

Использование технологий генеративного искусственного интеллекта в качестве ассистента геолога на примере оценки вероятности открытия месторождений УВС (gCoS)

Затраты на геолого-разведочные работы (ГРР) в России в реальном выражении снижаются, а объемы поисково-разведочного бурения уменьшаются. При этом средний объем извлекаемых запасов нефти открываемых месторождений составляет всего 4,2 млн т. Несмотря на это, в традиционных и новых, слабо изученных регионах сохраняется значительный ресурсный потенциал для открытий.

Для оценки перспективных объектов геологи тратят много времени на поиск данных и их анализ для субъективной оценки шансов геологического успеха (gCoS). Использование генеративного искусственного интеллекта (ГенИИ) должно повысить эффективность их работы и результативность ГРР. Для этого нами был разработан MVP GeoAssistant, включающий модули поиска релевантных исследований, предобработки исходных данных, извлечения структурированных данных и сравнения качества ответов различных БЯМ.

Бенчмаркинг показал, что коммерческие иностранные языковые модели лучше справляются с извлечением данных, чем опенсорсные. Доля корректного извлечения значений параметров, необходимых для расчета Р_{нм} (одного из факторов gCoS) у лучшей модели превысила 68%, а у худшей - 44%. Российские модели пока не пригодны для решения этой задачи. При определенной доработке ГенИИ-системы могут успешно применяться для создания структурированных баз данных, в качестве ассистентов эксперта и для генерации отчетов в секторе разведки и добычи.

Предисловие



ВЫГОН Григорий Вадимович
Управляющий директор, к.э.н.

Технологии генеративного искусственного интеллекта (ГенИИ) стремительно развиваются, однако вопросы его успешного практического применения в промышленности, в частности, нефтегазовой отрасли ГенИИ, еще остаются.

В инженерных задачах важно быстро обработать большой массив информации и собрать достоверные данные для дальнейшего анализа, например, для подбора аналогов, оценки вероятности открытий или объема запасов. Поэтому именно задача экстракции данных из геологических отчетов была нами выбрана для R&D в сфере ГенИИ, результаты которого приведены в данном отчете.

Качественное извлечение данных (например результатов измерений и оценок параметров пластов) из разных источников в их привязке к геологической онтологии (образец, бассейн, формация и тп), является нетривиальной задачей. В частности, предобработка исходных данных в процессе распознавания файлов в формате pdf, а также нахождения онтологических взаимосвязей между объектами, разбросанными по тексту и различным таблицам сильно влияют на качество извлечения. Эти вопросы мы решали при создании MVP ассистента геолога (GeoAssistant) для оценки геологического шанса успеха (gCoS) .

В России, по данным Роснедр, ежегодно открывается около 50 месторождений УВС, причем средняя величина технически извлекаемых запасов нефти открытого месторождения постепенно снижается и составляет всего около 4 млн т. При этом есть ряд новых регионов (глубокие горизонты Западной Сибири, Таймыр, Восточная Сибирь, шельфы северных морей и Дальнего востока), где открытия могут быть значительными. В виду их недостаточной изученности, инструмент для gCoS на базе ГенИИ может быть чрезвычайно полезен.

В перспективе ассистент геолога может использоваться, как частными нефтегазовыми компаниями, так и подведомственными структурами Роснедр (Росгеолфонд, ГКЗ, институты) для формирования структурированных баз данных на основе значительного объема накопленной геологической информации и создания инструментов экспертизы геологических отчетов, ПТД, а также для приоритизации региональных ГРП. В частности, полезной выглядит возможность использования ГенИИ для подбора месторождений-аналогов, технико-экономических параметров, оценки ресурсов. Эту гипотезу мы планируем проверить и опубликовать результаты до конца 2024 года.

Ключевые выводы: GenAI может извлекать структурированные отраслевые данные для оценки gCoS и использоваться в других задачах сегмента Рид

1. Затраты на ГРП в реальном выражении **снижаются**, как и **объемы поисково-разведочного бурения**, а **средний размер извлекаемых запасов нефти** открываемых месторождений **составляет всего 4,2 млн т**. При этом как в **традиционных**, так и в **новых**, слабо изученных регионах, **сохраняется значительный ресурсный потенциал** для новых открытий.
2. При проведении геологоразведки требуется **оценка перспективных объектов по шансам геологического успеха (gCoS)**. Геологи тратят **много времени на поиск данных и их анализ для субъективной оценки gCoS**. Использование **ГенИИ** должно **повысить эффективность их работы и результативность ГРП**.
3. Для **проверки этой гипотезы** было **разработано решение на базе ГенИИ (GeoAssistant)**, включающее следующие **модули: поиск** релевантных исследований, **предобработка** исходных данных, **извлечение** структурированных данных и **сравнение** (бенчмаркинг) качества извлечения **различных БЯМ**.
4. **Качество распознавания документов в формате PDF является ключевым фактором**, влияющим на **результат извлечения структурированных данных**. В **GeoAssistant** была **разработана схема данных**, а также **выгружены из интернета и обработаны статические данные с геологической информацией**: нефтегазоносные бассейны, формации и геохронология.
5. **Экстракция данных** проводилась по **более, чем 50 геологическим статьям**: определялись **16 типов параметров у 830 образцов**. **Всего база данных содержит 900 формаций 37 бассейнов Индии**, выставленных в рамках Программы лицензирования открытых площадей Индии (Open Acreage Licensing Program, OALP).
6. **Сложность экстракции** геоинформации заключается в **нахождении взаимосвязей между объектами**, разбросанными по тексту и **различным таблицам**. **Проведенная работа** позволила определить и классифицировать **основные виды ошибок при предобработке и экстракции данных с помощью БЯМ**.
7. **С помощью продукта GeoAssistant** для выбранного ЛУ в Индии были **отобраны подходящие параметры НГМП** - ожидается богатое **ОВ ранней зрелости с III типом керогена (генерация газа)**, **значение параметра Р_{нм}** (одного из пяти факторов gCoS) **оценено в 0,95**
8. Как показал **бенчмаркинг**, **коммерческие иностранные БЯМ лучше справляются с извлечением данных, чем модели с открытыми весами**. **Доля корректного извлечение значений 4х параметров образца, названий бассейнов и формаций, необходимых для расчета Р_{нм}**, у лучшей модели из выборки, **Claude 3.5 Sonnet, превысила 68%**, а у худшей (**Llama 3.1 70 B**) - **составила около 44%**. **Российские БЯМ пока с этой задачей плохо справляются и в бенчмаркинг не попали в силу недостаточного размера контекстного окна и плохого следования инструкциям**.
9. При определенном развитии **ГенИИ-системы** могут **успешно применяться в задачах по созданию структурированных (реляционных, графовых и др.) баз данных из неструктурированных данных**, в качестве **Ассистентов эксперта** и для **генерации отчетов в секторе разведки и добычи** профильными **органами государственной власти, нефтегазодобывающими и нефтесервисными компаниями, НТЦ**.



ЭФФЕКТИВНОСТЬ ГРР И РОЛЬ ШАНСА ГЕОЛОГИЧЕСКОГО УСПЕХА



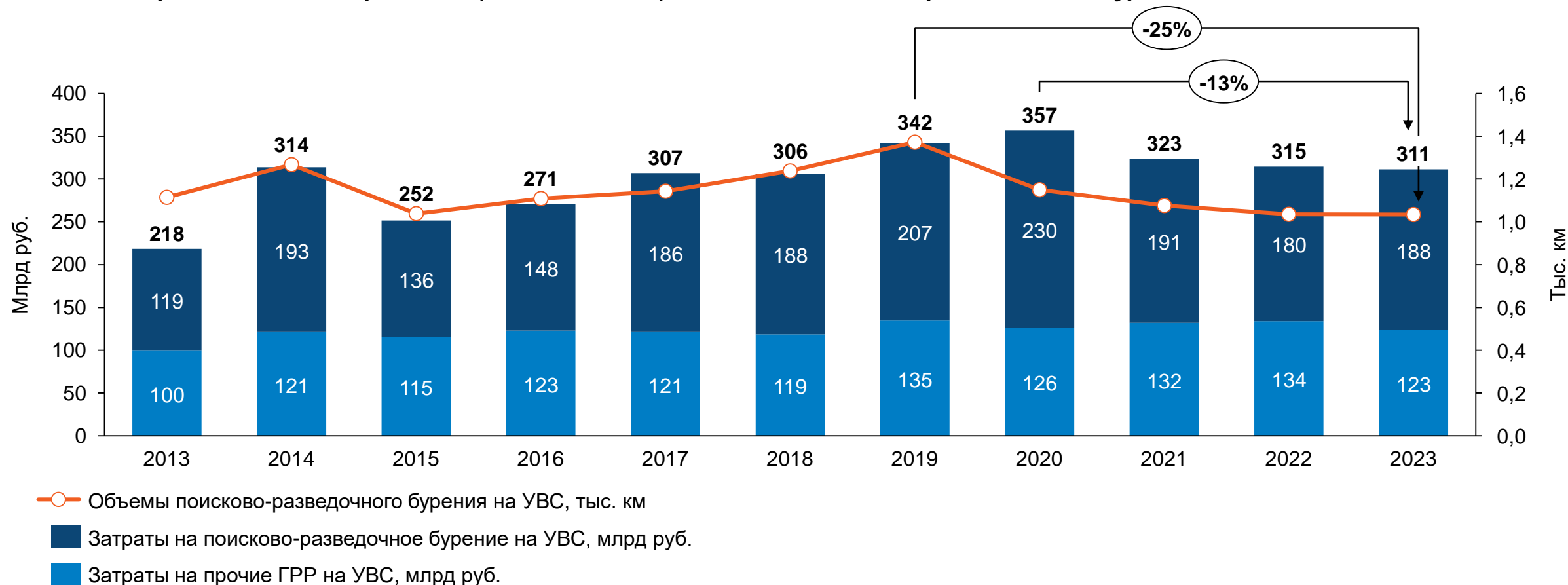
ОПИСАНИЕ РЕШЕНИЯ НА БАЗЕ ГЕНЕРАТИВНОГО ИИ



ОЦЕНКА РЕЗУЛЬТАТОВ И ПЕРСПЕКТИВЫ ПРИМЕНЕНИЯ

Затраты на ГРП в России за 10 лет находятся в диапазоне 220-340 млрд руб. , объемы поисково-разведочного бурения с 2019 г. снизились на 25%

Динамика затрат на ГРП на нефть и газ (в ценах 2023 г.) и объемы поисково-разведочного бурения на УВС в 2013-2023 гг.



В период 2013-2019 гг. наблюдалась отчетливая корреляция между затратами на геологоразведку и объемами поисково-разведочного бурения. С 2020 года объемы бурения начали резко снижаться, значительно опережая падение затрат на ГРП. Средняя стоимость бурения 1 метра поисково-разведочной скважины в реальном выражении выросла за 10 лет на 53% и составила в 2023 году - 162,2 тыс. руб./м.

Показатели прироста извлекаемых запасов **демонстрируют стагнацию:** **средний размер открываемых м/р за последнее десятилетие – 4,2 млн т,**

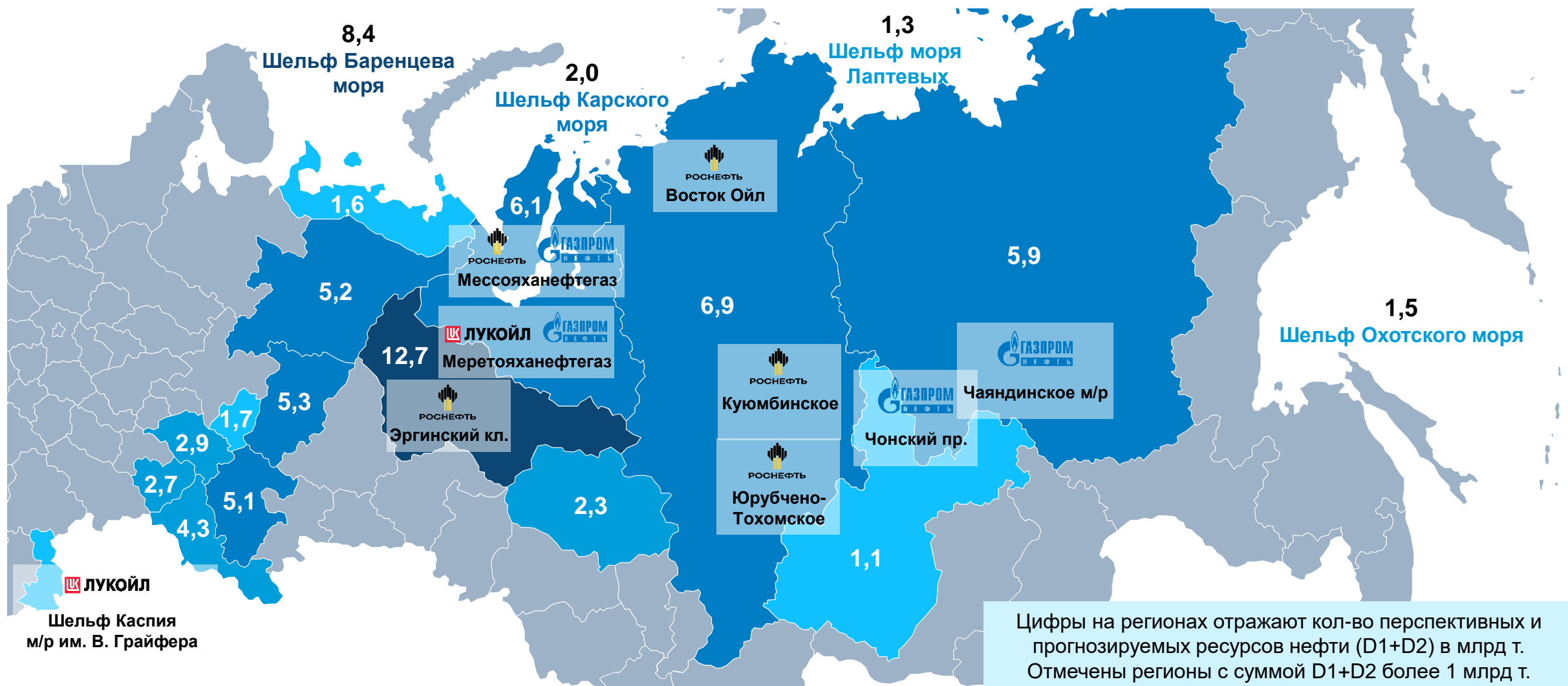
Динамика прироста запасов нефти за счет новых м/р и разведки (категории А+В1+С1) за 2013-2023 гг.



- Исключая аномальный 2020 год, прирост запасов нефти каждый год составлял от 43 до 263 млн т, при среднем значении в **123 млн т**.
- Средний размер ТИЗ новых м/р за последние годы – **4,2 млн т**

Как в традиционных, так и в новых, слабо изученных регионах, сохраняется значительный потенциал новых открытий

Фрагмент карты основных проектов освоения м/р нефти в РФ и перспектив открытия новых м/р

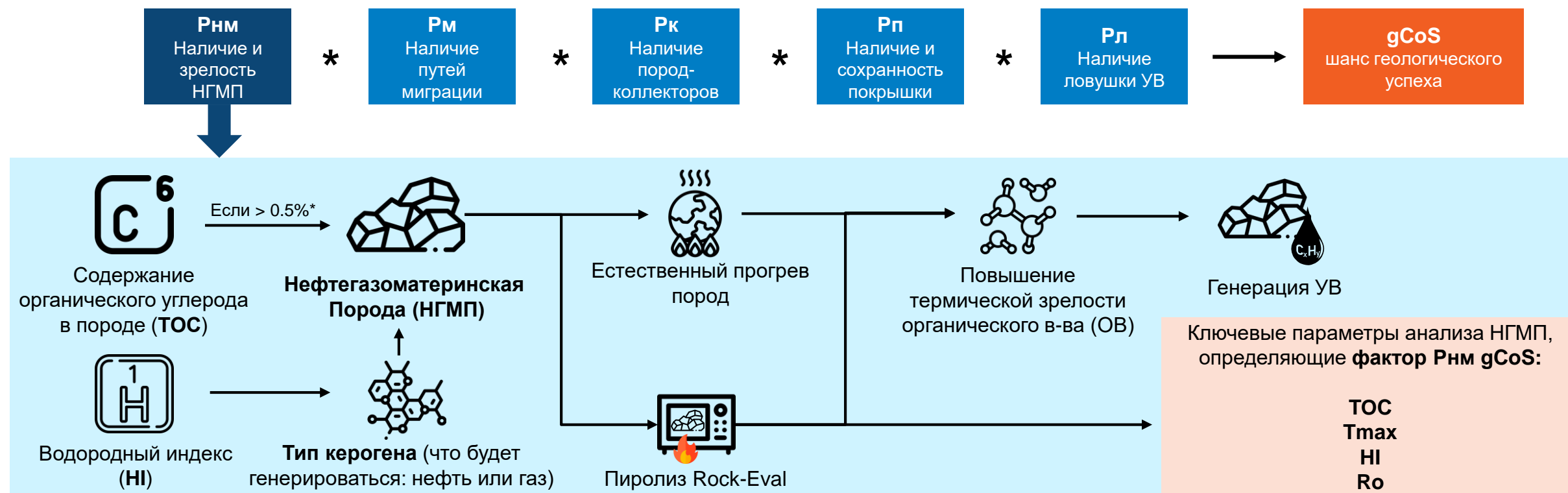


Для проведения **эффективной геологоразведки** требуется оценка перспективных объектов по **шансам геологического успеха**

Методика оценки шансов геологического успеха ($gCoS$, $K_{г\gamma}$)

Шанс геологического успеха ($gCoS$, $K_{г\gamma}$) – это вероятность открытия поисковой скважиной залежи УВ с достаточными объемами для получения устойчивого притока УВ на поверхность.

- Величина $gCoS$ может изменяться **от 0 до 1** и характеризует уверенность эксперта в открытии залежи;
- $gCoS$ **вычисляется для каждой ожидаемой залежи отдельно**, после чего может суммироваться по правилам теории вероятностей;
- $gCoS$ рассчитывается как **произведение 5-ти независимых геологических факторов**, определяющих успех или неудачу ГРП.



* Для карбонатных пород встречается граничное значение НГМП в 0,3%

Геологи тратят много времени на поиск данных и их анализ для субъективной оценки gCoS, использование ГенИИ должно повысить эффективность

Преимущества использования ГенИИ для решения задач нефтегазовой отрасли (на примере оценки gCoS)

Как сегодня производится оценка gCoS

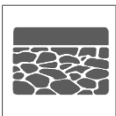
- Команда геологов оценивает вероятности каждого из параметров нефтегазовой системы исходя из **субъективного взгляда специалистов и степени их знаний**
- Оценка каждого фактора gCoS требует **анализа большого объема данных** для рассматриваемого региона и **сопоставления с мировыми аналогами**, что человек **делает долго и с ошибками**

Преимущества от внедрения ГенИИ в ГРП

- **Повышение эффективности работы** геологов, геофизиков и петрофизиков с большими неструктурированными данными из геологических отчетов, подсчетов запасов, книг, статей и т.д.
- **Возможность экспресс-оценки перспективности участков** на основе автоматического расчета показателя gCoS
- **Повышение эффективности ГРП** за счет приоритезации геологических объектов



- **Снижение степени субъективности** оценки gCoS, **автоматизация сбора, обработки и анализа** геологических данных **повысит эффективность ГРП**
- **Внедрение ГенИИ** в геологоразведку позволит **существенно улучшить работу с большими объемами геологической информации**, позволив специалистам-геологам оптимизировать свою работу, делегировав рутинные задачи на ИИ



ЭФФЕКТИВНОСТЬ ГРР И РОЛЬ ШАНСА ГЕОЛОГИЧЕСКОГО УСПЕХА



ОПИСАНИЕ РЕШЕНИЯ НА БАЗЕ ГЕНЕРАТИВНОГО ИИ



ОЦЕНКА РЕЗУЛЬТАТОВ И ПЕРСПЕКТИВЫ ПРИМЕНЕНИЯ

Для проверки гипотезы **возможности использования ГенИИ** в **прикладных задачах нефтегазовой геологии** нами было разработано решение на базе БЯМ

Цели и задачи исследования

ЦЕЛЬ ИССЛЕДОВАНИЯ:

Продемонстрировать возможности использования технологии ГенИИ в геологии на примере задачи экстракции данных из геологических отчетов для определения факторов gCoS (шанса геологического успеха)

ЗАДАЧИ ИССЛЕДОВАНИЯ:

- Оценить качество обработки неструктурированных данных из геологических отчетов
- Оценить точность и полноту экстракции онтологически связанных данных
- Вычислить параметр Рнм (один из факторов для расчета gCoS)
- Провести бенчмаркинг БЯМ

ОБЪЕКТ ИССЛЕДОВАНИЯ:

ЛУ MB-OSHP-2023/1

Географическое положение – **шельф Камбейского залива**

Нефтегазоносный бассейн – **бассейн Мумбайского шельфа**

Суббассейн - **суббассейн депрессии Сурат**

Площадь участка – **2 935 км²**

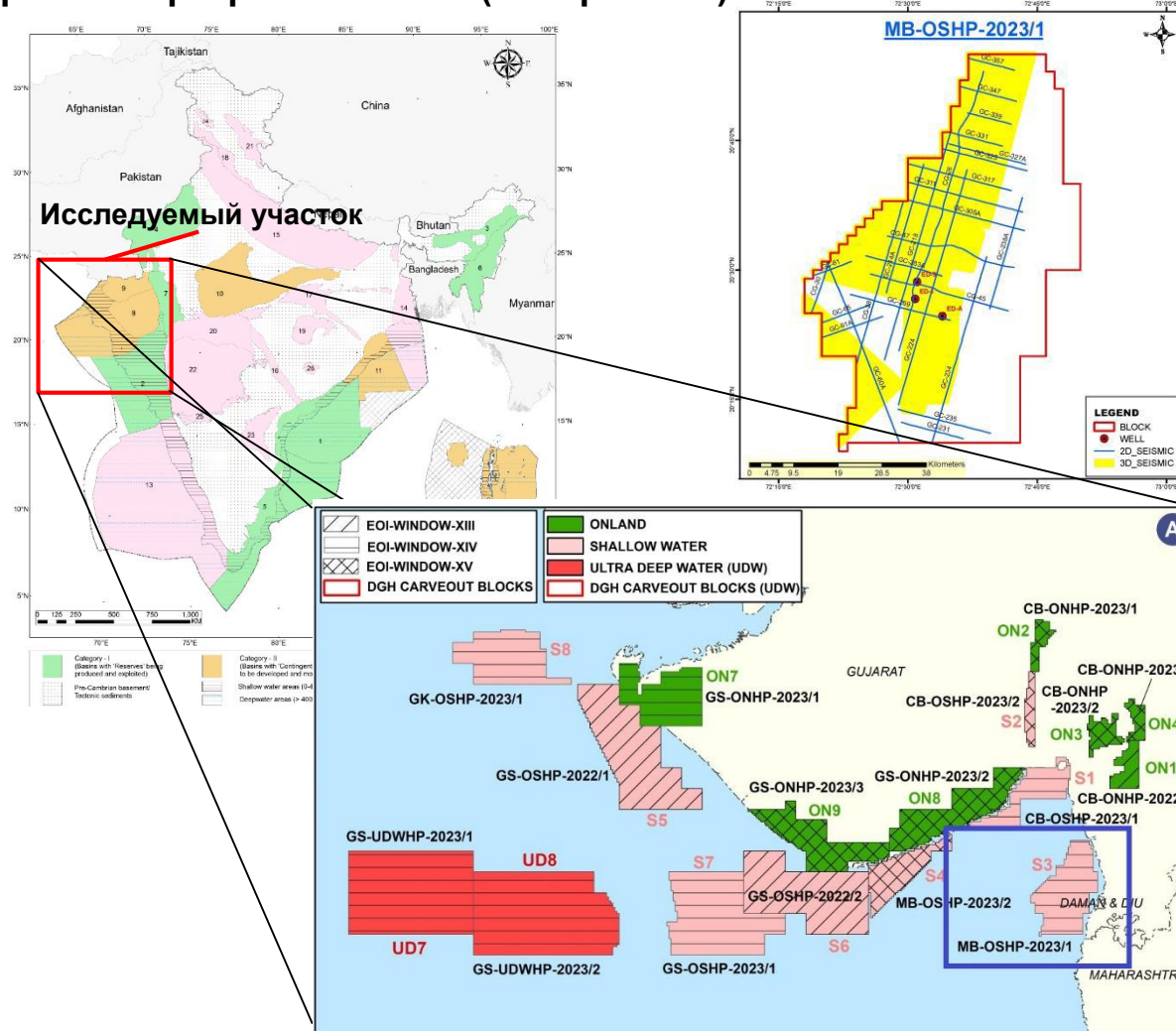
Покрытие сейсмикой 3D – **1 183 км²**

ИНСТРУМЕНТ ИССЛЕДОВАНИЯ:

Разработанное решение на базе ГенИИ, включающее следующие модули:

- **Поиск** релевантных исследований
- **Предобработка** исходных данных
- **Извлечение** структурированных данных
- **Сравнение** качества ответов различных БЯМ

Карта нефтегазоносных бассейнов Индии и ЛУ выставленных в рамках Программы OALP (январь 2024)



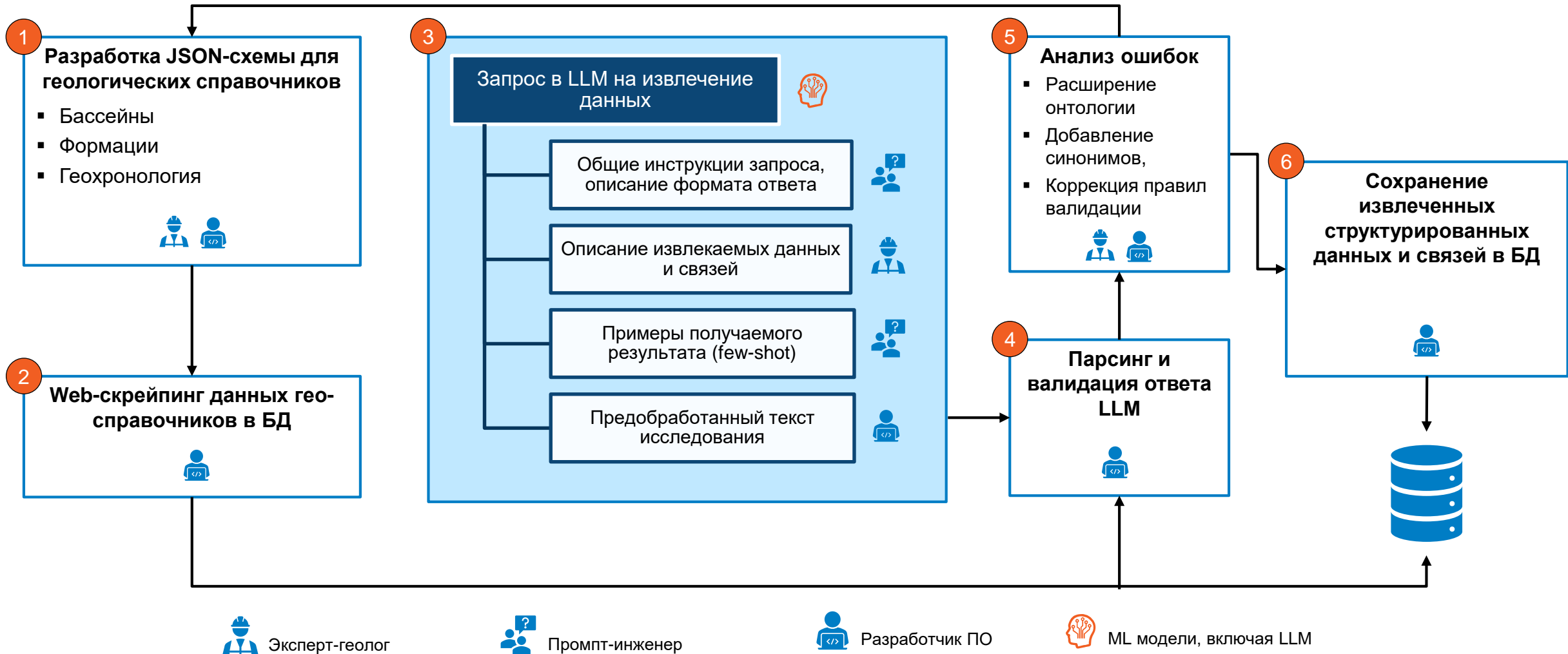
Качество распознавания документов в формате PDF является **ключевым фактором**, влияющим на **результат извлечения** структурированных данных

Поиск и препроцессинг неструктурированных данных геологических отчетов



Для **качественного извлечения** данных была **разработана JSON-схема** и **обеспечено взаимодействие экспертов с различными компетенциями**

Извлечение данных (extraction)



Для качественной работы модели были сформированы JSON-файлы с геологической информацией: нефтегазоносные бассейны, формации и геохрон

Отрывок JSON-шаблона геохронологической шкалы

```
{
  "name": "Bashkirian",
  "stratigraphic_type": "Stage",
  "geochronological_type": "Age",
  "beginning_date_ma": 323.2,
  "ending_date_ma": 315.2,
  "parent": "Pennsylvanian",
  "children": []
},
```

Carboniferous	Pennsylvanian	Upper	Gzhelian	303.7 ±0.1
			Kasimovian	307.0 ±0.1
		Middle	Moscovian	
		Lower	Bashkirian	315.2 ±0.2 323.2 ±0.4
	Mississippian	Upper	Serpukhovian	330.9 ±0.2
		Middle	Visean	346.7 ±0.4
		Lower	Tournaisian	358.9 ±0.4

Отрывок JSON-справочника по формациям Индии

```
{
  "name": "Panna Fm",
  "period": "Paleogene",
  "age_interval": "Paleocene to earliest Eocene",
  "type_locality": "Well Bassein \u2013 1, (depth interval 2070 \u2013 2145 m)",
  "lithology": {
    "description": "Sandy claystone. Represented by conglomerate and sandstones at the basal part, overlain by a section of coal-shale alternation in the middle and succeeded by shale section at the top.",
    "pattern": "Sandy claystone"
  },
  "fossils": "Top of the Panna Fm coincides with top of the Nummulites.",
  "age": {
    "description": "Mumbai stratigraphic chart indicates early Danian through mid-Ypresian",
    "age_span": {
      "beginning_stage": "Danian",
      "fraction_up_in_beginning_stage": 4.0,
      "beginning_date_ma": 6429.0,
      "ending_stage": "Ypresian",
      "fraction_up_in_ending_stage": 5.0,
      "ending_date_ma": 5204.0
    }
  }
},
```

В результате **экстракции данных** из статей **определялись параметры образцов** формаций и бассейнов, необходимых для **расчета gCoS**

Интерфейс окна извлеченных параметров образцов ПО GEOASSISTANT и статистика по обработанному датасету

PreprocessingPapersBasinsFormations**Samples**EvaluationsSettings

Filter by Basin...

Filter by Paper...

Filter by Lithology...

Filter by TOC...

Columns

Formation	Basin	Paper	Depth	Lithology	Th	Tmax	TOC	HI	OI	PI	PY
Panna	Mumbai Offshore		2375	Shale, Limestone	122	428	31.86	196.7	5.96	0.03	67.07
Panna	Mumbai Offshore		2410	Siltstone	126	435	4.95	117.6	18.4	0.03	5.98
Panna	Mumbai Offshore		2460	Siltstone, coal	148	430	33.25	243.5	5.5	0.05	85.46
Panna	Mumbai Offshore		2620	Siltstone, Shale, Limestone	105	437	3.06	61.43	107	0.03	1.94
Panna	Mumbai Offshore		2690	Siltyshale, Siltstone	132	440	1.63	14.72	116	0.07	0.26
Panna	Mumbai Offshore		2990	Shale	110	446	3.82	89.26	28	0.05	3.59
Panna	Mumbai Offshore		3060	Sandstone, Shale	120	447	1.36	63.97	107	0.16	1.04
Panna	Mumbai Offshore		3070	Siltstone, Shale	132	436	1.41	24.82	100	0.08	0.38
Panna	Mumbai Offshore		3100	Sandy siltstone, Siltstone	134	442	1.58	56.96	76.6	0.27	1.24
Panna	Mumbai Offshore		3140	Sandy siltstone, Shale, Sandstone	128	435	2.23	43.94	70.9	0.07	1.06

0 of 23 row(s) selected.

Rows per page10Page 1 of 3



Сложность экстракции геоинформации заключается в нахождении взаимосвязей между объектами, разбросанными по тексту и различным таблицам

Примеры таблиц в статьях с образцами анализов Rock-Eval и определения отражательной способности витринита

Table 1
Rock Eval pyrolysis results of shales from the Jharia sub-basin, Damodar Valley.

Block	Borehole/Depth (m)	S ₁	S ₂	PI	Tmax	S ₃	PC (%)	RC (%)	TOC(%)	HI	OI	MINC (%)
W. Mahuda	MMW-18											
R-1	151.8–152.2	0.61	15.43	0.04	442	0.29	1.4	6.35	7.75	199	4	0.79
R-2	193.8–194.2	0.95	24.75	0.04	444	0.13	2.17	9.12	11.29	219	1	0.52
R-3	226.9–227.2	0.42	3.54	0.11	450	0.15	0.41	3.76	4.17	85	4	0.88
R-4	272.1–272.4	1.19	68.69	0.02	439	0.2	5.86	17.23	23.09	297	1	4.59
R-5	330.1–330.5	0.29	2.67	0.1	455	0.07	0.36	3.04	3.4	79	2	1.29
R-6	369.1–369.4	0.27	2.15	0.11	452	0.16	0.25	2.62	2.87	75	6	0.81
R-7	416.5–416.8	2.47	17.01	0.13	445	0.24	1.76	10.21	11.97	142	2	1.69
R-8	416.0–416.5	2.76	19.71	0.12	450	0.34	2.02	10.78	12.8	154	3	1.6
R-9	450.1–450.6	1.47	8.45	0.15	464	0.35	1	6.88	7.88	107	4	2.11
Singra	MSG-34											
BM-1	409.8–410.1	0.35	3.63	0.09	455	0.19	0.37	3.99	4.36	83	4	0.54
BM-2	421.8–422.1	0.58	6.39	0.08	452	0.11	0.6	3.88	4.48	143	2	0.33
B-I	605.1–605.4	0.76	33.2	0.02	461	0.11	2.85	13.57	16.42	202	1	6.77
B-II	629.4–629.7	0.6	3.96	0.13	470	0.33	0.67	5.28	5.95	67	6	2.57
B-III	664.2–664.5	0.35	3.95	0.08	467	0.22	0.43	5.79	6.22	64	4	1.24
B-VA	1012.8–1013.2	0.87	7.3	0.11	485	0.21	0.76	10.55	11.31	65	2	7.34
B-VB	1048.8–1049.1	0.63	6.28	0.09	492	0.21	0.7	10.49	11.19	56	2	8.34
Kapurja	MKP-29											
BM-1	320.3–320.6	0.76	4.7	0.14	459	0.15	0.5	3.89	4.39	107	3	0.94
BM-2	351.0–351.3	0.66	4.08	0.14	457	0.25	0.48	3.8	4.28	95	6	1.42
B-I	540.3–540.7	0.56	6.15	0.08	473	0.19	0.6	7.51	8.11	76	2	1.57
B-II	572.9–573.2	0.69	3.31	0.17	473	0.21	0.43	4.67	5.1	65	4	1.57
B-III	641.3–641.6	0.42	6.48	0.06	476	0.18	0.62	8.24	8.86	73	2	1.22
B-V	876.3–876.7	0.19	2.9	0.06	493	0.16	0.32	7.16	7.48	39	2	1.47
B-VII	958.6–959.0	0.13	5.55	0.02	493	0.18	0.66	12.31	12.97	43	1	4.44
B-IX	1009.0–1009.3	0.12	2.99	0.04	509	0.27	0.38	8.34	8.72	34	3	5.55

S₁ = mgHC/gTOC; S₂ = mgHC/gTOC; PI = S₁/S₁ + S₂; Tmax = °C; PC = Pyrolysable carbon (%); RC = Residual carbon (%); HI = mgHC/g TOC; OI = mg CO₂/g TOC; MINC = mineral carbon (%); R = Ranigani Formation; BM = Barren Measure Formation; B = Barakar Formation.

Table 1
Location of the studied samples with depth and formation.

Sample Name	Colliery	Age	Depth(m)
Moon	Moonidih	Permian	421
PB	Parbatpur		180
WB	WestBokaro		80
NKP	North Karanpura		523
TKPDG	Makum	Late-Oligocene	63
TKPGS			150
TKPMD			182

Table 3
TOC and Rock eval pyrolysis data for Damodar Valley and Assam shale.

Sample	S1 (mg HC/g sample)	S2(mg HC/g sample)	GP (S1+S2) HC/g rock	PI	Tmax(°C)	EVRO(%)	TOC (%)	HI (mg HC/g TOC)	OI (mg CO ₂ /g TOC)
MOON	0.91	43.98	44.89	0.02	465	1.21	30.91	142	0.71
PB	1.03	21.03	22.06	0.05	440	0.76	12.18	173	0.903
WB	1.96	26.5	28.46	0.07	451	0.95	19.01	139	1.104
NKP	0.75	18.95	19.7	0.04	442	0.79	6.39	297	8.763
TKPDG	2.18	59.27	61.45	0.04	422	0.43	19.61	302	0.66
TKPMD	3.22	126.43	129.65	0.02	434	0.65	37.36	338	0.08
TKPGS	0.56	10.23	10.79	0.05	431	0.59	4.84	211	5.37

S1: Free hydrocarbon (HC) present in rock, mg HC/g rock; S2: HC that forms due to thermal cracking of kerogen, mg HC/g rock; S3: Carbon dioxide yield during thermal cracking, mg CO₂/g rock; HI: Hydrogen Index, mg HC/g TOC; Ro: Vitrinite reflectance (%); TOC: Total organic Carbon, wt. %; Tmax: Temperature at the crest of S2 peak denoting maximum rate of HC generation; PI: Production Index; OI: Oxygen Index, mg CO₂/g TOC.

Table 2
XRD analysis showing distribution of minerals in shale samples.

Sample ID	Basin name	Quartz (%)
MOON	Damodar Valley	25.45
PB		37.41
WB		36.88
NKP		26.88
TKPDG	Assam	44.79
TKPG		55.82
TKPMD		6.22

Название формации (выделено красным) закодировано в названии образца и расшифровано в сноске к ней. Ссылка на название бассейна в заголовке таблицы (выделено синим) и данные образца в самой таблице (выделено зеленым)

Название образца (подчеркнуто фиолетовой линией) встречается в различных таблицах с привязкой к названию бассейна (синее подчеркивание) и к параметрам образца (зеленое подчеркивание)

Проведенная работа позволила **определить и классифицировать основные виды ошибок** предобработки и экстракции данных **с помощью БЯМ**

Типичные виды ошибок, возникающие при предобработке и экстракции данных

 Ошибки предобработки данных	 Ошибки экстракции онтологические	 Ошибки экстракции технические
Таблицы с данными образцов распознаются неверно	Неверно определяется вид данных. Например, вместо названия формации приводится геохронологический период.	БЯМ не следует формату выходных данных
Таблицы встроены в рисунки и не распознаются	Неверно определяется значение параметра образца из таблицы - LLM путает колонки и использует данные из соседних колонок таблиц.	Уход в бесконечную (до заданного порога) генерацию "мусорных" токенов
	Неверно определяется бассейн или формация при сложных или неявных связях между ними и образцами.	
	Ошибки, допущенные авторами в статьях. Например, под формациями понимают названия угольных шахт.	
	Различные названия одного и того же объекта в разных источниках	



ЭФФЕКТИВНОСТЬ ГРР И РОЛЬ ШАНСОВ ГЕОЛОГИЧЕСКОГО УСПЕХА



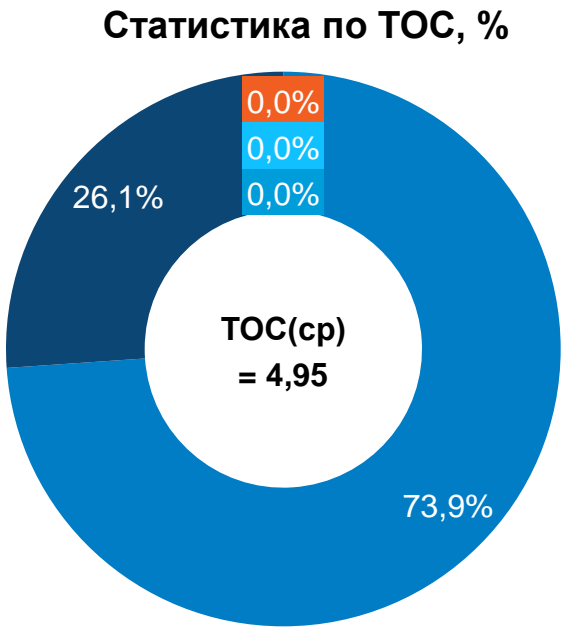
ОПИСАНИЕ РЕШЕНИЯ НА БАЗЕ ГЕНЕРАТИВНОГО ИИ



ОЦЕНКА РЕЗУЛЬТАТОВ И ПЕРСПЕКТИВЫ ПРИМЕНЕНИЯ

С помощью ПО GEOASSISTANT были отобраны подходящие параметры НГМП - ожидается богатое ОБ ранней зрелости с III типом керогена (генерация газа)

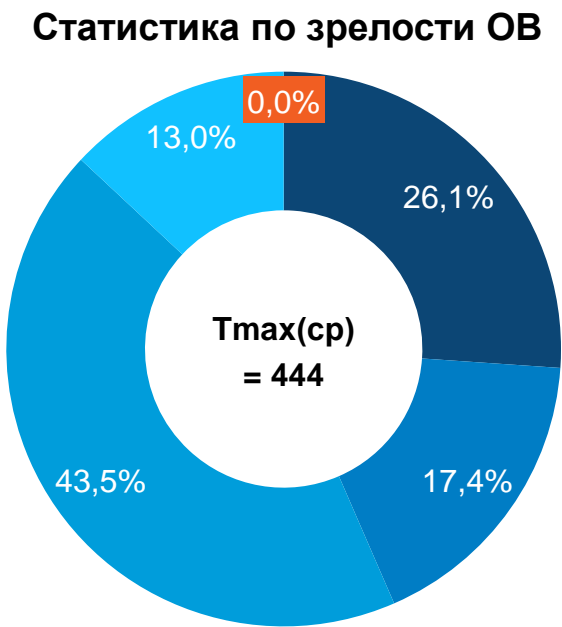
Статистика экстрагированных параметров для отобранных образцов, подходящих для выбранного ЛУ



Ожидаются НГМП **богатого** и **очень богатого** классов



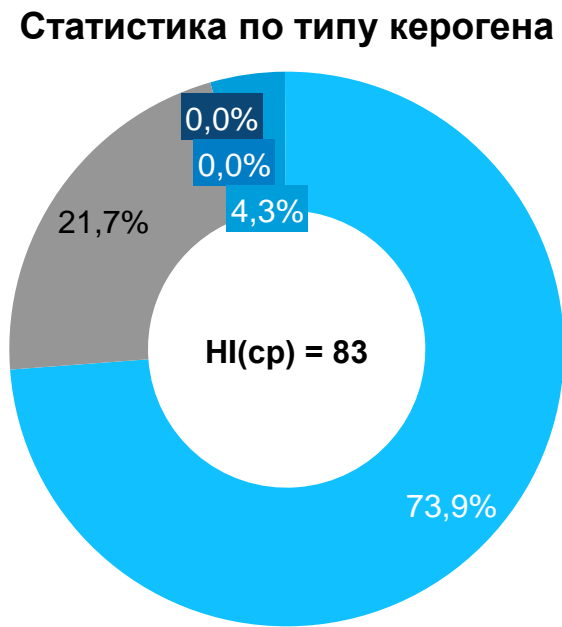
- Очень бедные
- Бедные
- Средние
- Богатые
- Очень богатые



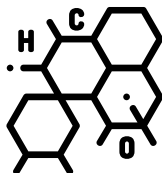
Наиболее вероятно обнаружение НГМП в **ранней стадии зрелости ОБ**



- Незрелые
- Ранней зрелости
- Пик нефтегенерации
- Поздней зрелости
- Сверхзрелое



Ожидается генерация преимущественно **газообразных УВ**



- I - нефть
- II - нефть
- II/III - нефть и газ
- III - газ
- IV - отсутствие УВ

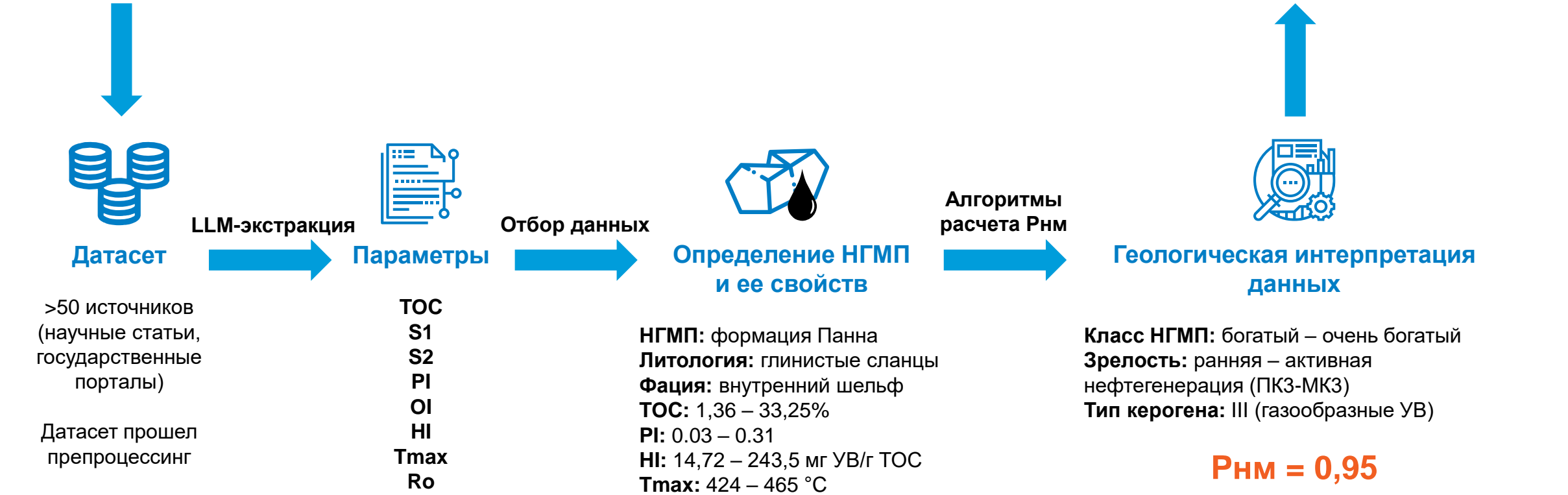
Использование ГенИИ позволило из неструктурированной геологической информации автоматически определить значение параметра R_{nm} gCoS

Запрос (prompt) адресованный GEOASSISTANT:

Определи значение фактора наличия и зрелости нефтегазоматеринской породы R_{nm} для ЛУ MB-OSHP-2023/1 нефтегазоносного суббассейна депрессии Сурат бассейна Мумбайского шельфа, расположенного в восточной части Камбейского залива

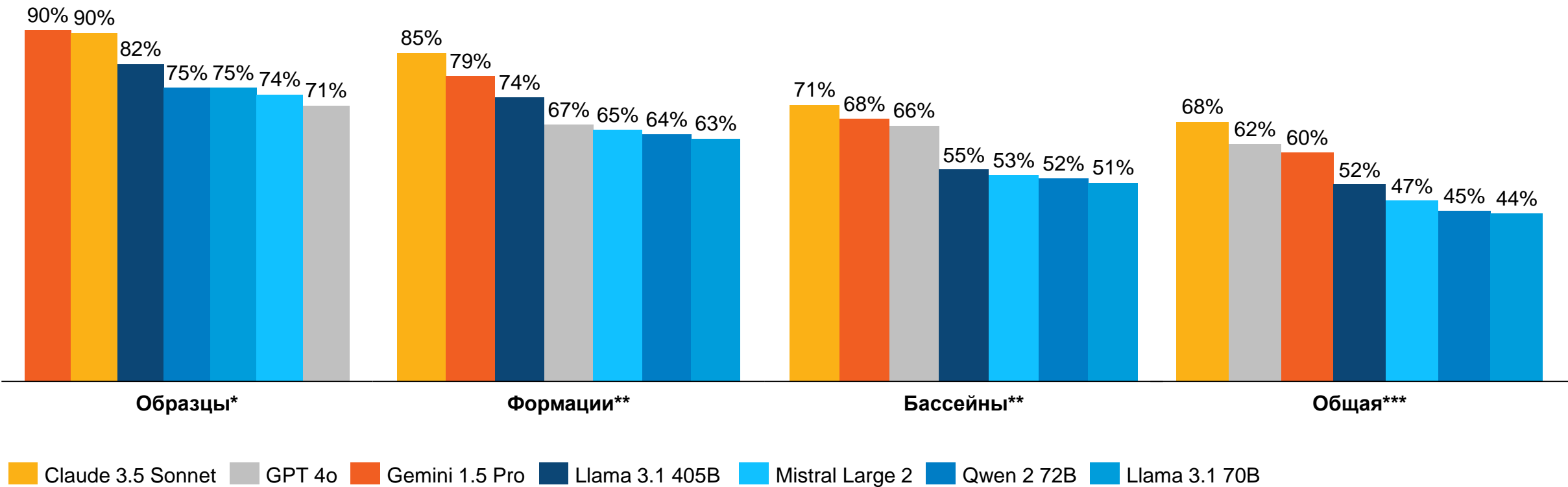
Ответ полученный от GEOASSISTANT:

Наличие НГМП крайне вероятно. Существуют НГМП с высоким ТОС и хорошей зрелостью (ПКЗ-МКЗ), генерация УВ либо идет в настоящее время, либо отложения почти вошли в главную зону нефтегенерации. Ожидается генерация, преимущественно, газообразных УВ.
Параметр R_{nm} составляет 0,95



Коммерческие иностранные БЯМ лучше справляются с извлечением данных, чем модели с открытыми весами, российские модели пока не работают

Бенчмаркинг корректности экстракции данных различными БЯМ по отдельным группам: значения параметров по образцам, названия бассейнов, формаций и общая корректность извлечения, %



В исследовании тестировались также отечественные модели ([YandexGPT Pro](#) и [GigaChat](#)), однако они не попали в итоговый бенчмаркинг, поскольку в первоначальном пайплайне они не справились со следованием инструкциям и генерацией в заданном формате (CSV), а в итоговом – не подошли из-за недостаточного размера контекстного окна

* Корректное извлечение значений 4х параметров (ТОС, Tmax, HI, Ro) образца
** Корректное извлечение значений 4х параметров образца и связанного с образцом названия бассейна или формации соответственно
*** Корректное извлечение значений 4х параметров образца, связанных с ним названий бассейна и формации

ГенИИ-системы могут применяться в задачах сектора разведки и добычи профильных госорганов, нефтегазовых компаний, «сервисников» и НТЦ

Возможные направления применения ПО GEOASSISTANT и ключевые заинтересованные стороны



Все материалы, представленные в настоящем документе, носят исключительно информационный характер, являются исключительно частным суждением авторов и не могут рассматриваться как призыв или рекомендация к совершению каких-либо действий.

ООО «ВЫГОН Консалтинг» и его сотрудники не несут ответственности за использование информации, содержащейся в настоящем документе, за прямой или косвенный ущерб, наступивший вследствие использования данной информации, а также за достоверность информации, полученной из внешних источников.

Любое использование материалов документа допускается только со ссылкой на источник - ООО «ВЫГОН Консалтинг».

123610, Россия, Москва,
Краснопресненская наб., д.12

телефон: +7 495 543 76 43

web: <http://vygon.consulting>

e-mail: info@vygon.consulting