Prometheus unbound: LLMs and the future of public health

Chris von Csefalvay

2024-01-01

UNCLASSIFIED

Introduction

Your speaker

I’m **Chris von Csefalvay** $^{\text{CPH FRSPH MTOPRA}}$. I’m a data scientist and computational epidemiologist. My work focuses on epidemic surveillance and computational modeling of infectious diseases. I wrote the book on that.

I’m Practice Director of Advanced Biomedical AI/ML at HCLTech, where I advise the world’s leading pharmaceutical and biotech companies on how to use AI/ML to improve their R&D and clinical operations.

DISCLAIMER

Statements made in this presentation are the author’s own in his personal capacity. They do not reflect the views of the United States Government, the Commonwealth of Virginia, HCL America Inc. or of any organisation, company or board he is associated with.

What we'll discuss

- What LLMs are (very, very briefly)
- Why they matter, esp. how they're different from other ML models (spoiler alert: they're not)
- How they can be used in public health
- How they can be abused in public health
- How we can use them responsibly

What we won’t discuss

- The effect of finite *vs.* infinite horizon discounted returns in policy optimisation for training LLMs using RLHF
- Generally any other technical point
- Whether LLMs are going to usher in the age of SkyNet and/or take our jobs
- Upstream ethical issues that would arise with any other model, e.g. misuse to support harmful policies

Assumptions

- You know the very basics of machine learning: testing, training, inference &c.
- You have a good understanding of the concerns of public health (as you should, you’re mostly all CPHs!).
- You can suppress visceral reactions to mathematics for one slide’s duration.

Room 161, Building 6, CDC

On 28 April 1981, Sandy Ford, a drug technician for the CDC, noted the spike in requests for pentamidine for PCP (now called *P. jirovecii*). This as the first real, data-driven signal to draw attention to what became the HIV/AIDS pandemic.

Room 161, Building 6, CDC (cont.)

By then, we had:

- the very early cases (pre-ZR59/DRC60)
- the earliest American cases (late 1950s-early 1960s)
- first paediatric cases in the US (1977/78)
- …and probably a lot more activity under the radar.

The requests for pentamidine created a convenient *quantitative* data point. We are rarely so lucky.

What if we could discern data points like this from EMR/EHR data? LLMs bridge the gap between free text information and quantitation.

Three tools

Instructions:

1. **Look** at these examples of LLMs in public health.
2. Think about your **‘knee jerk reaction’**. Would you be comfortable using (implementing) such a tool as a public health professional? Would you be comfortable using such a tool as an end user?
3. **Imagine** a world in which this tool doesn’t merely exist but is ubiquitous and widely accepted, maybe even mandatory and/or governmentally endorsed. How would you feel living in such a world?

1

A tool that uses social media data to isolate geographically co-occurring patterns of symptoms of infectious respiratory illness.

- Potential for early detection
- Rapid pandemic response to emerging pathogenic threats
- Privacy risks?
- Stigmatisation?

2

A pharmacovigilance tool that isolates symptoms from passive reporting systems and creates a DEW (Distant Early Warning) system for adverse events.

- Potential for early intervention, thus preventing harm
- Risk of getting it wrong… possibly quite badly (‘down the garden path’)?
- Potential for abuse (VAERS) and misinterpretation (also VAERS)?

3

A tool that helps social scientists automatically code qualitative data and interviews.

- Acceleration of qualitative research
- Reducing research cost and barriers to entry
- Garbage in, garbage out?
- "Black box" effects?

The word set free

> In the beginning was the Word, and the Word was with God, and the Word was God.
>
> – Gospel of John ch.1 v.1

What is a language model?

- A language model is simply a mathematical model (specifically, a sequence model) that operates on tokens (usually, words) in a sequence.
- It is trained on a corpus of text, which is a collection of sequences of tokens.
- Most often, the tokens are words, but they can be anything, including characters, phonemes, morphemes, or even whole sentences.

This is the only slide that will contain maths. I promise.

Training

$$\hat{\Phi} = \underset{\phi}{\operatorname{argmin}}\, J(f_{\phi}(k_1, k_2, \cdots, k_n), k_{n+1})$$

Inference

$$k_{n+1} = \underset{w \in W}{\operatorname{argmax}}\, p(k_{n+1} = w | k_1, k_2, \cdots, k_n, \Phi)$$

And now in human terms

Training

“Find me the parameters that describe a function that, for each sequence of tokens, maps the most likely next word to any sequence of words.”

(The maximum length of tokens n here is called the *context window*.)

Inference

“Given a sequence of words, what is the most likely next word?”

Some details on training

- How training **works**: **RLHF** (Reinforcement Learning with Human Feedback), a way of rewarding ‘correct’ guesses of the $n + 1$-th token.
- How training is **transferred**: **transfer learning** allows a model to leverage low level features (think: connectionism on multiple levels) of a foundation model to be used for more complex models.
- **What** these models are trained on: various source corpora – CommonCrawl, RefinedWeb, Twitter/X, Reddit, Youtube comments (seriously?)

Some details on training (cont.)

Each of these is meaningful:

- Different RLHF ‘policy’ preferences create different choices.
- Transfer changes context.
- Source corpora changes understanding/knowledge base (imagine training an LLM entirely on hardboiled detective fiction, then asking how it would resolve a problem).

Emergent magic

So far, so good. Where this becomes interesting is the fact that if you train these models on sufficiently large data sets, their answers will start to make sense.

(We’ve ignored a lot of what goes into this, including transformer architectures, attention &c. – these are useful from a technical perspective, but not a functional perspective.)

If a language model is trained on a sufficiently large corpus of text, it will start to give answers that reflect a kind of knowledge/understanding of the world.

Emergent knowledge

> Q: What is the capital of Hungary?
>
> GPT-4: The capital of Hungary is Budapest.

Note that GPT does not have an ‘understanding’ of what Budapest is, or a capital is, or Hungary is. It only has an understanding that given the token sequence

```
["the", "capital", "of", "hungary", "is"]
```

the most likely next token is

```
["budapest"]
```

Interlude: a note on semantics

- **Machine learning**: a branch of applied mathematics that deals with systems that improve with data
- **AI**: a (fairly ill-defined) branch of applied machine learning
- **LLMs**: the use of AI techniques to tackle language modeling tasks (comprehension, translation, entity extraction, response, summarisation, synthesis,…)
- **GPT** (Generative Pre-Trained Transformer): an extremely popular foundation model, currently on its 4th iteration (GPT-4)
- **ChatGPT**: an implementation that uses GPT to act as a chatbot.

Emergence phenomena

What we found is that **if you train a model on enough data, its responses will simulate a kind of knowledge**.[1]

(Or, less anthopomorphically put: given enough data, the model builds an association between Budapest" and "Hungary". We sometimes refer to this emergent understanding of the world as the model's "knowledge base", but is *still* nothing mystical – just weights & biases that encode the model's understanding of conditional probabilities of tokens.)

[1] For a given value of 'knowledge.

Consequences

A uniquely powerful tool…

LLMs are uniquely powerful tools for understanding language.

- parsing free text summaries in clinical records
- targeted entity extraction and concept extraction
- social listening using LLMs for e.g. epidemic signals

…with unique challenges…

LLMs also have unique challenges.

- They are not interpretable. LLMs are much less interpretable than even other neural network based models. We know the answers, but not *why* they’re the answers.
- They are not feasibly explainable (in the sense that they cannot be explained in terms of a set of rules or a decision tree).
- They make mistakes, e.g. hallucinations.

…requiring a unique approach.

The consequence is that there won’t be a one-size-fits-all ‘consumer’s guide to LLMs’.

- What tasks can be trusted to LLMs? What guardrails need to be present?
- Where don’t we mind the risks?
- Where do we need to be more careful?

Applications

Contact

- LLM based chatbots for public health functions:
  - Contact tracing
  - Countering vaccine hesitancy
  - Countering misinformation
  - Health education
  - Mental health support
- Rapid generation of accurate yet customised health content in a multiplicity of languages
  - Health education
  - Emergency response

Contact (cont’d)

- Effective communication using social media bots
  - Social listening using LLMs
  - Social media bots for health promotion
  - Countering misinformation
- AI driven avatars and other multimodal experiences

Analytical use of LLMs

- LLMs can be used to analyse free text data, e.g.:
  - Clinical records
  - Social media posts
  - Qualitative data
- Automated coding of qualitative data
- Entity and concept extraction from free text data

Public health surveillance

- Social listening using LLMs for prodromic signals
- Signal generation from qualitative data
- Mining EHRs/EMRs for unusual presentations

Risks and rewards

> Behind the black portent of the new atomic age lies a hope which, seized upon with faith, can work our salvation. If we fail, then we have damned every man to be the slave of Fear. – Bernard Baruch, Address to the UN on the Baruch Plan, 14 June 1946

Opportunities

- Capture a vast ocean of unstructured data and understand it better than ever, facilitating improved early reporting
- Respond faster, better and more directly to provide tailored information to the public and fight misinformation
- Accelerate research by directly capturing unstructured input
- Promote health behaviours and health education through generative natural language applications (chatbots, multi-language content generation,…)

Threats

- Privacy
- Equity
- Just allocation of scarce resources
- Algorithmically entrenched prejudices and preferences

Some questions

- How will this affect global health? Who is going to be left behind?
- How will this reinforce existing health disparities? Will a ‘digital divide’ emerge between patients of tech-savvy physicians at well-funded institutions and overburdened clinicians who barely manage their case loads and have no time to learn new tools?
- How will this affect the allocation of resources? Will we be able to use these tools to allocate resources more efficiently, or will they be used to reinforce existing biases?
- In short: are we headed for heaven or hell?

Some tentative conclusions

- We must all become, if not experts, at least expert users of such systems, able to critically reason about what goes in them and what comes out of them.
- The best perspective is a multidisciplinary one. We need to understand the technical aspects of these systems, but also the public health aspects.
- Multidisciplinary problems require multidisciplinary solutions.

A coalition of solutions

- The technical perspective: better guardrails, better explainability, better interpretability
- The social perspective: a strong ethical stance that supports innovation but fosters responsibility (to each other, to the public and to the planet)
- The public health perspective: bringing in the equity and justice values that have always been at the core of public health

Thank you for your attention.