Research Artifacts Submission – Static Analysis of Privacy Policies and Descriptions in Alexa IoT Skills

We are submitting the following artifacts related to our project on identifying privacy and risk concerns in Alexa Skills:

1. **IOT_project_data.xlsx**
   This is the **primary dataset** we collected. It includes the following columns for each Alexa skill:
   - Skill Name
   - Skill Description
   - Privacy Policy Text
   - Rating
   - Number of Reviews
   - Vendor Name
2. **IOT_Security_Project_(alexa skills).ipynb**
   This is our **project notebook**, executable in **Jupyter Notebook** or **Google Colab**.
   - We recommend running it with **T4 GPU on Google Colab** for better performance.
   - Upon execution, it will prompt for the input data file (IOT_project_data.xlsx).
   - Run all cells sequentially to view analysis outputs and visualizations.
3. **risk_scored_for_skills.xlsx**
   This **output file** contains detailed risk evaluation for each skill, including:
   - Data Completeness (missing values)
   - Necessary vs. Unnecessary Data Fields
   - Computed Risk Score
   - Risk Classification (High, Medium, Low)
4. **vendor_rating_review_risk_summary.csv**
   This **summary file** aggregates vendor-level insights:
   - Vendor Name
   - Average Rating
   - Average Number of Reviews
   - Average Risk Score
   - Count of Skills by Risk Level

5. **vendors_with_certifications.txt**
   This **text file** lists vendors that have **at least one Alexa skill mentioning a privacy standard certification**.
   - The presence of a certification indicates the vendor *claims* adherence to certain privacy or security standards.
   - This listing does **not evaluate whether these certifications are valid or properly implemented**, only that they are referenced in the data.

Required Python Libraries

The following libraries are required to run the notebook:

- ipython
- nltk
- numpy
- pandas
- scipy

You can install them using:

pip install ipython nltk numpy pandas scipy