module_1.md

Module 1: Jupyter-based integrated solution

Resources

- Blog Post: De-identify medical images with the help of Amazon Comprehend Medical and Amazon Rekognition
- Source code: GitHub: aws-samples/amazon-comprehend-medical-image-deidentification

Process

Step 1: Launch the CloudFormation Stack

- This CloudFormation stack deploys the infrastructure we will need in Module 2, and creates an S3 bucket we will use in this module.
- Reference: Implementation Guide, Step 1
- Sign in to the AWS console
- Click this CloudFormation console link to launch the CloudFormation stack:
  - The template file is pre-filled in the form (the template itself is `deployment/ai-powered-health-data-masking.template` in the source code for Module 2).
  - **Specify stack details** page: Enter a stack name
  - **Configure stack options** page: Accept defaults
  - **Review** page: Select Acknowledge box
- From the **Outputs** of the stack, make note of the **ImageBucketName**

Step 2: Upload the sample image file

- Download the sample image file `deidentify-medical-2.jpg` from this GitHub link by right-clicking on **Download** and saving to your local machine
- Open the S3 service console, and click on the bucket created in Step 1.
- Upload into this bucket the sample image file from above

Step 3: Create a SageMaker instance

- Open the SageMaker service console
- Select **Notebook** → **Notebook instances** from the left-hand sidebar
- Click **Create notebook instance** from the Notebook Instances page
  - Enter a suitable name
  - Select **ml.t3.medium** instance type (2 vCPU, 4 GiB memory, $0.0582/hr)
  - In **Permissions and encryption** → **IAM role**, select **Create a new role**
  - In the **Create an IAM role** popup, select **Any S3 bucket**
  - Note the name of the newly-created role (it will start with **AmazonSageMaker-ExecutionRole-**)
  - Click *Create notebook instance* to launch the instance (2 minutes)

Step 4: Add permission to use AWS services

- Open the IAM service console
- Locate the IAM role created in Step 2, and click on its name

- It contains these policies:
  - **AmazonSageMakerFullAccess** to run SageMaker
  - **AmazonSageMaker-ExecutionPolicy** allows access to S3
- Click **Attach Policies** and search for and select these policies:
  - **AmazonRekognitionFullAccess**
  - **ComprehendMedicalFullAccess**
- Click **Attach policy** and confirm that you now have 4 policies attached to this role

Step 5: Install the Jupyter notebook

- In the Notebook Instances tab, once the notebook instance is running, click on its name and click on **Open Jupyter** to open Jupyter in a new tab.
- Open another browser tab, open this GitHub link, right-click on **Download**, and save the file `medical_image_de_id.ipynb` to your local machine.
- Return to the Jupyter tab, upload the Jupyter notebook file
- Click **Upload**, locate and upload the notebook file
- Click on the notebook file to run the notebook in a new tab

Step 6: Edit and run the Jupyter notebook

- In the Jupyter notebook tab, edit the code cell with the **bucket** and **object** (path and name) of the image file in your bucket from Step 2
- Run the notebook by selecting **Cell** → **Run All** in the top menu bar
- The first time you run this, some software will be installed
- Experiment by adjusting the **phi_detection_threshold** to a value closer to 1.00