paralegal
paralegal

paralegal

is a local AI tool that anonymizes confidential information

Up to 80% time savings

0% leakage risk • Fully offline

Confidential data can be anonymized automatically —
and safely all processing happens on the client's device

Problem

Organizations of all sizes face a critical challenge

- Documents contain sensitive personal data: names, addresses, dates, ID numbers.
- They cannot be uploaded to cloud AI services — risk of data leakage and legal violations.
- Manual anonymization takes hours and leads to human errors.
- Companies must comply with internal policies, as well as national and international data protection laws (PDPA, GDPR).
- There are no safe, automated, fully local solutions. The market needs a fast, accurate, fully local anonymization tool.

Why now?

- The volume of legal and corporate documents grows every year.
- Regulators increase confidentiality requirements.
- Employees increasingly use cloud AI tools without IT/Compliance oversight.
- Data breaches have become much more expensive.
- Opportunity: local AI anonymization is an almost empty niche.

Potential users

Law firms, Legal departments

Banks & Fintech (KYC, Compliance)

Big Four, Auditors

Government sector

HR & corporate departments

AI teams needing anonymized datasets

Use cases

Safe use of cloud LLM services

Corporate specialists work with client documents daily. They copy fragments of contracts, letters, applications, and paste them into ChatGPT, Claude, or other cloud AI tools — **often without prior anonymization**.

This creates direct risks of personal data leaks, exposure of sensitive business information, and internal correspondence.

Preparing datasets for internal model training

AI teams train internal LLM/NLP models on corporate data — but raw documents contain extensive personal information.

Manual anonymization at scale is impossible, while regex-based scripts are low-quality and risky.

Companies often abandon valuable training data entirely, losing the opportunity to build accurate in-house models improving operations.

Sharing documents with external contractors

Legal, HR, audit, and analytics teams frequently send documents to consultants and partners. These documents often contain personal data, addresses, contract numbers, internal messages, or medical/financial information.

Sending them in raw form is prohibited. Manual anonymization is slow and error-prone.

Secure internal communication & document transfer

Teams constantly exchange documents across departments. PDFs, emails, chat fragments, comments — all may contain sensitive personal or corporate data.

This creates risk, especially in large organizations.

Mass document processing for audits & due diligence

During due diligence or external audits, companies must provide hundreds or thousands of documents — many containing personal data.

Manual redaction at this scale is unrealistic and outsourcing is expensive and slow.

Use in call centers, banking operations, and support teams

Support staff often copy customer dialogues into QA systems or AI analysis tools. These contain PINs, passport numbers, contacts, addresses, complaints, or financial/medical information.

Uploading such data to cloud systems is prohibited.

Solution

Paralegal is a local AI tool that automatically detects and removes personal data in any legal or corporate documents.

It works fully offline, enabling secure anonymization across all workflows — from contractor document sharing to dataset preparation and safe use of cloud LLMs.

Paralegal allows employees to safely use modern AI tools, and enables companies to integrate AI into daily processes without leakage risks.

It processes both single documents and large batches of thousands of files, producing legally safe versions for sharing, analysis, audits, and machine learning.

This reduces compliance risks, speeds up processes, and saves time — from days to hours.

Works fully offline • No clouds, APIs, or external servers.

Paralegal capabilities

- Detects names, addresses, birth dates, ID numbers, and other personal data
- Returns structured entities with precise span indices
- Replaces sensitive data with tokens ([NAME], [ADDRESS], [DOB])
- Produces only valid JSON
- Runs on a local LLM via LM Studio

Advantages Over Alternatives

Manual anonymization → slow, expensive, error-prone

Regex & classical NLP → low accuracy, misses context

Cloud AI services → confidentiality violations, unusable in banks & government

Paralegal → local, accurate, structured, safe

Mini demo

Before:

“Ivan Ivanovich Ivanov resides at 14 Navoi Street, has a diagnosis of hypertension, and lives with his wife Maria Ivanovna...”

After:

“[NAME] resides at [ADDRESS], has a diagnosis of [MEDICAL CONDITION], and lives with [FAMILY MEMBER]...”

JSON Output:

{ "entities": [{"kind": "PERSON NAME", "start": 0, "end": 21, "before": "Ivan Ivanovich Ivanov", "after": "[NAME]"}, {"kind": "ADDRESS", "start": 45, "end": 60, "before": "14 Navoi Street", "after": "[ADDRESS]"}, {"kind": "MEDICAL CONDITION", "start": 62, "end": 73, "before": "hypertension", "after": "[MEDICAL CONDITION]"}, {"kind": "FAMILY MEMBER", "start": 101, "end": 116, "before": "his wife Maria Ivanovna", "after": "[FAMILY MEMBER]"}] }

Architecture

Pipeline

- Document upload
- Local category preprocessing (strict typed masks)
- Semantic analysis by a local LLM using a system prompt
- Merge layer combines results and produces final JSON
- UI displays Before/After (human-in-the-loop)

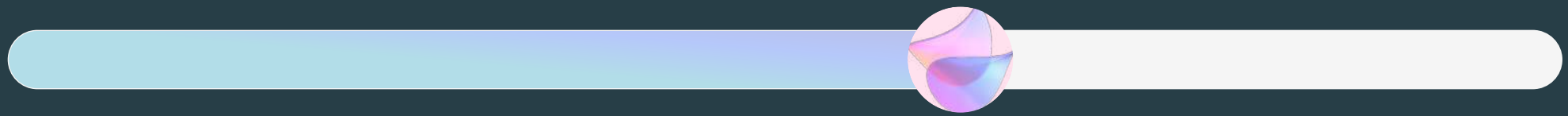
Technologies

- Frontend: React + TypeScript
- Backend: NestJS (Node.js)
- AI: LM Studio (local model)
- Zero-Cloud / Local-First
- Privacy-By-Design

Roadmap

Completed:

- API layer & JSON validation
- NestJS backend
- Before/After UI
- Local LLM integration
- Automated redaction pipeline



Current status: Prototype

Next steps:

Improve accuracy & JSON stability •
Desktop version + DOCX/PDF support •
Pilot client •

Why Us

- 12+ years of legal expertise (deep understanding of compliance & privacy)
- Production-level backend & frontend development
- Strong experience in LLMs, JSON validation & span indices
- Proven ability to build local AI systems
- Multidisciplinary team: PM, AI integrator, backend, frontend, architecture
- Unique combination: law + engineering + AI.

IGOR PAK

Project Manager

Python, LLM integration, legal expertise

GitHub: /iPakTulane

LinkedIn: /in/igor-pak/

EDEM VELIEV

Fullstack Developer

Node.js, TS - NestJS, React, LLM integration

GitHub: /Edem2000

LinkedIn: /in/edem-veliev/

All materials presented on this website are the intellectual property of the Stuntman team.