How does the AI community pronounce "epoch"?
A semirigorous sociolinguistic survey

Andre Ye

University of Washington
andreye@uw.edu

Introduction

To the uninitiated, an "epoch" merely means something like "a distinctive period of history". But to the enlightened folks in the AI community, an "epoch" refers to a complete model pass through the training dataset. However, it remains unclear how to pronounce the word. The official American pronunciation is somewhere between "eh - puck" and "eh - pick", and the British is "ee - pock". But f*ck the official rules. Any linguist worth their salt knows that the distinctions between "official" and "unofficial', "correct" and "incorrect', "language" and "dialect", and so on are largely sociopolitical. I watched a YouTube video once where Noam Chompsky said something along the lines of "the only thing that differentiates a language from a dialect is a lot of guns." I would cite the timestamp but I have too much work (read: I'm too lazy), so here's the full video: `youtube.com/watch?v=hdUbIlwHRkY`. It's well worth your time. Anyways, the meaning and practice of language is held in a community: language evolves with its constitutive community. In this paper, I present a semirigorous sociolinguistic survey of how individuals in the AI community pronounce the word "epoch", with convincing empirical evidence that the AI community represents a unique linguistic subculture within the United States.

Methodology

I made a survey on Google Forms, with the following questions and possible answers:

1. *How do you pronounce "epoch"?* Possible answers: "ee - pawk", "ee - pick", "eh - pick", "eh - puck", "eh - pawk", custom answer

2. *How strongly are you attached to this pronounciation? (How unwilling are you to change your pronounciation?)* Possible answers: scale from 1 – "ehh it's whatever (no attachment)" to 5 – "ride or die, baby! (highly attached)"

3. *You are an...* Possible answers: undergrad, grad student, industry (research), industry (non-research), post-grad (academia or research institution), magical unicorn

4. *You are mainly situated in which country?* Possible answers: US, Canada, Australia, UK, custom answer.

5. *Which statement about your language background best describes you?* Possible answers: "The only language I speak and have spoken is English", "My main language is English, but I grew up speaking other languages too", "I am equally fluent in English and another language(s)", "English is my second language, but I am fluent in it", "English is my second language, and I am not quite fluent in it", custom answer

6. *Your primary area of interest/expertise is... (select all that apply)* Possible answers: Natural Language Processing, Computer Vision, Reinforcement and Interactive Learning, Robotics, Machine Learning and Statistical Theory, Social Aspects of ML, custom answer

7. *What is your favorite generic ice cream flavor?* Possible answers: vanilla, chocolate, strawberry

Initially, I shared my survey in `r/machinelearning`, which seems like a natural place to reach people in machine learning. Besides being downvoted to hell, here are some of the comments my post received:

- "It's a ridiculous question that's a waste of everyone's time."

- "Really? On this sub? With [R] tag? A survey? Is it too hard to use google?"

- "Pronunciation is not subjective, it is either correct or wrong."

It seems that `r/machinelearning` wasn't much direct help for my survey, although I did find out that the subreddit has a lot of linguistic prescriptivists. I resorted instead to posting the survey in various university research groups' and research institutions' Slack servers. I would like to thank the many individuals who passed my survey along to their own communities. This strategy proved successful, and I received 162 responses in a one week period.

Results

The most popular pronunciation is "ee - pawk" (68.9%), with a significant lead over the alternatives "eh - pawk" (14.3%), "eh - puck" (10.6%), and "eh - pick" (6.2%) (Figure 1). This is interesting because nearly all of the respondents were from the United States (93.1%), where the official pronunciation is α"eh - puck" $+ (1 - \alpha)$"eh - pick" for some $0 < \alpha < 1$; and no respondents from the UK or Australia, where the official pronunciation *is* "ee - pawk".
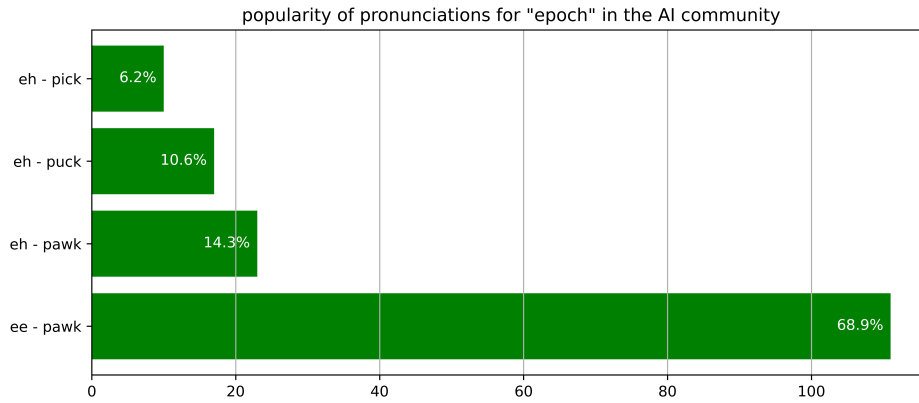
Figure 1: Number of respondents who pronounce "epoch" as "eh - pick", "eh - puck", "eh - pawk", and "ee - pawk".
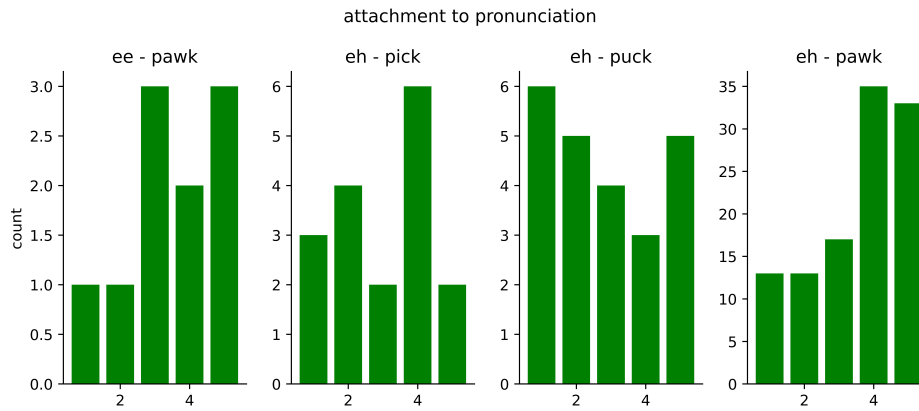

Figure 2: Attachment from 1 (least) to 5 (highest) of respondents to pronunciations (only individuals that favor pronunciation X are counted in the subchart for X).

Moreover, we find that the "ee - pawk" and "eh - pick" pronunciations both have the most dogmatic supporters (Figure 2). 29.7% of people pronouncing "eh - pawk" and 30.0% of people pronouncing "eh - pick" are maximally committed to their pronunciation. The mean attachment, from a 1 (lowest) to 5 (highest) scale, is 3.6 and 3.5, respectively. Meanwhile, "eh - pawk" and "eh - puck" have a mean attachment of 2.8 and 3.0, respectively. These results suggest that you don't want to be in the same room with both an "ee - pawk" and an "eh - pick" pronouncer.

We find that "ee-pawk" is much more popular among undergrads (75.0%), grad students (73.4%), industry researchers (77.8%), and magical unicorns (83.3%) than non-research industry workers (58.3%) and post-graduate academics (56.2%) (Figure 3). In particular, post-graduate academics disproportionately favor "eh - pawk" (21.9%) compared to other occupations. They probably think it makes them sound smarter. Interestingly, computer vision researchers disproportionately favor "ee - pawk" (86.5%), whereas NLP researchers disproportionately *disfavor* it (66.2%) (Figure 4). I was surprised to find that one of the survey respondents was studying "your mom", but regardless, if your research interest is my mom, then you are very

likely to pronounce "ee - pawk". We find that the popularity of "ee - pawk" is relatively constant across language background, but *twice as many* individuals speaking English as a second language than individuals only speaking English or as a first language favor "eh - pawk" (Figure 5a). Our results also show that individuals who like vanilla ice cream are less likely to pronounce "ee - pock" (Figure 5b). It also turns out that theory researchers disproportionately like vanilla ice cream and robotics researchers disproportionately like chocolate (Figure 6).

Conclusion

Our results show that the AI community in the United States has carved out a unique linguistic subculture in which "epoch" is pronounced "ee - pock". I have published the data at the following link: `andre-ye.github.io/data/epoch_pron_survey.csv`
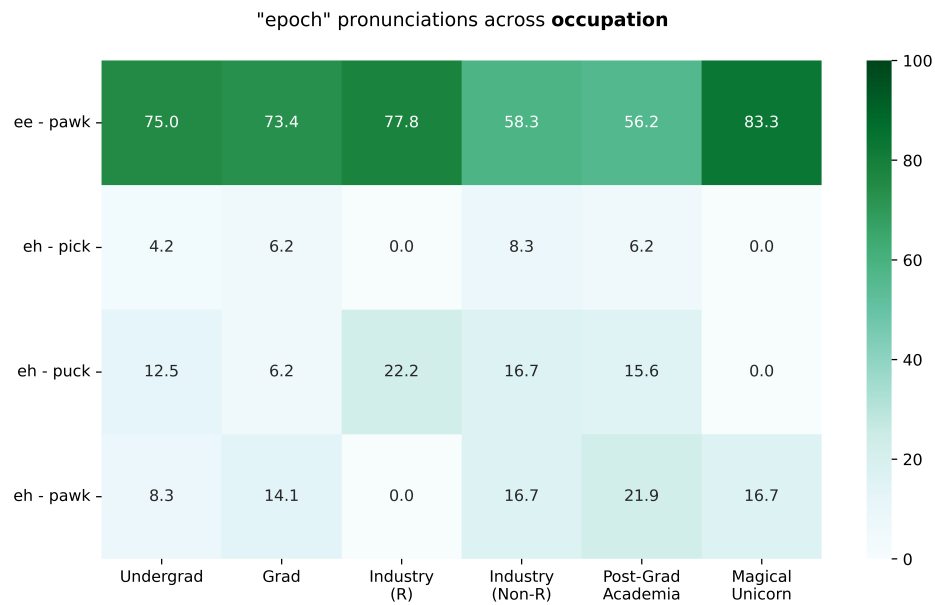
Figure 3: Pronunciation across different occupations. Percentages computed columnwise.
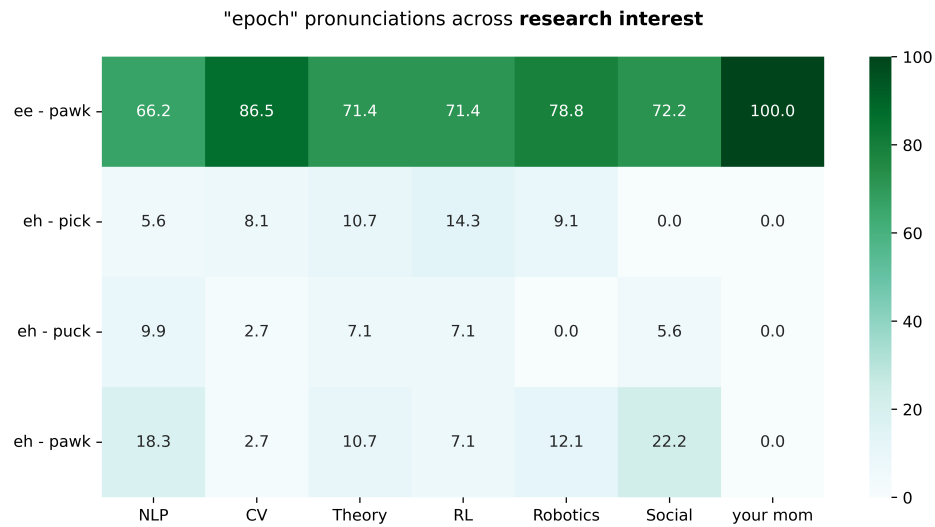

Figure 4: Pronunciation across different research interests. Individuals are represented across all of the research interests they indicated (possibly more than 1). Percentages computed columnwise.
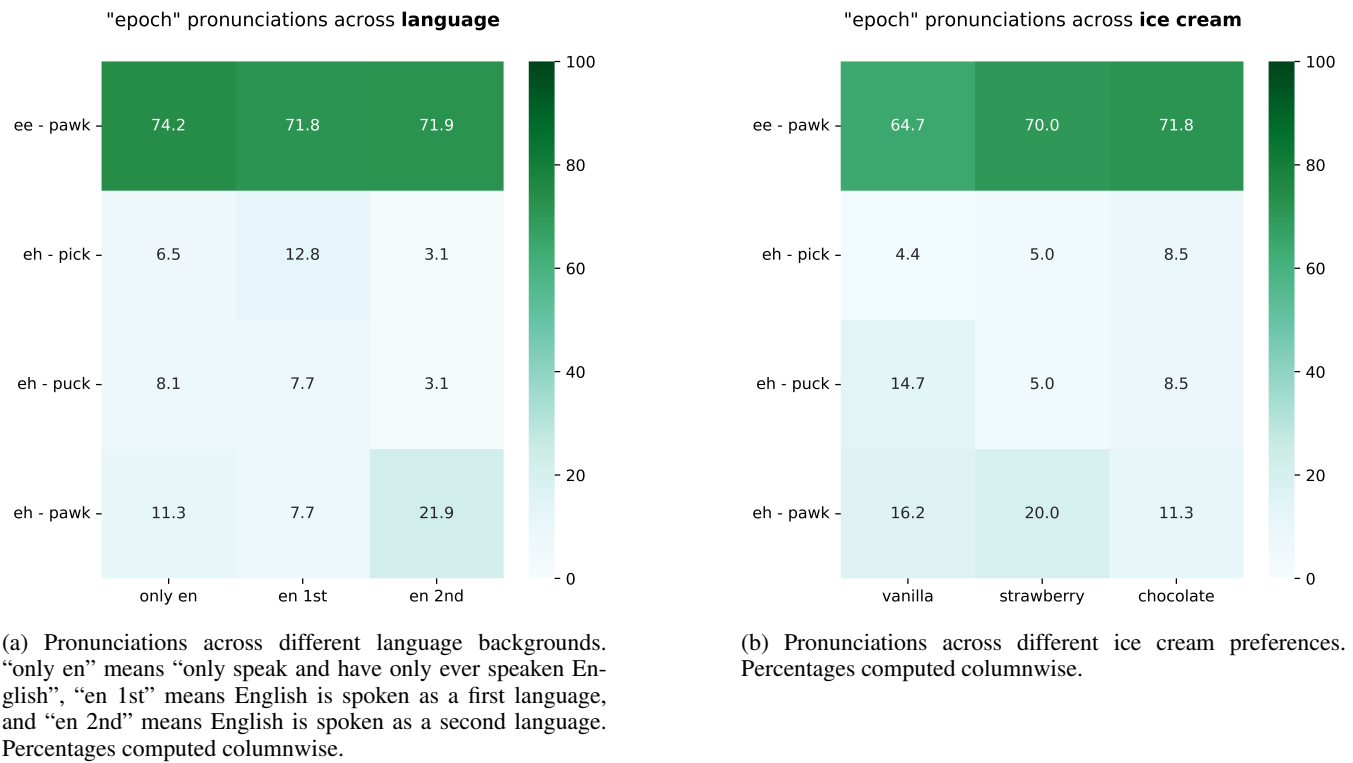
"epoch" pronunciations across **language**

"epoch" pronunciations across **ice cream**

(a) Pronunciations across different language backgrounds. "only en" means "only speak and have only ever speaken English", "en 1st" means English is spoken as a first language, and "en 2nd" means English is spoken as a second language. Percentages computed columnwise.

(b) Pronunciations across different ice cream preferences. Percentages computed columnwise.

Figure 5: I couldn't figure out how to get two independent figures side-by-side so this is a single figure with subfigures now.
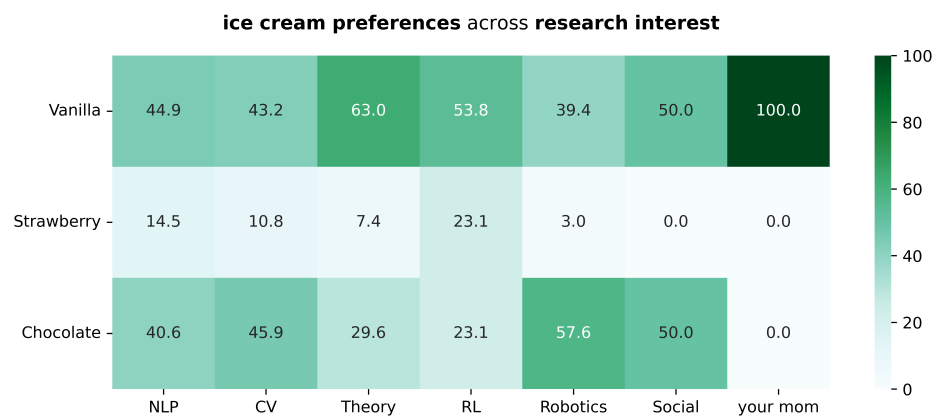

ice cream preferences across **research interest**

Figure 6: Ice cream preferences by research interest. Percentages computed columwise.