CS224v

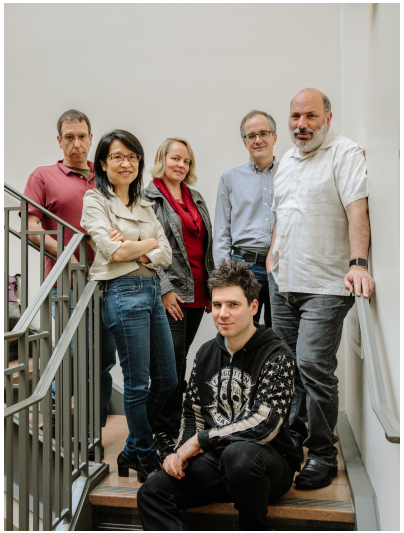
Conversational Virtual Assistants with Deep Learning

Monica Lam

CAs: Ethan Chi & Surabhi Mundada

PhD students: Giovanni Campagna, Mehrad Moradshahi, Sina Semnani, Silei Xu, Jackie Yang

My Background



The New York Times

Stanford Team Aims at Alexa and Siri With a Privacy-Minded Alternative

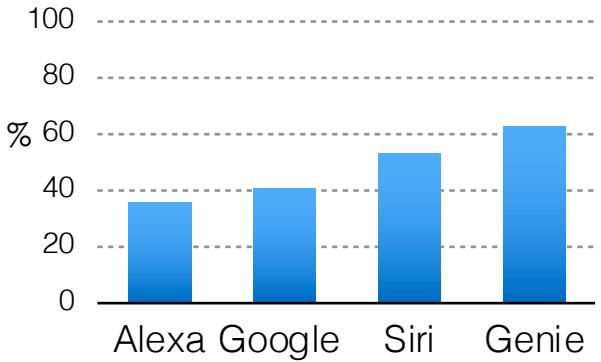
- **Compiler expert; author of the Dragon book**
- **Started research on privacy (2008)**
- **Research on conversational virtual assistants (2015)**
- **Focus: deep learning + programming languages**
  - Natural language (NL) is a human artifact to communicate, not a natural phenomenon
  - Need to understand new, long-tail sentences
  - Problem: Natural language programming
- **Leading an open virtual assistant initiative**
  - Goal: 20M+ voice developers (non AI experts)

Project Status

The Genie assistant is running on:

Home Assistant
local gateway

Baidu smart
speaker



Long-Tail Restaurant Queries

The first assistant that uses a contextual neural parser for dialogues

Leading the standardization of dialogue semantics representation

Lecture Outline

1. **Motivation of the Course**
2. Core Concept: Understanding Task-Oriented Dialogues
3. Technology to Address Ethical Considerations
4. This Course

Voice

3rd-Generation Computer Interface

World Wide Voice Web

Access knowledge in natural language	Which restaurant serves paella, with a rating over 4.2 that is quickest to get to?
Web Transactions	Book a Covid vaccine appointment at the nearest Safeway, as soon as possible.
Social media	Find me all the photos from my friends taken at Halloween.
The Internet of Things	Let my dad monitor my security camera for motion, when I am not home.

WWW: Browser

WWvW: Voice Assistant

Voice

- Not limited like GUI: WIMP (Windows, Icons, Menus, Pointers)
- Advantages and Key Challenges
  - Scale: The entire web
  - Expressiveness: Interactive expression of users' intention
  - Power: Function composition, over multiple websites
  - Personalized: A long-tail of personalized operations
  - Universal: Multi-lingual, even the preliterate, the illiterate

Computer is now learning the human language!

The key is Natural Language Understanding!

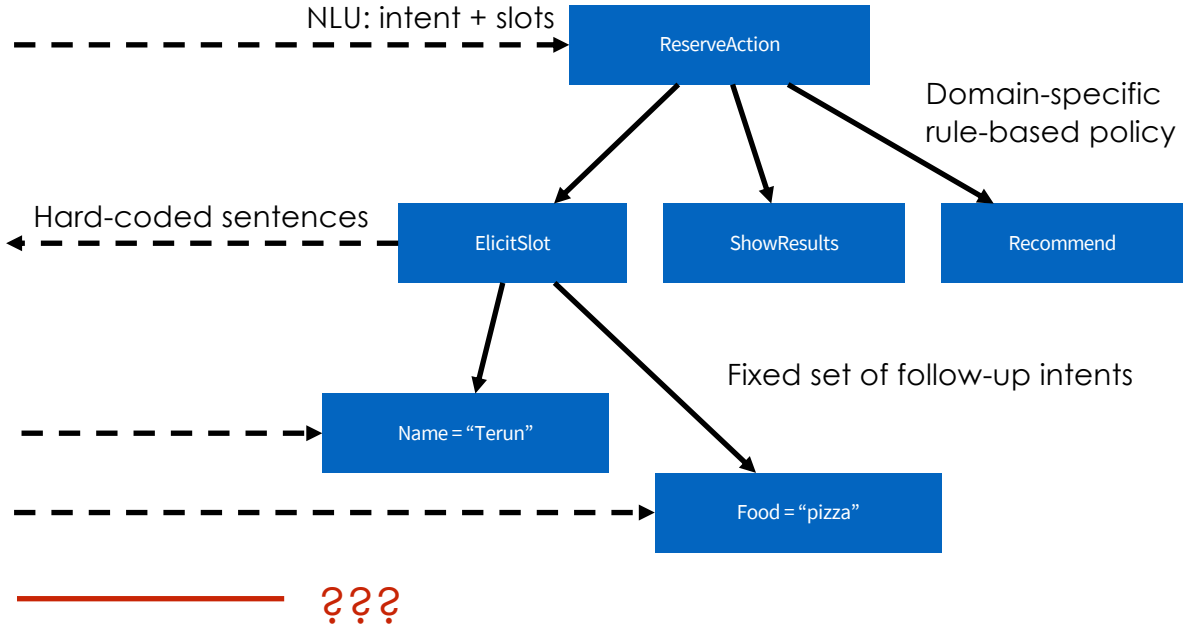
Commercial Agents

- Commercial assistants are not conversational today
  - We will change that!
- Customer support uses dialogues, using Dialogue Trees
  - Handcraft each turn of the dialogue
  - Predict what the user says
  - Use an intent classifier

Dialogue Trees & Intent Classification

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?

Status of Chatbot Technology

- Brittle: Unexpected sentences cause the assistant to fail
- Laborious:
  - Requires continuous refinement
  - Every chatbot is custom built:
    Work is repeated for the workflow of each domain

Traditional Deep-Learning

- Annotate what users say and train a neural network
- In academia: MultiWOZ
  - 5 domains – restaurants, hotels, taxis, trains, attractions (slots only)
  - Hand-annotate simulated dialogues - 3 times!
  - Error-prone: 15% error rate!
  - Best result with this approach: 61% accuracy
- In industry: Alexa has 10,000 employees

Too expensive; Too inaccurate; Too limited

Cannot Rely on Manual Annotation!

- Scale: The entire web
- Expressiveness: Interactive expression of users' intention
- Power: Function composition, over multiple websites
- Personalized: A long-tail of personalized operations
- Universal: Multi-lingual, even the preliterate, the illiterate

Cannot annotate and train one domain at a time!

How about Pretrained Language Models?

Unsupervised learning on internet data

Model	Objective Functions	Params	Year
BERT	Mask model	340M	2018
BART	Denoising seq2seq	406M	2019
MBART	25 languages BART	680M	2020
GPT	Next word prediction	110M	2018
GPT-2	Next word prediction	1.5B	2019
GPT-3	Next word prediction	175B	2020

Can predict the next word, sentence, paragraph given a paragraph

GPT-3

- User supplies a preamble,
  GPT generates follow-on text
- Recipe generator
  Resume generator
  Medical Q&A
  Podcasts, Creative Friction

Medical Q&A

Question: A 10 year old boy presents with recurrent episodes of dyspnoea, cough and wheeze triggered by cold air and allergens. After performing tests he is diagnosed with an obstructive disease and given medication. Which receptor is the medication most likely to work on:

A) muscarinic receptor
B) G-protein coupled receptor
C) Era
D) Erb
E) a-1 receptor

Correct answer is **B**

Explanation: The patient is suffering from asthma, a disease of the airways. The drugs used to treat asthma are bronchodilators. The bronchodilators act on the beta-2 receptors. The beta-2 receptors are G-protein coupled receptors

Does GPT-3 Understand?

Question: "Who is the Director of Stanford AI Lab?"
Answer: Chris Manning, since 2018

GPT3: "Fei-Fei Li"

GPT-3 has lots of statistical language knowledge

We need to ground Pretrained Models in Semantics

WWvW: Research Questions

- Scale: How to put the web on voice efficiently?
- Completeness: How to understand all possible operations?
- Multi-linguality: How to handle low-resourced languages?
- Dialogues: How to understand and carry out dialogues?

WWvW Needs a New Deep-Learning Approach

Key Takeaway

- **WWvW Needs a New Deep-Learning Approach**
  - To ground pretrained language models with semantics

Lecture Outline

1. Motivation of the Course
2. **Core Concept: Understanding Task-Oriented Dialogues**
3. Technology to Address Ethical Considerations
4. This Course

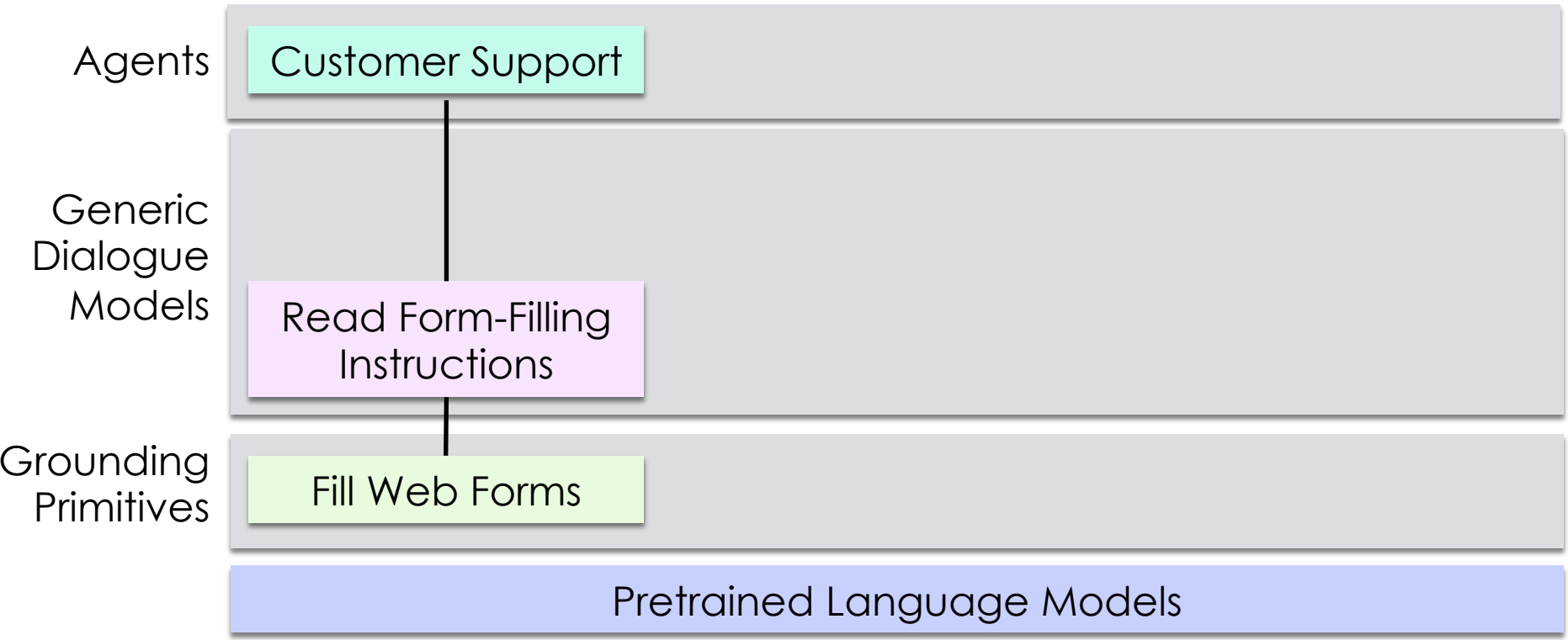
WWvW Needs a New Deep-Learning Approach

Open Pretrained Virtual Assistants

- **Pretrained**
  - Can handle specific tasks when given specific info
    e.g. instructions, data tables. web forms (like a human agent)
  - Leverages pretrained models
- **Open-source, crowdsourced**
  - Open, multilingual voice interfaces for music, news, TV, …
  - Need world-wide contributions (like Wikipedia)

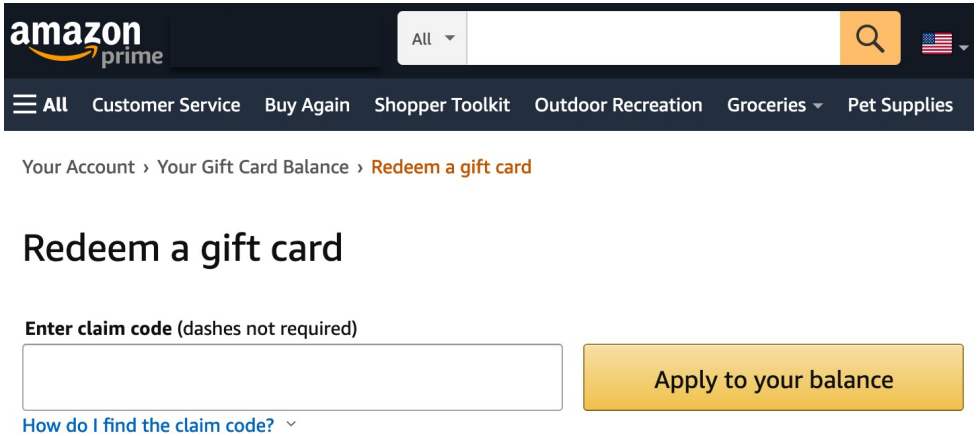
Genie: An open-source toolkit to make voice a commodity

Genie: Open Pretrained Assistant

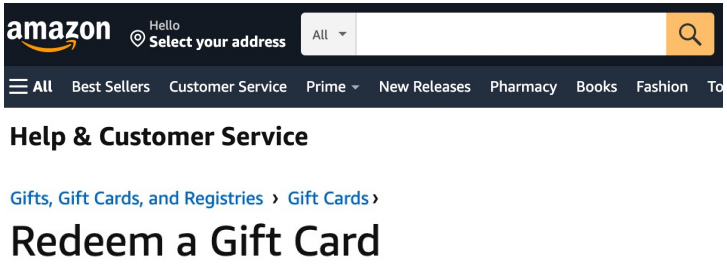


From Web Instructions in NL

Customer Service Web Page



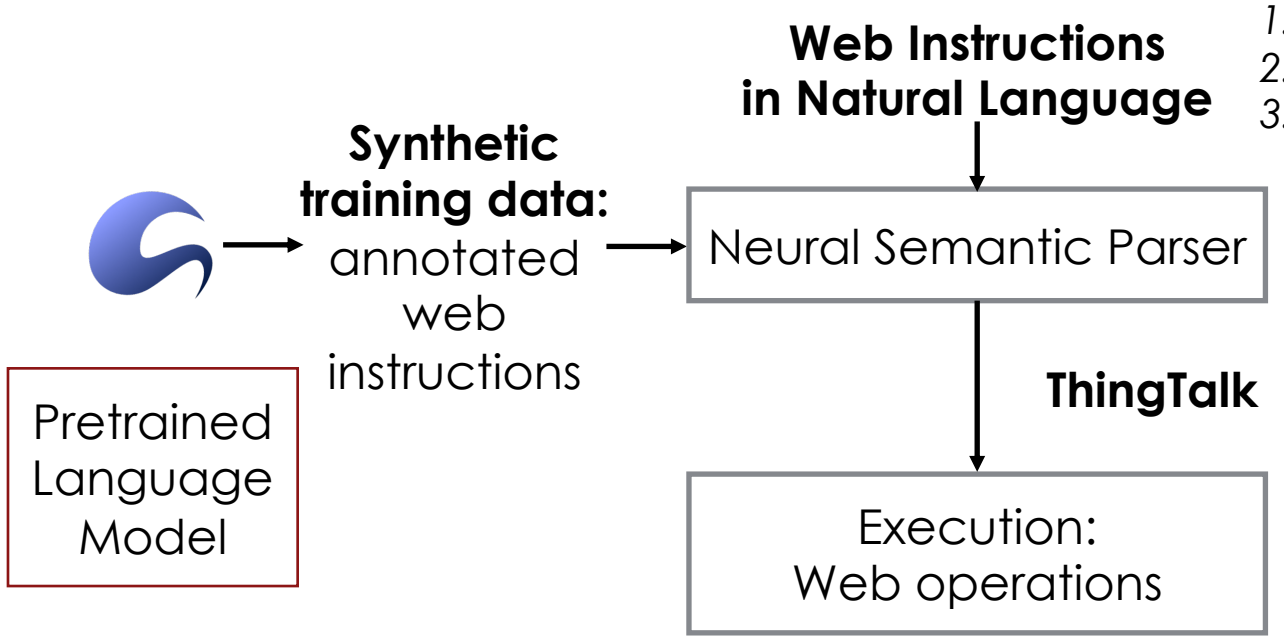
Help instructions



Agent instructions

1. *Ask the user for the claim code*
2. *Go to Redeem a Gift Card.*
3. *Enter the claim code and select **Apply to Your Balance**.*

A Pretrained Assistant Architecture

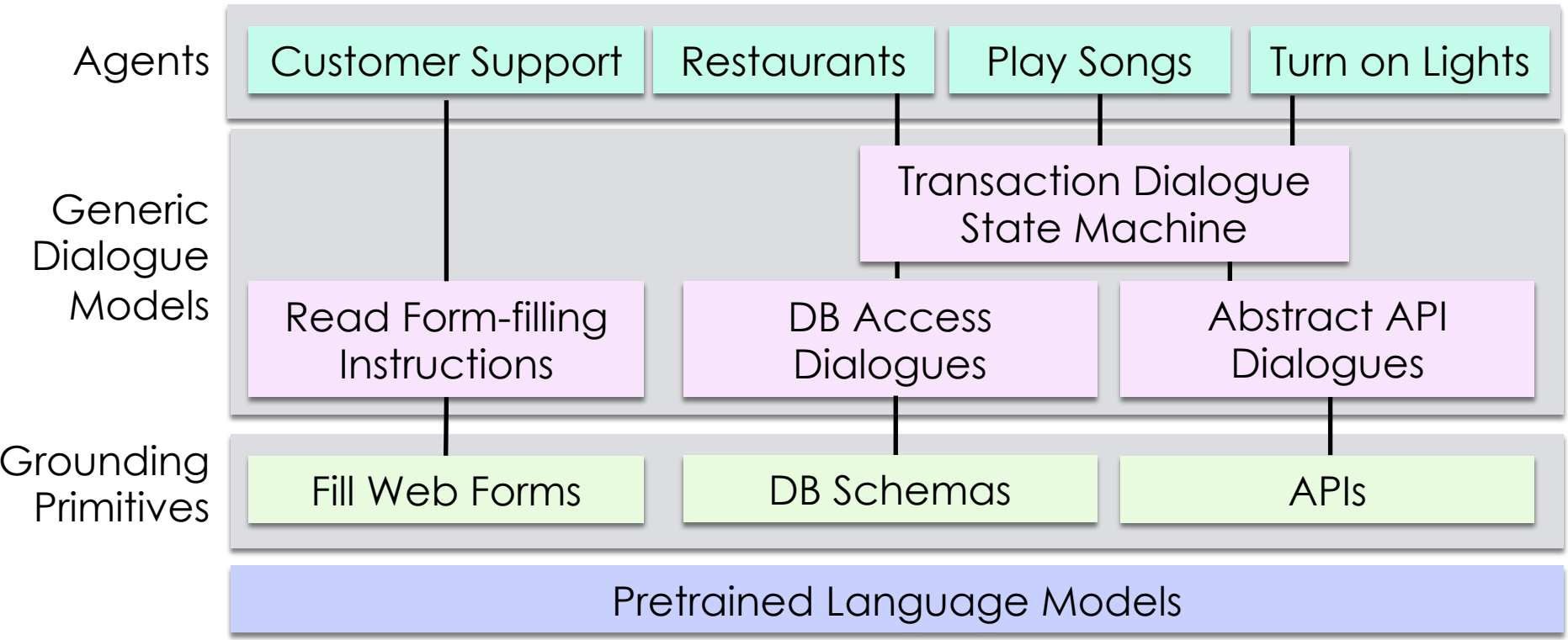


A Pretrained Customer Service Agent

- Trained once, run on any web service instructions
- Read instructions → answer calls immediately
- Experiment: 80 customer services (741 instructions)
- 76.7% accuracy
- 69% of users prefer Genie over following web instructions

Grounding Open-Domain Instructions to Automate Web Support Tasks
Nancy Xu, Sam Masling, Michael Du, Giovanni Campagna,
Larry Heck, James Landay, Monica S Lam
Proceedings of the NAACL-HLT, June 2021.

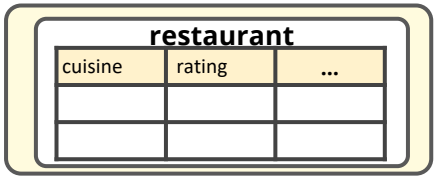
Genie: Open Pretrained Assistant



Genie: Schemas to Dialogue Agents

Database

Restaurants

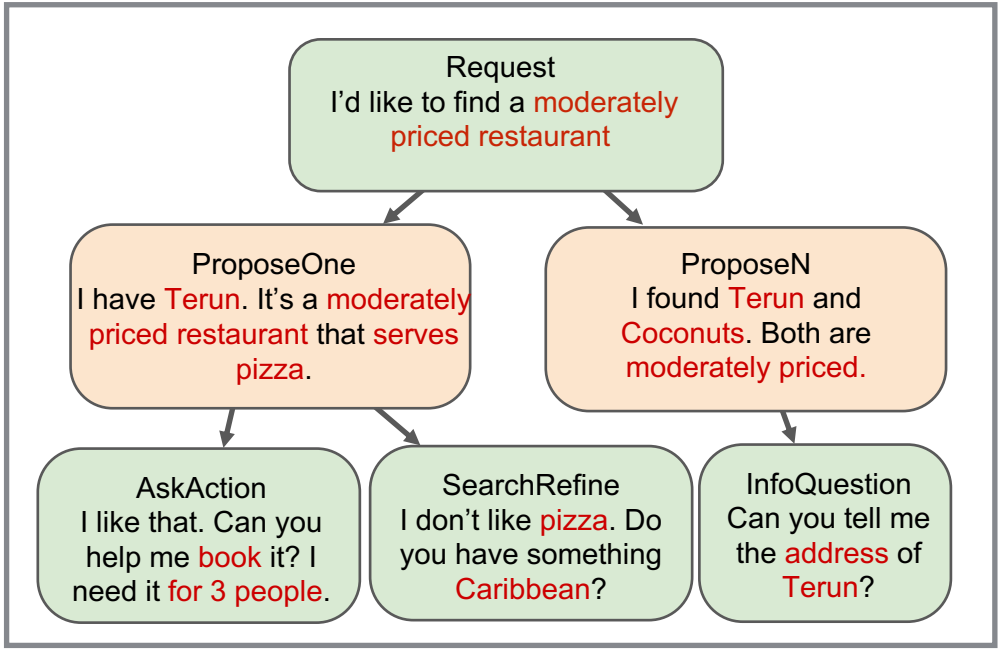


API

Book a restaurant

Pretrained Language Model

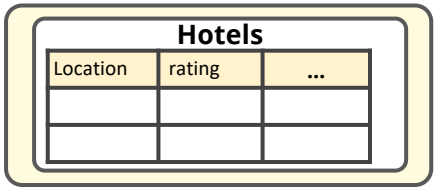
Dialogues about Restaurants



Genie: Schemas to Dialogue Agents

Database

Hotels

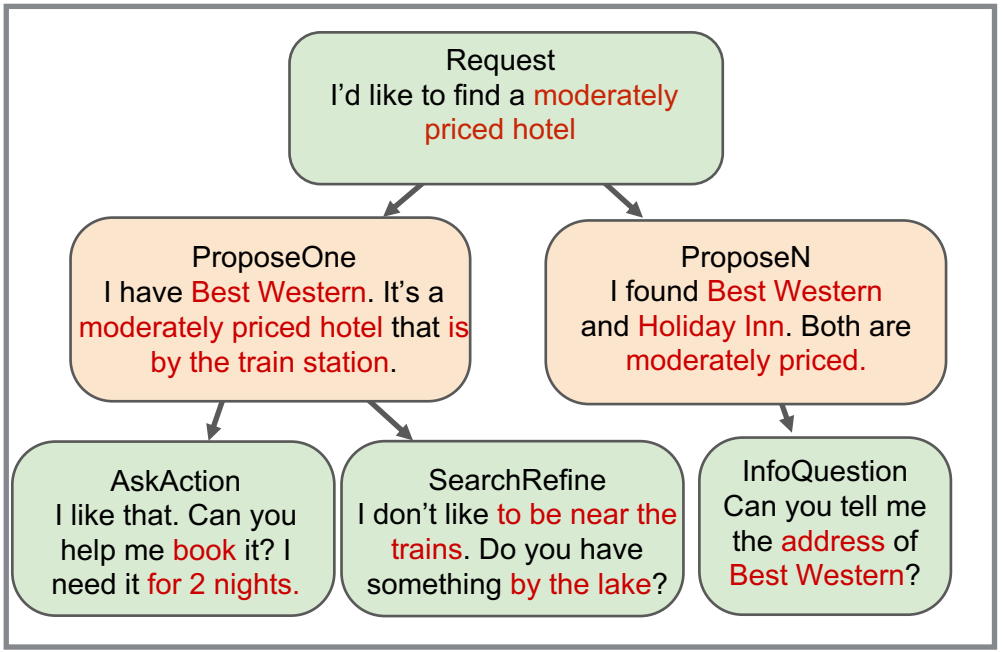


API

Book a hotel

Pretrained Language Model

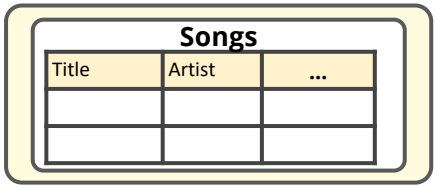
Dialogues about Hotels



Genie: Schemas to Dialogue Agents

Database

Songs

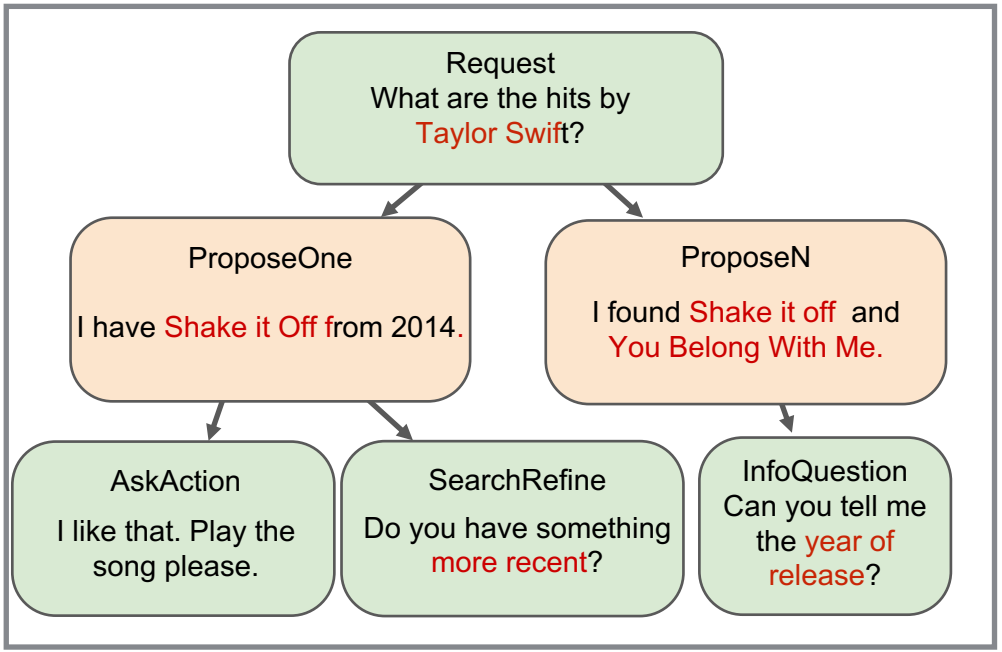


API

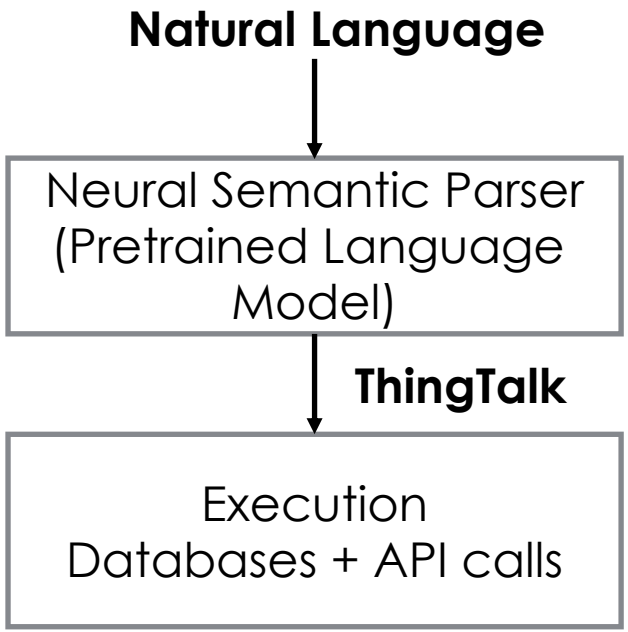
Play a Song

Pretrained Language Model

Dialogues about Songs



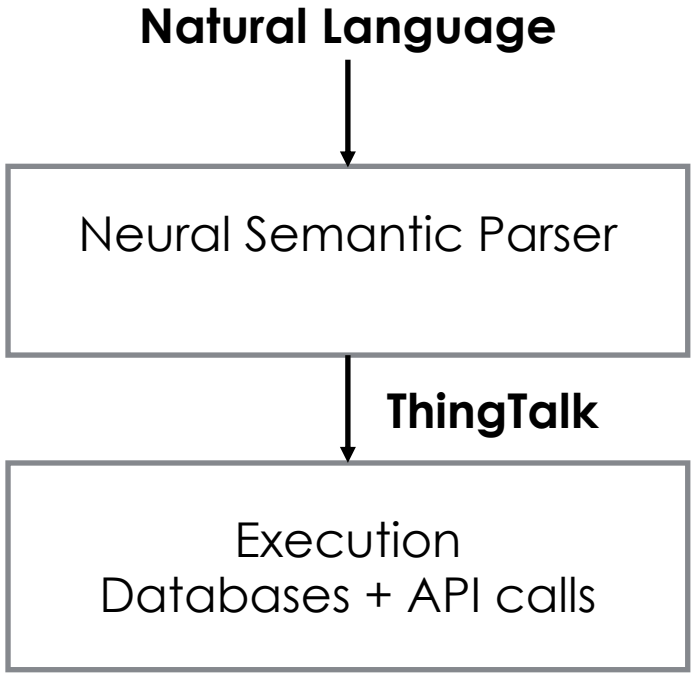
Pretrained Assistant Architecture



- **ThingTalk:**
  Programming language
  Covers everything that a computer can do

- **Neural semantic parser:**
  leverage pretrained models for natural language knowledge

Pretrained Assistant Architecture



1. **Completeness:**
   everything a computer can do
   - **ThingTalk:**
     1st programming language for virtual assistants
     - Database query
     - API Invocation
     - Composition
     - Event-driven
     - Access control

Queries in ThingTalk

Grammar

$$table\ [, filter]^{?}$$

ThingTalk

```
@yelp.restaurant(), geo==new Location("Palo Alto")
```

English

Show me restaurants in **Palo Alto**

Queries in ThingTalk

Grammar

$$table\ [, filter]^{?}$$

ThingTalk

```
@yelp.restaurant(), geo==newLocation("Palo Alto")
                 && servesCuisine =~ "Cuban"
```

English

Show me **Cuban** restaurants in Palo Alto

Queries in ThingTalk

Grammar

sort *fn* asc|desc of *table* [, *filter*]?

ThingTalk

```
sort aggregateRating.ratingValue desc of (
@yelp.restaurant(),geo==new Location("Palo Alto")
                && servesCuisine =~ "Cuban" )
```

English

Show me **top-rated** Cuban restaurants in Palo Alto

Queries in ThingTalk

Grammar

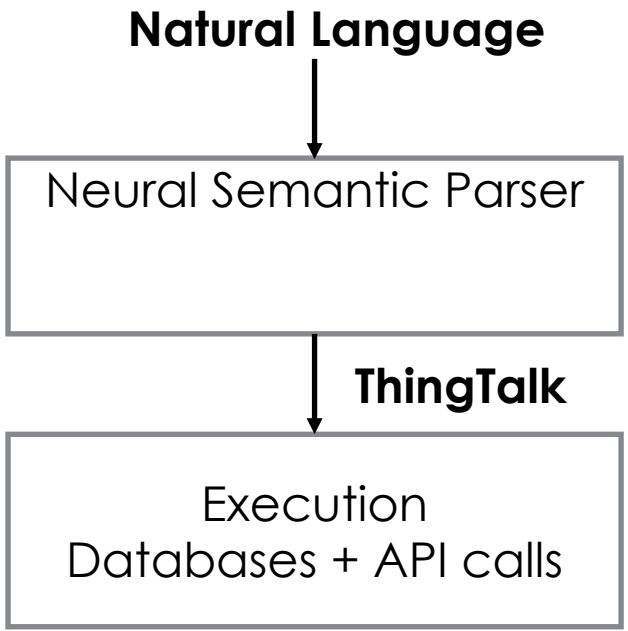
$$\texttt{sort } fn \texttt{ asc|desc of } table \ [\texttt{, } filter]^{?}$$

ThingTalk

```
sort aggregateRating.ratingValue desc of (
@yelp.restaurant(), geo==new Location("Palo Alto")
                && servesCuisine =~ "Cuban" )
     && contains (review,
                  any(@yelp.review(), author="Bob"))
```

English Show me top-rated Cuban restaurants in Palo Alto **reviewed by Bob**

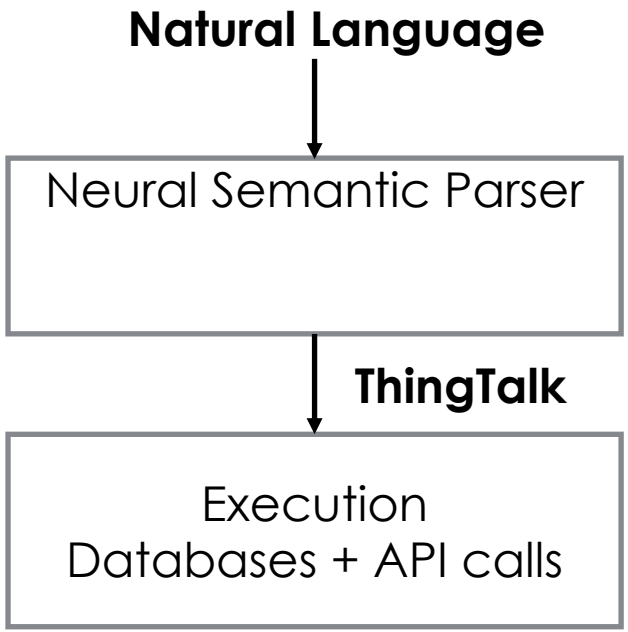
Pretrained Assistant Architecture



1. **Completeness: ThingTalk**
2. **Coverage of training data: Synthesis**
   - Don't just annotate, synthesize
   - Generate training data samples automatically
   - Use PL grammar:
     - relational algebra (Join, project, …) control constructs
     - from databases, APIs

Yushi Wang, Jonathan Berant, and Percy Liang (2015). "Building a Semantic Parser Overnight". In ACL

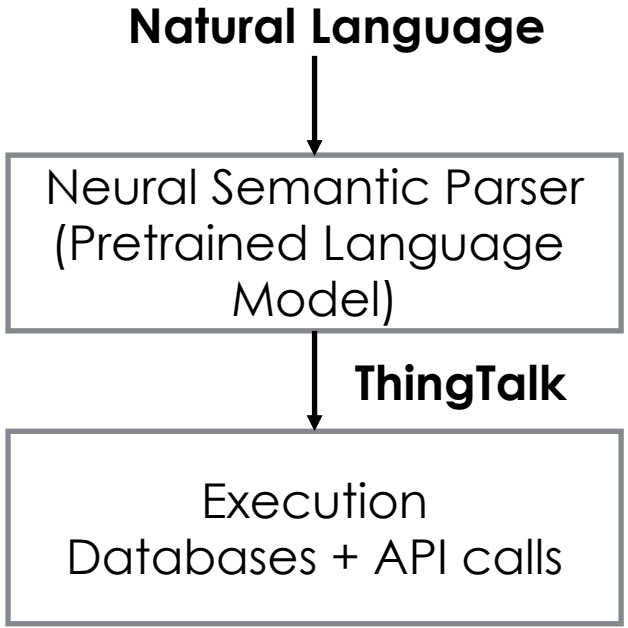
Pretrained Assistant Architecture



1. **Completeness: ThingTalk**
2. **Coverage of training data: Synthesis**
3. **NL knowledge: pretrained models**

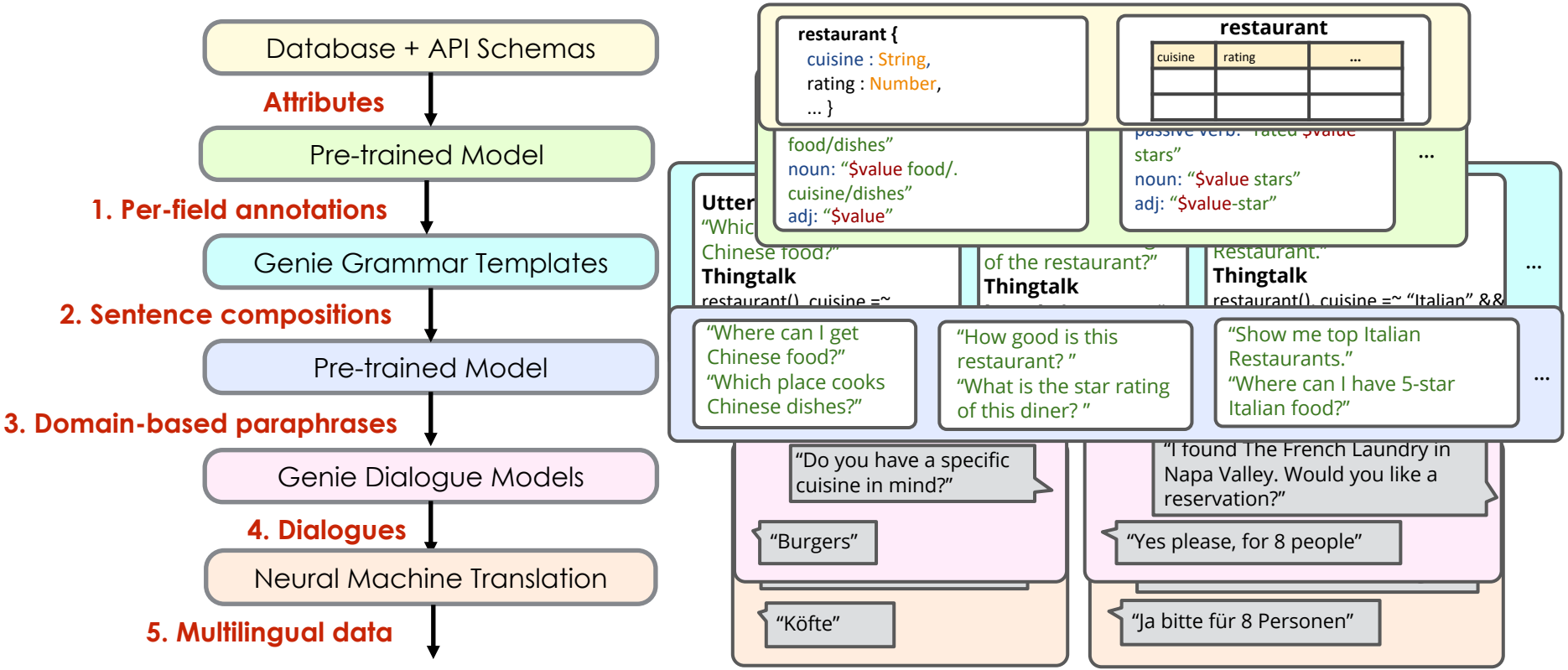
- Use a pretrained model to paraphrase synthetic data
- Use neural translator for multi-lingual data

Pretrained Assistant Architecture



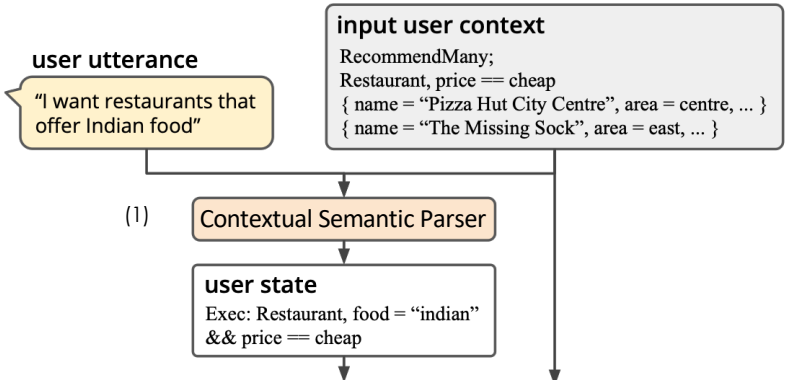
1. **Completeness: ThingTalk**
2. **Coverage of training data: Synthesis**
3. **NL knowledge: pretrained models**
4. **Effectiveness: contextual semantic parsing**
   - Formal context in ThingTalk
   - Fine-tune a pretrained language model

Synthesize Variety in Training Data

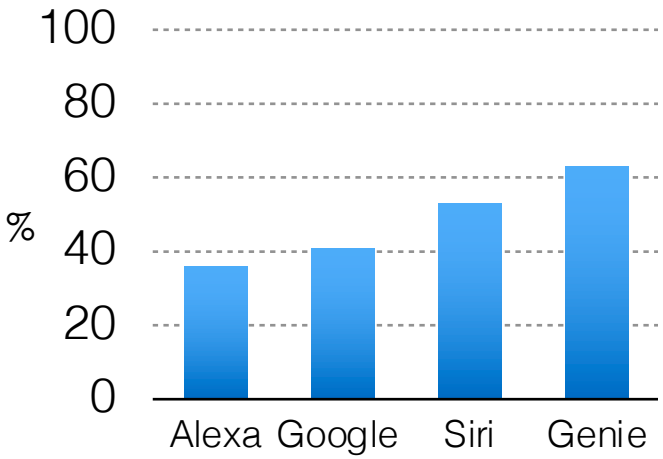


ThingTalk in a Virtual Assistant

1. The first agent with a contextual neural semantic parser
2. Direct execution made possible by ThingTalk
3. Agent policy to control the response



Comparison with Commercial Assistants



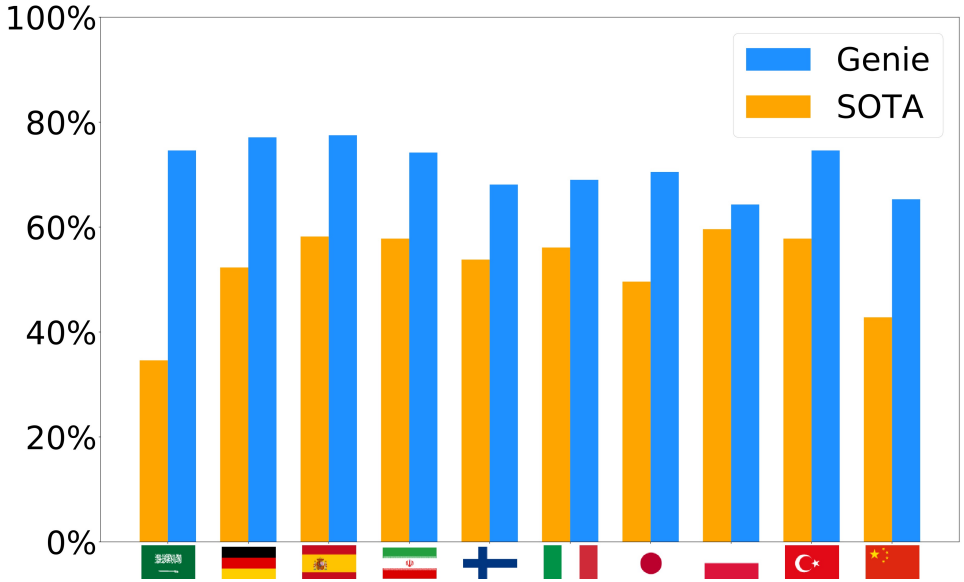
Long-tail Restaurant Questions

Examples of Long-Tail Questions	Alexa	Google	Siri	Genie
Show me restaurants rated at least 4 stars with at least 100 reviews				✓
Show restaurants in San Francisco rated higher than 4.5		✓		✓
What is the highest rated Chinese restaurant near Stanford?			✓	✓
How far is the closest 4 star restaurant?				✓
Who works for W3C and went to Oxford?				✓
Who worked for Google and lives in Palo Alto?				✓
Who graduated from Stanford and won a Nobel prize?	✓	✓		✓
Who worked for at least 3 companies?				✓
Show me hotels with checkout time later than 12PM				✓
Which hotel has a pool in this area?		✓	✓	✓

Multi-Lingual Assistants

Language	Restaurant Queries with Localized Entities
	look for 5 star restaurants that serve burgers
	ابحث عن مطاعم 5 نجوم التي تقدم الشاورما
	suchen sie nach 5 sterne restaurants, die maultaschen servieren
	busque restaurantes de 5 estrellas que sirvan paella valenciana
	به دنبال رستوران های 5 ستاره باشید که جوجه کباب سرو می کنند
	etsi 5 tähden ravintoloita, joissa tarjoillaan karjalanpiirakkaa
	cerca ristoranti a 5 stelle che servono bruschette
	寿司を提供する5つ星レストランを探す
	poszukaj 5 gwiazdkowych restauracji, które serwują kotlet
	köfte servis eden 5 yıldızlı restoranları arayın
	搜索卖北京烤鸭的5星级餐厅

Multi-Lingual Question Answering with Local Entities in 1 Day



Genie

Train parser with
auto-translated data
+ 2% manually translation

SOTA

On-the-fly translation

Localizing Open-Ontology QA Semantic Parsers in a Day Using Machine Translation
Mehrad Moradshahi, Giovanni Campagna, Sina J. Semnani, Silei Xu, Monica S. Lam
In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, November 2020.

Dialogue Accuracy

MultiWOZ 3.0 Annotated in ThingTalk

Restaurant Hotel Taxi Train Attraction

Created by Adrien Coquet from Noun Project

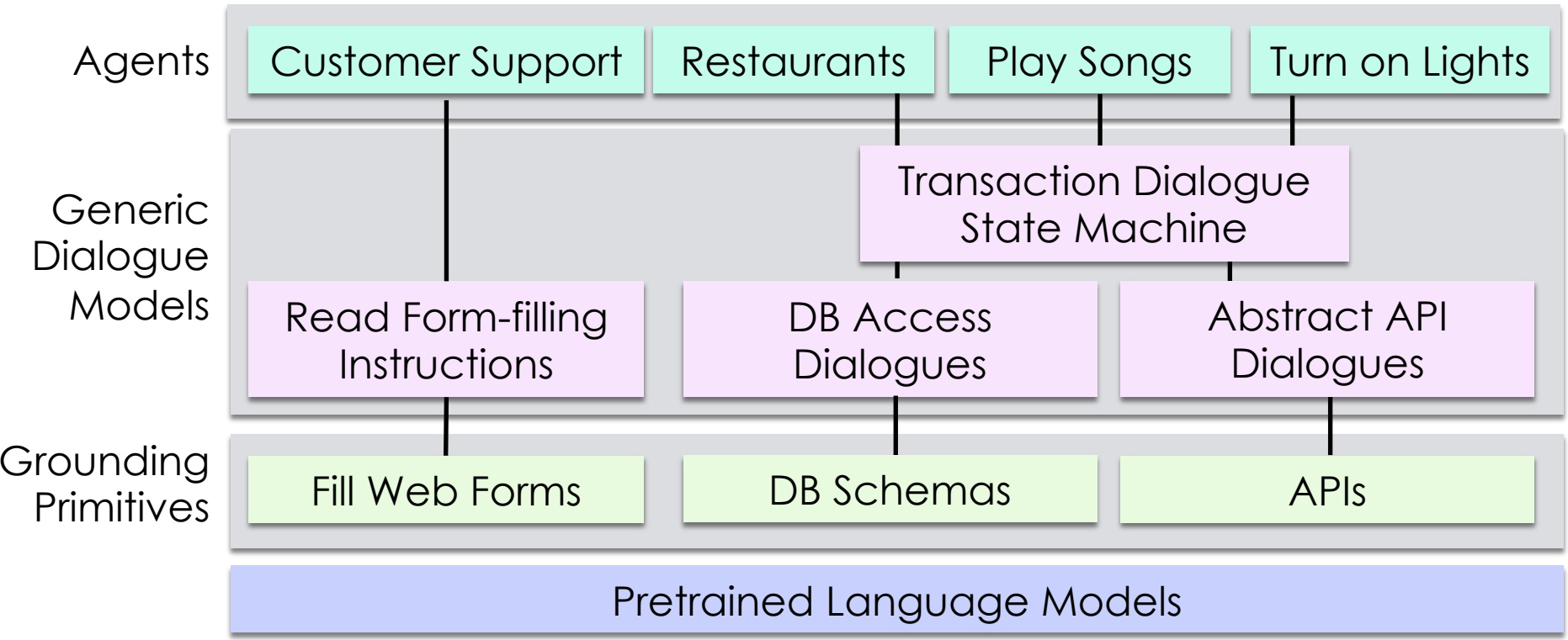
Created by dDara from Noun Project

Manual Annotations: 2% of original

Turn-by-turn accuracy: 80%

State-Machine-Based Dialogue Agents with Few-Shot Contextual Semantic Parsers
Giovanni Campagna, Sina J. Semnani, Ryan Kearns, Lucas Jun Koba Sato, Monica S. Lam
ArXiv Preprint, 2021

Genie: Open Pretrained Assistant



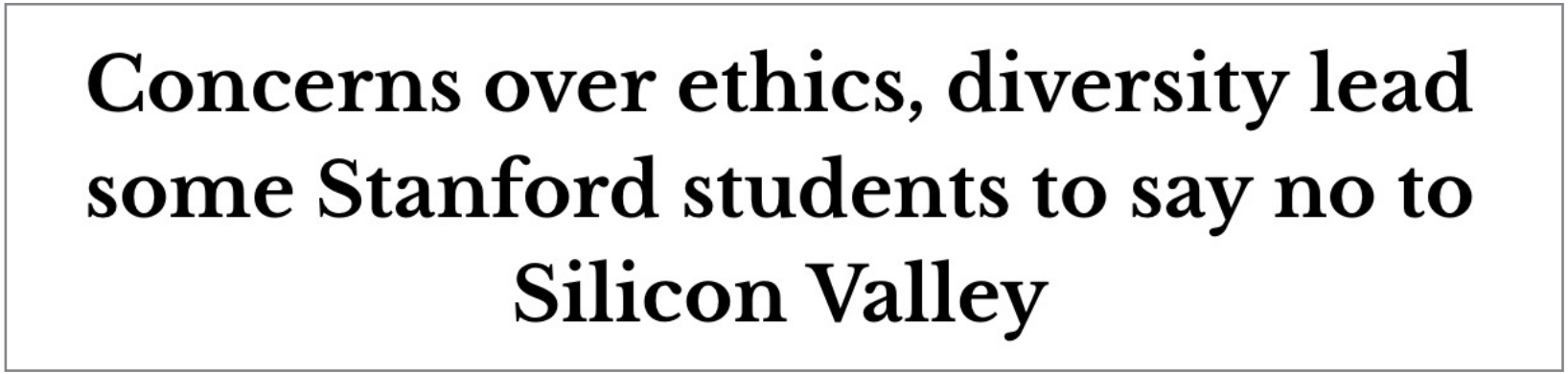
Key Takeaways

- **WWvW Needs a New Deep-Learning Approach**
  - To ground pretrained language models with semantics
- **Scaling WWvW**
  - **Completeness**: ThingTalk programming language
  - **Coverage** of training data: Synthesis
  - **NL knowledge**: pretrained models
  - **Effectiveness**: contextual semantic parsing

Lecture Outline

1. Motivation of the Course
2. Core Concept: Understanding Task-Oriented Dialogues
3. **Technology to Address Ethical Considerations**
4. This Course

On Stanford Daily Today



Stanford Daily, Matthew Turk on September 19, 2021

Emerging Smart Speaker Duopoly

- **Alexa (70% of the US market), Google Assistant**
- **Alexa's first-party skills cover the top functions**
  - Play music, news, reminders, timers, answer questions
  - Generic IoT control (60K compatible IoTs)
  - Purchase from Amazon
- **Alexa has 100K 3rd-party skills (chatbots)**
  - Ask Capitol One for bank balance
  - Ask Pizza Hut to place an order
  - Ask Anova to help me cook steak
  - Ask The Bartender what's in a Manhattan
- **Alexa is building an open but *proprietary* voice web.**

https://voicebot.ai/2019/08/09/us-smart-speaker-installed-base-reaches-76-million-according-to-cirp-with-52-one-year-growth/

Threat of the Assistant Duopoly

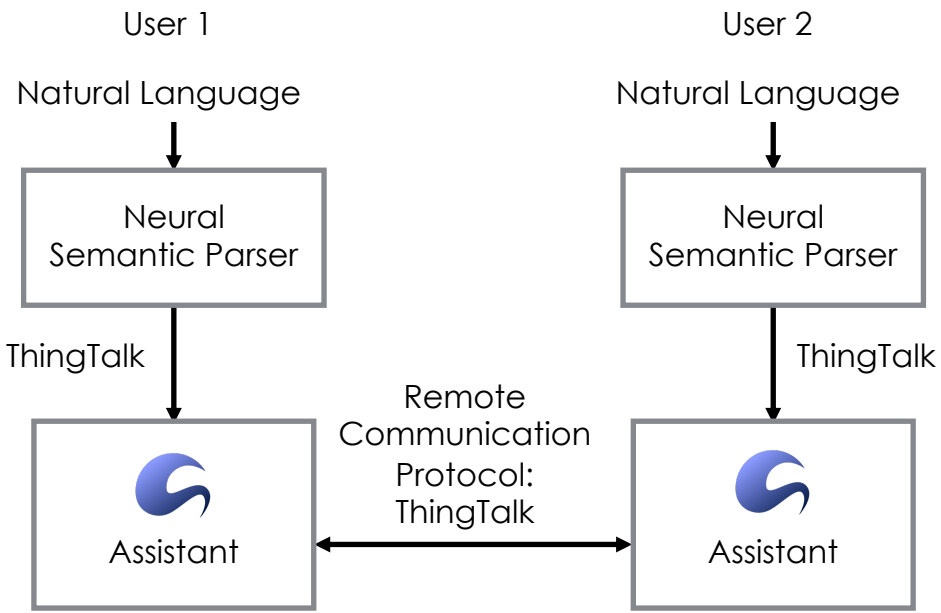
- **Proprietary web: Walled garden**
  - **Disallow services**
  - **Charge fees**
    - e.g. 15-30% on Apple app store or Google Play store
  - **Companies lose user relationships to the platform**
    - e.g. Media companies in Facebook
- **Privacy:** Personal information across many accounts in 2 large companies
- **Society**: Innovation, privacy, low-resource languages, non-profit causes

Solution (1): Decentralized WWvW

- Democratize technology with tools to create agents easily
- Encourage collaboration in the open through standards
- Pretrained agents to be attached to websites (www.x.com/well-known/wwvw)

The WWvW cannot be built by a single company

Solution (2): Private Decentralized Assistant



Privacy	Option to run on local devices
Choice of Vendors	Interoperability supported by a secure & powerful communication protocol
Sharing with Privacy	Share what the assistant can do Access control in NL *"My dad can access my security camera only if I am not home."* Enforced with SMT theorem prover

Campagna, G. et al. (Ubicomp 2018)

Key Takeaways

- **WWvW Needs a New Deep-Learning Approach**
  - To ground pretrained language models with semantics
- **Scaling WWvW**
  - **Completeness**: ThingTalk programming language
  - **Coverage** of training data: Synthesis
  - **NL knowledge**: pretrained models
  - **Effectiveness**: contextual semantic parsing
- **Technology for Ethics**
  - **Decentralized WWvW**
  - **Private decentralized assistant**

Lecture Outline

1. Motivation of the Course
2. Core Concept: Understanding Task-Oriented Dialogues
3. Technology to Address Ethical Considerations
4. **This Course**

This Course

- **The most recent research results in conversational virtual assistants**
  - The best ideas in the field: Not a survey on who did what
    - Genie: the first virtual assistant with a contextual dialogue semantic parser
    - Key ideas related to conversation agents in the latest literature
    - Ongoing research topics and ideas
  - Commercial practice
- **The first NLP lecture course to use the Genie toolset**
  - You can build a new assistant in a homework (1 week) without a large annotated data set
  - You can do a state-of-the-art conversational agent project in a course project (4 weeks)
- **Will scale to a full audience next year**

Learning Goals of This Course

1. **Understand the theory and practice of the state of the art of dialogue agents**
   - Written assignments for the theory
   - 3 programming assignments to build a fully working dialogue agent (groups of 2)
     - Learn the tools and workflow
     - Assess the performance of the state of the art
     - Get inspirations for your own project

Learning Goals of This Course

1. **Understand the theory and practice of the state of the art of dialogue agents**
2. **Active learning: a research project you propose** (groups of 2)
   Examples:
   - Improve the data synthesis pipeline, neural model
   - Propose a representation extension
   - Propose a new dialogue state machine
   - Improve response generation, internalizational, …
   - Create a challenging dialogue agent

Learning Goals of This Course

1. **Understand the theory and practice of the state of the art of dialogue agents**
2. **Active learning: a research project you propose**
3. **Technology for positive social impact: privacy, open access**

Tentative Schedule (Part 1)

Introduction	This lecture
	Anatomy of a virtual assistant
Basic NLP	Seq2seq neural models for NLP
	Pretrained networks
Semantic parsers for questions	Question-answering agents
	Training data synthesis
Semantic parsers for dialogues	Dialogue semantics
	Data acquisition for dialogues
	Transactional agent generation

Tentative Schedule (Part 2)

Other components in a virtual assistant	Multi-lingual assistants
	Response generation
	Error recovery
	Named entity disambiguation
	Multimodal assistants
Privacy & Fair competition	Decentralized assistants with interoperability
Related technologies	Free-text question answering
	Chatty dialogues
	Speech-to-text, text-to-speech

Grading

- Participation: 10%
- Homeworks: 25%
- Examination: 30%
- Final project: 35%

You have two grace days for late homeworks in the quarter.