

BERT-FlowDelta Improvement and NL2SQL Integration



Project Value: Our project improves a Q&A system model with access to Ford vehicle manuals that can answer customers questions about vehicle features and operation while capturing elements of human conversation

Mario Flores | Brendan Kerr
Pum Khai | Brandon Wu

BERT-FlowDelta Model and Improvement

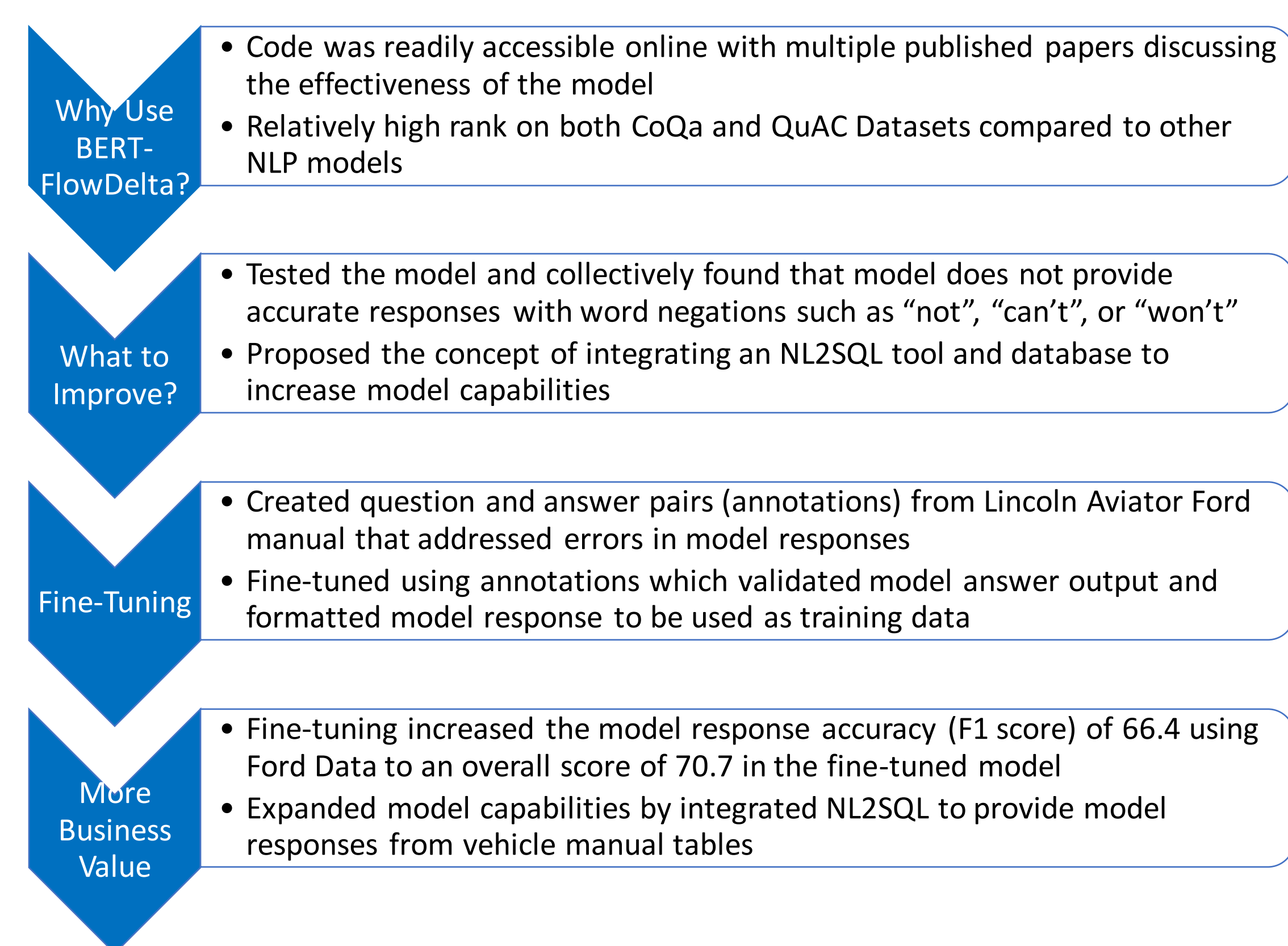
What is BERT-FlowDelta?

- A multiturn question and answer model that captures human conversation more accurately than single turn models
 - The flow of human conversation (i.e. how topics change) is captured by passing the difference of two previous hidden states
- The Modular Annotation team, built a tool to produce data to further fine-tune the model in the future

Model Improvements

- Originally, when provided with user-generated questions containing word negations, the model provided nonsensical and inaccurate responses
- To improve this, we gave the model user-generated questions that contained negations, validated model answer output, and formatted the answer output to be used as training data

Improvement Process

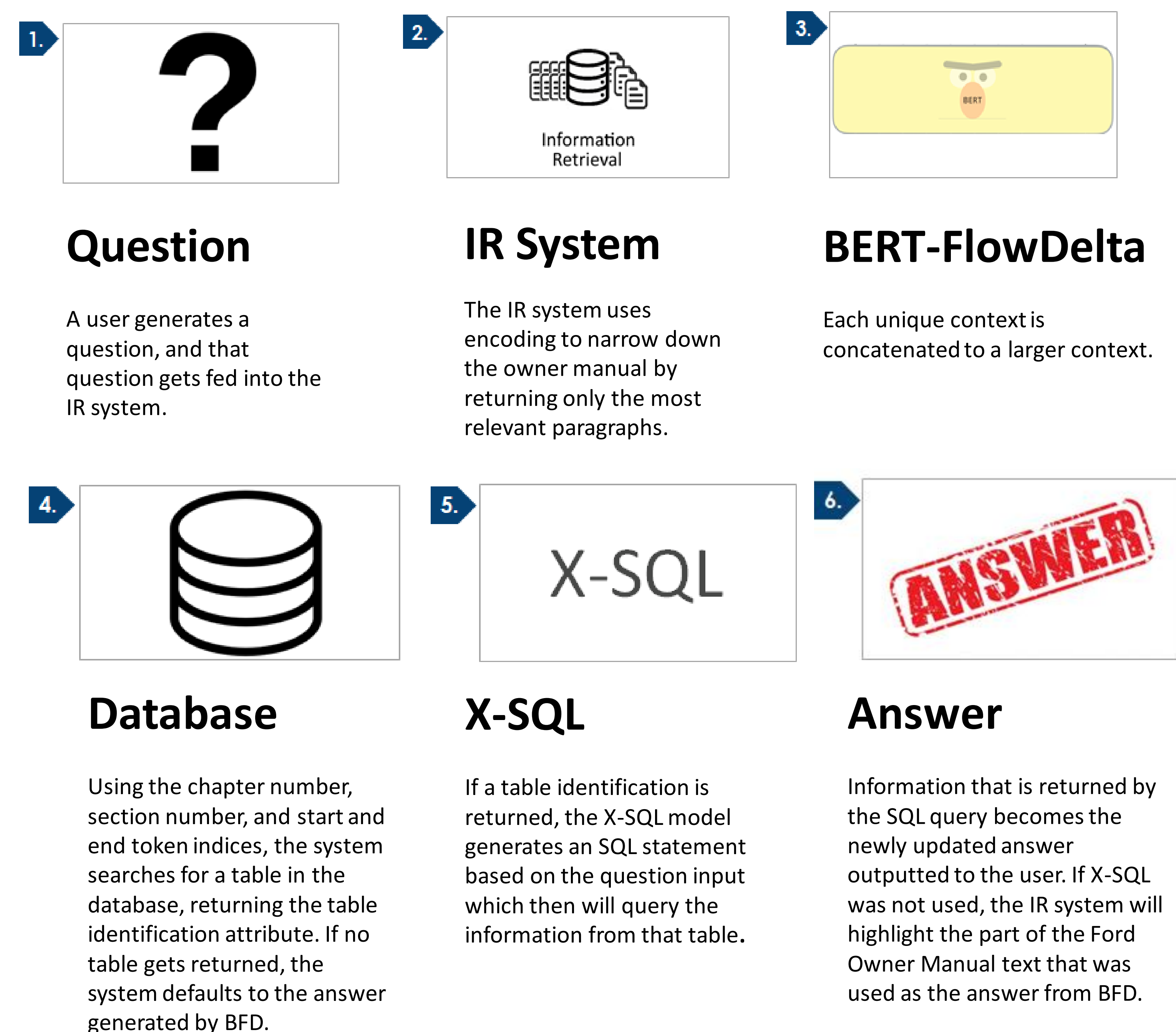


Integrating NL2SQL with BERT-FlowDelta

NL2SQL

- **Motivation:**
 - Ability to generate answers from the tables inside the Lincoln Aviator Manual
- **Requirements:**
 - **Database:** A collection of 176 tables extracted from the Lincoln Aviator Manual
 - **Metadata table** with entities that reference all other tables by a unique identifier (uuid)
 - **Natural Language question**
 - The corresponding **SQL query** (generated by X-SQL model)

Updated Question/Answer System Pipeline



NL Question
What **safety restraint** would you recommend for a **small child**?

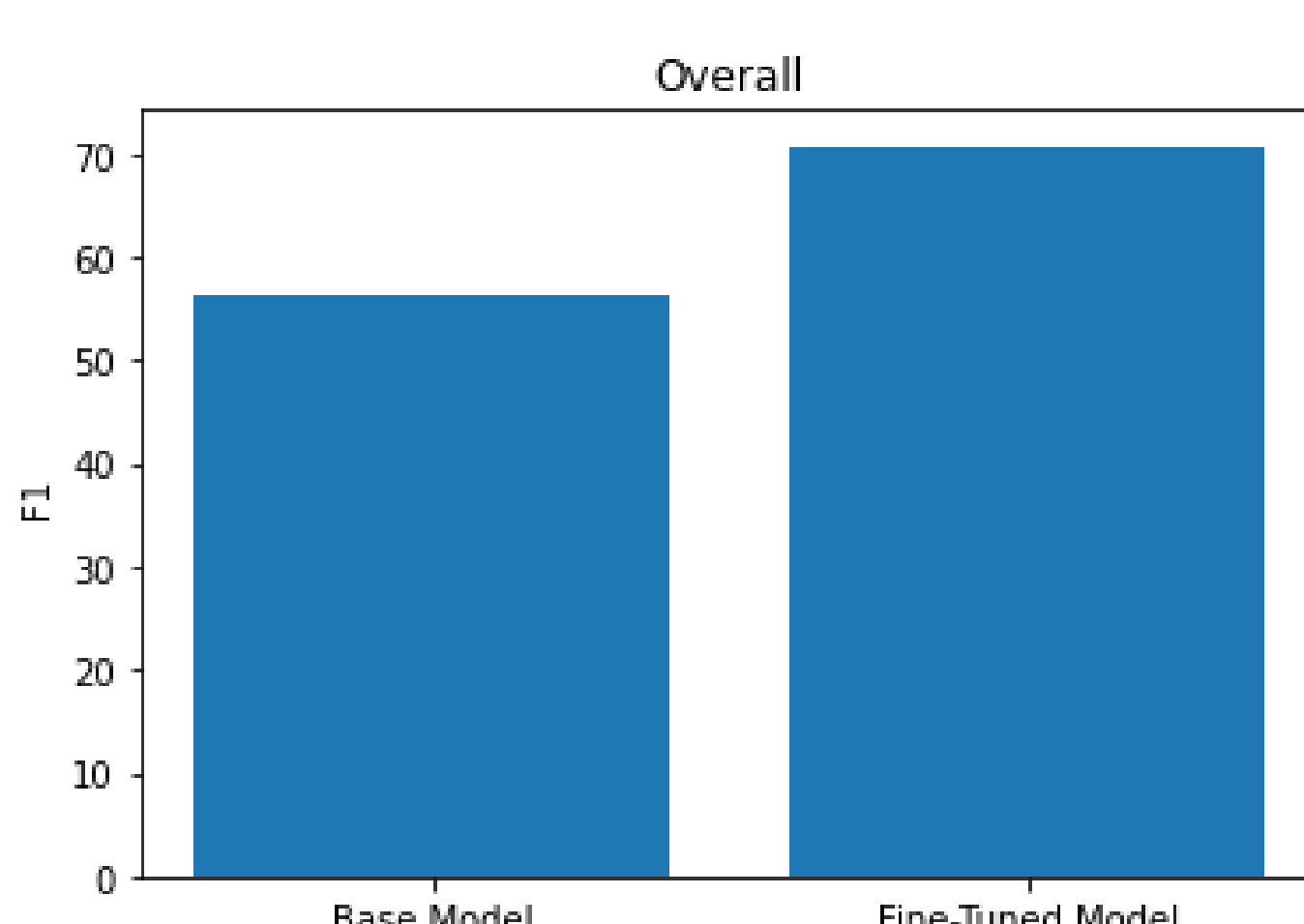
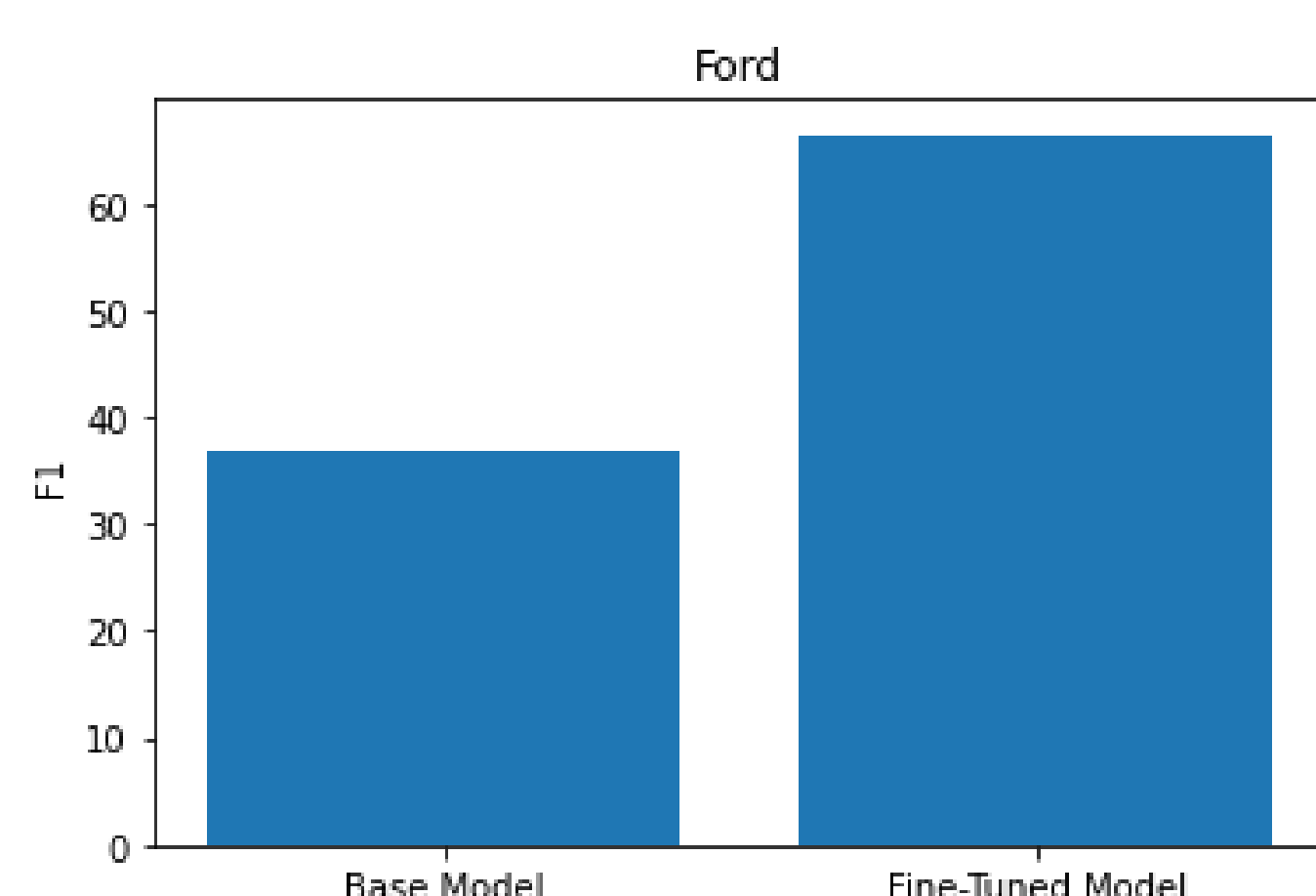


Desired:
SELECT Restraint FROM Child Safety Restraints WHERE Child="small child"

Model Improvement Results and Conclusions

Fine-tuning Results

- Assigned a score to the performance of the model question-response accuracy (F1 score).
- Base model (3 epochs)
 - Ford F1: 36.7
 - Overall F1: 56.4
- Fine-tuned model (3 epochs):
 - Ford F1: 66.4
 - Overall F1: 70.7



Conclusion

- The F1 score of the base model (pre-training data), fine-tuned model, and overall scores shows that the model was considerably improved due to the data produced by our team. In addition, the NL2SQL integration will help the model answer customers' questions better by gathering information from both the main text and the tables inside the Lincoln Aviator Manual.

Future Goals

- Merge the Modular Annotation Tool with BERT-FlowDelta
- SQL Query rejection model integration

Acknowledgements

Thank you, Sarah Rodenbeck, Justin Gould, the Ford Machine Learning group, Dr. Mark Daniel Ward, Ellen Gundlach, and Maggie Betz for providing the resources and support for the successful completion of our project.