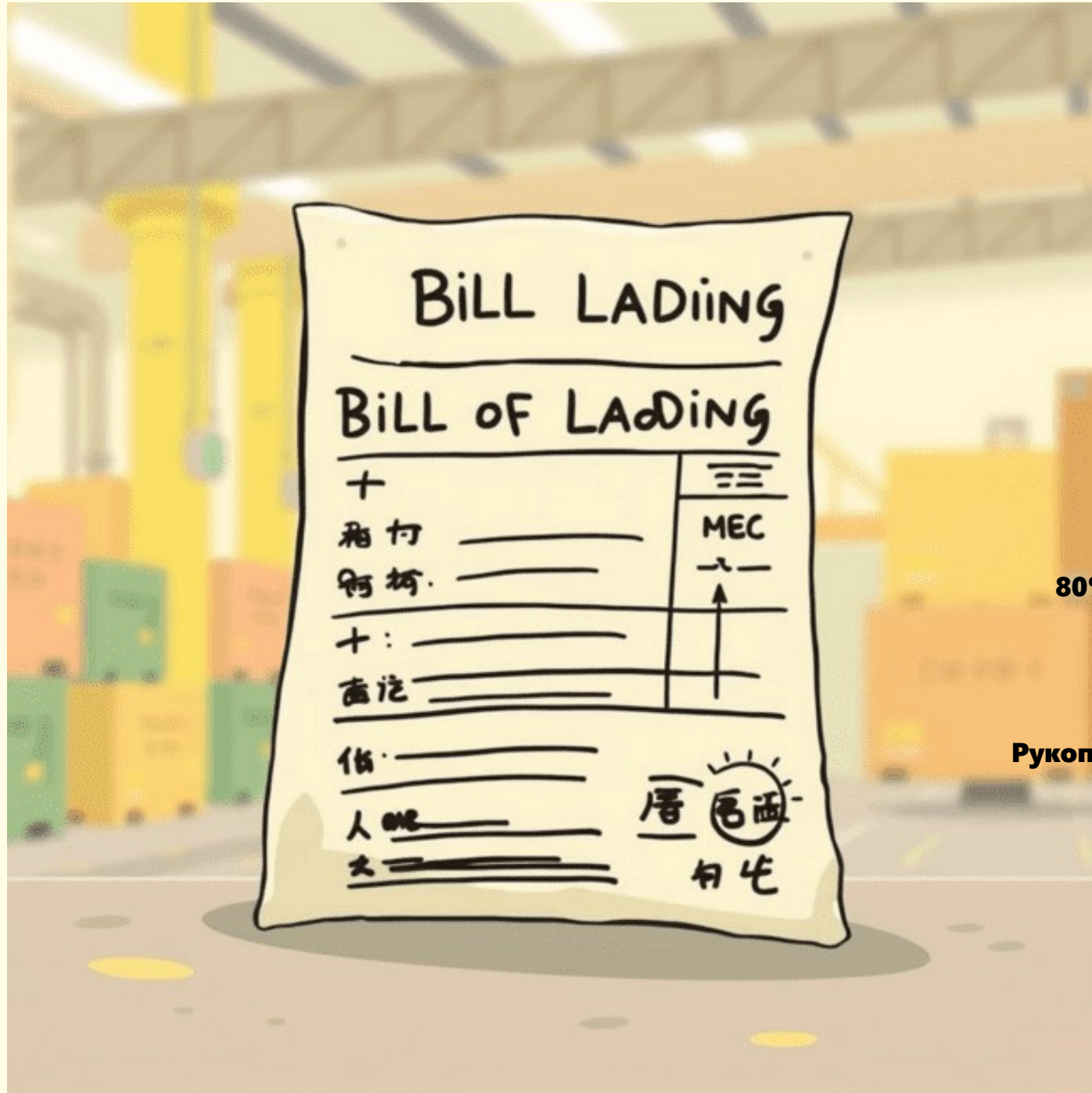


Розробка інтелектуальної системи управління даними для деревообробного підприємства на базі LLM/RAG, OCR/NLP та стеку n8n–Prometheus–Grafana

Виконав: Бакуновець Артем Віталійович

Науковий керівник: д.т.н., професор Єгорченкова Н.Ю.

Актуальність Проблеми: Неструктуровані Дані та Втрата Ефективності



Втрата Даних (Dark Data)

Неможливість доступу та аналізу цінної інформації, що міститься у паперових архівах.

Затримка Введення

Від 4 до 24 годин на ручне введення даних, що уповільнює бізнес-процеси.

Високий Рівень Помилки

3-5% помилок при ручному введенні, що веде до некоректних рішень та фінансових втрат.

80% документів — паперові та неструктуровані

що ускладнює їх аналіз та зберігання

Рукописні товарно-транспортні накладні (ТТН)

є яскравим прикладом даних, які не піддаються автоматичному аналізу, що призводить до значних втрат.

Мета та Об'єкт Дослідження

Мета: Підвищення ефективності управління даними шляхом впровадження AI-системи

01

Об'єктом дослідження є процеси управління інформаційними потоками та документообігом на деревообробному підприємстві

02

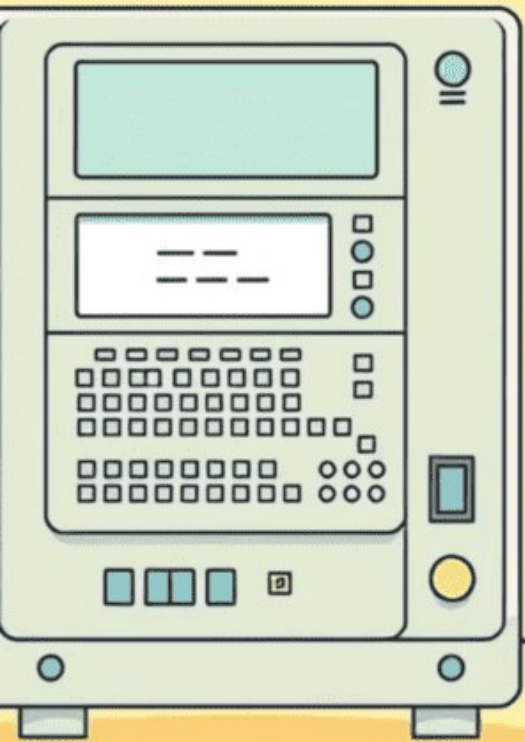
Предметом дослідження є методи та засоби створення інтелектуальних інформаційних систем, заснованих на технологіях LLM, OCR/NLP та low-code оркестрації.

03

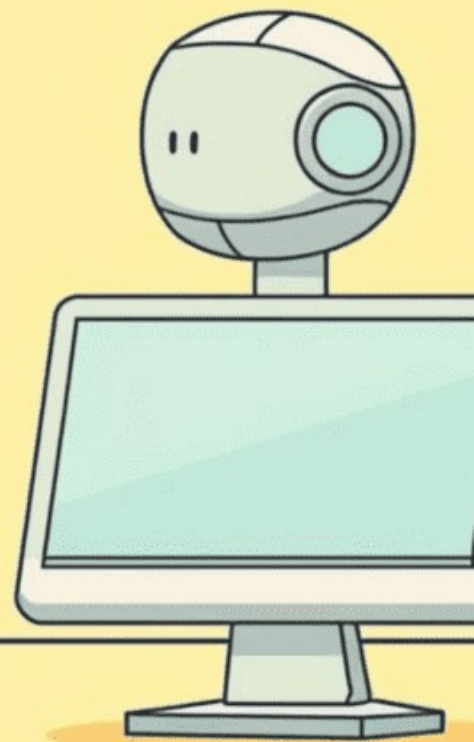
Наукова новизна одержаних результатів полягає у розробці комплексної архітектурної моделі, що є першою спробою інтеграції GPT-4o (як multimodal OCR/NLP), Qdrant (RAG-ядро) та n8n (оркестратор) для автоматизації обробки структурно-складних документів деревообробної галузі. Удосконалено методику моніторингу шляхом застосування Prometheus-Grafana для оцінки бізнес-метрик ефективності та вартості роботи ШІ (Cost-Per-Token).

Old vs. New

ERP



AI



Аналіз Рішень та Вибір Стеку

Вибір оптимальних технологій для розробки інтелектуальної системи — це ключ до успіху.

Традиційні ERP-системи (SAP/1C)



Дорого: Висока вартість ліцензій та впровадження.



Довго: Тривалий термін розгортання та адаптації.



Жорстко: Низька гнучкість до нестандартних форматів даних.

Запропоноване Рішення (n8n + AI)



Дешево: Економічно вигідна модель оплати та ліцензування.



Швидко: Швидке впровадження та масштабування.



Гнучко: Адаптація до будь-яких типів документів та даних.

Я обрав шлях інтеграції мікросервісів на базі **оркестратора n8n** та моделі **Pay-as-you-go**, що забезпечує доступність рішення для МСБ.

Чому n8n + AI?

Моє рішення базується на передових технологіях, що забезпечують безпеку, гнучкість та економічну ефективність.



Self-hosted

Повний контроль над даними та їх безпекою.

ОРЕХ модель

Плата за токени, а не за дорогі ліцензії, оптимізує витрати.

Гнучкість

Легка адаптація до нестандартних документів та бізнес-процесів.

Цей підхід робить інтелектуальну обробку даних доступною та ефективною для малого та середнього бізнесу.

Розділ 2: Архітектура Системи

Мікросервісна архітектура гарантує гнучкість, масштабованість та безпеку даних.

01

Мікросервісна Архітектура

Базується на контейнеризації Docker для ізоляції та ефективного керування компонентами.

02

Self-hosted Розгортання

Встановлення на власному Ubuntu VPS підприємства для максимального контролю та безпеки даних.

03

Надійність та Безпека

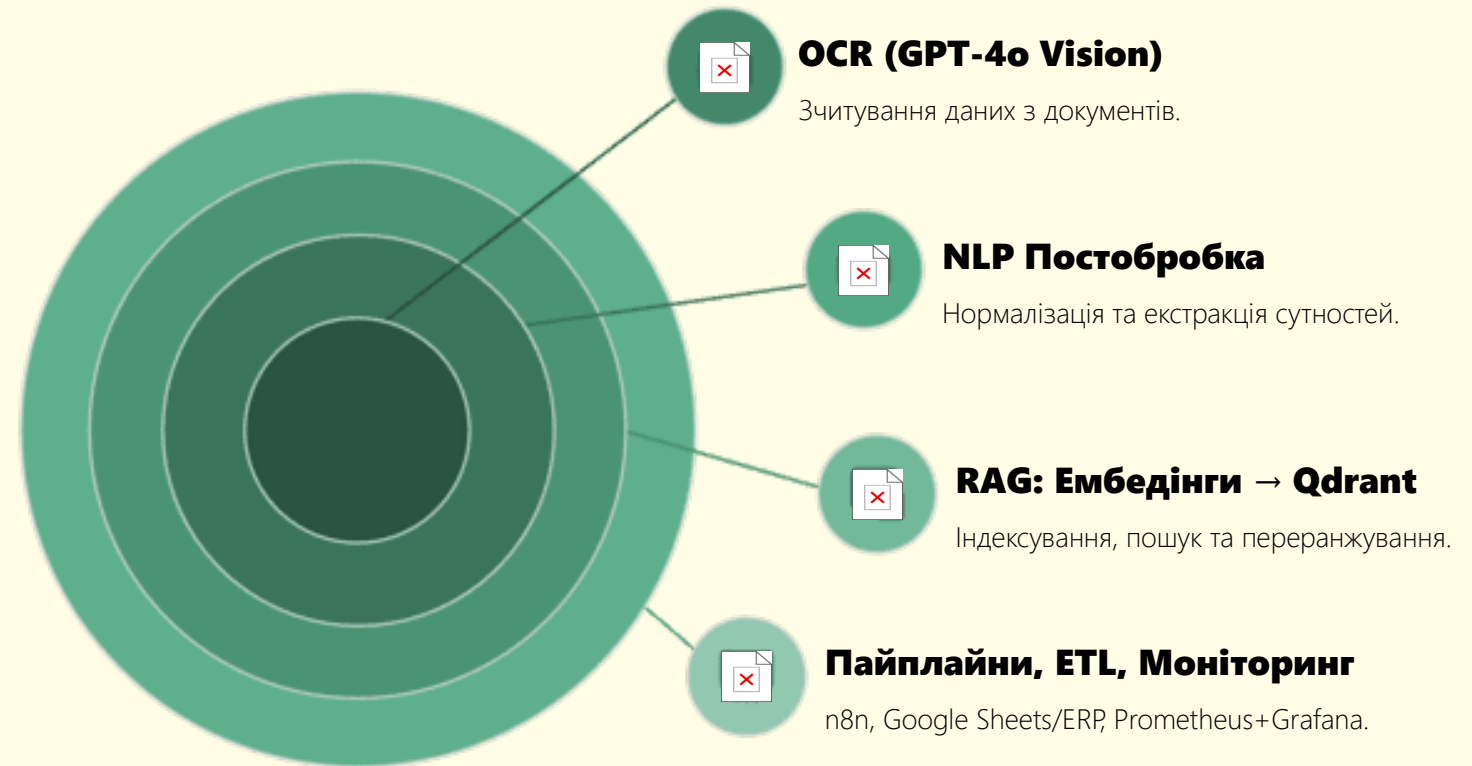
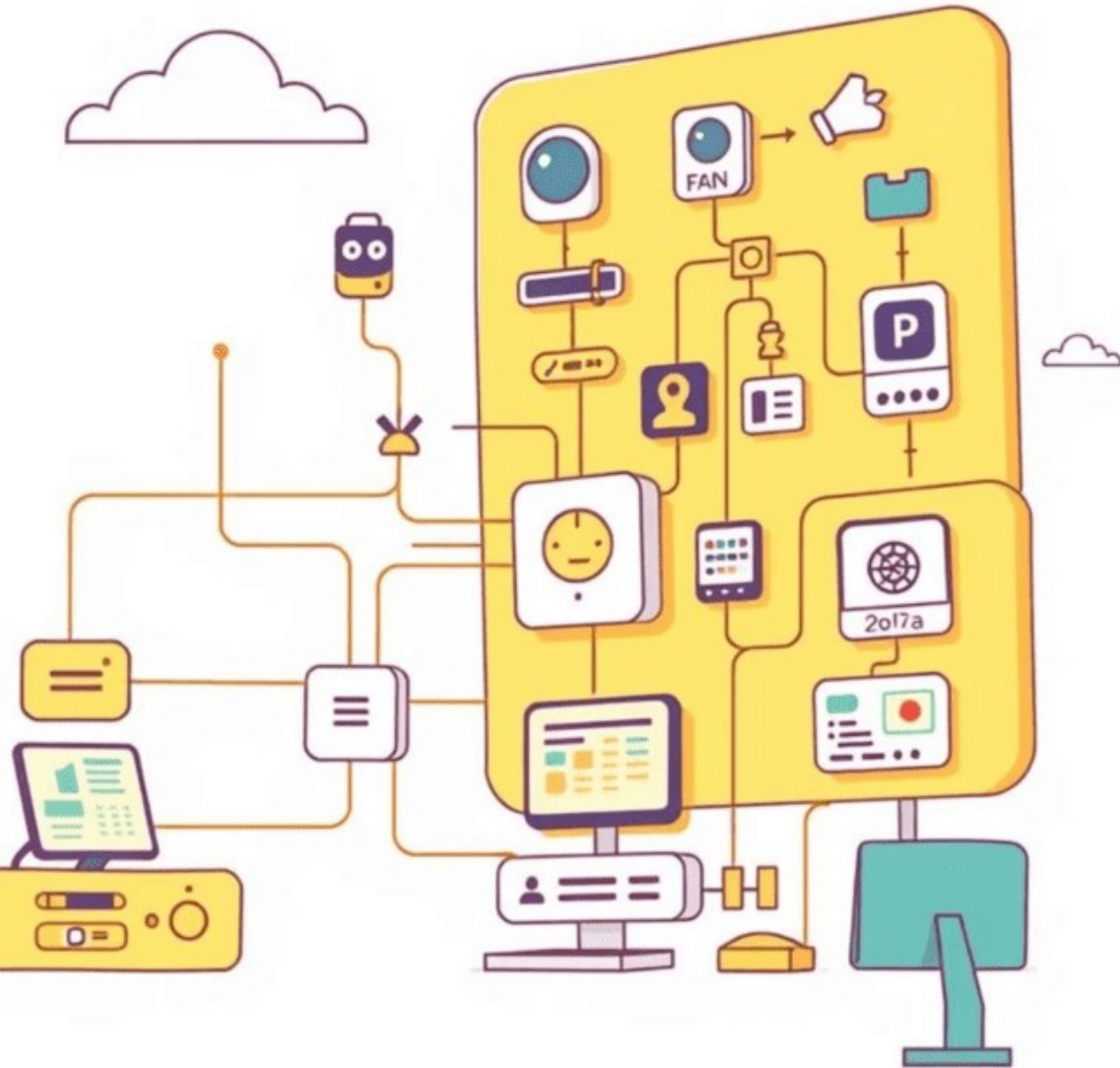
Забезпечується за допомогою Nginx та SSL-шифрування для захисту всіх транзакцій.

Центральний елемент — **оркестратор n8n**, який керує потоками даних: від отримання зображень до когнітивної обробки та запису результатів в облікову систему.



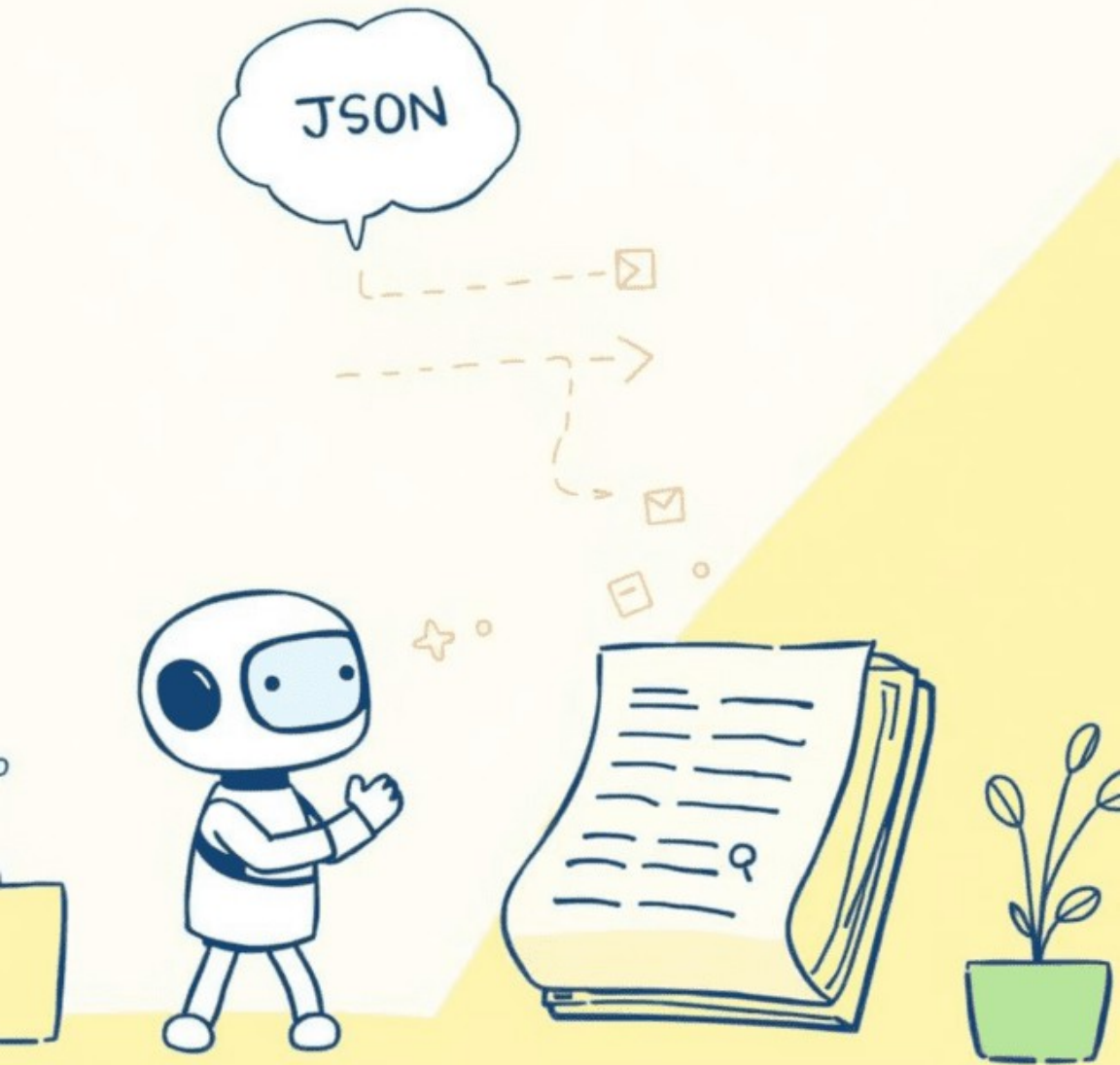
Загальна Архітектура Системи: RAG + LLM + n8n

Ця блок-схема відображає інтегрований підхід до автоматизації обробки даних, що поєднує можливості передових ШІ-інструментів з гнучкими робочими потоками.



Алгоритм Когнітивної Обробки (OCR/NLP)

Використання мультимодальної моделі GPT-4o для високої точності розпізнавання документів.



GPT-4o Vision

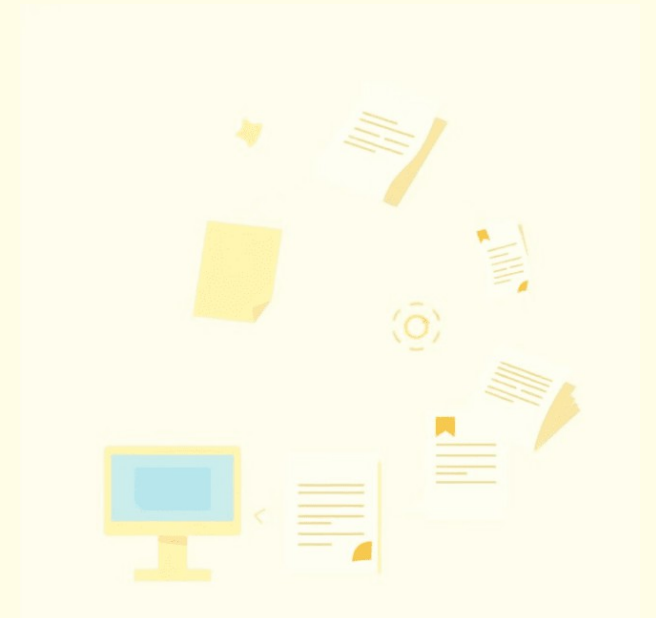
Замість традиційного Tesseract, що забезпечує значно вищу точність на складних документах.

Chain-of-Thought

Стратегія промптингу, яка змушує модель "міркувати" крок за кроком для перевірки логіки даних.

Валідація JSON-схеми

Забезпечення відповідності розпізнаних даних очікуваному формату, мінімізуючи помилки.

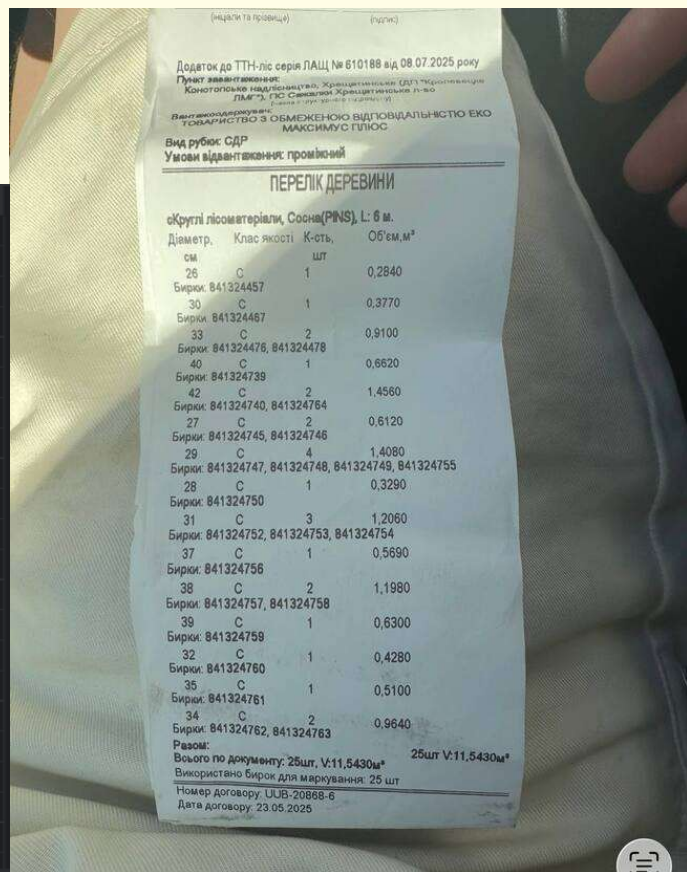


Цей підхід дозволяє досягти високої точності навіть на документах поганої якості, наприклад, рукописних накладних.

Приклад вхідного документів та результат обробки GPT та Tesseract

Tesseract

	c1	c2	c3	c4	c5
1	Діаметр	Клас якості	Кількість	Об'єм	Бирки
2	сКруглі	лісоматеріали,	<null>	<null>	6, I:, м., м?, Об'єм,
3	Діаметр,	<null>	Класякості	<null>	К-сть,
4	см	<null>	<null>	<null>	шт
5	43	<null>	В	<null>	0,7620, 1
6	Бирки:	271003406	<null>	<null>	<null>
7	37	<null>	В	<null>	1, 0,56890
8	Бирки:	27100340/	<null>	<null>	<null>
9	29	<null>	В	<null>	1,0560, 3
10	Бирки:	271003408,	<null>	271003418,	271003432
11	'35,,В,, "1, 05100"	<unset>	<unset>	<unset>	<unset>
12	Бирки:	271003409	<null>	<null>	<null>
13	728	<null>	в	991003412,	4, 13160
14	Бирки:	271003410,	<null>	<null>	271003414,, 271003428
15	32	<null>	В	<null>	2, 0,8560, -
16	Бирки:	272003411,	<null>	2/10034126	<null>
17	34	<null>	В	<null>	2, 0,9640
18	Бирки:	271003413,	<null>	271003419	0,7540
19	30	<null>	в	<null>	1, 2
20	Бирки:	271003415,	(ai	271003417	<null>
21	22	<null>	<null>	<null>	2, 0,4080
22	Бирки:	271003416,	в	271003421	<null>



GPT

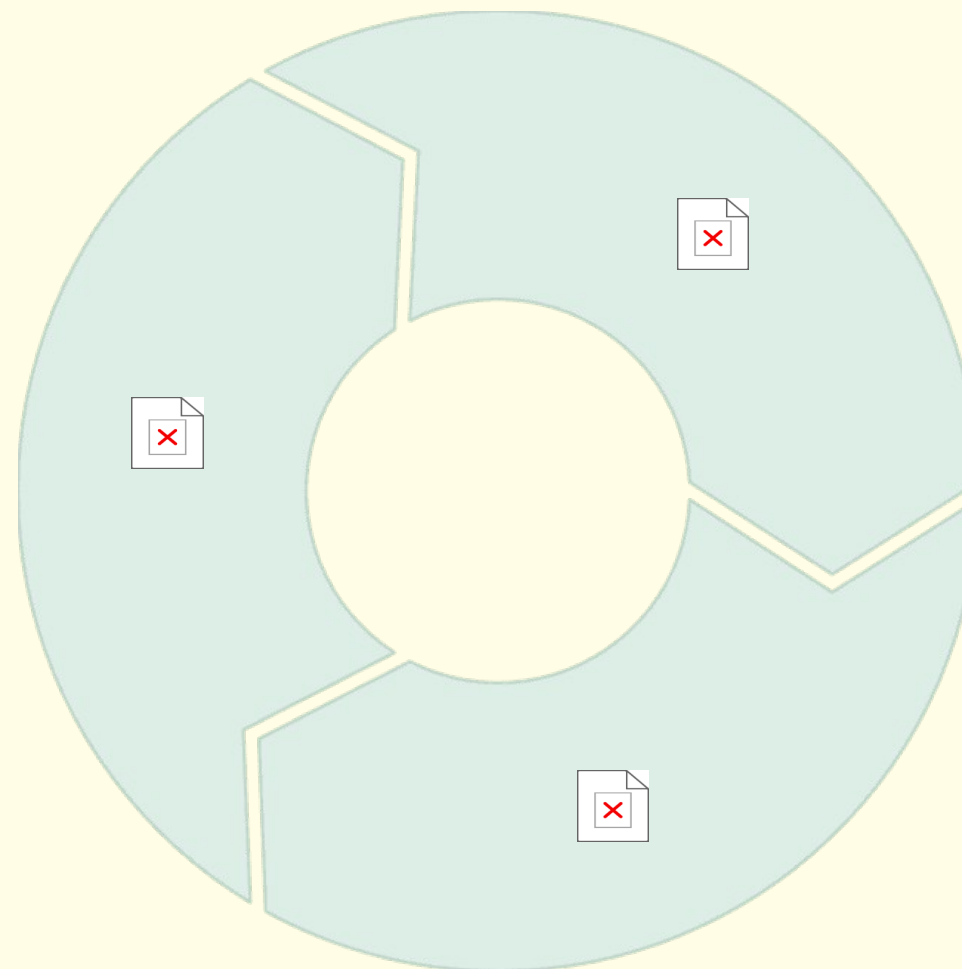
	Діаметр	Клас якості	Кількість	Об'єм	Бирки
1	42 С		5	3.6400	276014237, 27601423...
2	37 С		1	0.5690	276014244
3	27 С		2	0.6120	276014245, 276014714
4	28 С		2	0.6580	276014246, 276014441
5	29 С		1	0.3520	276014247
6	32 С		1	0.4280	276014248
7	36 С		1	0.5390	276014255
8	33 С		1	0.4550	276014443
9	40 С		1	0.6620	276014449
10	35 С		2	1.0200	276014712, 276014715
11	38 С		1	0.5990	276014713
12	39 С		1	0.6300	276014716
13	41 С		2	1.3900	276014717, 278461541

Підсистема RAG та Векторний Пошук

Для запобігання "галюцинаціям" штучного інтелекту ми впровадили архітектуру RAG.

База Знань Qdrant

Open Source векторна база даних, що є "довгостроковою пам'яттю" системи.



Гібридний Пошук

Поєднання векторного (змістовного) та ключового (BM25) пошуку для максимальної релевантності.

Захист від Галюцинацій

Система шукає схожі документи в архіві перед тим, як надати відповідь, забезпечуючи достовірність.

Гібридний пошук є критично важливим для точного розпізнавання технічної документації та артикулів.

Реалізація Сценаріїв в n8n

Програмна реалізація в n8n забезпечує надійність та ефективність обробки.

1

Основний Пайплайн

Автоматизований процес обробки документів від входу до інтеграції з обліковою системою.

2

Human-in-the-Loop

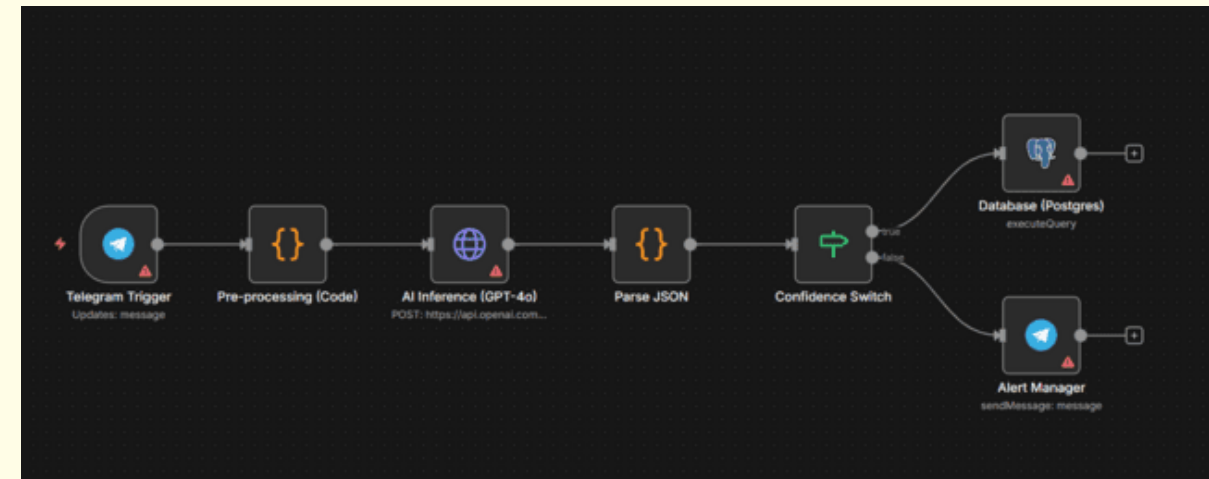
Якщо впевненість розпізнавання нижче 85%, документ надсилається менеджеру в Telegram для підтвердження.

3

Маршрутизація Помилки

Автоматична обробка та маршрутизація помилок для забезпечення безперебійної роботи.

Ця логіка гарантує **100% цілісність даних** в ERP-системі, мінімізуючи ризики помилок.





Моніторинг продуктивності: Prometheus та Grafana в дії

Для забезпечення стабільної та ефективної роботи нашої системи обробки документів ми впровадили комплексний моніторинг на базі Prometheus та Grafana. Це дозволяє в реальному часі відстежувати ключові показники та оперативно реагувати на будь-які відхилення.

Технічні та бізнес-метрики: Глибоке Занурення

Технічні Метрики

Ми відстежуємо критичні показники продуктивності системи:

Latency: час відгуку системи, що впливає на швидкість обробки.

Error Rate: відсоток помилок, що вказує на стабільність роботи.

Resource Utilization: використання CPU, пам'яті, диска для оптимізації інфраструктури.

Бізнес-Метрики

Крім технічних, ми контролюємо метрики, що безпосередньо впливають на бізнес-показники:

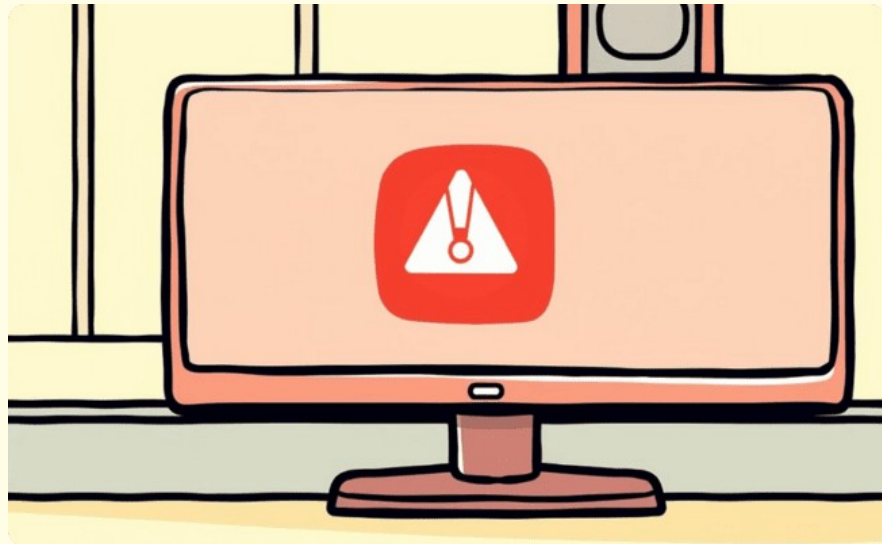
Cost-per-token: витрати на обробку одиниці даних, важливі для бюджетування.

Обсяг оброблених документів: ключовий показник ефективності автоматизації.

Time-to-process: час, необхідний для повної обробки документа.



Миттєве Реагування: Система Алертів



Налаштовані Пороги

Визначено критичні пороги для всіх ключових метрик, щоб ідентифікувати потенційні проблеми до того, як вони вплинуть на бізнес.



Автоматичні Повідомлення

У випадку перевищення порогів, система автоматично надсилає сповіщення відповідальним особам через Slack, email та інші канали зв'язку.



Швидке Вирішення

Оперативні алерти дозволяють команді швидко локалізувати та усунути несправності, мінімізуючи час простою та втрати.

Система алертів є невід'ємною частиною нашого моніторингу, забезпечуючи проактивне управління ризиками та безперебійну роботу.

Експериментальні Результати: Перетворення Процесів

Впровадження нової архітектури ІСУД принесло значні покращення у ключових аспектах обробки документів. Ось основні досягнення:

30-90 хв->3-5 хв

Зниження Часу Обробки

Феноменальне скорочення часу обробки з декількох годин до лічених хвилин.

98%

Точність OCR

Висока точність оптичного розпізнавання символів, що зменшує потребу у ручній корекції.

95%

Точність RAG

Вражаюча точність Retrieval Augmented Generation, що гарантує якісні відповіді та пошук.

80%+

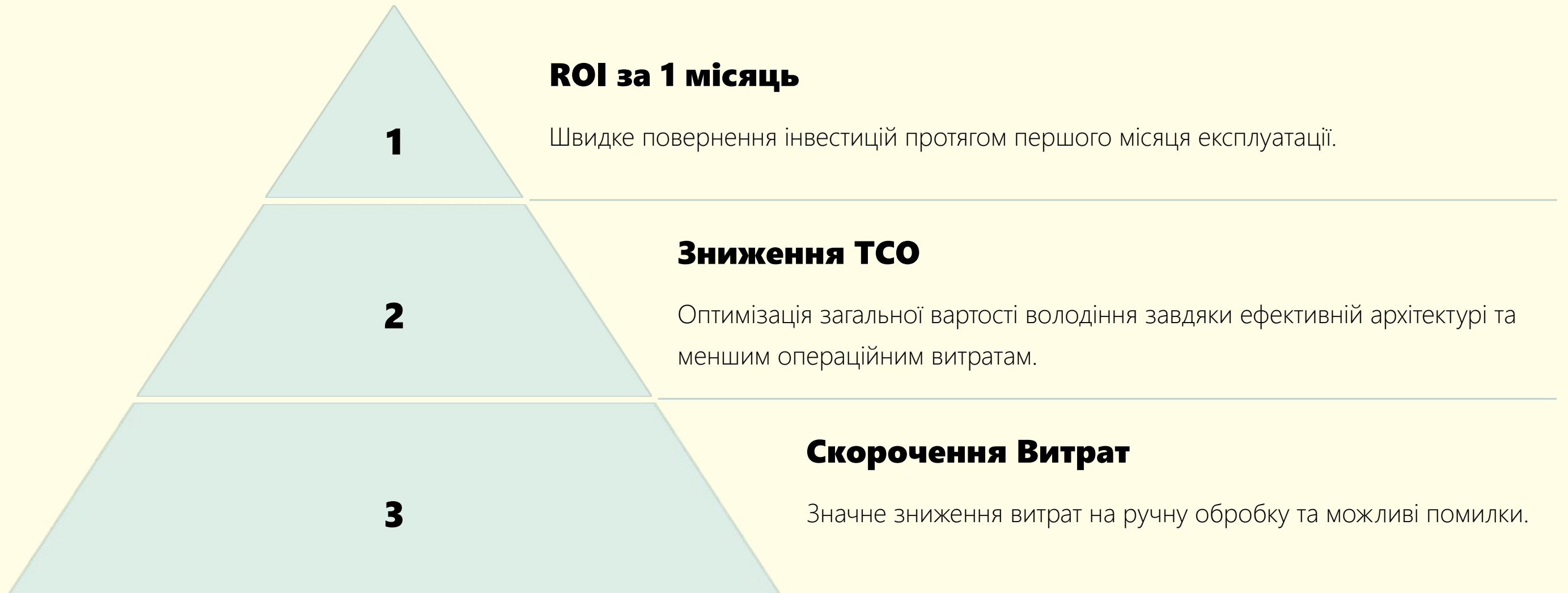
Економія Люд.-годин

Суттєва економія людських ресурсів завдяки автоматизації рутинних завдань.



Економічна Ефективність Впровадження

Впровадження нової системи не лише покращило оперативні показники, але й забезпечило значну економічну вигоду для компанії.



Ключові Досягнення Проекту

Цей проект успішно реалізував комплексний підхід до автоматизації документообігу, що принесло вагомні результати:



Створено Архітектуру ІСУД

Розроблено надійну та масштабовану архітектуру інтелектуальної системи управління документами.



Автоматизація Обробки ТТН

Повна автоматизація процесу обробки товарно-транспортних накладних, мінімізуючи ручну працю.



Реалізовано RAG та Сем. Пошук

Впроваджено Retrieval Augmented Generation та семантичний пошук для більш точного та швидкого доступу до інформації.



Побудовано Систему Моніторингу

Забезпечено комплексний моніторинг на базі Prometheus та Grafana для контролю продуктивності та стабільності.



Економічна Доцільність Підтверджена

Підтверджено високу економічну ефективність впровадження системи з чітким ROI.



Перспективи Розвитку Системи

Ми не зупиняємось на досягнутому і вже плануємо наступні кроки для подальшого розширення можливостей системи та її інтеграції в екосистему компанії.

Інтеграція з ERP

Глибока інтеграція з корпоративною ERP-системою для наскрізної автоматизації бізнес-процесів.



Автоматизація FSC-документів

Повна автоматизація обробки всіх документів FSC (Financial, Supply Chain), що вимагає специфічного підходу.



Мобільний Застосунок для Водіїв

Розробка зручного мобільного застосунку для водіїв для оперативного внесення даних та взаємодії.



Аналітичні Моделі Прогнозування

Впровадження аналітичних моделей для прогнозування обсягів документів, оптимізації маршрутів та ресурсів.

Дякую за увагу!

