Guidance for Intelligent Document Processing with AWS
Automate document processing with Amazon Textract, Amazon Comprehend, and Amazon A2I.

1. Documents are uploaded to Amazon S3, which invokes an AWS Lambda function, starting an Amazon Textract asynchronous detection job.
2. Amazon Textract outputs to an Amazon S3 bucket location, then sends a completion notification to Amazon Simple Notification Service (Amazon SNS).
3. The Amazon SNS topic sends the completion message to the Amazon Simple Queue Service (Amazon SQS) queue.
4. A Lambda function is invoked by the SQS queue to process and read the Amazon Textract output.
5. The Lambda function calls an Amazon Comprehend custom classifier async operation to classify the documents. Comprehend outputs to an S3 bucket.
6. A Lambda function is invoked by the S3 bucket. It sorts the input documents on the basis of classes determined by Amazon Comprehend and places the documents into another S3 bucket location.
7. The S3 bucket with the classified document invokes a Lambda function that can (a) call Amazon Textract sync or async APIs for data extraction, (b) call the Amazon Comprehend pre-defined or custom name entity recognizer to detect personal (PII) or health information (PHI), and (c) perform further document enrichment with medical insights with Amazon Comprehend Medical.
8. The data from the AI service calls and any enriched documents are saved in an S3 bucket location.
9. A Lambda function is invoked from the S3 bucket. The function runs review and validation on the data using predefined rules. It also checks accuracy scores and sends the information for human review if threshold scores are not met.
10. A human completes the review and uses Amazon Augmented AI (Amazon A2I) to update the appropriate information in to the S3 location, which initiates another validation using the Lambda Function.
11. A Lambda function stores the extracted and verified data in an Amazon DynamoDB table.
12. Lambda sends a notification that all rules were verified correctly or if any information needs further human review.