

Чек-лист проверки DeepSeek-OCR 2 перед внедрением

PDF для [downloads/deepseek-ocr-2-checklist.pdf](#). Его можно дать команде, которая сравнивает DeepSeek-OCR 2 с текущим OCR, IDP или VLM-пайплайном на собственных документах.

1. Соберите тестовый корпус

- Добавьте PDF и сканы из реальных рабочих процессов.
- Разделите документы по типам: договоры, счета, акты, инструкции, научные статьи, таблицы и формы.
- Отдельно включите документы с русским и английским текстом.
- Добавьте негативные кейсы: случайные строки, артикулы, серийные номера и короткие коды.
- Не используйте только документы, на которых модель уже выглядит хорошо.

2. Разметьте эталон

- Подготовьте эталонный текст для обычных абзацев.
- Отдельно разметьте таблицы, формулы, подписи и порядок чтения.
- Для номеров, кодов, сумм и дат используйте exact match.
- Сохраняйте эталон и результаты в версионизируемом наборе.
- Фиксируйте, какие ошибки считаются блокирующими для бизнеса.

3. Сравните с текущим пайплайном

- Сравните OCR-2 с текущим OCR, IDP или гибридной схемой.
- Отдельно оцените первую DeepSeek-OCR, если она есть в тестовой матрице.
- Проверьте CER, WER, edit distance и exact match.
- Замерьте качество reading order для многостолбцовых страниц.
- Проверьте устойчивость к поворотам, шуму, печатям и рукописным пометкам.

4. Проверьте эксплуатацию

- Замерьте задержку и потребление GPU-памяти.
- Проверьте стабильность на длинных PDF и массовой очереди задач.
- Настройте логирование входов, выходов и ошибок.
- Определите правила обработки персональных данных и коммерческой тайны.
- Подготовьте мониторинг качества и план отката на прежний пайплайн.