An insurance survAIyor (Assurant Challenge 1)

Kshitij Duraphe, Shaunak Joshi, Minyoung Park, Rong Li

Objective

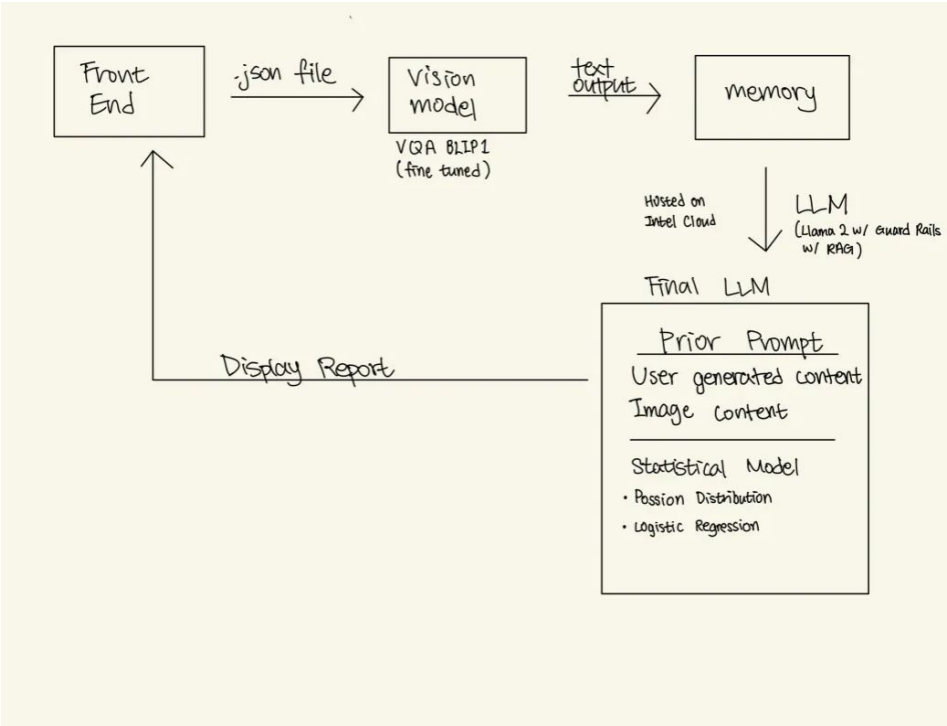
- Implement Multi-Mode Generative AI technology to evaluate the health of a property through image analysis and feature examination.
- **Practically -** Have a system that :

1. Analyzes images of a property
2. Feeds its analysis and any associated text into an AI
3. Uses language processing techniques to generate insights
4. Returns a report containing these insights, grade, and information

Our implementation

MVP:

1. Have a **frontend** for images/text
2. Send data to VQA
3. **VQA:** image-> text, concatenate
4. VQA Output+Risk assesment -> LLM
5. LLM Output: **Insight, Policy, Grade**



Technologies used

- Python Scientific Stack
- **VQA:** BLIP1 Model (finetuned on **iStockPhoto** dataset), LabelStudio
- **LLM:** Llama-2, PredictionGuard, Langchain, Retrieval Augmented Generation, Intel Cloud
- **Interface**: Node.js, PHP, SQL
- **Dataset:** iStockPhoto Interior+Exterior images of property damage

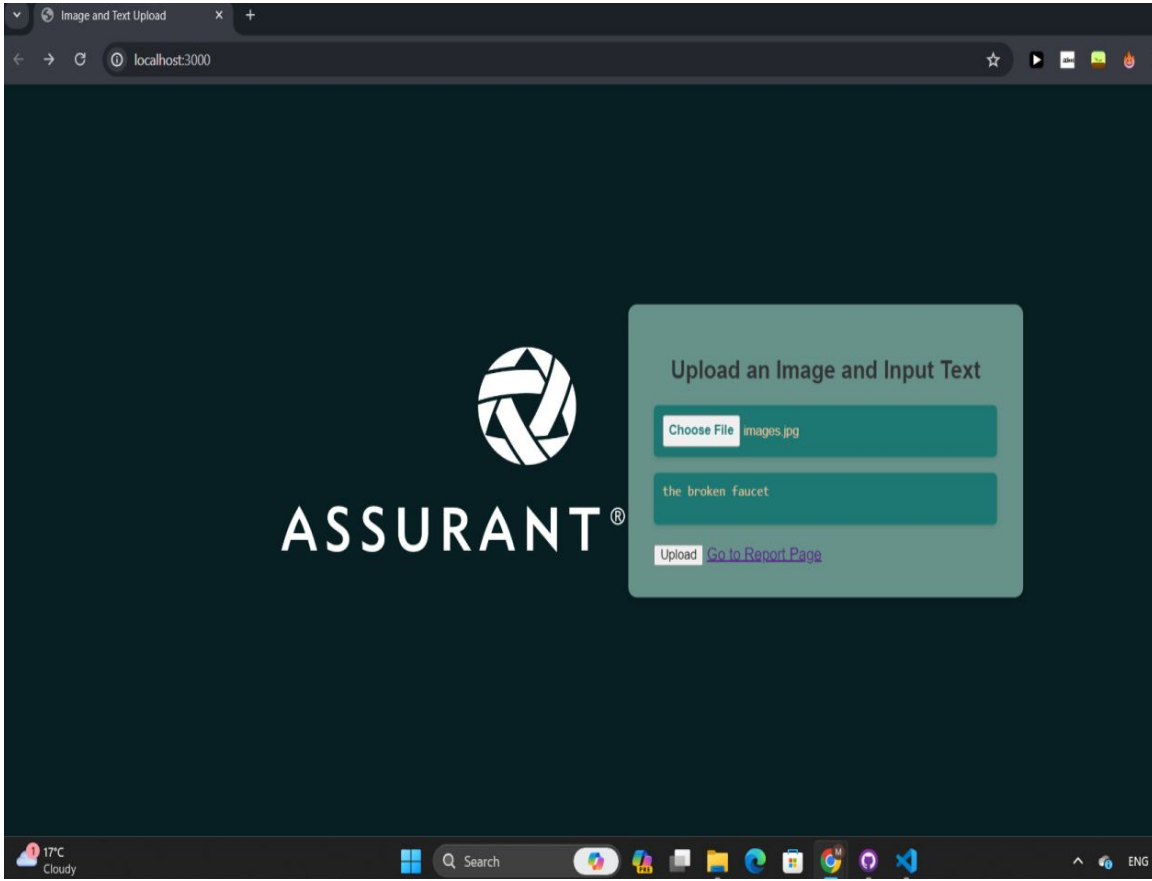
Frontend

What the user sees:

- Upload a photo
- Type in complaints or comments
- Get a **personalized report**

What the insurance company can do:

- Associate client to photos and complaints
- Speed up claim processing by generating reports



VQA

- **Foundational GenAI model:** Chosen because we wanted to extrapolate from a small dataset
- **Fine-tuning:** Shift distribution to images we want (property damage images)
- **Annotate** images with LabelStudio

```
answer: hurricane
istockphoto-1324873600-612x612.jpeg
```

```
load checkpoint from https://storag
answer: water damage
istockphoto-937598662-612x612.jpeg
```

```
load checkpoint from https://storage
answer: hurricane
istockphoto-1362610895-640x640.jpeg
```

Statistical Model

- **Factors considered:** Age of house, building material, length of cracks in the wall, width of the radius, proneness to catastrophes, location level (whether the surrounding locality is filled with crime), water leakage levels
- **Logistic regression** implemented as a test, with a Poisson distribution

```
we take an sample with the age_of_house: 30 years, material_factor: 0.2, cracks_length: 0,
 cracks_width:2,leakage_radius: 0, catastrophe_level: 0, location_level:0, to predict the probability of filing a file within 6 month
The probability of filing a claim within 6 month is: 0.39

Process finished with exit code 0
```

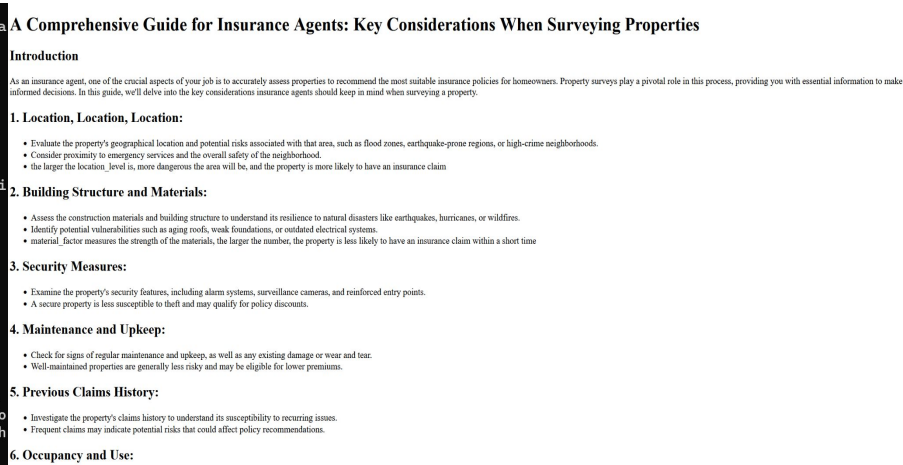
LLM

- **Nous-Llama2** with **guardrails (PredictionGuard)** and **Retrieval Augmented Generation** (RAG)
- **Guardrailing** done by hand as well. **RAG** samples from guide for insurance agents.
- Only 2 calls to the LLM to generate everything

```
ubuntu@kshitij-instance:~$ python3 rag.py
/home/ubuntu/.local/lib/python3.10/site-packages/langchain/__init__.py:29: UserWarning: Importing PromptTemplate from la
ngchain root module is no longer supported. Please use langchain.prompts.PromptTemplate instead.
  warnings.warn(
/home/ubuntu/.local/lib/python3.10/site-packages/langchain/__init__.py:29: UserWarning: Importing FewShotPromptTemplate
from langchain root module is no longer supported. Please use langchain.prompts.FewShotPromptTemplate instead.
  warnings.warn(
/usr/lib/python3/dist-packages/scipy/__init__.py:146: UserWarning: A NumPy version >=1.17.3 and <1.25.0 is required for
this version of SciPy (detected version 1.26.4
  warnings.warn(f"A NumPy version >={np_minversion} and <{np_maxversion}"
WARNING ⚠ user config directory '/home/ubuntu/.config/lancedb' is not writeable, defaulting to '/tmp' or CWD. Alternati
vely you can define a LANCEDB_CONFIG_DIR environment variable for this path.
reading vision data
reading frontend data

3 insights that we've generated based on your report are:
 Insight 1: Inspect and repair any earthquake damage promptly to prevent further damage and ensure safety.
Insight 2: Address the water damage immediately to prevent mold growth and structural damage.
Insight 3: Seal any entry points to prevent rodent infestation and maintain the property's integrity.
The full report is:
Rating: 4

Based on the insights provided, the property has a mix of positive and negative factors. The insights indicate that the
current has long cracks and wide cracks, which suggests that there may be potential issues with the home's structure. Ho
wever, the age of the housing is not a major concern in this case. Additionally, the probability of filing a claim in th
e next 3 to 6 months is relatively low at 0.37
```



A Comprehensive Guide for Insurance Agents: Key Considerations When Surveying Properties

Introduction

As an insurance agent, one of the crucial aspects of your job is to accurately assess properties to recommend the most suitable insurance policies for homeowners. Property surveys play a pivotal role in this process, providing you with essential information to make informed decisions. In this guide, we'll delve into the key considerations insurance agents should keep in mind when surveying a property.

1. Location, Location, Location:

- Evaluate the property's geographical location and potential risks associated with that area, such as flood zones, earthquake-prone regions, or high-crime neighborhoods.
- Consider proximity to emergency services and the overall safety of the neighborhood.
- the larger the location_level is, more dangerous the area will be, and the property is more likely to have an insurance claim

2. Building Structure and Materials:

- Assess the construction materials and building structure to understand its resilience to natural disasters like earthquakes, hurricanes, or wildfires.
- Identify potential vulnerabilities such as aging roofs, weak foundations, or outdated electrical systems.
- material_factor measures the strength of the materials, the larger the number, the property is less likely to have an insurance claim within a short time

3. Security Measures:

- Examine the property's security features, including alarm systems, surveillance cameras, and reinforced entry points.
- A secure property is less susceptible to theft and may qualify for policy discounts.

4. Maintenance and Upkeep:

- Check for signs of regular maintenance and upkeep, as well as any existing damage or wear and tear.
- Well-maintained properties are generally less risky and may be eligible for lower premiums.

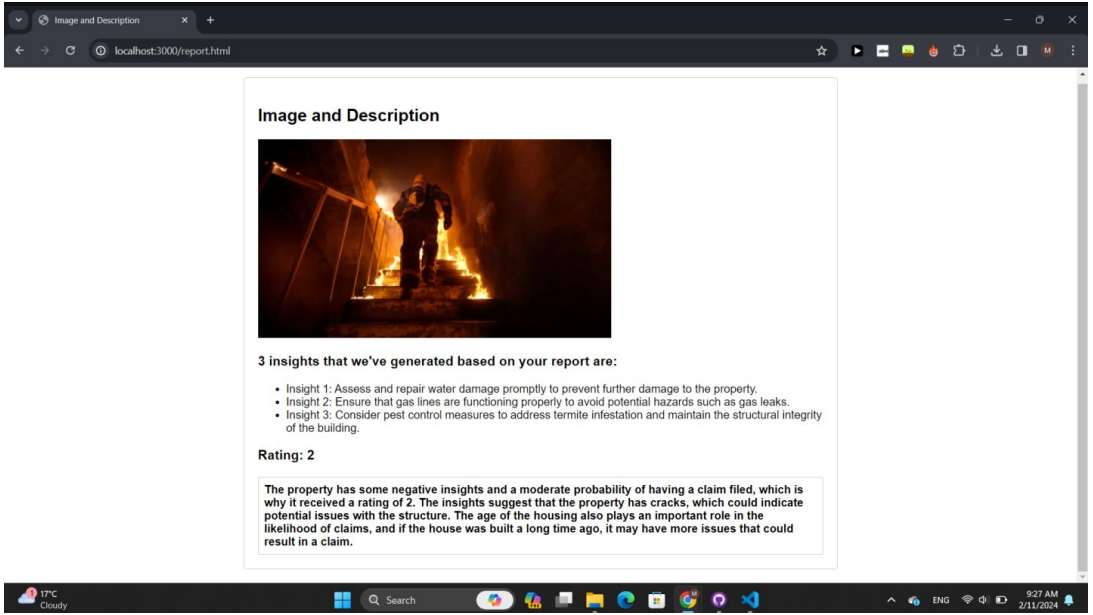
5. Previous Claims History:

- Investigate the property's claims history to understand its susceptibility to recurring issues.
- Frequent claims may indicate potential risks that could affect policy recommendations.

6. Occupancy and Use:

Final Output

- **Image** of claim
- **Insights**
- **Rating**
- **Report**



Challenges faced

- **Lack of computational power:** Intel VMs are slow
- **No real chance to optimize neural networks with Intel libraries**: The models we used were due to the lack of memory and storage space. Initial tests on GPT4 revealed much better VQA capabilities
- **Lack of good insurance datasets for our specific use case**: Most datasets are related to things like age, price of the property, habits, and so on. Therefore we had to come up with one from scratch

Future Scope

- Create some sort of real-time tool (like a phone app that you can take photos and send them to the cloud with)
- Add stronger models for better computational and generational capabilities
- Better risk assessment algorithms

Thank you!