



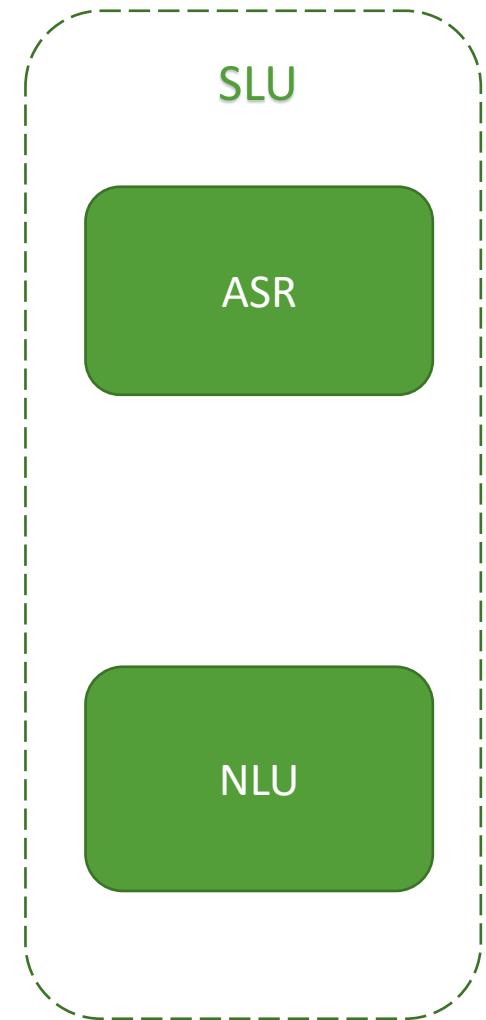
Fast and Scalable Expansion of Natural Language Understanding Functionality for Intelligent Agents

Anuj Goyal (joint work with Angeliki Metallinou & Spyros Matsoukas)

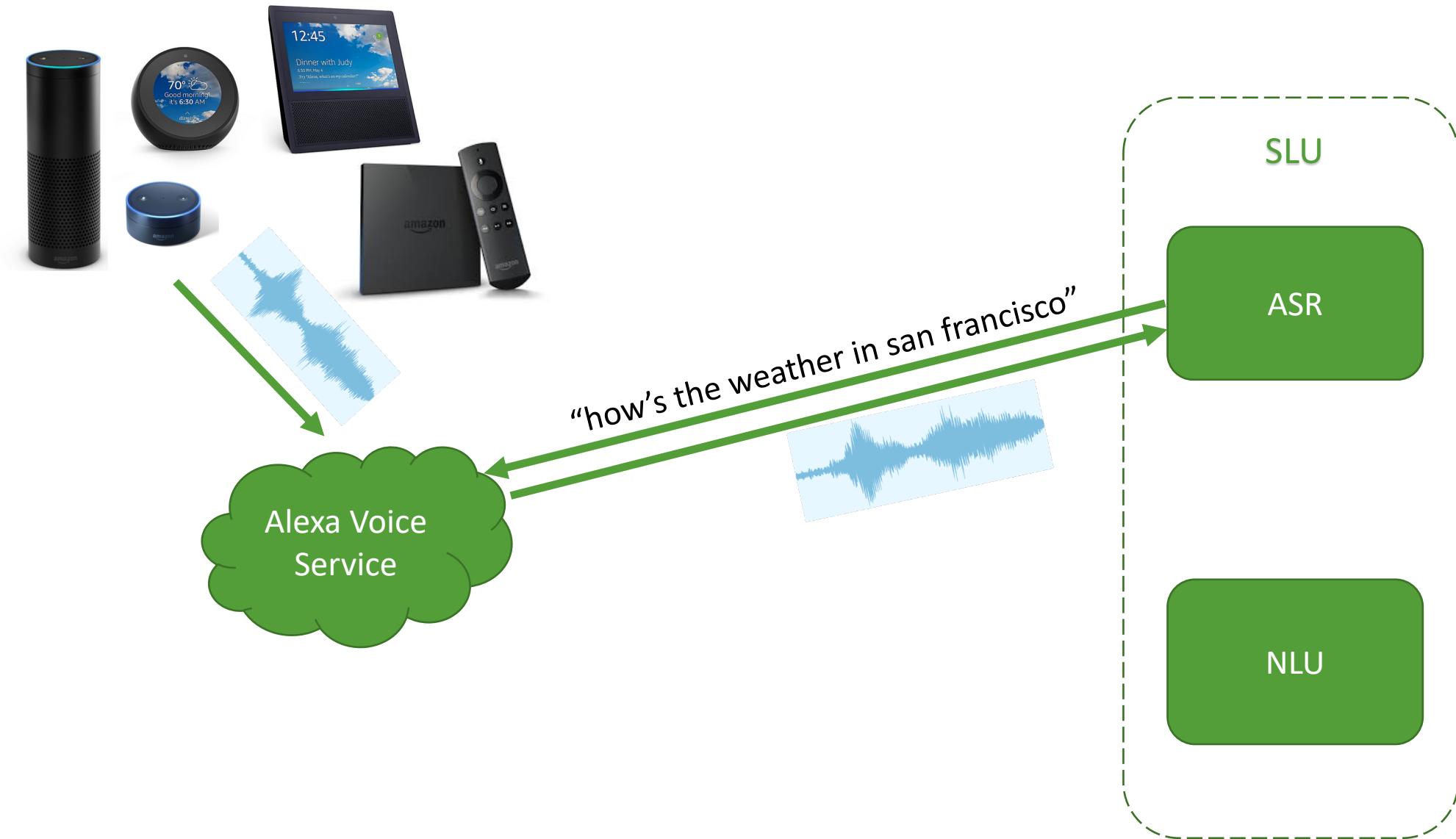
@anujgoya, @ametalli, @matsouka

Applied Scientist | Alexa

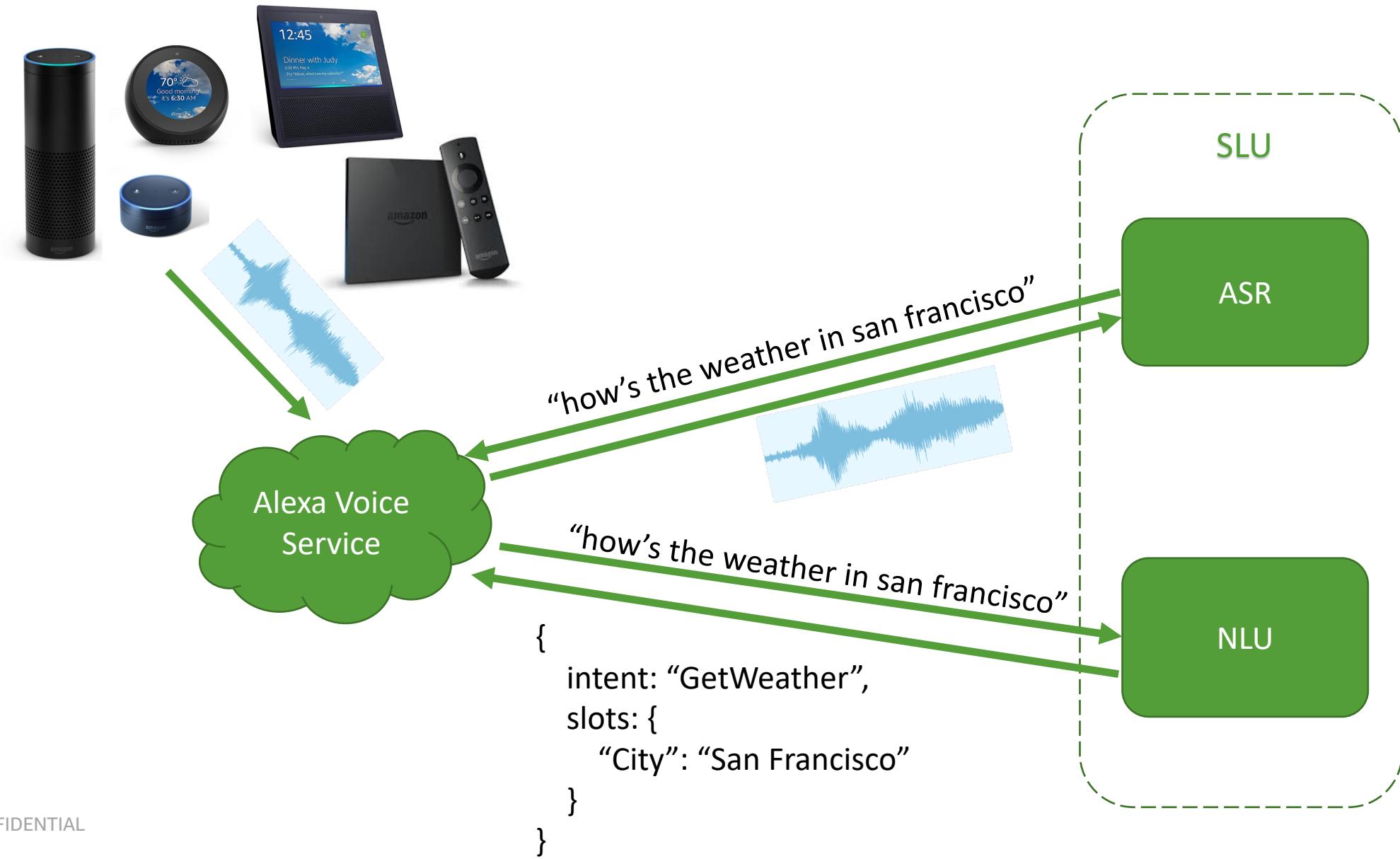
Where does NLU fit in?



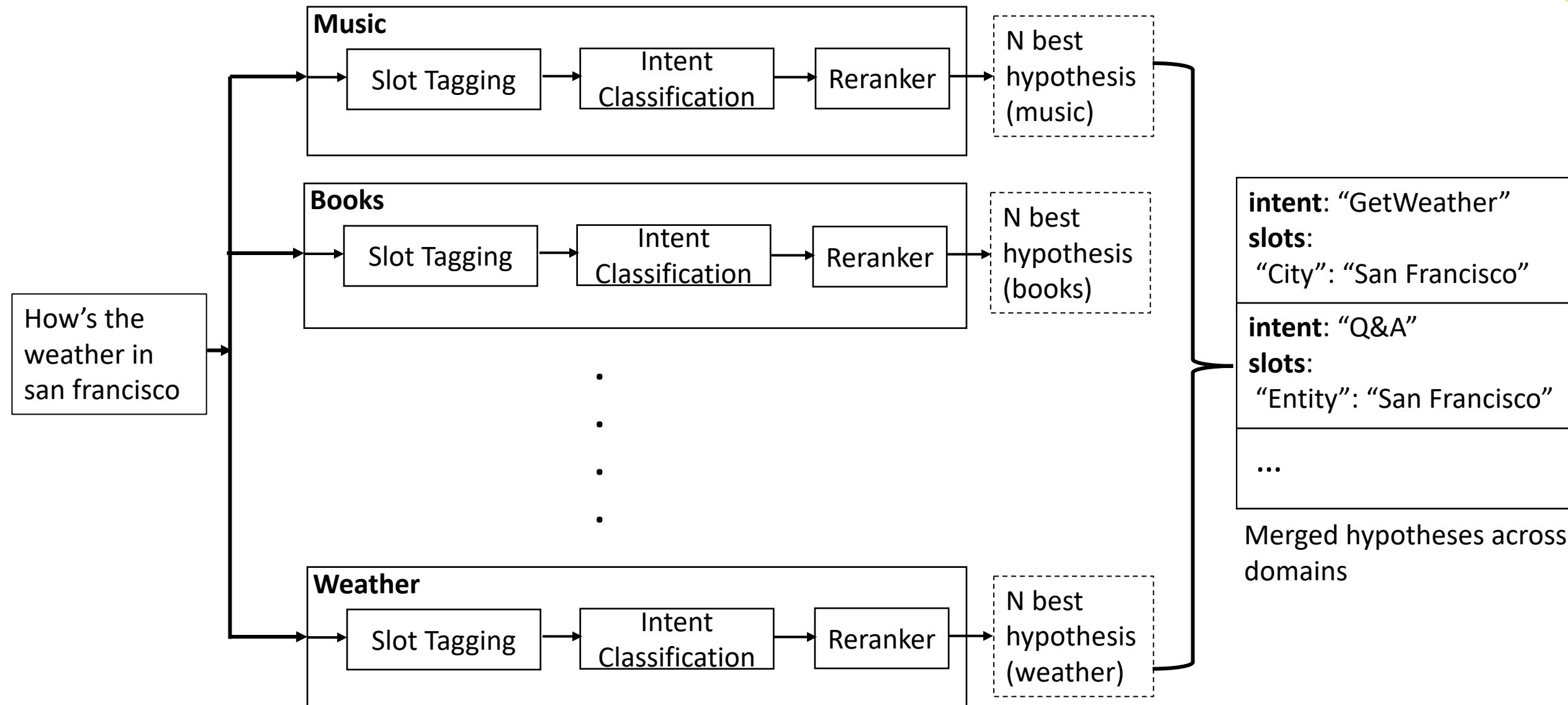
Where does NLU fit in?



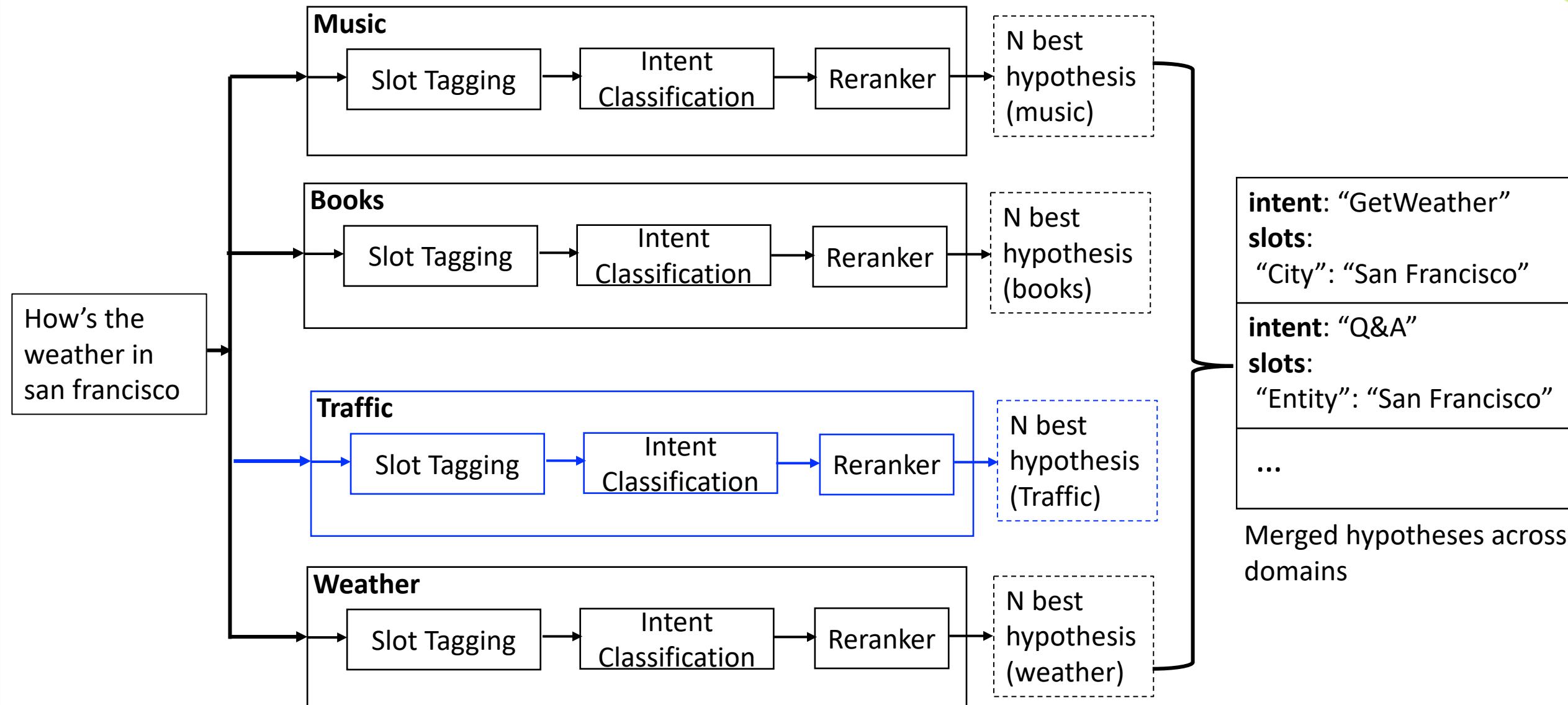
Where does NLU fit in?



Alexa Natural Language Understanding (NLU)



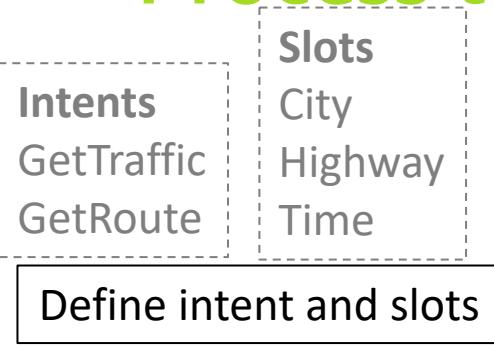
Alexa Natural Language Understanding (NLU)



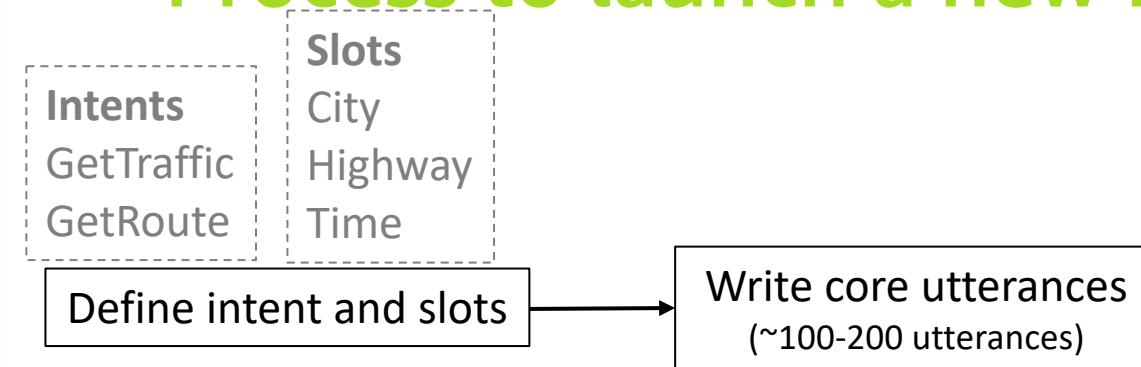
NLU – Built-in and External domains

- Built-in domains
 - Music, Books, Weather, Recipes etc.
- Custom developer domains (Skills)
 - Uber, Jeopardy, BBC, Short Bedtime Story, 7-Minute Workout etc.

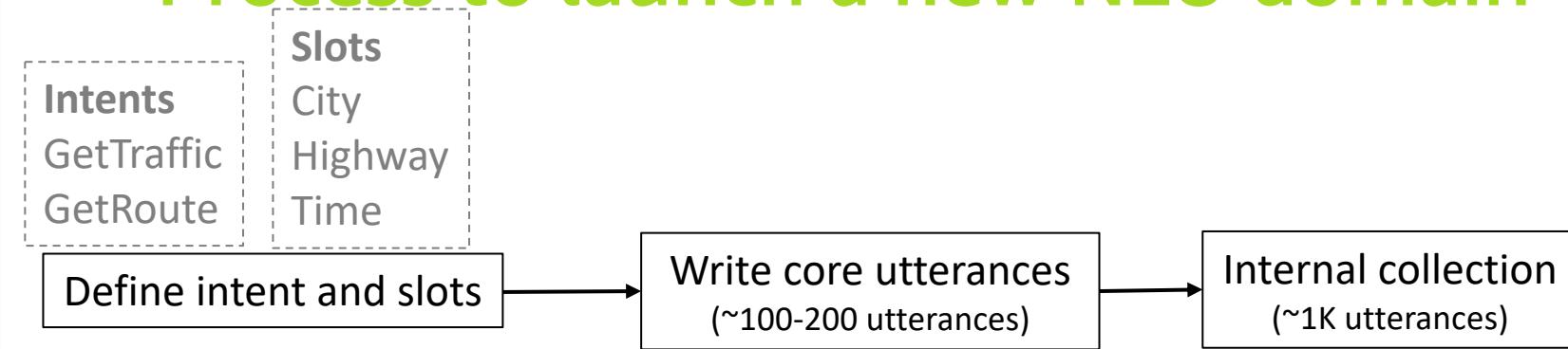
Process to launch a new NLU domain



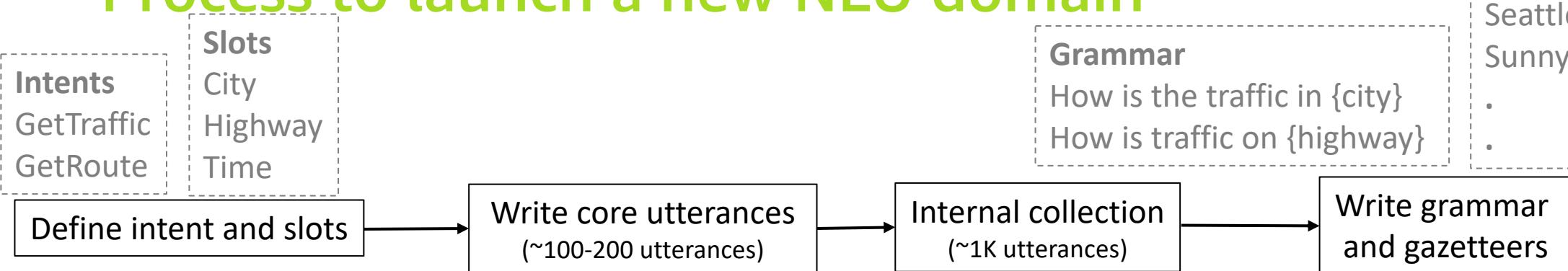
Process to launch a new NLU domain



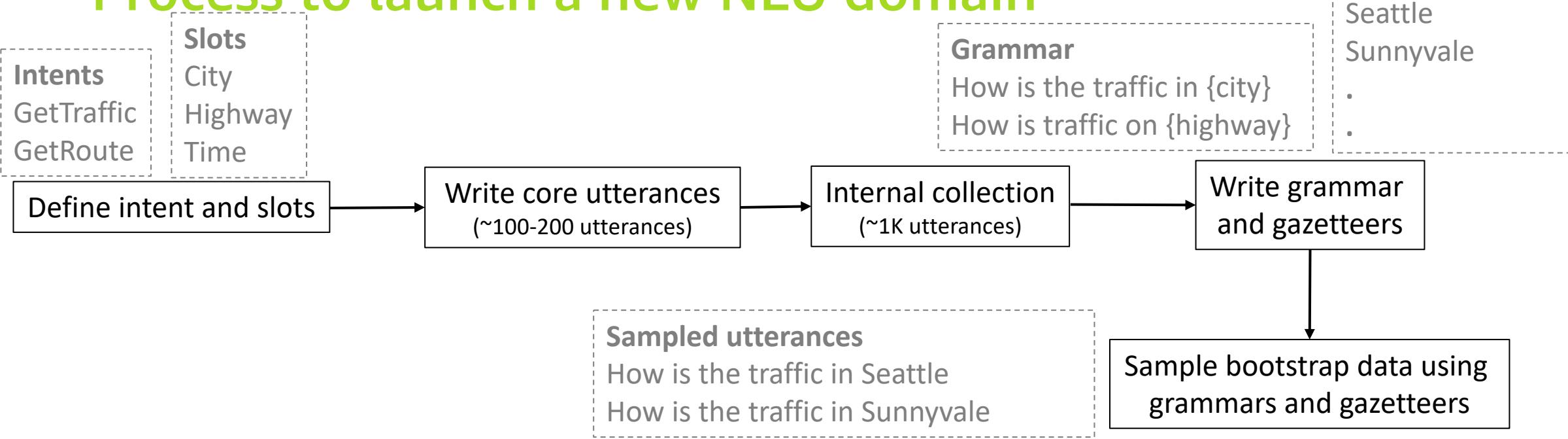
Process to launch a new NLU domain



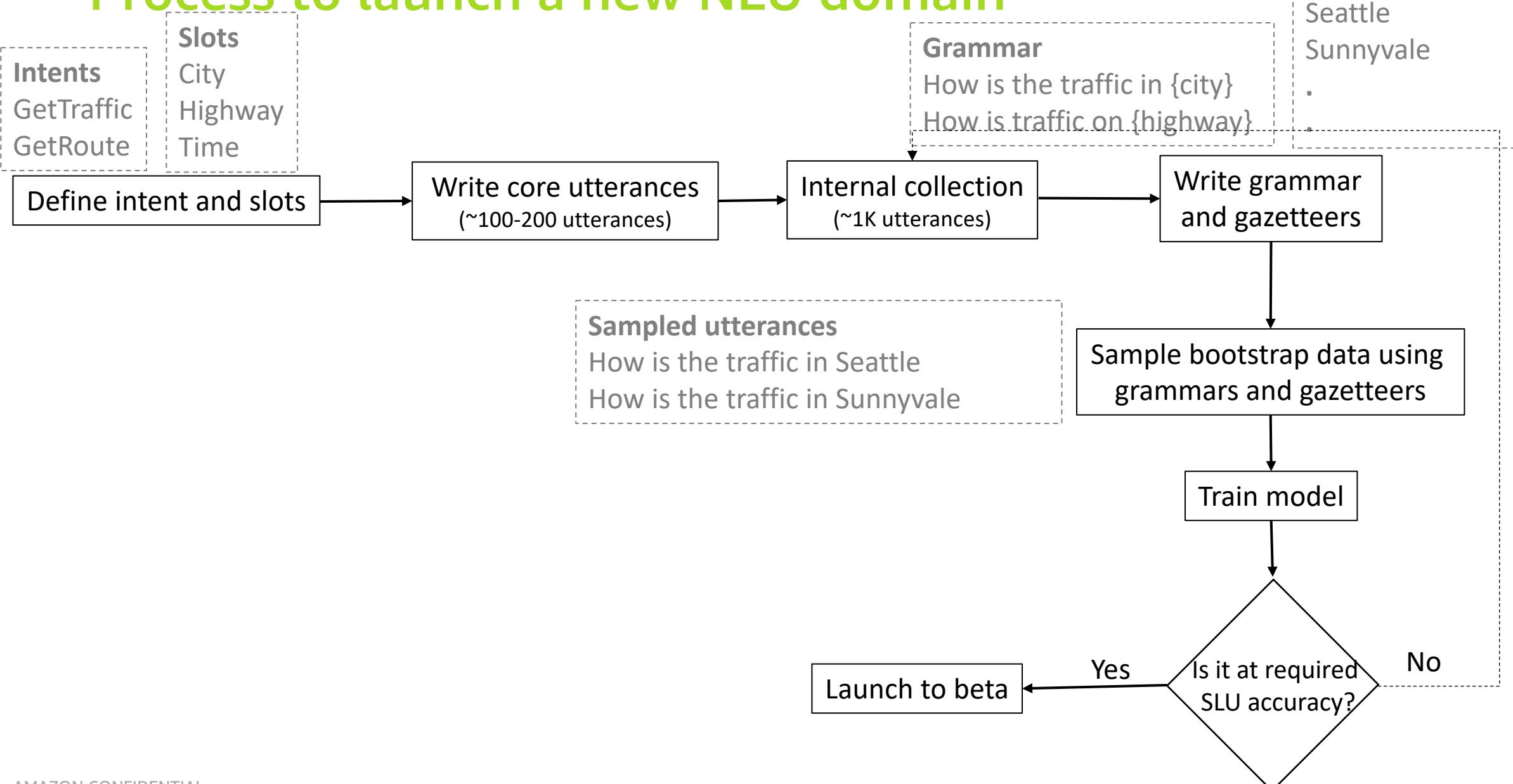
Process to launch a new NLU domain



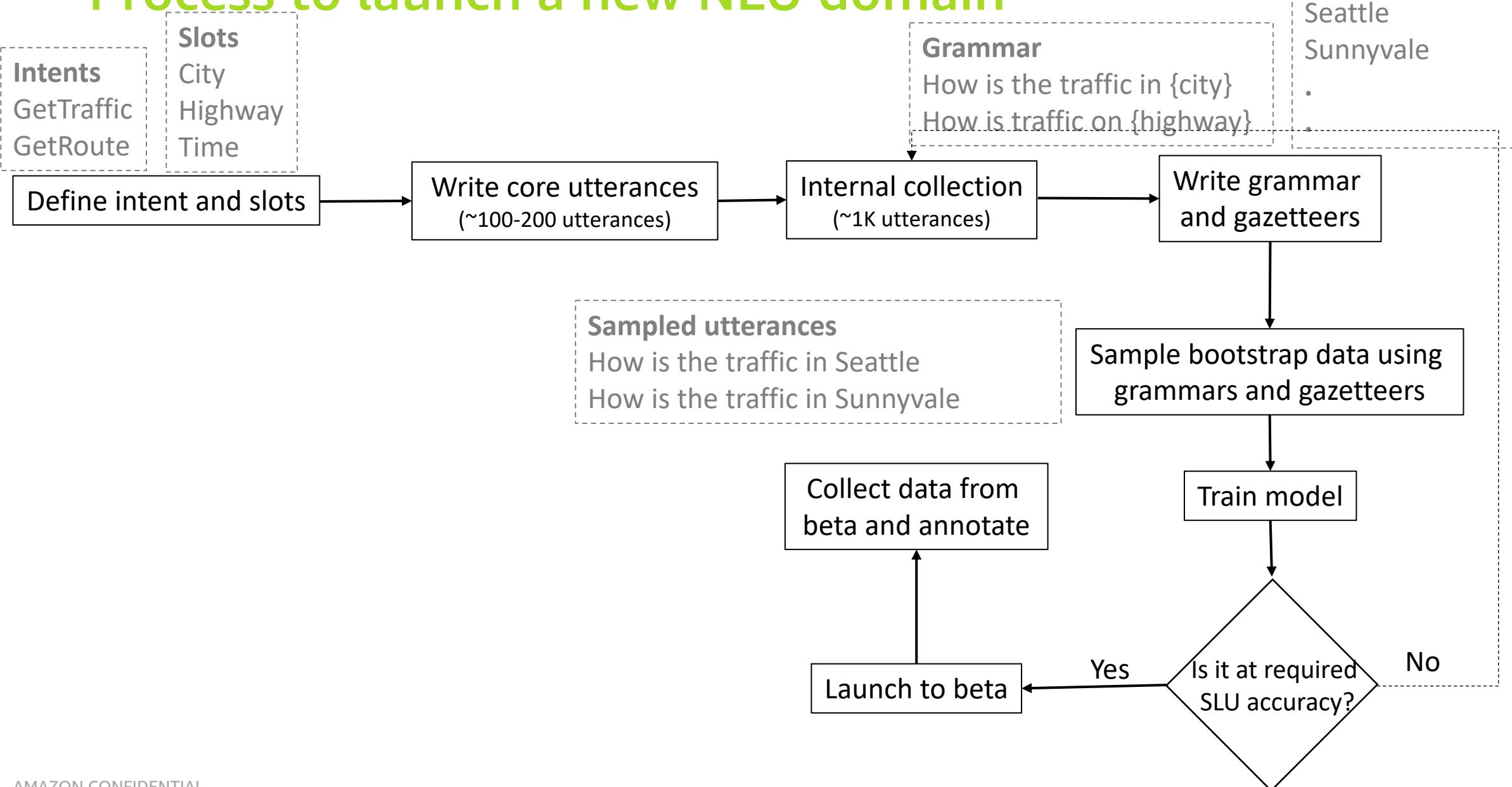
Process to launch a new NLU domain



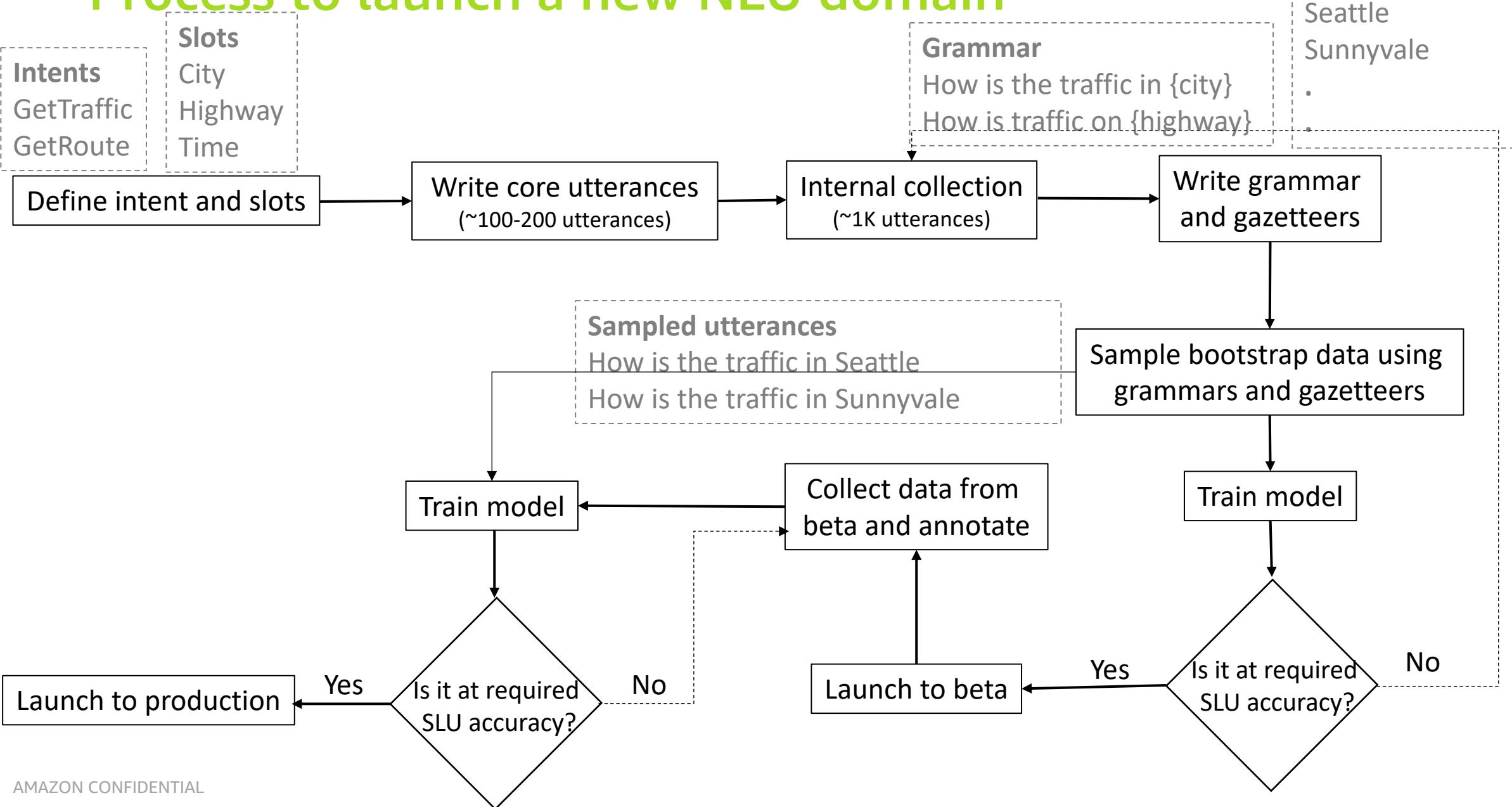
Process to launch a new NLU domain



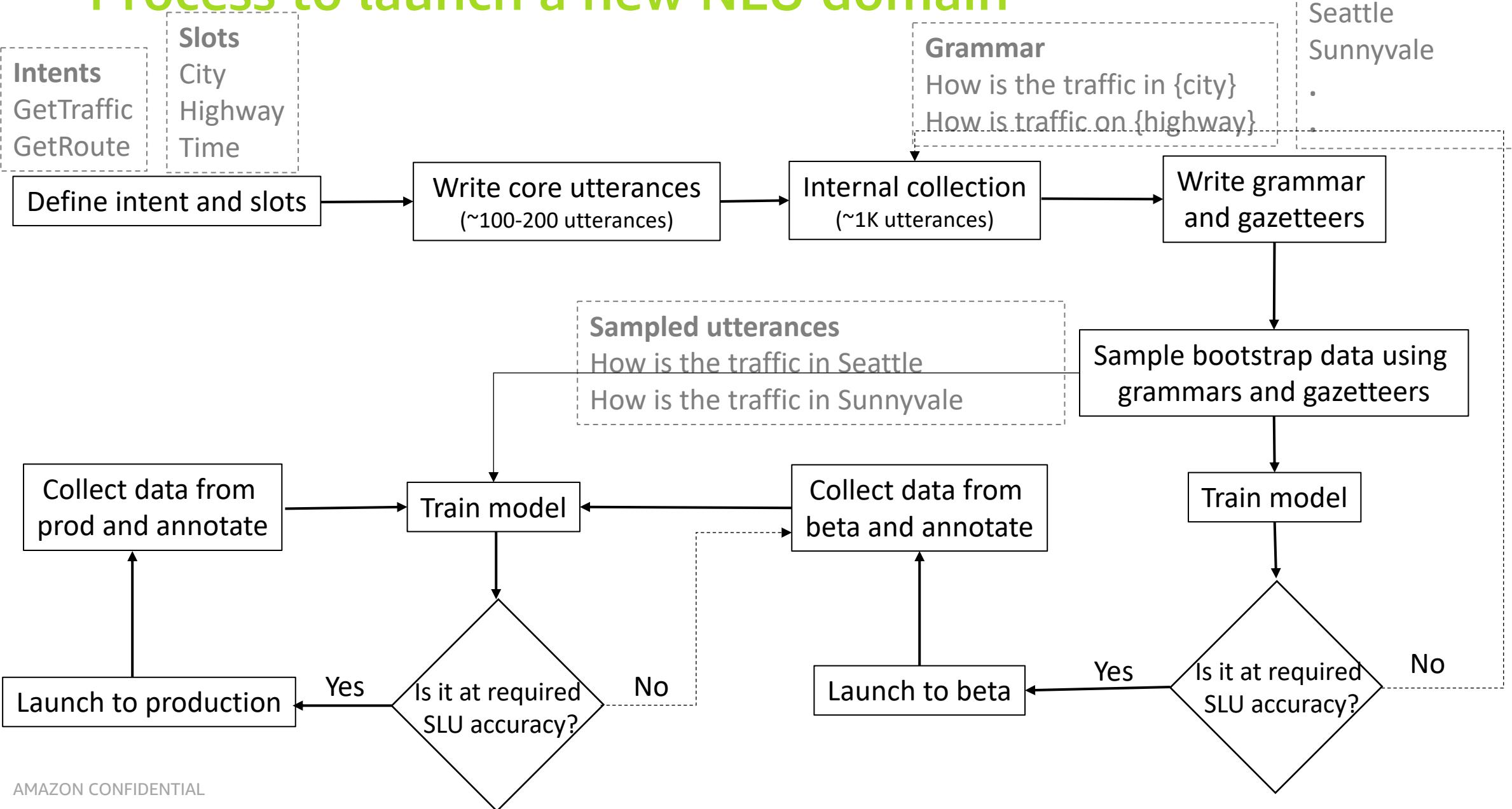
Process to launch a new NLU domain



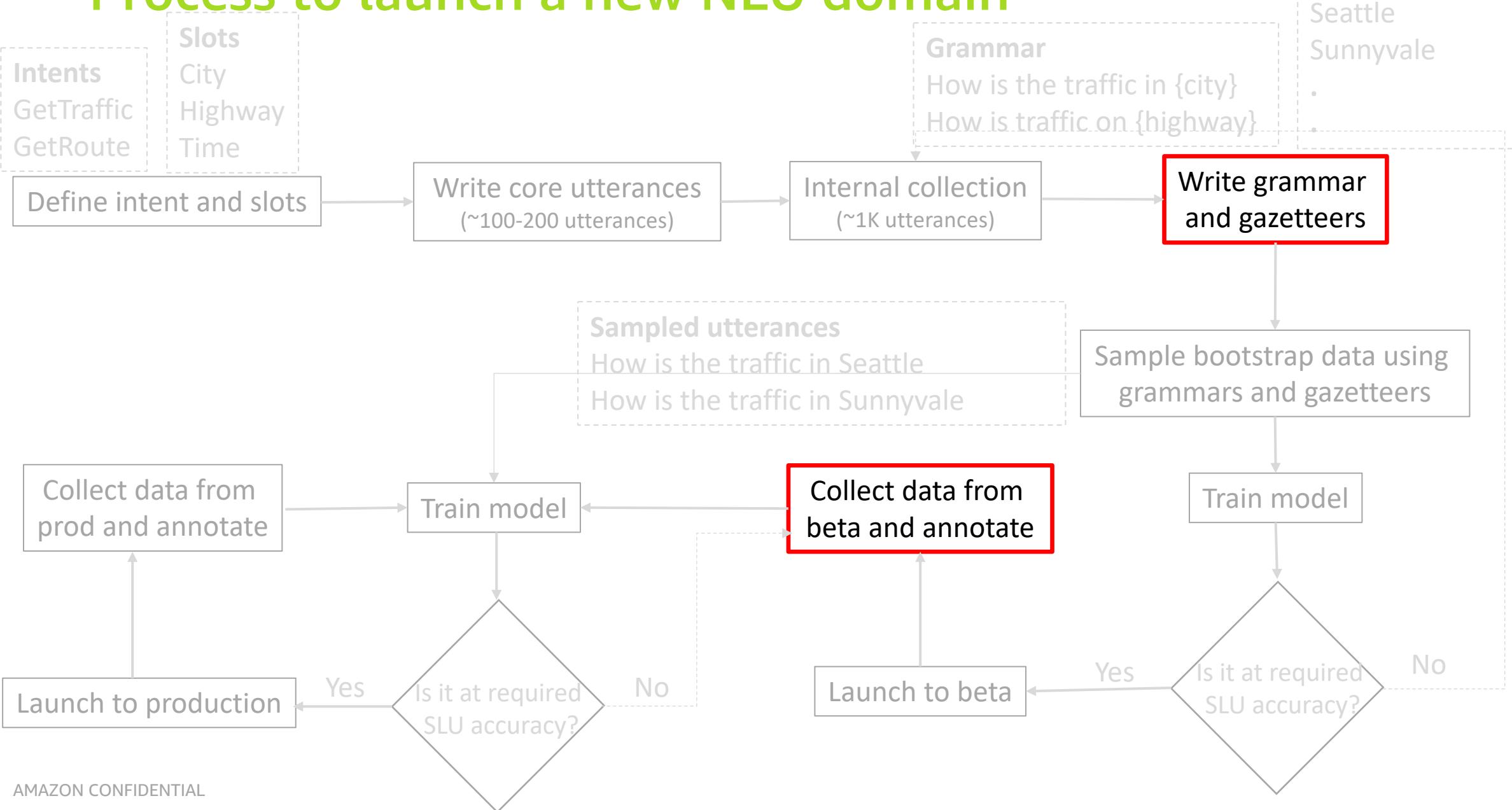
Process to launch a new NLU domain



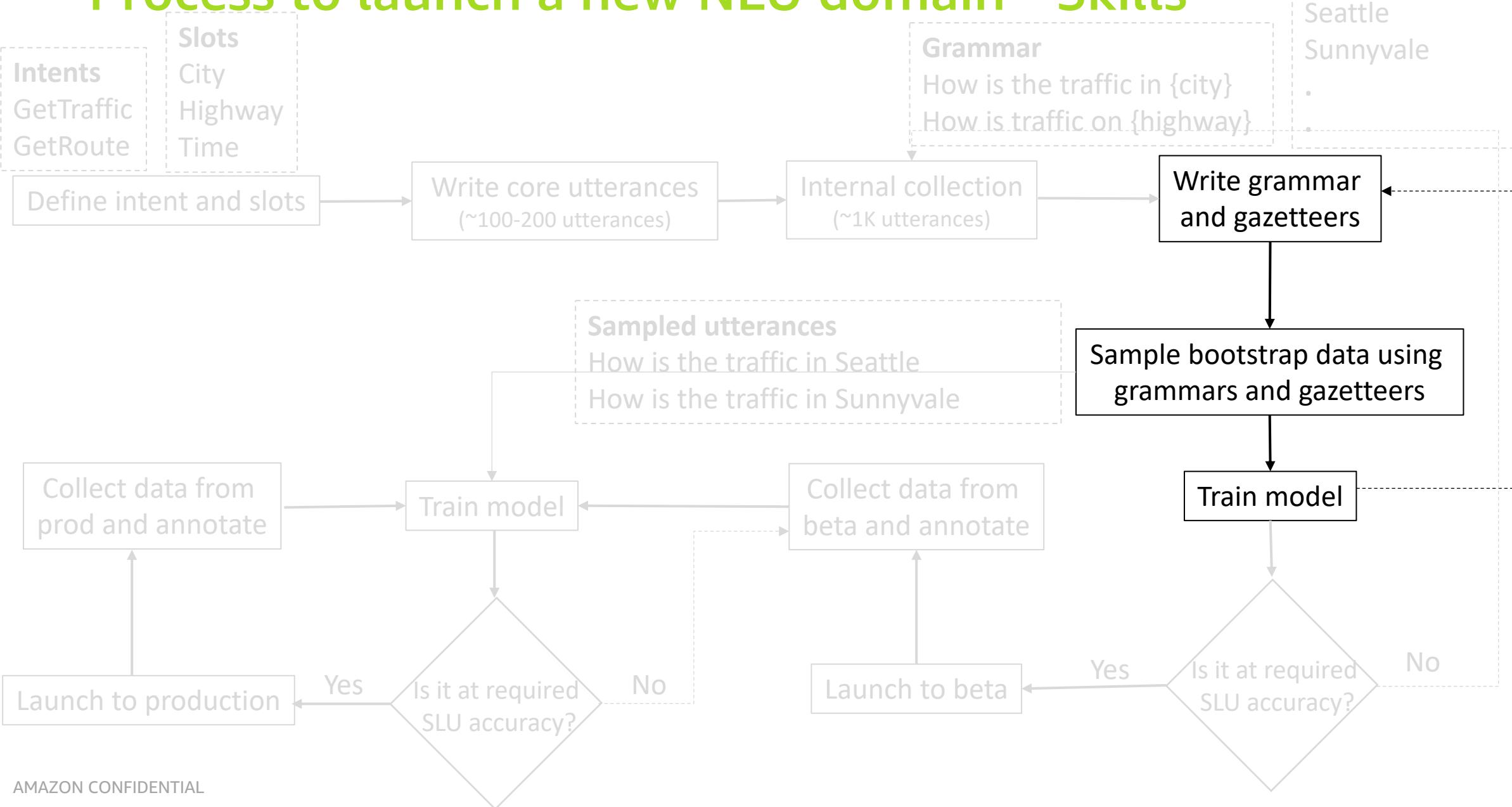
Process to launch a new NLU domain



Process to launch a new NLU domain



Process to launch a new NLU domain - Skills



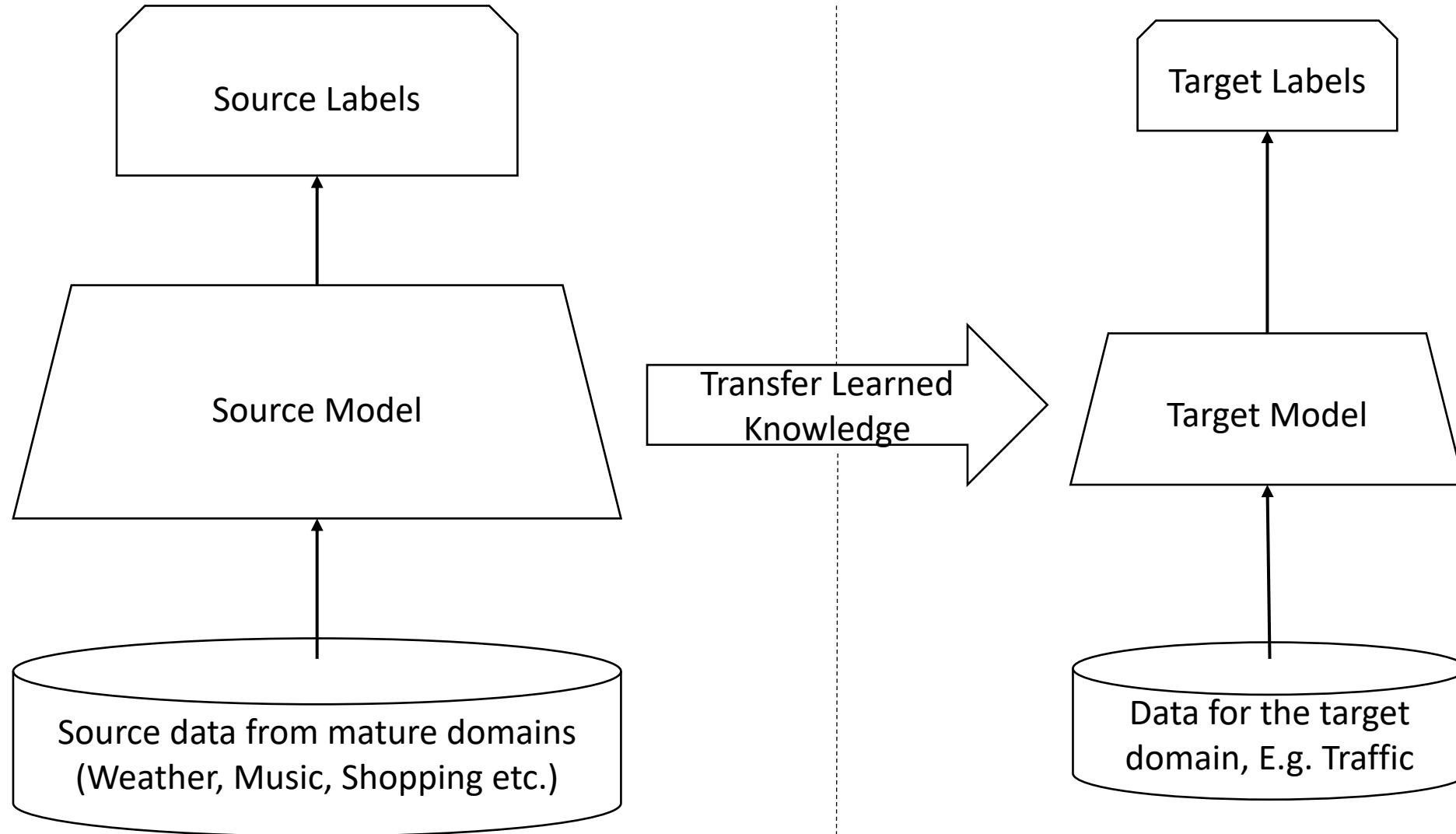
Challenges

- Annotated training data is limited during a domain bootstrapping.
- Collecting annotated data is a time and resource intensive process for both internal & external developers.
- Some external developers may not have expertise & resources to provide good data.

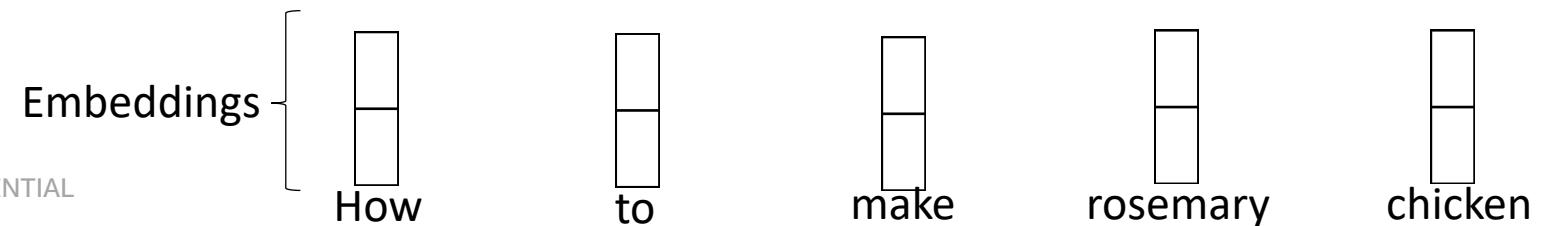
Some observations

- Already have millions of annotated utterances from mature domains
- There are many commonalities among the mature domains and a new domain
 - Slot similarities, e.g. “City” in Traffic and Weather domain
 - Utterance pattern similarities
 - “how's the weather in Seattle” (Weather domain)
 - “how's the traffic in Seattle” (Traffic domain)

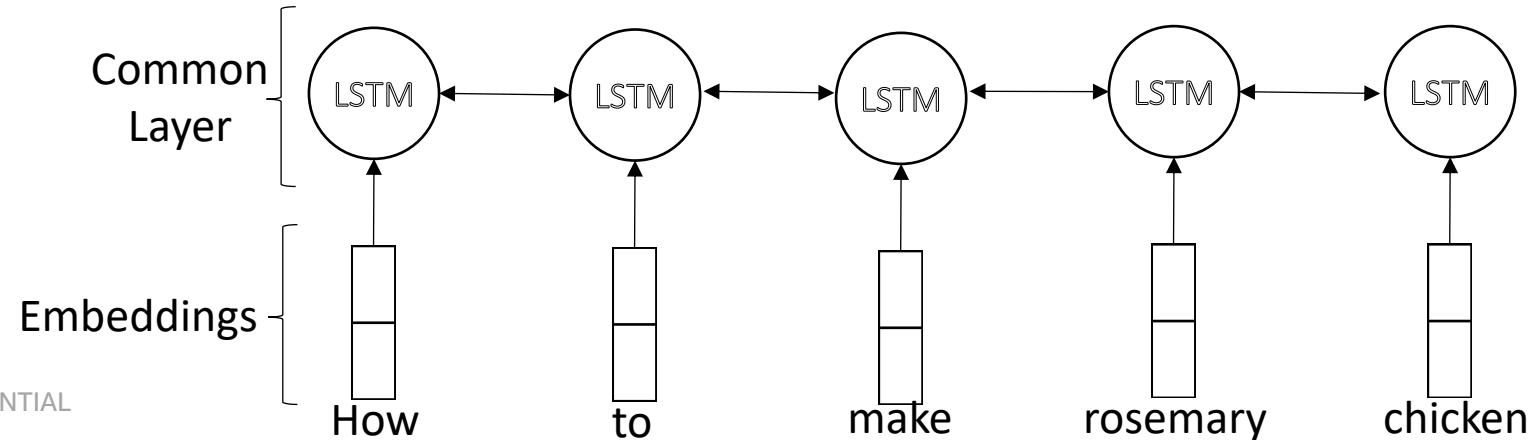
Transfer Learning



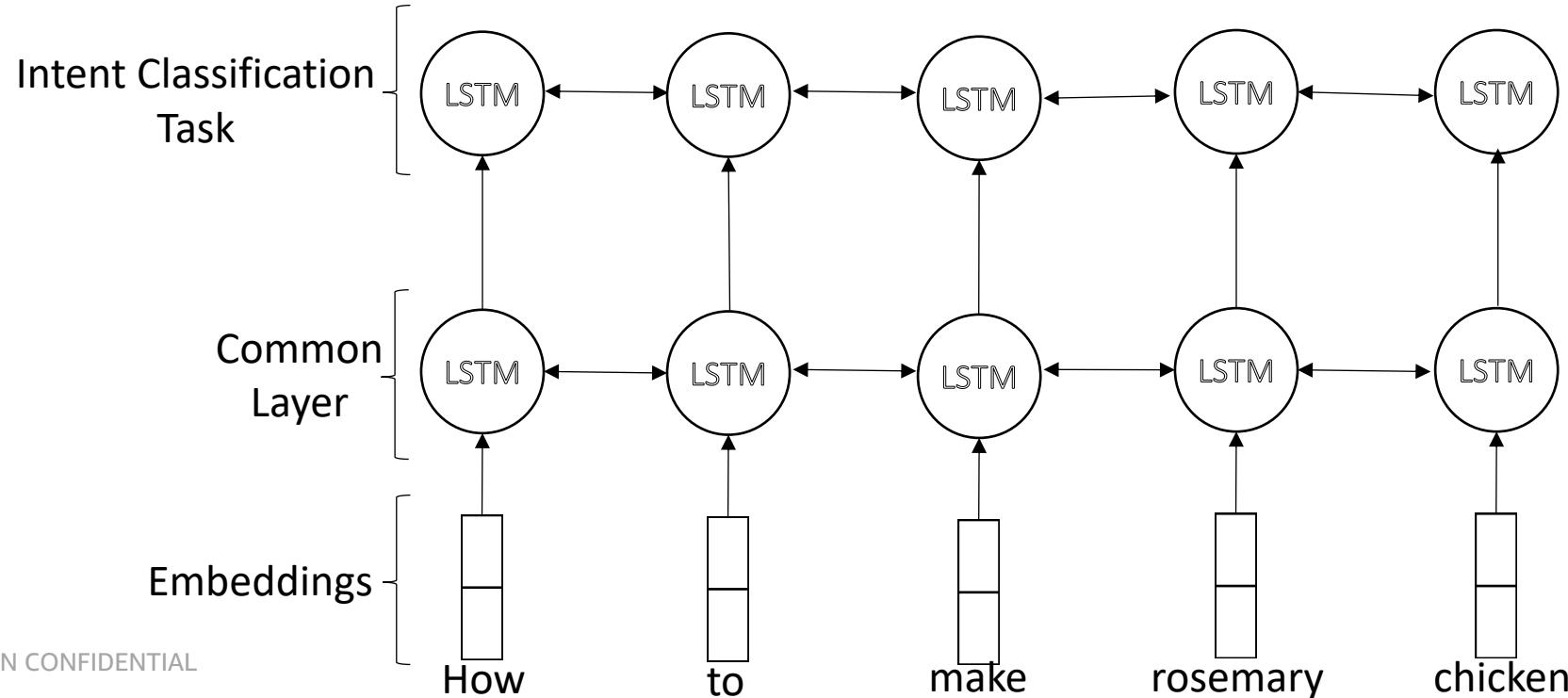
DNN model for NLU



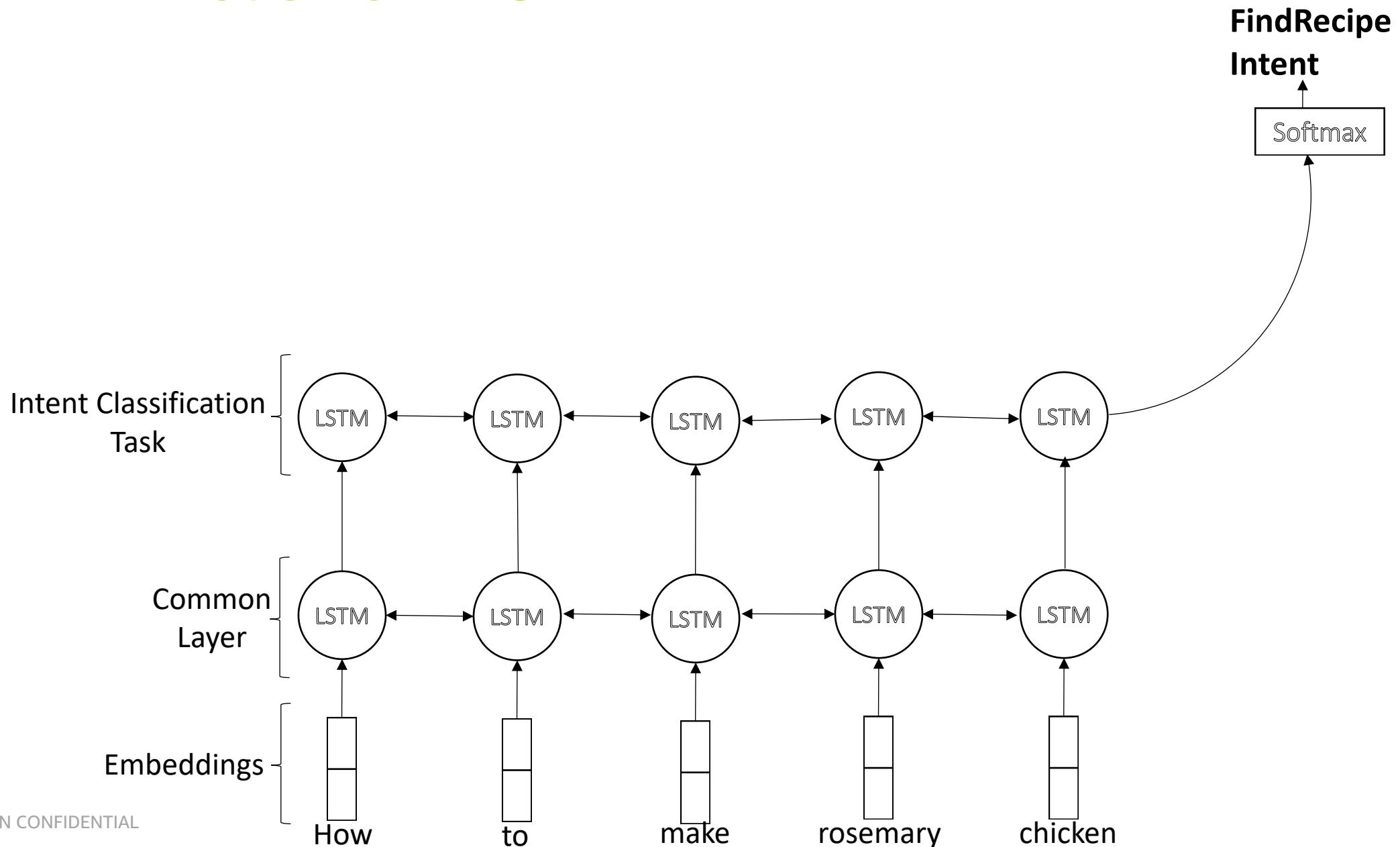
DNN model for NLU



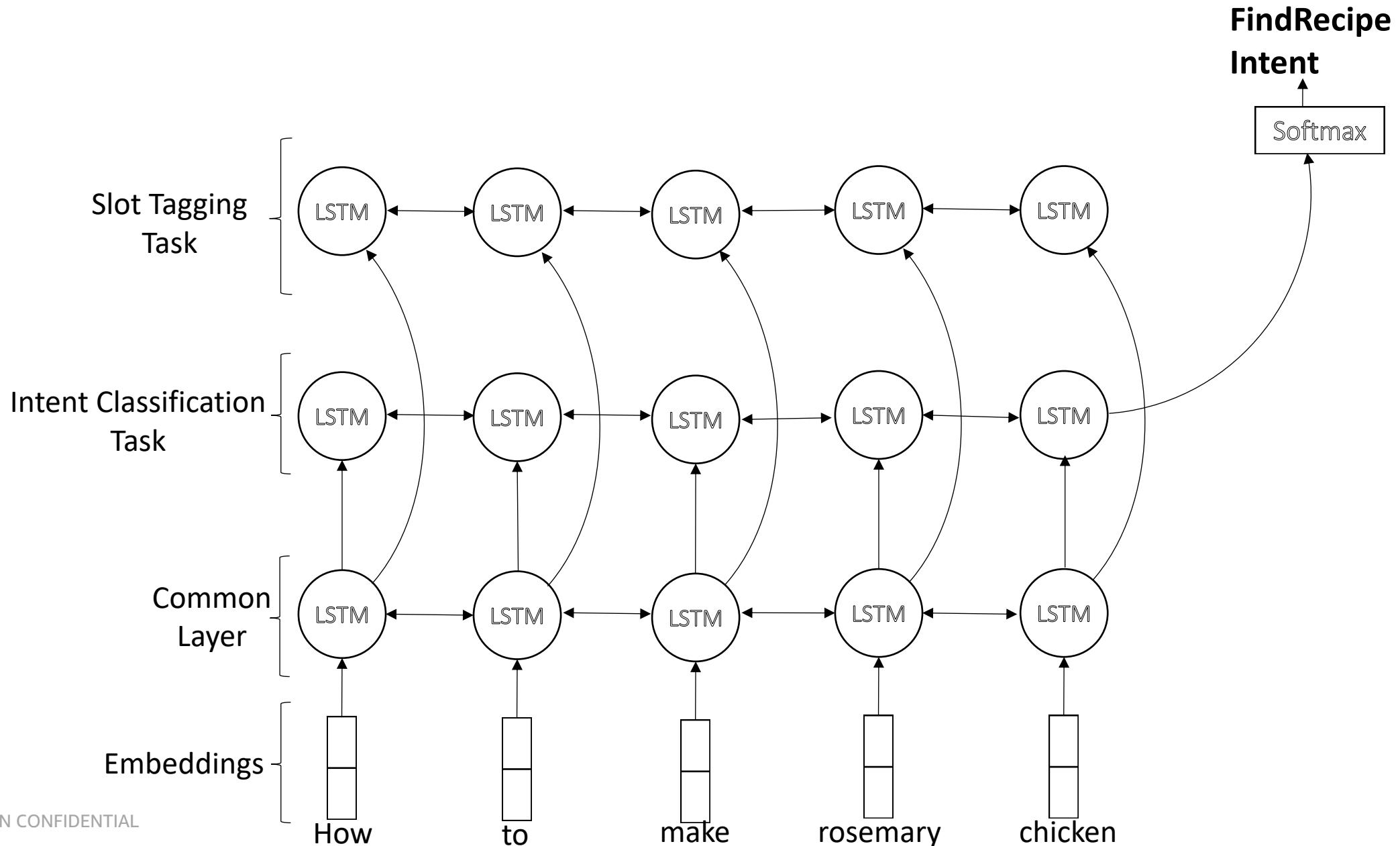
DNN model for NLU

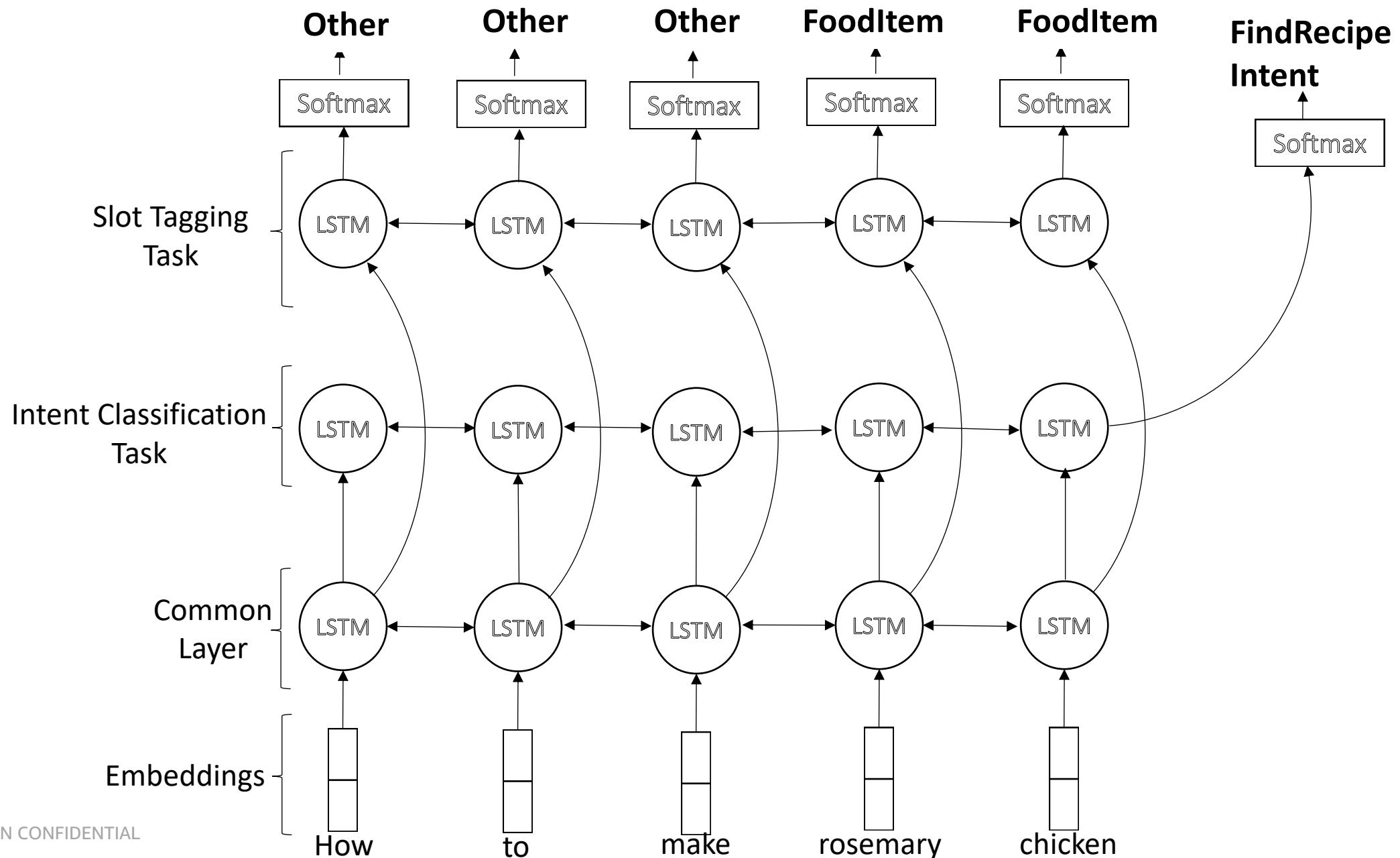


DNN model for NLU



DNN model for NLU



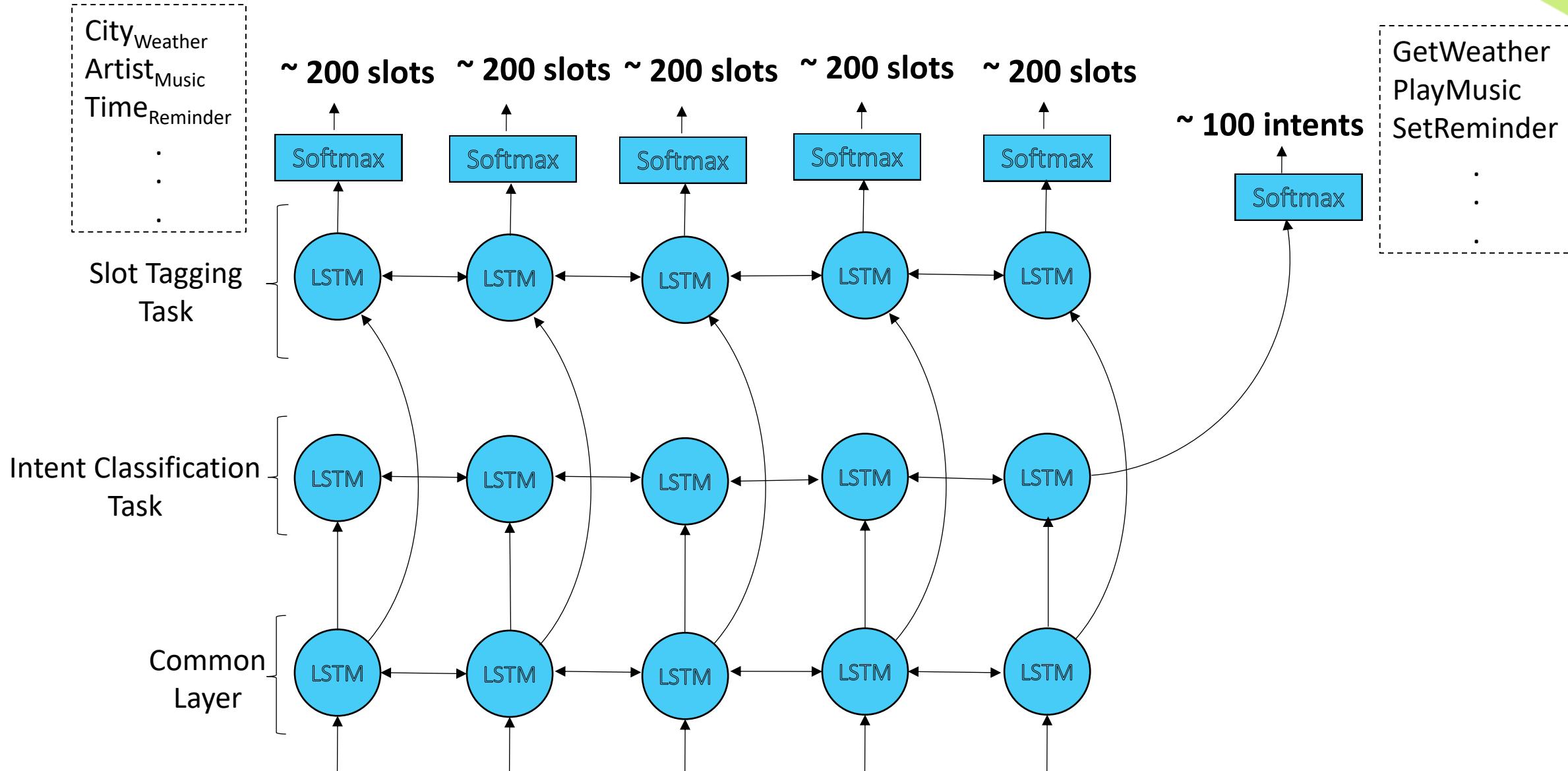




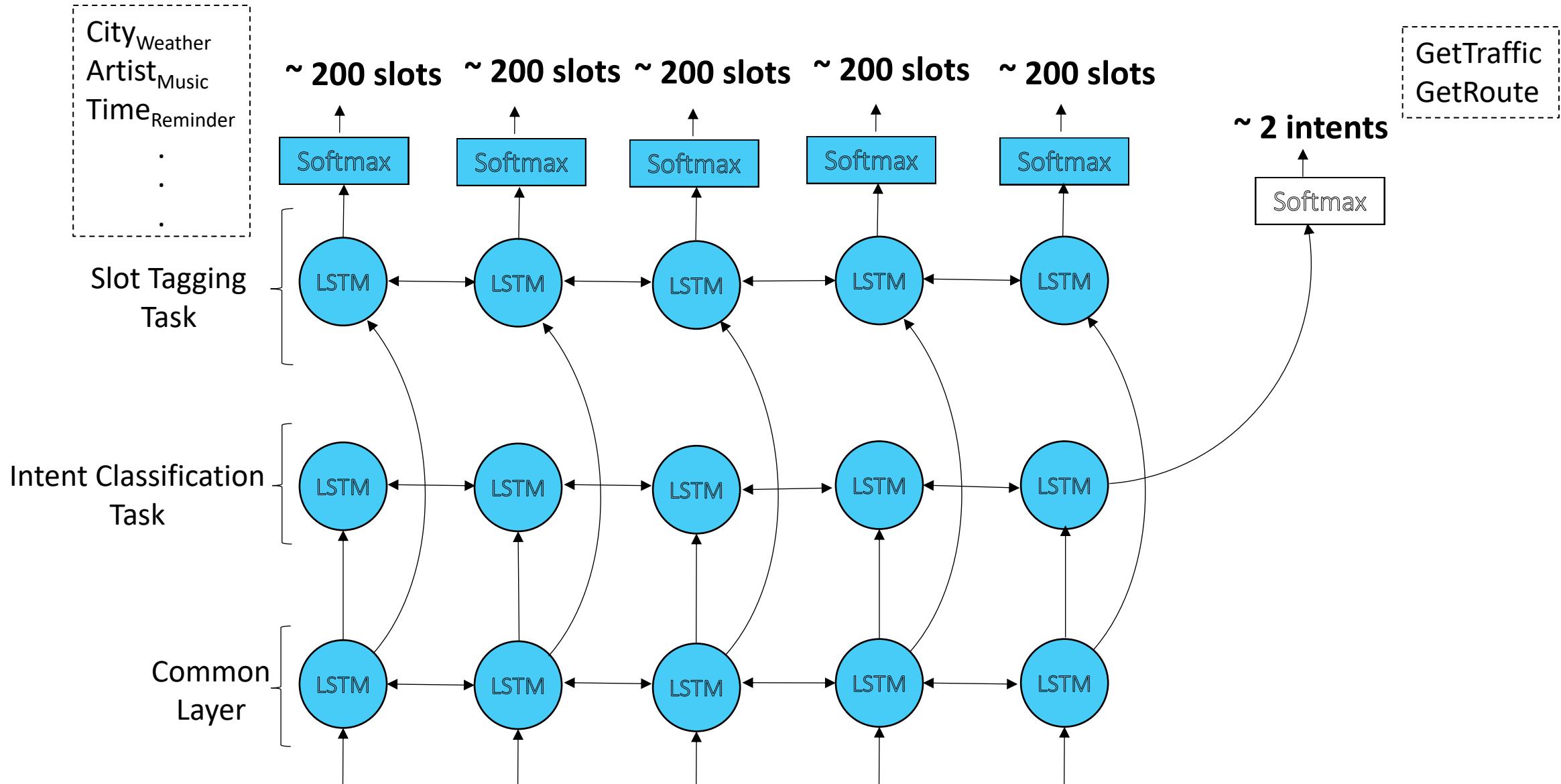
Pretraining

- Combine data from ~25 mature domains
- ~ 4 M training utterances
- ~ 100 Intents
- ~ 200 Slots

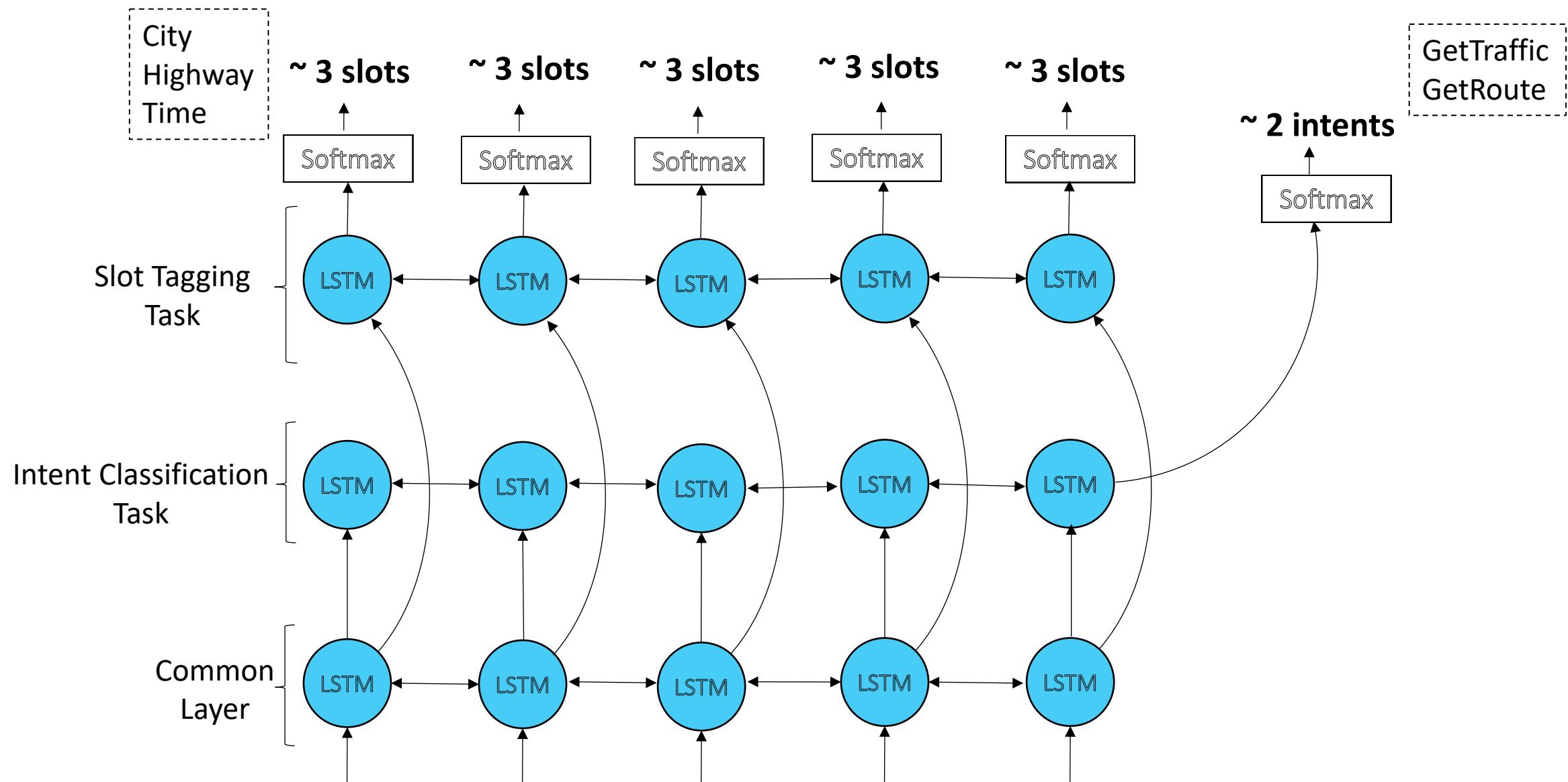
Fine-tuning for the target domain



Fine-tuning for the target domain



Fine-tuning for the target domain





Results

- Skills
- Built-in domains

Results for Skills (Custom developer domains)

Algorithm	F1 _{Intent}	F1 _{Slot}	SER*
CRF/MaxEnt Baseline	96.6	91.5	9.2
DNN (without transfer learning)	95.9	92.9	9.2
DNN (with transfer learning)	97.2	93.0	7.9

Median for ~200 Skills

$$*SER = \frac{\# \text{ utterances with at least one IC or ST Error}}{\# \text{ total utterances}}$$

Results for Skills (Custom developer domains)

Algorithm	F1 _{Intent}	F1 _{Slot}	SER*
CRF/MaxEnt Baseline	96.6	91.5	9.2
DNN (without transfer learning)	95.9	92.9	9.2
DNN (with transfer learning)	97.2	93.0	7.9

Median for ~200 Skills

$$*SER = \frac{\# \text{ utterances with at least one IC or ST Error}}{\# \text{ total utterances}}$$

Results for Skills (Custom developer domains)

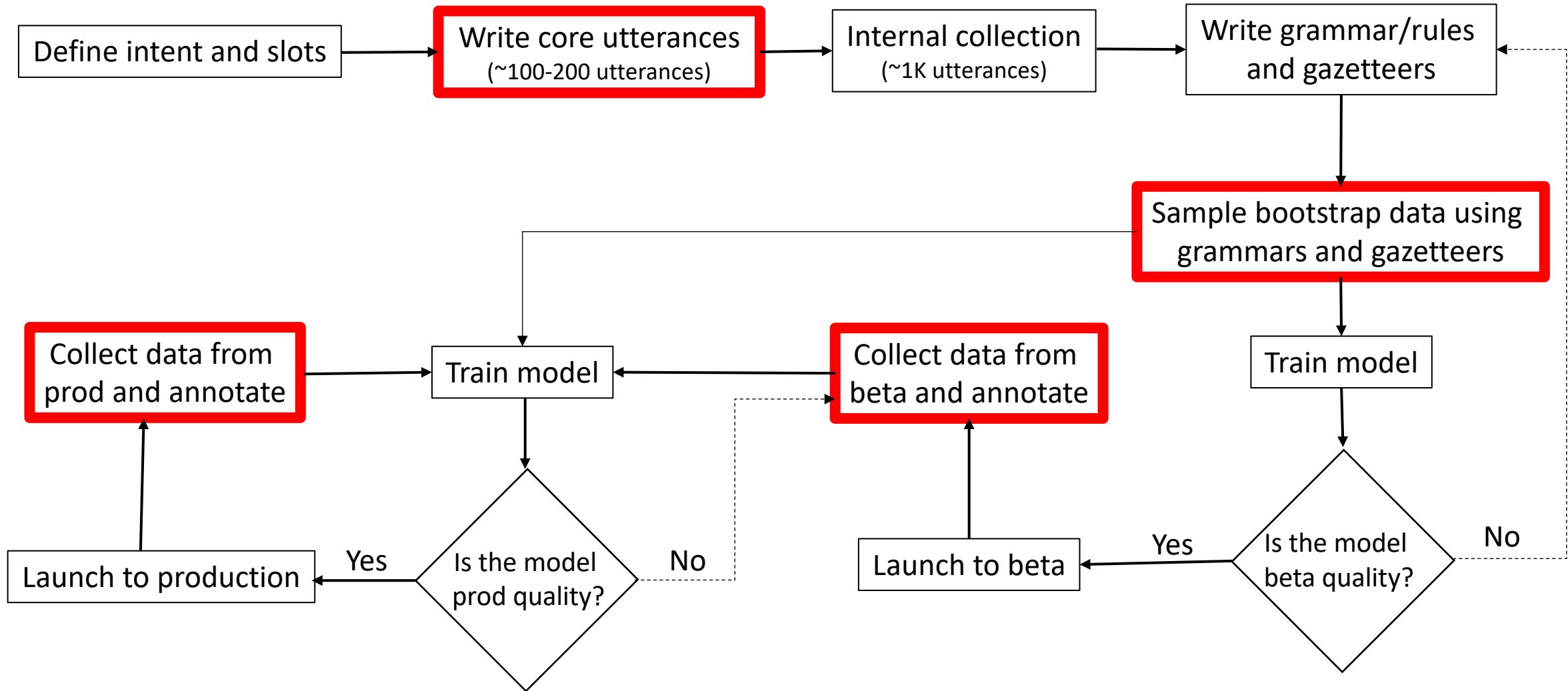
Algorithm	F1 _{Intent}	F1 _{Slot}	SER*
CRF/MaxEnt Baseline	96.6	91.5	9.2
DNN (without transfer learning)	95.9	92.9	9.2
DNN (with transfer learning)	97.2	93.0	7.9

+ 14%

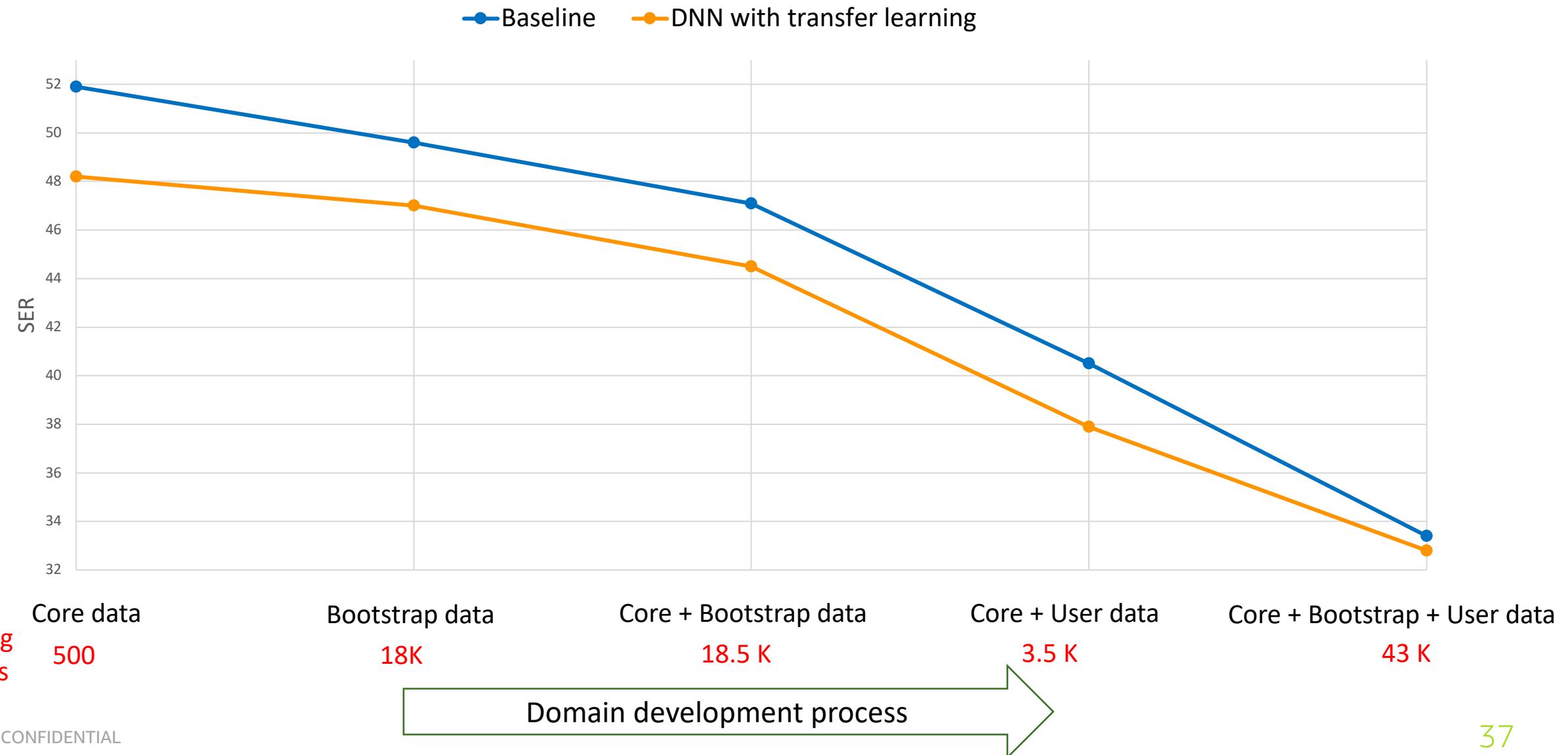
Median for ~200 Skills

$$*SER = \frac{\# \text{ utterances with at least one IC or ST Error}}{\# \text{ total utterances}}$$

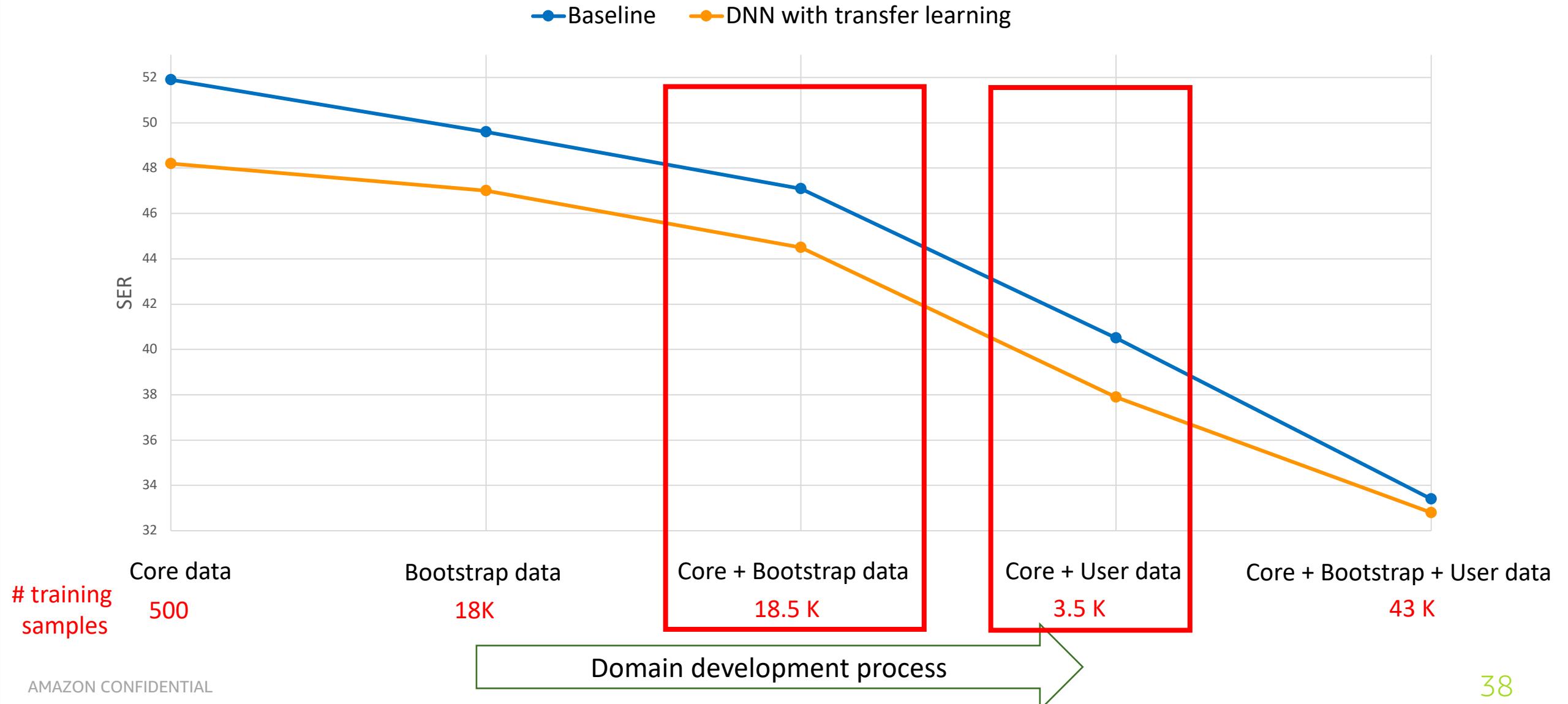
Experimental setup for built-in domains



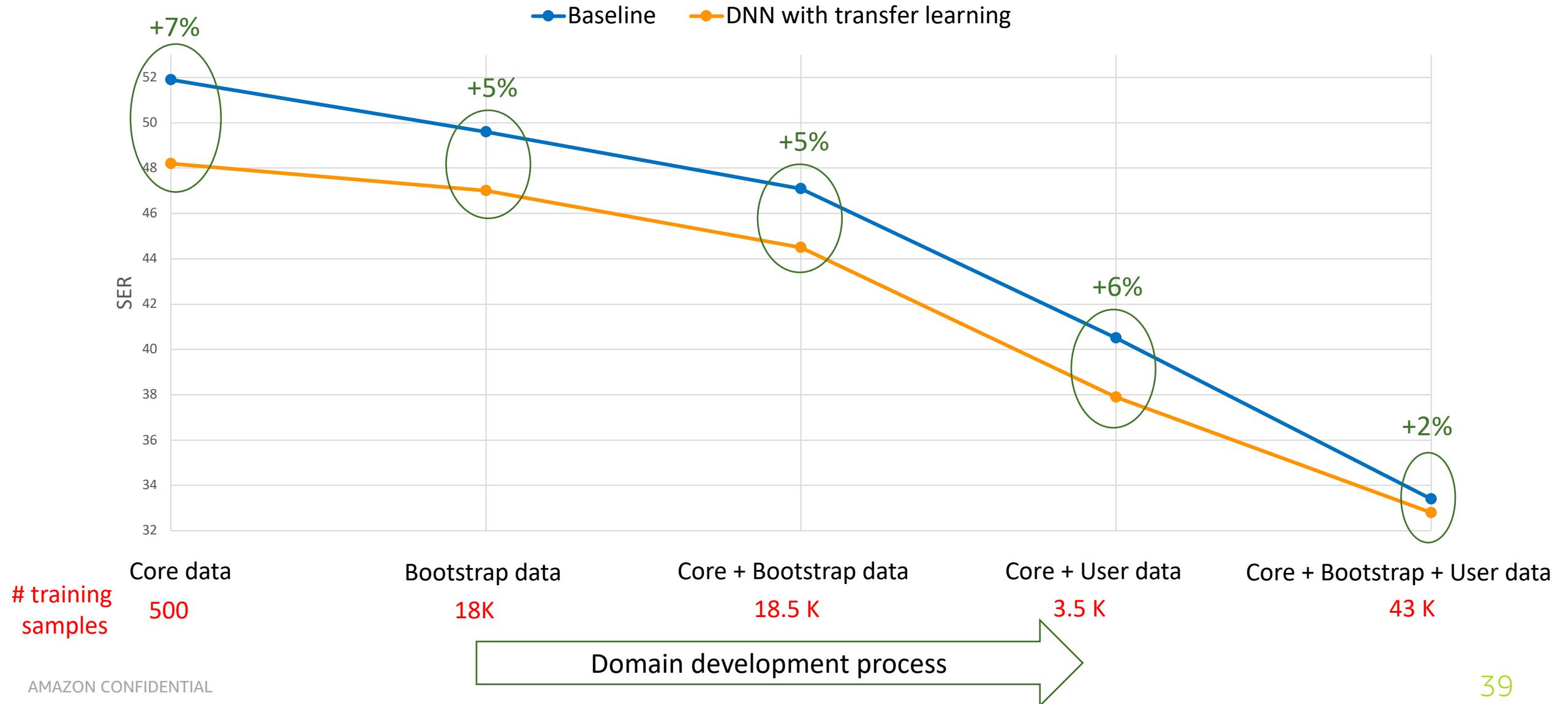
Results for built-in domains (Recipes)



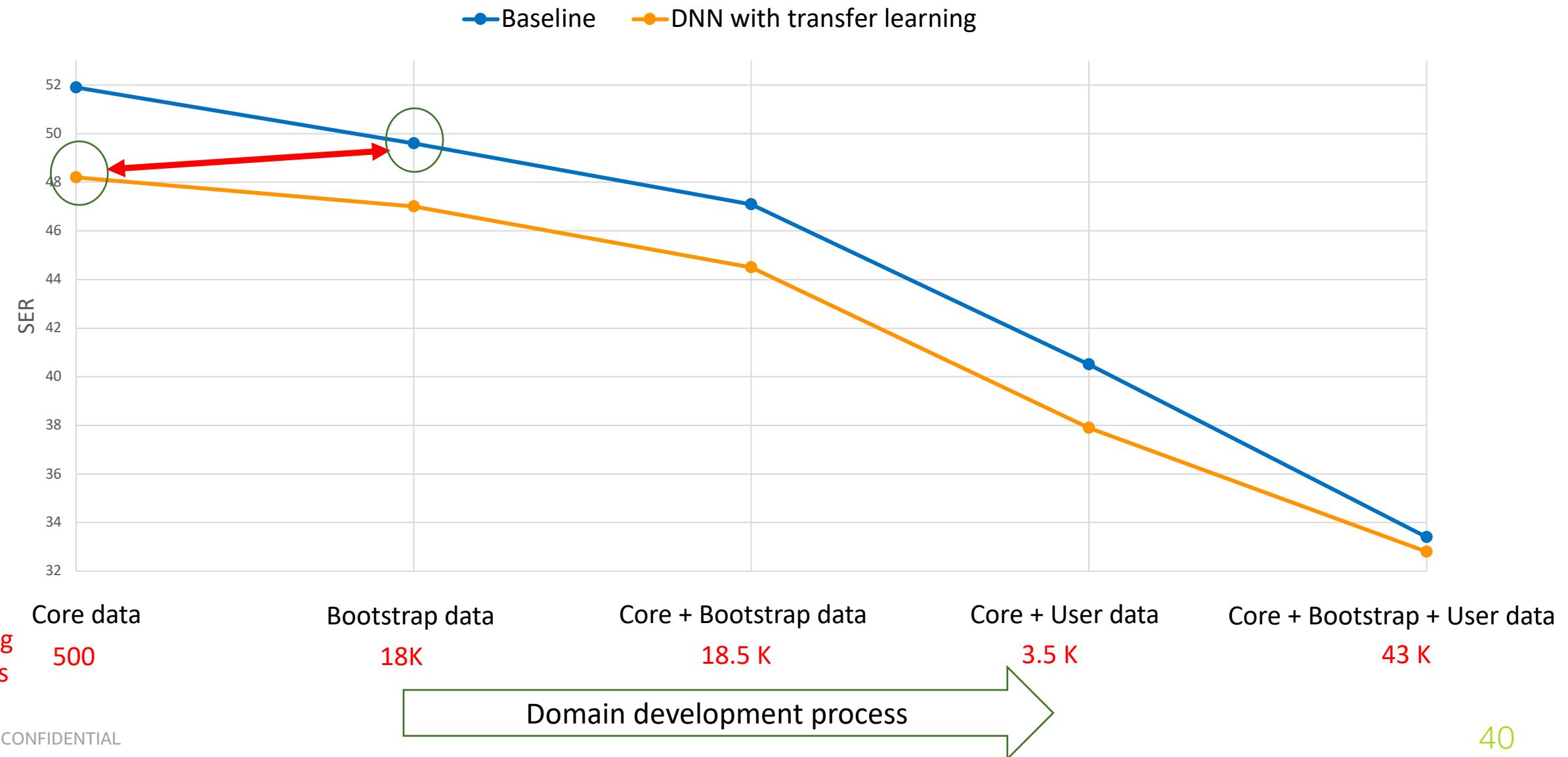
Results for built-in domains (Recipes)



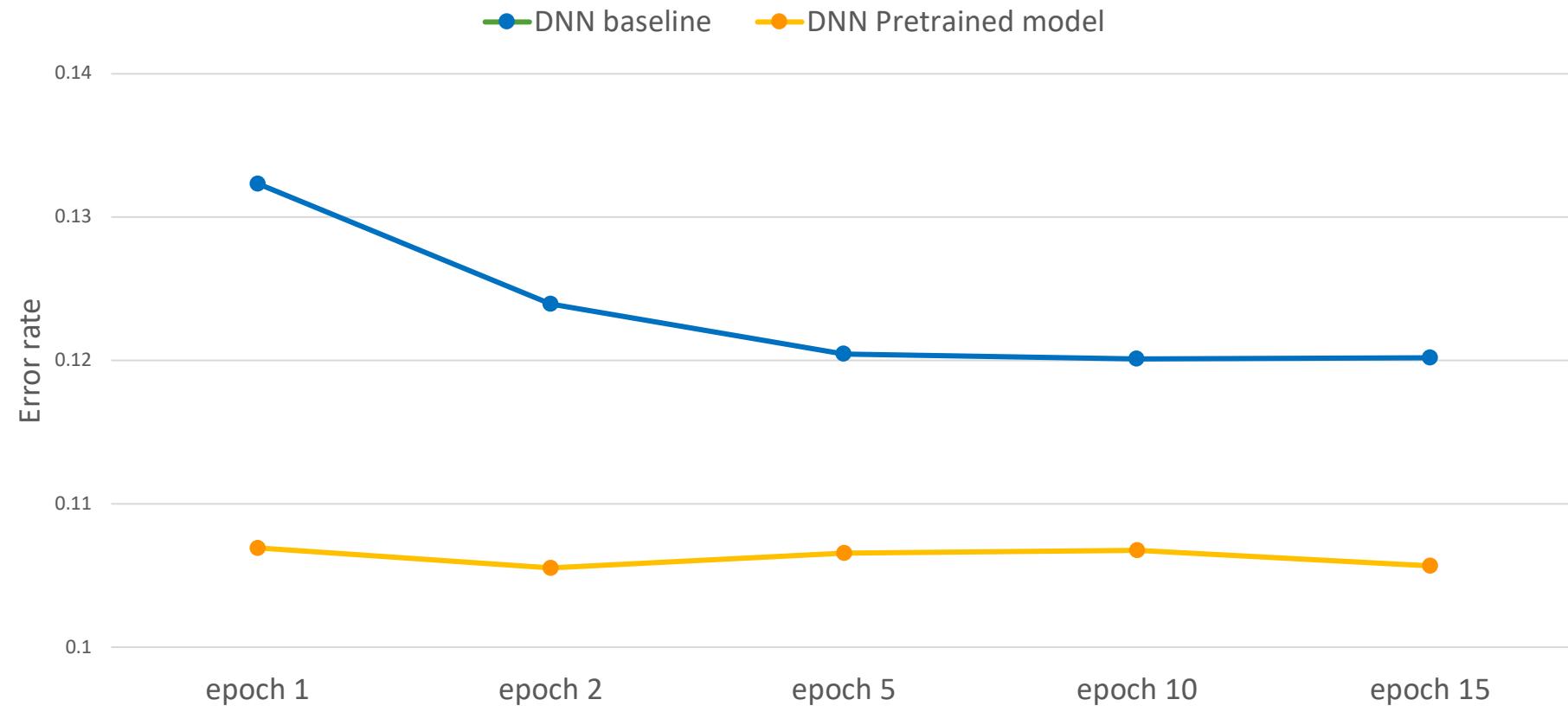
Results for built-in domains (Recipes)



Results for built-in domains (Recipes)



Training speed benefit of transfer learning



Training for pre-trained model is 5X faster than the training models from scratch

Conclusions

- Multi-task learning based DNN architecture can be used for NLU.
- The DNN model benefit from the transfer learning using data from resource-rich domains.
- Transfer learning gives ~10% relative gain in accuracy for both built-in domains and skills.
- We can reach at the required accuracy with a less amount of training data, which means saving of ~2 weeks/intent during domain expansion.

Ongoing/future work

- Try other DNN architectures
 - LSTM-CRF
 - Adversarial loss to learn domain agnostic features
- Explore relation between the pre-training data, target domain and performance
 - Preliminary results show that semantic similarity between source and target domain and transfer learning gains are correlated
- Use semi-supervised learning to select data for pre-training
 - Select target domain specific data for pre-training

Q&A



Anuj Goyal, Angeliki Metallinou, and Spyros Matsoukas. Fast and Scalable Expansion of Natural Language Understanding Functionality for Intelligent Agents. In NAACL 2018.



Appendix

Results for built-in domains (Bookings)

Train set	Baseline	Transfer learning	Relative improvement
Bootstrap data	10.1	6.3	37%
User data	8.2	7.1	13%
Bootstrap + User data	8.2	6.4	22%

Results for built-in domains (Local search)

Train set	Baseline	Transfer learning	Relative improvement
Core data	64.2	51.8	19%
Bootstrap data	64.0	62.8	2%
Core+ Bootstrap + User data	28.1	31.9	- 13%

Analysis

Baseline model: how|Other do|Other you|Other make|Other a|Other mango|FoodItem milkshake|FoodItem
Proposed model: how|Other do|Other you|Other make|Other a|Other mango|DrinkItem milkshake|DrinkItem

Baseline model: where|Other can|Other we|Other find|Other a|Other psychiatrist|PlaceType
Proposed model: where|Other can|Other we|Other find|Other a|Other psychiatrist|Profession

Skill Baseline model: DealIntent: play|Other again|Other
Proposed model: ContinuePlayingIntent: play|Other again|Other

Outline

- Where does NLU fit in Alexa SLU
- Overview of the NLU system
- Process and challenges in NLU domain expansion
- Transfer learning overview
- DNN architecture for NLU
- Transfer learning & DNN for NLU domain expansion
- Results & Analysis
- Future work & Conclusions

