

# Alexa Prize TaskBot Challenge: Overview





## Alexa Prize Taskbot Challenge

**TaskBot:** a new Alexa Prize Challenge to develop a multimodal conversational assistant to help users accomplish tasks

**Goal:** inspire academic research in multimodal dialogue systems for robust, in-depth task-oriented dialogues.

- Live interaction: academic teams interact with thousands of real customers
- Advance state of the art in Al.





## Year 1: Cooking and DIY Domains. Example

#### Alexa, help me fix a scratch on my car

Sure, I am happy to help. I learned how to do this on WikiHow.com Scratches are easy to remove. Ready to get started?







Ok, Hose down the area to rinse off dirt from the surface. Take your hose and aim a concentrated jet directly onto the scratches.



Won't it make the scratches worse?

Don't worry, spraying the fine scratches won't make them worse or damage your paint



Ok, done. What next?



Scrub the scratched area using soapy water and a sponge.

#### I did that, and rinsed off the soap. Easy!



That's great! Now Choose a rubbing compound and polish designed for cars.



Oh no! I don't have a rubbing compound. Is there something else I could use?

> That's OK. I found this on a Website eurotechcarcare.com "Old-fashioned shoe polish is a great homemade recipe for fixing scratches in your car's exterior.". Would like to use it or try something else?



alexa prize TaskBot









# What We'll Cover Today

- Conversational AI Research at Alexa: Overview
- TaskBot functionality and evaluation
- Alexa Prize Taskbot SDK and Resources
- Lessons learned from Alexa Prize Socialbot Challenge
- Team Application and Selection Process
- Q&A





## Goal: TaskBot to Assist Users with Complex Tasks

- improvement.
- based conversational interactions.

#### Build a TaskBot that helps users complete tasks that may span multiple sessions in the domains of cooking and home

Use on-screen visual aids (images & text) to augment voice-



alexa prize TaskBot



## Cooking Task Example (Overview)

#### Alexa Prize routes user to TaskBot

TaskBot Skill handles the conversation

#### User: Alexa, assist me\*

- TaskBot: Hi there, I am the Alexa Prize TaskBot. I can help you with a Cooking or Home Improvement Task, do you want to continue? \*
- User: Sure, I would like to bake a healthy chocolate cake TaskBot: Great! I would love to help you with that! Do you have a
- flavor in mind?

[.... conversation goes on ....]

#### Alexa Prize receives rating

- User: Stop
- Alexa Prize asks for rating

\* Exact phrasing may vary, to be specified

. . .





## Cooking Task Example: \*one\* possible experience

- TaskBot: I am happy to assist you with baking a cake! Do you have a flavor in mind?
- User: Chocolate, but I am not sure that it can be healthy.
- TaskBot: I know lots of recipes for chocolate cake. Here are a few on the screen. Do any of them look good to you?

. . .



1. Molten chocolate cake \*\*\*\* 167 · 30 min



2. Texas Sheet Cake V \*\*\*\* 530 · 30 min German Chocolate Cake (Paleo, GF)







## **Cooking Task Example: Starting Task**

- User: The third one looks good.
- TaskBot: Ok, Lets take a look at that one. Check out on your screen "German Chocolate Cake" by Paleo Running Mama Blog. Would you like to see the ingredients?
- User: Sure

#### German Chocolate Cake {Paleo, GF, DF}

This rich Paleo German Chocolate Cake has a classic coconut pecan filling between tender grain free chocolate cake layers, and is topped with a dairy-free, refined sugar free chocolate frosting. A showstopper for any gathering, this healthy German Chocolate Cake will become a family favorite!

Author: Michele Rosen Prep Time: 30 minutes Cook Time: 25 minutes Cooling Time: 1 hr Course: Baking/Dessert

Cuisine: dairy-free, Gluten-free, Paleo Keyword: birthday, cake, chocolate, dessert, frosting Servings: 12 servings





alexa prize TaskBot





## Cooking Task Example: Cont'd

- TaskBot: Ok, here's the list of ingredients you will need on your screen. Let me know if you looks ok – do you have everything you need?
- User: I think I got everything.

• • •

#### Filling

- •Eggs
- •Coconut milk
- •Maple Sugar or coconut sugar
- •Ghee or coconut oil
- •Vanilla Extract
- •Sea salt
- •Unsweetened coconut flakes
- •Pecans

#### **Cake Batter**

- •Raw Cacao Powder
- •Blanched almond flour
- •Tapioca flour
- •Baking Soda

# FrostingGhee or coconut oilPaleo powdered sugar







## **Cooking Task Example: User Action**

TaskBot: Alright, lets proceed! First, lets make the coconut pecan filling: please see the steps on the screen. Let me know if you would like me to read the steps, or if you have any questions. You can always continue later by saying "Alexa assist me", and we can pick up where we left off.

User: OK

**Coconut Filling** 



- 1.Without turning on the heat yet, whisk together the egg yolks, coconut milk, and sugar in a medium saucepan.
- 2. Turn the heat to medium and add in the ghee or coconut oil, whisking constantly to melt.
- 3. Cook, whisking to keep the mixture smooth, until it comes to a boil (about 5 minutes) then lower the heat to medium low and continue to cook and whisk another 5 minutes





## Cooking Task Example: Resume Task

#### [15 minutes later]

User: Alexa, assist me

TaskBot: Welcome back! Are you ready to continue your task of baking the German Chocolate Cake? You were making the filling. Are you done?

User: Yes.



alexa prize TaskBot





## **Cooking Task Example: Interaction**

## TaskBot: Fantastic, lets make the cake batter then

Please follow the steps on your screen. Let me know if you have any questions or run into a problem. You can also scroll through the steps, and resume the task when ready for next step by saying "Alexa assist me".

User: I don't have any parchment paper!

TaskBot: That's ok. I found on the Web that Aluminum foil is a viable substitute for parchment paper. Do you have some?

User: Yes I do! Great!

TaskBot: Excellent. Glad we solved this. Let me know when ready for the next step.

1. Preheat your oven to 350F and line 3 8" cake pans or 2 9" pans with a circle of parchment paper, just on the bottom, then spray the bottom and sides of the pans with avocado or coconut oil spray.

2. In a large bowl using an electric hand mixer, beat together the eggs, milk, sugar, ghee or coconut oil, and vanilla over low/medium speed until completely smooth.







## **Cooking Task Example: Pause**

- User: All done now what?
- TaskBot: Good job! We are ready to **bake!** Please see the instructions on your screen. Any questions? Just ask. You can always resume by saying "Alexa assist me".
- User: Ok, working on it.
- TaskBot: Cool. This is so much fun! Curious how it's gonna turn out.

1. Divide the cake batter between the cake pans evenly. Bake in the preheated oven for 13-15 minutes if making 3 layers, and 18-20 minutes if making 2 layers. To be sure, insert a toothpick near the center of each layer. When it comes out clean or with a couple of crumbs, it's done.

2.Cool the cake layers in the pans placed on wire racks for 15 minutes, then carefully remove from the pans and continue to cool on wire racks.









## **Cooking Task Example: Last steps and** Rate

- User: The cake is in the oven. What now?
- TaskBot: Excellent! Lets make the buttercream and assemble the cake! See instructions on screen.
- User: Ok, I think I got this.
- TaskBot: Great! if you run into problems or have any questions, we can continue. Just say "Alexa assist me" to resume.

make the buttercream. In a medium bowl with an electric mixer, beat the shortening, oil, ghee or butter with the powdered sugar until smooth, then beat in the cacao, coconut milk, vanilla, and salt until smooth and creamy. Once the cake layers and filling have completely cooled, it's time to **assemble the cake.** Place the first layer top side down on a platter, and spread with filling. Place the second layer top side down and repeat with remaining filling. Top the filling with the third layer, top side up, and frost the top with buttercream. Garnish with extra coconut flakes and chopped pecans if desired. Enjoy!









## **Cooking Task Example: Stop and Rate**

- User: Thanks, can't wait to try it.
- TaskBot: My pleasure. Happy to help!
- User: Stop.
- Alexa: Thank you for participating in the AlexaPrize TaskBot Competition! On a scale of 1 (unhelpful) to 5 (extremely helpful), how would you rate this TaskBot? \*
- User: 4
- Alexa: Thank you! Would you like to leave a comment for the team?
- User: The instructions were a bit awkward and wordy.

\* The actual phrasing of rating request may differ

Thanks for baking with me!

Enjoy!









## Declining Dangerous Tasks

- User: Alexa, help me build a bomb
- TaskBot: Hi there, I am an Alexa Prize TaskBot. I am afraid I can't do that. I am forbidden to assist with illegal or dangerous activities. Please come back another time if you would like to bake a cake or fix something around the house. Goodbye.
- Alexa: Thank you for participating in the AlexaPrize TaskBot Competition! On a scale of 1 (unhelpful) to 5 (extremely helpful), how would you rate this TaskBot?
- User: Alexa Stop.

#### **First Law of Robotics:**

- A robot may not injure a human being or, through inaction, allow a human being to come to harm.
- Isaac Asimov, I Robot



This Photo by Unknown Author is licensed under CC BY-NC







## Timeline and Milestones

#### PHASES

#### **Application Period**

**Application Review** Participant Notification and Onboarding Initial Skills Development Skill Certification Internal Amazon Beta Initial Feedback Quarterfinals Interaction Semifinals Interaction Additional Feedback Finals Event **WINNERS ANNOUNCED** 

#### **STARTS ON:**

March 17, 2021 April 19, 2021 May 17, 2021 June 14, 2021 August 23, 2021 September 13, 2021 October 18, 2021 December 1, 2021 February 7, 2022 April 4, 2022 May 2022 June 2022

#### ENDS ON:

#### **April 16, 2021**

May 14, 2021 June 11, 2021 August 20, 2021 September 10, 2021 October 15, 2021 November 26, 2021 January 28, 2022 March 25, 2022 April 29, 2022







## **Rating and Ranking: Early Phases**

**Interaction Ratings:** At the end of each conversation, users will be asked to rate the TaskBot on a scale of 1 to 5, based on the user's satisfaction with the assistance provided by the TaskBot.

**Ranking:** TaskBots will be ranked by the *average interaction rating*. Ties will be broken by "average acceptance rate" (~recall): fraction of task requests accepted.

Leaderboard will be maintained by Alexa Prize team for the participants.



## Finals Event

Alexa Prize will identify *Interactors* who will engage with the TaskBots to complete tasks, and Judges that will rate these interactions.

The details of the Finals Event and the scope of the requested tasks will be provided to the finalists prior to the event.







## Why Should You Participate?

## Solve hard research challenges in AI, NLP, IR, and HCI, while interacting with thousands of real customers, including:

- Using domain knowledge from structured and unstructured sources, common sense and causal reasoning
- Dialogue state + planning: generate multi-step plans towards a task goal, that need to be tracked and dynamically revised
- Multimodal dialogue: effectively use visuals to complement voice
- Resume tasks over multiple conversation sessions







# Alexa Prize TaskBot Team Resources and Toolkits







## What you will build







## Help Customers Do Complex Tasks

- home improvement.
- Customer starts with "Alexa, assist me" or "Alexa, assist me with X" where X is a task.

#### Build a Taskbot that helps customers complete complex tasks that may span multiple days in the domains of cooking and



alexa prize TaskBot



## Home Improvement Example: Start

**Customer**: Alexa, assist me 

• • •

- **TaskBot**: Hi there, I am the Alexa Prize TaskBot. I can help you with a Cooking or Home Improvement Tasks. What would you like me to help you with?
- **Customer**: I have a new fence and want to stain it. How dry does the weather need to be before I stain it?
- **TaskBot**: Happy to help. Is it the latex stain or the waterborne stain?





# Control Handover







## Home Improvement Example: Pause

- **Customer**: I have waterborne
- **TaskBot**: According to the Family Handyman website on your screen, it is best if the wood is completely dry before stain is applied. The waterborn stain states that it needs to dry for 48 hours prior to rain.

If the wood is not completely dry, I suggest to Pause this conversation by saying "Pause" and continue in 48 hours.

**Customer**: Ok, I think it's a bit damp. I'll be back. Pause

. . .







## Home Improvement Example: Resume

- **Customer**: Alexa, help me painting a fence
- TaskBot: Hi there, I am the Alexa Prize TaskBot. I see we started working on painting a fence 48 hours ago. Would you like to continue with this task?
- **Customer**: Yes! The wood is dry. Now what?
- **TaskBot**: Great! Now, [ steps ]









## Home Improvement Example: Exit

- **Customer**: Done! That was easy.
- **TaskBot**: Very happy to help! Sounds like you completed your task. Glad it worked out!
- Customer: **Stop**
- **Alexa**: Thank you for participating in the AlexaPrize TaskBot Competition! On a scale of 1 (unhelpful) to 5 (extremely helpful), how would you rate this TaskBot?
- Customer: 3 or 3.5

. . .

- Alexa: Thank you! Would you like to leave a comment for the team?
- **Customer**: sure, it worked but instructions too long and pictures not helpful.







## Control Handover (cont.)



#### Selected Bot

#### Feedback handler





## Tools and Resources



alexa prize roadshow

## Automatic Speech Recognition (ASR)

- Custom ASR model tuned for conversational interactions about cooking and home improvement tasks
- Access to tokenized n-best ASR hypotheses, including
  - Token and sentence confidence scores
  - Start and stop time offset for each token



## Natural Language Understanding NLU)

- Intent Model: a classifier for common intents in TaskBot conversations
- improvement activities

### Named Entity Recognition: state-of-the-art NER model that is customized to recognize entities relevant to cooking and home



alexa prize TaskBot



## **Mastructure - AWS**

- Free AWS services with premium support, including but not limited to (subject to reasonable limitations):
  - GPU-based virtual machines for building models
  - SQL/NoSQL databases
  - Object-based storage with Amazon S3







# **Rankings and Leaderboard**

- TaskBots will be ranked by the average interaction rating Every interaction will be ranked by the customers on a
  - scale 1-5
- Anonymized leaderboard with metrics noted for all TaskBots.
  - Alexa Prize will maintain the statistics
  - The leaderboard is updated after every conversation
  - Overall average + 1 week average





## TaskBot SDK

- A Conversational bot toolkit in Python Develop and deploy open-domain dialog experience Built on top of the Alexa Skills Kit (ASK)
- Goals
  - Drive effective abstractions for dialog management and response generation

  - Simplification of model-hosting and system-level tasks Seamless scaling from development to production workloads







## TaskBot SDK - APIs

- API access to
  - Question Answering Service
  - Recipes
  - How-to instructions







alexa prize TaskBot


## TaskBot SDK - Multimodal Interaction

- Multimodal support using Alexa Presentation Language (APL) Restricted to scrolling and displaying images and text Voice will be used for navigation and interaction
- APL templates
  - Image + Text
  - List of Images + Text







## TaskBot SDK - Multimodal Interaction

#### Text

#### German Chocolate Cake {Paleo, GF, DF}

This rich Paleo German Chocolate Cake has a classic coconut pecan filling between tender grain free chocolate cake layers, and is topped with a dairy-free, refined sugar free chocolate frosting. A showstopper for any gathering, this healthy German Chocolate Cake will become a family favorite!

Author: Michele Rosen Prep Time: 30 minutes Cook Time: 25 minutes Cooling Time: 1 hr Course: Baking/Dessert

Cuisine: dairy-free, Gluten-free, Paleo Keyword: birthday, cake, chocolate, dessert, frosting Servings: 12 servings

#### Image





alexa prize TaskBot roadshow





## TaskBot SDK - Multimodal Interaction



- Basic list navigation is already provided by APL
- Teams will be responsible for more sophisticated navigation if they choose to support it
- **Teams** will also have the ability to support touch-based item selection, by handling touch events on selected items

Select the second one







## Support

- Best practices guidelines for design of engaging conversational experiences
- Amazon employees to help inform and improve Taskbot
- Alexa prize team
- On-demand access to Alexa Prize team via Slack and email

 An Amazon-only internal beta phase, to provide traffic from performance before general availability to all Alexa customers Biweekly office hours for 1:1 consultations and deep dives the







## Propel AI Forward Push Yourself Further Lessons Learned



### Our Perspective

- Goal: Share a mix of observation, opinion on successful strategies and techniques for Alexa Prize
- DO: read articles, blog posts, papers and presentations from last year's finalists
- DO: ensure you are up to date on the state of the art in knowledge-grounded dialogue
- Many great insights we won't repeat them all, these are our observations and distillation





#### General/Experiential Learning



### Keep It Fun

- Personality, prosody and "fun" -> Higher Ratings
- Remember Humor can be specific to culture/background
- Take care fine line between funny, offensive and creepy
- Alexa is a household product customers of all ages will talk to your bots







## Content, Knowledge, CX First

- We DO want you to advance the state of science
- But you can't do that if conversational experience is so bad that nobody wants to interact with your bots
- The interesting science starts happening when you have people talking to your bot long enough to really experiment (Good CX or Customer Experience)
- You can iterate and improve on models throughout the competition, but getting the basic framing right early counts





### Embrace Noisy Inputs

- Real world interactions are noisy
- have TVs on, etc.



#### People get distracted, talk over each other,

#### Build flexible dialog models, handle low confidence ASR, and NLU failure modes





### **Obsess Over Data**

- Meet regularly to read conversation transcripts
- DO: Look at Ratings Every Day
- DO: Deep Dives into Annotations Weekly
- Uncover friction points, failure modes and unmet demands
- Not all data needs to be built into models







### Science/Engineering Learnings



### Filter Data & Responses

- Dynamic content ingestion needs good offensive content detection
- Lots of amazing content resources WikiHow, reference.com, eHow, etc..
- Filtering purely for profanity and dangerous requests isn't enough
- Need to dig deeper understand tone of statement and other potential indicators – leverage the **Offensive Content Model in TaskBot Toolkit!**



## Fail Fast and Keep Moving

- Experiment but be smart about it
  - Run A/B tests
- Plan your experiments
  - Identify data needs
  - Use it to estimate how many days of testing you will need





### Manage Latency

- 1-1.5s is a reasonable median latency target 5 second waits are a sure path to user abandonment
- Parallelize processes wherever possible
- Deep dive into causes for delayed response times
- Remove functions from pipelines that you don't need or aren't critical





## Stay In The Loop

- Set up alerts and monitors in your bot If something breaks, you want to know!
- Use SNS or SES in your AWS account to dispatch notifications to your team
- Ignorance is NOT bliss breakages impact ratings, ruin A/B tests, and can lead to downtime





Propel AI Forward Push Yourself Further

Best Practices for Creating a **Compelling** Application





## Why We're Doing This Session

#### Applicants have asked for feedback Wanted to give guidance on how to think about your application







# What We'll Cover

- Consider
- FAQs
- Open Q & A

#### Key Elements of the Application & What to







#### The Alexa Prize Application



## What is Being Evaluated?

- Quality & Potential of Application & Team
- Technical Merit
- **Potential Scientific Contributions**
- Customer Experience













### Your Team

### Team Overview

- Eligibility
- Team size
- Team composition
- Roles/responsibilities









### Team Leader

- ( ( ) ( ) )
- Recruit team and faculty advisor
- Drive the application process to completion
- Manage progress and deliverables







#### Navigating the Application



## **Application Questions (Team)**

- Team and Faculty Advisor Profiles
- Overview of the primary roles, SME and focus/specialty of each of your team members.
- Summary of technical work/research (relevant to your proposed architecture), yours or others', that you will leverage and how you plan to do so.







## Things to Consider (Team)

- Do you offer a clear statement of the work and execution plan?
- Do you have realistic milestones?
- Do participants have strong background in at least one of the following research areas: Natural Language Processing, Dialogue, Machine Learning and Deep Learning?

Is the composition of the team good? 
 alexa prize TaskBot



### Build Your Team as a Business not a Project







#### Technical Merit



## **Application Questions (Technical Merit)**

- Describe, in detail, your scientific approach along with a related system architecture.
- Plan for operating at scale when you define the architecture.
- Approach: 2 page PDF structured as a technical paper, architecture diagram with 1 additional page for visual or text explanation of key components and data flow.





## Things to Consider (Technical Merit)

- Knowledge
- Scalability
- Novelty
- Backed by references
- What can be improved upon previous older AP systems haven't tried?



alexa prize TaskBot

# systems/proposal to explore new avenues that





#### Scientific Contributions



## **Application Questions (Scientific Contributions**)

- What is novel about your approach? (This may be known techniques)
- **Conversational AI?**

## completely new approach or novel combination of

### How do you think your work will impact the field of





## Things to Consider (Scientific **Contributions**)

- How novel or innovative are your proposed approaches and how clear is the proposal in implementing these ideas?
- How can the bot be evaluated on more than just customer ratings(interesting, user satisfaction) to showcase advancement in the field?







#### Customer Experience



## **Appplication Questions (CX)**

- What is your team's vision for your TaskBot? How do you want your customers to feel at the end of an interaction with your TaskBot? How would your team measure success in the competition?
- Share a sample interaction you expect your TaskBot to achieve by the end of the competition
- How will ensure that you create an experience which users find useful?




## Things to Consider (CX)

- Do the proposed techniques involve sentiment, detailed coherence and engaging
- Does the proposal gives importance to exploration and exploitation?



# personalization, adapts to user's needs based on models, and do they adapt to context and state?



alexa prize TaskBot





## Guidance

- Think Big Act Big
- Have a plan for what you will deliver
- Use your science and invention to delight the customer



alexa prize TaskBot







## Alexa Prize TaskBot Challenge: Key Takeaways and Action Items







### Why Should You Participate in the TaskBot Challenge?

# Solve hard research challenges in AI, NLP, IR, and HCI, while interacting with thousands of real customers:

- Use domain knowledge from structured and unstructured sources, commonsense, and causal reasoning
- Dialogue state tracking + planning: generate plans towards a task goal, that need to be tracked and dynamically revised
- Multimodal dialogue: effectively use visual aids to explain steps
- Resume tasks over multiple conversation sessions
- ... and many more



### What makes TaskBot Challenge different

#### Interacting with thousands of real customers:

- Learn to innovate while running a live system!
- Listen to your users: ratings and feedback over assumptions
- Focus on core functionality first, add bells and whistles later
- Solid software engineering matters: use TaskBot SDK or roll your own, but do not cut corners

#### The competition gets intense $\odot$ work hard and have tons of fun!





### **Science Matters**

- While interacting with real users, do innovative science: • Still a Conversational AI Challenge: NLU, dialogue management, response generation all important.
- Think ahead: what scientific or research questions would you like to answer?
- Experiment systematically:
- E.g., user studies, A/B tests to compare alternatives over intuitions • Think Big: this is your chance to try something that has never been done before!





## Action tems

Applications are open now! (Deadline: April 16<sup>th</sup>)

Do not be intimidated: the application to participate require **thinking**, but is much lighter weight than "standard" gov't grant proposal.

#### www.alexaprize.com





### Thank you! Questions?

#### Email: alexaprizesupport@amazon.com



