CV Case Study:

Subjective Scene Classification

## for Improved
Search Experience

Project

A leading hospitality marketplace partnered with Alegion to build a scalable data labeling pipeline for 30+ workflows ranging from competitor cross-listing to building type validation. Among those projects were ones focused on determining if a listing location was a primary residence, a secondary residence like a vacation home, or a dedicated rental property. There was also a need to classify what type of building a listing was in and if a private room was provided. The goal was to ensure that listing information that customers saw online was accurate and provided the best possible search experience.

Challenge

Every use case comes with a unique set of labeling requirements, and it takes expertise to not only maintain, but also increase labeling quality as projects scale from pilot to production stages. In order for the machine learning model to be effective, the client team needed to maintain a minimum accuracy of 80%. Prior to working with Alegion, the client team was only able to achieve 60% accuracy. Getting to at least 80% accuracy would represent an increase in data quality while also scaling the volume of data annotated. They also needed the ability to test, iterate, and learn about category definitions without adding cycles in order to extract answers from data and speed up decision making.
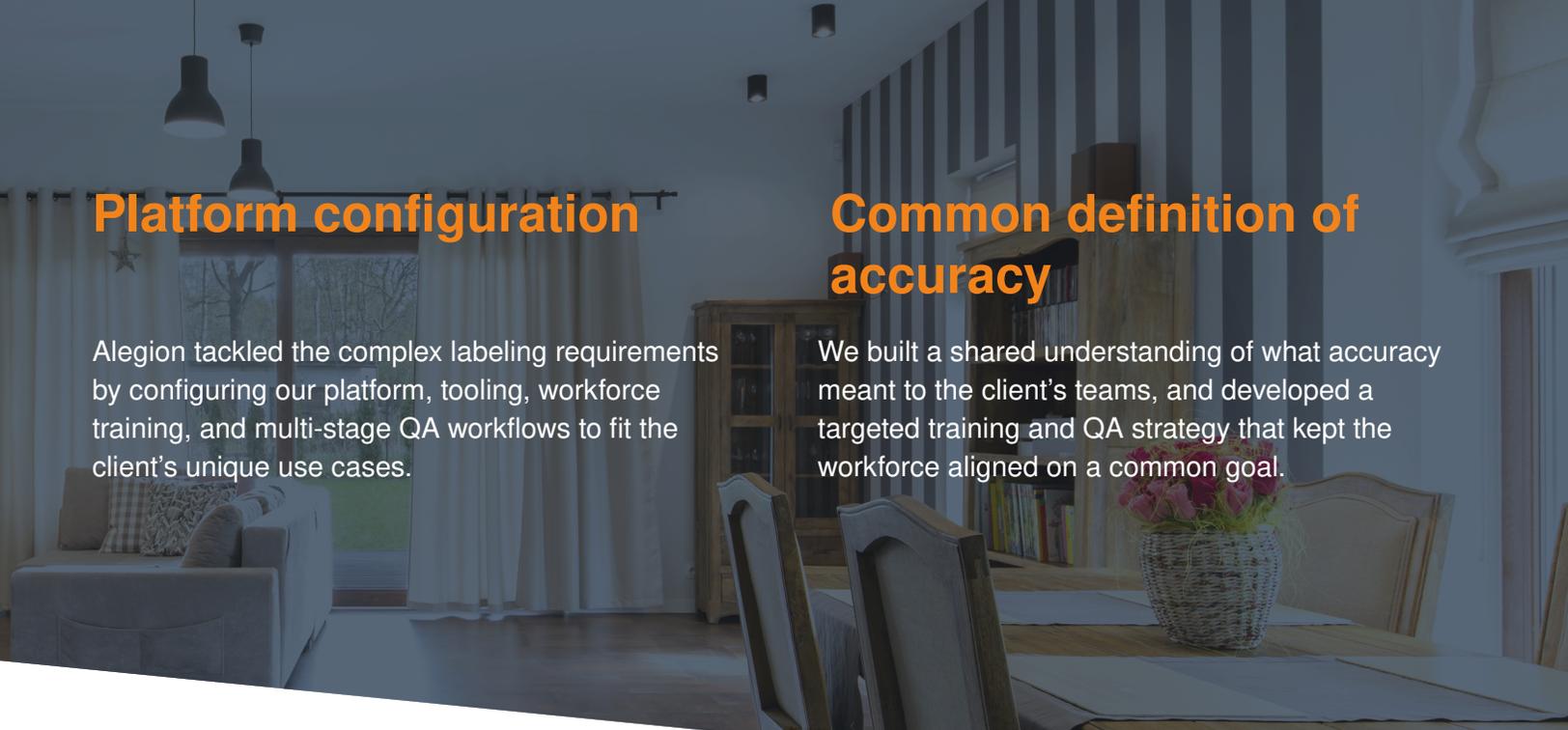
Platform configuration

Alegion tackled the complex labeling requirements by configuring our platform, tooling, workforce training, and multi-stage QA workflows to fit the client's unique use cases.

Common definition of accuracy

We built a shared understanding of what accuracy meant to the client's teams, and developed a targeted training and QA strategy that kept the workforce aligned on a common goal.

Data exchange automation

We streamlined the data exchange via API to replace CSV data transfers, mitigating the risk of errors and redundancy resulting from an otherwise manual process.

Training

It was imperative that annotators understand the client's definition of accuracy, including the annotation criteria and attribute taxonomies. Our customer success team screened, trained, and scored the workforce on Gold Data (known correct answers) to ensure quality output. Annotators were prequalified for the client's project. Scoring was based on Gold Data, time on task, judgement confidence levels, and ongoing annotator performance. Annotators that accurately labeled the Gold Data consistently stayed on the project, while annotators who's judgements were overturned incurred a negative score. High-performers were given more complex tasks and low performers were re-trained or removed from the task.
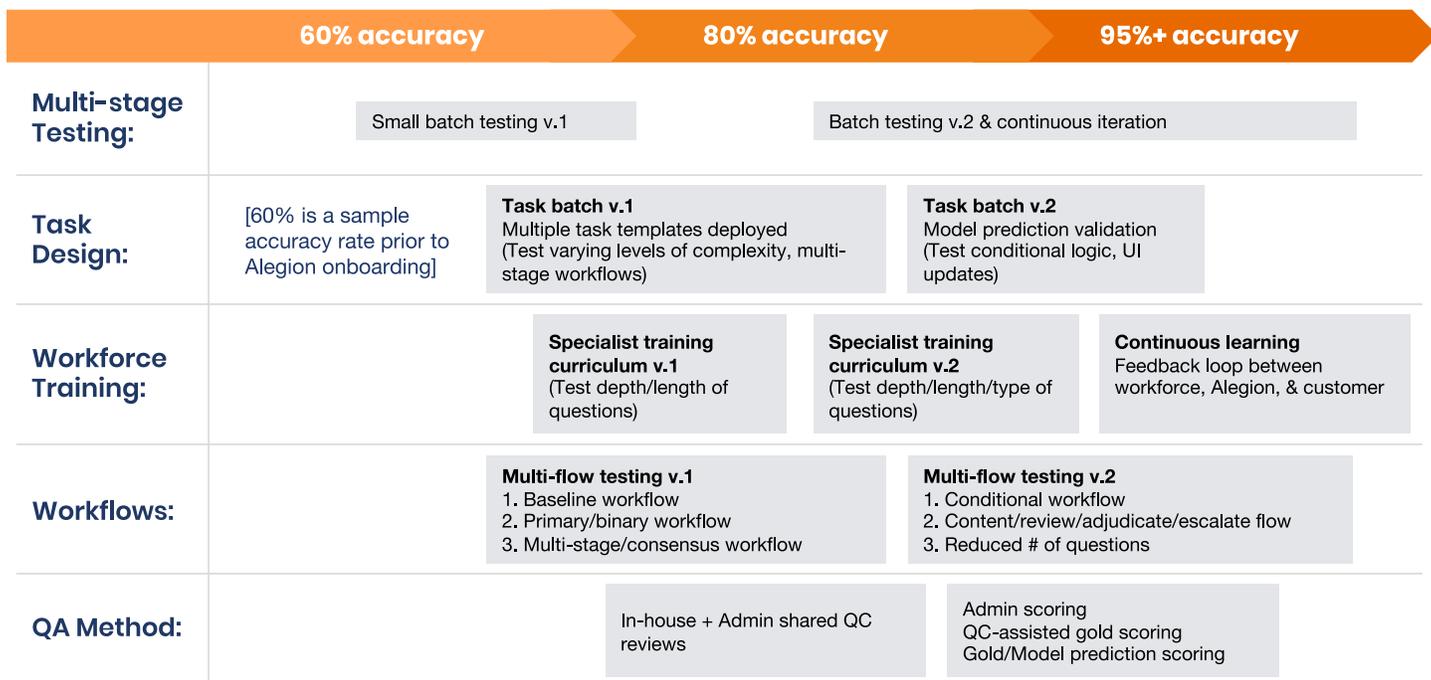
Process

Alegion began by testing 2 different use cases with a total of 5 different workflows, one control workflow and four experimental workflows using the same input records. One workflow added supporting questions to help deal with subjectivity, one provided the answer that needed to be validated, one provided an open-ended question, one provided the option to escalate at each stage rather than make a judgement, and finally the last experiment had 5 reviewers with consensus measured by 3/5 agreement.

The learnings from each experiment were then taken and applied, such as using supporting questions that helped improve quality and reducing the number of reviewers, as having too many caused indecision and resulting in lower quality. The final workflow reduced the number of supporting questions, made use of a double blind review, a QC adjudicator, and a final reviewer that compared the answers to the ML predicted result.

TESTING & ITERATION: a path to accuracy and consistency

	60% accuracy	80% accuracy	95%+ accuracy
Multi-stage Testing:	Small batch testing v.1	Batch testing v.2 & continuous iteration	
Task Design:	[60% is a sample accuracy rate prior to Alegion onboarding]	**Task batch v.1** Multiple task templates deployed (Test varying levels of complexity, multi-stage workflows)	**Task batch v.2** Model prediction validation (Test conditional logic, UI updates)
Workforce Training:		**Specialist training curriculum v.1** (Test depth/length of questions)	**Specialist training curriculum v.2** (Test depth/length/type of questions) — **Continuous learning** Feedback loop between workforce, Alegion, & customer
Workflows:		**Multi-flow testing v.1** 1. Baseline workflow 2. Primary/binary workflow 3. Multi-stage/consensus workflow	**Multi-flow testing v.2** 1. Conditional workflow 2. Content/review/adjudicate/escalate flow 3. Reduced # of questions
QA Method:		In-house + Admin shared QC reviews	Admin scoring QC-assisted gold scoring Gold/Model prediction scoring

//ALEGION

Quality Management

Labeling accuracy is the top priority. By testing QA methodologies and evaluating the results against gold data, Alegion QA, and model predictions, we were able to increase both labeling accuracy and throughput. Small-batch testing and iteration of the quality parameters enabled us to not only hit target thresholds but delivered insights that helped reshape the client's problem statements.

Impact

The result of this iterative testing process and strategically applying learnings was improving accuracy from 60% to 98%. The improvements in data quality improved model performance that drove the customer search experience. The additional insights gained have also allowed the client team to optimize internal resources, reducing operating costs and improving efficiency.