Chapter 1

Introduction

Increased vaccination rates over the past 50 years have saved hundreds of millions of lives worldwide [12]. Immunization has led to the eradication of smallpox, a disease that killed more than 300 million people between 1900 and 1977 alone [10]. It has led to the eradication of two of three wild poliovirus (WPV) serotypes and the massive reduction in global cases of WPV type 1, with just 12 reported cases in 2023 [5]. Recent vaccines for hepatitis B and human papillomavirus (HPV) have provided long-term protection against chronic diseases and have proved to be incredibly effective at preventing many types of cancers, including cervical, penile, and vaginal cancers [7]. While vaccines can be vital and powerful tools towards improving public health they often depend on high coverage to remain effective. Since the COVID-19 pandemic there is emerging evidence that global immunization rates have stalled, particularly for children [11]. Further, increases in vaccine hesitancy in developed countries have led to outbreaks in vaccine preventable diseases (VPDs) [1]. The resurgance of VPDs poses a significant threat to public health, and especially threatens vulnerable populations such as children and the elderly.

The decline in vaccination coverage is driven by many factors. Some of these factors can be solved in good faith, including the inaccessibility of vaccine services [3], high costs [3], fear of needles [14], and pain associated with vaccination [8]. Others are harder to address and may potentially worsen over time, such as ingrained personal or cultural attitudes [4], trust in public health institutions [13], exposure to misinformation [2], or simply lack of

knowledge. Given broad modern adoption of digital communication, and recent advances in natural language processing (NLP), this latter set of factors is suitable for computational solutions. These issues can be broadly characterized as *information factors* negatively affecting vaccine acceptance. When discussing such information factors and vaccines, there is a strong tendency to focus on misinformation. This is for good reason. The World Economic Forum's *2024 Global Risks Report* ranked mis- and disinformation as the most severe near-term global risk, rated specifically as the most probable cause of material crisis on a global scale [16]. Strong anti-vaccination (antivax) views grounded in misinformation have become a point of discussion within even the top ranks of US politics [6]. Loomba et al. [9] found that even for people who previously said they would "definitely accept" the COVID-19 vaccine, misinformation induced at least a 6% decline in intent to vaccinate. There has been an observed rise in vaccine misinformation across social media [2]. However, there are many information factors beyond misinformation which impact vaccine confidence. For example, Vasudevan et al. [15] found that maternal perceptions of whether their newborn babies were small at birth is negatively associated with timely vaccination in rural Bangladesh. While it's hard to track if these mothers had been exposed to any direct mis- or dis-information regarding infant size and vaccine safety, it's reasonable to assume that a lack of information entirely might produce such behaviors just as well. This could be due to faulty reasoning, a non-understanding of advanced biology and vaccine science, or simple emotions such as fear of potential harm to the child.

In this dissertation, I describe several key research advances that form a new technological foundation for the public health field on this problem. These research advances allow for the accelerated improvement of our understanding of information factors for vaccine acceptance and new tools to develop interventions. In particular, I highlight language models as an emerging technology platform with massive potential for use in interventions of vaccine concerns. These vaccine concerns, formally introduced and defined in **??**, represent articulable reasons for vaccine hesitancy, and allows this research a broader scope than just concerns raised directly by misinformation.

1.1 Challenges

Addressing information factors negatively affecting vaccine acceptance is unfortunately not a straight-forward problem. It has been my experience in this space that application of techniques from one discipline quickly surfaces substantial limitations which require diving into a different discipline entirely. For example, in one of our earliest projects on this topic, we attempted to build an exploration algorithm for social media communities in order to discover comments that might be causal of new vaccine concerns. The attempt was to apply principles from reinforcement learning (RL) to web data, but we quickly discovered that the perception of a social media comment (our reward function) is highly subjective; prompting us to switch focus towards the prediction of human opinion distributions, which required a deeper dive of techniques from NLP. The most difficult challenges in this area form two broad categories: (1) Data within this field is scarcely available and difficult to use. This creates challenges for language model performance and even greater challenges for evaluation. (2) There are very high standards in the public health domain to have a chance at meaningful system adoption. These standards warrant greater focus on under-explored topics of research within NLP. In Sections 1.1.1 and 1.1.2, I explain why challenges arise in each of these two categories and highlight why these obstacles make progress in this space particularly challenging for NLP-based solutions.

1.1.1 Data quality and availability

Unfortunately, there is not much when it comes to high quality datasets for real-life discourse on vaccination. This lack of data availability is for several reasons. Social media platforms have in the past five years began to significantly restrict the access to their data. For example, while Facebook used to provide academic researchers with application program interface (API) access to download data through a platform called CrowdTangle, this feature was ultimately terminated in August of 2024 . This follows a wider trend of making data less available and more expensive as demand for it has been driven up

by corporate interest in large language model training . This problem is retroactive, since many companies had policies of forbidding peer-to-peer data sharing except through unique identifiers, which require researchers to download the actual text content of a Tweet or Facebook post through the API themselves . When these APIs are made unavailable, datasets released in the past become unusable.

An important aspect for ML models to achieve generalization is to provide independently sampled distribution of training data . Getting an independent sample of anti-vaccination content is very difficult. Out of all content on social media, the actual relevant content regarding vaccines is only a tiny fraction of all social media posts . If you sample within communities such as active antivax groups, the types of concerns you will find may not generalize to a broader population and their concerns. The data is also polluted by artificial intelligence (AI) generated content. Therefore, one must be able to ingest huge amounts of data, and run high performance, unbiased relevance classification to filter out the samples of interest. Doing so without accidentally biasing relevance classifiers with any potentially unknown spurious correlation is a difficult technical challenge. The independence problem is further exacerbated by the fact that anti-vaccination content is sometimes in violation of the terms of agreement for social media platforms. Therefore, this content is often subject to removal by the site , biasing samples even further and potentially removing the most strongly held concerns from being available. This means that studying vaccine concerns on social media is complicated by having to fill in an ambiguous hole of missing content based on a proprietary filtering algorithm we simply don't have access to.

Labeling the data is not only hard (which it is), but also potentially sensitive. Social media data contains potentially private information, and must therefore be carefully handled to avoid leaks. Given that the topic of vaccines can be particularly polarizing, even public social media posts need to be properly de-identified to ensure any attention placed on them don't lead to unfair backlash for the author. Even the annotators which read this data could be exposed to large amounts of misinformation by labeling this data, requiring additional risk mitigations on their behalf. The last thing we would want is to

induce misinformed vaccine hesitancy in those involved with our project.

Solving all the above challenges, the resulting data in this domain is often far from what would be considered clean or high quality language text. Social media data is full of informal communication, relatively newly invented slang, obscure references, sarcasm, abbreviations, evolving use and misuse of emojis, and insincere (troll) comments that are meant to provoke reactions in others. When discussing topics that are known to get censored, some users employ sophisticated adversarial attack methods to evade text-based content filtering algorithms. Dealing with this data presents additional challenges to any language modeling algorithms used and may significantly impact performance.

Lastly, data in the domain of vaccine concerns is massively subjective. What seems credible to one person may not seem credible to another. What raises a strong concern regarding immunization for you may not cause much concern for me. Concerns regarding vaccines are a personal and evolving target. Each person may change their mind regarding vaccine throughout the years, and in fact they may even change their mind regarding a single text throughout the week! There are constant shifts in interactions with vaccine information. These shifts and subjective differences are not, however, noise. They are important signal that should be captured and understood in order to fully understand attitudes regarding vaccination. However, separating this signal from noise within the already difficult-to-access, low-quality, sensitive data is a very challenging task.

1.1.2 High standards for system adoption

In public health, the stakes of mistakes is high. Poor model performance is not just a dollar lost by a company, or some seconds wasted waiting for a search result to load. In public health errors may result in real harm to human health. It's therefore very understandable that any intervention must meet exceedingly high standards and rigorous evaluation to be adopted and trusted. There are several critical factors that make this challenging.

Most fundamentally, language model-based interventions must be easily understood and trusted by both healthcare professionals and the general public to gain traction. This

requires outputs and model decision's be grounded in understandable and widely agreed upon language and definitions. However, as covered in Section 1.1.1, vaccine concerns are highly personal and subjective. Technological solutions in this space therefore have to do both things at once: they must tailor to subjective differences and individual language, but be easily explainable to anyone through the use of a shared language and agreed upon definitions. From the perspective of NLP, this conflict of objectives is possibly one of the most challenging tasks within the domain.

The evaluation of eventual interventions will be very closely scrutinized. There must be rigorous evidence of impact and safety regarding tools for this space, something that is particularly hard given how nascent language models are. Misclassifications or the accidental spread of misinformation through language model outputs could exacerbate vaccine hesitancy instead of mitigating it. Language model outputs must be verified by qualified professionals as factual and unlikely to cause harm before being acceptable for use in human-subject experimentation. Even for well-defined task, such as the classification of vaccine attitudes in text, it is not just about the overall accuracy or performance of the models. There are many things about the use of language models that we do not yet understand at a level required for adoption of this technology in public health interventions. In the past two years, the use of LLMs as automatic evaluation metrics has increased geometrically , but there is little understanding of the confidence we can have in these judgements. Likewise, LLMs are used to replace human judgements in several tasks, but the potential biases and inconsistencies these models may exhibit are understudied. This point is especially important when viewed through the perspective of trust. Trust is crucial in public health, and remains an ongoing challenge in the use of LLMs, which tend to operate in ways that are not easily explainable to non-technical audiences.

Lastly, due to the massive scale of potential interventions, cost is a consideration that cannot be overlooked. Monitoring systems may need to work over terabytes of data every month . Today, state-of-the-art (SOTA) LLM APIs are prohibitively expensive and too slow to act as possible solutions for this type of scale. However, given the high standards

discussed in this section, performance cannot be easily traded for cost.

Bibliography

[1] Kerri-Ann M. Anderson and Nicole Creanza. "Internal and external factors affecting vaccination coverage: Modeling the interactions between vaccine hesitancy, accessibility, and mandates". In: *PLOS Global Public Health* 3.10 (Oct. 2023). Ed. by Julia Robinson, e0001186. ISSN: 2767-3375. DOI: 10.1371/journal.pgph.0001186. URL: http://dx.doi.org/10.1371/journal.pgph.0001186 (cit. on p. 1).

[2] Erika Bonnevie et al. "Quantifying the rise of vaccine opposition on Twitter during the COVID-19 pandemic". In: *Journal of Communication in Healthcare* 14.1 (Jan. 2021). Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/17538068.2020.1858222, pp. 12–19. ISSN: 1753-8068. DOI: 10.1080/17538068.2020.1858222. URL: https://doi.org/10.1080/17538068.2020.1858222 (visited on 07/27/2022) (cit. on pp. 1, 2).

[3] Peter Briss, Abigail Shefer, and Lance Rodewald. "Improving vaccine coverage in communities and healthcare systems: No magic bullets". In: *American Journal of Preventive Medicine* 23.1 (2002), pp. 70–71 (cit. on p. 1).

[4] Eve Dubé et al. "Vaccine hesitancy: An overview". In: *Human Vaccines amp; Immunotherapeutics* 9.8 (Aug. 2013), pp. 1763–1773. ISSN: 2164-554X. DOI: 10.4161/hv.24657. URL: http://dx.doi.org/10.4161/hv.24657 (cit. on p. 1).

[5] Keri Geiger et al. "Progress Toward Poliomyelitis Eradication — Worldwide, January 2022–December 2023". In: *MMWR. Morbidity and Mortality Weekly Report* 73.19 (May 2024), pp. 441–446. ISSN: 1545-861X. DOI: 10.15585/mmwr.mm7319a4. URL: http://dx.doi.org/10.15585/mmwr.mm7319a4 (cit. on p. 1).

[6] Anjali Huynh. *5 Noteworthy Falsehoods Robert F. Kennedy Jr. Has Promoted*. July 2023. URL: https://www.nytimes.com/2023/07/06/us/politics/rfk-conspiracy-theories-fact-check.html (cit. on p. 2).

[7] National Cancer Institute. *Human Papillomavirus (HPV) Vaccines*. [Accessed 12-10-2024]. May 2021. URL: https://www.cancer.gov/about-cancer/causes-prevention/risk/infectious-agents/hpv-vaccine-fact-sheet (cit. on p. 1).

[8] Allison Kennedy, Michelle Basket, and Kristine Sheedy. "Vaccine Attitudes, Concerns, and Information Sources Reported by Parents of Young Children: Results

From the 2009 HealthStyles Survey". In: *Pediatrics* 127.Supplement$_1$ (May 2011), S92–S99. ISSN: 1098-4275. DOI: 10.1542/peds.2010-1722n. URL: http://dx.doi.org/10.1542/peds.2010-1722N (cit. on p. 1).

[9] Sahil Loomba et al. "Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA". en. In: *Nature Human Behaviour* 5.3 (Mar. 2021). Number: 3 Publisher: Nature Publishing Group, pp. 337–348. ISSN: 2397-3374. DOI: 10.1038/s41562-021-01056-1. URL: https://www.nature.com/articles/s41562-021-01056-1 (visited on 06/29/2022) (cit. on p. 2).

[10] American Museum of Natural History. *Smallpox*. URL: https://www.amnh.org/explore/science-topics/disease-eradication/countdown-to-zero/smallpox (cit. on p. 1).

[11] World Health Organization. *Global childhood immunization levels stalled in 2023, leaving many without life-saving protection*. July 2024. URL: https://www.who.int/news/item/15-07-2024-global-childhood-immunization-levels-stalled-in-2023-leaving-many-without-life-saving-protection (cit. on p. 1).

[12] World Health Organization. *Global immunization efforts have saved at least 154 million lives over the past 50 years*. [Accessed 12-10-2024]. Apr. 2024. URL: https://www.who.int/news/item/24-04-2024-global-immunization-efforts-have-saved-at-least-154-million-lives-over-the-past-50-years (cit. on p. 1).

[13] Pieter Streefland, A.M.R Chowdhury, and Pilar Ramos-Jimenez. "Patterns of vaccination acceptance". In: *Social Science amp; Medicine* 49.12 (Dec. 1999), pp. 1705–1716. ISSN: 0277-9536. DOI: 10.1016/s0277-9536(99)00239-7. URL: http://dx.doi.org/10.1016/S0277-9536(99)00239-7 (cit. on p. 1).

[14] Anna Taddio et al. "Survey of the prevalence of immunization non-compliance due to needle fears in children and adults". In: *Vaccine* 30.32 (July 2012), pp. 4807–4812. ISSN: 0264-410X. DOI: 10.1016/j.vaccine.2012.05.011. URL: http://dx.doi.org/10.1016/j.vaccine.2012.05.011 (cit. on p. 1).

[15] Lavanya Vasudevan et al. "Maternal determinants of timely vaccination coverage among infants in rural Bangladesh". In: *Vaccine* 32.42 (2014), pp. 5514–5519. ISSN: 0264-410X. DOI: https://doi.org/10.1016/j.vaccine.2014.06.092. URL: https://www.sciencedirect.com/science/article/pii/S0264410X14009761 (cit. on p. 2).

[16] NSaadia Zahidi. "World Economic Forum Global Risk Report 2024". In: *World Economic Forum, Geneva.* 2024 (cit. on p. 2).