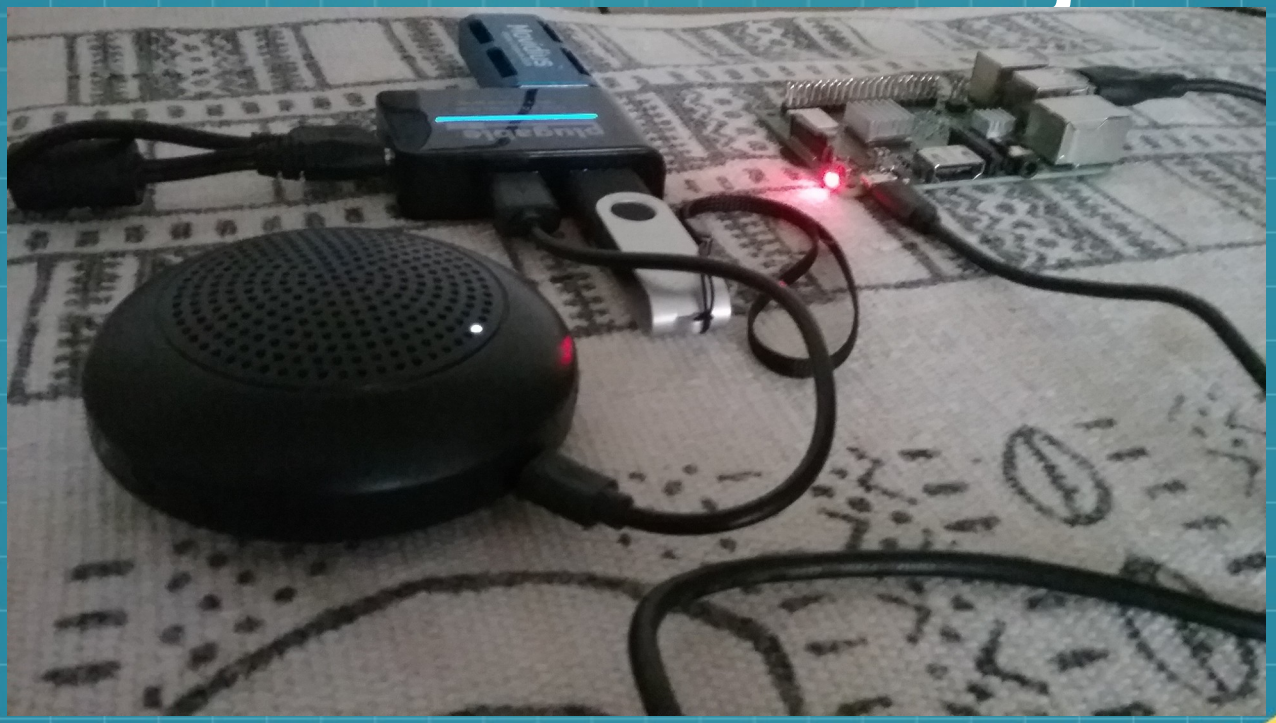
The Naomi Project



Outline

- Why now?
  - Convenience
  - Safety
  - Privacy
- Notable OpenSource Projects
- Parts of a voice assistant
  - ASR
  - Intent Parsing
  - Text to Speech
- Naomi
  - Collect and format samples for training
  - Simple, and developer friendly
  - How can I help?

Why Now?

- A verbal interface is helpful any time you can’t use your hands or your eyes
- A verbal interface has been a goal forever
- Commercial Assistants have a lot of room for improvements

I grew up with R2D2 and K-9

The goal of Smart Speakers

These and other tech corporations have grand ambitions. They want to colonize space. Not interplanetary space. Everyday space: home, office, car. In the near future, everything from your lighting to your air-conditioning to your refrigerator, your coffee maker, and even your toilet could be wired to a system controlled by voice.

-SHULEVITZ, J. (2018). “Alexa, HOW WILL YOU CHANGE US?” *Atlantic*, *322*(4), 94–104.

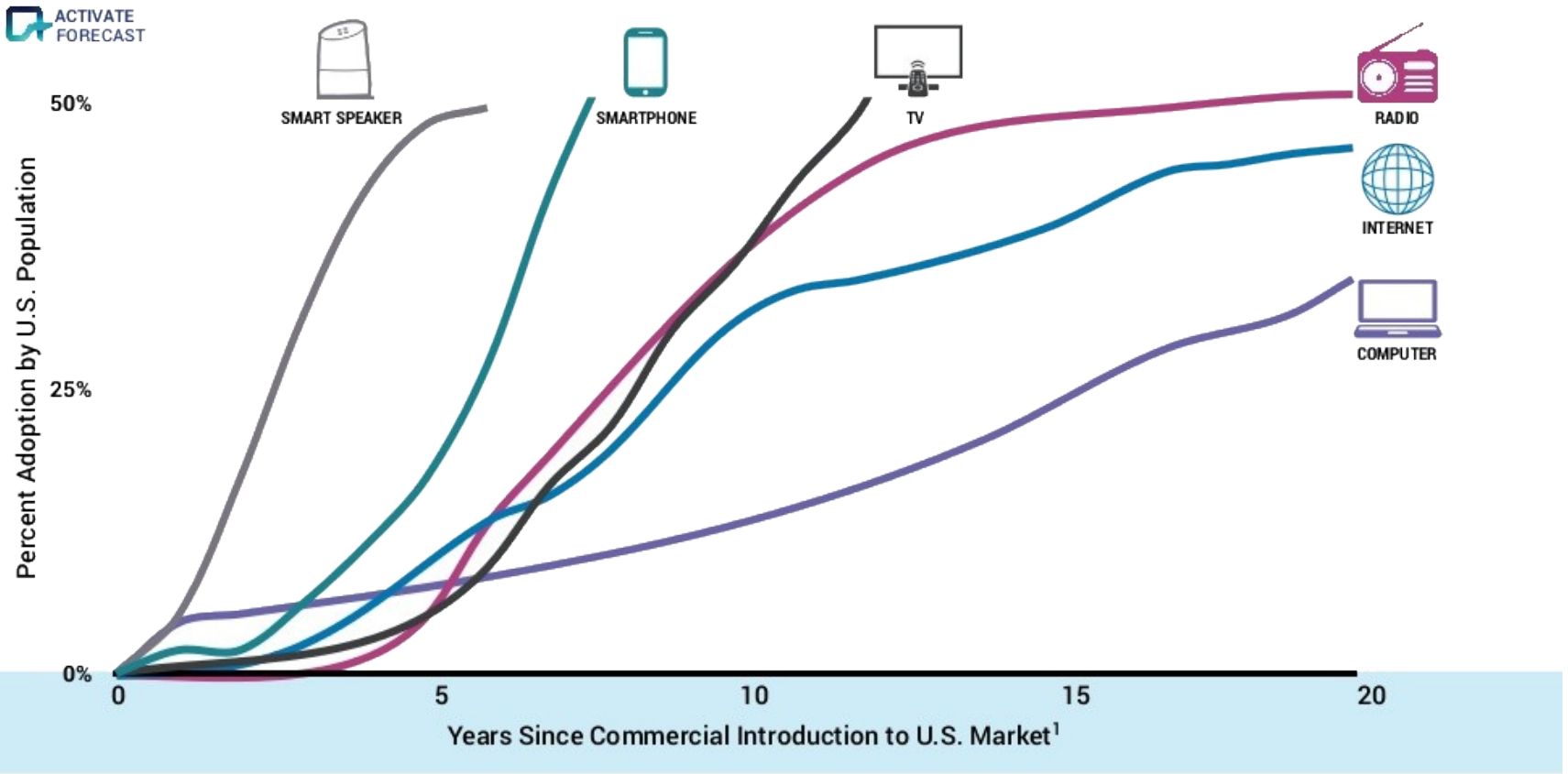
If you designed a speech assistant that you would need to work in order to flush your toilet

- Reliability
- Security

Adoption rates

We forecast that smart speaker adoption is likely to be faster than any other consumer device; however sales will not grow to the sky

SMART SPEAKER PENETRATION, U.S., YEARS FROM INCEPTION, % POPULATION



1. Radio, TV, and computer measured as share of U.S. households. Smartphone and Internet measured as share of U.S. population.
Sources: Activate analysis, U.S. Census Bureau, World Bank

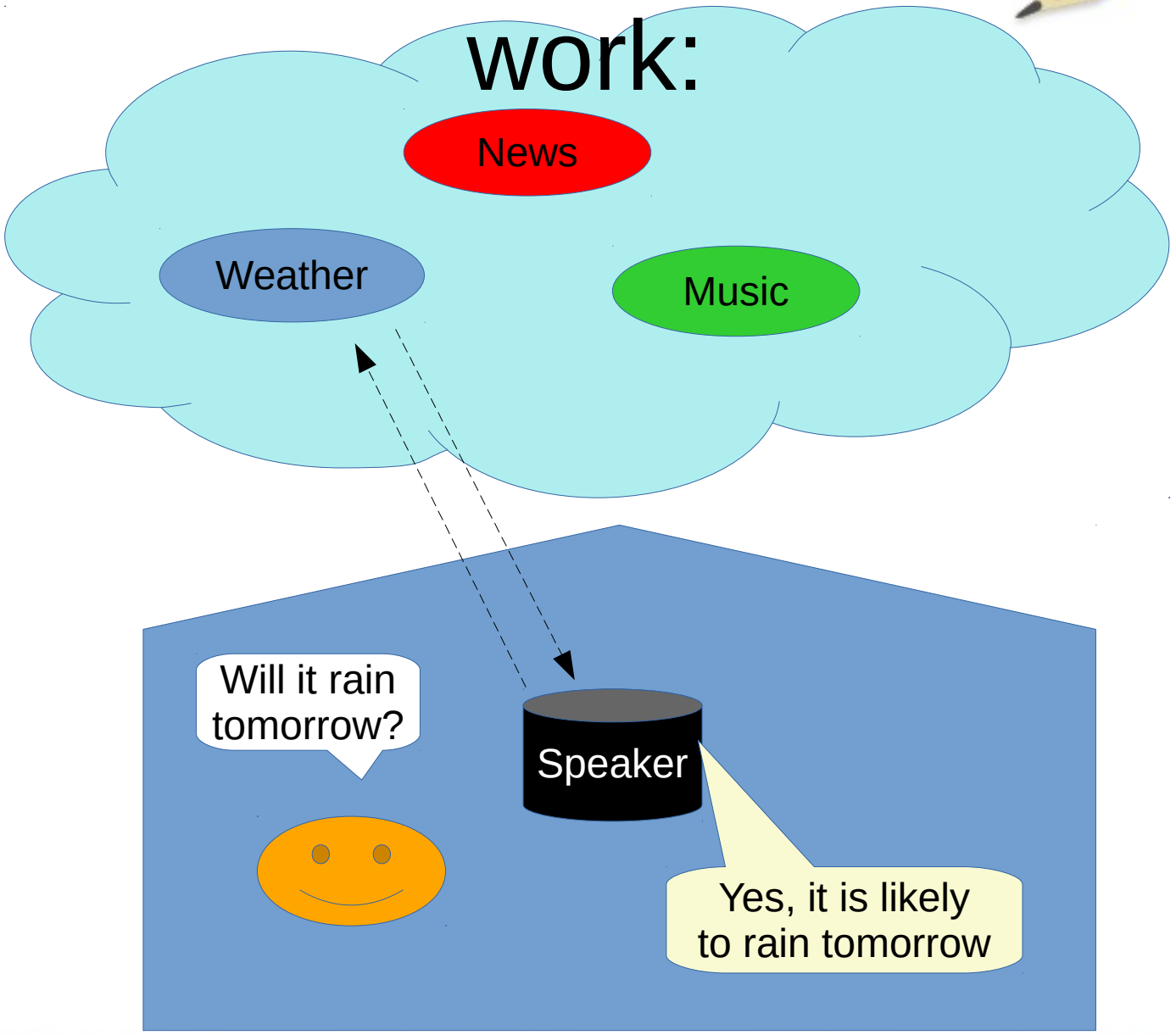
activate

How the public thinks smart speakers work

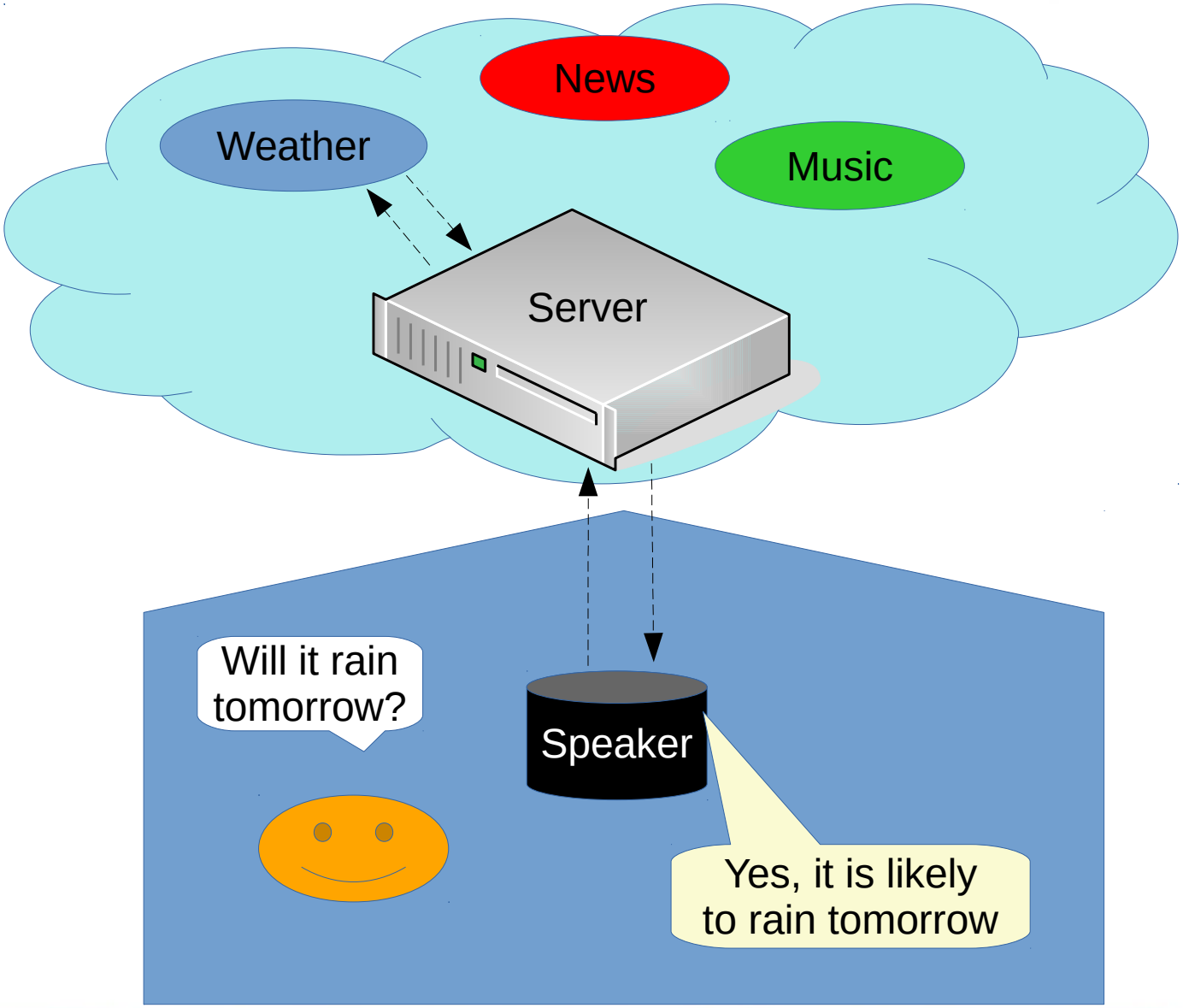
"...the speakers possess artificial intelligence, can conduct basic conversations, and are hooked up to the internet, which allows them to look stuff up and do things for you."

-SHULEVITZ, J. (2018). "Alexa, HOW WILL YOU CHANGE US?" *Atlantic*, *322*(4), 94–104.

How the smart speaker makers want you to think they work:



How they actually work



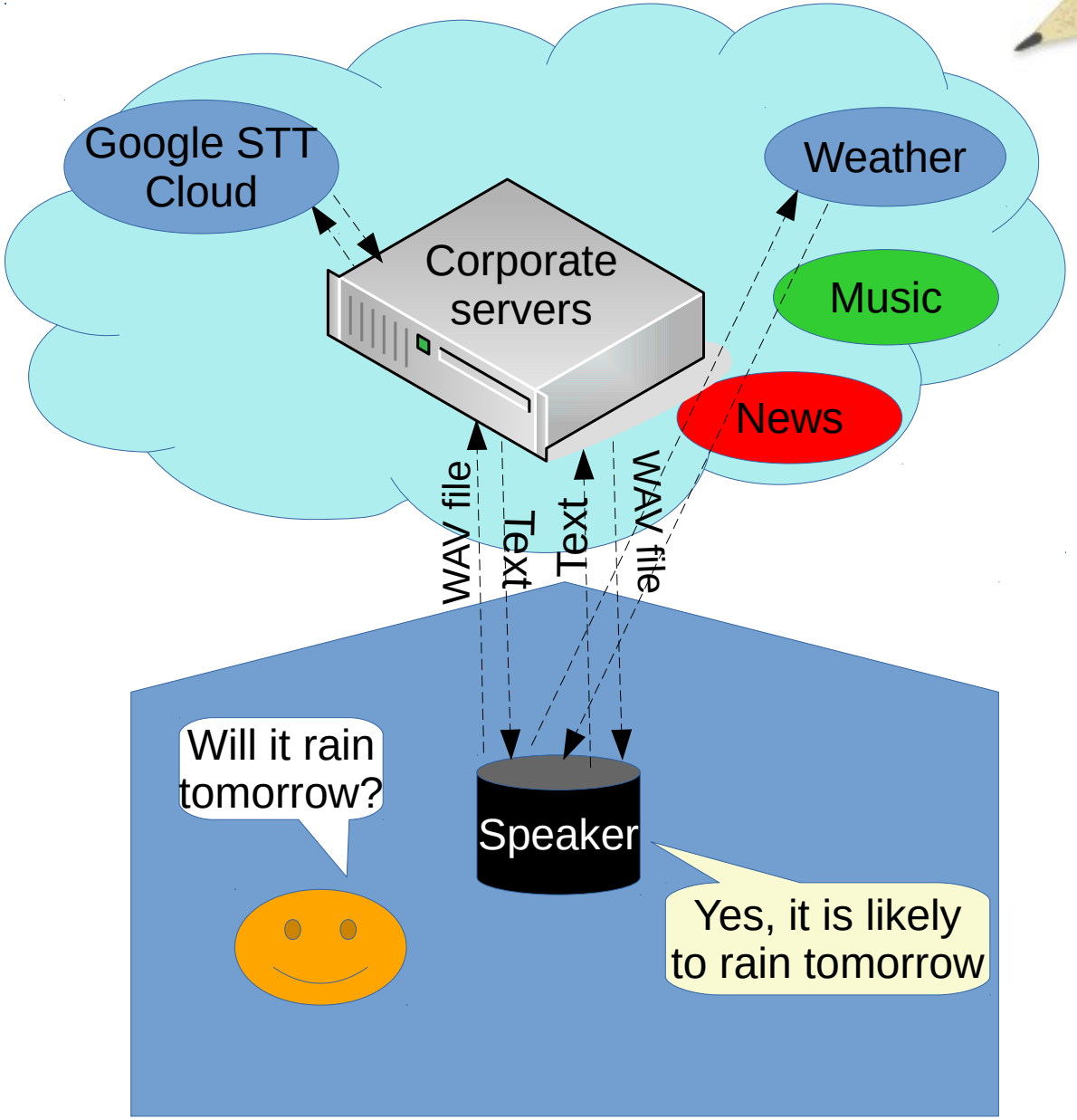
Notable Open Source Assistants

- Mycroft (https://mycroft.ai/)
- Rhasspy (https://rhasspy.readthedocs.io/)
- Almond (https://almond.stanford.edu)
- Jasper (https://jasperproject.github.io/)
- Naomi (https://projectnaomi.com)

Mycroft

- Simple to set up
- Compatible with Node Red, Hass.io, Home Assistant
- Voice recognition and speech generation happen in the cloud

How Mycroft Works



Rhasspy

- Well documented
- Works with Node-Red, Hass.io and Mozilla Home Assistant
- MIT license
- Uses pre-built models

Almond

- Available as an application for Gnome desktop or Android
- More of a chat bot
- Fairly new
- Works with Mozilla Home Assistant

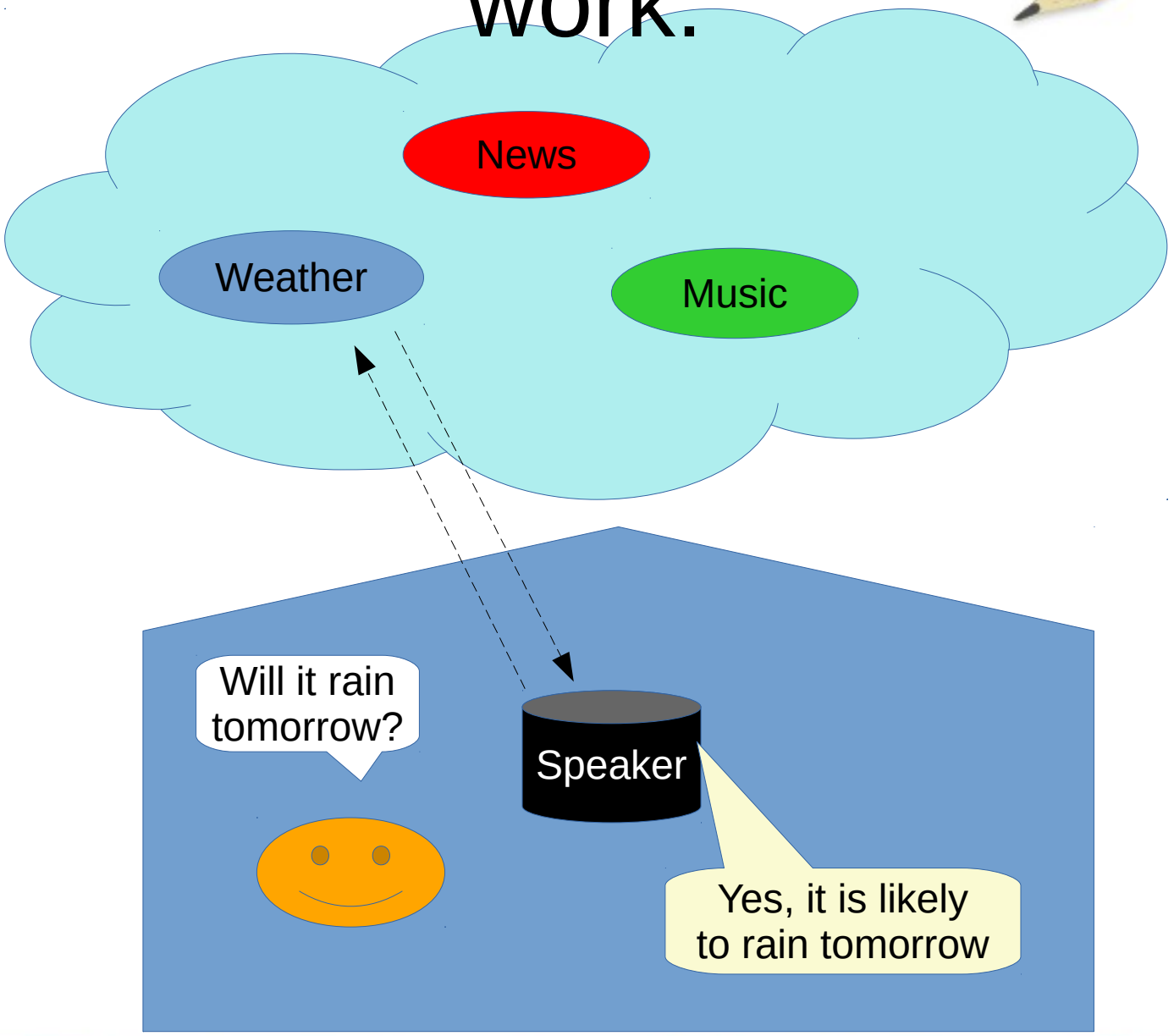
Jasper

- Plugin based
- 100% python
- Source code available and pretty simple
- Development stopped in 2017

Naomi

- Based on Jasper-dev branch
- 100% Python
- Simple
- Personal

How Naomi and Rhasspy work:



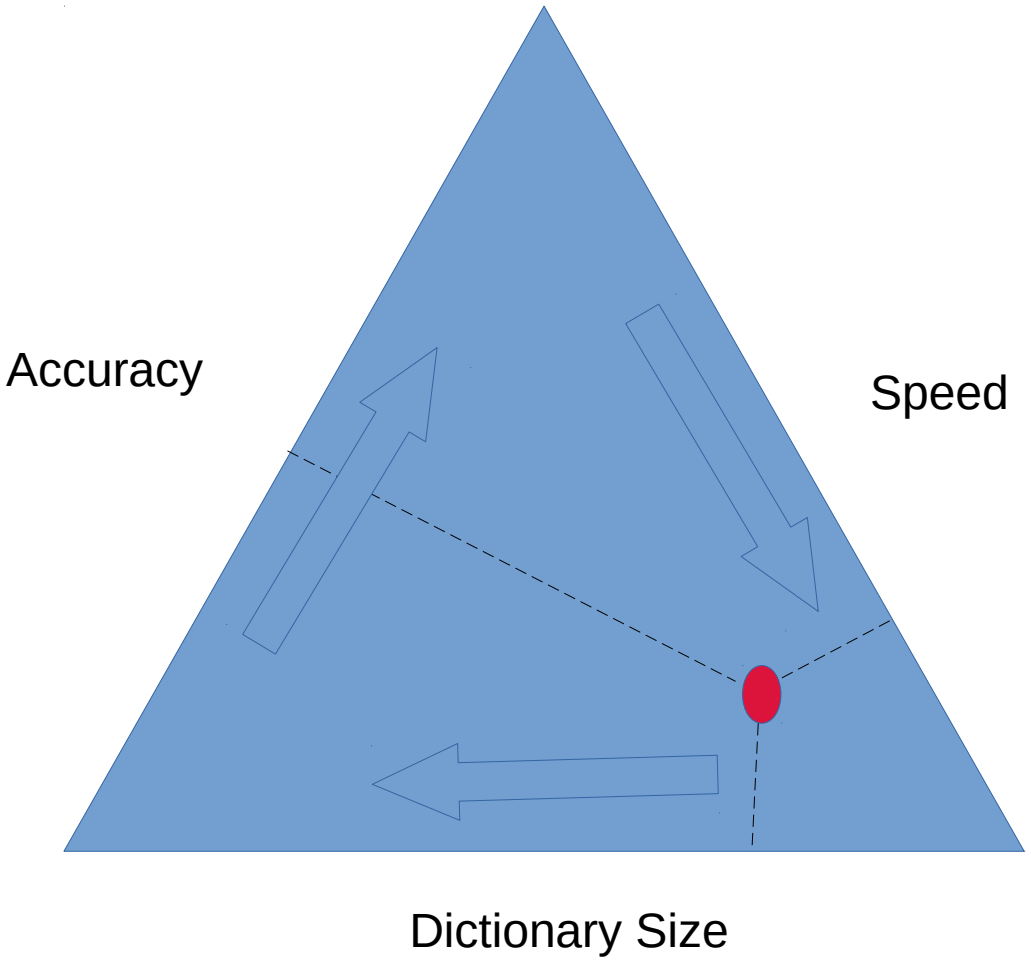
Parts of a Voice Assistant

- Speech Recognition
- Intent Parsing
- Speaking result

STT Engines

- Kaldi – Needs to be simplified for use
- PocketSphinx – Very fast, not good with large vocabularies
- Mozilla DeepSpeech – Slow, difficult to install, uses a non-standard version of tensorflow. Based on Baidu’s DeepSpeech papers
- PaddlePaddle DeepSpeech – Based on the same research paper as Mozilla DeepSpeech, but this one uses a standard PaddlePaddle engine.
- Wav2Letter++ - Brought to you by the fine folks at Facebook

STT Trade-offs



Intent Parsers

- Naomi (based on edit distances and weighted keywords)
- Adapt (keyword matching)
- Padatious (neural network)
- FuzzyWuzzy (edit distance matching)
- Fsticuffs (FST based, words must appear in the correct order to match, very fast)

Text to Speech

- Festival
- Flite
- Espeak
- Mary TTS
- Mimic

Demonstration

- Options (./Naomi.py --help)
- Voice control
- Training
- Store

Wishlist

- Testing
- Better documentation
- Simpler/automated setup process
- More tools for Speech to Text training
- Intent parsing plugins
- WebRTC interface
- Support for ESP-01/other devices
- Pulseaudio
- Help building voices
- Help user submit recordings and transcriptions to VoxForge and Common Voice
- Help forming a nonprofit LLC

References

Adobe Digital Insights (2019) State of the Voice Assistants 2019. Retrieved from https://www.slideshare.net/adobe/state-of-voice-assistants-2019

Edara, K. (2014) *U.S. Patent No. 20,140,337,131* Cupertino, CA: U.S. Patent and Trademark Office

Finley, K. (2016) Nest's Hub Shutdown Proves You're Crazy to Buy Into the Internet of Things. Wired. Retrieved from https://www.wired.com/2016/04/nests-hub-shutdown-proves-youre-crazy-buy-internet-things/

Gaida, C., Lange, P., Petrick, R., Proba, P., Malatawy, A., Suendermann-Oeft, D. (2014) Comparing Open-Source Speech Recognition Toolkits. *Siani Diagnostics*. Retrieved from http://sinaidiagnostics.com/su/pdf/oasis2014.pdf

Hern, A. (2019) Apple contractors 'regularly hear confidential details' on Siri recordings. *The Guardian*. Retrieved from https://www.theguardian.com/technology/2019/jul/26/apple-contractors-regularly-hear-confidential-details-on-siri-recordings

Kelly, M (2019) Amazon confirms it holds on to Alexa data even if you delete audio files. *The Verge*. Retrieved from https://www.theverge.com/2019/7/3/20681423/amazon-alexa-echo-chris-coons-data-transcripts-recording-privacy

Reid, K. (2020, February). Open Source Voice for Makers. *Make Magazine*, 72, 38–45. Retrieved from https://makezine.com/

Meyer, R. J. (2019) Multi-task and Transfer Learning in Low-resource Speech Recognition. Retrieved from https://jrmeyer.github.io/misc/MEYER_dissertation_2019.pdf

Rizk, Basem. (2019) Evaluation of State Of Art Open-source ASR Engines with Local Inferencing. 10.13140/RG.2.2.34901.37603. Retrieved from https://www.researchgate.net/publication/335524542_Evaluation_of_State_Of_Art_Open-source_ASR_Engines_with_Local_Inferencing

SHULEVITZ, J. (2018) "Alexa, HOW WILL YOU CHANGE US?" *The Atlantic*, 322(4), 94–104. Retrieved from https://wgu.idm.oclc.org/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=132171212&site=eds-live&scope=site

Stucke, M. E., & Ezrachi, A. (2017) How Digital Assistants Can Harm Our Economy, Privacy, and Democracy. *Berkeley Technology Law Journal*, 32(3), 1239–1299. https://doi.org/10.15779/Z383B5W79M