Document classification for a global procurement services provider

AI that labels every file across a company's shared drives for category, retention and sensitivity, checks each label twice and takes the safe option when in doubt.



INDUSTRY

Procurement Services

PUBLISHED

February 2026

READ ONLINE

samuylov.ai · Document classification for a global procurement services provider

SERVICES

Document AI · LLM-powered classification · Audit trail and human review · Cloud deployment and hand-over

Client

A global procurement services provider, running sourcing and supplier operations on behalf of large enterprises across dozens of countries.

Challenge

Procurement runs on documents: contracts, amendments, invoices, supplier correspondence, certificates. They accumulate for years across shared drives and collaboration tools, in inconsistent formats and several languages, and each one carries obligations – how long it must be kept, who may see it, which process owns it.

The files sat across shared drives and team collaboration sites: more than a million of them, to be sorted into 14 business categories and five sensitivity levels. Every file had to be listed, read and labelled, in runs that could stop and resume without losing work.

The client's existing classifier depended on metadata that many of these files lacked. And a wrong label is not cosmetic: it can delete a file that must be kept, or expose one that must not be seen. The system could not guess, and with a fixed compliance deadline and this many files, labelling by hand was not an option.

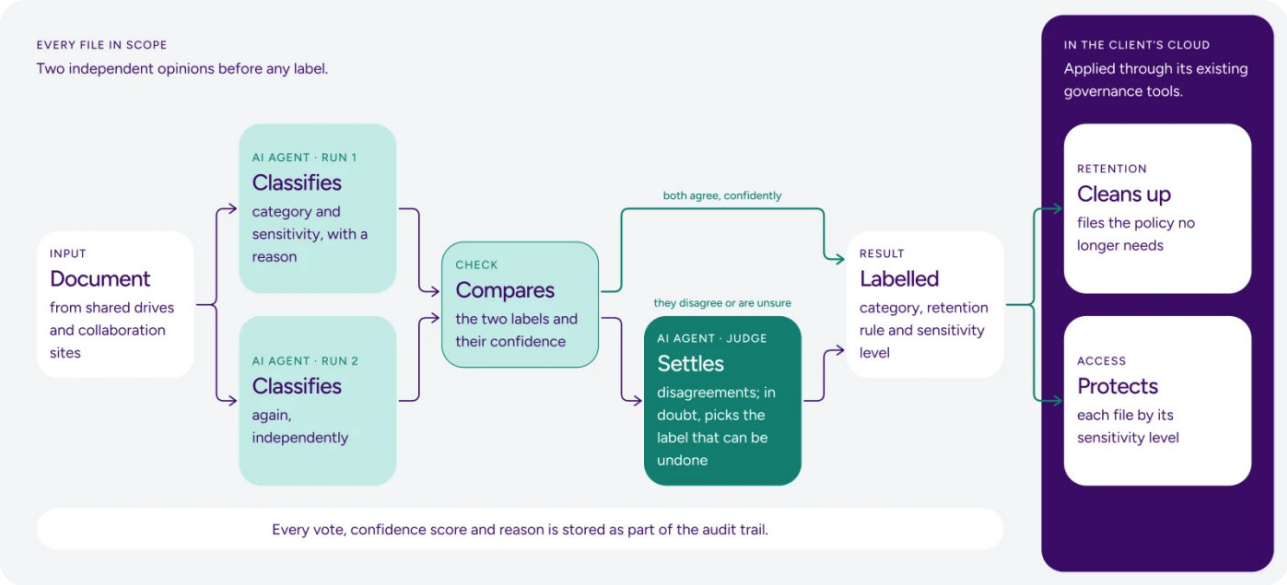
Our approach

We built the system so that confidence decides what happens to each file.

Two opinions, then a judge. Two AI agents classify every document independently. When they agree with confidence, the label stands. When they disagree or are unsure, a third agent, the judge, weighs both explanations and decides.

Doubt defaults to the safe side. When the evidence is ambiguous, the system picks the label that can still be undone. Every agent's vote and reasoning is stored in an audit trail, so people can review any decision and nothing irreversible rests on a guess.

It runs in the client's cloud. The system is set up from code in the client's own cloud infrastructure, and writes labels through the governance tools the client already uses. Runs pace themselves to the API limits, pick up where they stopped after an interruption, and produce the reports compliance needs.



Every file gets two independent opinions before it is labelled. Disagreements go to a judge, and doubt always lands on the label that can still be undone.

Results

Every shared drive and collaboration site in the project was scanned and labelled, and around a hundred thousand outdated files were then deleted under the client's retention policy. We kept no copy of the documents.

What transfers

Label every file once, with a recorded reason, and retention and access rules can run automatically from then on.

Confidence should decide what happens to a file, not just sit in a log. Decide up front which labels are applied automatically, which fall back to the safe option, and how each decision is recorded. That record is what makes a label defensible in an audit.



Denis Samuylov
– Founder & CEO

Founder-led, team-delivered – Denis leads the engagement and stays on it after the pitch.

contact@samuylov.ai +41 79 136 38 26 Book an intro call →