

AI-Powered Data Extraction Saves Hundreds of Hours

AT A GLANCE

Challenges

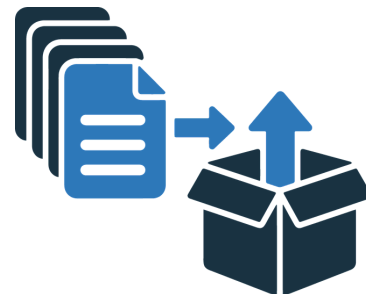
- Unorganized Alignment data
- Content unsuitable for fixed-rule parsing
- Large files exceeding AI context limits
- Risk of inaccurate AI-generated results

Methods

- AI-assisted extraction from Excel and PDF files
- Model-based review of extracted information
- Source citations and deterministic validation
- Automated correction of failed results

OBJECTIVES

iCEV needed to extract lesson-to-standard alignment data from hundreds of Excel and PDF files and convert it into a structured format ready for database ingestion. The files varied significantly in structure and often contained freeform text combining lesson names with location references such as presentation slide numbers. A fixed extraction process could not reliably handle that variation, while manual processing would have required hundreds of hours. Because the resulting data would support iCEV's AI-assisted alignment processes and the creation of alignment documents, the solution also needed safeguards against inaccurate AI-generated results.



SOLUTIONS

IntelliTect developed a C# and .NET application using Microsoft Foundry and OpenAI models to interpret unstructured and semi-structured data across iCEV's file library. For large files, intelligent chunking preserved the hierarchy and context needed to interpret individual records. Parallel processing allowed the application to work through those sections efficiently. IntelliTect also built verification into the workflow. Along with each extracted lesson, the model returned a location in the source file. Deterministic code checked the lesson against that citation. Results that could not be verified were automatically returned to the model for correction, while an additional model-based review checked the data for consistency. The application produced structured output designed for direct database ingestion. IntelliTect completed and delivered the solution in 56 hours over two weeks.



BENEFITS



Hundreds of Hours Saved

Automated extraction replaced a repetitive process that would have required hundreds of hours review.



Greater AI Confidence

Source citations, deterministic checks, and model-based review help to identify questionable results early.



Structured Data Ready for Use

Easy data ingestion into iCEV's database, where it could support alignment and document creation.

