

# AviationQA

DATASET

## About

AviationQA is introduced in the paper titled-

**There is No Big Brother or Small Brother: Knowledge Infusion in Language Models for Link Prediction and Question Answering**

We create a synthetic dataset, AviationQA, a set of 1 million factoid QA pairs from 12,000 National Transportation Safety Board (NTSB) reports using templates. These QA pairs contain questions such that answers to them are entities occurring in the AviationKG (Agarwal et al., 2022). AviationQA will be helpful to researchers in finding insights into aircraft accidents and their prevention.

Examples from dataset:

1. What was the Aircraft Damage of the accident no. ERA22LA162?  
Substantial
2. Where was the Destination of the accident no. ERA22LA162?,  
Naples, GA (APH)

## Download

**Uncompressed csv file:**

[https://drive.google.com/file/d/17K5l4UFZStijFcZL-p1XOORI\\_TBOxj3W/view?usp=sharing](https://drive.google.com/file/d/17K5l4UFZStijFcZL-p1XOORI_TBOxj3W/view?usp=sharing)

**Compressed file(<10 MB):**

<https://drive.google.com/file/d/1u0eeq1Jg3BofCw1m9ozSXWIT23hYiLO/view?usp=sharing>

**Command to extract the csv file from compressed tar.xz file:**

```
tar -Jxvf aviation.tar.xz
```