

TDC SUMMIT SÃO PAULO



Como aplicar Inteligência Artificial na gestão de documentos utilizando o OCI





Graduada em Ciência da Computação e Pós Graduada em Data Science e Ux Design. Possuo 09 anos de experiência em BI e gosto de trabalhar com Visualização de Dados, ajudando as empresas a se tornarem Data Driven.

Tayná Salvador
Oracle
Senior Cloud Solution Engineer

CUSTOMER INNOVATION OFFICE



PROPÓSITO

Escutar os reais **desafios** do nosso Cliente e **cocriar** soluções **inovadoras** que **melhorem** o seu dia-a-dia.

* () >
MEU TIME

AUMENTANDO A EFICIÊNCIA

Com o uso de IA conseguimos reduzir o trabalho de dias para alguns minutos



BENEFÍCIOS DE UTILIZAR AI



Melhora a Tomada de Decisão

Aumento da Automação

Redução de Erros, Riscos e Custos Operacionais

Comodidade e Escalabilidade

Atendimento Otimizado ao Público

ORACLE APPLICATIONS, SERVICES, DATA PLATFORMS AND INFRASTRUCTURE AI

Applications

Fusion Applications

NetSuite

Fusion Analytics

Industry Applications

3rd Party Applications

AI Services

NEW



OCI Generative AI

NEW



OCI GenAI Agents



Digital Assistant



Speech



Language



Vision



Document Understanding

Data Platforms



OCI Data Science



AI Vector Search in Oracle Database



MySQL HeatWave Vector Store



OCI Data Labeling

Data

AI Infrastructure



Compute bare metal instances and VMs with NVIDIA GPUs



OCI Supercluster with RDMA networking

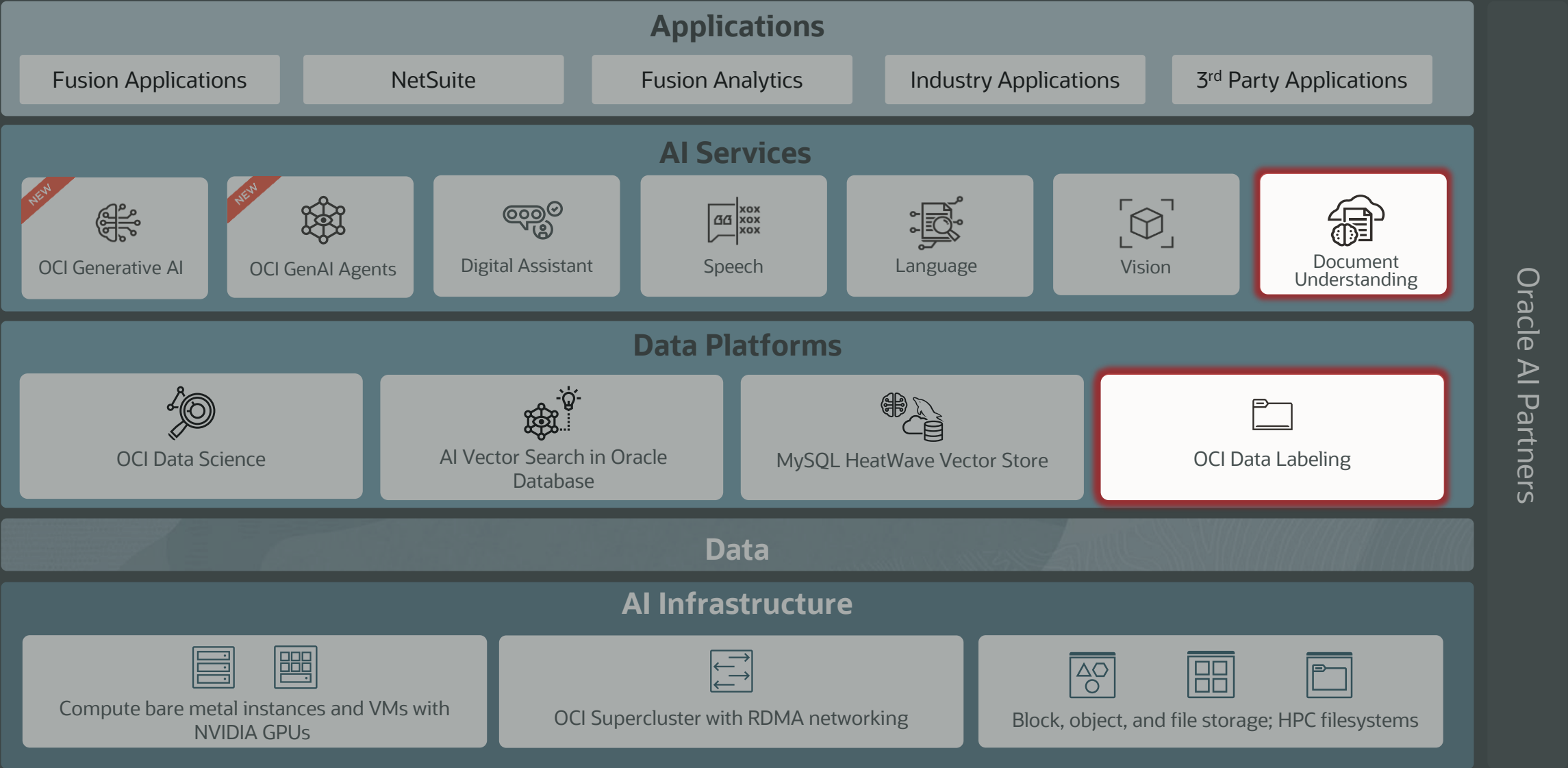


Block, object, and file storage; HPC filesystems

Oracle AI Partners



ORACLE APLICATIONS, SERVICES, DATA PLATFORMS AND INFRASTRUCTURE AI



DOCUMENT UNDERSTANDING

Automatize a extração de texto, estrutura e informações importantes de documentos comerciais



<https://www.oracle.com/artificial-intelligence/document-understanding/>

DOCUMENT UNDERSTANDING

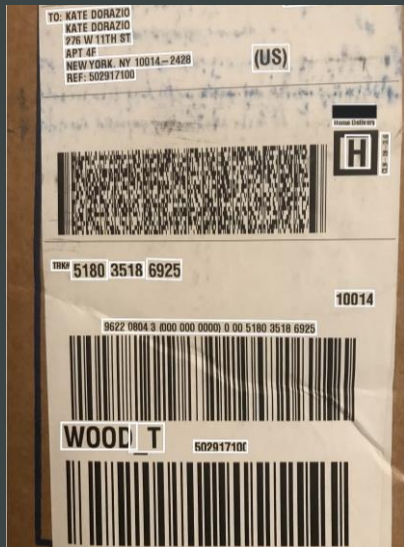
Capacidades

Extração de
Texto

Classificação de
Documentos

Extração de
Tabelas

Extração de
campos Chave



DOCUMENT UNDERSTANDING

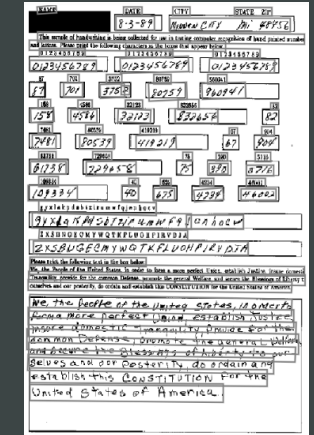
Extração de Texto (OCR)

Extraia recursos textuais de documentos, incluindo:

- Documentos impressos e manuscritos
- Documentos inclinados e sombreados
- Documentos girados

Aplicações comuns:

- Relatórios de negócios em formato PDF
- Recibos e faturas
- Seguro ou outros formulários de inscrição
- Etiquetas de produtos
- Informações sobre pacotes e equipamentos
- Conteúdo enviado pela força de trabalho para dispositivos móveis



DOCUMENT UNDERSTANDING

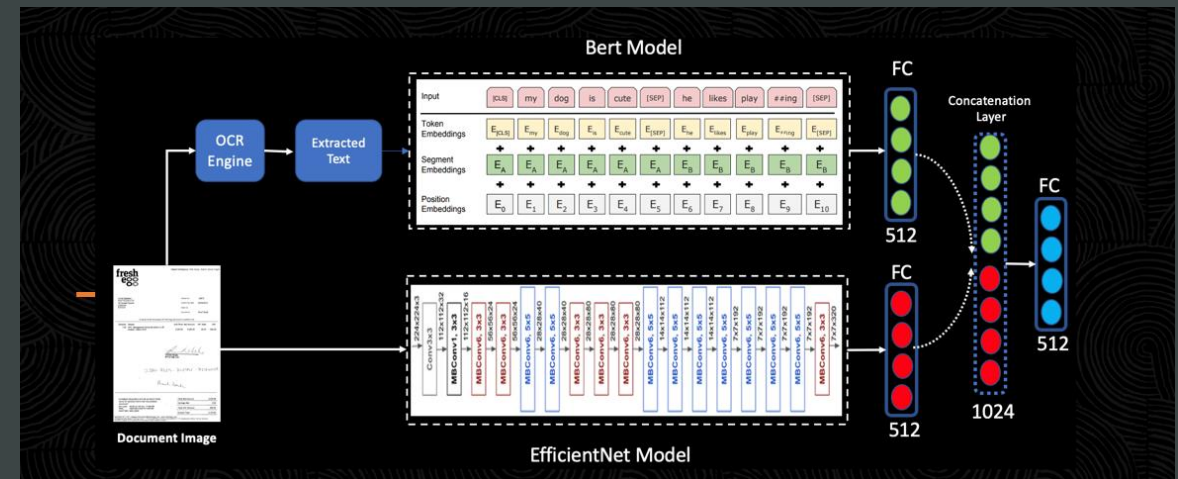
Classificação de Documentos

Classifica documentos em diferentes tipos com base na aparência visual, recursos de alto nível e palavras-chave extraídas.

- Modelo Bert (baseado em texto) + EfficientNet (baseado em imagem)
- A combinação de modelos baseados em imagem e texto torna o híbrido robusto e aumenta a precisão

As aulas de classificação pré-treinadas incluem:

- Fatura
- Recibo
- Extrato bancário
- Currículo



DOCUMENT UNDERSTANDING

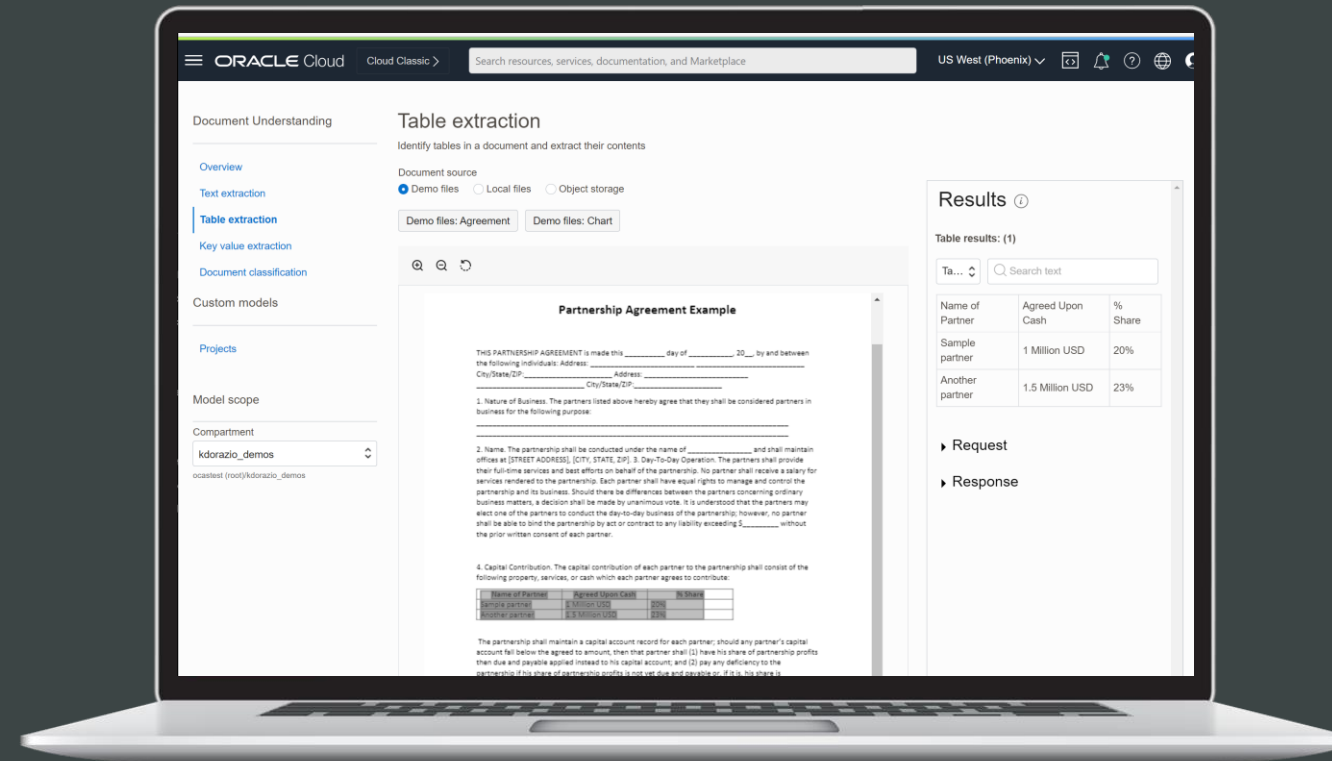
Extração de Tabelas

Extraia texto em formato tabular com informações de linha/coluna por célula.

- Funciona em mesas com e sem bordas
- Funciona para tabelas com e sem cabeçalhos
- Detecta cabeçalhos de tabelas

Aplicações comuns:

- Detalhes em nível de item em recibos
- Faturas
- Ordens de compra
- Documentos de relatório de mercado



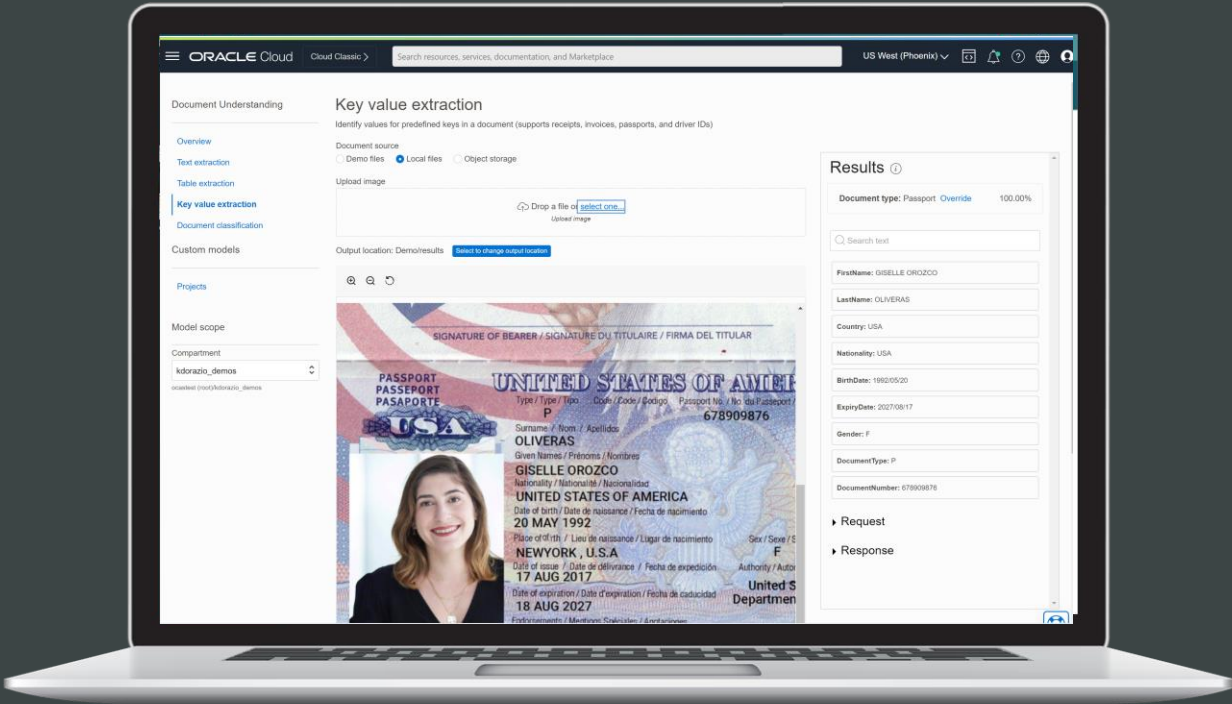
DOCUMENT UNDERSTANDING

Extração de Campos Chave

Identifique valores para campos comuns em passaportes (sem necessidade de treinamento).

Inclui campos como:

Field	Description
FirstName	First name (given name) listed on the document.
LastName	Last name (family name) listed on the document.
Country	Country listed on the document.
Nationality	Nationality of the document owner.
BirthDate	Date of birth.
ExpiryDate	Date of expiration listed on the document.
Gender	Gender listed on the document.



DOCUMENT UNDERSTANDING

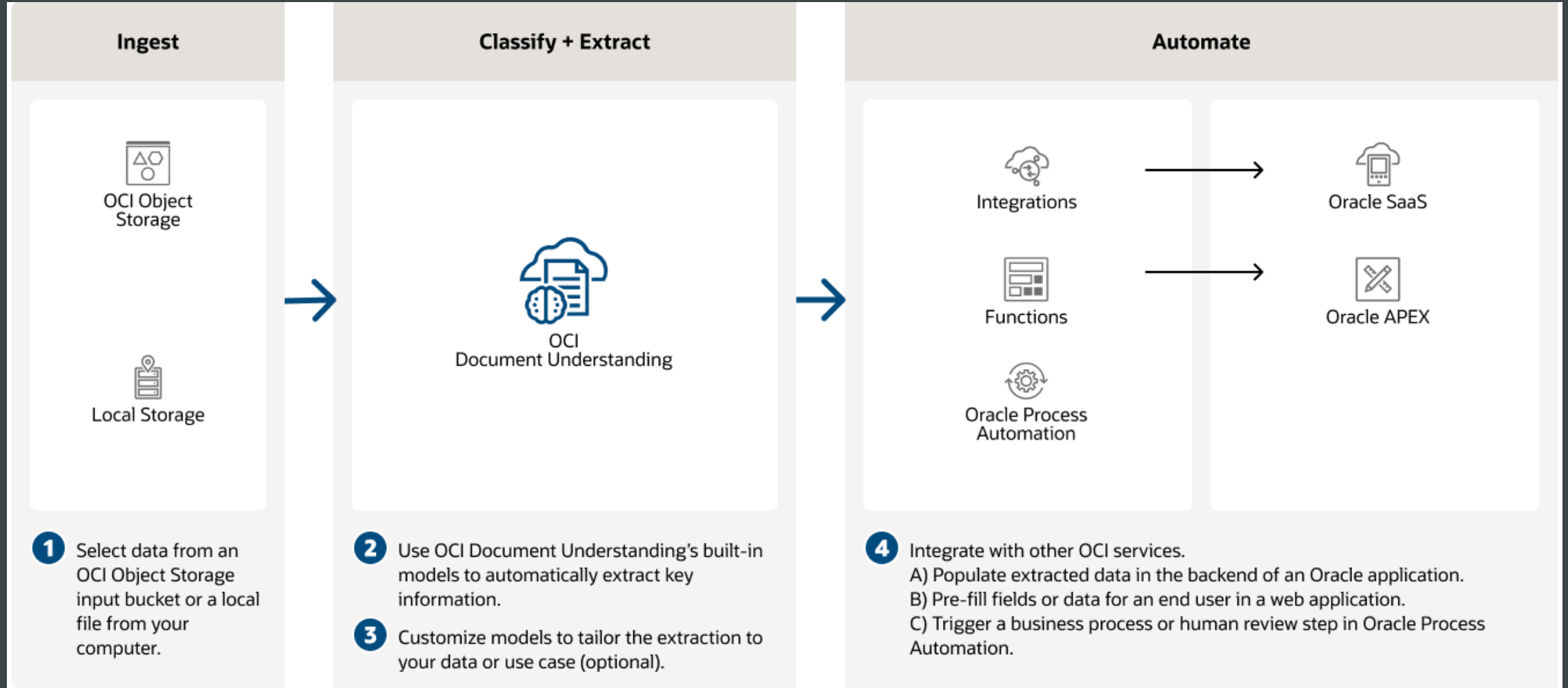
Casos de Uso

- **Pesquisa inteligente:** enriqueça os arquivos de documentos com metadados, incluindo tipo de documento e campos-chave para facilitar a recuperação
- **Automação de faturas:** extraia dados de faturas, como valores, datas e condições de pagamento para processar automaticamente as faturas para pagamento. Isso pode economizar tempo e esforço e melhorar a precisão e a confiabilidade do processo de faturamento.
- **Análise de documentos:** extraia dados de documentos como faturas ou sinistros de seguros para identificar tendências ou oportunidades de economia de custos.
- **Detecção de fraude:** extrair dados de documentos e detectar padrões ou anomalias que possam indicar fraude. Gestão de recursos humanos: extraia dados de registros de funcionários, como nomes, endereços e cargos para preencher um sistema de gestão de recursos humanos. Isso facilita o armazenamento e o gerenciamento das informações dos funcionários e pode reduzir o risco de erros.
- **Gerenciamento da cadeia de suprimentos:** extraia informações de documentos de remessa, como conhecimentos de embarque, e acompanhe o status das remessas. Isto pode ajudar as empresas de logística a gerir a sua cadeia de abastecimento e a garantir que os produtos são entregues a tempo.



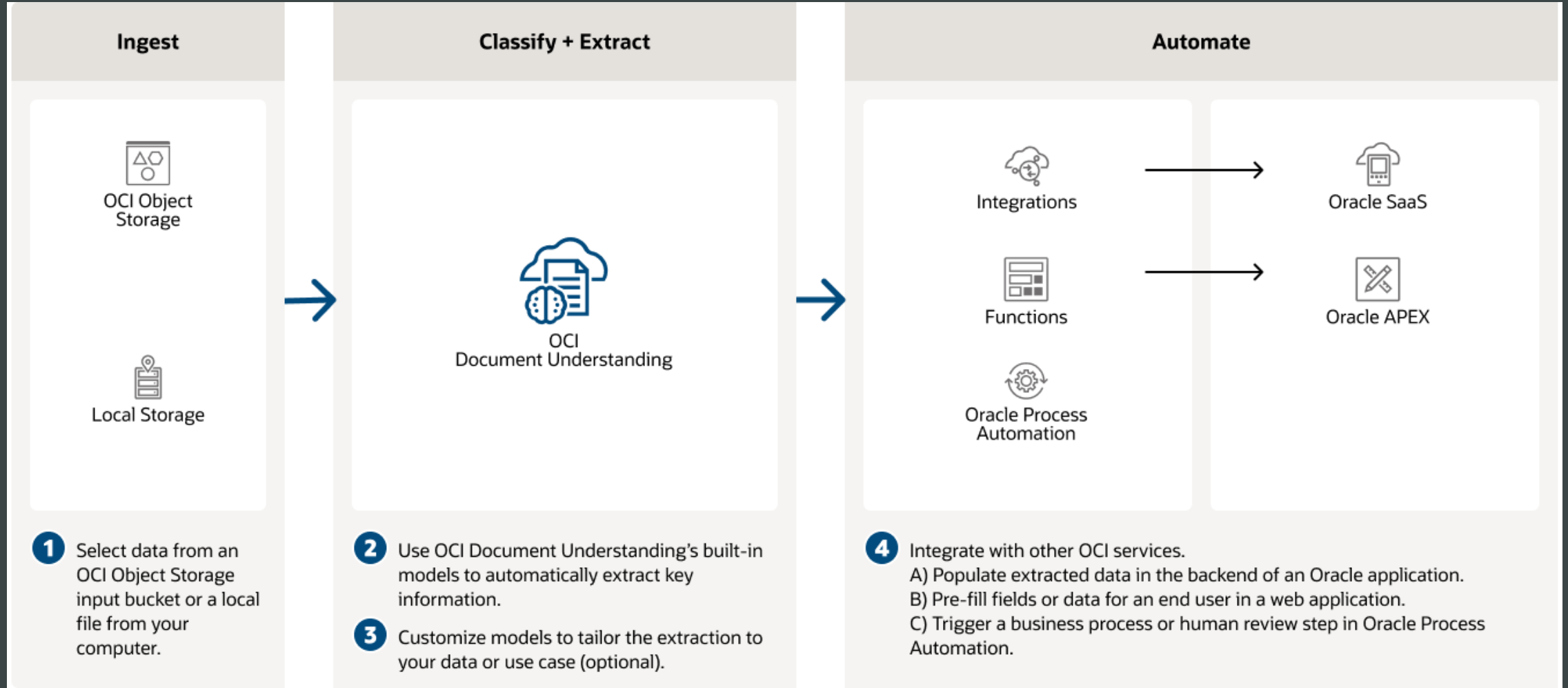
DOCUMENT UNDERSTANDING

Arquitetura – Analise dos Dados presentes nos Documentos



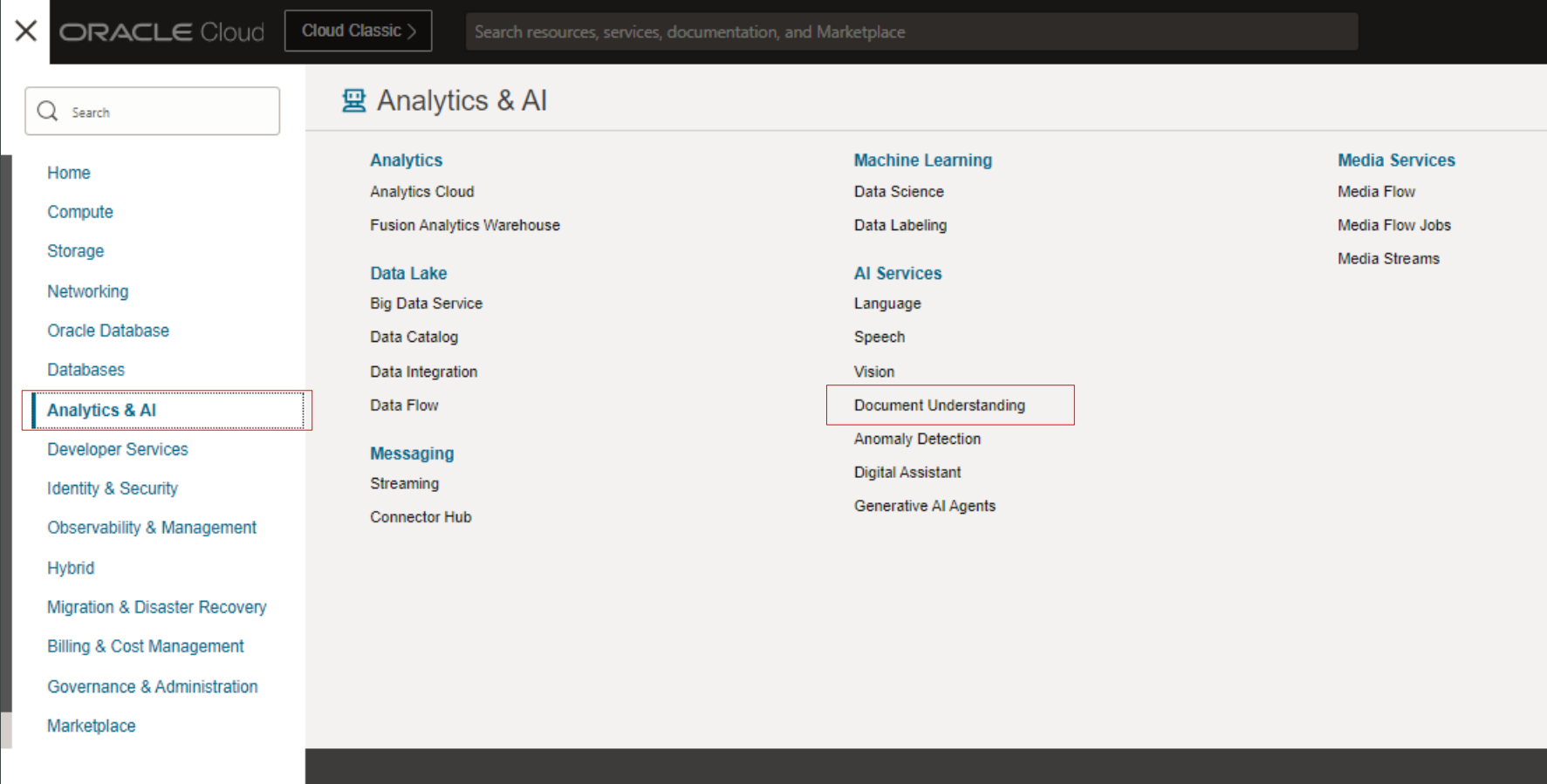
DOCUMENT UNDERSTANDING

Arquitetura - Automação



DOCUMENT UNDERSTANDING

Instanciando o Serviço









DOCUMENT UNDERSTANDING

Instanciando o Serviço

ORACLE Cloud

Cloud Classic >

Search resources, services, documentation, and Marketplace

US West (Phoenix) 

Document Understanding

Overview

Text extraction

Table extraction

Key value extraction

Document classification

Document elements


Custom models

Projects


Document Understanding

Extract insights from document data


Use Oracle Cloud Infrastructure Document Understanding for document analysis tasks like extracting text and recognizing tables. Document Understanding service is a fully managed, multi-tenant, native cloud service that helps you with all your common Document Understanding tasks.

 Watch service tour

Get started

 Video tours

Overview

 Essentials

Document Understanding overview

Policies for Document Understanding

Resources

[Self-paced workshop](#)

[Documentation](#)

[REST API reference](#)

[Sample code](#)

[Oracle Cloud Customer Connect](#)

Feature

Text extraction

Provides word-level and line level text as well as the bounding box coordinates of where the text is located. Optionally, you can create a searchable PDF which embeds a transparent layer on top of a document image in PDF format to make it searchable by keywords.

Table extraction

Identifies tables and individual cells in order to extract content in tabular format.

Key value extraction

Identifies a predefined set of key fields from documents such as receipts, invoices, driver licenses, health insurance ID and passports. Note: some supported documents may be in limited availability (LA) only.

Document classification


Classifies documents into different types based on their visual appearance and high-level features, including invoice, receipt, bank statement, health insurance ID, driver license, passport, tax form, and resume.

Custom key value extraction

Key value extraction model trained on your own labeled dataset for documents like domain-specific forms and intake documents.

Custom document classification

Document classification model trained on your own labeled dataset for industry-specific document types or granular types of one document.





MODELOS PERSONALIZADOS

—
Customize um modelo para ser aderente ao seu
ramo de atuação

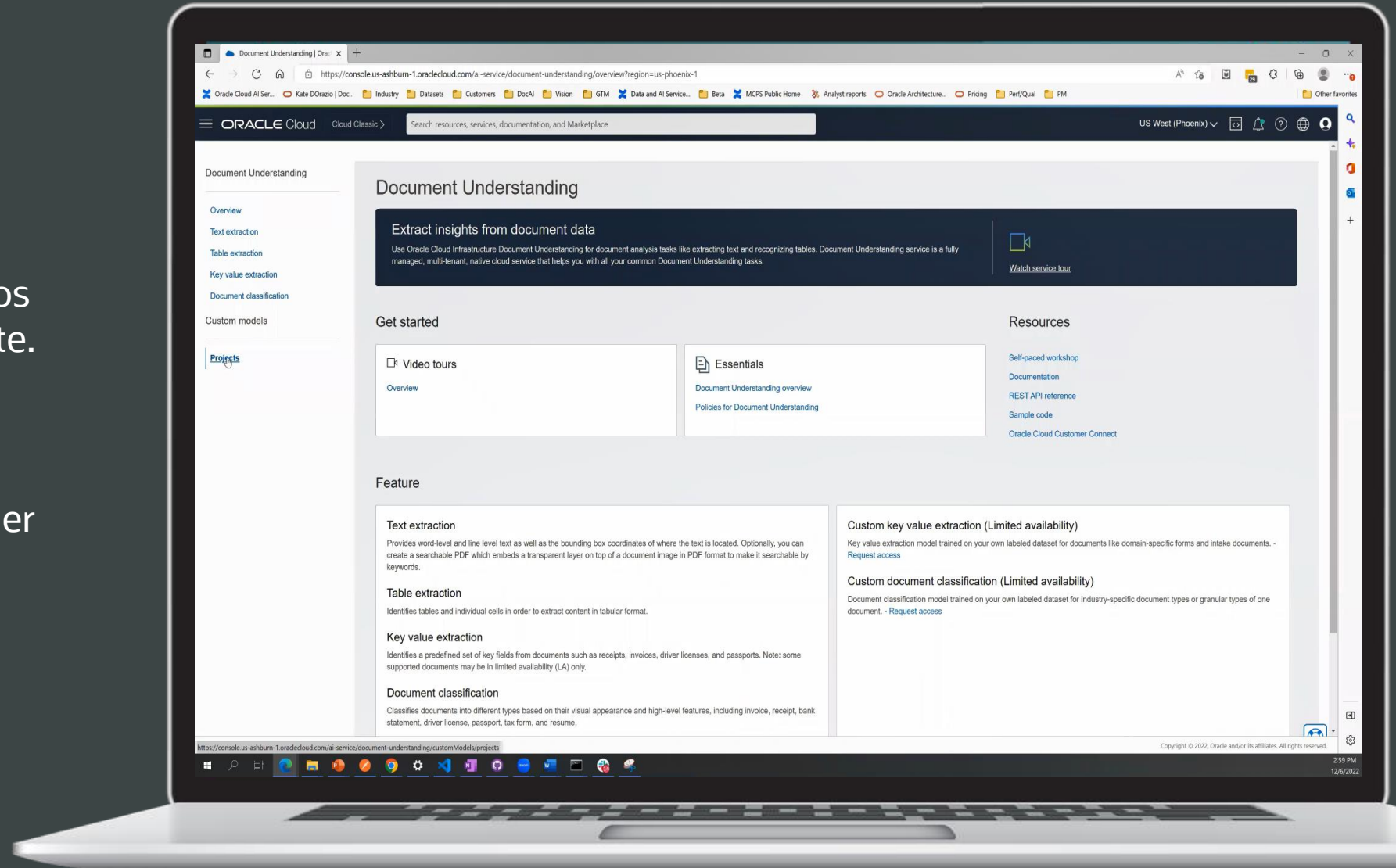


MODELOS PERSONALIZADOS

Ajuste nossos modelos pré-treinados com base em seus dados

Adapte abordagens de aprendizagem por transferência para ajustar nossos modelos de aprendizagem profunda aos dados específicos do cliente.

Não há necessidade de gerenciar recursos computacionais ou qualquer outro aspecto da infraestrutura do modelo personalizado



MODELOS PERSONALIZADOS

Passo 01 - Escolha entre classificação de documentos e extração de valor-chave

ORACLE Cloud

Cloud Classic >

Search resources, services, documentation, and Marketplace

US West (Phoenix) >

Create and train model

1 Select data

2 Train model

3 Review

Train a custom model using a training dataset. A unique ID is created so you can identify the model afterward.

Choose model type to Train

Type

Key value extraction

Training data

Choose existing dataset

Use an existing dataset to train your model

✓

Create a New Dataset

Create a new dataset in Data Labeling Service to train your model (redirects to new page)

Data Source

☒ Data Labeling Service

☐ Object Storage

The chosen dataset must contain a minimum of 10 images. Create and choose a dataset with appropriate format of schema. [Please read this for accepted schema formats \(redirects to new page\)](#)

Choose a dataset in **kdorazio_demos** [\(Change Compartment\)](#)

Demo Overview

Show advanced options

Next

Cancel



MODELOS PERSONALIZADOS

Passo 01 – Selecione o Tipo de Modelo a ser treinado

ORACLE Cloud Cloud Classic > Search resources, services, documentation, and Marketplace US West (Phoenix) [Icons]

Create and train model

1 Select data 2 Train model 3 Review

Train a custom model using a training dataset. A unique ID is created so you can identify the model afterward.

Choose model type to Train

Type

Key value extraction

Training data

Choose existing dataset

Use an existing dataset to train your model ✓

Create a New Dataset

Create a new dataset in Data Labeling Service to train your model (redirects to new page)

Data Source

☒ Data Labeling Service ☐ Object Storage

i The chosen dataset must contain a minimum of 10 images. Create and choose a dataset with appropriate format of schema. [Please read this for accepted schema formats \(redirects to new page\)](#)

Choose a dataset in **kdorazio_demos** (Change Compartment)

Demo Overview

Show advanced options

Next Cancel

Escolha entre
classificação de
documentos e extração
de valor-chave

MODELOS PERSONALIZADOS

Passo 02 – Selecione dados de treinamento já rotulados existentes

ORACLE Cloud Cloud Classic > Search resources, services, documentation, and Marketplace US West (Phoenix) [Icons]

Create and train model

1 Select data 2 Train model 3 Review

Train a custom model using a training dataset. A unique ID is created so you can identify the model afterward.

Choose model type to Train

Type

Key value extraction

Training data

Choose existing dataset

Use an existing dataset to train your model

Create a New Dataset

Create a new dataset in Data Labeling Service to train your model (redirects to new page)

Data Source

☒ Data Labeling Service ☐ Object Storage

i The chosen dataset must contain a minimum of 10 images. Create and choose a dataset with appropriate format of schema. Please read this for accepted schema formats (redirects to new page)

Choose a dataset in **kdorazio_demos** (Change Compartment)

Demo Overview

Demo Overview

LibertyTest

KVTest

UHC Test

Show advanced options

Next Cancel

Se você tiver um conjunto de dados de treinamento existente, selecione no menu suspenso

MODELOS PERSONALIZADOS

Passo 02 – Rotule seu próprio conjunto de dados com o Data Labeling Service

Training Data

Choose Existing Dataset

Use an existing dataset to train your model

Create a New Dataset

Create a new dataset in Data Labeling Service to train your model

ORACLE Cloud

Cloud Classic >

Search resources, services, documentation, and Marketplace

US East (Ashburn)

Data Labeling > Dataset list > DemoOverview1 > Add labels

Add labels: CF2.PNG

Page < 1 of 1 >

Shortcuts

Skip

Save

Save & next

Cancel

Summary ⓘ

Select Labels...

Recipient: Jane Hill

STRAIGHT BILL OF LADING - SHORT FORM

Carrier Name: XeDei

Phone: 121-555-8569

SHIPMENT IDENTIFICATION NO: 54DFC445D6FC5

Carrier Address: 500 4th St

Date: 12/5/22

FREIGHT BILL PRG NO

City: Seattle

State and Zip: Washington

SAC

DUNS

TO: Jane Hill

TRAILER CARRIER NUMBER: 4

Consignee Address: 320 4th St

ROUTE

City: Seattle

State and Zip: 11812

FROM:

SPECIAL INSTRUCTIONS

Shipper Address: 320 4th St

SHIPPER'S INTERNAL DATA

City: Seattle

State and zip: 11812

FOR PAYMENT SEND BILL TO:

SE NO

Name: N/A

Address: 320 4th St

City: Seattle

State & Zip

Labels (4)

Search label

Recipient

CarrierName

ShippingId

TrailerNo

Info

Text : Hill

Label : Recipient

OCR confidence : 98.91%

Records (0 of 5 labeled)

Status All

Se você tiver documentos sem rótulos, anote-os facilmente com o Data Labeling Service



MODELOS PERSONALIZADOS

Passo 03 – Treine seu Modelo

ORACLE Cloud

Cloud Classic >

Search resources, services, documentation, and Marketplace

US West (Phoenix)

Create and train model

1 Select data

2 Train model

3 Review

Train a custom model using a training dataset. A unique ID is created so you can identify the model afterward.

Model display name

OverviewKVModel

Model description Optional

Training duration

☒ Recommended training (May take up to 24 hours)

☐ Custom duration

Show advanced options

Previous

Next

Cancel

Nomei o seu modelo e especifique a duração do tempo de treinamento



MODELOS PERSONALIZADOS

Passo 04 – Revisão

ORACLE Cloud

Cloud Classic >

Search resources, services, documentation, and Marketplace

US West (Phoenix) v

Create and train model

1 Select data

2 Train model

3 Review

Train a custom model using a training dataset. A unique ID is created so you can identify the model afterward.

Training data asset information

Compartment: klorazio_demos

Dataset: Demo Overview

Edit

Model information

Model: OverviewKVModel

Description:

Model type: KEY_VALUE_DETECTION

Training duration (in hours): 24

Edit

Previous

Create and train

Cancel

Revise os detalhes antes
de enviar o trabalho para
treinamento



MODELOS PERSONALIZADOS

Passo 05 – Veja as métricas do Modelo

ORACLE Cloud

Cloud Classic >

Search resources, services, documentation, and Marketplace

US West (Phoenix)

Artificial Intelligence > Document Understanding > Projects > Project details: models > Model details

MOD

ACTIVE

Edit

Move Resource

Delete

Model information

Tags

General information

Description:

Model type: KEY_VALUE_DETECTION

OCID: ...nqpkta Show Copy

Training data asset: ...wvlnjzbdvloxofg3uq Show Copy

Created by: kate.dorazio@oracle.com

Created: Mon, Dec 5, 2022, 18:45:30 UTC

Max Training duration: Recommended mode (24 hours)

Training metrics

Precision: 1.0000

Recall: 1.0000

Accuracy: 1.0000

Threshold: 0.5000

Total documents: 5

Test documents: 3

Trained duration in hours: 0.2842

Resources

Analyze

Work request logs

Training metrics

List scope

Compartment

kdorazio_demos

ocastest (root)/kdorazio_demos

Name

State

Model metrics in kdorazio_demos Compartment

Key-value	Precision	Recall	Accuracy	Test documents	Threshold
CarrierName	1.0000	1.0000	1.0000	3	0.5000
ShipmentID	1.0000	1.0000	1.0000	3	0.5000
Recipient	1.0000	1.0000	1.0000	3	0.5000
TrailerNo	1.0000	1.0000	1.0000	3	0.5000

Showing 4 Items < 1 of 1 >

Revise as pontuações de precisão, recall e exatidão por campo de valor-chave ou classe de documento



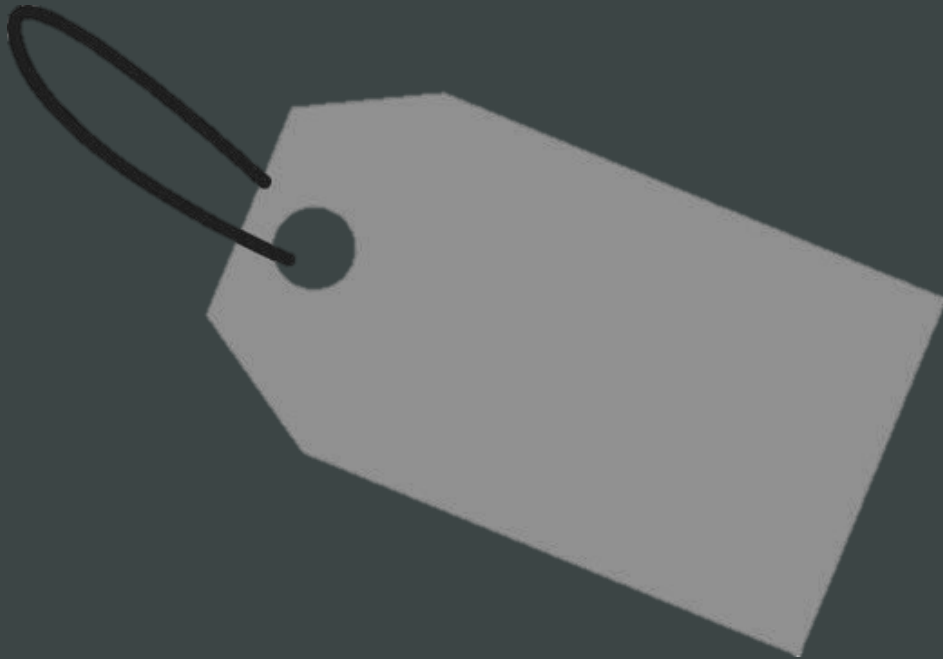
DATA LABELING

Rotule seus dados para que possam ser utilizados para treinar modelos de IA e Machine Learning com mais precisão.



DATA LABELING

O que é?



Data Labeling, ou Rotulagem de Dados, é o processo de adicionar TAGs ou Labels (Rótulos) aos dados brutos, como Imagem, Texto ou Documentos.

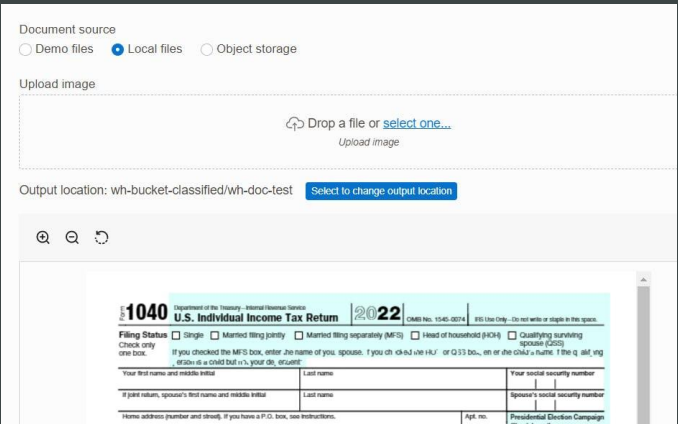
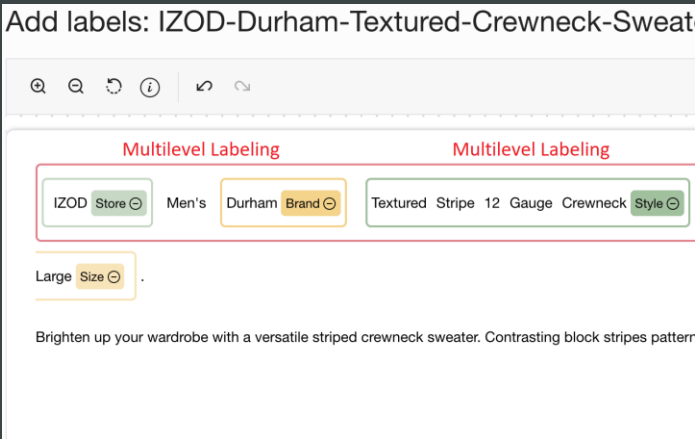
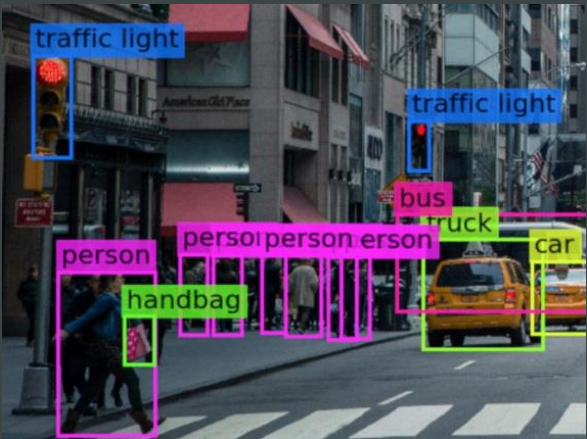
DATA LABELING

Capacidades

Imagem

Texto

Documentos



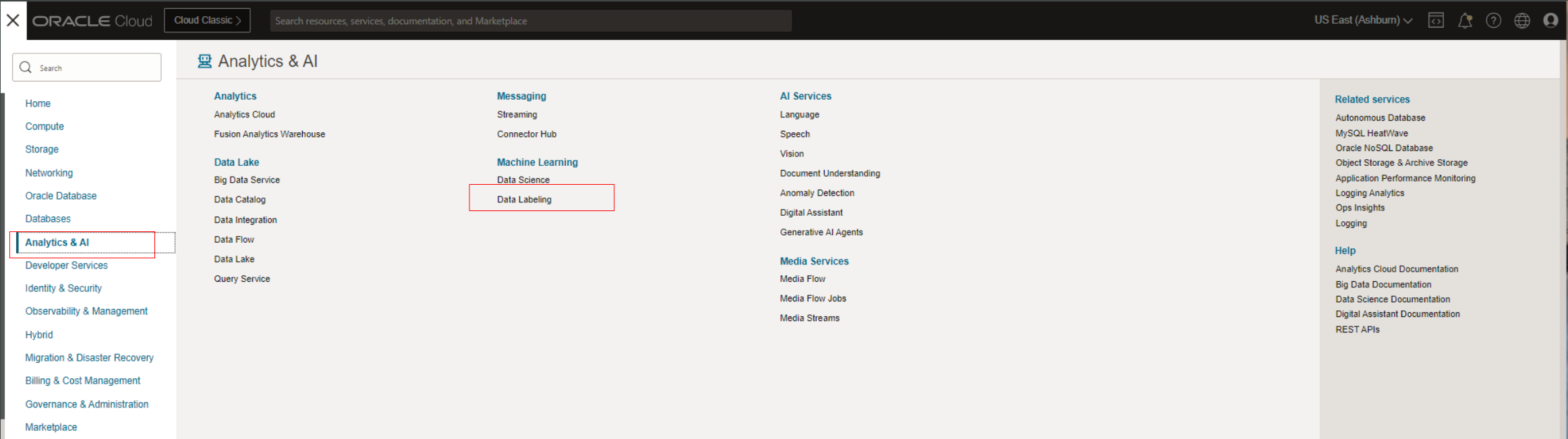
DATA LABELING

Casos de Uso

- **Classificação de imagens:** Rotule conjuntos de dados que consistem em logotipos, silhuetas e cores de roupas populares, tipos de produtos e imagens médicas. Use esses conjuntos de dados rotulados para planejamento de estoque, categorização de produtos, gerenciamento de prateleiras e diagnósticos médicos.
- **Deteção de irregularidades:** Rotule imagens irregulares para criar modelos que automatizam elementos de verificação de qualidade do produto, detecção de defeitos, vigilância de segurança e gerenciamento de estoque.
- **Processamento de formulários:** Rotule documentos para facilitar a extração de informações valiosas para o processamento de solicitações de pacientes, diagnóstico de relatórios médicos e pesquisa celular.
- **Assistência virtual:** Classifique conjuntos de dados, incluindo recibos, faturas e pedidos para chatbots de suporte ao cliente e arquivamento automatizado de despesas.
- **Extração de informações:** Marque agrupamentos de palavras e atribua rótulos. Conjuntos de dados de texto rotulados podem ser usados para análise de pesquisa de clientes, modelagem de tópicos e suporte ao cliente.

DATA LABELING

Instanciando o Serviço



DATA LABELING

Instanciando o Serviço

ORACLE Cloud

Cloud Classic >

Search resources, services, documentation, and Marketplace

US East (Ashburn) > <> 🔔 ? 🌐 👤

Data Labeling

Overview

Datasets

Work requests

Data Labeling

Getting started

The OCI Data Labeling service is an OCI native service that allows customers to create and browse datasets, view data records (text, images) and apply labels for the purposes of building AI/ML models. The service also provides interactive user interfaces designed to aide in the labeling process. Once records are labeled, the dataset can be exported as line-delimited JSON for use in AI/ML model development.

Watch service video

Features

Image labeling

Categorize images into single or multiple label classes or draw bounding boxes around specific objects.

Text labeling

Categorize text files into single or multiple label classes or highlight and label text.

Document labeling

Categorize documents into single or multiple label classes.

Resources

All documentation

Tutorials

Videos

What's new

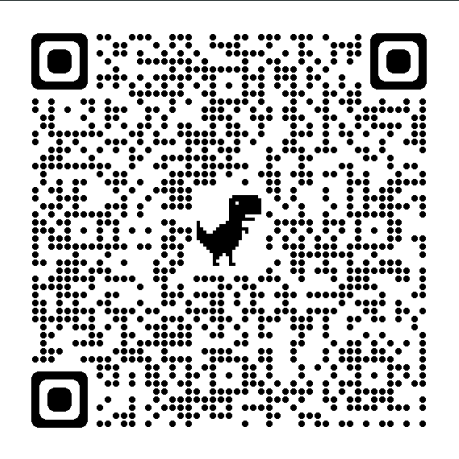
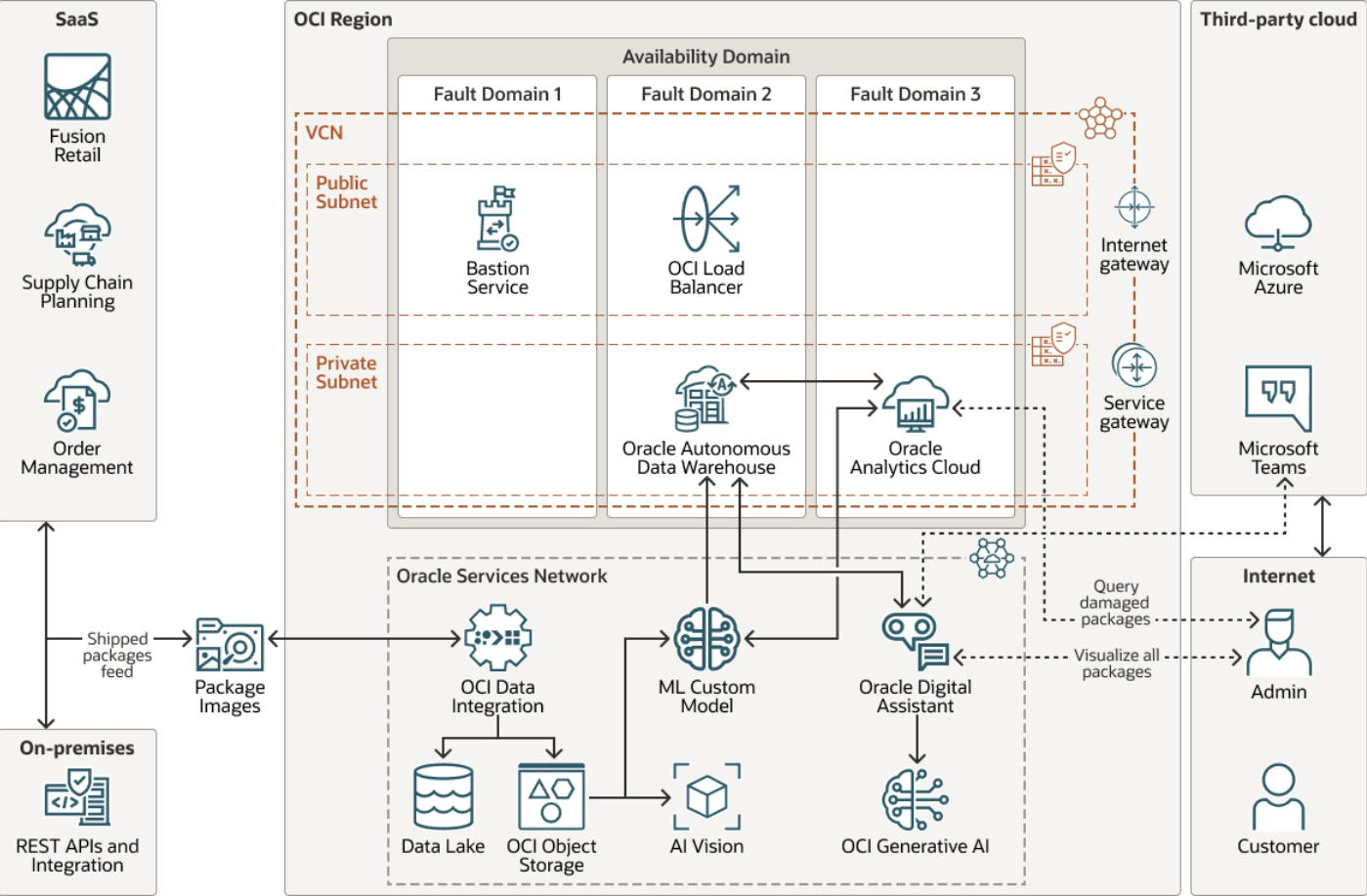
Release notes

33

Copyright © 2024, Oracle and/or its affiliates | Confidential: Internal/Restricted/Highly Restricted

ARQUITETURA DE REFERÊNCIA

Automatize e simplifique seu fluxo de trabalho de classificação de imagens com os serviços OCI AI



CONTÉUDO DE APOIO

Documentação

<https://docs.oracle.com/en-us/iaas/Content/document-understanding/using/overview.htm>

<https://docs.oracle.com/en-us/iaas/Content/data-labeling/using/home.htm>

Polices

https://docs.oracle.com/en-us/iaas/Content/document-understanding/using/about_document-understanding_policies.htm#about_vision_policies



Video Tours

<https://www.youtube.com/playlist?list=PLKCK3OyNwlzt1x62El9gGGeNaQr0va58c>

<https://www.youtube.com/playlist?list=PLKCK3OyNwlzsgkJ4CKrkyCx334Veo6SSP>



Experimente o poder da IA gratuitamente na nuvem

Conheça os recursos de IA e ML da Oracle Cloud com o **Modo Gratuito**.

São **US\$ 500** em créditos para você usar durante 30 dias – uma condição **EXCLUSIVA** para visitantes do TDC Summit São Paulo!



ORACLE
DevLive
São Paulo

Junte-se a nós em São Paulo para criar novas possibilidades com dados e IA

03 de abril de 2024
World Trade Center São Paulo

Inscreva-se já!
oracle.com/devlive-sao-paulo/



ORACLE



Thank you!

