



Министерство инноваций, цифрового  
развития и инфокоммуникационных  
технологий Республики Саха (Якутия)



# ОПЫТ ИСПОЛЬЗОВАНИЯ НЕЙРОСЕТИ ДЛЯ КЛАССИФИКАЦИИ ДОКУМЕНТОВ В СЭД

Якутск | 2021



ЕСЭД работает с 01.01.2019 года на базе СЭД «ДЕЛО»



37

ОГВ РС(Я) [100%]

445

ОМСУ РС(Я) [100%]

529

ГУ [100%]

1349

МУ [60 %]



69

деревьев в месяц  
сохранено\*



13 700

пользователей

5,2 Тб

Объем базы данных

1,9 Тб

в год прирост БД



3600

документов  
в день

1700

Проектов документов  
в день

9500

поручений  
в день

\*1 дерево = 17 пачек бумаги ([esquire.ru/ecology-27](http://esquire.ru/ecology-27))





## ЦЕЛЬ:



Повышение эффективности работы с документами:

- Ускорение обработки документов в СЭД
- Автоматизация рутинных процессов

## ЗАДАЧИ:



- Применимость технологии машинного обучения в документообороте.
- Повышение эффективности работы технологии.
- Тиражирование технологии.

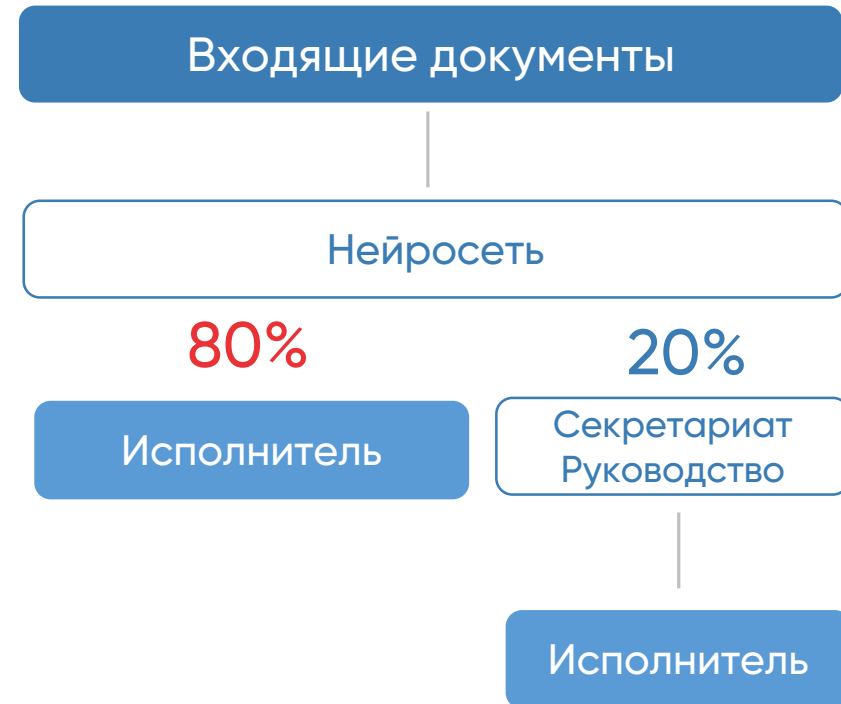
# КЛАССИФИКАЦИЯ ВХОДЯЩИХ ДОКУМЕНТОВ



Путь документов в системе документооборота **сейчас**



Путь документов в системе **в будущем**





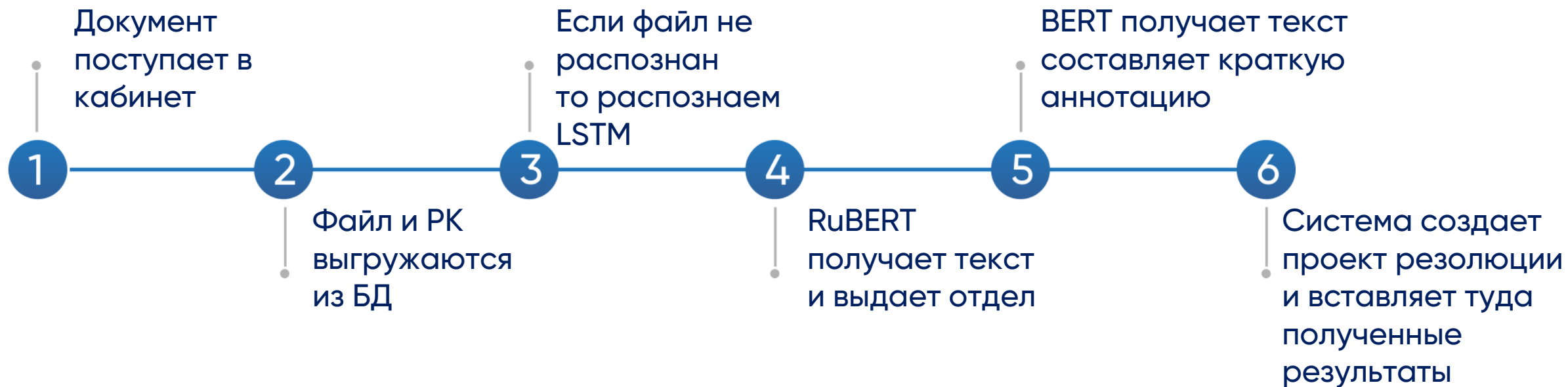
## Преимущества BERT:

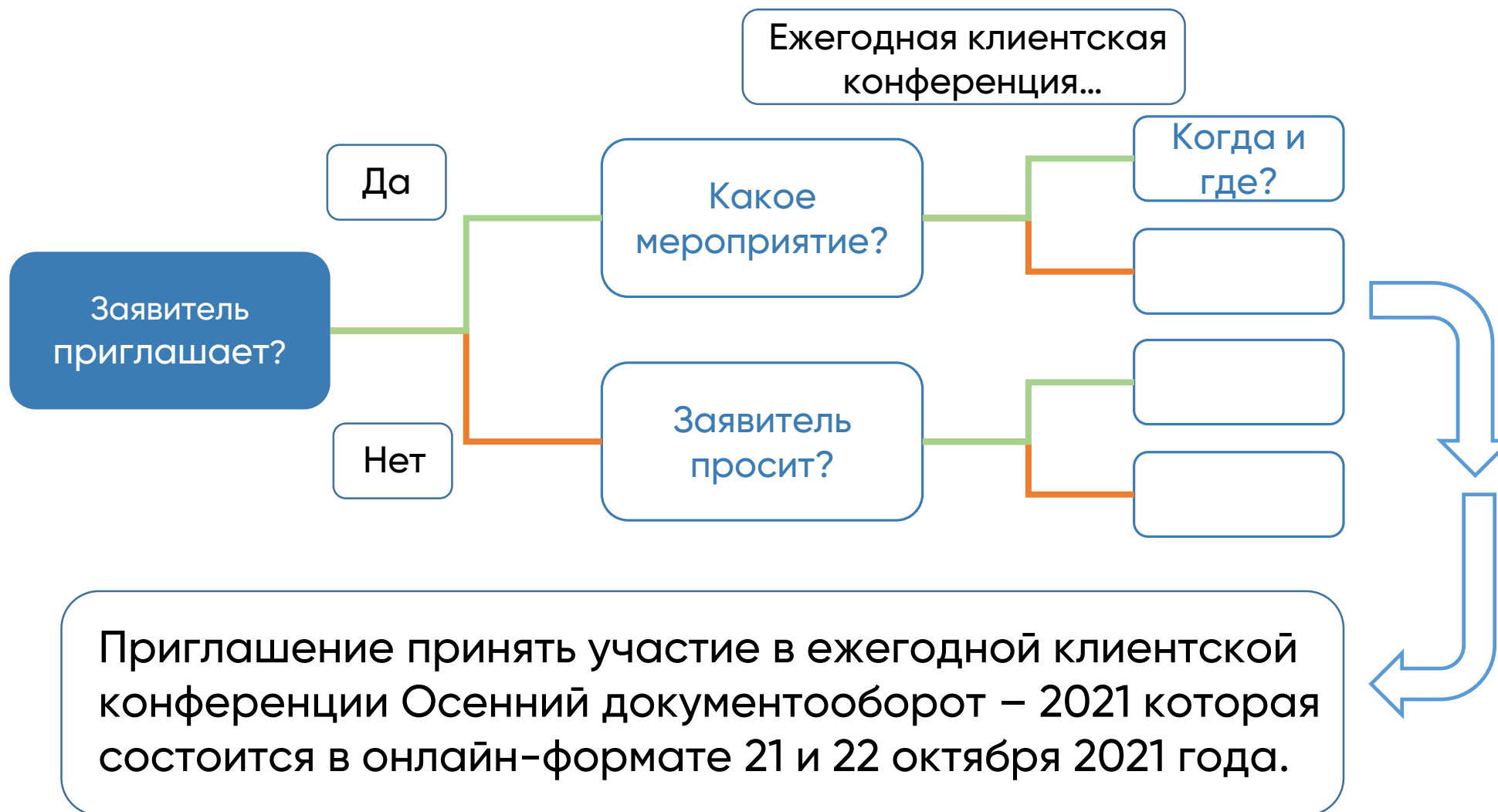
- Учитывает контекст слов
- Предобученная нейросеть





## Весь цикл - 3 минуты









## Данные для обучения

Входящие документы Мининноваций РС(Я) с 2019 года		Входящие документы ГБУ РЦИТ с 2020 года	
PK с нераспознанными PDF файлами	PK с распознанными PDF файлами	PK с нераспознанными PDF файлами	PK с распознанными PDF файлами
~12000 PK	~3000 PK	~4000 PK	~1000 PK

Справочно: в ГБУ РЦИТ поступает 60 документов в день, в министерство 40 документов в день

# ОЦЕНКА НЕЙРОСЕТИ ПОСЛЕ ОБУЧЕНИЯ



Результаты обучения на документах министерства

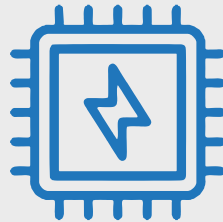
МОДЕЛЬ	Точность	Полнота*	macro-F1**
TF-IDF+RF	0.83	0.43	0.56
Word2Vec+RF	0.86	0.50	0.63
Single BERT multilabel	0.65	0.24	0.35
Multiple BERT	0.90	0.60	0.72
Multiple BERT only scanned	0.92	0.68	0.78
Multiple BERT only scanned and conf > 0.9	<b>0.94</b>	<b>0.71</b>	<b>0.81</b>



Одна модель для классификации занимает ~**750 МБ.**

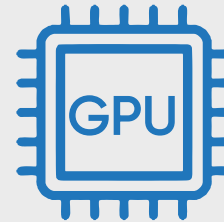
Общая модель для аннотирования занимает **2.5 ГБ**

Для запуска модели в качестве вычислительного аппарата можно выбрать:



**CPU + RAM**

**1-3 мин** на документ.



**GPU + GPU RAM**

**0.8 - 1.2 сек** на документ

# КАК ЭТО ВЫГЛЯДИТ В СЭД



На утверждении	Без приоритета	ГБУ "РЦИТ" - Официальный ящик		(Проект) Приглашение 28 сентября 2021 г. в 17ч 00м на совещание у руководителя Департамента водного и ж/д транспорта Министерства транспорта и дорожного хозяйства Республики Саха (Якутия) Коркина С.Н. на презентацию концепции программно-аппаратного решения (автоматизации) для паромной переправы в п.Нижний Бестях докладчик компания СОЛВО. Совещание будет проходить в режиме ВКС:	(2)
На утверждении	Без приоритета	ГБУ "РЦИТ" - Официальный ящик		(Проект) Просят учесть данные изменения и применять новые реквизиты при составлении договорной документации отправке почтовых сообщений и оформлении прочих документов.	(1)
На утверждении	Без приоритета	ГБУ "РЦИТ" - Официальный ящик	Александрова Анна Германовна - Начальник отдела [...]	(Проект) Уведомляют об оплате обязательно в противном случае не гарантируется наличие товара на складе.	(1)
На утверждении	Без приоритета	ГБУ "РЦИТ" - Официальный ящик		(Проект) [...]	(1)
На утверждении	Без приоритета	ГБУ "РЦИТ" - Официальный ящик		(Проект) Для работы	(3)
На утверждении	Без приоритета	ГБУ "РЦИТ" - Официальный ящик		(Проект) Просят ограничить доступ сотрудникам ведомства в разделе регистрации документов в части регистрация приказов по основной деятельности кроме сотрудников отдела финансово-правового обеспечения государственной службы и организационного сопровождения.	(1)
На утверждении	Без приоритета	ГБУ "РЦИТ" - Официальный ящик		(Проект) Просят предоставить средства вычислительной техники (СВТ) из обменного фонда средств вычислительной техники органов государственной власти РС(Я) во временное пользование Министерство	(1)

# КАК ЭТО ВЫГЛЯДИТ В СЭД



**В** РК № 4087 от 27.09.2021 (501.Входящие) 0501.ГБУ РС(Я) "РЦИТ" / Кабинет ГБУ "РЦИТ"

Вести резолюцию | Контроль исполнения | Вести отчёт | Протокол | Протокол поручения | Добавить в личную папку | Редактировать поручение | Удалить РК | Сводный журнал | Ещё ▾

Краткие сведения  
Основное  
Поручения  
Связки  
Адресаты

⊕ Все ⊖ Ветка ⊖ Мои По автору ▾

**В** 4087 от 27.09.2021

⊕ Кривошапкин Иван Александрович

⊖ ГБУ "РЦИТ"

▶ ⊖ Проект (1) Приглашение принять участие в ежегодной клиентской кон...  
└ Кривошапкин Иван Александрович

▼ Проект

**ГБУ РС(Я) "РЕСПУБЛИКАНСКИЙ ЦЕНТР ИНФОКОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ"**  
**ОФИЦИАЛЬНЫЙ ЯЩИК**

Кривошапкину Ивану Александровичу

Приглашение принять участие в ежегодной клиентской конференции Осенний документооборот – 2021 которая состоится в онлайн-формате 21 и 22 октября 2021 года.

Контроль

Примечание:  
Данный проект поручения был создан нейросетью.

Приглашение на конференцию Осенний д... ▾

1 / 2 | - + | ↺ ↻ | ⏴ ⏵ | ⏶ ⏷

**ЭОС** **ПК-6**

Россия 107113 Москва, ул. Шумкина, д.20, стр. 1 | Телефон: 8(495) 221-24-31

Онлайн-конференция пользователей «Осенний документооборот-2021» Исх. № 590-21 от 20.09.2021

ГБУ РС (Я) "Республиканский центр инфокоммуникационных технологий" Директору Кривошапкину И.А.

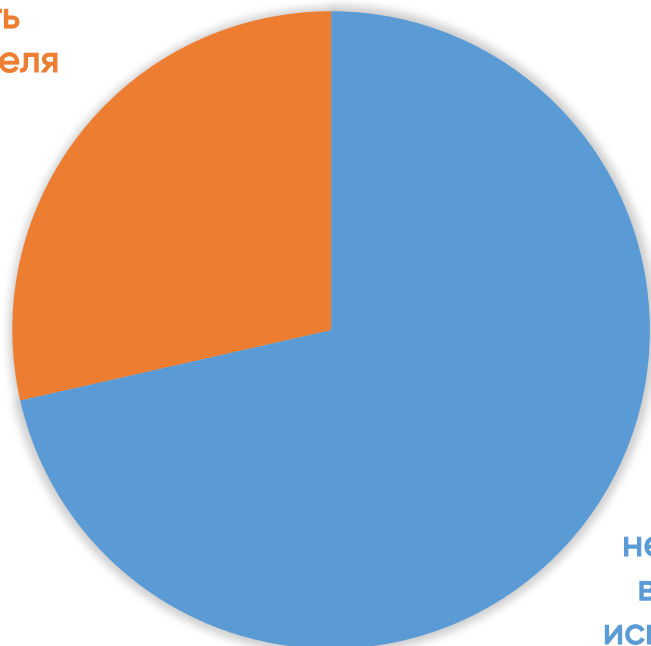
**Уважаемый Иван Александрович!**

От лица лидера российской ЕСМ-отрасли компании «Электронные Офисные Системы» и технического подкомитета по стандартизации документов (ПК 6) приглашаем Вас принять участие в ежегодной клиентской



## 550 ПОРУЧЕНИЙ

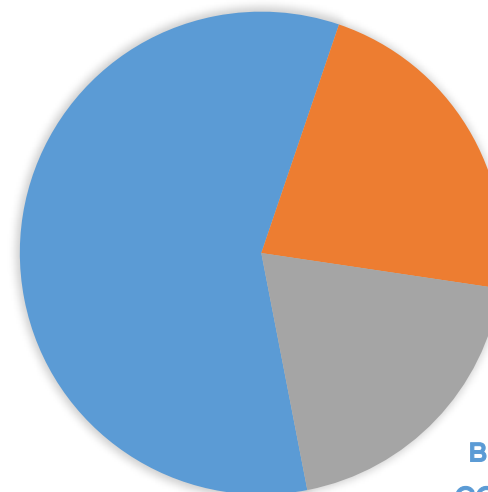
нейросеть  
не смогла  
выбрать  
исполнителя  
**29%**



нейросеть  
выбрала  
исполнителя  
**71%**

## 393 ПОРУЧЕНИЯ С ОТВЕТОМ

выбор  
полностью  
совпадает  
с выбором  
человека  
**58%**



выбор  
частично  
совпадает  
**22%**

выбор не  
совпадает  
**20%**

Данные с 08.09.2021 по 07.10.2021

76% Средняя точность  
68% Средняя полнота



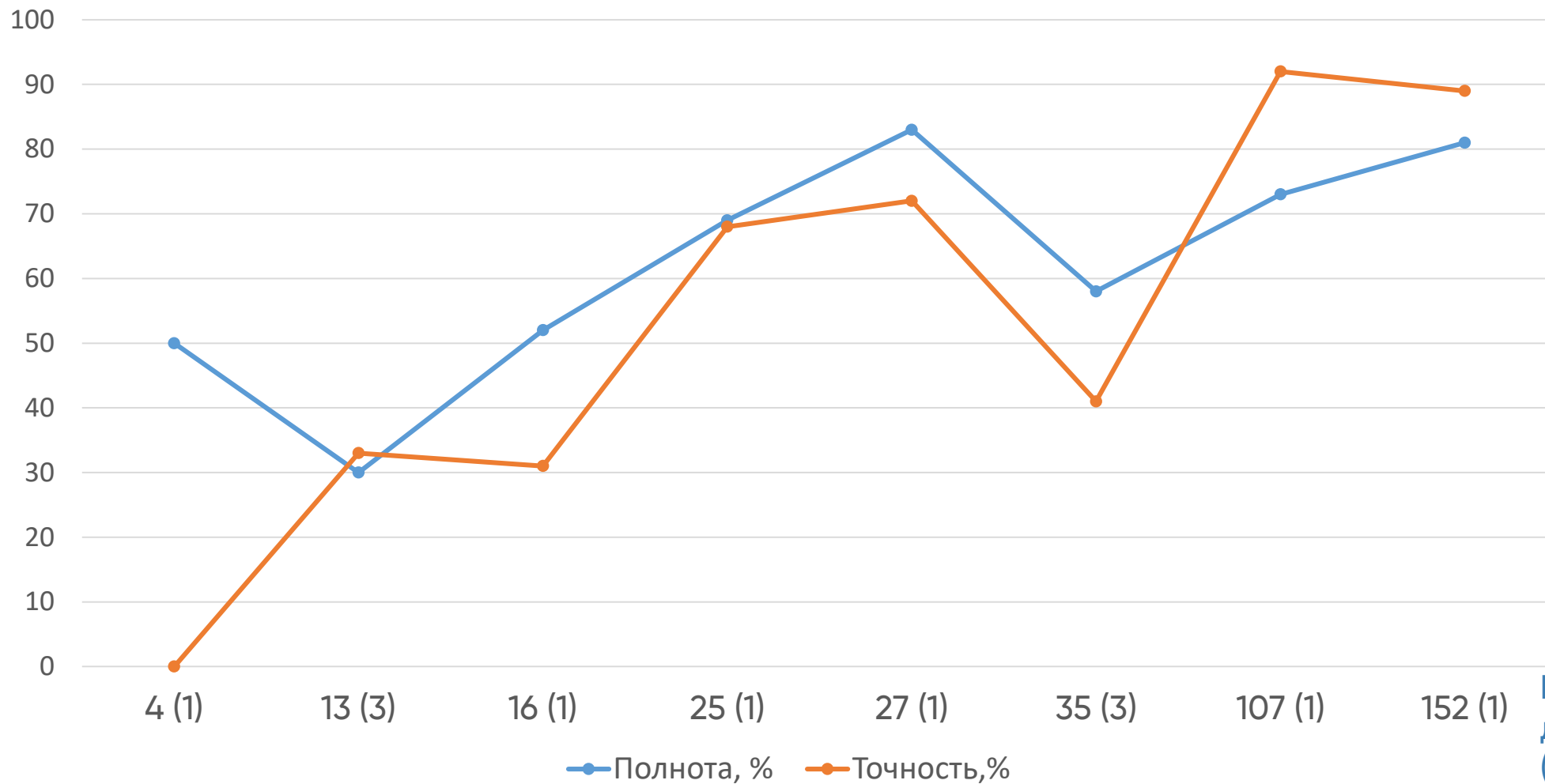
Данные с 11.10.2021 по 20.10.2021

# СТАТИСТИКА ПОРУЧЕНИЙ В ГБУ РЦИТ ПО ОТДЕЛАМ



Процент, %

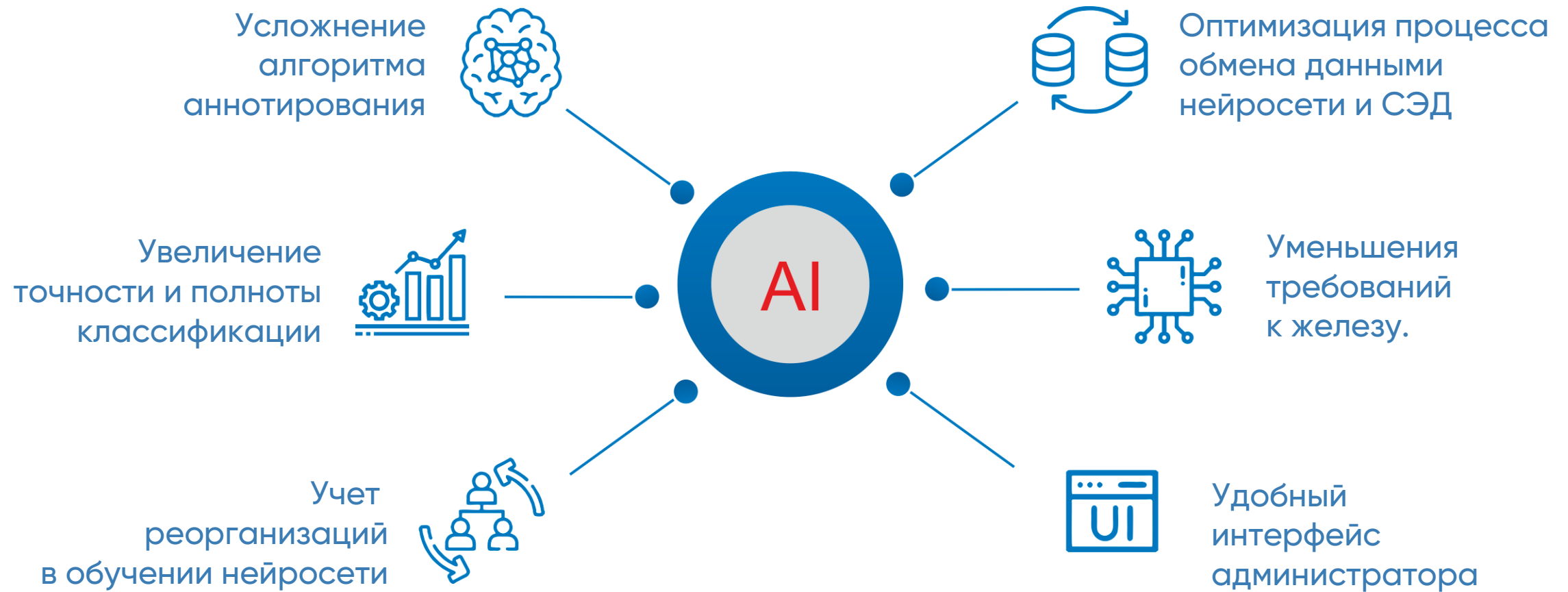
Точность ответов нейросети и количество документов



Количество документов (классов)



# КЛАССИФИКАЦИЯ ВХОДЯЩИХ ДОКУМЕНТОВ





Министерство инноваций, цифрового развития и  
инфокоммуникационных технологий Республики Саха (Якутия)



ГБУ РС(Я) «Республиканский центр инфокоммуникационных  
технологий»



ООО «Инфосистемы-КС» – партнер ЭОС



**СПАСИБО ЗА ВНИМАНИЕ!**