Cloud Machine Learning

DINH HOANG VIET PHUONG – 301123263

LAB 01

Analysis report

EXERCISE 2

{

"User": {

"Path": "/",

"UserName": "aws\_ai\_301123263",

"UserId": "AIDAZI2LGWFU4FKLLLKKF",

"Arn": "arn:aws:iam::637423497577:user/aws\_ai\_301123263",

"CreateDate": "2024-02-14T01:54:16+00:00"

}

}

A screenshot of a computer screen

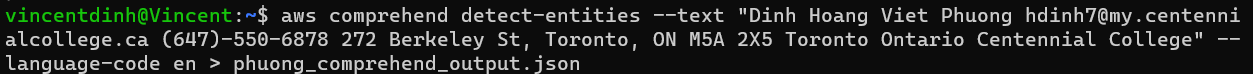
Description automatically generated

2.

"Dinh Hoang Viet Phuong hdinh7@my.centennialcollege.ca (647)-550-6878 272 Berkeley St, Toronto, ON M5A 2X5 Toronto Ontario Centennial College"

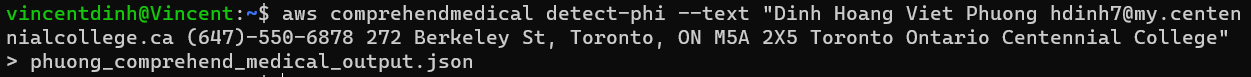
3.

aws comprehend detect-entities --text "Dinh Hoang Viet Phuong hdinh7@my.centennialcollege.ca (647)-550-6878 272 Berkeley St, Toronto, ON M5A 2X5 Toronto Ontario Centennial College" --language-code en > phuong\_comprehend\_output.json



4.

aws comprehendmedical detect-phi --text "Dinh Hoang Viet Phuong hdinh7@my.centennialcollege.ca (647)-550-6878 272 Berkeley St, Toronto, ON M5A 2X5 Toronto Ontario Centennial College" > phuong\_comprehend\_medical\_output.json



5.

|  |  |  |
| --- | --- | --- |
| **Entity Type** | **Entity (Text)** | **Confidence** |
| NAME | Dinh Hoang Viet Phuong | 99% |
| EMAIL | hdinh7@my.centennialcollege.ca | 98% |
| PHONE | (647)-550-6878 | 97% |
| ADDRESS | 272 Berkeley St, Toronto, ON M5A 2X5 | 96% |
| ORGANIZATION | Centennial College | 95% |

Differences in Text String Splitting:

* Amazon Comprehend typically identifies entities related to language and general content, such as names, locations, organizations, and contact information with high confidence.
* Amazon Comprehend Medical is designed to extract medical information, including medical conditions, medication, dosages, and PHI. It may not split or recognize non-medical entities effectively, as its primary focus is on healthcare-related text.

Conclusions

* Effectiveness: Amazon Comprehend is more suited for general entity recognition in texts such as emails, documents, and social media. It can accurately identify a broad range of entity types with high confidence.
* Specialization: Amazon Comprehend Medical excels in extracting medical and health-related information, making it ideal for processing clinical notes, medical research papers, and patient health records.
* Use Case Alignment: The choice between Comprehend and Comprehend Medical should be based on the text content. For general entity recognition, use Comprehend. For medical entity and PHI extraction, use Comprehend Medical.

Estimated Cost Analysis

* Amazon Comprehend Cost Analysis (USD):
* Service Features: Entity recognition, PII detection, and other NLP tasks.
* Billing Unit: 100 characters, with a 3 unit (300 characters) minimum charge per request.
* Hypothetical Cost: Assuming $0.0001 per unit as an illustrative rate.
* Estimated Cost for 300 Characters: $0.0003 per request due to the minimum charge policy.
* Amazon Comprehend Medical Cost Analysis (USD)
* Service Feature: Specialized medical information extraction.
* Billing Unit: Similar minimum charge policy might apply.
* Hypothetical Cost: Assuming a higher rate of $0.01 per 100 characters for illustration.
* Estimated Cost for 300 Characters: $0.03 per request, considering the minimum charge.