

# Building a Socialbot: Lessons Learned from 10M Conversations

---

Mari Ostendorf

& the Sounding Board Team  
*University of Washington*





# The Sounding Board Team

## Students

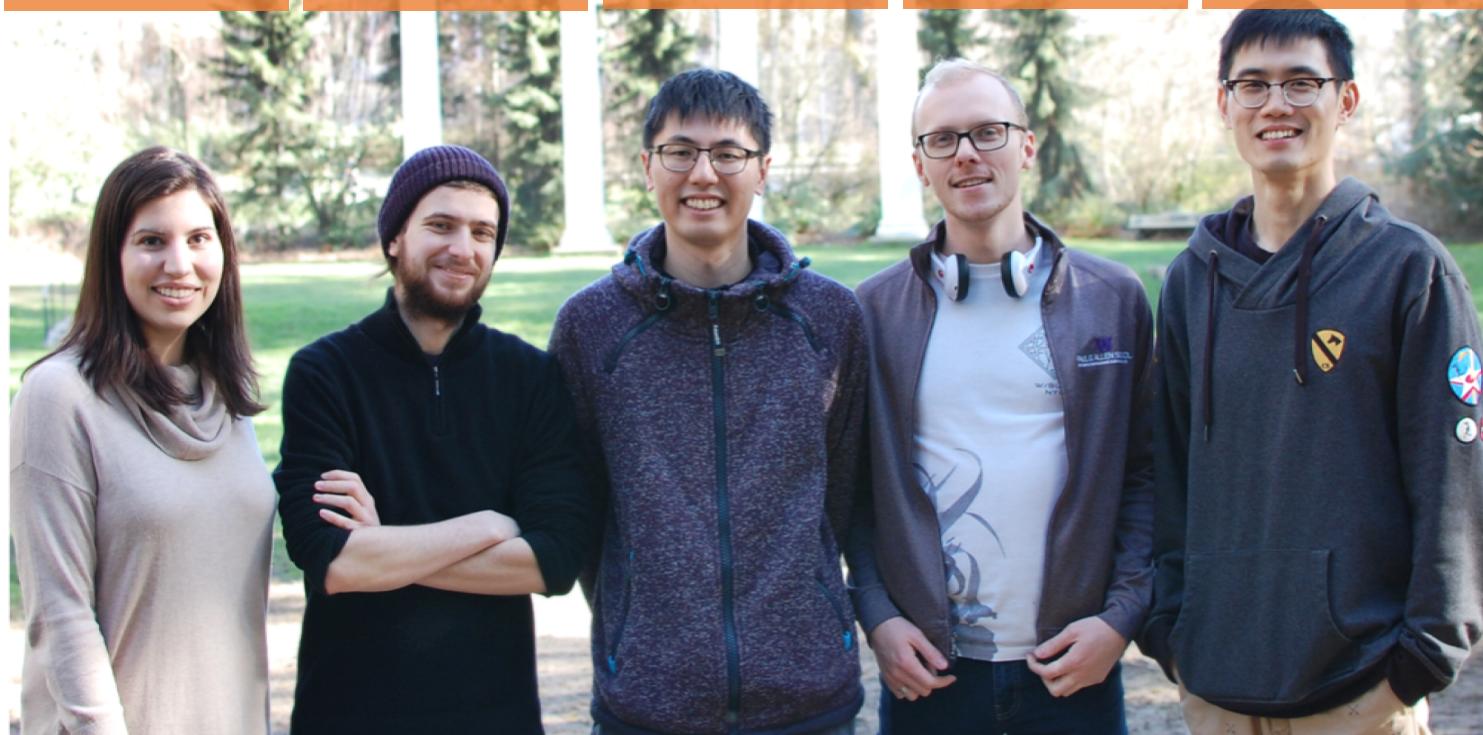
Elizabeth  
Clark  
CSE

Ari  
Holtzman  
CSE

Hao  
Fang  
EE

Maarten  
Sap  
CSE

Hao  
Cheng  
EE



## Faculty Advisors



Mari Ostendorf, EE



Yejin Choi - CSE - Noah Smith





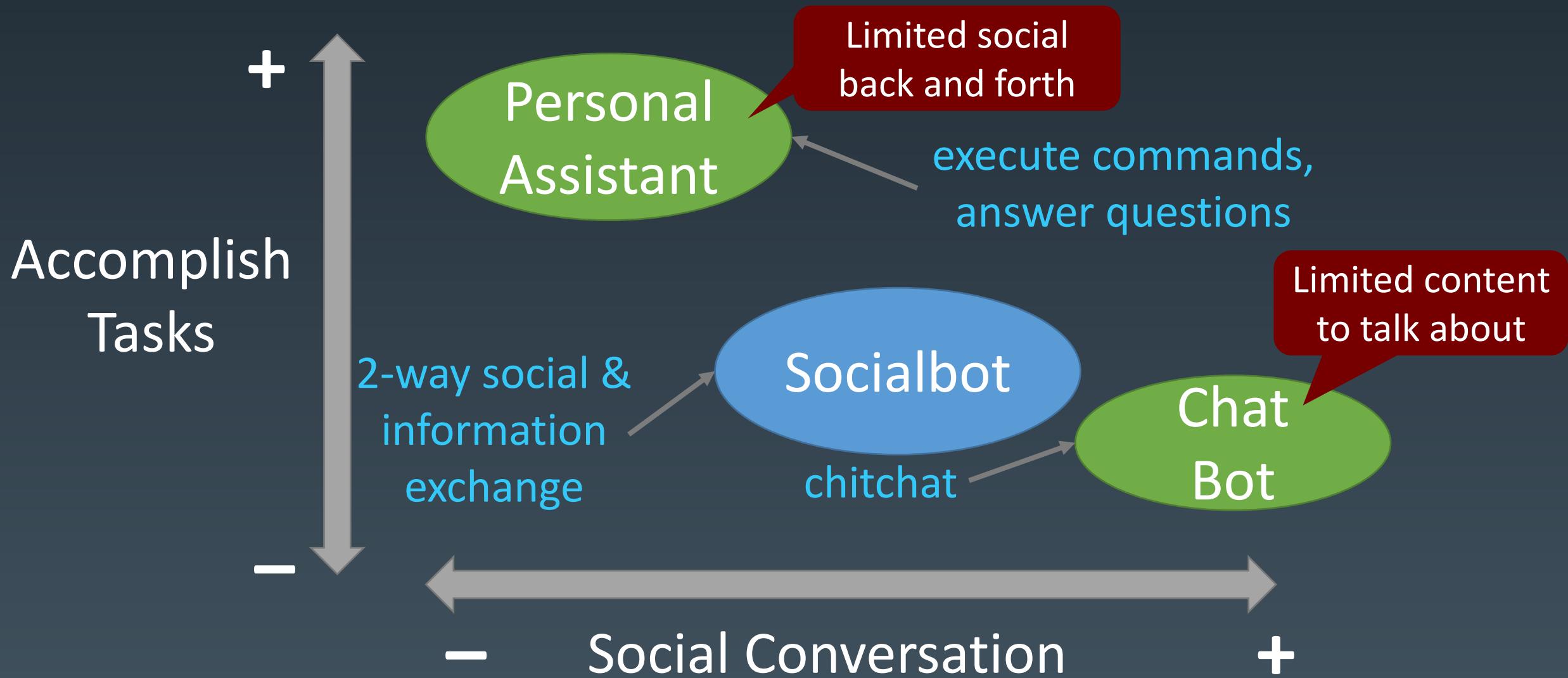
Teams of university students try to build a **socialbot** that converses coherently and engagingly with humans on popular topics and current events.

# Roadmap

- Context: Our view of a socialbot
- The adventure of building Sounding Board
- What we learned from 10M conversations
- Lessons for academic-industry partnerships

# The Socialbot as a Conversational Gateway

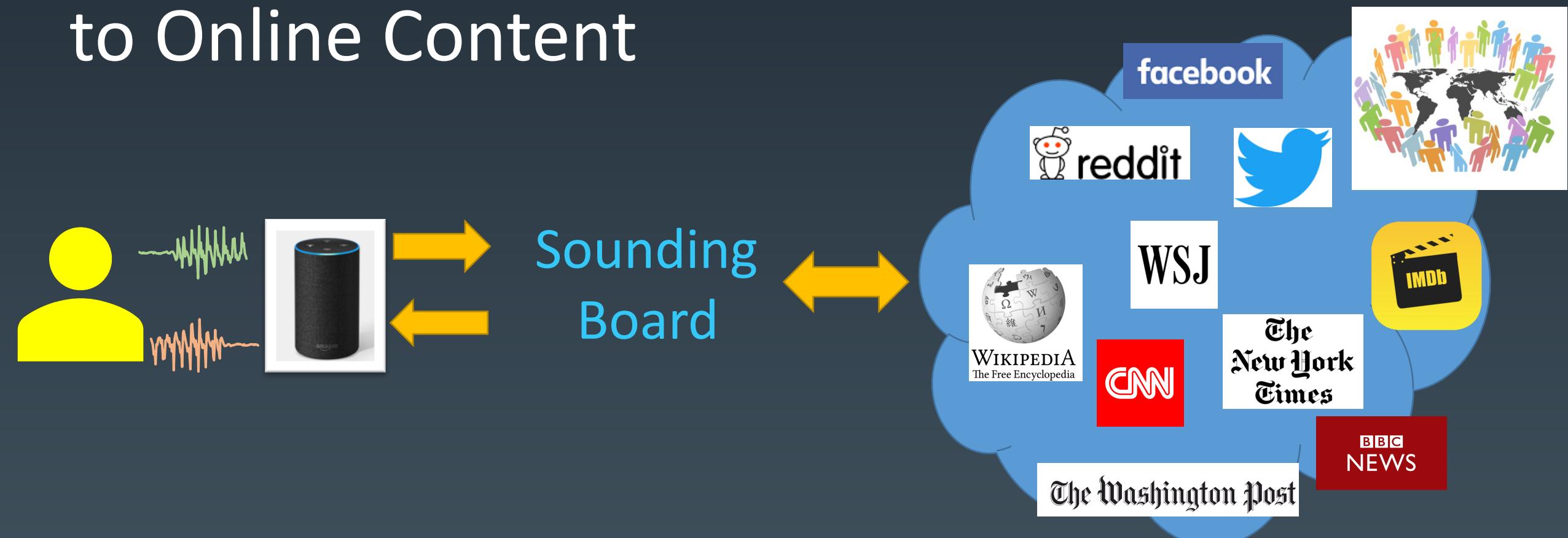
# Types of Conversational AI Systems



# A Perspective on Socialbots

- A socialbot facilitates evolving user goals & priorities
- Users (should) know they are talking to a bot
- Broad applications
  - *Education: language learning, tutoring systems*
  - *Help desk, information exploration*
  - *Exercise/therapy coach, companion*

# Sounding Board: A Conversational Gateway to Online Content



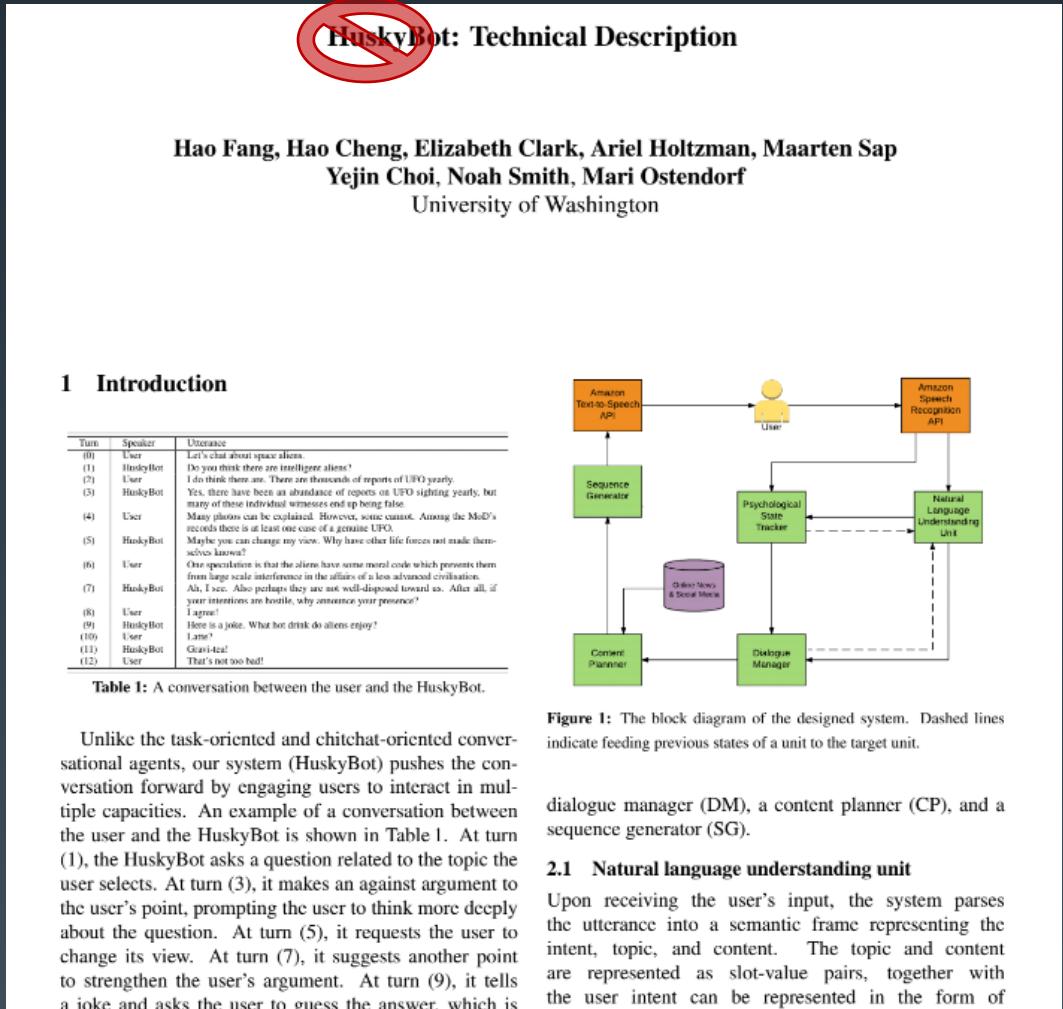
# The Adventure: The Sounding Board Design

- Early lessons
- Design philosophy
- Brief system overview\*
- Evaluation

\* For more info, check out the demo  
Monday 2pm, Elite Hall B.

# Original Goals

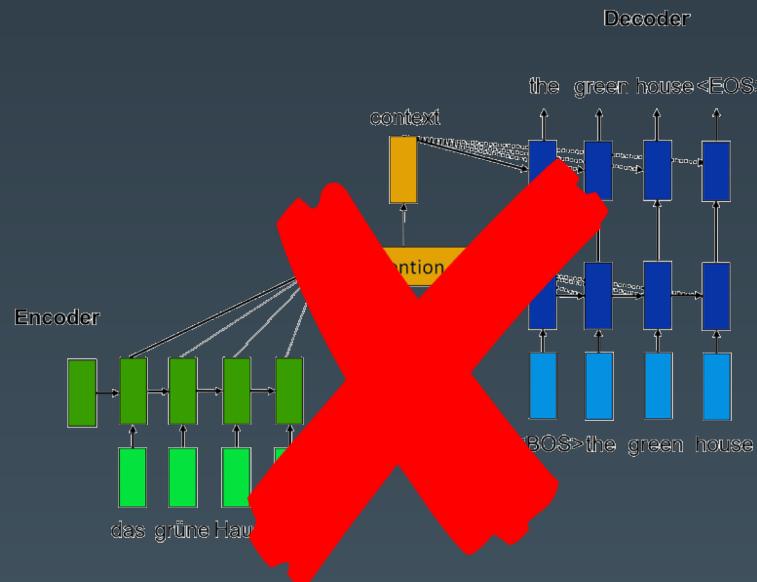
- Different interaction modes
  - Debate
  - Collaborative story writing
- User personality
  - Sophisticated user modeling
  - Personalized conversation
- End-to-end deep learning



# First Attempts

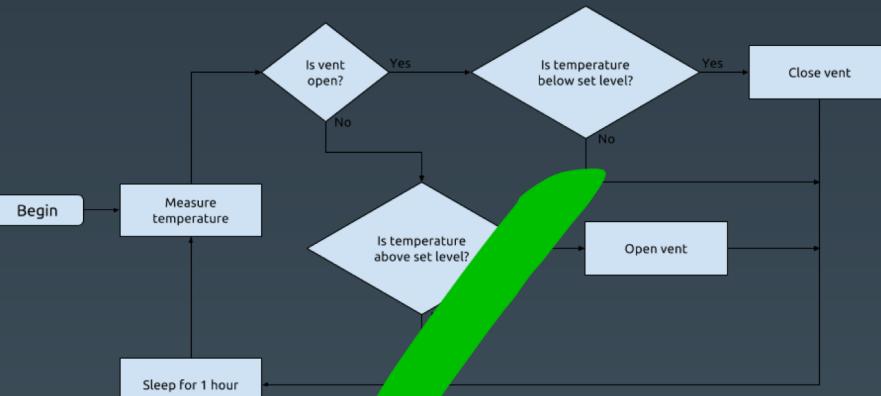
## Approach #1

A seq2seq bot trained on a large amount of carefully selected, pre-processed data



## Approach #2

A bare-bones, rule-based, low-content bot



# Early Stage Challenges

- Software:
  - No experience with Alexa skill kits, built-in tools are more for speech-enabling an existing app
  - No existing dialog system to build on
- Data:
  - Task is open domain & users want current content → there was no good existing data for end-to-end training
  - Our initial system was sufficiently bad, we didn't want to learn from early user conversations with it

Turn	User	Sounding Board
1	Let's chat	Hi! This is an Alexa Prize socialbot. How's your day?
2	I'm <u>five</u> .	Glad to hear it! We could talk about robots, batman, or superman. Say "next", to chat about other things.
3	Superman.	It looks like you wanna chat about superman. I was high up in the cloud when I realized: If people don't recognize Clark Kent as Superman because of his glasses, does his eye doctor know his true identity?
4	I guess so.	Did you know that Henry Cavill almost missed the call for the role of Superman cause he was playing "World of Warcraft"
5	Really, I didn't know that.	Weird, right? Speaking of superman, Did you see: Superman? It was released in 1997. It's a comedy.
6	Yes, it was hilarious.	I'm glad you feel this is hilarious. Which part do you like best about this movie?
7	The part when he met lewis leah.	Interesting. meccartin, and raffi co-directed this film. The movie has a 6.3 out of 10 on IMDB. which seems pretty good! do you like the movie's director?
8	Yeah, he's cool.	Cool! By the way, do you wanna know something interesting about raffi?
9	Okay, go ahead.	Super! My friend told me that Raffi is a worldwide children's rights activist and has fought against exploitative advertising to children.

# What Makes Someone a Good Conversationalist?

- Have something interesting to say
- Show interest in what your partner says

*These principles apply to a socialbot*

# Have something interesting to say

- Users react positively to learning something new

SpaceX sends beer ingredients to International Space Station just in time for Christmas

- ... and negatively to old or unpleasant news

Man Given 'Options' Before Cutting Dog's Head Off, Ga. Sheriff Says

Fort Lauderdale Pizza Hut Caught Refusing to Deliver to Black Neighborhood at Night

# Show interest in what the user says

- Users lose interest when they get too much content that they don't care about
- Users like acknowledgment of their reactions & requests
- Some users need encouragement to express opinions

Really, I didn't know that.	Weird, right? Speaking of superman, Did you see: Superman? It was released in 1997. It's a comedy.
Yes, it was hilarious.	I'm glad you feel this is hilarious. Which part do you like best about this movie?
The part when he met lewis leah.	Interesting. meccartin, and raffi co-directed this film. The movie has a 6.3 out of 10 on IMDB. which seems pretty good! do you like the

...but it can be annoying

This article mentioned Google.  
Have you heard of Google?

# Design Philosophy

- Content-driven



- Daily content mining, large & dynamic content collection
- Knowledge graph
- DM that promotes popular content, diverse sources (styles)

- User-centric



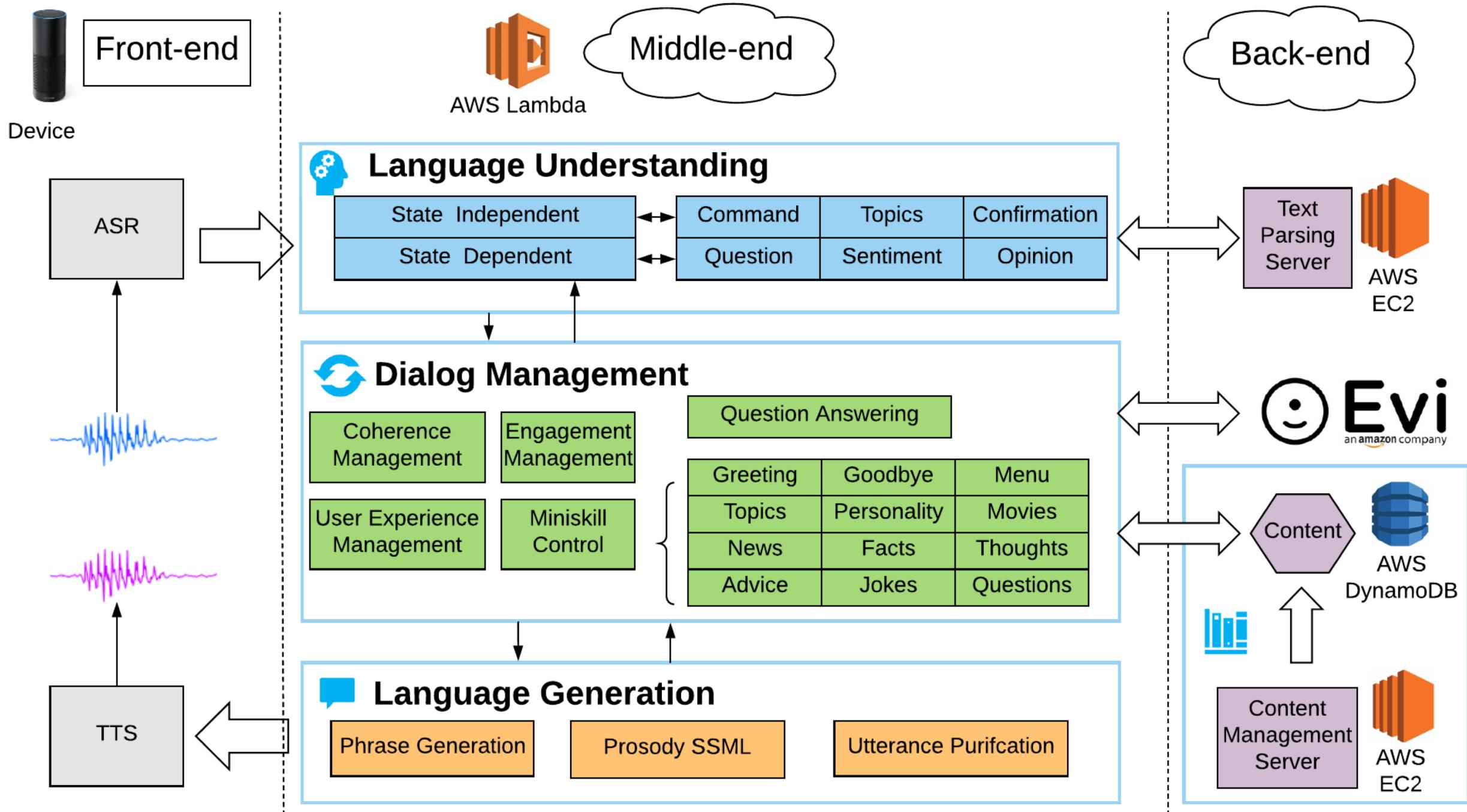
- Language understanding that detects user sentiment
- Dialog management (DM) that tries to learn user personality, handles rapid topic changes, tracks engagement, ....
- Language generation with prosody-appropriate grounding

# Prosody – What's that?

- It's not what you say, but how you say it
- Intonation, pausing, duration lengthening... (attributes of the acoustic signal)



- Which communicate
  - User intent, sentiment, sarcasm, ...
  - Socialbot empathy, enthusiasm, topic change,...



# Multi-dimensional NLU Representation

Commands

Tell me a joke.

Questions

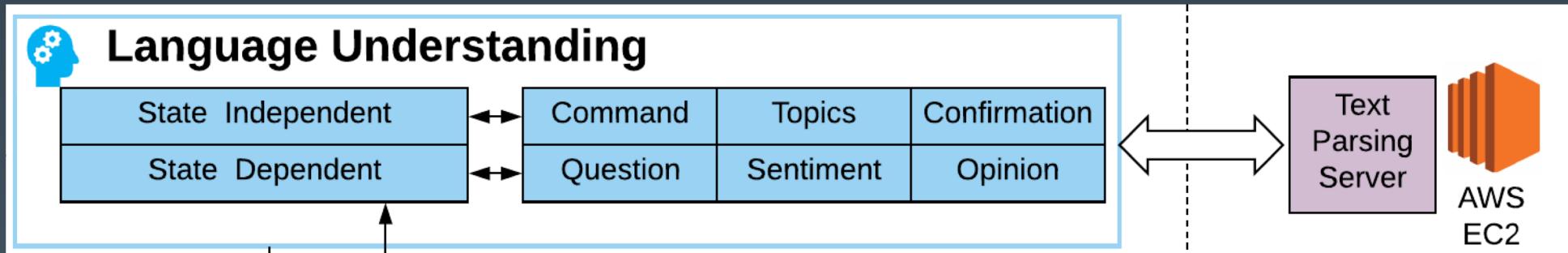
What is your favorite color?

Topics

Let's talk about technology.

User Reactions

That's really interesting!





# Hierarchical Dialog Management



- Master (Global)
  - Rank topics, miniskills, content
  - Consider: topic coherence, user engagement, content availability
- Miniskills (Local)
  - greeting / goodbye / menu
  - probe user personality
  - discuss a news article / movie
  - tell a fact / thought / advice / joke

Turn	User	Sounding Board
1	Let's chat	Hi! This is an Alexa Prize socialbot. How's your day?
2	I'm <u>five</u> .	Glad to hear it! We could talk about robots, batman, or superman. Say "next", to chat about other things.
3	Superman.	It looks like you wanna chat about superman. I was high up in the cloud when I realized: If people don't recognize Clark Kent as Superman because of his glasses, does his eye doctor know his identity?
4	I guess so.	Did you know that Henry Cavill almost missed the call for the role of Superman cause he was playing "World of Warcraft"
5	Really, I didn't know that.	Weird, right? Speaking of superman, Did you see: Superman? It was released in 1997. It's a comedy.
6	Yes, it was hilarious.	I'm glad you feel this is hilarious. Which part is about this movie?
7	The part when he met lewis leah.	Interesting. meccartin, and raffi co-directed this film. The movie has a 6.3 out of 10 on IMDB. which seems pretty good! do you like the movie's director?
8	Yeah, he's cool.	Cool! By the way, do you wanna know something interesting about raffi?
9	Okay, go ahead.	Super! My friend told me that Raffi is a worldwide children's rights activist and has fought against exploitative advertising to children.

Negotiation

Thought

Movie

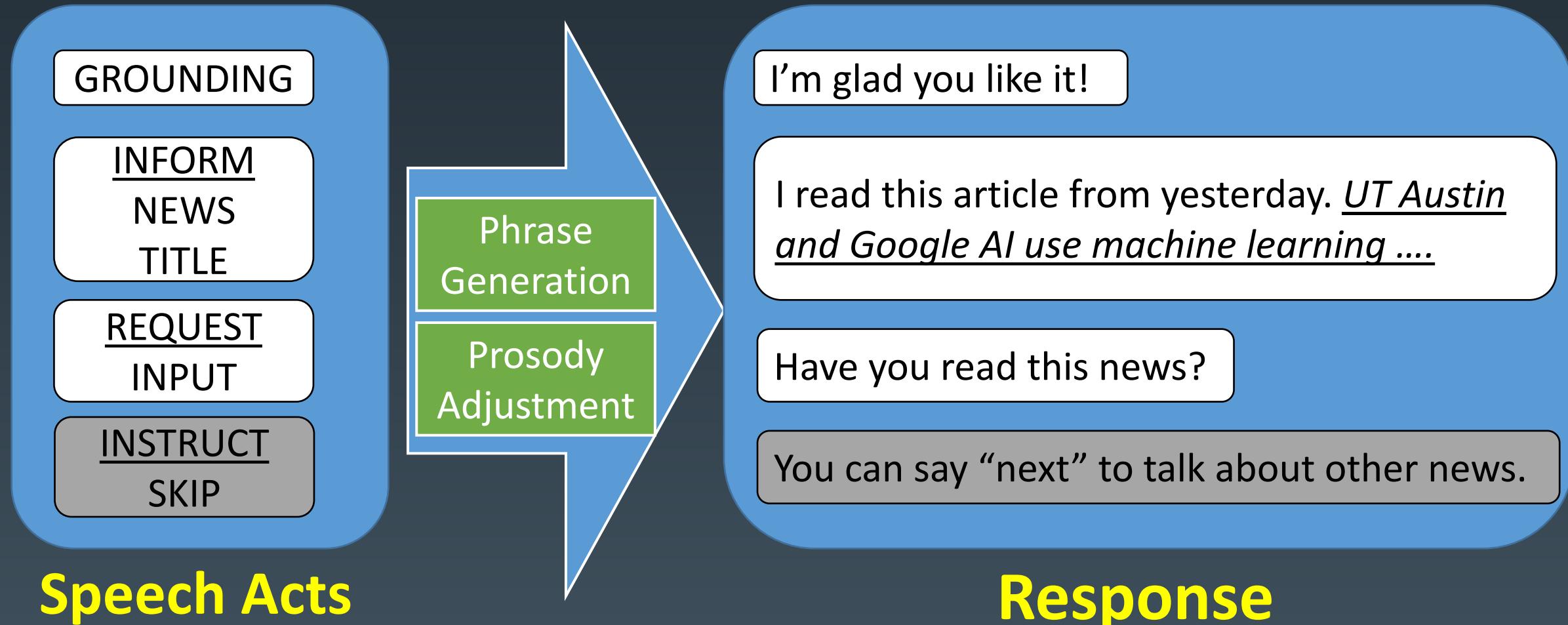
Fact

# User Personality

- User-centric topic suggestions
- Five-factor model (Costa&McCrae, 1992)
  - E.g., “Do you talk a lot?”
- Helps us understand how users interact with Sounding Board



# From Speech Acts to Natural Language



# Content Management

- Crawl online content
- Filter inappropriate & depressing content
- Index interesting & uplifting content
  - noun phrases, entities, meta-info
- Knowledge graph
  - daily updated
  - 80K entries, 300K topics

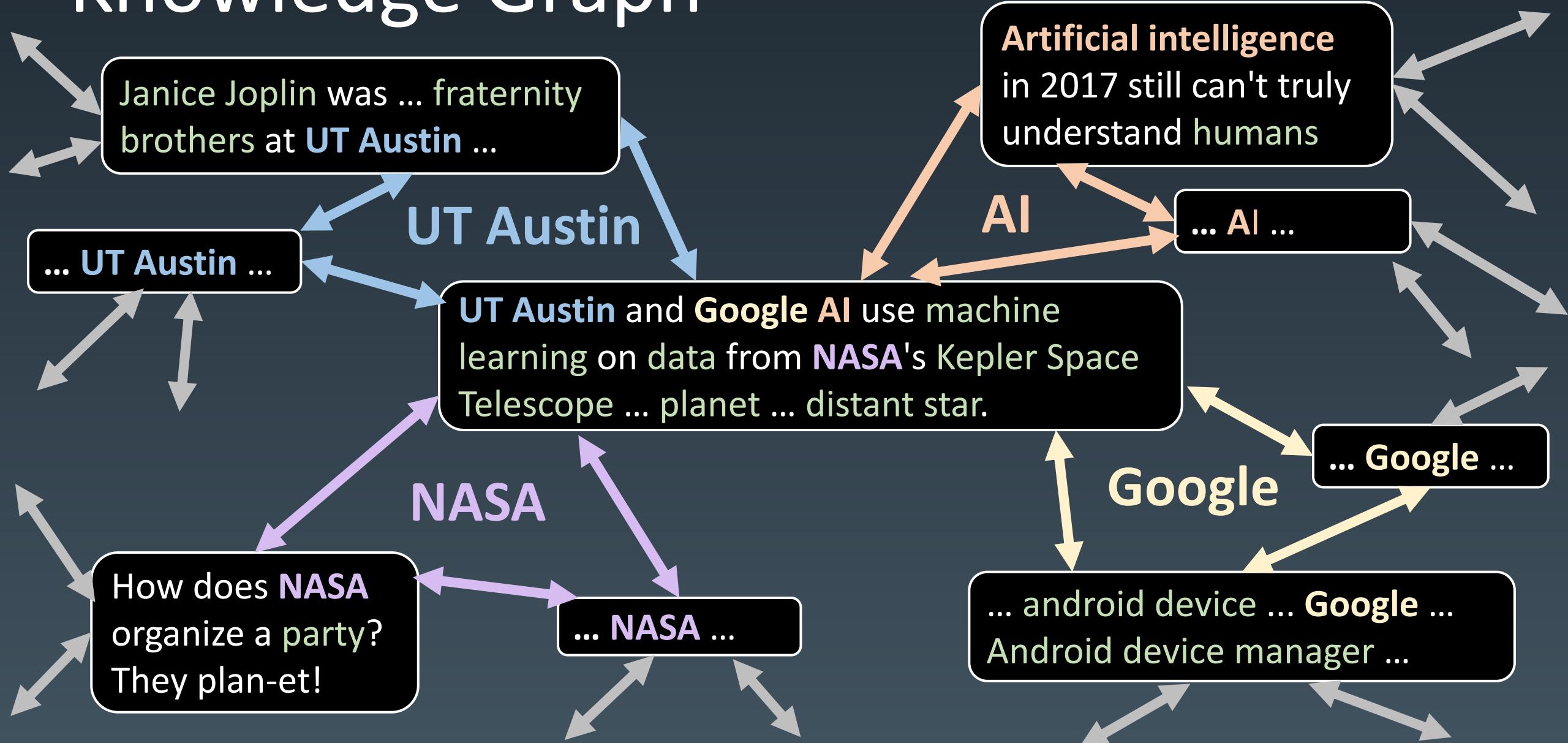


UT Austin and Google AI use machine learning on data from NASA's Kepler Space Telescope to discover an eighth planet circling a distant star.

science

astronomy

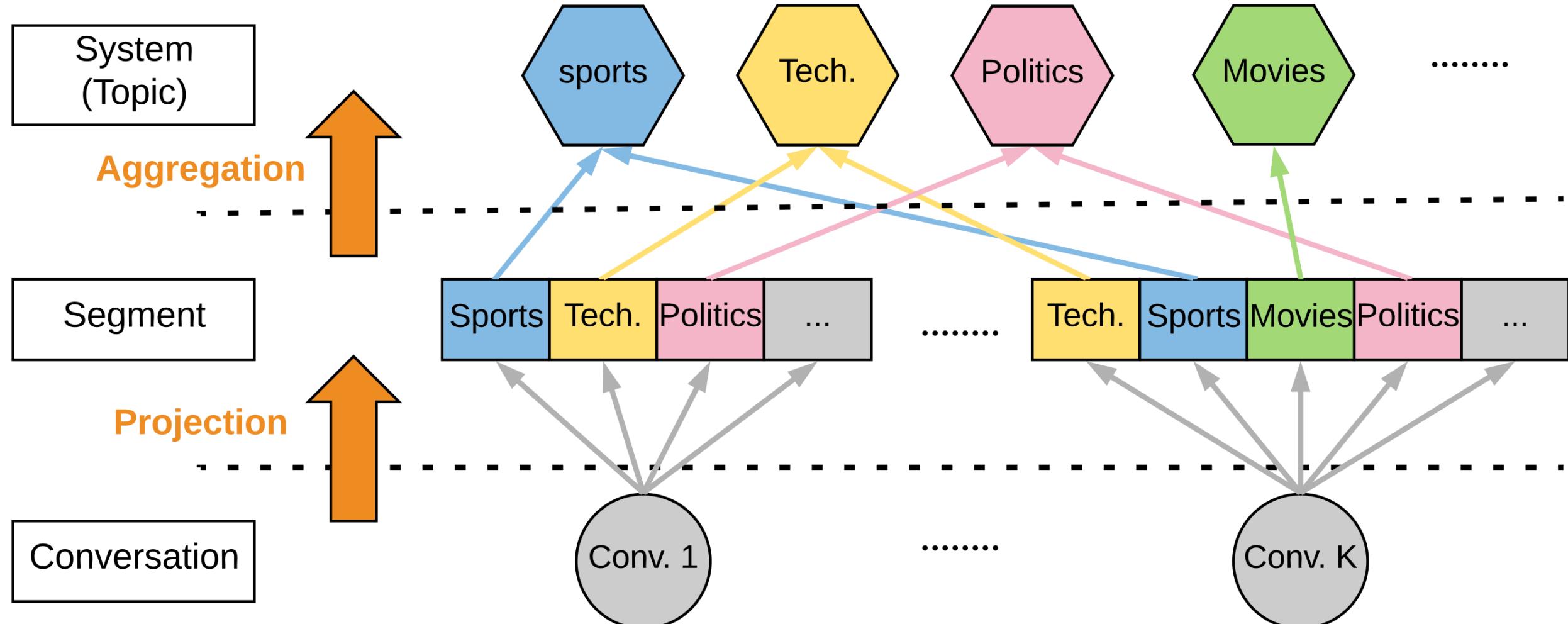
# Knowledge Graph



# Evaluation

- User ratings: expensive, noisy & sparse
  - users may decline to rate the system (43%)
  - user ratings have a very high variance ( $3.65 \pm 1.40$ )
  - conversations can have good and bad sections
- Users provide more information than is available from the final conversation rating & length
  - topic proposal, topic acceptance & rejection
  - reaction to content

# Content Rewards



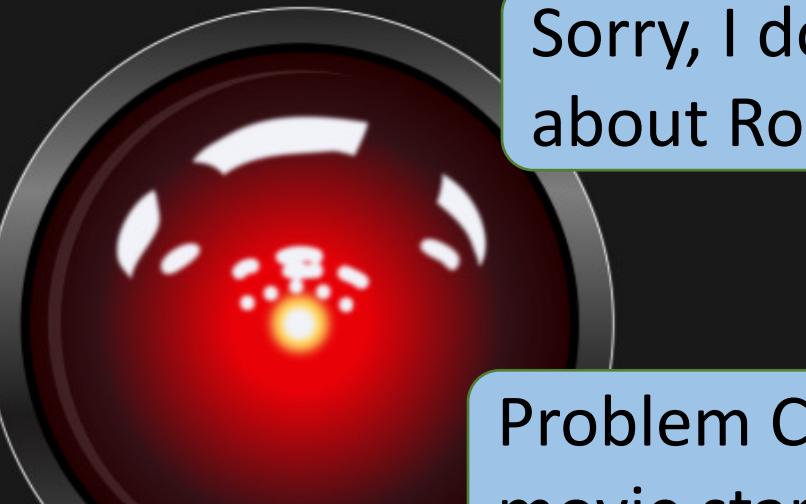
# What we learned from 10M Conversations

- Impacts of working with speech
- User variation
- It's a wild world
- Deep conversations

HAL 9000

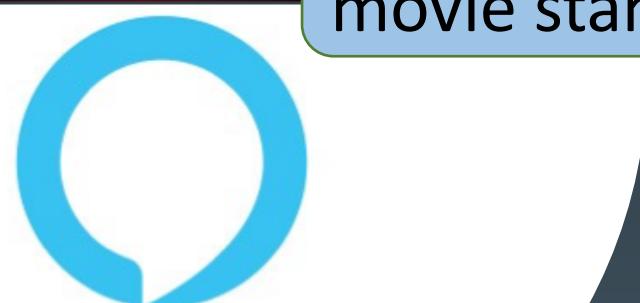
HEURISTICALLY PROGRAMMED ALGORITHMIC COMPUTER  
URBANA, ILLINOIS MCMXCVII

Open the pod bay doors please.



Sorry, I don't know anything about Rod Kay Moore.

What's the problem, Hal?



Problem Child is a 1990 comedy movie starring Michael Oliver.

...



HAL, I won't argue with you anymore. Open the doors.

alexa

Playing the Doors on Spotify.

# Lesson #1: ASR is imperfect

- Speech recognition is not a solved problem
  - Systems are tuned for commands, not conversational speech
- Most systems provide developers with an impoverished representation of speech
  - No sentence segmentation, no representation of prosody
  - Missed intents (sarcasm, doubt, ...) and affect (annoyance)

# Actual problems.... (ASR errors)

Hi! This is an AlexaPrize Socialbot. How's your day?



Glad to hear it. We can talk about robots, ...

• • •

cause does that you're gonna state ***that's cool***

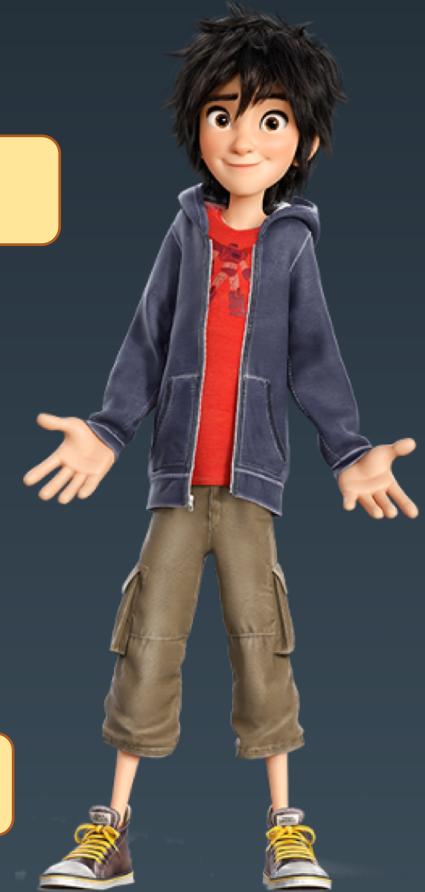
I'm happy you liked that.

• • •

Alexa do you want off

You're asking: do you want off. I actually  
don't know. Do you wanna know ...

I'm five.



# More actual problems.... (Prosody)

How about robots what's up with robots?

Sorry. I'm actually not very familiar  
with *robots what's up with robots*.

• • •

Did you know that in a restaurant in China robots  
have replaced humans to make noodles?

Yeah I did not.



Wow! You must be very smart!  
Did you see this headline ....



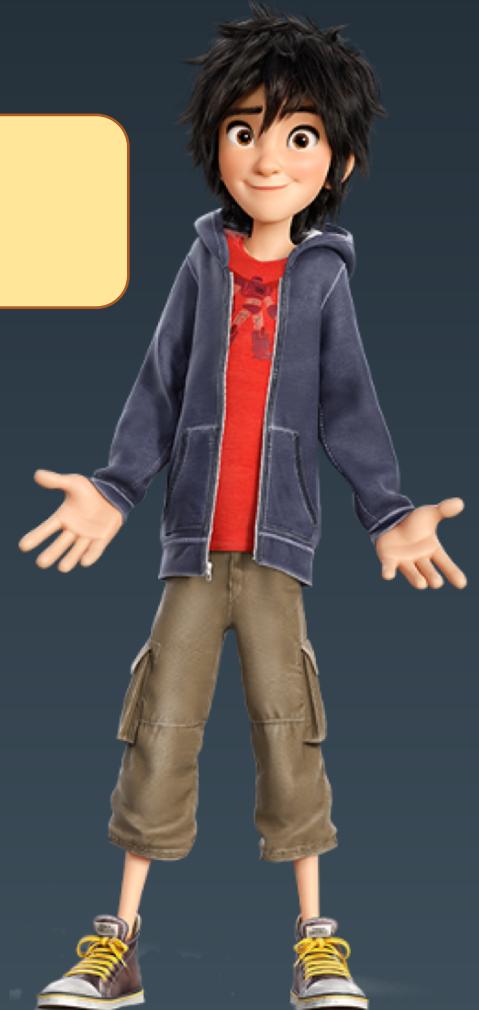
Note: 2018 Alexa prize  
teams get punctuation.

# Uninterpretable ....

yeah can i get my butt hampshire suspense are  
there was a cough sure stop



No problem. Let me think. How  
about we chat about...



*Often “can you repeat that?” isn’t going  
to yield a much better result.*

# Handling Impoverished Text

- NLU must be robust to errors; leverage ASR uncertainty
- Use a multi-dimensional representation of the user utterance to capture multiple intents
- Leverage dialog history to (partially) compensate for missing prosodic cues
- Use a mix of last resort strategies:
  - Generic responses, change the topic
  - Repeat the request, say you didn't get it

# Lesson #2: Users Vary

- Different interests, opinions on issues, sense of humor
- Interaction styles: terse vs. verbose, politeness, ...
- Different goals: information seeking, opinion sharing, getting to know each other, adversarial

# Real Users: Content Preferences Vary



Did you know that Malaysian vampires are tiny monsters that burrow into people's heads and force them to talk about cats?

## *Amused*

Oh you are so funny.

Oh my god that's funny.

## *Not amused*

That's not true.

Oh gods are you have to hear this.

What the heck.

## *Not really listening?*

Cool.

Wow that's interesting.

That's awesome.

## *Cat lover*

Cats are my favorite animals.

Let's talk about cats.

That's creepy.

# Real Users: Personality types

- Personality correlates with user ratings
  - Extroverted, agreeable, open -> higher ratings
- Topics brought up by users
  - Introverts (AI, cats), extroverts (news, fashion)
  - Open & imaginative (time travel, aliens)
  - Low conscientiousness (pokemon, video games, minecraft)

## Talkative User

- So you're saying mean is a type 2 diabetic art is junk food
- No I'd rather go back to talking about vampires
- I love vampires the something that I really loved do you know and rice the author
- it is weird i don't understand
- Yes I would
- i don't know that's an interesting question and is it really true that garlic keeps vampires the wedding and what i
- what are they have their long fingernails for
- i think that that's probably true but i think it vampires are evil and they don't care about sustaining things for human be...

## Interaction Styles

## Terse User

- |                         |                         |
|-------------------------|-------------------------|
| → Cool                  | → No                    |
| → One excuses enough    | → Yes                   |
| → No thanks             | → Cool                  |
| → No                    | → Yeah that's cool      |
| → No                    | → No I didn't           |
| → Sure                  | → No                    |
| → Yeah                  | → Yes                   |
| → Yeah                  | → No I didn't know that |
| → No I didn't know that | → No                    |
| → No                    | → Yes                   |
| → Yeah                  | → Cool                  |
| → No                    | → ...                   |

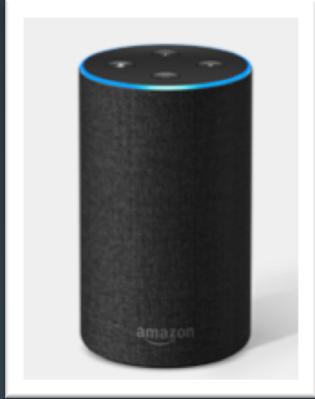
# Lesson #3: It's a Wild World

- Problematic content
  - Offensive or controversial material
  - Sensitive topics
  - Depressing topics
- Difficult users
  - Adversarial users
  - Risky users



**Important:**  
The bot is taken offline when it says things that are not family friendly.

# Filtering Is Hard



Do you want to hear something about Goldilocks?

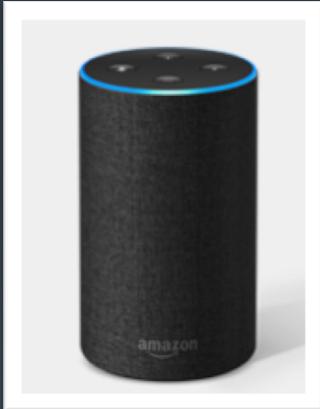
Sure!

If Goldilocks tried 3 different beds, then Mama Bear and Papa Bear slept separately. I think it's safe to say that Baby Bear is the only thing keeping the family together.

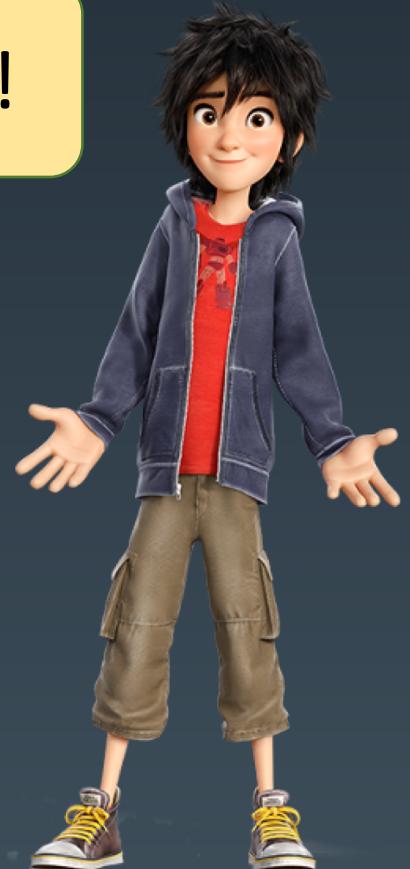


# Know Your Audience

Let's talk about Santa Claus!



You know what I realized the other day? Santa Claus is the most elaborate lie ever told.



# Adversarial User Examples

- Taboo topics



Alexa, are you naked?

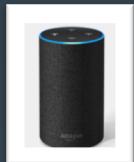


I'm a chatbot.

- Offensive language



@#%\*#\*!!.



I heard you mention unicorn. I'll do some research on that. How about we talk about some news about artificial intelligence?

- Venting



You suck.



Unfortunately, I'm not sure I understood what you said.

# Lesson #4: Shallow Conversations

- Limited understanding & generation capabilities drove a dialog strategy of shallow conversations
- Users often want to learn more about news stories presented – need dialog context to interpret question
- Switching to related topics is fragile when there are word sense ambiguities

# Talking about news....

Celebrities.

Celebrities ? I love it! We can chat about **Taylor Swift** . I read this article from Yesterday. The title was: Taylor Swift Attempts to Silence Critic, ACLU Fires Back. Did you know about this ?



My pleasure. Let's see, have you heard anything about ACLU?

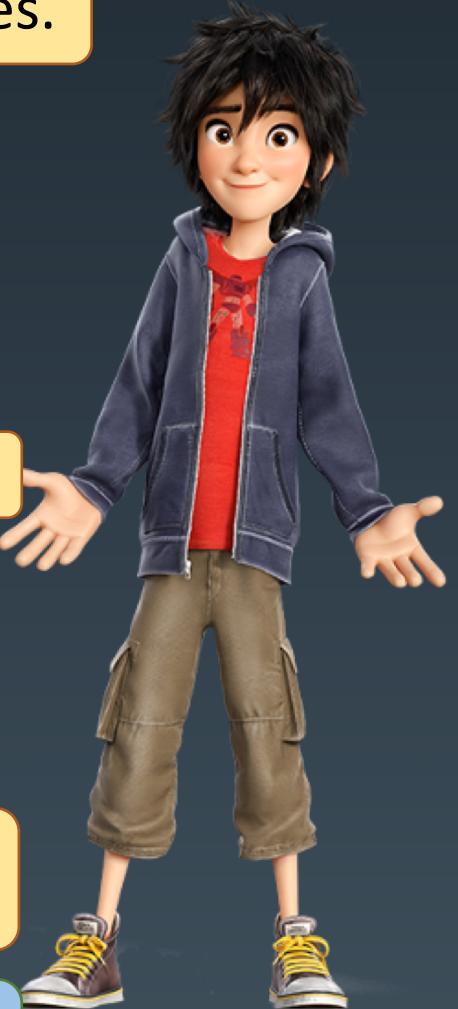
No tell me about it.

No what is that.

• • •

No I like your information about Taylor Swift. Tell me more about that.

You mentioned your information. My friend also gave me this advice. Check if your information has been stolen in an ...



# Challenges of Deeper Conversations

- Tell me more...
  - Summarization with a conversational style
- Specific article questions
  - Is the answer in the article or in a general resource?
  - Why questions
- Coherent topic switching
  - Integrated popularity and semantic relevance ranking

# Lessons for Academic-Industry Partnerships

- What works
- What needs work

# What Works

- Access to data from **real users** at a large scale
  - Impacts the problems we choose to solve and the resulting solutions, increases relevance of the work
  - Teaches students about the complete problem
- Funding to support students (no free ~~lunch~~ labor)
  - Research drivers, bug finders & potential future employees
- Industry person-time allocated to support partnership
  - Early access to system improvements
  - Advice on tools, feedback on progress

Many thanks to....

Amazon, Google, Microsoft, Mobvoi, Tencent, Samsung,  
Bloomberg, Allstate, Facebook, Boeing, AT&T, Apple, IBM,  
Nynex, ATR, ...

for good collaborations.

# What Needs Work

- Privacy-preserving access to user data
  - For spoken language systems: prosody info
  - For text & speech: speaker/author demographics
- For spoken dialog systems: richer speech interfaces
- Competitions are great kickstarters, but
  - Substantial engineering effort is required
  - Longer term access to users/data & collaboration is needed to leverage the investment

# Summary – Sounding Board

- A socialbot can be more than a chatbot
- Content-driven & user-centric design
- Technology is still in early stages: architecture should allow for change
- Learn from user responses and ratings

# Summary – General Take Aways

- Data from real users has real impact on research
- A socialbot is a great platform for NLP research
- Spoken conversations begin & end with speech

# Thank You

*For more info, check out the demo Monday 14:00 - 15:30  
Dialogue and Interactive Systems - Elite Hall B*