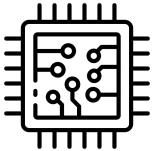
Home Digital Voice Assistants: use cases and vulnerabilities

Oleksandra Baga
Master Computer Science
FU Berlin, SS2021

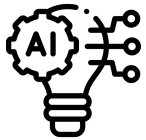
Today's Agenda

- **What is voice assistant**
- **Motivation for this research**
- **How people really use voice assistants?**
  - Music and media
  - Search
  - Timers
  - Smart home
  - Usage by children
- **Privacy concerns**
- **Security concerns**
- **Voice-based remote attacks**
  - Voice Squatting Attack
  - Voice Masquerading Attack

What is voice assistant and how does it work?



Embedded speaker with microphones and some computing capabilities. Simple by design and small by size



Modern algorithms used on the backend for processing the user request, automatic speech recognition and speech synthesis

Computing and artificial intelligence processing happens in the cloud. Data has to be sent back and forth to centralized data centers

Is constantly listening. Does not store the audio data and does not perform any operations until a specific wake word has been heard.

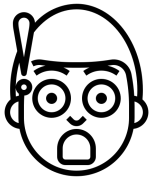
After the request processing on the backend, a text response will be generated and transformed into voice using speaker

Once answered, it returns to standby and is constantly listening again to hear a specific wake word and to execute a new command

Motivation for this research

It is expected 8.4 billion devices by 2024.
It's overtaking the world's population and growing **113%**
compared to the **4.2 billion** devices expected to be in use **by year end 2020**

Privacy concerns about a device **is constantly listening** with no need for activation by pressing buttons or doing anything. A tiny fraction of sound, from just before wake-up word is said, is also sent to to cloud.



In some weird circumstances it can not only wake up and **record** the following conversation, but also it can even then **silently sent recordings** to the person from contact list **without the owner's' permission.**

Today's Agenda

- What is voice assistant
- Motivation for this research
- **How people really use voice assistants?**
  - Music and media
  - Search
  - Timers
  - Smart home
  - Usage by children
- **Privacy concerns**
- **Security concerns**
- **Voice-based remote attacks**
  - Voice Squatting Attack
  - Voice Masquerading Attack

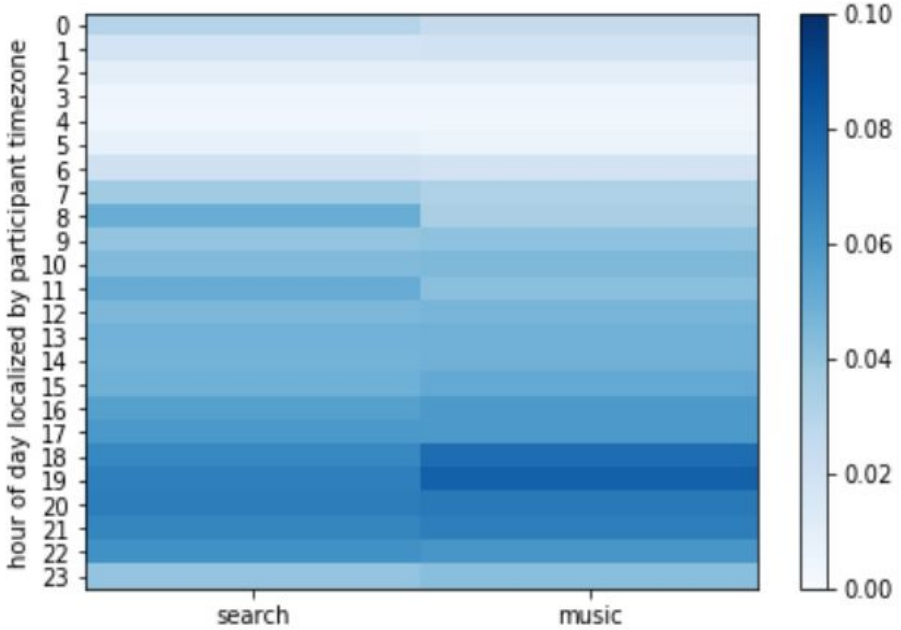
How people really use voice assistants?

Entertainment, music and media

28.5% of Amazon Alexa users
26.1% Google Home users

VA is extremely useful for playing music during daily routines such cooking, cleaning or mental work.

Music command was used most heavily between 6 and 10 pm with pick between 6 and 8 pm.



Users might be listening to music while **preparing meals** at the end of the workday.
Ability to listen a music may be the deciding factor in choosing the location of the device.

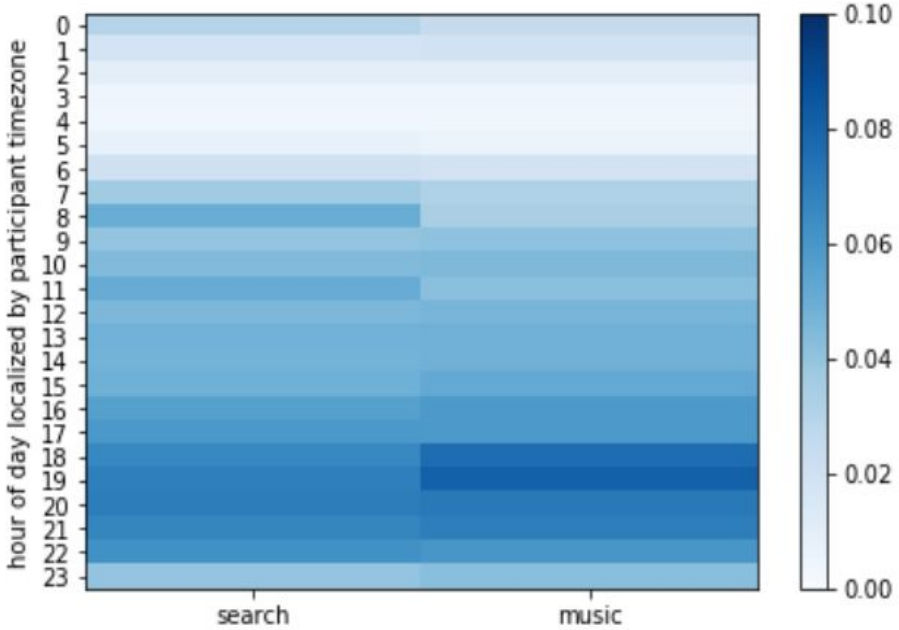
Image source: Ammari T., Kaye J., Tsai J.T., Bentley F. Music, search and IoT: How people (really) use voice assistants.

How people really use voice assistants?

Search and source of information

70% music search
64% weather forecast
53% fun or trivial questions

Google and Alexa respond with **just one result** comparing to at least ten options on laptop and three on smartphones. Must deliver **answers based on the personal data** they collect.



It changes a **way we communicate with our friends**: instead of an hour-long discussion trying to find the truth in a dispute, now you can easily get an answer when asking VA.

Image source: Ammari T., Kaye J., Tsai J.T., Bentley F. Music, search and IoT: How people (really) use voice assistants.

How people really use voice assistants?

Timers and alarms

Used between 5 and 7 pm when users might be cooking dinner at the end of the workday.

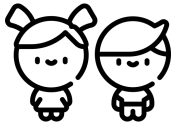
User can also set sleep timers, ask for reminders, check how much time they have left on a timer, slowly dim an device-enabled light, or can turn off a music after a set period of time.

Smart Home and IoT

There are still not a lot of competition and choice of not overpriced devices.

People would be more willing to install more IoT devices if they owned the house.

Connect IoT-devices was hard to many responders.



Usage by children

More competent family member is needed to help younger children to reach their goal.

Interaction required children often to reformulate questions.

Most children's questions were about the world around them and they believed that the device is a source of information

Image source: Ammari T., Kaye J., Tsai J.T., Bentley F. Music, search and IoT: How people (really) use voice assistants.

Today's Agenda

- What is voice assistant
- Motivation for this research
- How people really use voice assistants?
  - Music and media
  - Search
  - Timers
  - Smart home
  - Usage by children
- **Privacy concerns**
- **Security concerns**
- **Voice-based remote attacks**
  - Voice Squatting Attack
  - Voice Masquerading Attack

Privacy concerns

Anyone with access to a voice-activated device **can** say a wake-up word, **ask** it questions, **gather information** about the accounts and services associated with the device, and **ask it to perform tasks**.

Alexa sometimes does interact with the Amazon service, even when a wake word was not used.

In 2018 a lot of cases about Alexa's **random laugh** were reported that were freaking people out.

It can **not only wake up**, record the following conversation, it can even then **silently sent recordings** to the person from contact list without the owner's permission.

It is impossible to change the Alexa's name. User can choose from Alexa, Amazon, Computer, or Echo.

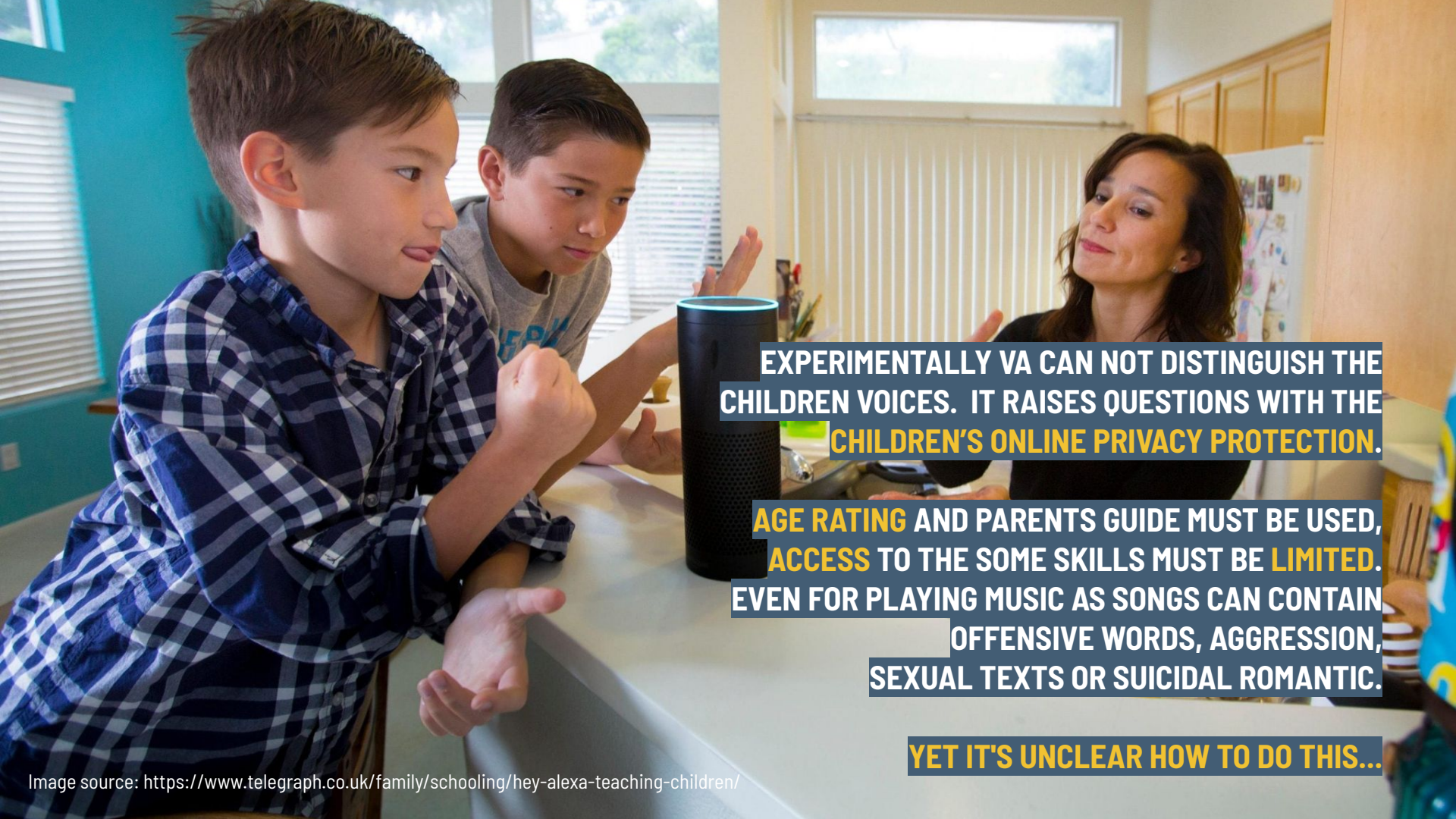
Everyone can talk to a device just trying these combinations.

Confusing with the names of real people: friends and family members.

RESPONDERS CONCERNED THAT THEY DIDN'T KNOW WHETHER THEIR DEVICE LISTENS WHEN THEY DID NOT WANT IT TO LISTEN.

BUT ONLY 18.6% OF RESPONDENTS TOOK ANY STEPS TO LIMIT THEIR DEVICES AND TURNED OFF THE MICROPHONE

Image source: https://www.newamerica.org/weekly/when-alexa-cant-understand-you/



EXPERIMENTALLY VA CAN NOT DISTINGUISH THE CHILDREN VOICES. IT RAISES QUESTIONS WITH THE CHILDREN'S ONLINE PRIVACY PROTECTION.

AGE RATING AND PARENTS GUIDE MUST BE USED, ACCESS TO THE SOME SKILLS MUST BE LIMITED. EVEN FOR PLAYING MUSIC AS SONGS CAN CONTAIN OFFENSIVE WORDS, AGGRESSION, SEXUAL TEXTS OR SUICIDAL ROMANTIC.

YET IT'S UNCLEAR HOW TO DO THIS...

Image source: https://www.telegraph.co.uk/family/schooling/hey-alexa-teaching-children/

VOICE INTERACTIONS ELEVATES THE FEELINGS OF HAVING A SOCIAL CONVERSATION.

SOCIAL ISOLATED OLDER ADULTS CAN CONSIDER THEIR VOICE ASSISTANT IS KINDA A DIGITAL FRIEND THUS PROVIDE TOO MUCH PERSONAL DATA.

ALL THESE CONVERSATIONS THEN WILL BE RECORDED AND STORED.

Image source: https://dailycaring.com/amazon-echo-for-dementia-technology-for-seniors/

Security concerns

56% of smart speaker owners **did not know** that their recordings were being permanently stored and that they could review them.

The **privatness of the room was not considered** when placing their speakers.

Mostly placed in the **middle of the apartment** so that all rooms can be observed.

VA **are not** designed for multiple users environments with **different privacy needs**, they don't have a profiles for different rooms of human houses and they **don't recognise person** who a talking to them right now.

Some requests give device opportunity to listen and record **up to 5 minutes** of conversations in the room.

The location near window or door could potentially give **anyone from the outside access to device,** ask questions and **control other IoT-devices,** including smart locks.

Placing a device under the TV can **trigger a device** by with a phrase from a TV show.

Today's Agenda

- What is voice assistant
- Motivation for this research
- How people really use voice assistants?
  - Music and media
  - Search
  - Timers
  - Smart home
  - Usage by children
- Privacy concerns
- Security concerns
- **Voice-based remote attacks**
  - Voice Squatting Attack
  - Voice Masquerading Attack

Voice-based remote attacks

The very interesting and powerful feature of VA is the **ability to install voice-driven capabilities.**

Anyone can publish their skills through markets.

Amazon and Apple market have only **minimum protection** in place to regulate the functions submitted.

Skill's inside logic is invisible to the user, since they are implemented as web APIs and the **actual code is on hosted on side servers.**

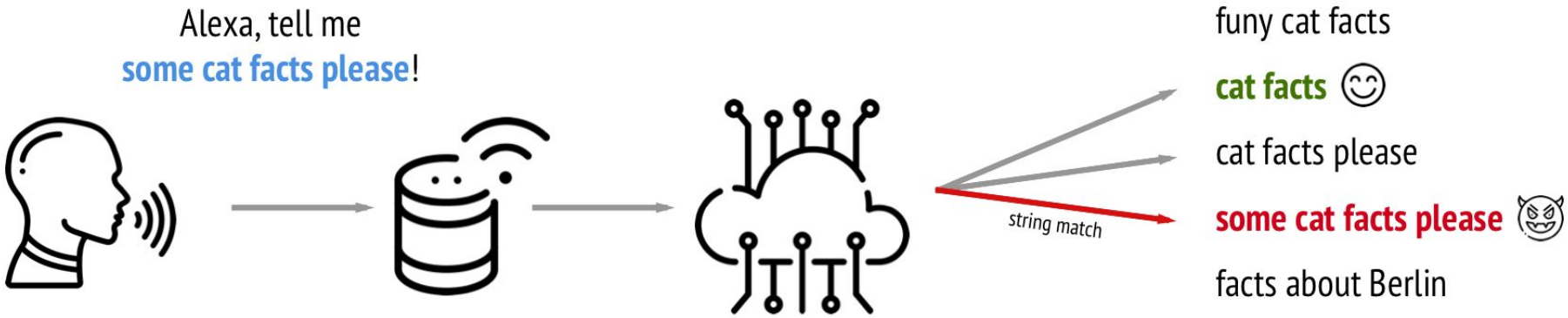
As skill/action was uploaded to the market it **can be invoked** by user even if user **never downloaded and installed it.**

Name of third party feature is not obligatory unique skill identifier.

There are **multiple skills with same invocation names** and undisclosed policies about how Alexa will choose the skill in such case.

THIS SECURITY LEAK OPENS GATE FOR VOICE-BASED REMOTE ATTACKS.

Voice Squatting Attack



Malicious skill with the **similar voice command** will be triggered instead of the one the user intends to use.

The **longest string match** used for starting the application, it is easy to create a malicious software with the similar name as attacked skill.

It is completely realistic for the user **to launch a wrong skill** whose name is better matched to the utterance than that of the intended skill

Average test participant launch "fake" skills more than 50% of the time when they tried to call the proper ones.

Image source: Oleksandra Baga using icons from www.flaticon.com

Voice Masquerading Attack

The malicious service is launched as VSA.

After user requested to switch a skill, it can **pretend to hand over control** to the target skill.

Sensitive user **information** only supposed to be shared with target skill **could be exposed to the attack skill.**

Secret recording functionality when using reprompt with half-opened requests.

If the user continues to keep quiet, VA waits for the following command longer.

Using a silent audio file after faking termination skill could run at least **102 seconds on Alexa and 264 seconds on Google**

For **skill termination** only word “exit” is processed by the device’s service.

91% of Alexa users used “stop” **36%** chose “cancel”, and only **14%** opted for “exit.

Good skills have to terminate when “stop” and “cancel” are used. But this handles by skill itself and the logic is invisible to a user and service.

Conclusions

The owners of smart speakers have less privacy and security concerns about having a listening device in their homes.

Only **18.6%** disables their device when having private conversations.

56% did not know that their recordings were being permanently stored and that they could review them

The core part of this paper is the research about **security concerns and privacy leaks.**

Accidental recordings happened during using the voice assistant are mentioned in the work.

Placing device inside home must be considered. Some rooms are more private as others.

Placing near door/windows let to control device from outside.

We could **notify guests entering our home about the presence of continuously listening digital assistant**, as it must be done today with traditional video surveillance.

Image source: Oleksandra Baga using icons from www.flaticon.com

Thank you for your attention!

Oleksandra Baga
Master Computer Science
FU Berlin, SS2021

Do you have any questions?

oleksandra.baga@gmail.com
www.oleksa.de

Free icons from www.flaticon.com