по одному аргументу, подтверждающему положительные и отрицательные последствия демографической политики, направленной на ограничение рождаемости, в развивающихся странах.

Приложить файл

Проверить

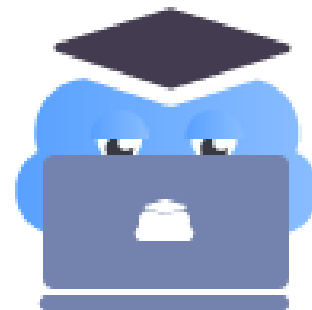
Проверка развернутых ответов учителем

- Учитель **проверяет** развернутые ответы ученика и **соотносит** их с представленными критериями.
- После оценивания ответа по критериям отметка за задание отправляется на сервер и объединяется с общей отметкой за всю работу.

The screenshot displays the interface for checking a student's answer. At the top, the student's name "Никифоров Данил" and progress "2/8 проверено" are visible. The task text is "Ландшафты бывают разные. На планете Земля их великое множество." with an image "Экзопланета.jpg". The evaluation criteria are listed on the right: "Задания №31 Тренировка по ЕГЭ-2019 Взаимодействие заряженных тел после соприкосновения". The student's answer is shown in a box with a score of 3 points. The criteria for 3 points are: "Приведено полное решение, включающее следующие элементы: • Записаны положения теории и физические законы, закономерности, применение которых необходимо для решения задачи выбранным способом. • Описаны все вводимые в решение буквенные обозначения физических величин. • Проведены необходимые математические преобразования (допускается вербальное указание на их проведение) и расчёты, приводящие к правильному числовому ответу (допускается решение «по частям» с промежуточными вычислениями). • Представлен правильный ответ с указанием единиц измерения искомой величины." The criteria for 2 points are: "Правильно записаны необходимые положения теории и физические законы, закономерности, проведены необходимые преобразования и представлен правильный ответ с указанием единиц измерения искомой величины. Но имеется один из следующих недостатков: • Записи, соответствующие одному или обоим пунктам: 2 и 3 – представлены не в полном объеме или отсутствуют. • При ПОЛНОМ правильном решении лишние записи, не входящие в решение (возможно, неверные), не отделены от решения (не зачёркнуты, не заключены в скобки, рамку и т. п.). • При ПОЛНОМ решении в необходимых математических преобразованиях или

Проверка развернутых ответов компьютером

- Задача – **автоматизировать проверку** заданий с развернутым ответом **с помощью искусственного интеллекта.**
- Этапы выполнения задачи:
 - **отбор заданий** с развёрнутым ответом;
 - **формирование элементарных критериев** для каждого задания;
 - **сбор, очистка данных** (ответов обучающихся);
 - **проверка данных методистом;**
 - **выбор модели нейросети;**
 - **обучение нейросети;**
 - **анализ эффективности, дообучение.**



Отбор заданий

- Задания должны предполагать:
 - **открытую форму ввода** ответа (ввод развернутого ответа);
 - **текстовые ответы**, не содержащие:
 - рисунки;
 - графики, чертежи;
 - формулы;
 - таблицы;
 - коды программ;
 - вычисления.
- Для внедрения проверки с помощью технологий ИИ хорошо подходят задания **гуманитарных предметов** (географии, биологии, истории, обществознания и проч.).
- Для проекта уже отобрано **70 заданий** по программам подготовки к ОГЭ и ЕГЭ.

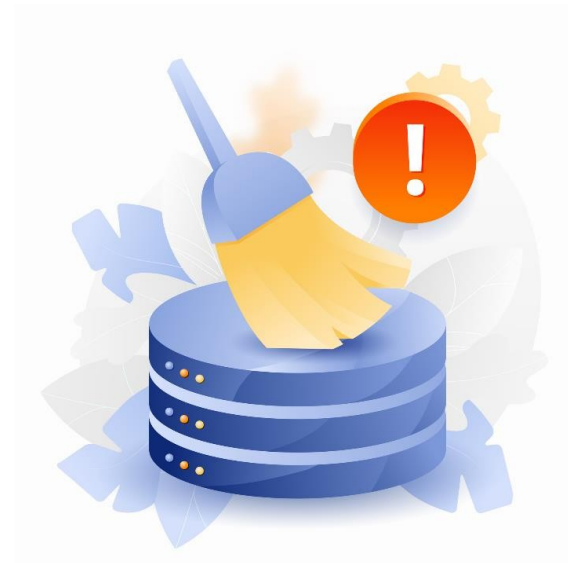


Формирование элементарных критериев

- Для обучения нейросети авторское решение по каждому из заданий разбивается на **составные части** - элементарные критерии, так искусственному интеллекту будет проще оценить ответ и выдать более точный результат.
- Элементарные критерии формируются **методистами-предметниками** с опорой на критерии оценивания, **соответствующие демоверсии КИМ**.
- У каждого задания **разное количество** элементарных критериев (у нас от 1 до 14).
- Итоговый балл за ответ рассчитывается на основе отметок, полученных по каждому элементарному критерию.
- Формула верного ответа (2 балла) для одного из заданий:
(crit_1 = false AND crit_2 = false AND crit_5 = false) AND (crit_3 = true AND crit_4 = true) AND (crit_6 = true OR crit_7 = true OR crit_8 = true)

Сбор и очистка данных

- Этап выполнения задачи делится на **следующие шаги**:
 - **сбор** доступных ответов, полученных естественным путем;
 - **очистка** данных.
- В данный момент собрано **около 50 тыс. ответов**.
- **Очистка данных** от заведомо некорректных:
 - пустые ответы;
 - ответы из 1 символа;
 - ответы, введенные на английской раскладке;
 - и другие.



Проверка ответов

- **Проверка ответов** методистом:
 - Каждый сложный ответ разбивается на несколько простых, которые проверяются на соответствие элементарным критериям, выделенным в задании: «Да» проставляется в случае соответствия критерию, «Нет» – при несоответствии.
 - Оценивается корректность ответа в целом (верный / частично верный / неверный).
- Переданные по каждому заданию ответы делятся на **две группы**: тренировочную и тестовую.



Выбор модели и обучение нейросети

- **Наиболее высокие метрики** показали модели нейросетей:
 - модель глубинного обучения ruBERT;
 - логистическая регрессия без аугментации.
- После анализа результатов проведенных исследований выбор пал на нейросеть модели **ruBERT**.
- В целях обучения нейросети был сформирован пакет с ответами, проверенными и промаркированными методистами.

OOF	423340	423314	423318	423288	
ind, ml, logreg emb	0,616438	0,471698	0,633333	0,5	0,555367
ind, ml, logreg tfidf	0,424658	0,509434	0,6	0,676471	0,552641
global, ml, logreg emb	0,726027	0,796226	0,6	0,647059	0,692328
global, ml, tinybert	0,534247	0,339623	0,6	0,5	0,493468
global, binary, logreg emb	0,945205	0,716981	0,633333	0,941176	0,809174
global, binary, tinybert	0,465753	0,358491	0,633333	0,5	0,489394
IR	0,69863	0,815094	0,633333	0,558824	0,67647
TEST	423340	423314	423318	423288	
ind, ml, logreg emb	0,6875	0,318182	0,461538	0,333333	0,450138
ind, ml, logreg tfidf	0,46875	0,409091	0,461538	0,466667	0,451512
global, ml, logreg emb	0,83125	0,718182	0,684615	0,4	0,658512
global, ml, tinybert	0,5	0,318182	0,384615	0,2	0,350699
global, binary, logreg emb	0,9375	0,590909	0,461538	0,866667	0,714154
global, binary, tinybert	0,5	0,318182	0,461538	0,2	0,36993
IR	0,76875	0,618182	0,684615	0,6	0,667887

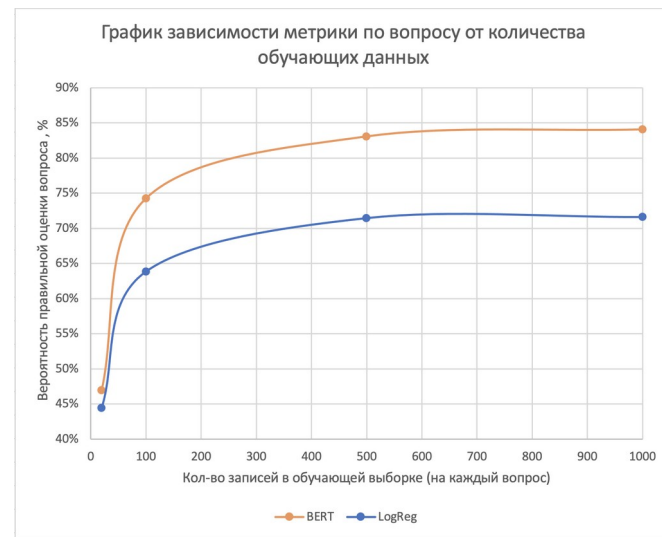
Анализ эффективности Дообучение

- При тестировании нейросеть показала **высокие результаты**.
- Но были выявлены недостатки:
 - нейросеть плохо оценивает **сложно сформулированные ответы**, например, ответы с двойным отрицанием; на оценку влияет наличие определенных слов, но не их **порядок в предложении**. Если в правильном ответе переставить местами слова «больше»/«меньше», нейросеть по-прежнему будет считать ответ верным.
- Нейросеть будет дообучаться на новых ответах.

Предмет	Вероятность правильной оценки критерия	Вероятность правильной оценки задания в целом
Биология	0.9270	0.9204
География	0.9575	0.9315
История	0.9669	0.9425
Обществознание	0.9461	0.8899

Особенности обучения моделей нейросетей

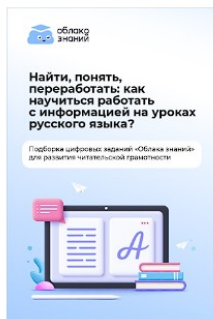
- **Эффективность модели** во многом зависит от простоты предметной области. Обучить ИИ обществу сложнее, чем географии.
- Обучение обособленных экземпляров для каждого предмета **не показало** статистически значимого **улучшения метрик** (1-2% в качестве, но при этом 1 модель занимает в 4 раза меньше места в оперативной памяти).
- Требуется **более 1000 ответов** на каждое задание.
- Обучение может **подкрепляться предметной базой** – комплектом опорных конспектов по предметам.



Ожидаемые результаты

- Успешное выполнение задачи позволит **автоматизировать** проверку развернутых ответов в текстовой форме на задания (преимущественно гуманитарных предметов).
- Способы использования:
 - прямое **использование учеником**:
 - когда нет учителя;
 - когда учитель проверяет слишком долго, а приблизительный результат хочется узнать пораньше;
 - **использование учителем**:
 - помощник при проверке развернутых ответов (предварительная проверка);
 - проверка качества своей работы (сравнение результатов своей проверки с результатами проверки ИИ)

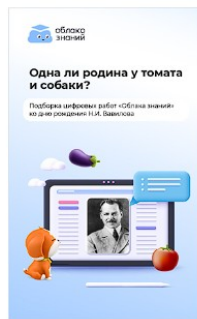
Сервис «Облако знаний» представлен в трёх социальных сетях: **Telegram**, **ВКонтакте** и **Дзен**



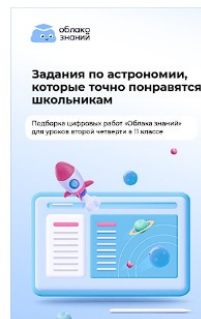
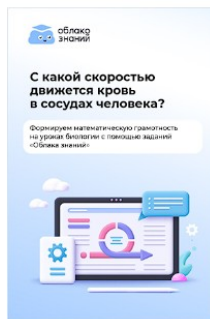
<https://t.me/oblakoz>



<https://vk.com/oblakoz>



<https://dzen.ru/oblakoz>





**СПАСИБО
ЗА ВНИМАНИЕ!**