

HM LAND REGISTRY: USING AI FOR INTELLIGENT DOCUMENT COMPARISON

Find out how we used artificial intelligence and machine learning to automate manual tasks, improving efficiencies and reducing costs.

About HM Land Registry

HM Land Registry (HMLR) safeguards land and property ownership valued at £7 trillion, enabling over £1 trillion worth of personal and commercial lending to be secured against property across England and Wales. The Land Register contains more than 25 million titles showing evidence of ownership for more than 87% of the land mass of England and Wales.

HMLR employs approx. 4,300 case workers across 14 UK locations to process and review applications and registrations. HMLR has a strategic objective to become a leading global digital land registry and has specific strategic aims to automate aspects of simple casework, where possible. A significant amount of the casework carried out is highly manual and time-consuming and is ideally suited to automation driven by Artificial Intelligence (AI).

Manually comparing documents

When a new building development is planned, a precedent transfer deed is agreed with HMLR, with the intention that the transfer deeds submitted for registration on each individual plot sale should match that precedent. At present, the transfer submitted for registration must be compared to the precedent transfer to ensure there are no significant differences.

Deed comparison is a manual task requiring a caseworker to examine both documents in detail, identify differences and assess the impact. To further complicate the task, deeds are often long and contain complicated legal terminology. The differences between documents are often minor, and pinpointing these is time-consuming, repetitive and can understandably be error prone.

As a longstanding digital partner to HMLR, Kainos' objective was to use AI to automate this document comparison task, freeing up caseworkers to focus on more valuable work, while also speeding up the review time and easing the entire process for applicants.

Creating an AI solution

The goals of the project included:

- Creating an AI solution that could flag discrepancies in what were expected to be matching documents
- Identifying the precise locations of differences in the documents, to help ease the burden on caseworkers
- Seamlessly integrating a solution into the caseworkers' workflow to ease the burden of existing operational processes

Our approach to this problem followed a journey from exploration to solution:

- 1. Exploratory data analysis:** Building familiarity with the data, exploring document types, differences in quality and potential challenges.
- 2. OCR:** Investigating the accuracy and reliability of available cloud OCR tools and services across a range of HMLR data sets.
- 3. Image analysis:** Investigation using digital fingerprinting to analyse the visual structure of HMLR data to aid in comparison.
- 4. Image-based classification:** Development of custom AI modules capable of identifying HMLR specific templates in documents.
- 5. Legal language intelligence:** Investigation of finding, counting, and classifying different types of HMLR specific legal language, e.g. easements, covenants.
- 6. Document comparison:** Determining an approach to successfully compare documents using AI.

This journey included exploring whether an image-based recognition approach (page-by-page comparison) using Siamese neural network or an OCR-based recognition approach using AWS Textract and AWS Comprehend would be more effective. We found that an OCR approach leveraging AWS Textract was the most straight-forward to configure and delivered the best results out of the potential tools and services identified.

The resulting solution was development of a web application for caseworkers that allowed upload of transfer documents as PDFs. The web application used a range of machine learning (ML) and AI techniques to intelligently identify significant differences within complex legal language across documents and highlight these for on-screen review for caseworkers to action. This considered and handled multi-format differences, including: sections, text, signatures, checkboxes and dates across multiple document versions.

The AI document comparison tool is now in production and in use by caseworkers across HMLR. A key component built into the solution is capturing data as it is used. In the future, this will enable the ML models to be iteratively trained to assess the significance of a difference based on caseworker feedback.

Benefits



Improved quality and efficiency: HMLR staff no longer have to manually compare thousands of documents each week.



Reduced risk of indemnity claims: Intelligent automation detects 100% of the differences between documents.



Significant time and cost savings: manual document review can take a caseworker 10-15 minutes. Automating it reduces this time by half.

// As part of our ongoing strategic aims to become the world leading digital land register, this project has been a successful first step in our journey towards embracing artificial intelligence to improve our service and enhance how our employees work.

HMLR Representative