Fireworks Project

Understanding the project:

* Problem: Extract relevant information across various types of documents such as Passports and Drivers licenses to create an Identity Verification solution for a FSI enterprise (Know Your Client protocol)
* Core Components:
  + Document upload (images – passports/licences)
  + Document data extraction
    - PNG, JPG, and JPEG images
  + Identity validation
    - Valid form of identification (Passport or Drivers License)
    - Complete and correct (in right format) data
    - Customer Identification (CIP) requires validation of name, date of birth, address, and identification number

\*\* The key component is extracting the data correctly

\* My first thought is to use image to text – let’s see how that works

Testing the Llama 3.2 90B Vision Instruct model in the Fireworks playground

\*\* Image needs to be before text in User Content – Seems off in docs! \*\*

Building the PoC

Stripe Identity

Design Requirements

* + Document upload (images – passports/licences)
  + Document data extraction
    - PNG, JPG, and JPEG images
  + Identity validation
    - Valid form of identification (Passport or Drivers License)
    - Complete and correct (in right format) data
    - Customer Identification (CIP) requires validation of name, date of birth, address, and identification number

Dev Process:

* Image MUST be before the user prompt – seems off in the documentation but I saw the warning at the top of the docs
* Llama 3.2 90B needed clarification to ensure it doesn’t hallucinate/use it’s own knowledge to fill data (Ben Franklin)
* Using 405B for verification right now since it is able to understand more complexities – may bump down
  + 3B not doing well – could finetune
* Which model to use for parsing out information from the chat to store?
  + Function calling v2 seems to work the best
  + Is prompting the problem?
  + Feed the conversation data / data schema in differently?

TODO:

* Clean up all functionality
* Value prop => costing => for FSI enterprise
* Create video documenting development/solution
* Clean code
* Publish Git Repo
  + Add Markdown
    - Include video
* Email Janelle

Design Choices:

* Using a compound approach for verifying users
  + Parse data from image using Llama 3.2 90B
  + Verify if that data is valid/complete with Llama 3.1 405B – Opportunity for finetuning – Was getting inconsistency with smaller models
* Using a compound approach for gathering data via a chat interface
  + Llama 3.2 3B for chatting with the user based on data needed
  + Llama 3.1 405B for parsing data from conversation – Another opportunity to improve prompting and/or finetune a smaller model
* Using 405B for parsing out data from chat – reliability but opportunity to fine-tune

Cost Analysis:

A close up of a number

Description automatically generated

Chat – more intuitive comfortable experience for the user than large forms

* Also may get additional details if you make it easier for the user