VoiceID on the fly:

A Speaker Recognition System that Learns from Scratch

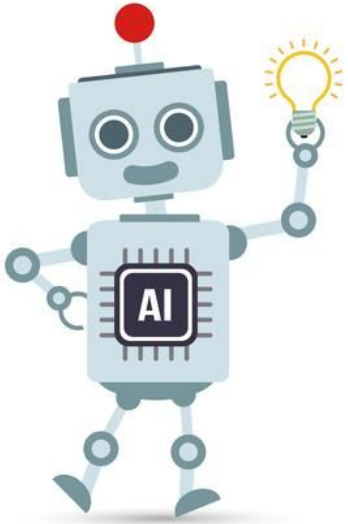
Baihan Lin, Xinxin Zhang
University of Washington, Seattle

Existing systems in speaker recognition

Training

Require tons of audio data to train

Require the users to preregister



Deployment

Old voiceprints can’t transfer to new users

No learning after deployed

Users can’t interact to correct the system

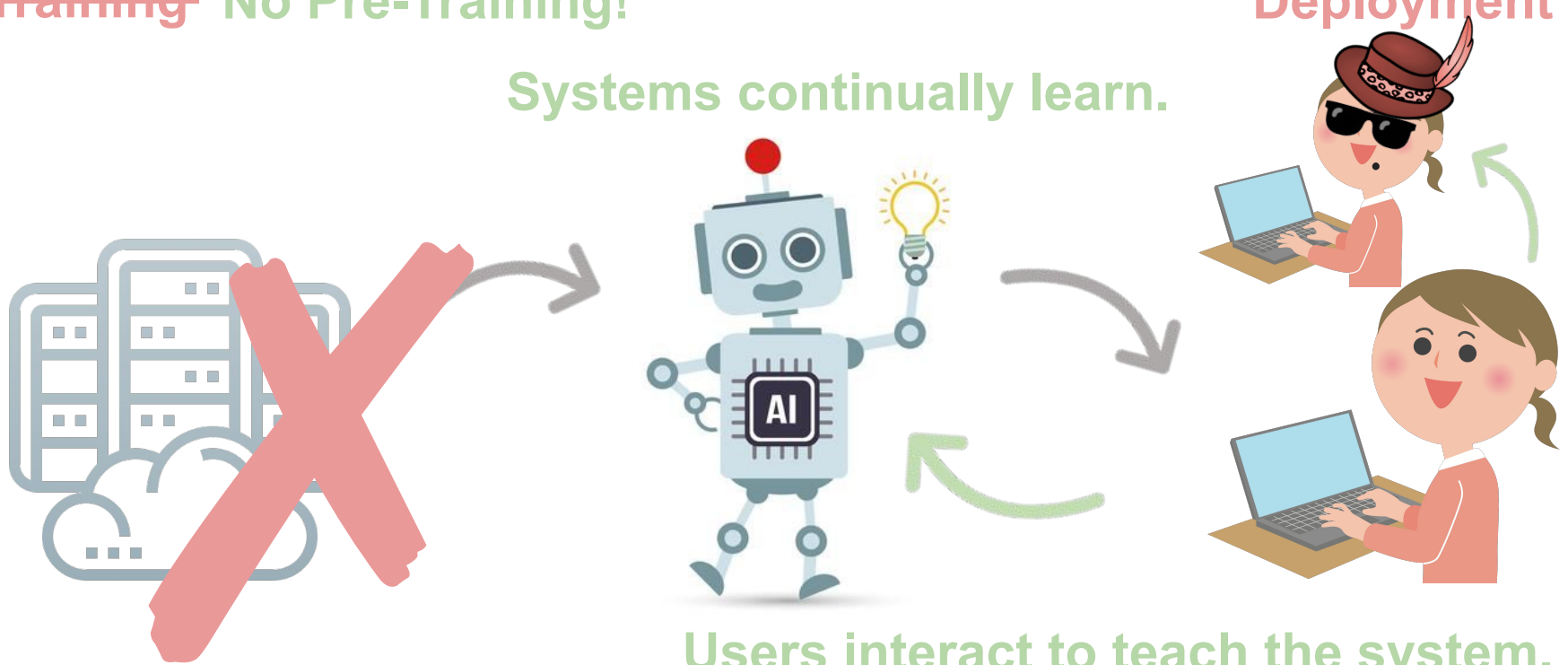
~~Existing~~ Systems in Speaker Recognition

Similar profiles transfer to new users.

~~Training~~ No Pre-Training!

Deployment

Systems continually learn.

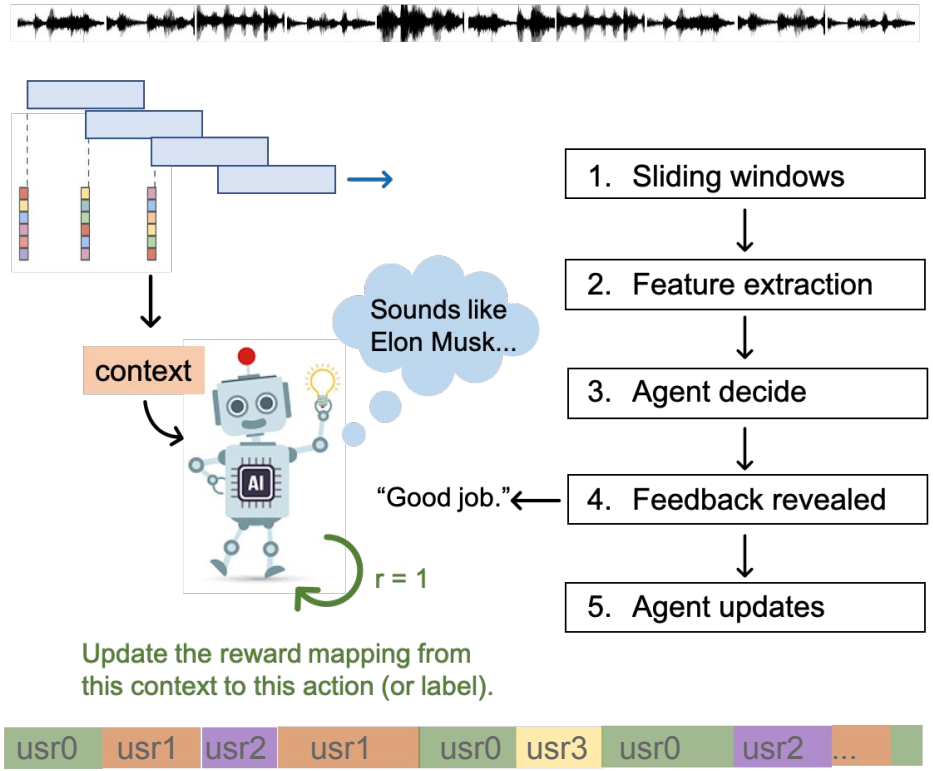


Users interact to teach the system.

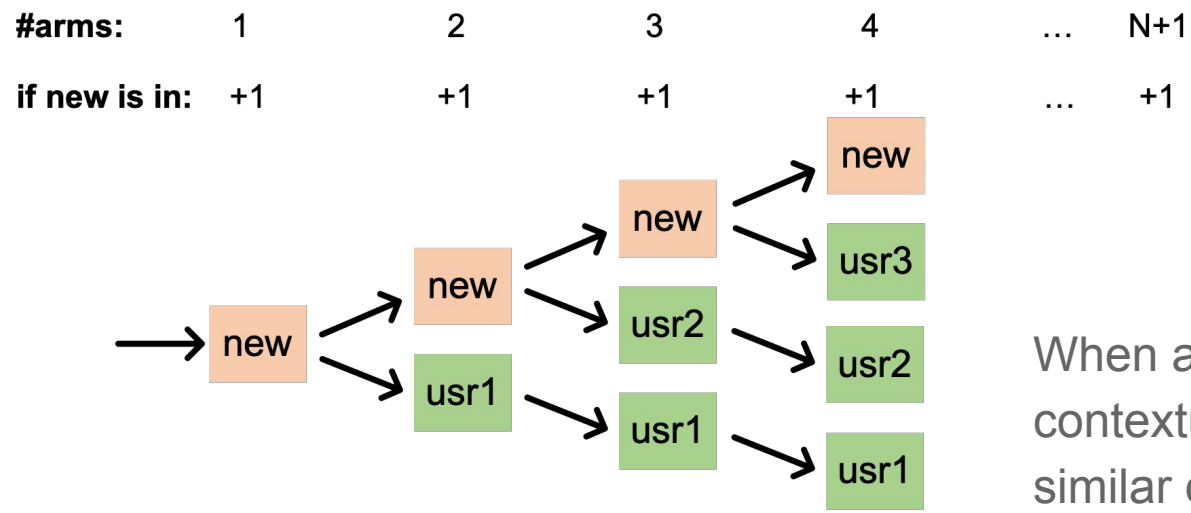
Online Learning Flowchart

Our system learns from scratch!

- ☒ user registrations
- ☒ pre-training in advance
- ☑ real-time new user registration
- ☑ transfer voiceprint information from old users to new ones



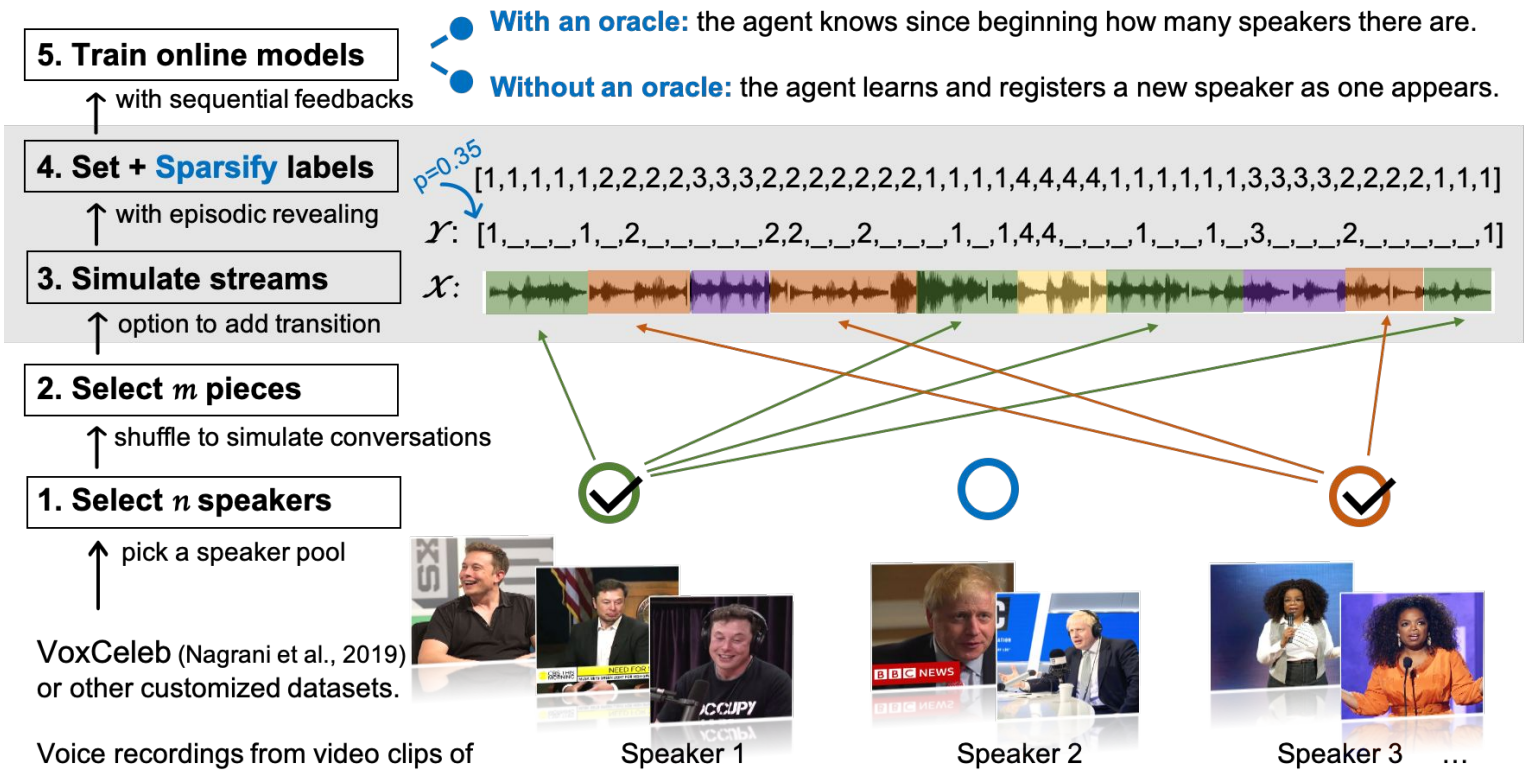
Solution: online learning with Episodic Rewards



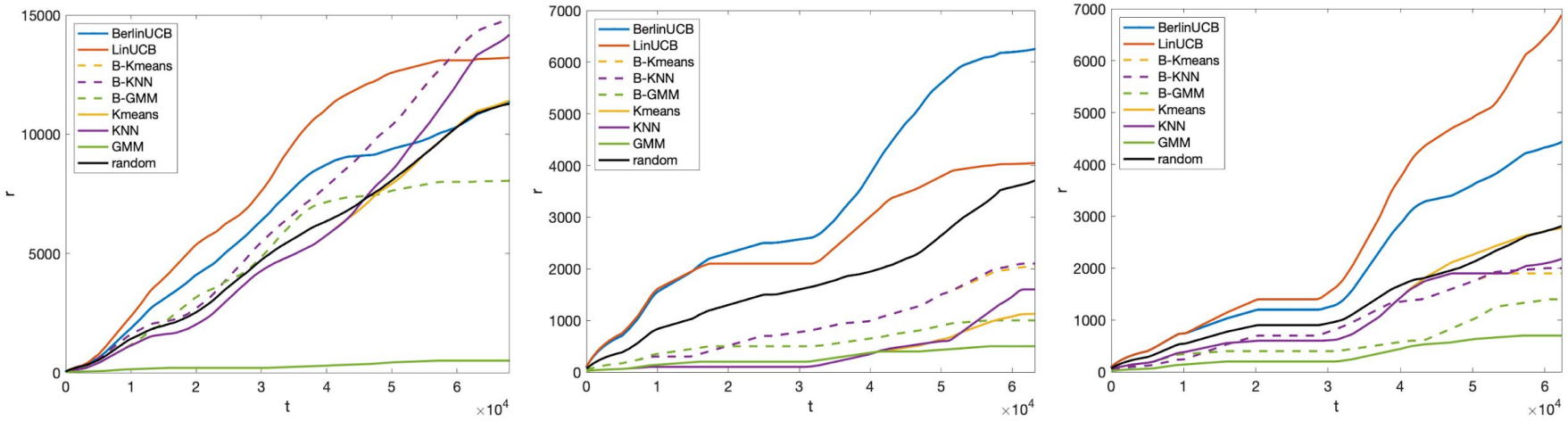
When a new user emerges, the contextual bandit copies the most similar one to the new arm.

When there is no feedback (from the user) as the label, the system creates a self-supervised label with clustering, as a pseudo bandit feedback.

Empirical evaluations - the MiniVox benchmark



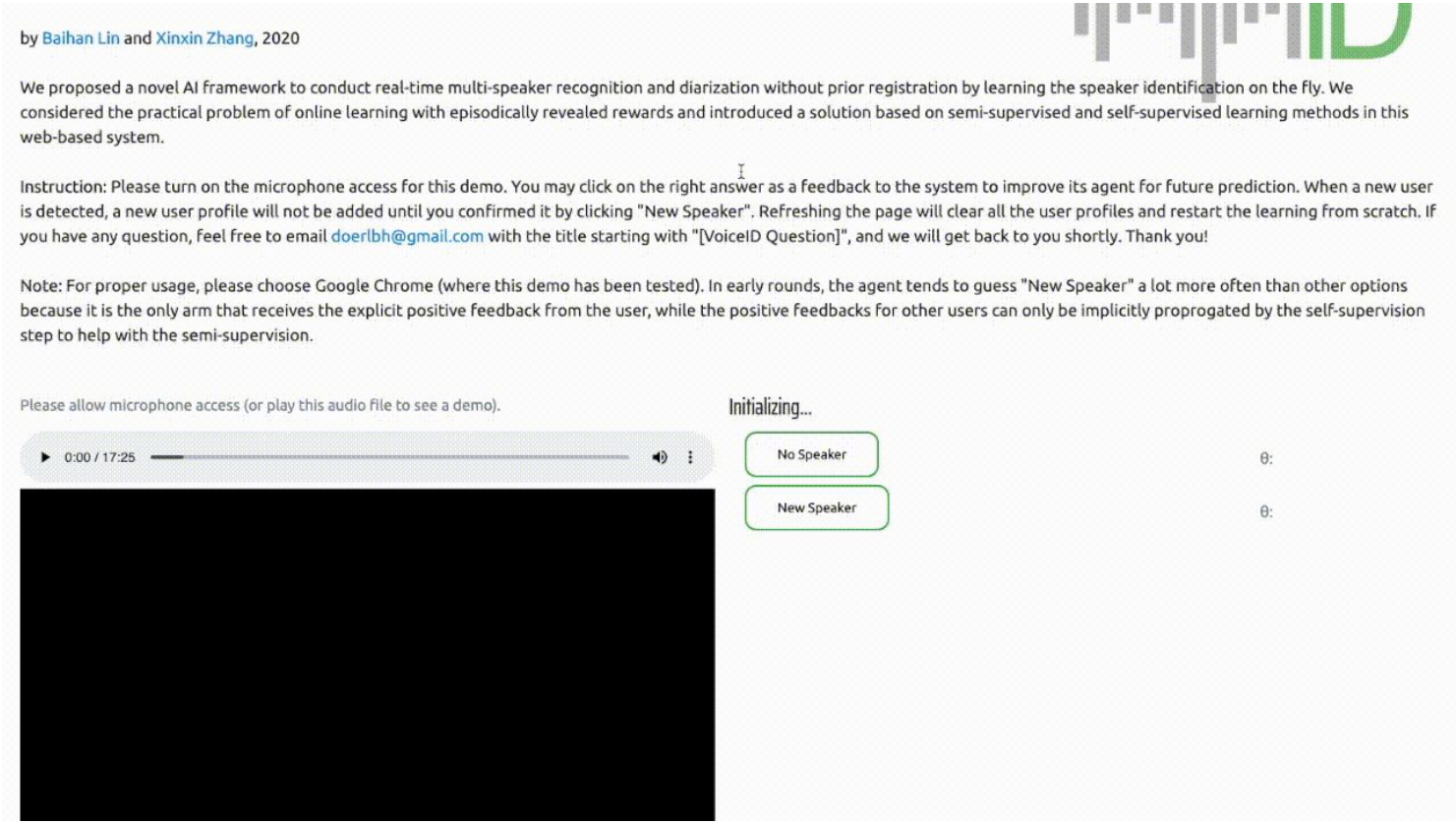
Empirical evaluations - the MiniVox benchmark



In MiniVox benchmark with MFCC as features, when the labels are only revealed 1.0% of the time (with 5, 10, and 20 speakers), our algorithm learns the fastest.

This suggests a smooth deployment of our system and a good user experience.

Demo system in action



Thank you!

If you have any question, feel free to contact:

Baihan Lin (baihan.lin@columbia.edu)
Xinxin Zhang (zhangx43@uw.edu)

System: https://www.baihan.nyc/viz/VoiceID/