D001 Economic Analysis of Non-Standard Data

Benjamin W. Arold

11. Ethical Considerations

Now, we move from "What is?" to "What ought to be?"

- Most of the content discussed so far concerns positive questions
  - How does a given model work
  - What are we measuring, etc.
- Today, we will discuss the ethics of text-, image- and audio-data related questions
  - How should models be built from a social viewpoint
  - How should we (not) use them, etc.
  - How should textual, visual or audio data (not) represent our society
- Ethics discussions are necessarily influenced by moral considerations
  - What is good or bad? Just or unjust

Scientists do not operate outside of society

- Positive science is interested in discovery and facts (the "truth" about the world)
- However, positive science interacts with normative considerations in many ways, for example:
  - Which questions do we deem important to research
  - What impact will our discovery have on society
  - What cost are we willing to carry for discovery (not just monetary)
  - What responsibilities do we (not) have as scientists
  - Further examples
- Values and other context-specific characteristics interact with the entire social science pipeline:
  - Research question $\rightarrow$ (data collection) $\rightarrow$ theoretical or empirical analysis $\rightarrow$ interpretation

Let us discuss a historic case study: Census data

- A government wants to establish a comprehensive population registration system for administrative and statistical purposes
- The system should simplify administration and offer perspectives for social research
- Questions:
  - Do you have any ethical concerns
  - Would you be more/less concerned if it is a company/researchers/another actor constructing the comprehensive database? Why?

History has seen horrible examples of unanticipated secondary use

- During the Second World War: Dutch census data used to facilitate the genocide against Jews and others:
  - Data collection during peaceful times
  - Trusted by citizens
  - Later misuse by the Nazis
  - Extremely high death rate of Dutch Jews: 73%
- More: Seltzer, W., & Anderson, M. (2001). The Dark Side of Numbers: The Role of Population Data Systems in Human Rights Abuses. Social Research, 481-513

Let us discuss further case studies

- Surveillance technology
  - Should governments use face recognition technology? If so, for which applications?
  - How does this question relate to economics research?
- Representation of different groups (e.g., by ethnicity, gender, age)
  - From a normative viewpoint, what is the "right" amount and kind of representation of different societal groups in the media
  - How does this question relate to economics research?
- Accents and labor market outcomes
  - (Under what circumstances) should employers be allowed to consider a worker's accent in their hiring decision
  - How does this question relate to economics research?

The Environmental Impact of LLMs

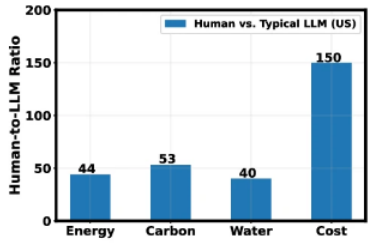
- LLMs require a significant amount of energy at all stages
  - ChatGPT is estimated to consume the energy of 33,000 households
  - Google carbon emissions have increased by 50% to build and sustain new data centers (source)

The Environmental Impact of LLMs

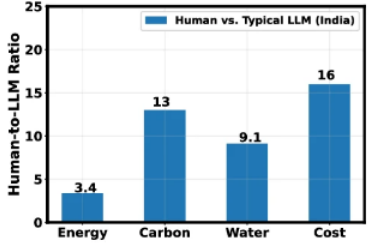
- LLMs require a significant amount of energy at all stages
  - ChatGPT is estimated to consume the energy of 33,000 households
  - Google carbon emissions have increased by 50% to build and sustain new data centers (source)
  - While training GPT-4, OpenAI consumed around 6% of West Des Moins (Iowa) district in a month

The Environmental Impact of LLMs

- LLMs require a significant amount of energy at all stages
  - ChatGPT is estimated to consume the energy of 33,000 households
  - Google carbon emissions have increased by 50% to build and sustain new data centers (source)
  - While training GPT-4, OpenAI consumed around 6% of West Des Moins (Iowa) district in a month
- Recent studies have highlighted how LLMs are a more efficient alternative to human labor



(a) Typcial LLM (U.S.)



(b) Typical LLM (India)

The Environmental Impact of LLMs

- LLMs require a significant amount of energy at all stages
  - ChatGPT is estimated to consume the energy of 33,000 households
  - Google carbon emissions have increased by 50% to build and sustain new data centers (source)
  - While training GPT-4, OpenAI consumed around 6% of West Des Moins (Iowa) district in a month
- Recent studies have highlighted how LLMs are a more efficient alternative to human labor. However:
  - LLMs size is growing $\Rightarrow$ energy demand will increase
  - In the short term, LLMs are unlikely to substitute humans fully

So What Do We Do?

- Given the fast evolution of these technologies it's hard to make recommendations

So What Do We Do?

- Given the fast evolution of these technologies it's hard to make recommendations
- **Users:** Responsible use of GenAI
  - Revert to Google query for web searches (more reliable and consumes 10 times less than ChatGPT per query)
- **Companies**: invest in research for *greener* AI systems