Towards an Aviation Large Language Model by Fine-tuning and Evaluating Transformers

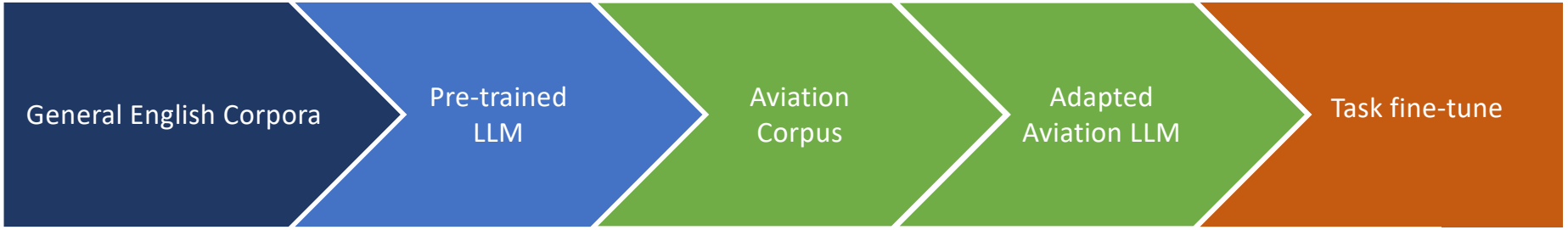
David Nielsen (KBR, Inc.),
Stephen S.B. Clarke, and Krishna Kalyanam (NASA)
NASA Ames Research Center, Moffett Field, CA, 94035, USA

Background and motivation

Why are we using large language models (LLMs)?

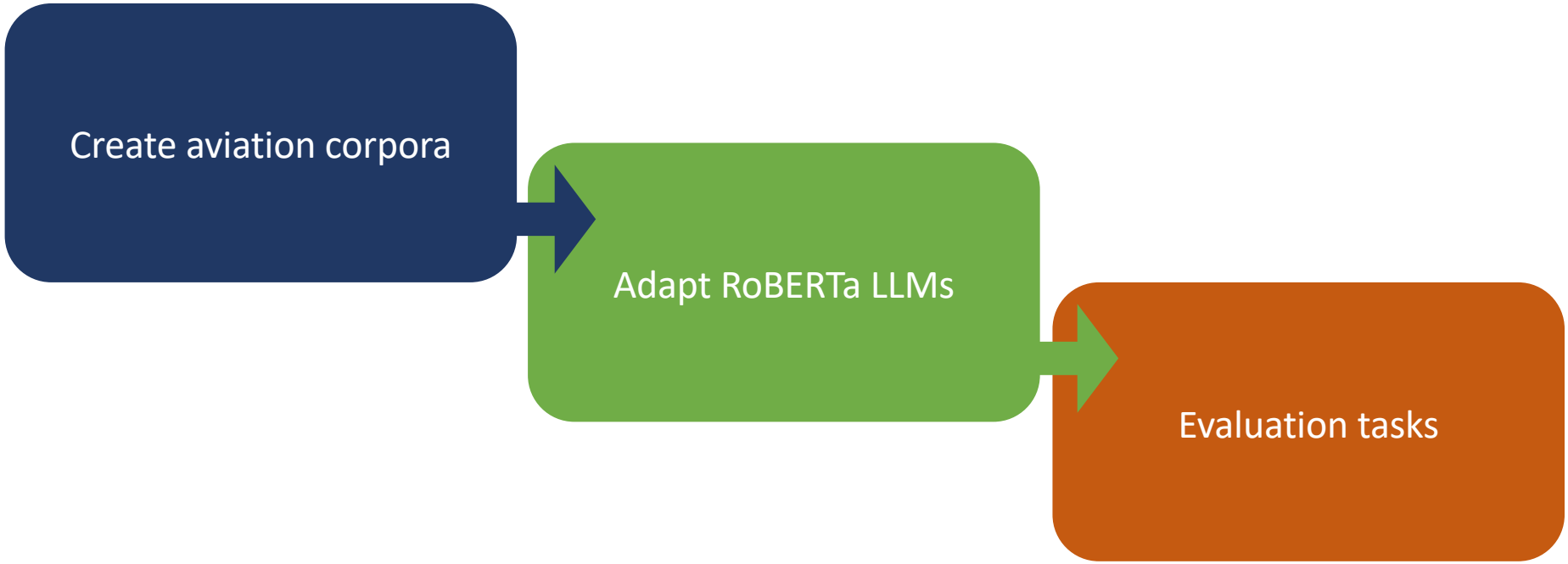
- What is an LLM?
  - Large: millions to trillions of parameters, large amount of training data
  - Language model: mathematical model of natural language
- LLMs provide pre-trained starting point for many NLP tasks
  - Fine-tuning vs training from scratch
  - Outperform many other language models
- Many different architectures, our focus has been RoBERTa (Robustly Optimized BERT Pretraining approach)
  - BERT: Bidirectional Encoder Representations from Transformers

Adapting an LLM



- Pre-trained LLMs can be adapted to a specific application domain to improve performance over base models
  - Med-BERT and BioBERT
  - LEGAL-BERT
  - Aviation-BERT
- Can we use Aviation domain specific text to get better LLM performance?

Methodology Outline



Aviation corpora creation

Corpora built from Letters of Agreement (LOA)

- Formal agreements between two or more parties about how airspaces will be used.
- Representative of the National Airspace System (NAS) terminology
- Parties can be FAA organizations, private companies, or the military
- Contain the constraints aircraft must follow

(Name) Center/Approach Control and (Name) FSS

LETTER OF AGREEMENT

EFFECTIVE: _____

SUBJECT: Special VFR Operations within (Name) Airport Surface Area

1. PURPOSE: To provide operating procedures for Special VFR flight handling in the (name) surface area without individual coordination.

2. SCOPE: The procedures outlined herein are for use in the conduct of Special VFR operations within the (name) Airport surface area at or below _____ feet. These procedures are applicable only to aircraft equipped with functioning 2–way radio in order to effect a recall when required by traffic or weather conditions.

3. RESPONSIBILITIES: Upon request by the (name) FSS, the Center/Approach Control Facility may authorize Special VFR operations in the (name) Airport surface area for specific periods of time. The Center/Approach Control Facility must retain the authority to withdraw the provisions of this agreement at any time.

4. PROCEDURES:

a. Local Special VFR operations. The (name) FSS must not authorize more than one aircraft to operate simultaneously in the surface area unless pilots agree that they will maintain visual separation with other aircraft operating in the surface area.

b. IFR Arrivals and Departures. Special VFR operations must be controlled by the (name) Center/Approach Control during the following periods:

(1) From 10 minutes prior to the estimated time of arrival of an IFR aircraft over the approach fix until it is on the ground (IFR arrivals must not be cleared for an approach until the FSS confirms that there are no Special VFR operations in progress.)

(2) From 10 minutes prior to the estimated time of departure of an IFR aircraft until it departs the surface area.

Air Traffic Manager, (Name) FSS

Air Traffic Manager, (Name) ARTCC/Approach Control

https://www.faa.gov/air_traffic/publications/atpubs/foa_html/chap4_section_3.html

Aviation corpus creation

- Two corpora created
  - Air Route Traffic Control Centers (ARTCC) LOAs: LOAs with an ARTCC as a party
  - Full LOA: all LOA pdfs including Terminal Radar Approach Control (TRACON), airports, etc.
- Allows assessment of corpora size on classification task performance

Current data	Number of PDFs	Number of pre-training documents
ARTCC LOA	1,595	7,057
Full LOA	7,497	29,904

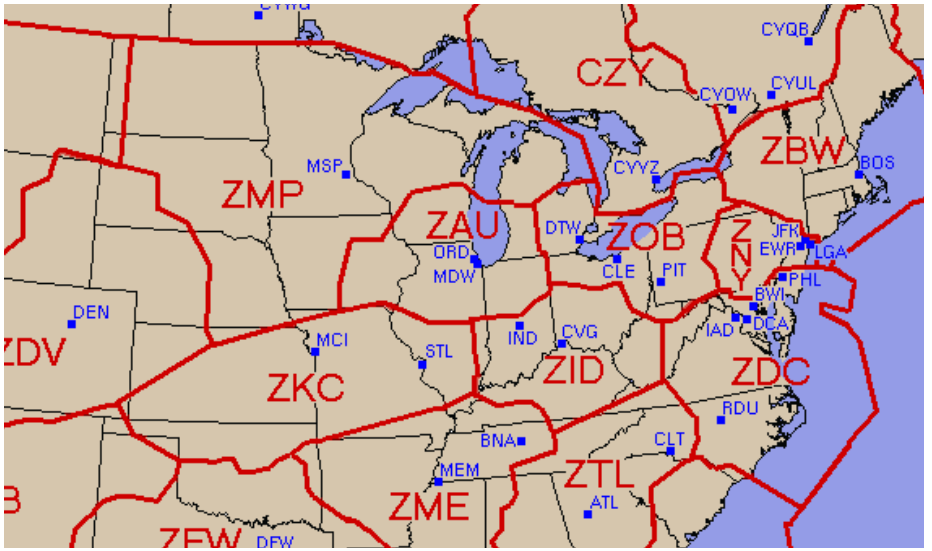


Image credit: FAA
https://www.fly.faa.gov/ois/tier/themap.htm

RoBERTa adaptation

RoBERTa Adaptation

- Continue masked language modeling (MLM) pre-training task with corpora
  - Mask text as seen on the right
  - Model predicts what words should be in the **[MASKED]** slot
- Adapted RoBERTa models for 30 different combinations of two hyperparameters
- From these 30 models, which are the 'best' performing ones?

Masked langue modeling example	
Raw text	Departures: Tower must: Assign altitudes as follows: Assign 16,000ft MSL for aircraft requesting FL170 or higher. Advise aircraft to expect clearance to filed altitude 10 minutes after departure.
Masked text	Departures: **[MASKED]** must: Assign altitudes as follows: Assign 16,000ft MSL for aircraft requesting FL170 or higher. Advise aircraft to **[MASKED]** clearance to filed altitude 10 minutes after departure.

Evaluation

Evaluation overview

- Two evaluation tasks built from held-out LOA data
  - Document Classification: civil vs non-civil document
  - Constraint Classification: trajectory constraint vs no-constraint in a line of text
- Used embeddings from adapted RoBERTa LLMs as input to a logistic regression classifier model for both tasks
  - Data split 90/10% into train and test

Task	Positive examples	Negative examples
Document classification	222	271
Constraint classification	129	370

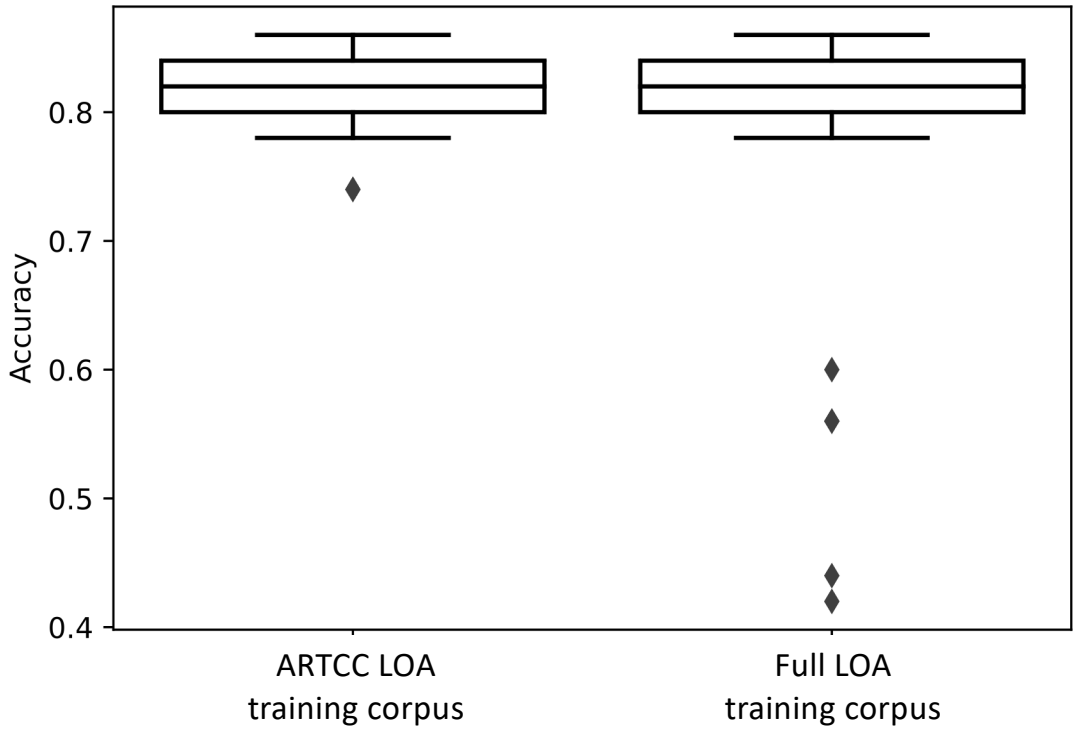
Document classification

- Task: can our model correctly label documents as ‘civil’ or ‘not-civil’?
  - Civil documents *only* have FAA signatories
  - Not-civil documents have at least one non-FAA signatory
- Subject matter experts (SMEs) labeled documents from the Dallas-Fort Worth ARTCC (ZFW)
- Classifier trained for each adapted RoBERTa model and evaluated on the two tasks

Civil	Not-civil	Total documents
222	271	493

Results: document classification

Results across all models



Model	Adaptation dataset	Test F1 civil	Test accuracy
Base RoBERTa	None	0.81	0.82
Best ARTCC model	ARTCC LOAs	0.83	0.84
Best Full LOA	Full LOAs	**0.85** (+0.04)	**0.86** (+0.04)

Constraint classification

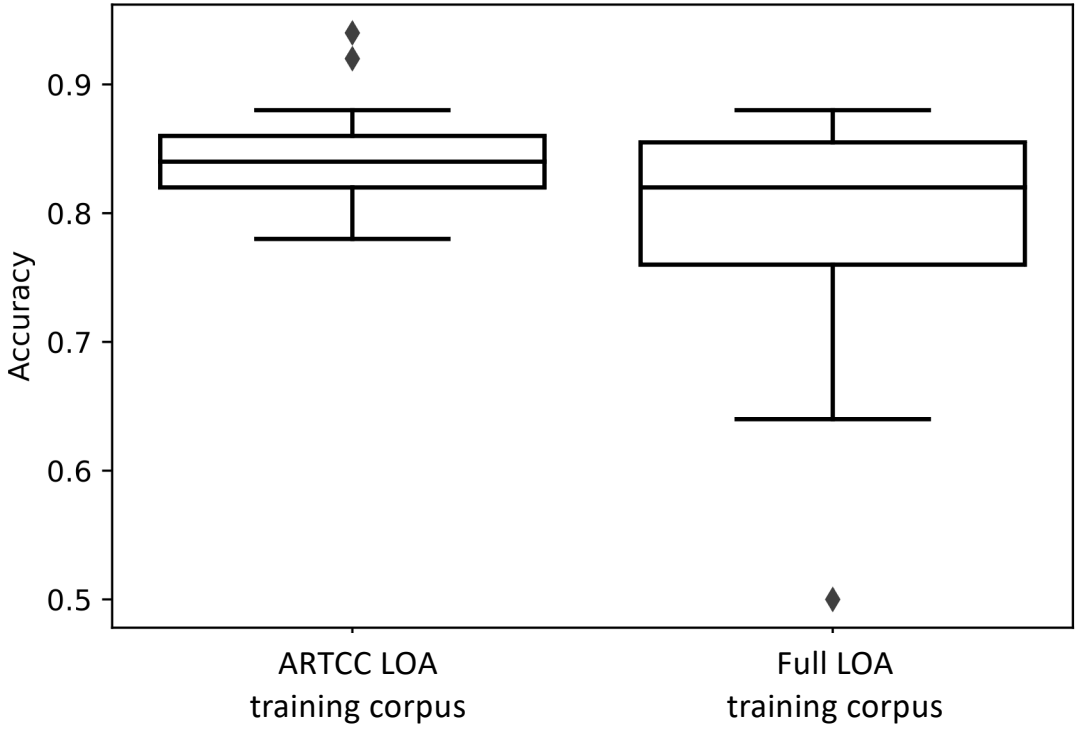
- Task: can our classifier correctly identify a line of text containing a trajectory constraint?
- SMEs labeled lines from ZFW LOAs and the same training and evaluation was performed

Constraint	Not-constraint	Total documents
129	370	499

Example line text	Constraint label
Departures: Tower must: Assign altitudes as follows: **Assign 16,000ft MSL** for aircraft requesting FL170 or higher	1
Departures: Tower must: Assign altitudes as follows: Aircraft requesting FL170 or higher will be handed off to ARTCC1.	0

Results: constraint classification

Results across all models



Model	Adaptation dataset	Test F1 constraint	Test accuracy
Base RoBERTa	None	0.74	0.82
Best ARTCC model	ARTCC LOA	**0.91** (+0.17)	**0.94** (+0.12)
Best Full LOA model	Full LOA	0.82	0.88

Conclusions

- Domain adaptation using aviation corpora improves performance across the board on both tasks.
- Larger corpora improved document classification performance
- Adapted models performed better on constraint classification
  - Specific ARTCC adaptation was more useful for this task than simply increasing the data size (i.e., number of documents)
  - Non-ARTCC LOAs contained less relevant information about constraints than the ARTCC LOAs

Future work

- Create larger aviation corpora
  - Use the public data from Aviation-BERT which contains safety reports
  - Additional FAA datasets (e.g., Standard Operating Procedures (SOPs))
- Add additional domain evaluation tasks and evaluate performance
- Compare BERT-base architectures with Generative Pre-trained Transformer (GPT) LLMs

Questions?

David Nielsen: david.l.nielsen@nasa.gov