CS 294S/294W
Democratizing Virtual Assistants
A Social-Good Research Project Course

Monica Lam
Stanford University
lam@cs.stanford.edu
Why a Remote Research Course?

A welcomed change from Zoom lectures.

Expose students to the exciting world of research.
This Class

1. Introduce an exciting research agenda
2. Explain the course design
3. Overview of the new methodology
4. Suggest research topics
5. Gather initial interest / Get to know each other
Exciting Time to Do CS Research

Computers get a new interface: Voice!

Talking Wikipedia
- General knowledge Q&A in all languages
- Add meaning to pretrained NL models

Pervasive Dialogue Agents
- A new software development toolset
- 20M web developers $\rightarrow$ 20M NL developers!

End-user NL Programming
- Consumers/professionals automate their tasks
- Long-tail programming
OVAL: An Open-Source Initiative

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

Sponsors
NSF    Alfred P. Sloan Foundation    Stanford Human-centered AI

Computer Science Faculty
Michael Bernstein  Dan Boneh  Monica Lam  James Landay
Fei-fei Li  Chris Manning  David Mazieres  Chris Re

Philanthropy & Digital Society  Internet & Society Center
Lucy Bernholz  Jen King

Students
Giovanni Campagna  Michael Fischer  Ranjay Krishna
Mehrad Moradshahi  Sina Semnani  Silei Xu  Jackie Yang

The New York Times
An Open-Source Virtual Assistant Platform

**GENIE**
Virtual Assistant 2.0 Tools

Today:
Affordable only by the largest companies (Alexa: 10K employees)

Goal:
Democratize with affordable methodology & effective toolsets

---

**THINGPEDIA**
Crowdsourced Skill Repository

Today:
Proprietary voice web (Alexa: 100K 3rd party skills)

Goal:
Inter-operable skills open to all virtual assistants

---

**ALMOND**
Privacy-protecting assistant

Today:
Virtual assistants are ultimate surveillance tools

Goal:
A federated virtual assistant architecture that allows local execution.

Opportunities for many AI, HCI, Systems Research Projects
This Year’s Infrastructure Goal

• An open privacy-preserving virtual assistant with the top 10 skills
• Experimental research platform
• An alternative for consumers (like Firefox)
• To be released in June 2021
This Class

1. Introduce an exciting research agenda
2. Explain the course design
3. Overview of the new methodology
4. Suggest research topics
5. Gather initial interest / Get to know each other
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 give mini-lectures on their research topic
    (an important part of research training)

• Final project presentation and report
# A Tentative Schedule

<table>
<thead>
<tr>
<th>Week</th>
<th>Tuesday</th>
<th>Thursday</th>
<th>Due (10:30am)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sep 15, 17</td>
<td>Course Introduction</td>
<td>Schema → Q&amp;A (HW)</td>
<td>9/17: Student profile</td>
</tr>
<tr>
<td>Sep 22, 24</td>
<td>Schema → Dialogues</td>
<td>Project Discussions</td>
<td>9/24: HW due</td>
</tr>
<tr>
<td>Sep 29, Oct 1</td>
<td>Project Discussions</td>
<td>NL Primer</td>
<td></td>
</tr>
<tr>
<td>Oct 6, 8</td>
<td>Proposals</td>
<td>Proposals</td>
<td>10/6: Project Proposal</td>
</tr>
<tr>
<td>Oct 13, 15</td>
<td>Group Meetings</td>
<td>Students’ Mini-lectures</td>
<td></td>
</tr>
<tr>
<td>Oct 20, 22</td>
<td>Group Meetings</td>
<td>Students’ Mini-lectures</td>
<td>10/19: Weekly Update</td>
</tr>
<tr>
<td>Oct 27, 29</td>
<td>Group Meetings</td>
<td>Students’ Mini-lectures</td>
<td>10/26: Weekly Update</td>
</tr>
<tr>
<td>Nov 3, 5</td>
<td>Group Meetings</td>
<td>Students’ Mini-lectures</td>
<td>11/2: Weekly Update</td>
</tr>
<tr>
<td>Nov 10, 12</td>
<td>Group Meetings</td>
<td>Students’ Mini-lectures</td>
<td>11/9: Weekly Update</td>
</tr>
<tr>
<td>Nov 17, 19</td>
<td>Final Project Presentation</td>
<td>Final Project Presentation</td>
<td>11/20: Project Report</td>
</tr>
</tbody>
</table>
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%
This Class

1. Introduce an exciting research agenda
2. Explain the course design
3. Overview of new methodology
4. Suggest research topics
5. Gather initial interest / Get to know each other
Paradigm Shift

Existing approach

1. Hand-annotated training data
   • Coverage, compositionally, cost, correctness
   • Alexa: 10,000 employees

Virtual Assistant 2.0

1. Mostly synthesized training data, using pretrained language models

2. Brittle dialogue trees
   Intent classifier per utterance

2. High-level programming
   One contextual neural network
Can you help with information regarding a food place? I need to book at 15:45.

How about the restaurant with name La Tasca and Italian food?

Can you find something which serves seafood?

What date are you looking for? Thursday please.

How about the Copper Kettle? It is a food place with seafood food.

What is the price range and the area?

The Copper Kettle is a moderately priced restaurant in the north of the city.

Would you like a reservation? No, thanks.

Can I help with you anything else? Thank you, that will be it for now.
Contextual Pure-Neural Semantic Parser

Exec: query Restaurant, food==“Indian” & & …

Pointer-Generator

Attention

Pool

BERT Encoder

Feed Forward

LSTM

Embedding

Utterance

Does it serve Indian food?

Result State

Restaurant(food = …)

Agent State

RecommendOne

User State

(Autoregressive)

Exec: query Restaurant …
# Dialogue State Tracking

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joint Accuracy (MultiWOZ 2.1)</td>
<td></td>
</tr>
<tr>
<td>TRADE (Wu et al., 2019)</td>
<td>45.6</td>
</tr>
<tr>
<td>SUMBT (Lee et al., 2019a)</td>
<td>46.7</td>
</tr>
<tr>
<td>DSTQA (Zhou and Small, 2019)</td>
<td>51.2</td>
</tr>
<tr>
<td>DST-Picklist (Zhang et al., 2019a)</td>
<td>53.3</td>
</tr>
<tr>
<td>SST (Chen et al., 2020)</td>
<td>55.2</td>
</tr>
<tr>
<td>TripPy (Heck et al., 2020)</td>
<td>55.3</td>
</tr>
<tr>
<td>SimpleTOD (Hosseini-Asl et al., 2020)</td>
<td>55.7</td>
</tr>
<tr>
<td>Turn-By-Turn Accuracy (Cleaned Test Set)</td>
<td>Genie</td>
</tr>
</tbody>
</table>

Genie

- Trained with only synthesized data
- Perfect annotations
- Validate and test with real data
- Need to track only the user state, one turn at a time
Answering Complex Questions

<table>
<thead>
<tr>
<th>Queries</th>
<th>Alexa</th>
<th>Google</th>
<th>Siri</th>
<th>Genie</th>
</tr>
</thead>
<tbody>
<tr>
<td>Show me restaurants rated at least 4 stars with at least 100 reviews</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Show restaurants in San Francisco rated higher than 4.5</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>What is highest rated Chinese restaurant in Hawaii?</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>How far is the closest 4 star and above restaurant?</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Find a W3C employee that went to Oxford</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Who worked for both Google and Amazon?</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Who graduated from Stanford and won a Nobel prize?</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Who worked for at least 3 companies?</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Show me hotels with checkout time later than 12PM</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Which hotel has a swimming pool in this area?</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>
New-Generation HCI: Voice

<table>
<thead>
<tr>
<th>FRONT ENDS</th>
<th>INTERFACE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Menus Forms</td>
<td>Hardcoded</td>
</tr>
<tr>
<td>Keyword Search</td>
<td>Hardcoded</td>
</tr>
<tr>
<td>NL Automation</td>
<td>Compiled</td>
</tr>
<tr>
<td>NL Dialogues</td>
<td>Interactive program</td>
</tr>
<tr>
<td>DATABASE</td>
<td></td>
</tr>
<tr>
<td>Database</td>
<td></td>
</tr>
<tr>
<td>API Calls</td>
<td></td>
</tr>
<tr>
<td>FAQs</td>
<td></td>
</tr>
<tr>
<td>Free Text</td>
<td></td>
</tr>
</tbody>
</table>

NL Automation (User driven)
- Turn on the lights
- When apple stock drops to $100, buy 3 shares
- Find a Spanish restaurant that is open at 10pm in Palo Alto

NL Dialogues
- User-driven: reservations
- 2-way: doctor appts
- Agent-driven: Online teaching
MVC (Model View Controller) → MRP (Model Response Parser)
This Class

1. Introduce an exciting research agenda
2. Explain the course design
3. Overview of the new methodology
4. Suggest research topics
5. Gather initial interest / Get to know each other
## Research Projects

<table>
<thead>
<tr>
<th>Problem</th>
<th>Area</th>
<th>Goal</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wikidata in NL</td>
<td>Systems</td>
<td>Scalability</td>
<td>Develop methodology &amp; tools to cover Wikidata</td>
</tr>
<tr>
<td>AI</td>
<td>Scalability</td>
<td></td>
<td>Zero-shot learning using type information</td>
</tr>
<tr>
<td>Usable Dialogue Agents</td>
<td>AI</td>
<td>Breadth</td>
<td>Generalize a contextual neural network from 5 (Multiwoz) to 11 domains (SGD)</td>
</tr>
<tr>
<td>(Transactions)</td>
<td>Accuracy</td>
<td></td>
<td>Named entity disambiguation in the wild (Bootleg)</td>
</tr>
<tr>
<td></td>
<td>Error detection</td>
<td></td>
<td>Neural network to identify likely correct components</td>
</tr>
<tr>
<td></td>
<td>Response fluency</td>
<td></td>
<td>Use Bart to generate fluent responses</td>
</tr>
<tr>
<td></td>
<td>Multilingual:</td>
<td>Localization</td>
<td>Use machine translation with entities in target languages</td>
</tr>
<tr>
<td></td>
<td>Localization</td>
<td></td>
<td>(Chinese Multiwoz, CrossWoz)</td>
</tr>
<tr>
<td>HCI</td>
<td>Usability</td>
<td></td>
<td>Conversational Q&amp;A dialogue design for music, movies, etc</td>
</tr>
<tr>
<td></td>
<td>Design</td>
<td></td>
<td>Dialogue to support function discovery</td>
</tr>
<tr>
<td></td>
<td>Multimodal</td>
<td></td>
<td>Combining the best of voice and text in assistants</td>
</tr>
<tr>
<td></td>
<td>Systems</td>
<td>Knowledge</td>
<td>Representation (time, location)</td>
</tr>
</tbody>
</table>
## Multi-disciplinary Research Projects

<table>
<thead>
<tr>
<th>Problem</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced Agents</td>
<td>Generic FAQ dialogue models</td>
</tr>
<tr>
<td></td>
<td>Personalized agents with users’ history &amp; profile (e.g. ordering food)</td>
</tr>
<tr>
<td>End-user programming</td>
<td>A gentle way to introduce end-users to creating skills: cron jobs, monitors, comparison shopping</td>
</tr>
<tr>
<td></td>
<td>Automate end-user routines with demonstrations (e.g. workout assistants)</td>
</tr>
<tr>
<td>End-to-end skills</td>
<td>Home Automation.</td>
</tr>
<tr>
<td></td>
<td>IoTs for 1000 devices (with tens of abstract devices)</td>
</tr>
<tr>
<td></td>
<td>Almond is the voice interface for Home Assistant</td>
</tr>
<tr>
<td></td>
<td>News, sports, radios, podcasts: Listening + asking questions</td>
</tr>
<tr>
<td></td>
<td>Safe voting, legal advice, personal finance</td>
</tr>
</tbody>
</table>