CS 294S/294W
Building the Best Virtual Assistant
A Research Project Course

Monica Lam

Stanford University
lam@cs.stanford.edu

Supported by NSF Grant #1900638

Why a Remote Research Course?

A welcomed change from Zoom lectures.

Expose students to the exciting world of research.

Virtual Assistants!

A once-in-20-years research opportunity	Mainframe, PCs, web, mobile/ubiquitious Entire web available by voice in all languages
Vision	23M voice interface developers
New technical approach	Annotating real data → training-data engineering
A new NLP data engineering tool chain	Virtual assistant programming language Grammar-driven data synthesis Neural language models, machine translation
Multidisciplinary research	HCI, ML, NLP, programming languages Driving applications

We need open-world collaborative research!

A Research Course for Beginners

- Hardest part of a PhD: how to select a topic
  - Apprentice under a thesis supervisor
- A true and tried technique for junior researchers
  - Work with a professor, senior graduate students in a small group
  - Choose from an identified research project: meaningful and doable
    - Or suggest a new topic
- Groups of 2 or 3

Course Design

- Background
  - Lectures on basic technology and hands-on experience (2 homeworks)
- Project proposal (Discussions)
  - Proposed research projects in Google docs (on the website)
  - Your ideas are welcome
- 5-week projects
  - Due Mondays: Weekly status updates
  - Tuesday class: small group feedback
  - Thursday class: students take turns in giving mini-lectures on their research topic (an important part of research training)
- Final project presentation and report

A Tentative Schedule

Week	**Tuesday**	**Thursday**	**Due** (10:30am)
April 7, 9	Course Introduction	Schema → Q&A (HW1)	4/ 9: Student profile
April 14, 16	Schema → Dialogues	Tutorial & Discussion (HW2)	4/16: Homework 1
April 21, 23	Multimodal Assistants	Project Discussions	4/23: Homework 2
April 28, 30	Project Discussions	ML for NLP Primer	4/30: Project Proposal
May 5, 7	Group Weekly Meetings	Students' Mini-lectures	
May 12, 14	Group Weekly Meetings	Students' Mini-lectures	5/11: Weekly Update
May 19, 21	Group Weekly Meetings	Students' Mini-lectures	5/18: Weekly Update
May 26, 28	Group Weekly Meetings	Students' Mini-lectures	5/25: Weekly Update
June 2, 4	Group Weekly Meetings	Students' Mini-lectures	6/ 1: Weekly Update
June 9	Final Project Presentation	—	6/10: Project Report

Grading

- Attendance is mandatory
  - please let us know if you can't make it to class
- In-class participation: 15%
- Homework: 15%
- Final project: 70%

Let's Get to Know Each Other

Overview

Conventional Wisdom

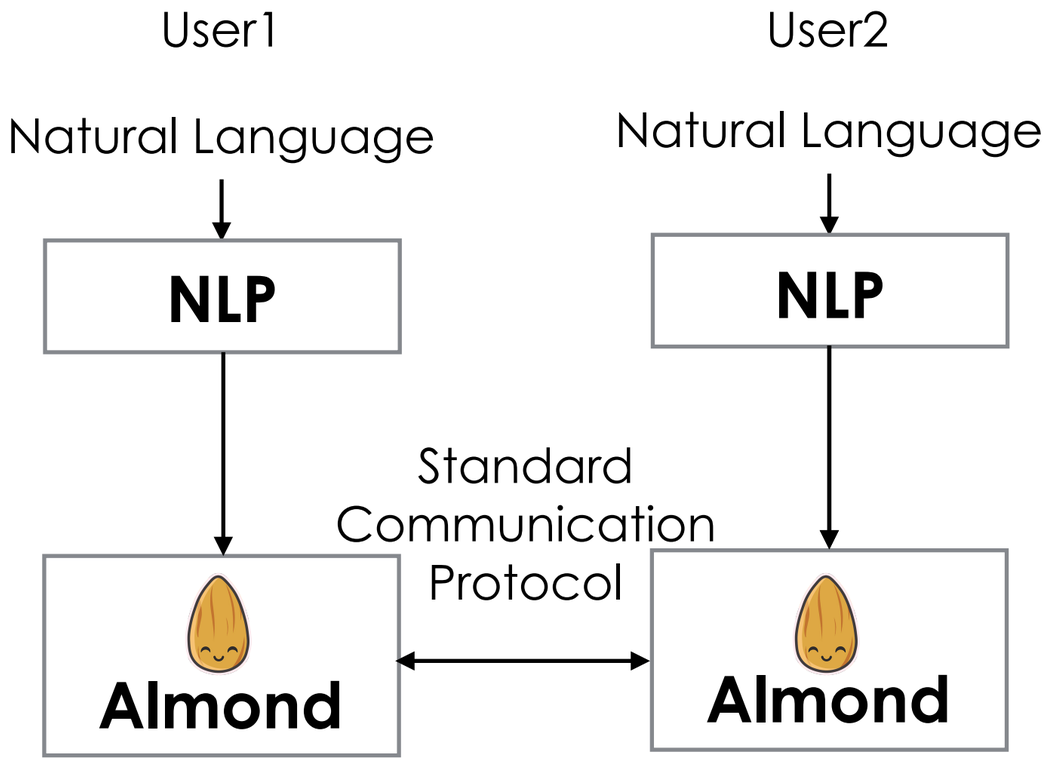
- Natural language processing needs a neural network
- Neural network needs well-annotated real users' training data
  - Pre-requisite: Millions of real users
  - Cost: 10,000 Alexa employees annotating real user data
  - Coverage: Millions still don't have enough coverage
  - Robustness: Dialogue trees, how to handle change of topics?
  - Accuracy: Annotation errors: 30% errors (Multi-Oz)
  - Bootstrapping: How do you start?
  - Scalability: 1.8 B web pages, exponential number of dialogues, thousands of natural language

Metrics: CCRABS

Problem 1

- Will the linguistic technology, web be owned by a duopoly?
  - Alexa: 70% of the 76M installed base of owners in the US
    - 100,000 3rd-party skills, 60,000 compatible IoT devices
- Will it cover the entire web (incl. non-profit)? Rare languages? Is it feasible? Is it profitable?
- Monopolies hurt consumers
  - Privacy, open competition, innovation, quality of service

Protect Privacy with an Open Federated Architecture



A fully-functional research prototype is available as Almond for Android/web.

- **NLP**
  - training in the cloud (currently)
  - inference locally (in the future)
- **Almond: Privacy-preserving assistant**
  - Keeps users accounts & data local
  - Communicate/share with each other (like email)
  - Users share in natural language
- **Integrated with Home Assistant**

Problem 2

- Purely neural approach is prohibitively expensive

Vision of the Future Virtual Assistants

- The entire Web is going voice-accessible!
- Redefine Search
  Based on history, emails, calendar, articulated user preference
- Automation: Natural language programming
  - Personal: order groceries, food every week or evening, pay bills ..
  - Doctors, stock brokers, loan officers
- Advisors Behavior influence/manipulation
  - Fitness, bodybuilding, finances, education, careers

We need a new methodology that is open to all!

Alexa: Syntax-Dependent Representation

Search for an upscale restaurant and then make a reservation for it

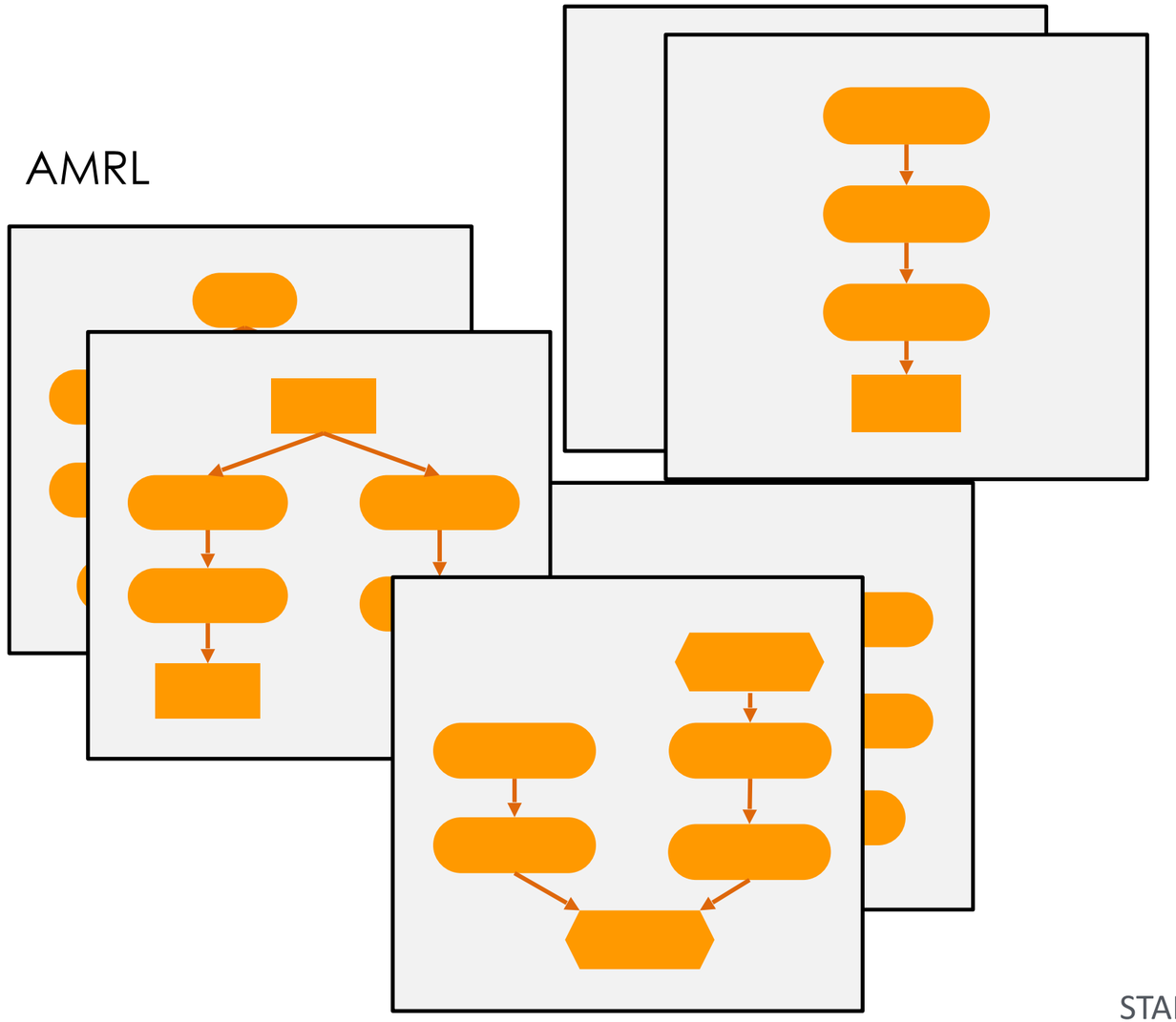
Reserve a high-end restaurant for me

Can you reserve a restaurant for me? I want an upscale place.

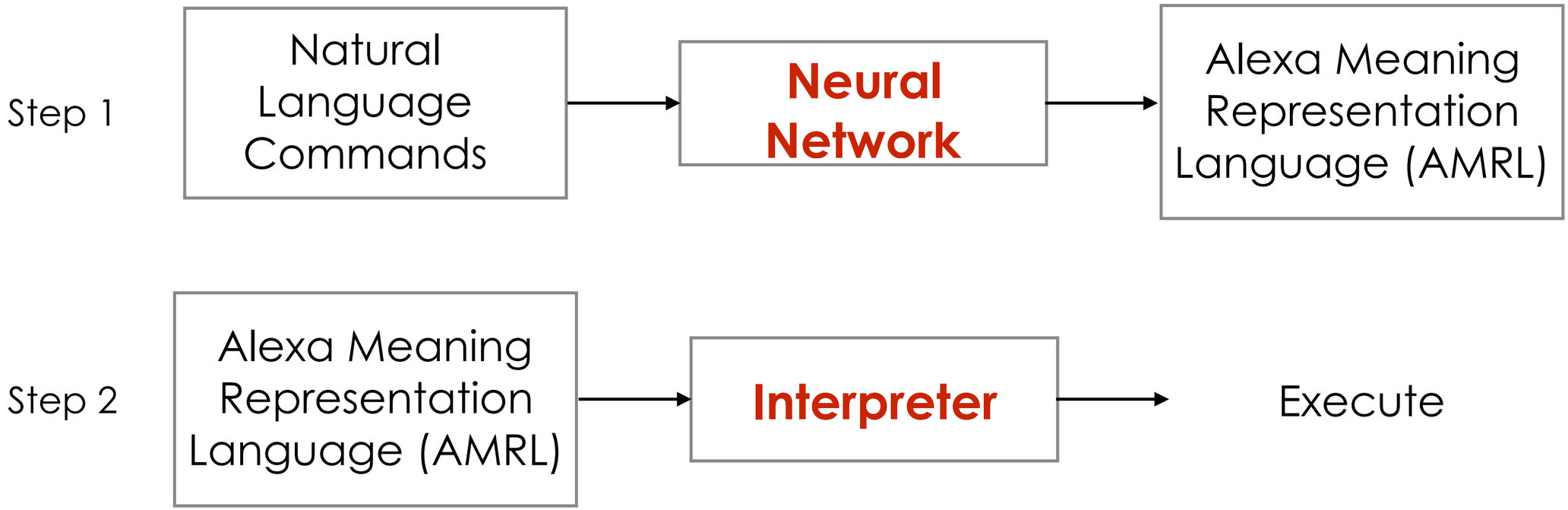
找一家高档餐厅，然后帮我预约

我想预约一个高级餐厅

یک رستوران خوب پیدا کنید و برای من قرار ملاقات بگذارید



Alexa's 2-Step Approach



Idea 1: End-to-End Translation

- Human-computer communication
  - Easier than understanding human-human communication.
- ThingTalk:
  formal virtual assistant programming language
  - Capture full capability
  - Independent of language syntax, source natural language
- End-to-end translation
  - Let neural network figure out the intermediate representation

Text

Search for an upscale restaurant and make a reservation for it

Meaning: ThingTalk code

```
now => @com.yelp.Restaurant(),
price == enum(expensive)
=> @com.yelp.reserve(restaurant=id)
```

Unique Semantic Representation

```
now
=> @com.yelp.Restaurant(),
   price == enum(expensive)
=> @com.yelp.reserve
   (restaurant=id)
```

Reserve me a luxury restaurant

Could you please get me a restaurant that is upscale? want to reserve one.

给我找一家高级餐厅并预约

یک رستوران مجلل جستجو کنید و سپس برای آن رزرو کنید

E ʻimi i kahi hale ʻaina hulahula a laila hana iā ia no ka mālama ʻana iā ia

高級レストランを検索してから予約する

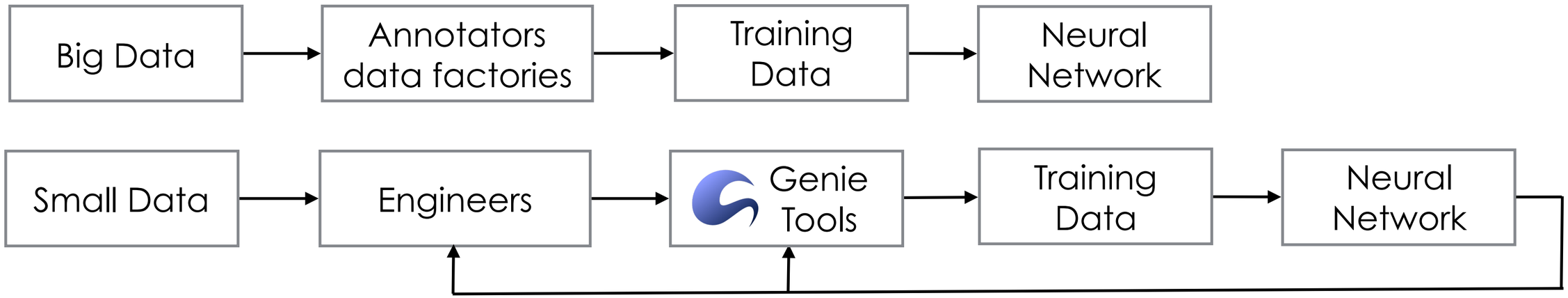
Cerca un ristorante di lusso e dammi la prenotazione

Per favore riesci a trovarmi un ristorante? Ho bisogno di qualcosa di lussoso.

Prenotami un ristorante da lusso

Idea 2: Training-Data Engineering

- Tools to address CCRABS
  - cost, coverage, robustness, accuracy, bootstrapping, scalability
- Apply CS engineering approach to AI training data



Q&A

Alexa
User hand-codes question/code 1 by 1

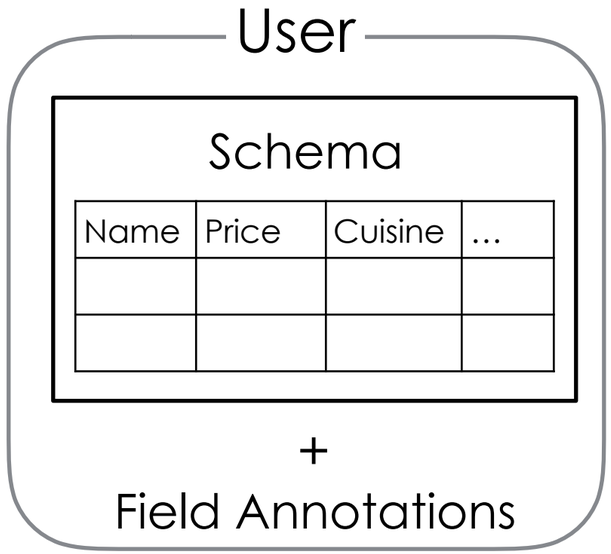
get me an upscale restaurants

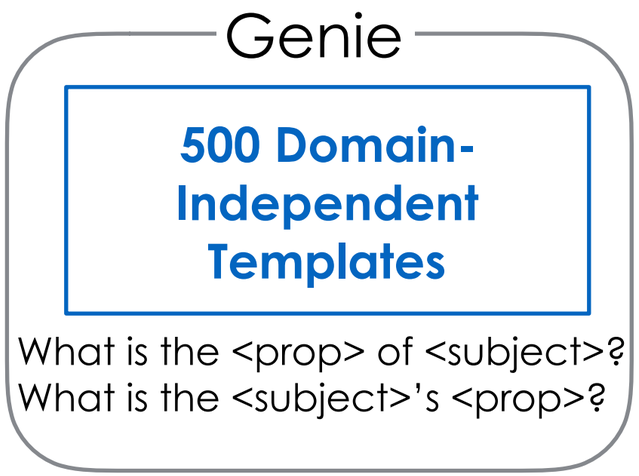
What are the restaurants around here?

What is the best restaurant?

search for Chinese restaurants

Genie: Synthesizes question/code from a schema





get me an upscale restaurants
What are the restaurants around here?
What is the best restaurant?
search for Chinese restaurants
What is the best restaurant within 10 miles?
Find restaurants that serve Chinese or Japanese food
What is the best non-Chinese restaurant near here?
Show me a cheap restaurant with 5-star review.
Are there any restaurant with at least 4.5 stars?
What is the phone number of Wendy's?
I'm looking for an Italian fine dining restaurant.
Give me the best Italian restaurant.
Find me the best restaurant with 500 or more reviews
Show me some restaurant with less than 10 reviews
...

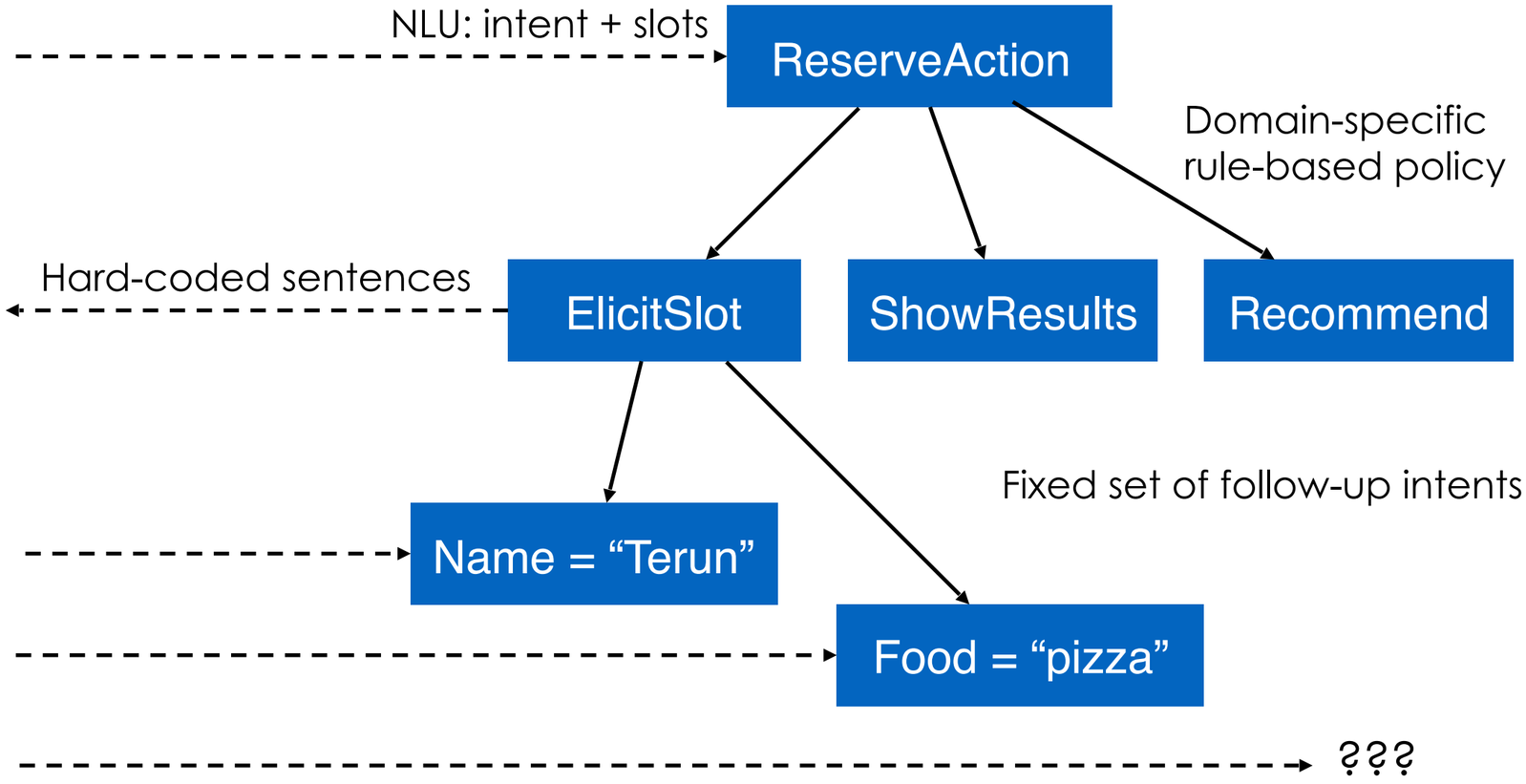
Today's Dialogue Trees: Laborious & Brittle

A: Hello, how can I help you?

U: I'm looking to book a restaurant for Valentine's Day

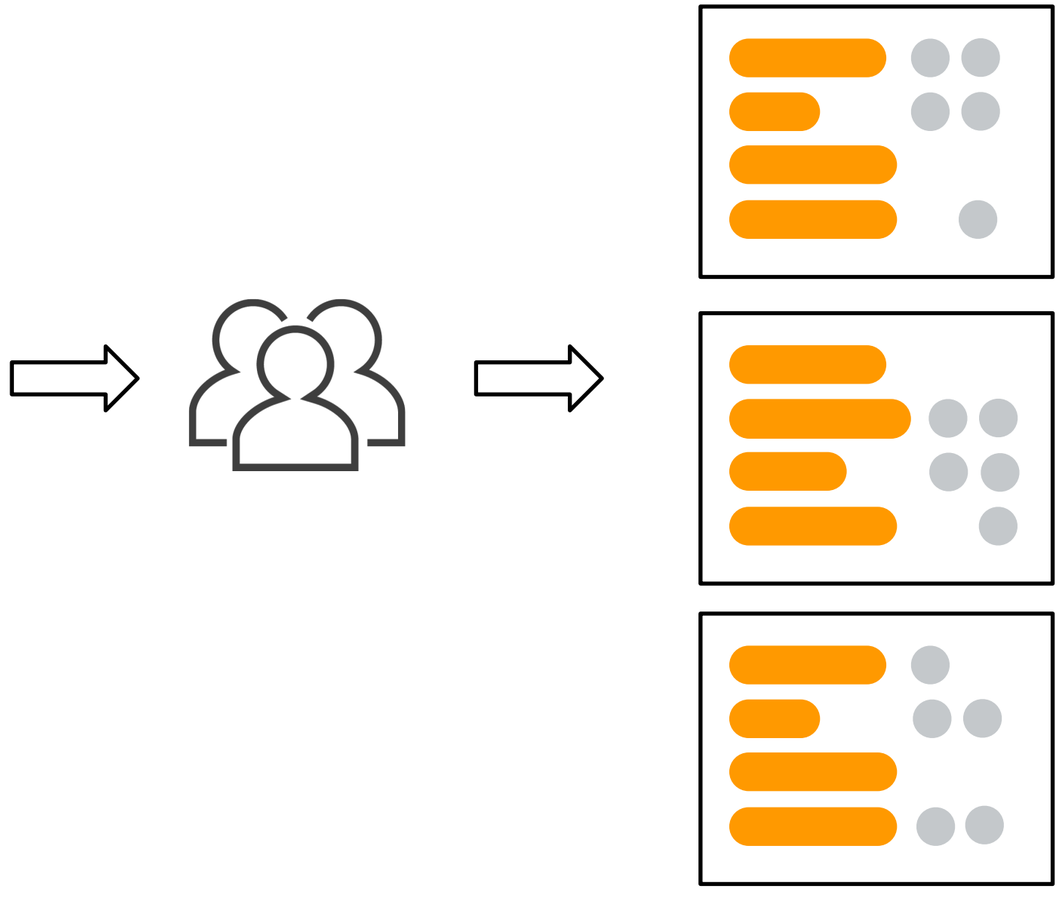
A: What kind of restaurant?

U: Terun on California Ave
-- or –
U: Something that has pizza
-- or –
U: I don't know, what do you recommend?

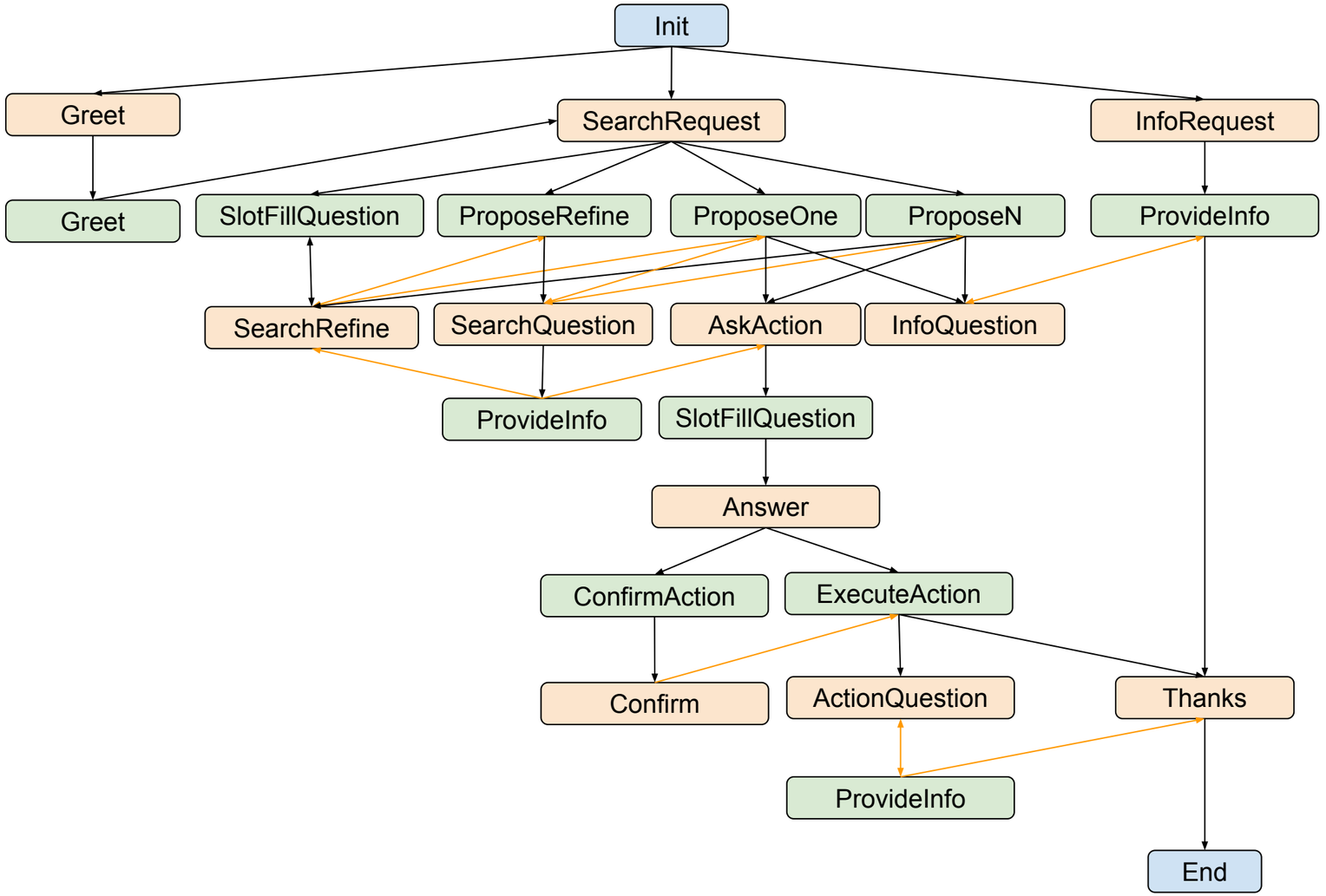


Alexa: Annotate 1 Dialogue at a Time

Annotation of intents and slots



Genie: Transaction Dialogue State Machine



Technology Stack

Businesses

Restaurant Table

Domains

Schema

Name	Price	Cuisine	...

Restaurant Reservation API

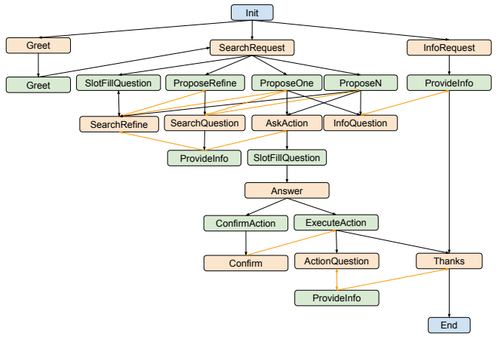
Annotated Small Data

Sentence Templates

What is the <prop> of <subject>?
What is the <subject>'s <prop>?

Dialogue Models

Transaction Dialogue State Model



Synthesis

Training Data

StateResult
Restaurant, price == moderate && geo == "Palo Alto"
{ id = "Terun", price = moderate, cuisines = ["pizza"], … }
{ id = "Coconuts", price = moderate, cuisines = ["caribbean"]}

ProposeOne
I have Terun. It's a moderately priced restaurant that serves pizza.
+code

ProposeN
I found Terun and Coconuts. Both are moderately priced.
+code

AskAction
I like that. Can you help me book it? I need it for 3 people.
+code

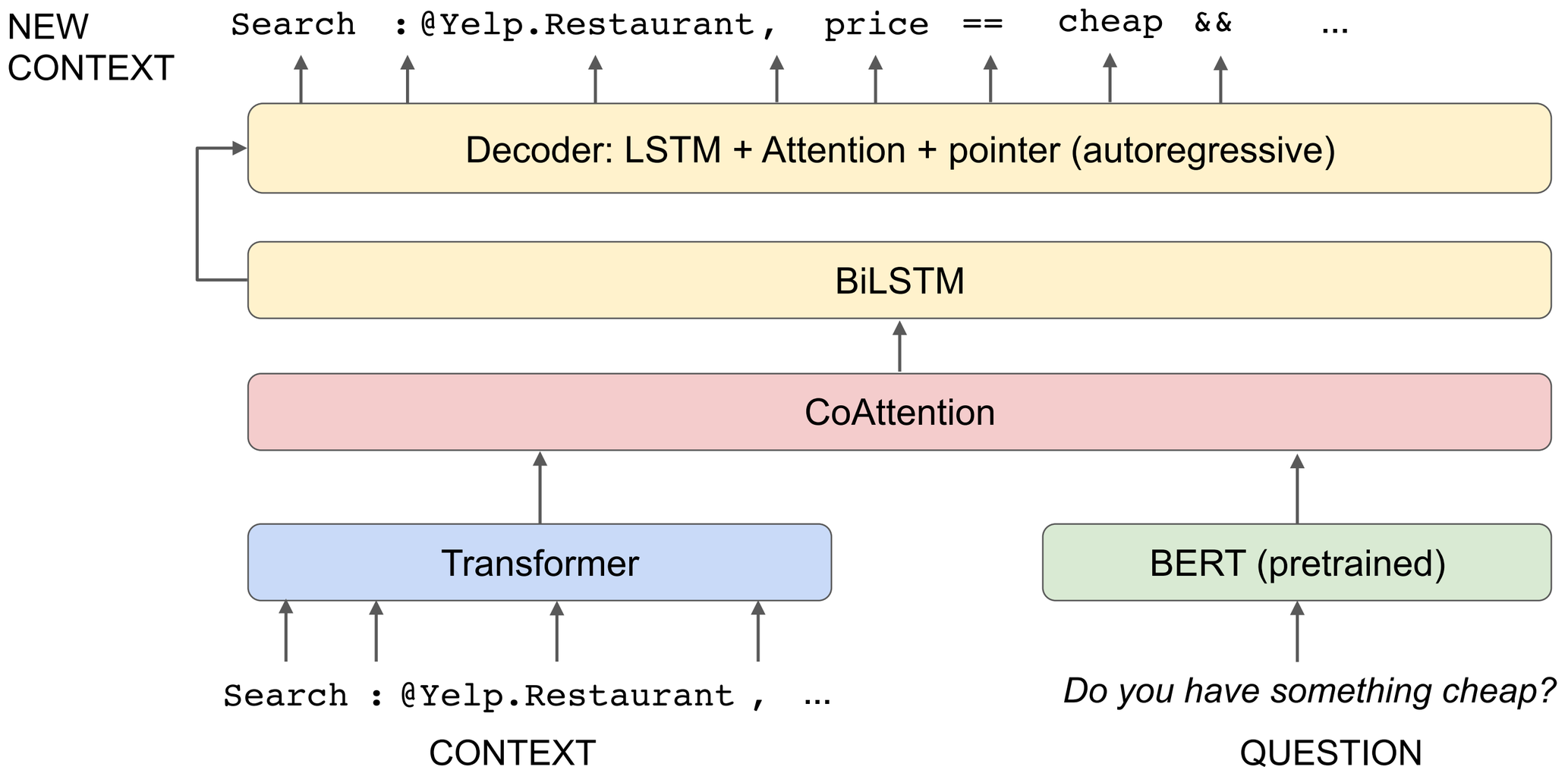
SearchRefine
I don't like pizza. Do you have something Caribbean?
+code

InfoQuestion
Can you tell me the address of Terun?
+code

Neural Network

Restaurant Reservation Agent

Contextual Language Understanding Model



Preliminary Results

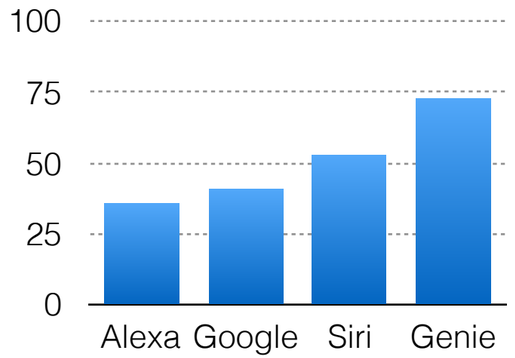
API annotations → multi-domain event-based actions
When Apple's stock drops to $200, buy $10,000

API annotations → Access control
My dad can view my security camera if I am not home.

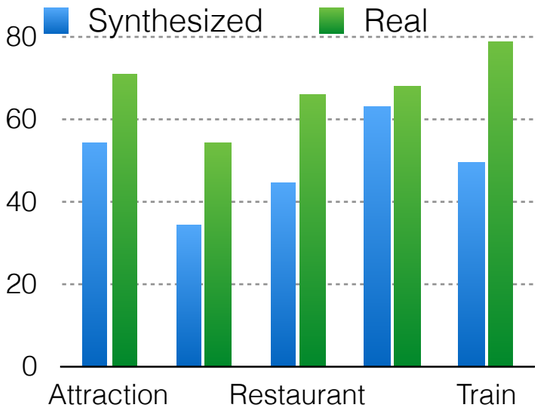
Schema annotations → accurate complex queries
Find a Spanish restaurant open at 10pm

Schema annotations → Neural dialogue acts + agent
61% turn-by-turn accuracy on restaurants in MultiWoz

Transfer learning to new domains (MultiWoz dialogues)
Synthesized data training achieves 73% of real data



Complex Queries



Domain Transfer
for Dialogues

Potential Projects

Discipline	Examples
Applications	Assistants: Social, Music, COVID-19, Minecraft for Autistic Children
Multi-disciplinary	Two-Way Conversations
HCI + NLP	Program by Example + Voice
ML	Improvement with User Feedback
	Neural Model Experimentation for Assistants
	Multi-Lingual Assistants
	Controllable and Natural Response Generation
	Multi-Domain Transactional Dialogues
Systems	Automatic Template Creation
	Completeness of Template-Based Question Synthesis