INTRODUCTION

Goal Statement:
- In 2021, nearly 91 million smart speakers were installed in the US. Given the prominence of smart speakers, data privacy and security have become major societal concerns. This project focuses on the Amazon Echo device. We analyzed online ads and checked for targeted ads to evaluate the collection, usage, and sharing of voice data with third parties.
- This project strives to increase transparency and better understand how data is used. This increases consumer awareness and ensures that company policies align with actual practices and the federal law. We hope to make an effort toward promoting ethical corporate standards.

Skills Developed and Purpose
- Religion and Spirituality Persona
  - Methods (using MyCroft):
    - Use Mycroft Skills Manager and PuTTY to add skills from Mycroft and GitHub
  - Methods (using Amazon Echo):
    - Run Python scripts that automate downloading scripts on Amazon website ( Marionette tool) and tell Alexa to run skills using automated voice
  - Skills: Internet radio, audio bible, random bible verse, prayer time reminder, translator, etc.
  - Purpose: Develop persona to analyze voice data usage, collection, and sharing
    - Check for trackers in network traffic (DNS packets)
    - Check for advertisers syncing cookies with Amazon

- Alexa Data
  - Cost Per Mille (CPM) Bid Values:
    - Price advertisers are willing to pay per thousand ads seen
  - Run Python scripts that automate process of visiting websites and collect ad data/images
  - Analysis:
    - For many personas, the CPM is higher than the vanilla persona
    - Other tests (i.e. Mann-Whitney U test*) demonstrate that interest personas have higher bids than the vanilla persona
  - Findings: Amazon uses voice data to determine user interests and create targeted ads.

Building Mycroft (Pi Experience)

- Features: Wake word (e.g. “Hey Mycroft”), speech-to-text (cloud), skills, voice/debug log (“mycroft-cli-client” command)
- Why project is unique/organic: Open Source software allows features to be built from scratch/customized. It can be compared with commercial IoT devices.
- Future innovations/interests:
  - Breadth: Comparison across different IoT devices and IoT devices on mobile phones
  - Depth: Breakdown of personas (multivariate) into age group, sex and/or gender
  - Solutions: Method to process/store voice data that satisfies both companies and consumers
- Impact of experience: We learned how voice assistants use data, research methods, tools (e.g. tcpdump, PuTTY), and teamwork!

ACKNOWLEDGEMENTS

Special thanks to Ernest Garrison, Rahmadi Trimananda, Dr. Umar Iqbal, Janus Varmarken, Dr. Athina Markopoulou, Marilyne Tamayo, and ProperData guest speakers!