

Intelligent Document Extraction with OpenAI GPT & LangChain

OpenAI API | GPT | Vision Models | MongoDB | LangChain | FastAPI

Get Started



Invoice

Tax Invoice/Bill of Supply/Cash Memo
(Original for Recipient)

Sold By :
CLICKTECH RETAIL PRIVATE LIMITED
Survey Number 99/1, Mamidipally Village,
Shamshabad
HYDERABAD, TELANGANA, 500108
IN

PAN No:AAJCC9783E
GST Registration No:36AAJCC9783E1Z8
Dynamic QR Code:

Billing Address :
Ramya Devi
K-809, ABC Apartments, Madhapur,
Mehdipatnam, Main Road
HYDERABAD, TELANGANA, 500089
IN
State/UT Code:36

Shipping Address :
Ramya Devi
K-809, ABC Apartments, Madhapur,
Mehdipatnam, Main Road
HYDERABAD, TELANGANA, 500089
IN
State/UT Code:36
Place of supply:TELANGANA
Place of delivery:TELANGANA
Invoice Number :HYD3-1279928
Invoice Details :TG-HYD3-297683823-2425
Invoice Date :30.11.2024

Order Number:407-7033233-4757902
Order Date:30.11.2024

Sl. No	Description	Unit Price	Qty	Net Amount	Tax Rate	Tax Type	Tax Amount	Total Amount
1	Corsair Katar Pro Ultra-Light Optical USB Gaming Mouse with Backlit RGB LED, 12400 DPI (Black) B08KSV3SR (B08KSV3SR) HSN:84716060	₹1,143.22	1	₹1,143.22	0%	CGST	₹102.89	₹1,349.00
TOTAL:						0%	SGST	₹102.89
							₹205.78	₹1,349.00

Amount in Words:
One Thousand Three Hundred Forty-nine only

For CLICKTECH RETAIL PRIVATE LIMITED:

Authorized Signatory

Whether tax is payable under reverse charge - No

*ASPL-Amazon Seller Services Pvt. Ltd., ARPL-Amazon Retail India Pvt. Ltd. (only where Amazon Retail India Pvt. Ltd. fulfillment center is co-located)

Customers desirous of availing input GST credit are requested to create a Business account and purchase on Amazon.in/business from Business eligible offers

Please note that this invoice is not a demand for payment

Page 1 of 1

Receipt

The helpful place.

Thank You for Shopping at
Ace Hardware of Navarre
8188 Navarre Pkwy, Navarre, FL 32566
Phone: (850) 939-9800

11/14/2024, 5:46:31 PM

Reg: 2 ACE Rewards # : 1928777
Customer: J Manchester: Since 2011

2489540 DeWalt 20V MAX
Cordless Brushed 2 Tool Kit

2029735 DeWalt Maxfit
Screwdriver Bit Set 30 pc

CLDBVG COKE 20oz

\$219.99

\$19.98

\$1.99

Subtotal

Tax

Total

\$241.97

\$16.93

\$258.90

Card number **** * 3081
Card type Debit
Card entry Chip
Date/time 11/14/2024 5:46 PM
Reference # 62845289260246240685C
Status APPROVED

THANK YOU
HAVE A NICE DAY

Business Card

DAVID PACKARD ELECTRICAL ENGINEERING
350 SERRA MALL, ROOM 170
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-9505

RAFAEL ULATE
DIRECTOR OF ADMISSIONS
DEPARTMENT OF ELECTRICAL ENGINEERING

TEL: (650) 725-9327
FAX: (650) 723-1882
ulate@ee.stanford.edu

Government ID Cards

भारत सरकार
GOVERNMENT OF INDIA

NAME : XXXX
DOB : XXXX
GENDER : XXXX

4444 3333 6666 8888

आधार - भारतीय विशिष्ट पहचान प्राधिकरण

Bank Documents

भारतीय स्टेट बैंक
STATE BANK OF INDIA

पासबुक
PASS BOOK

किओस्क बैंक
KIOSK BANK

Companies handles

- Thousands of invoices
- Thousands of Receipts
- Many PDF files
- Many Scanned documents

Can Computer understand documents like humans ?

No

Human Brain Naturally Understand documents

Seller Name


Seller Address

Items / Description

amazon.in

Tax Invoice/Bill of Supply/Cash Memo
(Original for Recipient)

Sold By :
CLICKTECH RETAIL PRIVATE LIMITED
*Survey Number 99/1, Mamidipally Village,
Shamshabad
HYDERABAD, TELANGANA, 500108
IN

PAN No:AAJCC9783E
GST Registration No:36AAJCC9783E1Z8
Dynamic QR Code:


Billing Address :
Ramya Devi
K-809, ABC Apartments, Madhapur,
Mehdipatnam, Main Road
HYDERABAD, TELANGANA, 500089
IN
State/UT Code:36


Shipping Address :
Ramya Devi
K-809, ABC Apartments, Madhapur,
Mehdipatnam, Main Road
HYDERABAD, TELANGANA, 500089
IN
State/UT Code:36
Place of supply:TELANGANA
Place of delivery:TELANGANA

Order Number:407-7033233-4757902
Order Date:30.11.2024

Invoice Number :HYD3-1279928
Invoice Details :TG-HYD3-297683823-2425
Invoice Date :30.11.2024

Sl. No	Description	Unit Price	Qty	Net Amount	Tax Rate	Tax Type	Tax Amount	Total Amount
1	Corsair Katar Pro Ultra-Light Optical USB Gaming Mouse with Backlit RGB LED, 12400 DPI (Black) B08KSZV3SR (B08KSZV3SR) HSN:84716060	₹1,143.22	1	₹1,143.22	0%	CGST	₹102.89	₹1,349.00
					0%	SGST	₹102.89	
TOTAL:								₹205.78 ₹1,349.00

Amount in Words:
One Thousand Three Hundred Forty-nine only

For CLICKTECH RETAIL PRIVATE LIMITED:

Authorized Signatory

Whether tax is payable under reverse charge - No

Buyer Name

Buyer Name

Invoice Number

Invoice Date

Total Amount

Currency

*ASSPL-Amazon Seller Services Pvt. Ltd., ARIP-Amazon Retail India Pvt. Ltd. (only where Amazon Retail India Pvt. Ltd. fulfillment center is co-located)

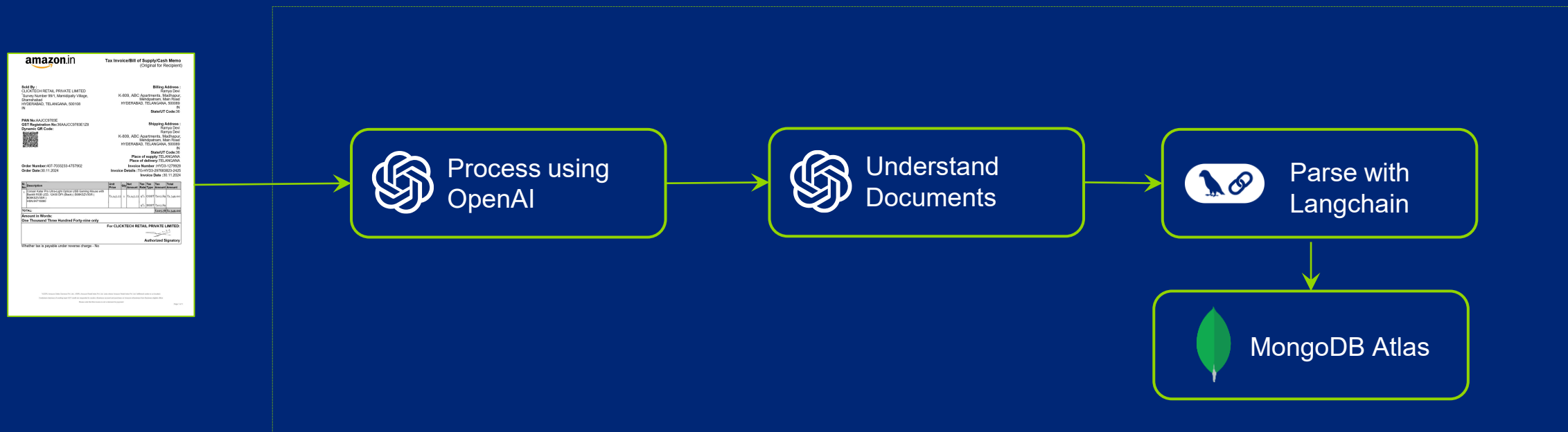
Customers desirous of availing input GST credit are requested to create a Business account and purchase on Amazon.in/business from Business eligible offers

Please note that this invoice is not a demand for payment

Page 1 of 1

VEDICSKILL

What you will learn



Next



Course Plan



Intelligent Document Extraction with OpenAI GPT & LangChain

OpenAI API | GPT | Vision Models | MongoDB | LangChain | FastAPI

Course Plan

VEDICSKILL.

amazon.in Tax Invoice/Bill of Supply/Cash Memo (Original for Recipient)

Sold By : CLICKTECH RETAIL PRIVATE LIMITED
Survey Number 99/1, Mamidipally Village, Shamshabad
HYDERABAD, TELANGANA, 500108 IN

Billing Address : Ramya Devi
K-809, ABC Apartments, Madhapur, Mehdiapattam, Main Road
HYDERABAD, TELANGANA, 500089 IN
State/UT Code:36

PAN No:AAJCC9783E
GST Registration No:36AAJCC9783E1Z8
Dynamic QR Code:

Shipping Address : Ramya Devi
K-809, ABC Apartments, Madhapur, Mehdiapattam, Main Road
HYDERABAD, TELANGANA, 500089 IN
State/UT Code:36

Place of supply:TELANGANA
Place of delivery:TELANGANA

Order Number:407-7033233-4757902
Order Date:30.11.2024

Invoice Number :TG-HYD3-297683823-2425
Invoice Date :30.11.2024

Sl. No	Description	Unit Price	Qty	Net Amount	Tax Rate	Tax Type	Tax Amount	Total Amount
1	Corsair Katar Pro Ultra-Light Optical USB Gaming Mouse with Backlit RGB LED, 12400 DPI (Black) B0BK5ZV3SR (B0BK5ZV3SR) HSN:84716060	₹1,143.22	1	₹1,143.22	9%	CGST	₹102.89	₹1,246.11
					9%	SGST	₹102.89	
TOTAL:								₹1,246.11

Amount in Words:
One Thousand Three Hundred Forty-nine only

For CLICKTECH RETAIL PRIVATE LIMITED:

Authorized Signatory

Whether tax is payable under reverse charge - No

Page 1 of 1

ACE
The helpful place.

Thank You for Shopping at
Ace Hardware of Navarre
8188 Navarre Pkwy., Navarre, FL 32566
Phone: (850) 939-9800

11/14/2024, 5:48:31 PM

Reg: 2 ACE Rewards # : 1928777
Customer: J Manchester: Since 2011

2489540 DeWalt 20V MAX \$219.99
Cordless Brushed 2 Tool Kit
2029735 DeWalt Maxfit \$19.98
Screwdriver Bit Set 30 pc
CLDBVG COKE 20oz \$1.99

Subtotal \$241.97
Tax \$16.93

Total \$258.90

Card number **** * 3081
Card type Debit
Card entry Chip
Date/time 11/14/2024 5:48 PM
Reference # 62845289260246240685C
Status APPROVED

THANK YOU
HAVE A NICE DAY

RAFAEL ULATE
DIRECTOR OF ADMISSIONS
DEPARTMENT OF ELECTRICAL ENGINEERING

DAVID PACKARD ELECTRICAL ENGINEERING
350 SERRA MALL, ROOM 170
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-9505

TEL: (650) 725-9327
FAX: (650) 723-1882
ulate@ee.stanford.edu

भारत सरकार
GOVERNMENT OF INDIA

NAME : XXXX
DOB : XXXX
GENDER : XXXX

4444 3333 6666 8888

आधार - भारतीय विशिष्ट पहचान प्राधिकरण

SBI भारतीय स्टेट बैंक
STATE BANK OF INDIA

PASS BOOK

YONO SBI

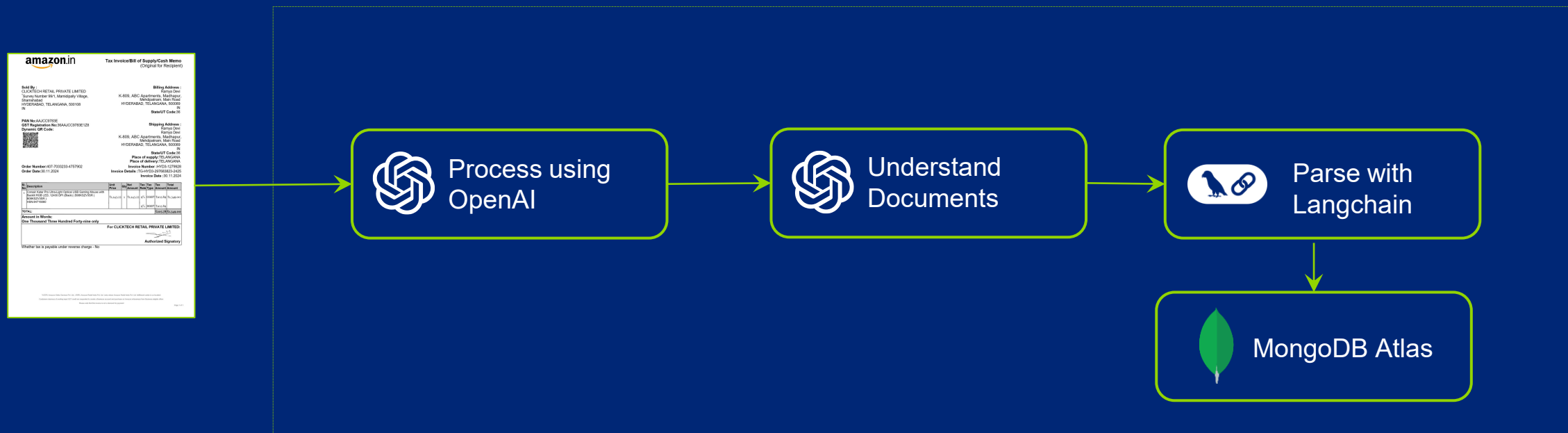
किरकोक बैंक
KIOSK BANK

Documents

Most common documents in our life

Business & Commercial	Personal Identify Documents (India)	Customer Details
Invoice	Aadhaar Card	Electrical/Water/Telephone Bill
Shipping Invoice	PAN card	Gas Bill
Bill of Lading	Voter ID	Property Tax
Purchase Order	Driving License	Rental Agreement
Proforma Invoice	Ration Card	Agreement of Sale
Receipt of payment	Bank Passbook	Pay slips
Debit/Credit Note	Employee Photo ID card	
Worker Order	Government Issue Card	
Contract or Agreement	Health Insurance Card	
Business card		




Common Framework



Documents

In this course we will focus on two documents

Invoice

		Tax Invoice/Bill of Supply/Cash Memo (Original for Recipient)																														
Sold By : CLICKTECH RETAIL PRIVATE LIMITED Survey Number 99/1, Mamidipally Village, Shamshabad HYDERABAD, TELANGANA, 500108 IN		Billing Address : Ramya Devi K-809, ABC Apartments, Madhapur, Mehdipatnam, Main Road HYDERABAD, TELANGANA, 500089 IN State/UT Code: 36																														
PAN No: AJJC93783E GST Registration No: 36AAJCC9783E128 Dynamic QR Code: 		Shipping Address : Ramya Devi Ramya Devi K-809, ABC Apartments, Madhapur, Mehdipatnam, Main Road HYDERABAD, TELANGANA, 500089 IN State/UT Code: 36 Place of supply: TELANGANA Place of delivery: TELANGANA																														
Order Number: 407-7033233-4757902 Order Date: 30.11.2024		Invoice Number : HYD3-1279928 Invoice Date : 30.11.2024 Invoice Details : TG-HYD3-297683823-2425																														
<table><thead><tr><th>Sl. No</th><th>Description</th><th>Unit Price</th><th>Qty</th><th>Net Amount</th><th>Tax Rate</th><th>Tax Type</th><th>Tax Amount</th><th>Total Amount</th></tr></thead><tbody><tr><td>1</td><td>Conair Katar Pro Ultra-Light Optical USB Gaming Mouse with Backlit RGB LED, 12400 DPI (Black) B08KSZVSJR (B08KSZVSJR) HSN:84716000</td><td>₹1,143.22</td><td>1</td><td>₹1,143.22</td><td>0%</td><td>GST</td><td>₹120.80</td><td>₹1,149.00</td></tr><tr><td colspan="2">TOTAL:</td><td></td><td></td><td></td><td>0%</td><td>GST</td><td>₹120.80</td><td>₹1,149.00</td></tr></tbody></table>		Sl. No	Description	Unit Price	Qty	Net Amount	Tax Rate	Tax Type	Tax Amount	Total Amount	1	Conair Katar Pro Ultra-Light Optical USB Gaming Mouse with Backlit RGB LED, 12400 DPI (Black) B08KSZVSJR (B08KSZVSJR) HSN:84716000	₹1,143.22	1	₹1,143.22	0%	GST	₹120.80	₹1,149.00	TOTAL:					0%	GST	₹120.80	₹1,149.00	Amount in Words: One Thousand Three Hundred Forty-nine only For CLICKTECH RETAIL PRIVATE LIMITED:  Authorized Signatory			
Sl. No	Description	Unit Price	Qty	Net Amount	Tax Rate	Tax Type	Tax Amount	Total Amount																								
1	Conair Katar Pro Ultra-Light Optical USB Gaming Mouse with Backlit RGB LED, 12400 DPI (Black) B08KSZVSJR (B08KSZVSJR) HSN:84716000	₹1,143.22	1	₹1,143.22	0%	GST	₹120.80	₹1,149.00																								
TOTAL:					0%	GST	₹120.80	₹1,149.00																								
Whether tax is payable under reverse charge - No																																

©2025 Amazon Seller Services Pvt. Ltd., ARSP, Amazon Retail India Pvt. Ltd. (only where Amazon Retail India Pvt. Ltd. fulfillment center is co-located).

Customer's acceptance of printed label GST credit is required by Seller. It becomes invalid and applicable for Amazon reimbursements from Seller's eligible offers.

Please note that this invoice is not to be used for payment.

Page 1 of 1

Business Card





Tax Invoice/Bill of Supply/Cash Memo
(Original for Recipient)

Sold By :

CLICKTECH RETAIL PRIVATE LIMITED
Survey Number 99/1, Mamidipally Village,
Shamshabad
HYDERABAD, TELANGANA, 500108
IN

PAN No:AAJCC9783E

GST Registration No:36AAJCC9783E1Z8

Dynamic QR Code:



Order Number:407-7033233-4757902

Order Date:30.11.2024

Billing Address :

Ramya Devi
K-809, ABC Apartments, Madhapur,
Mehdipatnam, Main Road
HYDERABAD, TELANGANA, 500089
IN
State/UT Code:36

Shipping Address :

Ramya Devi
K-809, ABC Apartments, Madhapur,
Mehdipatnam, Main Road
HYDERABAD, TELANGANA, 500089
IN

State/UT Code:36

Place of supply:TELANGANA

Place of delivery:TELANGANA

Invoice Number :HYD3-127992

Invoice Details :TG-HYD3-297683823-2425

Invoice Date :30.11.2024

Sl. No	Description	Unit Price	Qty	Net Amount	Tax Rate	Tax Type	Tax Amount	Total Amount
1	Corsair Katar Pro Ultra-Light Optical USB Gaming Mouse with Backlit RGB LED, 12400 DPI (Black) B08KSZV3SR (B08KSZV3SR) HSN:84716060	₹1,143.22	1	₹1,143.22	0%	CGST	₹102.89	₹1,349.00
					0%	SGST	₹102.89	

TOTAL: ₹205.78 ₹1,349.00

Amount in Words:

One Thousand Three Hundred Forty-nine only

For CLICKTECH RETAIL PRIVATE LIMITED:

Authorized Signatory

Whether tax is payable under reverse charge - No

Seller Name

Seller Address

Items / Description

Buyer Name

Buyer Name

Invoice Number

Invoice Date

Total Amount

Currency



Tax Invoice/Bill of Supply/Cash Memo
(Original for Recipient)

Sold By :

CLICKTECH RETAIL PRIVATE LIMITED
*Survey Number 99/1, Mamidipally Village,
Shamshabad
HYDERABAD, TELANGANA, 500108
IN

Billing Address :

Ramya Devi
K-809, ABC Apartments, Madhapur,
Mehdipatnam, Main Road
HYDERABAD, TELANGANA, 500089
IN
State/UT Code:36

PAN No:AAJCC9783E

GST Registration No:36AAJCC9783E1Z8

Dynamic QR Code:



Shipping Address :

Ramya Devi
K-809, ABC Apartments, Madhapur,
Mehdipatnam, Main Road
HYDERABAD, TELANGANA, 500089
IN

State/UT Code:36

Place of supply:TELANGANA

Place of delivery:TELANGANA

Order Number:407-7033233-4757902

Order Date:30.11.2024

Invoice Number :HYD3-1279928

Invoice Details :TG-HYD3-297683823-2425

Invoice Date :30.11.2024

Sl. No	Description	Unit Price	Qty	Net Amount	Tax Rate	Tax Type	Tax Amount	Total Amount
1	Corsair Katar Pro Ultra-Light Optical USB Gaming Mouse with Backlit RGB LED, 12400 DPI (Black) B08KSZV3SR (B08KSZV3SR) HSN:84716060	₹1,143.22	1	₹1,143.22	9%	CGST	₹102.89	₹1,349.00
					9%	SGST	₹102.89	

TOTAL:

₹205.78 ₹1,349.00

Amount in Words:

One Thousand Three Hundred Forty-nine only

For CLICKTECH RETAIL PRIVATE LIMITED:

Authorized Signatory

Whether tax is payable under reverse charge - No

*ASSPL-Amazon Seller Services Pvt. Ltd., ARPL-Amazon Retail India Pvt. Ltd. (only where Amazon Retail India Pvt. Ltd. fulfillment center is co-located)

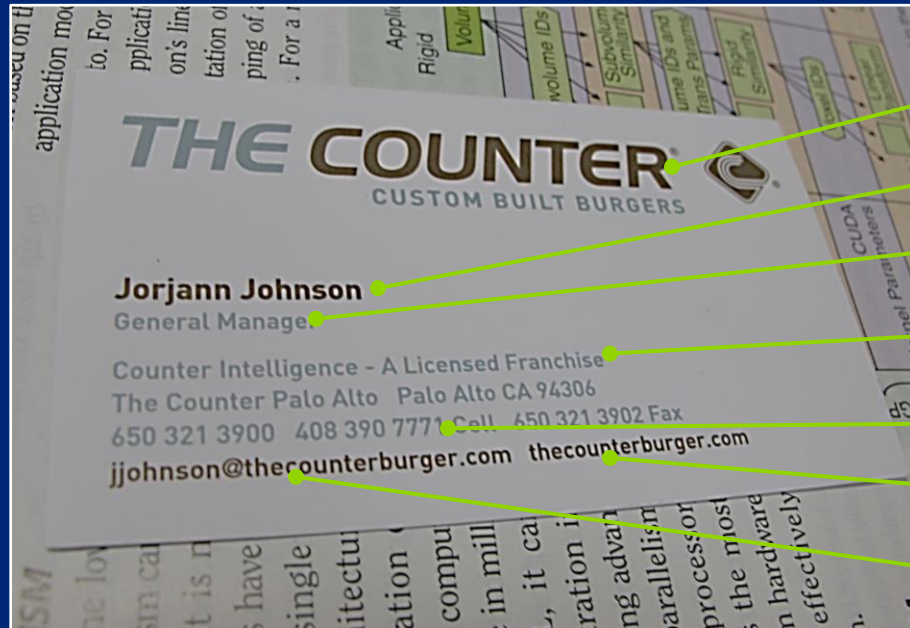
Customers desirous of availing input GST credit are requested to create a Business account and purchase on Amazon.in/business from Business eligible offers

Please note that this invoice is not a demand for payment

Page 1 of 1

Output from OpenAI GPT Model

```
{
  "invoice_number": "HYD3-1279928",
  "invoice_date": "30.11.2024",
  "seller_name": "CLICKTECH RETAIL PRIVATE LIMITED",
  "buyer_name": "Ramya Devi",
  "seller_address": [
    "SURVEY NUMBER 99/1, Mamidipally Village,
    Shamshabad, HYDERABAD, TELANGANA,
    500108,IN"
  ],
  "buyer_address": [
    "K-809, ABC Apartments, Madhapur,
    Mehdiapatnam, Main Road, HYDERABAD,
    TELANGANA, 500089, IN"
  ],
  "total_amount": "1349.00",
  "currency": "INR",
  "items": [
    {
      "item_number": "1",
      "description": [
        "Corsair Katar Pro Ultra-Light Optical USB
        Gaming Mouse with Backlit RGB LED, 12400 DPI
        (Black)"
      ],
      "quantity": "1",
      "item_amount": "1349.00"
    }
  ]
}
```



Company

Name

Designation/Title

Address

Phone Numbers

Website

Email

Business Card



Business Card

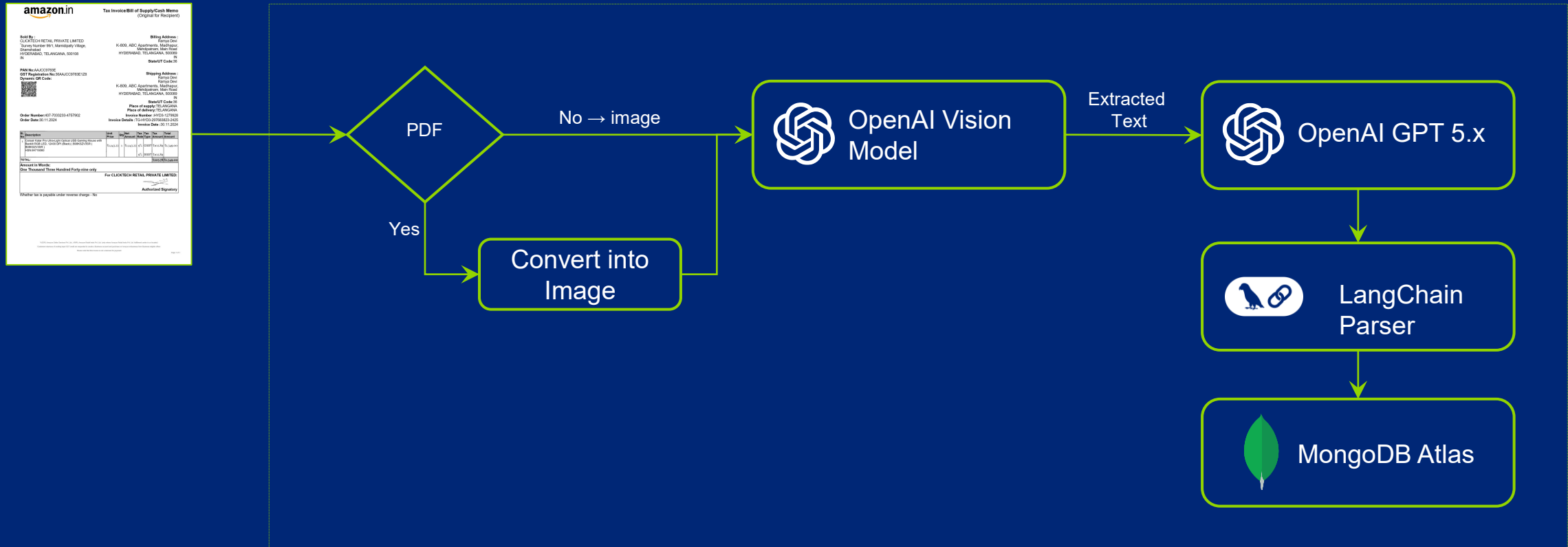


```
{
  "name": "Jorjann Johnson",
  "title": "General Manager",
  "company": "The Counter",
  "phone_numbers": [
    "650-321-3900",
    "408-390-7771",
    "650-321-3902"
  ],
  "emails": [
    "jjohnson@thecounterburger.com"
  ],
  "website": "thecounterburger.com",
  "address": [
    "The Counter, Palo Alto, CA 94306"
  ],
}
```

Architecture

Intelligent Document Extraction with OpenAI GPT & LangChain

→ Python, OpenAI API, Langchain, MongoDB, FastAPI



Road Map of the Course

MongoDB Atlas

- MongoDB Atlas Setup
- PyMongo
- CRUD Operations
Create, Read, Update & Delete

1

FastAPI

- API and HTTP
- Basic of FastAPI
- Pydantic for data validation
- Connecting FastAPI with MongoDB
- Render Templates

2

LangChain & OpenAI GPT

- Introduction to LLMs
- OpenAI GPT and models
- Prompt engineering basics
- Structured output with LLMs
- Parsing LLMs output

3

Capstone Project

- Intelligent Invoice Extraction
- Extract text using GPT vision models
- Entity extraction
- Validate output using Pydantic
- Store extract data in MongoDB

4

End to End Application

- Integrating all components
- MongoDB
+ FastAPI
+ LangChain
- Final Web Application

5



Setting up Course

→ Pre-requisite: Python

Software

- Python
- IDE
 - Jupyter Notebook
 - VS Code

Setup

- Download Folder Structure
- Install Poetry → Dependency Management.
- Required Installations.
- Base Path & Create .env file → “secrets.env”

Download Folder Structure

→ **GitHub:** <https://github.com/vedicskill/udemy-intelligent-doc-openai>



Install Poetry

Python Packaging and Dependency Management



Install Libraries

Dependencies Libraries for the Course

Libraries

Modules	Packages
IDE, ENV	jupyter python-dotenv
Mongodb	pymongo[srv]
FasAPI	fastapi[standard] uvicorn httpx httpie requests
LangChain & OpenAI	langchain langchain-openai openai

Next



MongoDB Atlas