Efficient Automatic Speech Recognition (ASR)
(Technical Topic)

An Analysis on the Role of Algorithmic Models in Lending Discrimination and Solutions in Informed Model Development
(STS Topic)

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By
Aidan Hijazi-Klop

October 27, 2022

Technical Team Members:
Felix Lin, Ph. D

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

ADVISORS

Caitlin Wylie, Department of Engineering and Society

Felix Lin, Department of Computer Science

Introduction

Machine learning as a field – for a large portion of its study and development – saw little scholarly consideration for the ethical implications until the turn of the century, with it only becoming a major focus of publications come 2016. While there are many facets of ethics that could be studied with regards to machine learning, accessibility is the unifying factor between the technical capstone and socio-technical research topic in question here. Accessibility in the context of this work is to be understood as the characteristic or quality of something, an entity or system, that makes it possible to engage with, access, or otherwise use it with reasonable ease. To this end, my technical capstone focuses on user accessibility of "end-to-end (E2E) modeling for automatic speech recognition (ASR)" (Li, 2022, p. 1); that is, my research seeks to explore how speech recognition models can work on low power devices to improve general availability. Conversely, the topic of the socio-technical research paper will focus on housing accessibility – specifically access to equitable mortgage loan rates – through the lens of discriminatory lending algorithms which operate on the assumption of pristine data; in reality said data is far from unbiased and is directly the "result of [a wealth of] historically accumulated injustices... including redlining, racial covenants, exclusionary zoning, and predatory inclusion" (So et al., 2022, p. 1). Improving accessibility – the ability for individuals to equitably engage with a culture's available systems and technologies – is a concept that is, and should continue to be, persistent in conversation of social justice, especially in a western context.

Technical Topic

Historical speech or "conversational" AI models – voice assistants like Siri or Amazon Alexa, automated call centers, speech dictation systems, etc. – have utilized deep learning[1] architectures, with Recurrent Neural Networks (RNN) and Hidden Markov Models (HMM) being the leading choices. The primary concern in the use of such architectures lies in the accuracy of speech processing, the common metric to measure this being word error rate[2] (WER); this is especially true when extending use of models beyond the English language. For context regarding accuracy, "the widely cited figure [for human's WER in conversation] is 4%" (Stolcke & Droppo, 2017, p. 1). In response to this, recent development – as described in Li's work published at the well established Asia Pacific Signal and Information Processing Association's *Transactions on Signal and Information Processing Conference* – has shifted focus from "hybrid [or deep learning] models to end-to-end (E2E) models for [ASR]" with three leading approaches: "RNN transducer (RNN-T), RNN attention-based encoder-decoder (AED), and Transformer-AED" (Li, 2022, p. 1). Automatic speech recognition refers to a focus within machine learning drawing from computational linguistics that focuses on the development of the recognition and translation of spoken language into text by automatic computational processes.

---

[1] Deep learning is a type of machine learning and artificial intelligence (AI) that imitates the way humans gain certain types of knowledge, using multiple layers of processing which extract and learn from features in a dataset

[2] WER is the measure of the number of incorrect words divided by the total number of words in the speech segment

Introducing these E2E ASR pipelines – while fixing accuracy concerns – produced a new issue: lack of necessary computing power on edge devices[3], the typical hosts for speech recognition machine learning models. And this lack of necessary computational resources is what produces the aforementioned accessibility issue. If ASR models that have high enough accuracy to be considered conversational – that is achieving a sub 5% WER – and as such be deterministically useful in environments like doctor's offices or call centers, they require hardware to be run-on that is orders of magnitude more powerful than anything present in such environments. To this end the technical capstone here will focus on the improvement of end-user accessibility of E2E ASR models.

Working with Associate Professor Felix Lin, Ph.D, there are two pertinent avenues of research, more likely to be explored serially than in parallel given their nature. The first is the profiling of the existing state-of-the-art models on a variety of systems, working down to the level of an edge device (or analogous amount of computational power). Based on the research done to date, the natural approach to this seems to be to create a model agnostic benchmarking script. And from there, use a combination of server-class resources and System-on-Chip devices[4] to run said script across a chosen selection of models that, ideally, utilize different underlying architectures. Throughout this process, opportunities for architecture specific profiling may be discovered at which point discussions can be held on value in pursuing those as well. The second avenue of research is looking at real-time speech decoding. Initially the thought was to look at

[3] Edge device refers to a hardware element that controls data flow at the boundary of a network; common examples include routers, network switches, mobile phones, and broadly IoT architectures (including smart home devices).

[4] System-on-Chip (SoC) refers to a class of integrated circuits that combines all or most of the components of a computer or electronic system into a singular silicon chip

direct waveform to text (to be used in a NLP model) classification; however, brief research proved this to be out of scope for this capstone, as – per Chen et al. (2021) and similar – the only forays into this are neuro-biological in nature and based on high-density electrocorticography. However, the queries were not without fruit as ML researchers do have some established work on the viability of "streamable Transformer-Transducer (T-T) models" (Chen et al., 2021, p. 1). This refocuses the second avenue instead on exploration, through profiling or otherwise, of the implications of streamable ASR on computational performance for edge devices. In addition to the well-defined paths above, a "moonshot" goal of sorts of this research project is to explore architecture optimization of the batch or E2E ASR models. The hope of the combination of these research efforts is to layout a path forward for accessible ASR on edge devices and take the first steps in such development.

Science, Technology, and Society Topic

Pivoting to the socio-technical topic under analysis, with the lens of the relational view framework, accessibility in this context is being discussed in terms of housing – or specifically mortgage lending discrimination. The nature and extent of systemic inequalities in the United States are varied, but one topic that has been of particular note – especially following predatory housing investments by firms including, but not limited to, Pretium and Blackstone (Fields, 2022) – is housing access. Mortgage lending, while only a small element of the housing issue in the U.S., is one directly influenced by machine learning "as pricing [and lending] increasingly relies on intelligent algorithms that extract information from…real-world mortgage data" (Gillis & Spiess, 2022, p. 1). With situations as statistically significant as "Latinx and Black borrowers [paying] 4.7–4.9 basis points more in interest for GSE and FHA home-purchase loans and 1.5–1.6 basis points more for FHA and GSE refinance loans," (Bartlett et al., 2019, p. 2), the current system is intrinsically detrimental to accessibility. When risk analysis for potential lending candidates is conducted based on data that is biased by decades of public and private policies designed to reinforce racial inequalities, "the use of algorithms can also lead to inadvertent discrimination" (Bartlett et al., 2019, p. 1). And it certainly would be simple to lay blame on the financial institutions willingly perpetuating the use of these discriminatory algorithms, that would be ignoring the broader relationship these entities have with the regulatory system at large. With "federal housing laws in the United States [failing] to catch up to technology" (Rodriguez, 2020, p. 6) the fault seems to extend beyond the lenders as they are not the only parties ignorant of, or apathetic towards, the implications that go into the data being used. It only seems natural within the capitalist construction of the U.S. economy for lending insitutions to willfully ignore the ethical considerations of data used in life-changing decisions, especially if those decisions can positively the lenders' bottom line.

The question, however, of how to adjust lending algorithms without crippling a function that modern banking systems have come to rely on is a key one; an analysis of a situation such as this is only so valuable without a proposal for a path forward. And this is the question aimed to be addressed within the STS research topic established here. The approach to doing so will begin with using historical analysis and case studie, drawing upon both Bartlett et al. (2019) and Gillis & Spiess (2022) as well as additional journals to paint a clear picture of the United States's history of housing injustices and their role in modern lending data. From there, largely inspired by work in So et al. (2022), a discussion of the use of "reparative algorithms" will be pursued:

both in their use to detect bias in existing data as well as inform refinement of future lending models to be more bias-conscious. And this principle of reparative algorithms is not without some scholarly basis: recent work from MIT's Computer Science and Artificial Intelligence Laboratory produced a classifier of existing loaning algorithms which "can mitigate bias within data with multiple sensitive parameters and sensitive options while maintaining high levels of accuracy" (Singh et al., 2022, p. 12). While this research will not delve into the development of such reparative algorithms, it will extensively discuss how those models can be actualized upon be that from a public policy perspective or otherwise. One would be remiss to mention that certain authors like Lee & Floridi in their 2021 work argue that "algorithm[ic solutions] are not necessarily more discriminatory than existing[, human-based,] processes" (p. 24); such stances could potentially be used to argue against the need for algorithmic reparations. This being said, the general consensus of law, finance, and computer science scholars is that "algorithms…may reinforce existing biases in the data, system, and society and exacerbate racial inequality" (Lee & Floridi, 2021, p. 24). As such, the research question proposed above of how do we address LatinX and Black borrowers being disenfranchised by the algorithmic based elements United States's mortgage lending system remains.

Conclusion

The core assertion being brought here is that machine learning, without proper consideration for bias in training data, can not only just perpetuate systemic accessibility issues but potentially reinforce them as well. While one would be remiss to not mention the social implications of ASR accessibility and lending accessibility wildly differ in scale – improved usability of speech recognition inspecific use cases versus life altering decisions regarding housing and wealth – intentional analysis of the social conditions is pertinent to both. Properly conducted analysis of these technical and STS domains have the potential to move the needle on technical innovation from the health to smart home industry and at the very least reframe the conversation about systemic housing injustices in the United States; the hope for the latter is to provide a usable framework for future and retrospective ethical consideration of data used in machine learning solutions deployed in public and private domains.

References

Bartlett, R., Morse, A., Stanton, R., & Wallace, N. (2019). Consumer-Lending Discrimination in the FinTech Era (Working Paper No. 25943). *Journal of Financial Economics*. https://doi.org/10.3386/w25943

Chen, X., Wu, Y., Wang, Z., Liu, S., & Li, J. (2021). Developing Real-Time Streaming Transformer Transducer for Speech Recognition on Large-Scale Dataset. *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5904–5908. https://doi.org/10.1109/ICASSP39728.2021.9413535

Fields, D. (2022, January 4). Perspective | Tech and finance firms buying up homes doesn't bode well for everyone else. *Washington Post*. https://www.washingtonpost.com/outlook/2022/01/04/corporate-landlords-silicon-valley/

Gillis, T. B., & Spiess, J. L. (2022). Big Data and Discrimination. *The University of Chicago Law Review*, 31.

Lee, M. S. A., & Floridi, L. (2021). Algorithmic Fairness in Mortgage Lending: From Absolute Conditions to Relational Trade-offs. *Minds and Machines*, *31*(1), 165–191. https://doi.org/10.1007/s11023-020-09529-4

Li, J. (2022). Recent Advances in End-to-End Automatic Speech Recognition. *APSIPA Transactions on Signal and Information Processing*, *11*(1). https://doi.org/10.1561/116.00000050

Rodriguez, L. (2020). All Data Is Not Credit Data: Closing the Gap Between the Fair Housing Act and Algorithmic Decisionmaking in the Lending Industry. *Columbia Law Review*, *120*(7), 1843–1883.

Singh, A., Singh, J., Khan, A., & Gupta, A. (2022). Developing a Novel Fair-Loan Classifier through a Multi-Sensitive Debiasing Pipeline: DualFair. *Machine Learning and Knowledge Extraction*, *4*(1), 240–253. https://doi.org/10.3390/make4010011

So, W., Lohia, P., Pimplikar, R., Hosoi, A. E., & D'Ignazio, C. (2022). Beyond Fairness: Reparative Algorithms to Address Historical Injustices of Housing Discrimination in the US. *2022 ACM Conference on Fairness, Accountability, and Transparency*, 988–1004. https://doi.org/10.1145/3531146.3533160

Stolcke, A., & Droppo, J. (2017). Comparing Human and Machine Errors in Conversational Speech Transcription. *Interspeech 2017*. 141. https://doi.org/10.21437/Interspeech.2017-1544