Meals on Wheels Portal
(Technical Paper)

Language Models and Online Bot Revolution
(STS Paper)

A Thesis Prospectus Submitted to the
Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia
In Partial Fulfillment of the Requirements of the Degree
Bachelor of Science, School of Engineering

Maxwell Patek
Fall, 2019

Technical Project Team Members
Michael Benos
Alex Hicks
Kyle Leisure
Kevin Naddoni
Josh Santana
Nate Strawser

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

Introduction

Increasingly, our lives hinge on the communication that occurs online. While video media and audio media like podcasts are easy and fun to engage in, much of the important, real-time communication online takes the form of text. From news articles to Tweets to blogs to Reddit threads, online text is the front lines of online society, reacting to current events almost in real time. However, due to the fast-paced and anonymous nature of online text, the integrity of the data is in question. How do we know that what we are reading is trustworthy? Who said it? What authority do they have? Is the information accurate? Until now, these questions have been restrained to the case where we are interacting with other humans online. While there have been online bots, they have been relatively easy to identify and not convincing. However, there is new cutting-edge research being conducted on machine learning models for predicting and generating realistic human text. The research has progressed so far that in a paper by OpenAI earlier this year, experts were able to create a model so good at creating realistic human-like text that they decided to withhold the model from the public, releasing only the algorithm (Radford et al. 2019). This is unprecedented in the field of Natural Language Processing. I want to investigate what would happen when this technology is recreated by the public.

While the internet is a platform to share and interpret information, it also acts as a tool for social change and action. Meals on Wheels Charlottesville/Albemarle is a local nonprofit organization that uses the internet for action. They provide "nutritious meals to the ill, aging, and convalescing residents of the Charlottesville/Albemarle," leveraging donations and volunteers to maintain a flexible cost structure ("About Meals on Wheels of Charlottesville/Albemarle", 2019). Meals on Wheels staff use software to organize the customers, volunteers, routes, and food, and the software is hosted on the online cloud. In order to allow the organization to grow while

keeping costs low, the Meals on Wheels Charlottesville/Albemaurle organization needs a new online portal. A UVA capstone team will be working with the Meals on Wheels staff to satisfy the staff needs and to unroll the technical debt accrued by the previous two capstone teams to work on the application.

Technical Topic

Meals on Wheels is an organization with a complex set of needs and interactions. The non-profit's main activity is delivering meals to customers in the Charlottesville-Albemarle Area who do not have the means to get meals themselves, mostly due to disability or age. The organization operates with the help of volunteers who pack the meals prior to distribution and volunteers who drive delivery routes to the customers. There are also a handful of shuttle routes that the organization runs that are also driven by volunteers.

The organization is administrated by a small team of paid staff. The staff organize the volunteers, assuring that every job is filled every day. They also manage the customers, maintaining a database of current and prospective clients, what their needs are, and what route delivers to them.

The resulting complexity comes from the combination of regular and irregular volunteers and the combination of regular and irregular customer needs. The food labels and exact routes depend on what customers need what amount of food on the given day, and the volunteers executing those tasks depend on the combination of weekly, biweekly, monthly and one-time assignments that are constantly changing.

While the greater US Meals on Wheels Organization sells professional software that helps staff manage this complexity, the cost is outside of the small regional organization's

budget. Thus, the Charlottesville-Albemarle Meals on Wheels managed volunteers, customers, and routes by hand until 3 years ago, when a University of Virginia CS Capstone team created a Django web app. This web app stores what was previously stored on paper in a database, allowing computer servers to do validation, and automation of tedious manual tasks.

The previous Meals on Wheels app is functional enough for the small staff, but there are major shortcomings. First, the staff complains about loading speed. Over time, the web app has become slower and slower. This is probably due to a combination of undeleted data and inexpensive data hosting solutions. Second, the staff complained about several organizational oddities with the layout of the app, making theoretically simple tasks much more difficult than they should be. Finally, there are both unimplemented requested features, such as searching assignments based on typically assigned volunteers instead of actual assigned volunteers, and implemented unused features, such as the "whiteboard report," which can be removed.

The staff needs require an update. However, the technical debt, which is "when engineers take shortcuts that fall short of best practices," accumulated by two separate capstone teams developing separate features over a two-year period necessitates a rewrite of the application (Allman, 2012). The following is the UML diagram of the previous team's database schema.

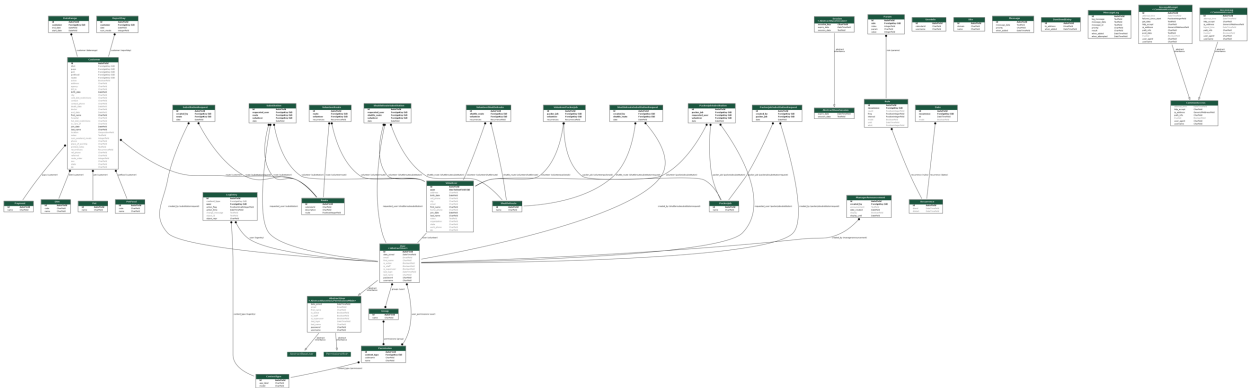

Figure 1: UML Diagram of original Portal data model.

To contrast, the following is the UML diagram for the proposed database schema for the new portal.

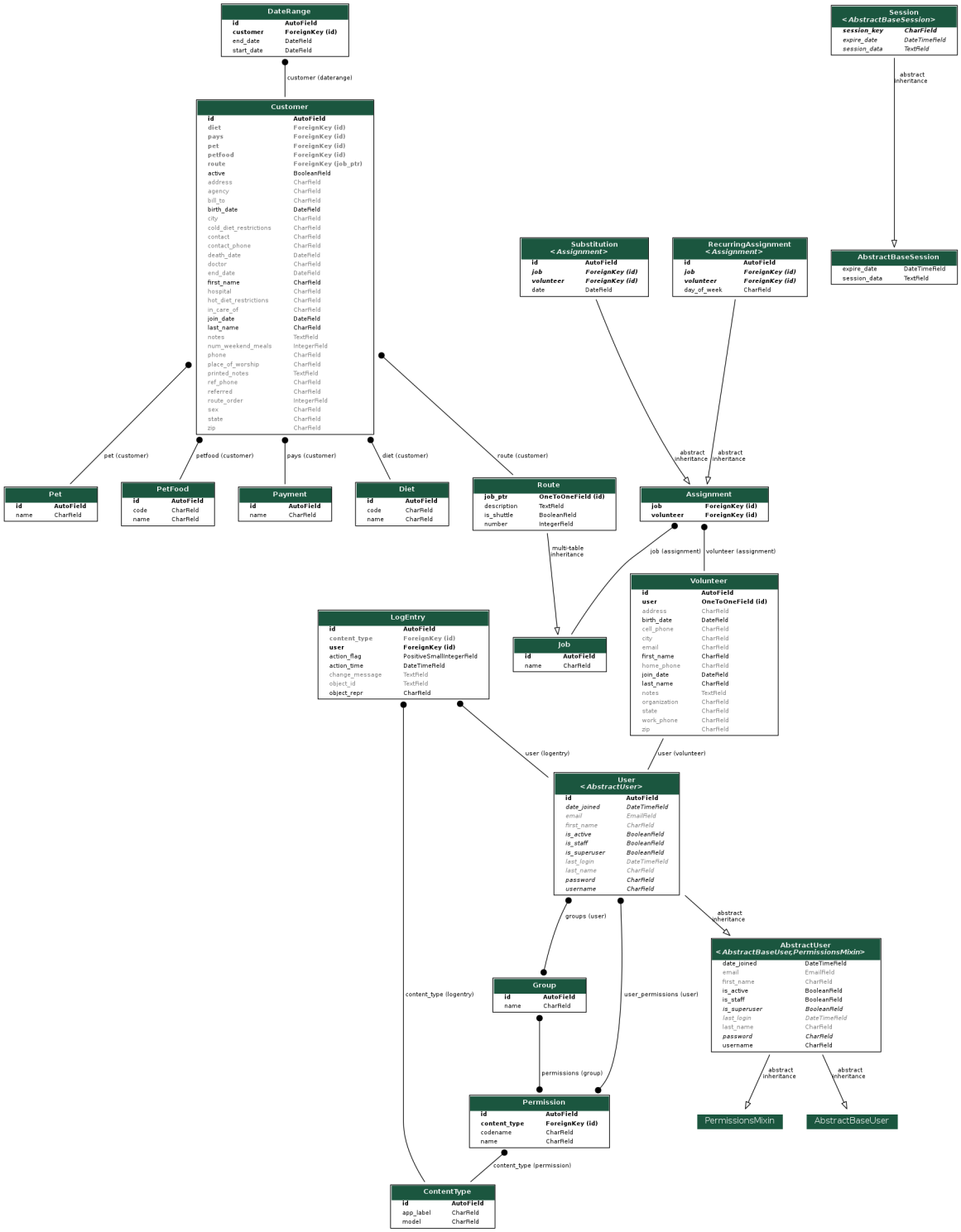

Figure 2: UML diagram of proposed replacement data model.

This application rewrite is the goal of the 2019-2020 capstone team, including myself, Michael Benos, Alex Hicks, Kyle Leisure, Kevin Naddoni, Josh Santana, and Nate Strawser. While addressing the immediate needs of the customer, the team will provide a more reasoned and maintainable application backend for long-term deployment. This includes state-of-the-art package management via Docker, normalized database models, and cost-effective cloud deployment via Amazon Web Services. By redesigning and modernizing from the ground up, the capstone project should enable to Meals on Wheels to operate at lower costs, freeing up funds to help more customers in need.

STS Topic

Machine learning has become one of the most powerful tools online. The algorithms that determine the content we see in our feeds, in our search results, and in our advertisements are dominated by state-of-the-art machine learning. This technology has allowed online platforms like Facebook, Google, and Twitter to grow and become indispensable tools in our lives. Machine learning has facilitated a shift towards online, text-based communication and information sharing.

However, this same technology can be used to distort the intended usages of these platforms. Bots have proliferated social media platforms. The technologists that create the online platforms and the adversarial users who write algorithms and bots to invade the platform have waged war since the world wide web began (Lazer, et. al. 2018). Adversaries invent some simulation of a human that can be replicated and released on the platform, attempting to enter circles of real humans and disrupt or influence them. The engineers that develop the platform then design countermeasures that are able to identify the bots or otherwise make them

transparent to users. Adversaries fight back with more and more advanced bots, and the engineers build more and more advanced detections.

Society has been increasingly dependent on online platforms. Many Americans get their news from social media. Even the president of the United States uses Twitter extensively for news delivery. In developing countries, online forums have allowed for otherwise voiceless groups of people to rally and enact change in their governments. As our dependence on the platforms increases so does our need to validate, and verify information we consume. When the actors in forums were strictly human, only the information and its acquisition was in question. However, because of the anonymous nature of the internet, the immediate source of the information is also in question. Beyond anonymity, bots introduce the possibility that the information we see is actually not from a human.

This causes many problems, such as the possibility for demographics, views, and groups to be distorted by groups of "people" that don't actually exist. These imaginary groups can push agenda, sway opinion, and drive real people to action. For example, bots were used during the 2016 presidential election to influence voters (Bessi, Ferrara, 2016). Bots can amplify fake news, which also influenced voters in the election (Allcott, Gentzkow, 2017). They also undermine the economic backbone of the platforms, marketing, advertising, ecommerce, and original online content. Bots can inflate advertising impression metrics, skewing costs and making analytics impossible. Bots can flood ecommerce sites and pages, inhibiting or preventing users from purchasing things or creating fake reviews that mislead customers (Adelani, et. al., 2019). They can also automatically steal original content, detracting from the revenue that the content creator would otherwise receive.

These problems are exasperated as bots become more advanced. Recent development in language model research has created bots that are able to, given a prompt, generate text that is so similar to human speech that it is often indistinguishable from real human speech. These bots may never be detectable online. As these models proliferate, society needs to adapt. I am going to use Paradigm Shift and Actor Network Theory to analyze how this adaptation might unfold.

American Physicist and Philosopher, Thomas Kuhn's Paradigm Shift Theory is an STS theory that describes time periods where the accepted rules and norms of a scientific discipline change (Kuhn, 1964). This may be taking place with the research scientists who design the algorithms that generate and detect bot text. We have long worked under the assumption that it is possible to distinguish between a human and an AI pretending to be human. The classical Turing Test is one measure of whether this assumption still holds. We may be nearing the end of the period when we can make that assumption. If it is impossible for humans to distinguish between bots and humans, is it reasonable to think that bots can make such a distinction? Even if they supposedly could, we may have no way of validating them. This shift will drastically change the approach that platforms must take to fighting bot activity.

While Paradigm Shift is an effective way to analyze the recent developments in language model research, there are critics of the theory. British Philosopher Martin Cohen claims that paradigm shift is inadequate for not acknowledging the greater shortcomings of scientific research in general (Cohen, 2015). He notes that scientific research is transitory anyway, and that identifying shifts is almost redundant. This is a valid criticism, but computer science research in machine learning is somewhat immune since it acknowledges the transitory truth of the effectiveness of one model over another.

Actor-Network Theory, invented by sociologists Michel Callon, Bruno Latour, and John Law, is a method of describing the interactions between human and non-human actors over time (Cressman, 2009). This theory is especially relevant because of the ability for non-human actors, bots, to be extraordinarily active. While the theory will be effective for describing the network of stakeholders, it receives criticism for being too generic and too easily incorrectly applied. Care will have to be taken to identify all actors and to avoid punctualizing important subnetworks.

Research Question and Methods

The question that the STS research paper will answer is: how can the software engineers and research scientists that support online forums adapt to the proliferation of bots that are indistinguishable from humans without infringing on free speech and privacy?

Several methods will be used to research this question. First, Documentary Research will be conducted, followed by Network Analysis, and concluded with Wicked Problem Framing. Documentary analysis will be useful because of the vast selection of published language model research papers, demonstrating use cases both in creating convincing bots and in detecting bots. Network Analysis will be used to illuminate the interactions between both the many stakeholders of online forums (human) and the technology itself (non-human, although possibly passing as human). Wicked Problem Framing will be used to search for deeper problems with internet, anonymous communication, and the philosophy of the mind.

Conclusion

The outcome of the STS research outlined above will be a comprehensive report on the potential outcomes of the proliferation of advanced language models and on the actions required

by the software engineers and research scientists that are responsible for the maintenance of online forums. The discussion will involve an analysis of the state-of-the-art in language model research, including demonstration of both beneficial and nefarious use cases, a framing of the problem as both an actor network involving the stakeholders in the online forum system and human-like non-human technological actors, and a wicked problem framing, resulting in an ultimate paradigm shift. These discussions will inform how engineers can adapt to this new technology.

While discussing the social implications of bots online, the thesis will explain our technical deliverable, a new version of the Meals on Wheels Web Portal, complete with documentation, automatic testing, and user manuals and tutorials. This new app will meet the needs of the customer while also providing a stronger base for future iteration.

References

Adelani, D. I., Mai, H., Fang, F., Nguyen, H. H., Yamagishi, J., & Echizen, I. (2019). Generating Sentiment-Preserving Fake Online Reviews Using Neural Language Models and Their Human- and Machine-based Detection. *ArXiv:1907.09177 [Cs]*. Retrieved from http://arxiv.org/abs/1907.09177

Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, *31*(2), 211–236. https://doi.org/10.1257/jep.31.2.211

Allman, E. (2012). Managing technical debt. *Communications of the ACM*, *55*(5), 50. https://doi.org/10.1145/2160718.2160733

Bellegarda, J. R. (2004). Statistical language model adaptation: Review and perspectives. *Speech Communication*, *42*(1), 93–108. https://doi.org/10.1016/j.specom.2003.08.002

Bessi, A., & Ferrara, E. (2016). *Social Bots Distort the 2016 US Presidential Election Online Discussion* (SSRN Scholarly Paper No. ID 2982233). Retrieved from Social Science Research Network website: https://papers.ssrn.com/abstract=2982233

Cambria, E., & White, B. (2014). Jumping NLP Curves: A Review of Natural Language Processing Research [Review Article]. *IEEE Computational Intelligence Magazine*, *9*(2), 48–57. https://doi.org/10.1109/MCI.2014.2307227

Caruana, R. (1997). Multitask Learning. *Machine Learning*, *28*(1), 41–75. https://doi.org/10.1023/A:1007379606734

Cohen, M. (2015). *Paradigm Shift: How Expert Opinions Keep Changing on Life, the Universe, and Everything*. Retrieved from http://ebookcentral.proquest.com/lib/uva/detail.action?docID=4393934

Cressman, D. (2009). *A Brief Overview of Actor-Network Theory: Punctualization, Heterogeneous Engineering & Translation*. Retrieved from https://summit.sfu.ca/item/13593

Kuhn, T. S. (1964). *The Structure of Scientific Revolutions*. Retrieved from https://search.lib.virginia.edu/catalog/u3307668

Lau, R. Y. K., Liao, S. Y., Kwok, R. C. W., Xu, K., Xia, Y., & Li, Y. (2011). Text mining and probabilistic language modeling for online review spam detecting. *ACM Transactions on Management Information Systems*, *2*, 1–30.

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., … Zittrain, J. L. (2018). The science of fake news. *Science*, *359*(6380), 1094–1096. https://doi.org/10.1126/science.aao2998

McCann, B., Keskar, N. S., Xiong, C., & Socher, R. (2018). The Natural Language Decathlon: Multitask Learning as Question Answering. *ArXiv:1806.08730 [Cs, Stat]*. Retrieved from http://arxiv.org/abs/1806.08730

Meals on Wheels Charlottesville. (n.d.). Retrieved October 31, 2019, from Meals on Wheels of Charlottesville / Albemarle website: https://www.cvillemeals.org/about-1

Murthy, D., Powell, A. B., Tinati, R., Anstead, N., Carr, L., Halford, S. J., & Weal, M. (2016). Automation, Algorithms, and Politics| Bots and Political Influence: A Sociotechnical Investigation of Social Network Capital. *International Journal of Communication*, *10*(0), 20.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (n.d.). *Language Models are Unsupervised Multitask Learners*. 24.

Seering, J., Flores, J. P., Savage, S., & Hammer, J. (2018). The Social Roles of Bots: Evaluating Impact of Bots on Discussions in Online Communities. *Proc. ACM Hum.-Comput. Interact.*, *2*(CSCW), 157:1–157:29. https://doi.org/10.1145/3274426

Zellers, R., Holtzman, A., Rashkin, H., Bisk, Y., Farhadi, A., Roesner, F., & Choi, Y. (2019). Defending Against Neural Fake News. *ArXiv:1905.12616 [Cs]*. Retrieved from http://arxiv.org/abs/1905.12616