



# FormX.ai

A Low Code Document Extraction Platform



# FormX.ai

A Low Code  
Document Extraction Platform

1

Pre-built AI for extracting  
**Financial Statement, IDs,  
Application Docs, Report,  
Invoices, Receipts**

2

Help our clients  
**Save manual effort,  
cross check input error,  
streamline the workflow**

3

With ROI

**>80%**

# We turn unorganised data from OCR engines to structured data in JSON

From...


"人票易通知書 Cheque Deposit AVICE\nHSBC匯豐\n戶口資料\nAccount Information\n**WP Monthly Parking Fee**\nSTIC  
\* \* ANY PA克\* \* \*\nD\n日期及時間\n銀行備考\nTVr201卷  
15:19:21\nBank Reference 3\n存入支票數目\nCheque  
scepted\n金額總數\n| Total Amount\n**S3213**\n客戶備註  
\nCustomer Reference\n20001:227555550282\n交易狀況  
\nTransaction Status\n接\n\*1max,\n定心事29c\nnm\n1: 14:  
\_\_\n14.iiiiiiiiii;\nna.alist\n','\nnt\n\_{\niti  
Pial\n.\n.\n111'\n,\n\n\nni=R\n二\n:\nbar;\n!\n中国在2030,32\n註  
\n:\n\"。 \n.3.3.5\n1\n1\nn",

(Data got from an OCR engine)

To...

```
{  
  "content": "\"WP Monthly Parking Fee",  
  "id": "fa6174b6-d530-4a63-887e-2d4122c18bb5",  
  "label": "account",  
  "languages": ["eng"]  
},  
{  
  "content": "$321.30",  
  "id": "e772d1fb-b8ed-4425-a10c-06d246cef7b0",  
  "label": "amount",  
  "languages": ["eng"]  
}
```

(Structured data returned by FormX)



# Why clients Adopt FormX.ai

---

# Client's Goal



## Enhance UX

Users no longer need to type input for applications or submission. All they need is take a photo on relevant documents, then FormX will extract the information and fill the form for the users.



## Release Manpower

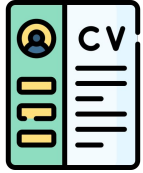
No need to work on the programming and no need to work on manual input. The IT team can focus on more important application and the user can self-serve using AI for the tedious work.



## Speed Up Process

FormX shortens the time and minimize manual input error, Client can speed up lots of work process, application process to provide a better service for internal terms or customers.

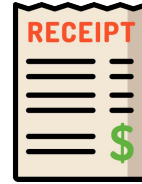
# Use Cases by FormX



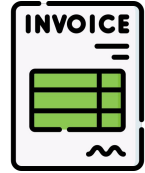
Resumes



Certificates



Receipts



Invoices



Bank Statements



Passports and IDs

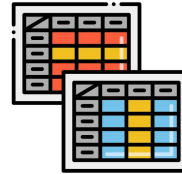


Table Forms



Contracts/Agreements

**And with GPT**

---

# FormX & GPT-4 Integration



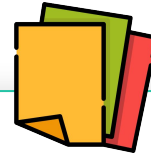
## Increase Accuracy

This integration results in a more comprehensive understanding of data, which enables organizations to process data extracted more accurately and efficiently



## Automate more complicated use case

This integration has accelerated digital transformation for all types of documents from simple format to complex format documents



## Train with fewer Samples

Creating your own extractor has become much more efficient with minimum training samples



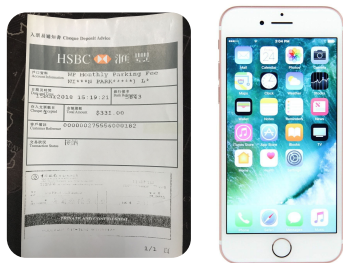
# How FormX works?

---

1

### Upload a photo

Your users scan and upload a form via your app.



2

### Pre-process

Optimise the scanned photo for light condition, skew correction, edge detection, etc.

3

### Pass it to OCR

Pass it to Google or Azure Vision or other 3rd party OCR engines to recognise the characters of each field.

5

### Return a JSON

A key-value pair JSON that contains the relevant fields will be returned to your app.

4

### Extract information

We extract data by semantic, layout and other info with ML models, and also auto-correct the data

# Use Our Pre-Built Extractors

## Fixed Layout

A document with exact same format and layout for all copies.

HKID	HK BR	Macau ID
NRIC Singapore ID	{any country}'s ID	
{any country}'s driving license		

## Variable Layout

A document type with different appearance for different document copies.

Receipt	Address Proof
International IDs	Boarding Pass
Invoice (Beta)	Passport

# Build Your Own Extractor

## Fixed Layout

Configure with 1 neat Document Copy on our Web portal. You can set up the fields you would like to extract easily.

Upload **1** Copy to Setup

## Variable Layout

Prepare 50 - 100+ copies (depending of the diversity of the documents) for training your own AI Extractor without a developer or AI engineer.

**Enhance** the Accuracy of Pre-built Extractor for your Use Case

**Train** a completely **new** Extractor

# Pre-processing Features

1

## **Adjust the angle and quality of an image**

FormX pre-process uploaded photos to improve the OCR accuracy. Photos are optimised for lighting condition, contrast, skew correction, edge detection, etc.

2

## **Classifier for Multiple Documents**

Our **Extractor Group** helps you classify the document. All uploaded documents will be automatically assigned to the best-matched document extractor.

3

## **Split a page of multiple documents**

If users put everything (such as passport & driving license) together in a page of document, FormX helps you split it into multiple images for further extraction.

## Other Features

4

### **Logos / Portrait Extraction**

Besides texts, FormX is able to recognise any images such as logo, portrait, signature, etc.

5

### **Desktop & Webhook Support**

Besides JSON API, you can also use our desktop apps or webhook feature to integrate with G Sheet / Airtable to batch process the documents.

6

### **Multiple OCR engines support**

FormX currently supports Google Vision API, Azure Computer Vision OCR and Tesseract. It also means we support a lot of different languages.

# Request A Demo From



**May Yeung**

FormX.ai Project Lead

---

[may.yeung@oursky.com](mailto:may.yeung@oursky.com)