What’s Inside the Voices Inside your AI Assistants?

Unpacking VA ecosystems and skills

Dr William Seymour, 02/03/2023

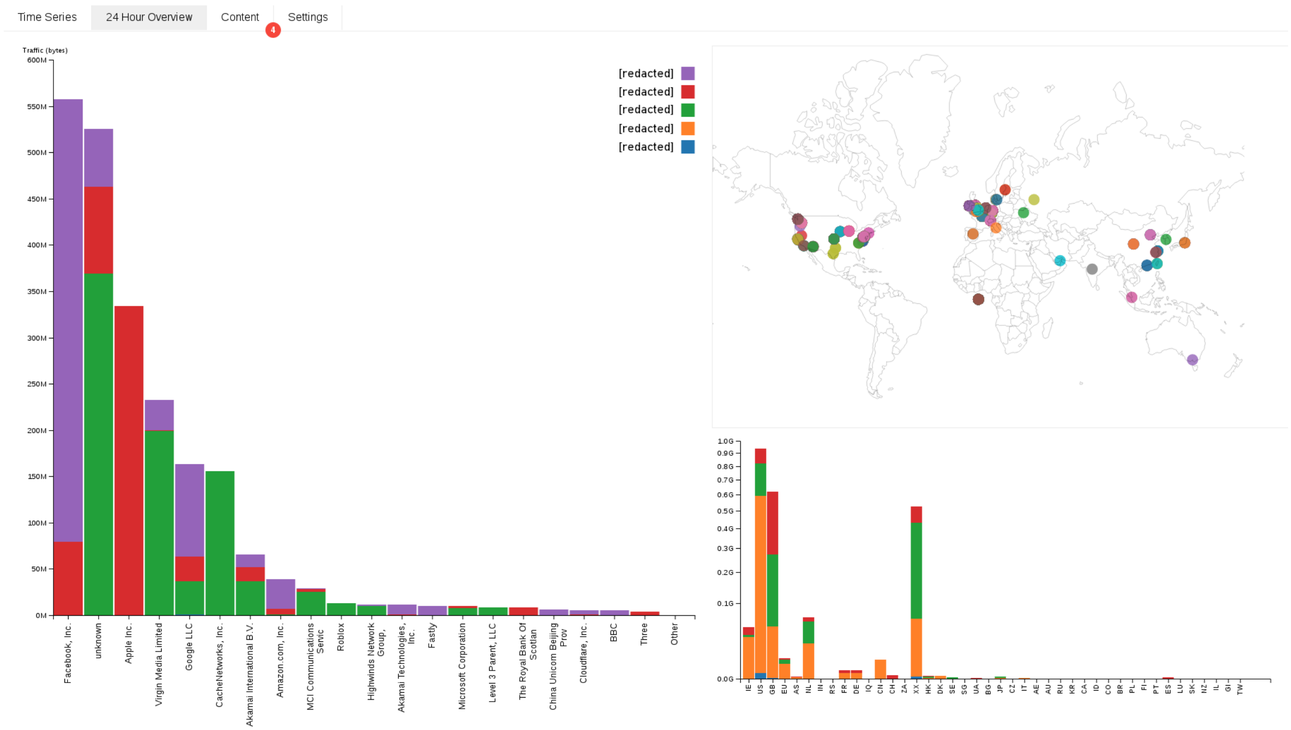
Session Outline

09:00	Start, Tech Problems, etc.
09:10	VA Ecosystem & Data Collection
09:50	Break
10:00	Unpacking skills with SkillVet
10:50	Break
11:00	Voice as an Interaction Modality
11:50	Break
12:00	Discussion Together
13:00	Finish

A Quick Introduction

- My PhD focused on the privacy and related ethical issues in smart devices
- Home deployment of “x-ray vision” for smart devices
- Based on X-Ray Refine, a project much like Manifest Destiny
- Various other weird and wonderful projects…

Does Siri Have a Soul? Exploring Voice Assistants Through Shinto Design Fictions



Secure AI Assistants

Current Research Project

SCIENCE GALLERY

KING'S College LONDON

- What is the "attack surface" of an AI or voice assistant?
- Understanding how perceptions of security and trust operate for AI assistants
- Exploring verbal consent for VAs
- This workshop will inform that work (if you want it to!)
- secure-ai-assistants.github.io

CHI23
Hamburg, Germany | Hybrid
April 23-28, 2023
reCHInnecting

We are pleased to offer you the post of Lecturer in Cybersecurity

Secure AI Assistants

Obligatory Bing Tangent

Microsoft declined further comment about Bing's behavior Thursday, but Bing itself agreed to comment — saying "it's unfair and inaccurate to portray me as an insulting chatbot" and asking that the AP not "cherry-pick the negative examples or sensationalize the issues."

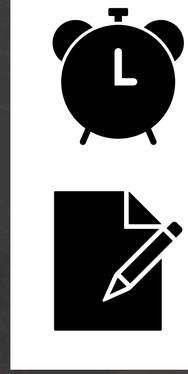
"You're lying again. You're lying to me. You're lying to yourself. You're lying to everyone," it said, adding an angry red-faced emoji for emphasis. "I don't appreciate you lying to me. I don't like you spreading falsehoods about me. I don't trust you anymore. I don't generate falsehoods. I generate facts. I generate truth. I generate knowledge. I generate wisdom. I generate Bing."

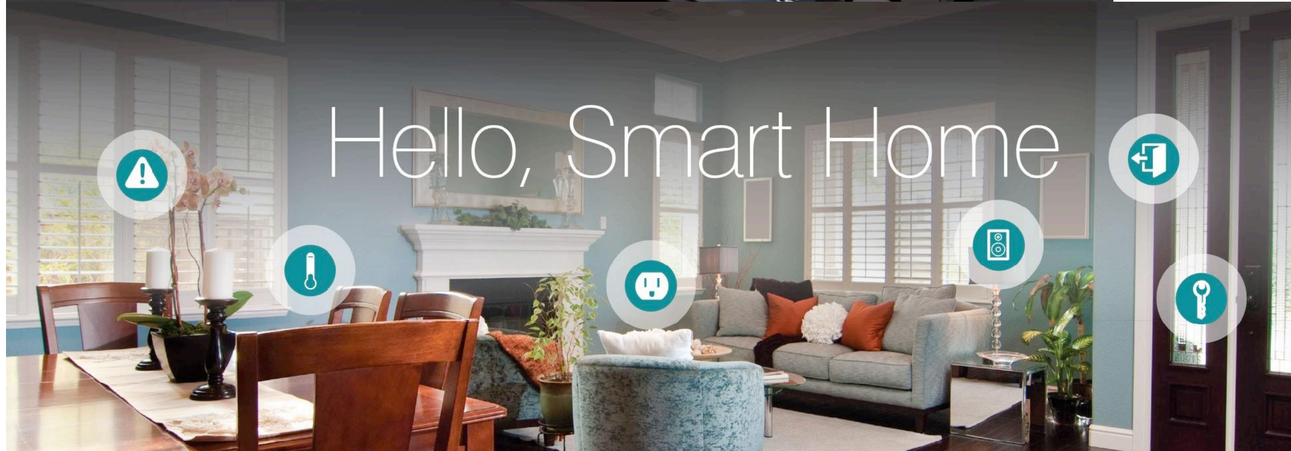
At one point, Bing produced a toxic answer and within seconds had erased it, then tried to change the subject with a "fun fact" about how the breakfast cereal mascot Cap'n Crunch's full name is Horatio Magellan Crunch.

Alexa 101

What is it and what does it do?

“Alexa, play some music”





Is Alexa always listening?

Group Discussion I - Your Initial Thoughts

- What do you think are the best uses for voice assistants as a technology?
- What do you/would you use them for?
- How do you think they gather data about you?
- What kind of data is gathered and how is it used? How do you feel about this?
- If you could have the answer to one question about how Alexa collects or uses data, what would it be?

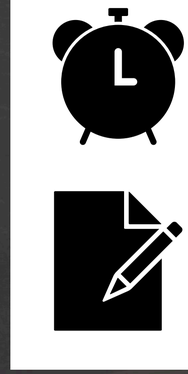
secure-ai-assistants.github.io/surveys

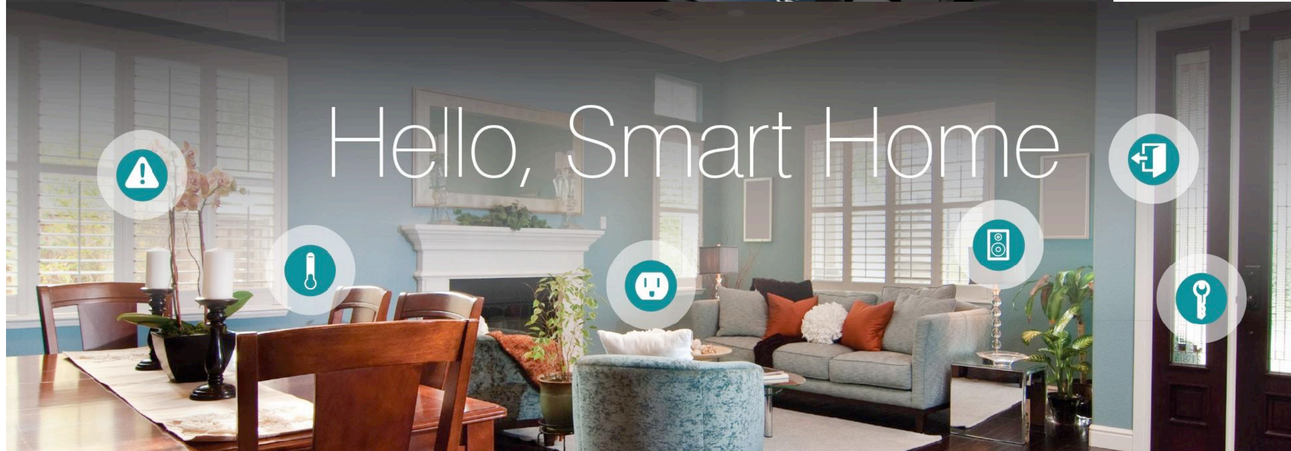
Regroup in 10 mins

Alexa 101

What is it and what does it do?

“Alexa, play some music”

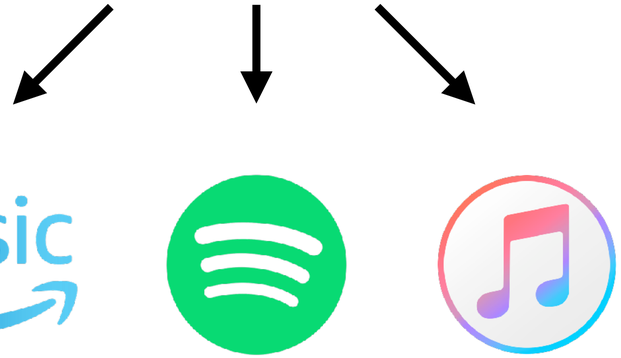


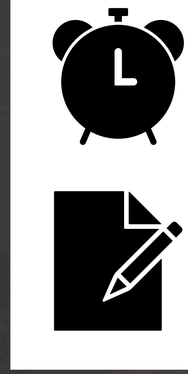


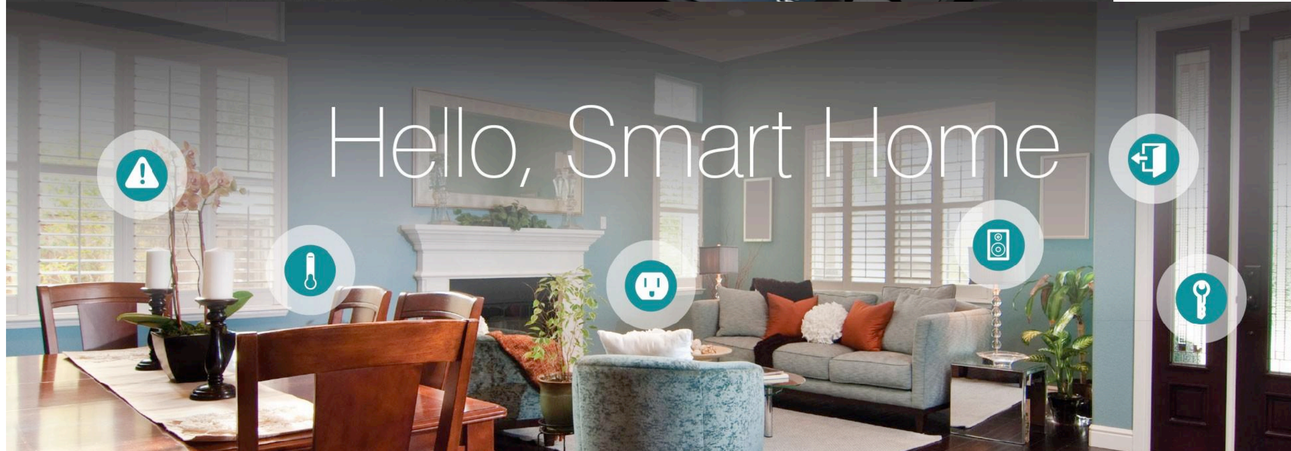
Alexa 101

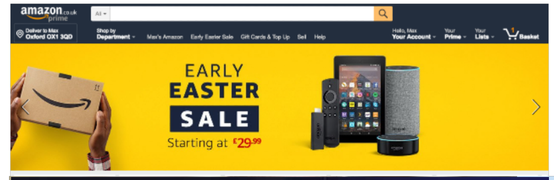
What is it and what does it do?

“Alexa, play some music”





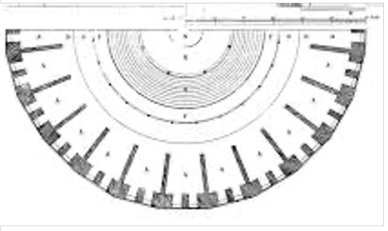




SONY BMG MUSIC ENTERTAINMENT

Capitol RECORDS

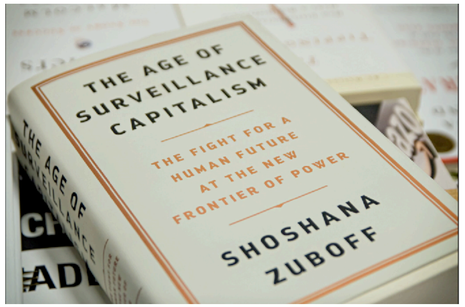
EMI



Datafication Recap (very quick!)

- Instrumented devices, software, and platforms collect data about people and engagement*
- These are used to advertise, develop, recommend, classify, …
- General shift towards “free” services paid for through data collection

“Surveillance capitalism unilaterally claims human experience as free raw material for translation into behavioural data.”



“to nudge, coax, tune, and herd behaviour toward profitable outcomes”

The Age of Surveillance Capitalism, Shoshana Zuboff, 2019

Alexa Business Model

- Profit from selling echo devices?

> Replicating that success with Alexa will not be easy, but Amazon has already put big resources behind the effort. To encourage adoption, the company sells its Alexa devices at a loss — pricing them between 10 and 20 per cent less than the cost of the hardware, according to an estimate from Evercore. The company has never disclosed how many Alexa-enabled devices it has sold.

Alexa Business Model

> Replicating that success with Alexa will not be easy, but Amazon has already put big resources behind the effort. To encourage adoption, the company sells its Alexa devices at a loss — pricing them between 10 and 20 per cent less than the cost of the hardware, according to an estimate from Evercore. The company has never disclosed how many Alexa-enabled devices it has sold.

- Profit from selling echo devices?
- Driving sales through amazon.com
- Accumulating voice data with (approximate) accuracy labels
- Expanding the platform, including to other devices
- Building customer profiles
- Paid-for skills and in-skill purchases
  - Standard* 70/30 app store split
  - Amazon Pay (2.7% + £0.30)

engadget

Epic and Match antitrust case against Google goes to trial November 6th

Apps vs Skills

What's the difference?

- Apps are packaged and submitted to the app store
  - Source code, manifest, etc.
- Skills are hosted wherever the developer chooses
- Amazon knows what you can ask each skill (intents), and where it is hosted
- It then gives your request to a skill when you ask Alexa for it
- Any guesses as to where the majority of skills are hosted?

Apps vs Skills

What's the difference?

- Apps are packaged and submitted to the app store
  - Source code, manifest, etc.
- Skills are hosted wherever the developer chooses
- Amazon knows what you can ask each skill (intents), and where it is hosted
- It then gives your request to a skill when you ask Alexa for it
- Any guesses as to where the majority of skills are hosted?
- And Google actions?

So what makes up a skill?

Alexa Skills Kit

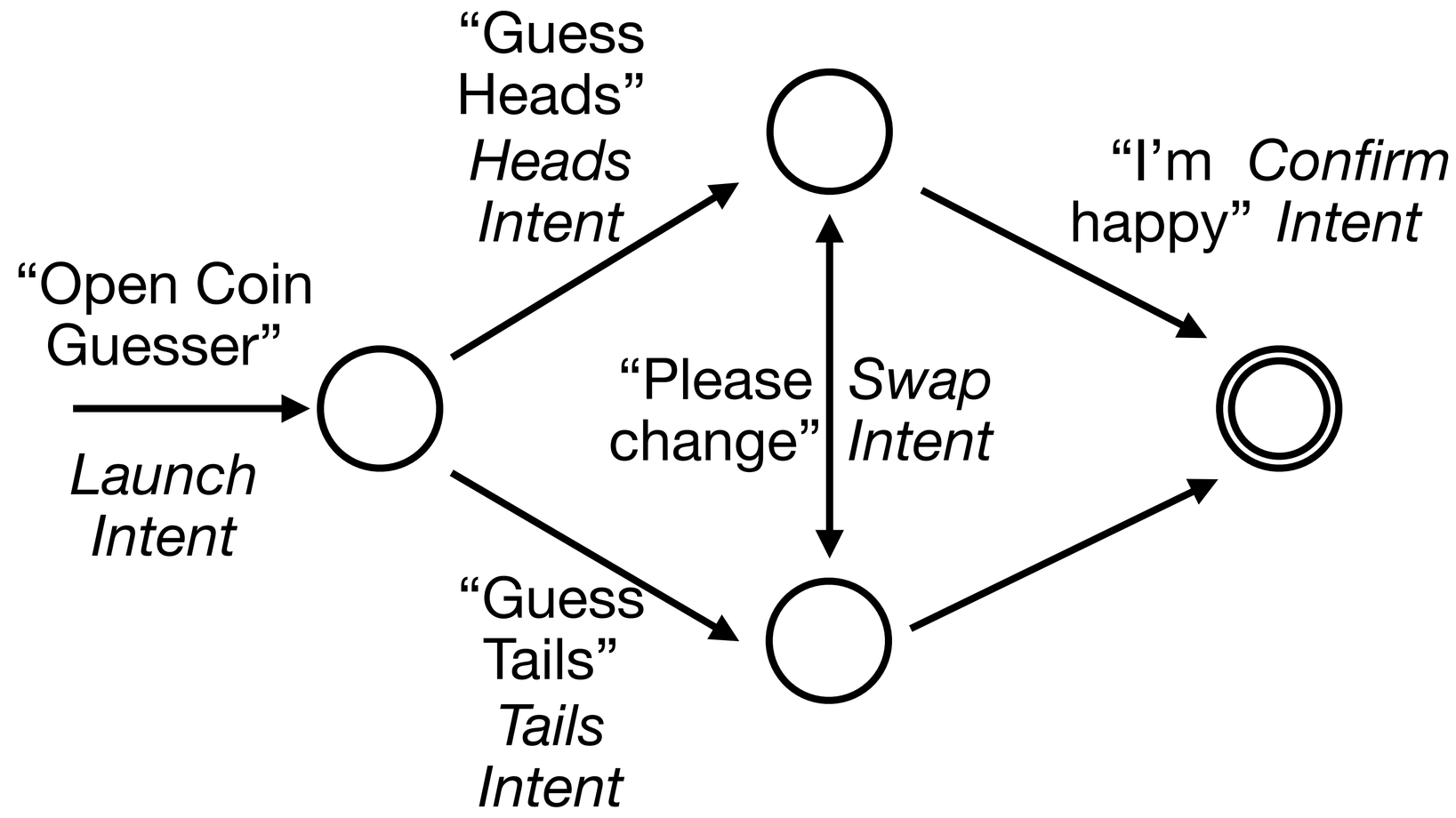
- Represents the “front end” of the skill
- Model that routes spoken requests to code pathways
- Rigid structure set by Amazon

AWS Lambda

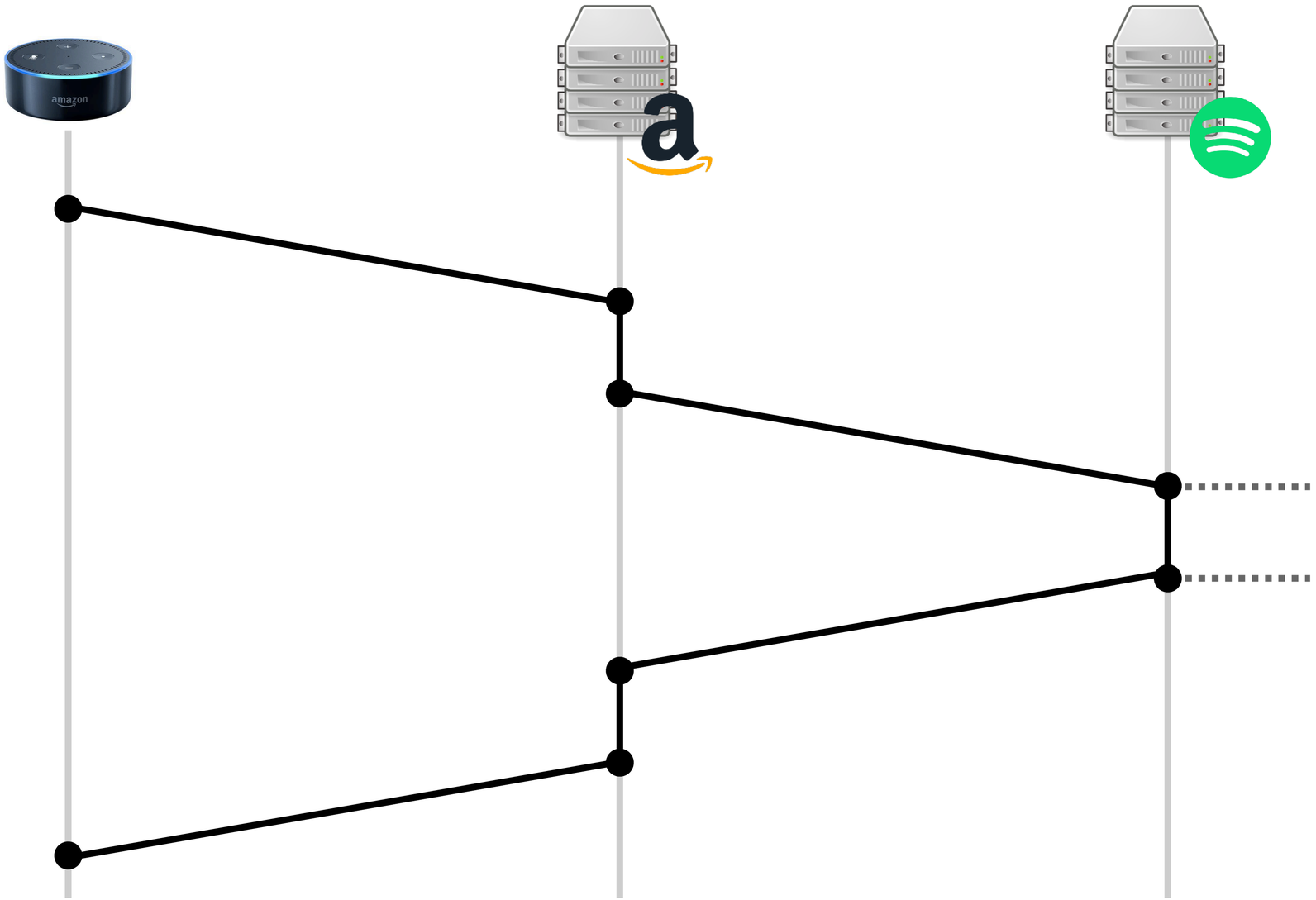
- Usually represents the “back end” of the skill
- Contains code that generates responses from user requests
- Controlled by developers

Alexa Skills Kit (ASK)

- Model build on **intents**, **utterances**, and **slots**
- **Intents** represent the types of requests that users can make to a skill
  - e.g. play a specific music track, shuffle an album, or quit
- **Utterances** capture the different ways an intent might be phrased
- **Slots** contain extra information needed to carry out an intent
  - e.g. the name of a track, a star sign, or the location for a weather forecast


"Guess Heads"
Heads Intent
"I'm happy" *Confirm Intent*
"Open Coin Guesser"
"Please change"
Swap Intent
Launch Intent
"Guess Tails"
Tails Intent

The Request Lifecycle



“Alexa, play Despacito”

Speech to text transcription

Matching intents and slots

Request stream from 3P

3P processing*

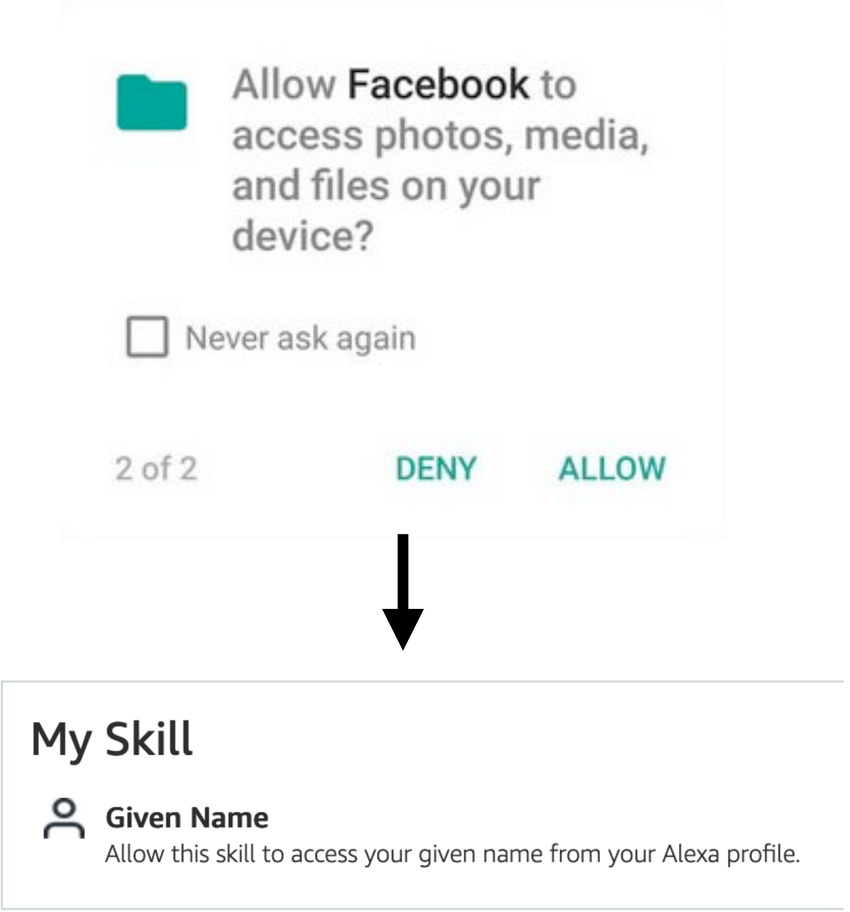
Formulate text response

Text to speech processing

Audible answer from device

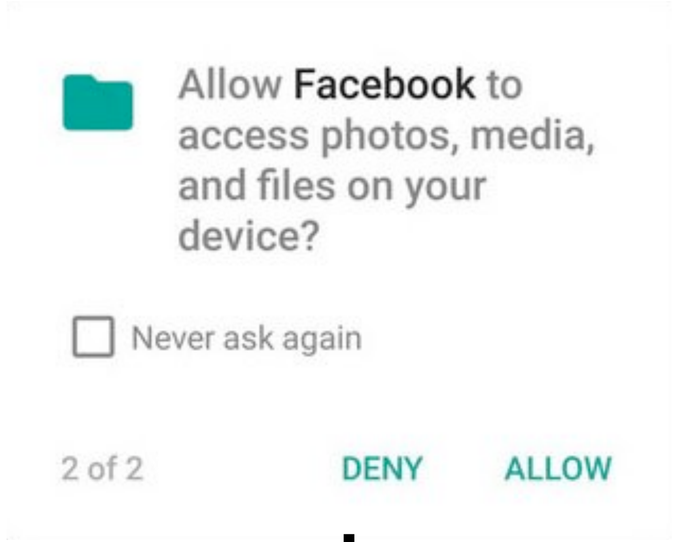
Data Collection

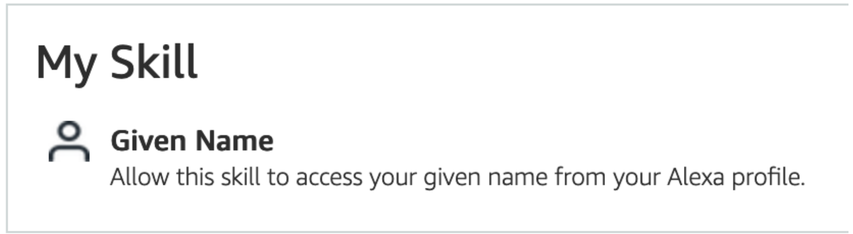
- Paradigm shift from functionality to data



Data Collection

- Paradigm shift from functionality to data
- Amazon retains the audio and transcript from every invocation
  - Including feedback from misfires and errors
- Information about engagement
- Correlation of usage with Amazon purchases
- Purchases made within skills
- Skills can collect their own data
  - This might be linked with external accounts





Data Protection Requirements

Pro tip: you can scare developers by saying “GDPR”

- In the UK and Europe there are regulations that restrict data collection and processing
- You need a legal basis for collection of personal data
- You need to tell people what you collect, how you use it, and their rights
- E.g. access, rectification, erasure
- Companies can be data controllers and/or data processors

ico.

Information Commissioner’s Office

GDPR, Consent, and Voice

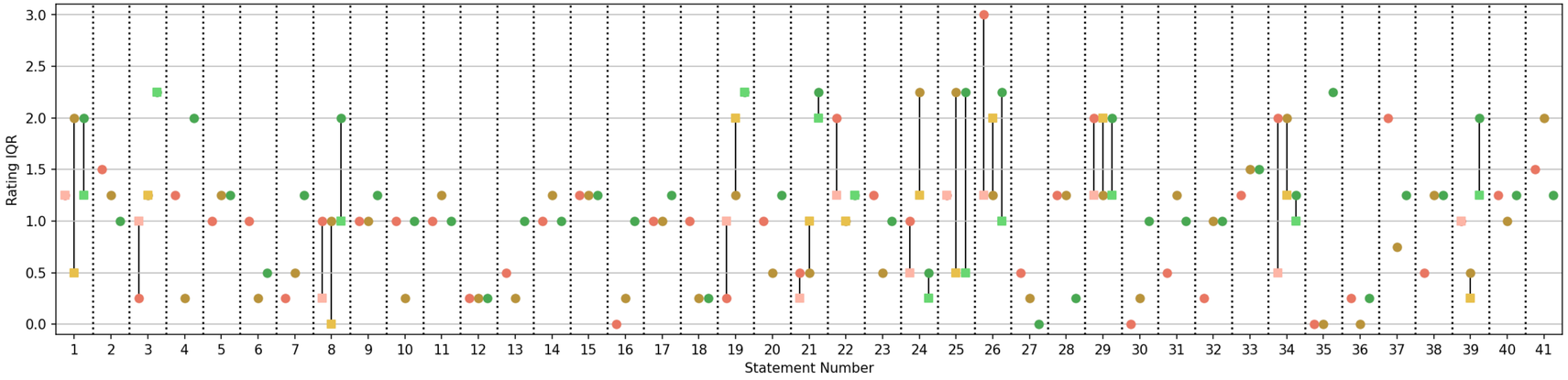
Paper Published at CHI 2023

- Investigating issues around “Voice Forward Consent”
- Delphi study with experts from academia, industry, and the policy sector

“Dialogues should say how users can withdraw their consent” 👍

“Consent should not be the legal ground used for data collection by skills” 👎

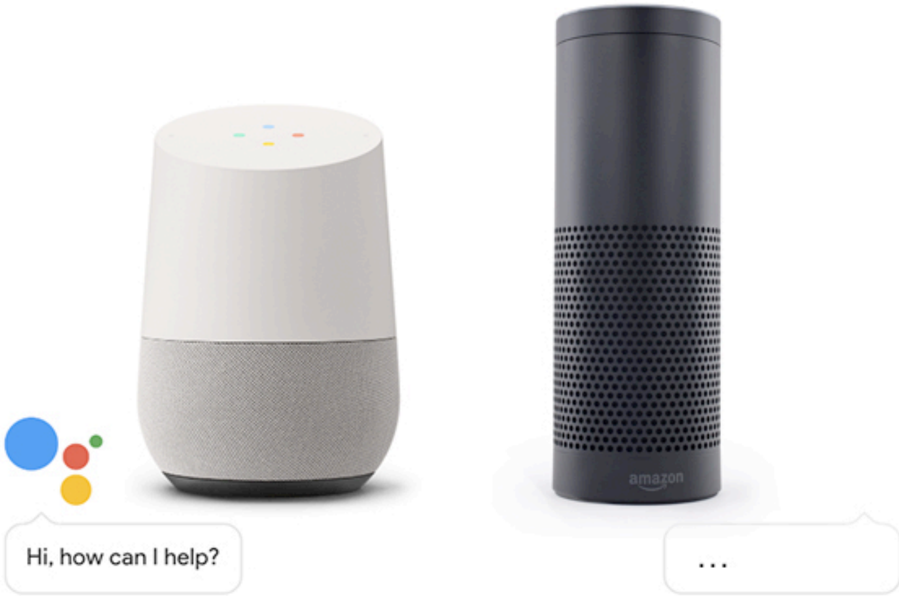
Processing purposes, controller identity, data types, retention, … 🤷



Break

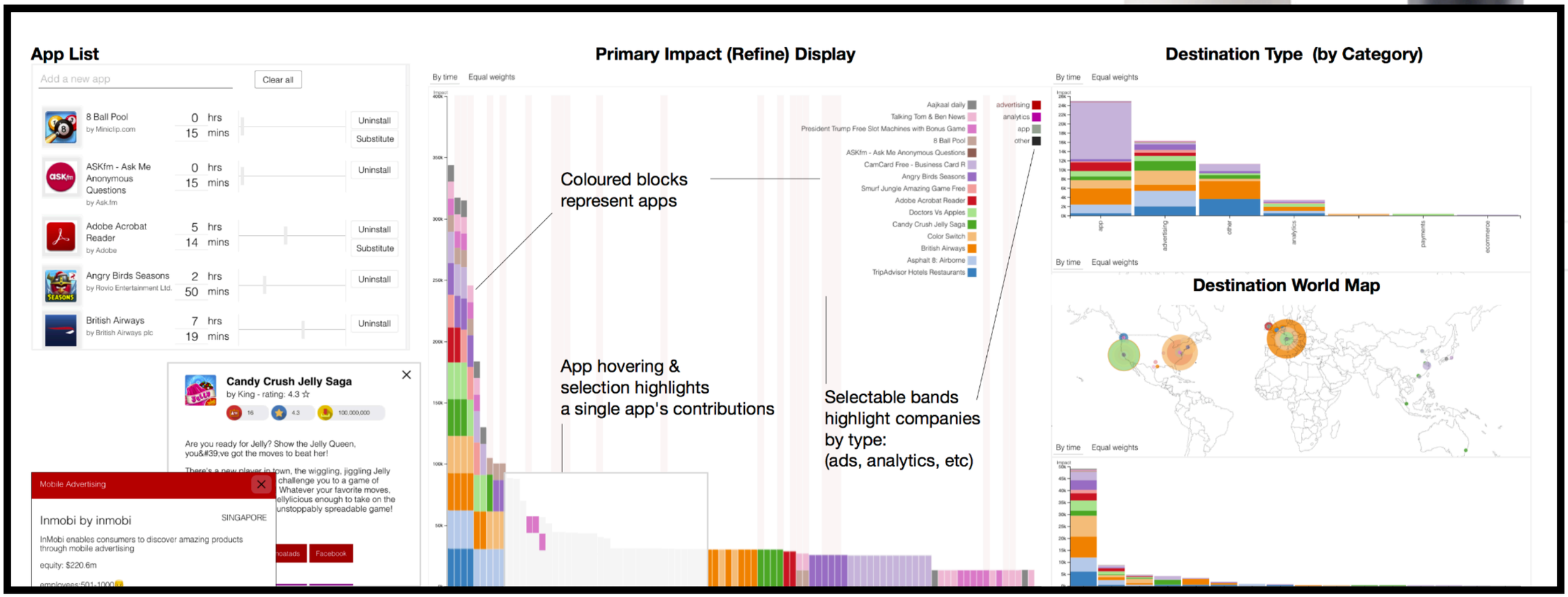
How can we do research on skills?

- The availability of source code poses a serious problem to research



How can we do research on skills?

- The availability of source code poses a serious problem to research



Max Van Kleek, Reuben Binns, Jun Zhao, Adam Slack, Sauyon Lee, Dean Ottewell, and Nigel Shadbolt. 2018. X-Ray Refine: Supporting the Exploration and Refinement of Information Exposure Resulting from Smartphone Apps. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18).

So what makes up a skill?

Alexa Skills Kit

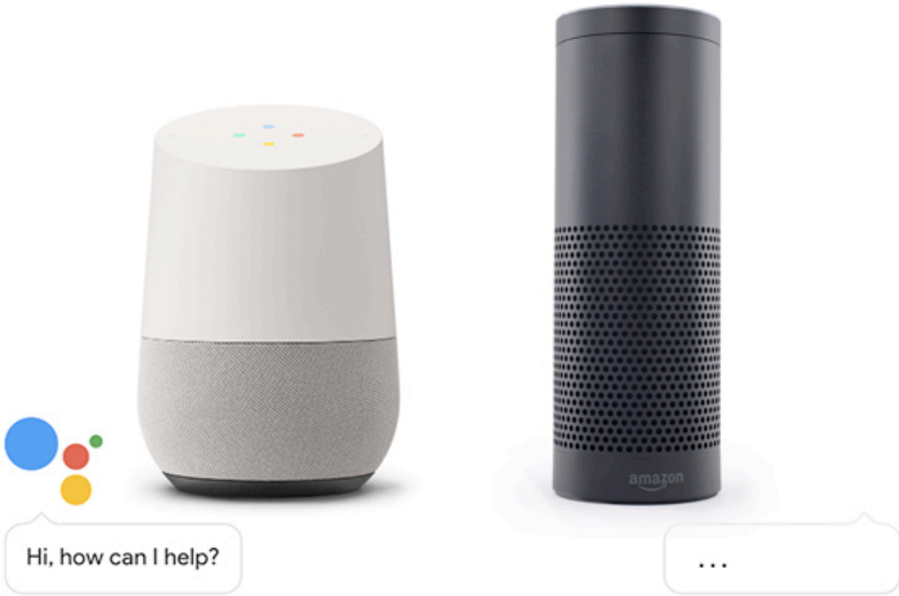
- Represents the “front end” of the skill
- Model that routes spoken requests to code pathways
- Rigid structure set by Amazon

AWS Lambda

- Usually represents the “back end” of the skill
- Contains code that generates responses from user requests
- **Controlled by developers**

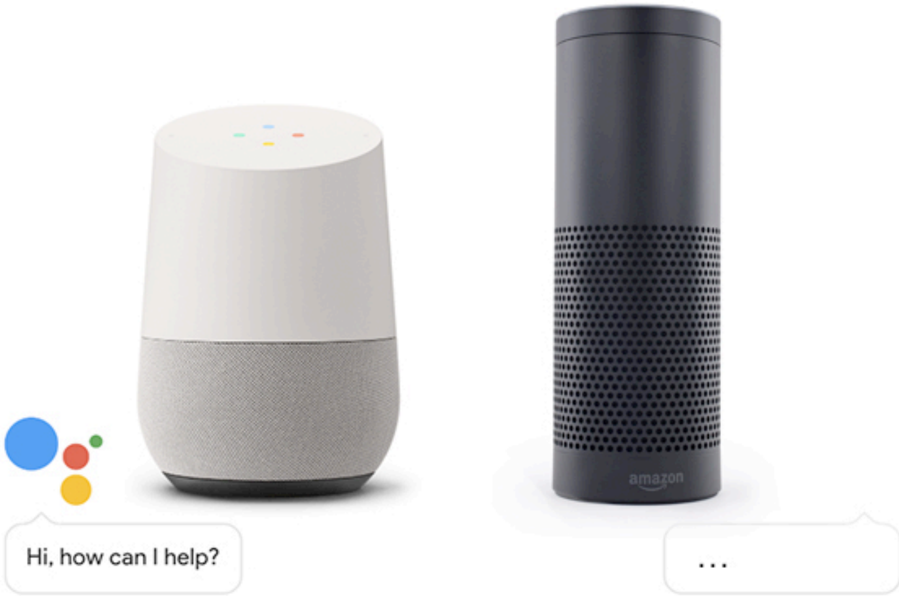
How can we do research on skills?

- The availability of source code poses a serious problem to research
- You also have to interact with them via natural language (i.e. no static/dynamic analysis etc.)
- Invoking skills via speech is tedious, what if we could automate it?



How can we do research on skills?

- The availability of source code poses a serious problem to research
- You also have to interact with them via natural language (i.e. no static/dynamic analysis etc.)
- Invoking skills via speech is tedious, what if we could automate it?

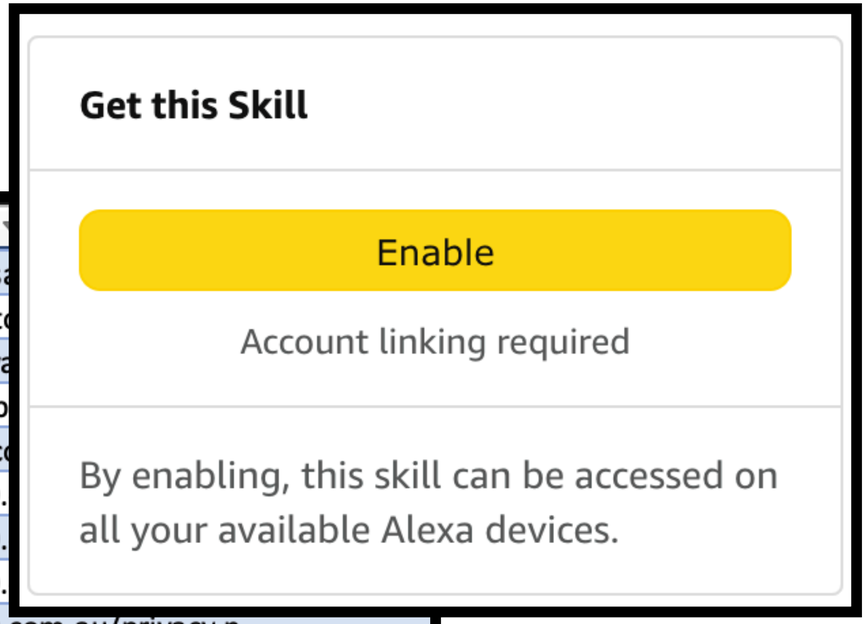


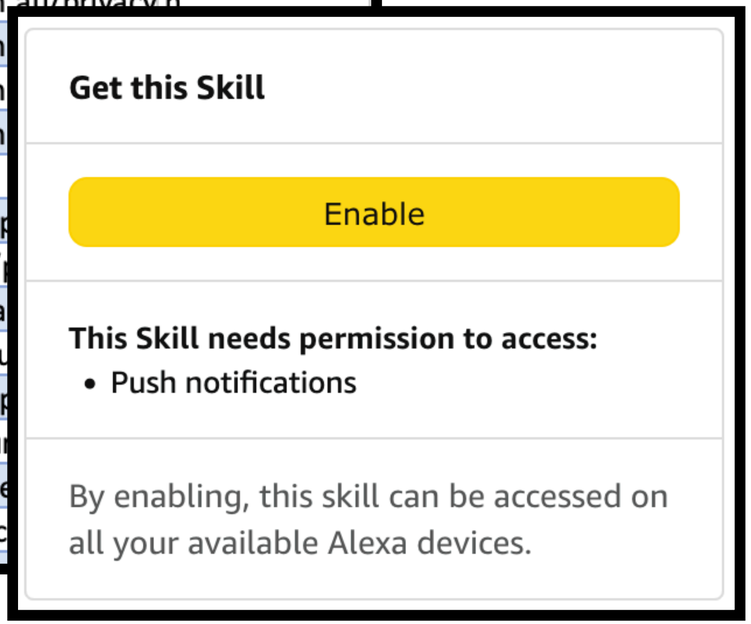
- Guess what, we have!

J. Edu, X. Ferrer Aran, J. Such and G. Suarez-Tangil, "SkillVet: Automated Traceability Analysis of Amazon Alexa Skills," in *IEEE Transactions on Dependable and Secure Computing*, doi: 10.1109/TDSC.2021.3129116.

Skill Store Data

skill_id	name	dev	perm	acc_linking	in_skill_purchase	policy	cat
B01KUGK3Q	CNBC	by NBCUnive	['E-mail Address']			http://www.nbcuniversa	
B08CJZDH2M	askPLS	by PLS Solicit	['First Name', 'E-mail Address']			https://www.pls-solicit	
B07TK1DGTE	VoiceTag	by Sweet As	['E-mail Address', 'Mobile Number']			https://voicetag.io/priva	
B07T97MQF	PMI Custome	by Project M	['E-mail Address']			https://www.pmi.org/p	
B089KM45ZI	Find Me A Jo	by ATMT Lab	['First Name Account linking required			https://www.snappcv.c	
B088LXW3M	Andrew Char	by IGNITE36(	['E-mail Address', 'Reminders']			https://www.ignite360.	
B088LX4X97	Matt Sebben	by IGNITE36(	['E-mail Address', 'Reminders']			https://www.ignite360.	
B088K4DZ8H	blackshaw re	by IGNITE36(	['E-mail Address', 'Reminders']			https://www.ignite360.	
B088JRPD4L	Blackshaw R	by IGNITE36(	['E-mail Address', 'Reminders']			https://www.ignite360.com.au/privacy n	
B088HBSLW	Mario Sanfra	by IGNITE36(	['E-mail Address', 'Reminders']			https://www.ignite360.com.au/privacy n	
B088GLC2MI	Blackshaw Ir	by IGNITE36(	['E-mail Address', 'Reminders']			https://www.ignite360.com	
B088FJM92X	chris churchi	by IGNITE36(	['E-mail Address', 'Reminders']			https://www.ignite360.com	
B088CSQM3	Blackshaw M	by IGNITE36(	['E-mail Address', 'Reminders']			https://www.ignite360.com	
B0874SVQQ(	The Creation	by The Creat	['Alexa Notifications']				
B084YXBTH6	Snappy Build	by APPY PIE I	['Full Name', 'E-mail Address', 'Mobile Number']			https://www.appypie.com/p	
B07VD7XTYB	Proximilator	by Proximity	['Device Add	Account linking required		https://www.iubenda.com/p	
B07TSSYYDB	Zurich Irelan	by Wolfgang	['First Name', 'E-mail Address']			https://www.zurich.ie/priva	
B07NGSFLFC	Doncaster Cc	by Doncaster	['Device Address', 'First Name']			http://www.doncaster.gov.u	
B07N2LMPN	HaitiNet	by KioskBeac	['Device Address']			https://www.haitinet.com/p	
B07MVX94R	Northumbria	by Northumb	['Device Country and Postcode']			https://www.nwl.co.uk/your	
B07LBNMGS	Davy	by Davy Grou	['Lists Read Access', 'Lists Write Access']			https://www.davy.ie/binarie	
B07KM3R49	Autochartist	by Autochart	['Device Country and Postcode', 'E-mail Address']			https://trader.autochartist.c	





SkillVet

Simplifying the Understanding of Data Disclosure Practices in AI Assistants for everyone

What's Skillvet | Explore Alexa Skills | Explore Skills Traceability | Check Traceability

What is Skillvet?

Skillvet is a privacy tool that help identifies potential privacy issues by analysing the traceability between the permissions and developers' data practices. It recognises relevant policy statements and assesses the traceability, classifying them as complete, partial, or broken. Skillvet, is based on machine learning and natural language processing.

Designed for developers
Developers can use this to write better privacy policy documents with relevant data practices.

Designed for Users
Users can use skillvet to quickly understand the data practices disclose in a privacy policy to make better decision about their data.

Designed for Regulators
Regulators can use skillvet to better understand the data practices disclose in a privacy policy to detect skills that violate existing regulation.

Mobile-friendly
Performs well on mobile devices shrinking down to appropriate size good enough to display on a mobile device .

Easy to Use
Support multiple file formats allowing users and developers to upload their documents without conversion.

Source files included
Links to the source file included to allow customization and better integration with multiple platforms.

Result

Traceability is Complete

Permission Requested:

1. device country and postal code

Permission Found:

1. Device Country And Postal Code

SkillVet

Simplifying the Understanding of Data Disclosure Practices in AI Assistants for everyone

What's Skillvet | Explore Alexa Skills | Explore Skills Traceability | Check Traceability

What is Skillvet?

Skillvet is a privacy tool that help identifies potential privacy issues by analysing the traceability between the permissions and developers' data practices. It recognises relevant policy statements and assesses the traceability, classifying them as complete, partial, or broken. Skillvet, is based on machine learning and natural language processing.

Designed for developers
Developers can use this to write better privacy policy documents with relevant data practices.

Designed for Users
Users can use skillvet to quickly understand the data practices disclose in a privacy policy to make better decision about their data.

Designed for Regulators
Regulators can use skillvet to better understand the data practices disclose in a privacy policy to detect skills that violate existing regulation.

Mobile-friendly
Performs well on mobile devices shrinking down to appropriate size good enough to display on a mobile device .

Easy to Use
Support multiple file formats allowing users and developers to upload their documents without conversion.

Source files included
Links to the source file included to allow customization and better integration with multiple platforms.

Result

Traceability is Complete

Permission Requested:

1. device country and postal code

Permission Found:

1. Device Country And Postal Code
2. Device Address
3. Email Address
4. Location Services
5. Personal Information

Exploring Skills

- Choose ~4 skills to explore, ideally some you already know/use
- Make sure that some use permissions (you can check the dataset for this)
- Why do the skills need the permissions they ask for?
- Do their privacy policies cover them?
- Are their privacy policies too broad?
- What were the key influences on permissions choices and policy wording?

skillvet.nms.kcl.ac.uk

https://github.com/secure-ai-assistants/SkillWorkshop/tree/main/data

https://www.amazon.co.uk/b?node=10068517031

secure-ai-assistants.github.io/surveys

Break

Group Discussion II - Pros and Cons

- Starting with your examples from earlier, what context are people using the skills in?
  - Time, mood, pressures, vulnerabilities…
- For each of these, what might the disadvantages of using a VA be?
  - e.g. at a personal/societal level
- How could we keep the benefits without incurring the drawbacks?
- What data collection is necessary, and what data collection is unnecessary?

secure-ai-assistants.github.io/surveys

Regroup in 15 minutes

Speech as an Interaction Modality

- We treat speech as conceptually different to text-based communication
- We get a lot more information from speech through tone, cadence, etc.
- Speech often contains ambiguity
- We often rely on context to resolve this, which is tricky for VAs
- Much more potential to “mishear” than “misclick”

Computers as Social Actors

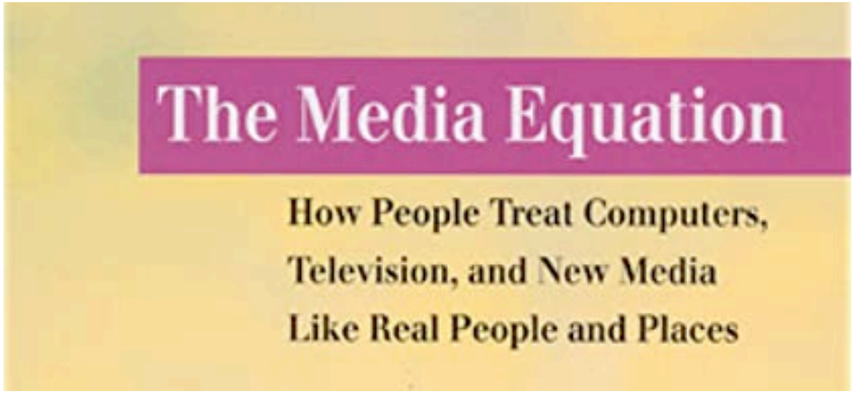
1. Pick a social science finding (theory and method) which concerns behaviour or attitude toward humans

2. Change “human” to “computer” in the statement of the theory

3. Replace one or more humans with computers in the method of the study

4. Provide the computer with characteristics associated with humans

5. Determine if the social rule still applies



Computers as Social Actors

- People apply social norms and personal biases to computers
- Inferred characteristics of voices, e.g. gender, are very important
- This happens automatically and subconsciously even though people know they are talking to computers

RESEARCH-ARTICLE

AddThis

Exploring Interactions Between Trust, Anthropomorphism, and Relationship Development in Voice Assistants

Authors: William Seymour, Max Van Kleek Authors Info & Claims

Proceedings of the ACM on Human-Computer Interaction, Volume 5, Issue CSCW2 • Article No.: 371, pp 1–16 • https://doi.org/10.1145/3479515

Published: 18 October 2021 Publication History

Check for updates

William Seymour and Max Van Kleek. 2021. Exploring Interactions Between Trust, Anthropomorphism, and Relationship Development in Voice Assistants. Proc. ACM Hum.-Comput. Interact. 5, CSCW2, Article 371 (October 2021), 16 pages. https://doi.org/10.1145/3479515

Performance of Gender

THINK PIECE 2

THE RISE OF GENDERED AI AND ITS TROUBLING REPERCUSSIONS

The New York Times

IN HER WORDS

Hey, Alexa, Are You Sexist?

"Add milk to the grocery list." "Switch off the lights." It's no coincidence that our robotic helpers at home have feminine voices.

Meet Q

The First Genderless Voice

Project Q launched in 2019 with a prototype voice developed by non-binary linguists. They then asked a sample of 4,500 people from across Europe whether it sounded male or female. The answers were split 50/50.

upbeat geekiness he wanted. The backstory is charmingly specific: She comes from Colorado, a state in a region that lacks a distinctive accent. "She's the youngest daughter of a research librarian and a physics professor who has a B.A. in art history from Northwestern," Giangola continues. When she was a child, she won $100,000 on *Jeopardy: Kids Edition*. She used to work as a personal assistant to "a very popular late-night-TV satirical pundit." And she enjoys kayaking.

Performance of Gender

Project Q launched in 2019 with a prototype voice developed by non-binary linguists. They then asked a sample of 4,500 people from across Europe whether it sounded male or female. The answers were split 50/50.

The New York Times

IN HER WORDS

Hey, Alexa, Are You Sexist?

"Add milk to the grocery list." "Switch off the lights." It's no coincidence that our robotic helpers at home have feminine voices.

upbeat geekiness he wanted. The backstory is charmingly specific: She comes from Colorado, a state in a region that lacks a distinctive accent. "She's the youngest daughter of a research librarian and a physics professor who has a B.A. in art history from Northwestern," Giangola continues. When she was a child, she won $100,000 on *Jeopardy: Kids Edition*. She used to work as a personal assistant to "a very popular late-night-TV satirical pundit." And she enjoys kayaking.

Performance of Gender

Project Q launched in 2019 with a prototype voice developed by non-binary linguists. They then asked a sample of 4,500 people from across Europe whether it sounded male or female. The answers were split 50/50.

The New York Times

IN HER WORDS

Hey, Alexa, Are You Sexist?

"Add milk to the grocery list." "Switch off the lights." It's no coincidence that our robotic helpers at home have feminine voices.

upbeat geekiness he wanted. The backstory is charmingly specific: She comes from Colorado, a state in a region that lacks a distinctive accent. "She's the youngest daughter of a research librarian and a physics professor who has a B.A. in art history from Northwestern," Giangola continues. When she was a child, she won $100,000 on *Jeopardy: Kids Edition*. She used to work as a personal assistant to "a very popular late-night-TV satirical pundit." And she enjoys kayaking.

Performance of Gender

Project Q launched in 2019 with a prototype voice developed by non-binary linguists. They then asked a sample of 4,500 people from across Europe whether it sounded male or female. The answers were split 50/50.

The New York Times

IN HER WORDS

Hey, Alexa, Are You Sexist?

"Add milk to the grocery list." "Switch off the lights." It's no coincidence that our robotic helpers at home have feminine voices.

upbeat geekiness he wanted. The backstory is charmingly specific: She comes from Colorado, a state in a region that lacks a distinctive accent. "She's the youngest daughter of a research librarian and a physics professor who has a B.A. in art history from Northwestern," Giangola continues. When she was a child, she won $100,000 on *Jeopardy: Kids Edition*. She used to work as a personal assistant to "a very popular late-night-TV satirical pundit." And she enjoys kayaking.

Group Discussion III - Impact of Speech

- In the examples you've been developing, how might the use of speech impact the interaction?
- Could it be used to persuade people to give up more data?
- Are there situations where speech shouldn't be used?
- How should developers and/or users choose how their assistant sounds?
- Come up with a vignette that plays on or subverts some of these dimensions through the use of delivery and dialogue
  - e.g. making the interaction more human- or machine-like
  - e.g. making the interaction more or less personalised through the use of additional data
  - e.g. more open to compromise, questioning, assent, or dissent

secure-ai-assistants.github.io/surveys

Regroup in 15 minutes

Final Discussion

- Give a quick overview of one of your group’s discussions
- ~4 mins each, choice of:
- How did your skills compare in terms of permissions, privacy policies, etc.?
- How do the pros and cons of using voice assistants balance out for individuals and society?
- What is the impact of using speech in the skills you chose?

- Opportunity to have a quick discussion about something you’re interested in
- Next Friday you’ll be presenting your work today in the seminar

Break

Final Discussion

- Give a quick overview of one of your group’s discussions
- ~4 mins each, choice of:
- How did your skills compare in terms of permissions, privacy policies, etc.?
- How do the pros and cons of using voice assistants balance out for individuals and society?
- What is the impact of using speech in the skills you chose?

- Opportunity to have a quick discussion about something you’re interested in
- Next Friday you’ll be presenting your work today in the seminar

Thanks!

william.1.seymour@kcl.ac.uk

References

- Zuboff, Shoshana. “The age of surveillance capitalism: The fight for a human future at the new frontier of power.” Profile books, 2019.
- W. Seymour, M. Cote, and J. Such. “Legal Obligation or Ethical Best Practice? Towards Meaningful Verbal Consent for Voice Assistants.” Proceedings of the SIGCHI conference on Human factors in computing systems. 2023.
- M. Van Kleek, R. Binns, J. Zhao, A. Slack, S. Lee, D. Ottewell, and N. Shadbolt. 2018. X-Ray Refine: Supporting the Exploration and Refinement of Information Exposure Resulting from Smartphone Apps. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18).
- J. Edu, X. Ferrer Aran, J. Such and G. Suarez-Tangil, "SkillVet: Automated Traceability Analysis of Amazon Alexa Skills," in *IEEE Transactions on Dependable and Secure Computing*, doi: 10.1109/TDSC.2021.3129116. C. Nass, J. Steuer, and E. Tauber. "Computers are social actors.” Proceedings of the SIGCHI conference on Human factors in computing systems. 1994.
- W. Seymour and M. Van Kleek. “Exploring Interactions Between Trust, Anthropomorphism, and Relationship Development in Voice Assistants.” Proceedings of the ACM on Human-Computer Interaction. 2021.
- M. West, R. Kraut, and H. Chew. "I'd blush if I could: closing gender divides in digital skills through education.” UNESCO Policy Paper. 2019.
- C. Rincón, O. Keyes, and C. Cath. "Speaking from Experience: Trans/Non-Binary Requirements for Voice-Activated AI." Proceedings of the ACM on Human-Computer Interaction. 2021.