Amazon Textract AnalyzeID

AWS AI Service Cards

AWS AI Service Cards: Amazon Textract AnalyzeID

Table of Contents

Amazon Textract AnalyzeID

An AWS AI Service Card explains the use cases for which the service is intended, how machine learning (ML) is used by the service, and key considerations in the responsible design and use of the service. A Service Card will evolve as AWS receives customer feedback, and as the service progresses through its lifecycle. AWS recommends that customers assess the performance of any AI service on their own content for each use case they need to solve. For more information, please see AWS Responsible Use of AI Guide and the references at the end. Please also be sure to review the AWS Responsible AI Policy, AWS Acceptable Use Policy, and AWS Service Terms for the services you plan to use.

This Service Card applies to the release of Textract AnalyzeID that is current as of November 7, 2022.

Overview

The Amazon Textract service extracts printed text, handwriting, and structured data from images of documents. Within this service, the AnalyzeID feature reads and extracts structured text data from images of identity documents, currently including US driver's licenses and US passports. This feature makes it easier for customers to automate and expedite their document processing.

AnalyzeID operates on the text that appears in an identity document to predict explicit and implied key-value pairs. Analyze ID can extract explicit key-value pairs, where a key ("Date of Issue") appears on the document and is aligned with its value ("03/18/2018"), and implied key-value pairs that may not have explicit keys appearing next to them ("María" appears in the center of a license, but is not marked as "First Name"). The service normalizes key-value pairs into a common taxonomy of 21 known keys, so that customers can compare information across ID types. For example, the service extracts the LIC# of a driver's license and Passport No. from a US passport, labeling both as "Document ID Number." To evaluate the accuracy of AnalyzeID, we compare these predictions to ground truth. Ground truth keys and values have been corrected by a human annotator. Each predicted key-value pair is a hit if the prediction matches ground truth, and a miss otherwise. Quality metrics like precision, recall, and F1 depend on the number of hits and misses.

Identity documents differ between jurisdictions (Virginia driver's licenses are distinct from California licenses) and within jurisdictions, since each jurisdiction evolves their documents over time. Each version of an identify document can differ by the keys included and by the values permitted for each key. There are also factors (called “confounding variation”) that make recognition difficult. The document designs can have complex graphic designs backing the text, and glossy plastic or other laminates overlaying the text. Documents can suffer wear and tear, e.g., from being carried in a purse or pocket, which obscures key information. Finally, the document can be poorly lit, occluded (such as from a portion of the hand holding the document during imaging), or not well-focused when imaged. AnalyzeID is designed to recognize text in these document images, ignoring the confounding variation.

Intended use cases and limitations

AnalyzeID is intended for use on driver’s licenses issued by US states and passports issued by the US government. It has not been trained for use on documents issued by territorial governments (for example, Puerto Rico) or on other forms of identification such as global entry cards or birth certificates. AnalyzeID supports documents issued in the last fifteen years (2007). This covers roughly three iterations of state-level design updates, which happen approximately every five years. This time frame supports all unexpired documents; US driver's licenses expire in at most twelve years, and US passports expire in at most ten.

AnalyzeID enables the text extraction step in a variety of customer-developed applications. These applications typically support end users in completing online tasks. For example, a financial service application could enroll new users with less typing and errors by allowing the user to scan the contents of their driver’s license. Similarly, a healthcare application could allow users to confirm their address or other account information more rapidly and with fewer errors when scheduling appointments. Applications differ primarily by 1/ the key-value pairs that are relevant, 2/ the image capture process employed, and 3/ the resolution of images submitted. When building AnalyzeID into any application workflow, customers should assess the need for human oversight and support the review of AnalyzeID output by human reviewers as needed.

Design of Textract AnalyzeID

Machine learning

> AnalyzeID is built using ML and optical character recognition (OCR) technologies. It works as follows: AnalyzeID takes an image of an identity document as input. An OCR model identifies text in the document. A second machine-learned model processes the full document image plus OCR output to return field names and contents as key-value pairs. For more information, see Analyzing Identity Documents in the *Amazon Textract Developer Guide*.

Performance expectations

Confounding variation will differ between customer applications. This means that performance will also differ between applications. Consider two different name and address verification applications A and B. Application A enables a building security agent to compare the name and address on a visitor's driver's license with the name and address of the person expected to visit the site. Application B enables a recruiter to collect personal information from an applicant during a video interview. With A, the building security agent uses an enterprise ID document scanner to capture license images that are well lit, sharply focused, and unoccluded. With B, the interviewee uses their own webcam to capture an image of their license as they hold it, increasing the risk of image blur, glare and occlusions. Because A and B have different input image qualities due to different image capturing devices and processes, they will likely have differing error rates, even assuming that each application is deployed perfectly using Textract.

Test-driven methodology

We use multiple datasets to evaluate performance. No single evaluation dataset provides an absolute picture of performance. That's because evaluation datasets vary based on their demographic makeup (the number and type of defined groups), the amount of confounding variation (quality of content, fit for purpose), the types and quality of labels available, and other factors. We measure Textract performance by testing it on evaluation datasets containing images of identity documents. The overall performance on a dataset is represented by the F1 score (F1), which balances the percentage of predicted fields that are correct (precision) against the percentage of correct fields that are included in the prediction (recall). F1 scores are bounded by the range [0,1]. Changing the confidence thresholds on the key-value pairs changes the F1 score. Groups in a dataset can be defined by key attributes (such as jurisdiction, length of last name), confounding variables (such as graphic design layout, image quality), or a mix of the two. Different evaluation datasets vary across these and other factors. Because of this, the F1 scores– both overall and for groups – vary from dataset to dataset. Taking this variation into account, our development process examines AnalyzeID's performance using multiple evaluation datasets, takes steps to increase F1 for groups on which AnalyzeID performed least well, works to improve the suite of evaluation datasets, and then iterates.

Fairness and bias

Our goal is for AnalyzeID to extract text fields from identity documents with high accuracy irrespective of the jurisdiction of the license or the demographic attributes of the person represented by the document. To achieve this, we use the iterative development process described above. As part of this process, we build datasets to capture the range of jurisdictions

(US states) and templates addressed by AnalyzeID, under a range of conditions for image quality. We routinely test on datasets of document images for which we have reliable key-value pairs. We find that AnalyzeID performs well across jurisdictional and demographic attributes. As an example, on an internal dataset composed of the front sides of US driver's licenses from 50 states, the lowest F1 accuracy between states is 95%, and the lowest F1 for demographic groups defined by age, veteran status and length of last name is 99%. Because results not only depend on AnalyzeID, but also depend on the customer workflow and the evaluation dataset, we recommend that customers test AnalyzeID on their own content.

Explainability

Customers have access to the confidence scores for each text field, which they may leverage for confidence thresholding as well as better understanding of AnalyzeID output. The predicted key provides insight into the prediction for the value.

Robustness

We maximize robustness with a number of techniques, including using large training datasets that capture many kinds of variation across many documents. Ideal inputs to AnalyzeID contain images that are relatively free from shadow, glare, or other obstructions, with the document oriented upright within the image frame. However, AnalyzeID models are trained to be resilient even when inputs vary from ideal conditions.

Privacy and security

AnalyzeID captures and processes text. Inputs and outputs are never shared between customers. Customers can opt out of training on customer content via AWS Organizations or other opt out mechanisms we may provide. For more information, see Section 50.3 of the AWS Service Terms and the AWS Data Privacy FAQs. For service-specific privacy and security information, see Amazon Textract FAQs and the Amazon Textract Security.

Transparency

Where appropriate for their use case, customers who incorporate AnalyzeID in their workflow should consider disclosing their use of ML to end users and other individuals impacted by the application, and give their end users the ability to provide feedback to improve workflows. In their documentation, customers can also reference this AI Service Card.

Governance

We have rigorous methodologies to build our AWS AI services in a responsible way, including a working backwards product development process that incorporates Responsible AI at the design phase, design consultations and implementation assessments by dedicated Responsible

AI science and data experts, routine testing, reviews with customers, and best practice development, dissemination, and training.

Deployment and performance optimization best practices

We encourage customers to build and operate their applications responsibly, as described in the AWS Responsible Use of AI Guide. This includes implementing responsible AI practices to address key dimensions including fairness and bias, robustness, explainability, privacy and security, transparency, and governance.

Workflow Design: We define performance as the experience of end users who interact with a customer-developed application that includes AnalyzeID for text extraction. The performance of any application using AnalyzeID depends on the design of the customer workflow, including: (1) image variation, (2) confidence thresholding, (3) human oversight, (4) workflow consistency, and (5) periodic testing for performance drift.

1. **Image variation:** Ideal images are relatively free from shadow, glare, or other obstructions, with the document captured at a direct angle, and oriented upright within the image frame. Customers can support their end users with appropriate guidance for capturing good images.
2. **Confidence thresholding:** Customers may tune performance by setting a filter or threshold on key-value pairs that AnalyzeID produces, based on the confidence score assigned to that pair. For better precision, choose a high threshold. For better recall, choose a lower threshold. To set an appropriate threshold, a customer may collect a representative set of inputs, label the text fields of each, and try higher or lower thresholds until satisfied with the user experience.
3. **Human oversight:** If a customer's application workflow involves a high risk or sensitive use case, such as a decision that impacts an individual's rights or access to essential services, human review should be incorporated into the application workflow where appropriate. Automatic key-value extraction with AnalyzeID can serve as a tool to reduce the effort incurred by fully manual solutions, and to allow humans to expeditiously review and assess identity documents.
4. **Consistency:** Customers should set and enforce policies for the kinds of input images permitted, and for how humans combine the use of confidence thresholding and their own judgment to determine final results. These policies should be consistent across demographic groups. Inconsistently modifying input images or confidence thresholds could result in unfair outcomes for different demographic groups.
5. **Performance drift:** A change in the kinds of images that a customer submits to AnalyzeID, or a change to the service, may lead to different outputs. To address these changes, customers

should consider periodically retesting the performance of Textract, and adjusting their workflow if necessary.

Further information

- For service documentation, see Analyzing Identity Documents.
- For related features, see Amazon Textract Documentation.
- For details on privacy and other legal considerations, see the following AWS policies: Acceptable Use, Responsible AI, Legal, Compliance, and Privacy.
- For help optimizing workflows, see Generative AI Innovation Center, AWS Customer Support, AWS Professional Services, Ground Truth Plus, and Amazon Augmented AI.
- If you have any questions or feedback about AWS AI service cards, please complete this form.

Glossary

Controllability: Steering and monitoring AI system behavior.

Privacy & Security: Appropriately obtaining, using and protecting data and models.

Safety: Preventing harmful system output and misuse.

Fairness: Considering impacts on different groups of stakeholders.

Explainability: Understanding and evaluating system outputs.

Veracity & Robustness: Achieving correct system outputs, even with unexpected or adversarial inputs.

Transparency: Enabling stakeholders to make informed choices about their engagement with an AI system.

Governance: Incorporating best practices into the AI supply chain, including providers and deployers.