

Artificial Intelligence

Lecture 12-2

Computers & Society (CPSC 430)

Hedayat Zarkoob

<https://www.students.cs.ubc.ca/~cs-430/>

Original slides by Kevin Leyton-Brown

Concern: Education

- LLMs will **profoundly affect education**
 - ChatGPT generates decent (~B) essays easily
 - Can also rewrite text, critique arguments, solve problem sets, answer comprehension questions
- We will need to **reconsider what we teach**
 - is the ability to generate fluent prose still important?
 - analogous to long division when students have calculators
 - what critical thinking skills become important when we assume that students have access to AI?
 - ultimately, our need for education won't go away
- We will need to **reimagine evaluation**
 - currently, assignments based on what is hard to do, easy to grade
 - that balance just changed!
 - ultimately, one can foresee LLMs serving as part of the solution



Concern: LLMs in the workplace

OpenAI report:

- How “exposed” are different occupations to automation with large language models (LLMs)?
- ~80% of the US workforce could have at least 10% of their work tasks affected by LLMs
- ~19% may see at least 50% of their tasks impacted
- Most affected tasks: writing and programming.
- Higher-income jobs (such as translators, tax consultants, and web designers) potentially face greater exposure

GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models

[Tyna Eloundou, Sam Manning, Pamela Mishkin, Daniel Rock](#)

We investigate the potential implications of large language models (LLMs), such as Generative Pre-trained Transformers (GPTs), on the U.S. labor market, focusing on the increased capabilities arising from LLM-powered software compared to LLMs on their own. Using a new rubric, we assess occupations based on their alignment with LLM capabilities, integrating both human expertise and GPT-4 classifications. Our findings reveal that around 80% of the U.S. workforce could have at least 10% of their work tasks affected by the introduction of LLMs, while approximately 19% of workers may see at least 50% of their tasks impacted. We do not make predictions about the development or adoption timeline of such LLMs. The projected effects span all wage levels, with higher-income jobs potentially facing greater exposure to LLM capabilities and LLM-powered software. Significantly, these impacts are not restricted to industries with higher recent productivity growth. Our analysis suggests that, with access to an LLM, about 15% of all worker tasks in the US could be completed significantly faster at the same level of quality. When incorporating software and tooling built on top of LLMs, this share increases to between 47 and 56% of all tasks. This finding implies that LLM-powered software will have a substantial effect on scaling the economic impacts of the underlying models. We conclude that LLMs such as GPTs exhibit traits of general-purpose technologies, indicating that they could have considerable economic, social, and policy implications.

 The report answers the question “which jobs do **people think** LLMs will automate” and not “which jobs will LLMs automate”.

Concern: LLMs in the workplace

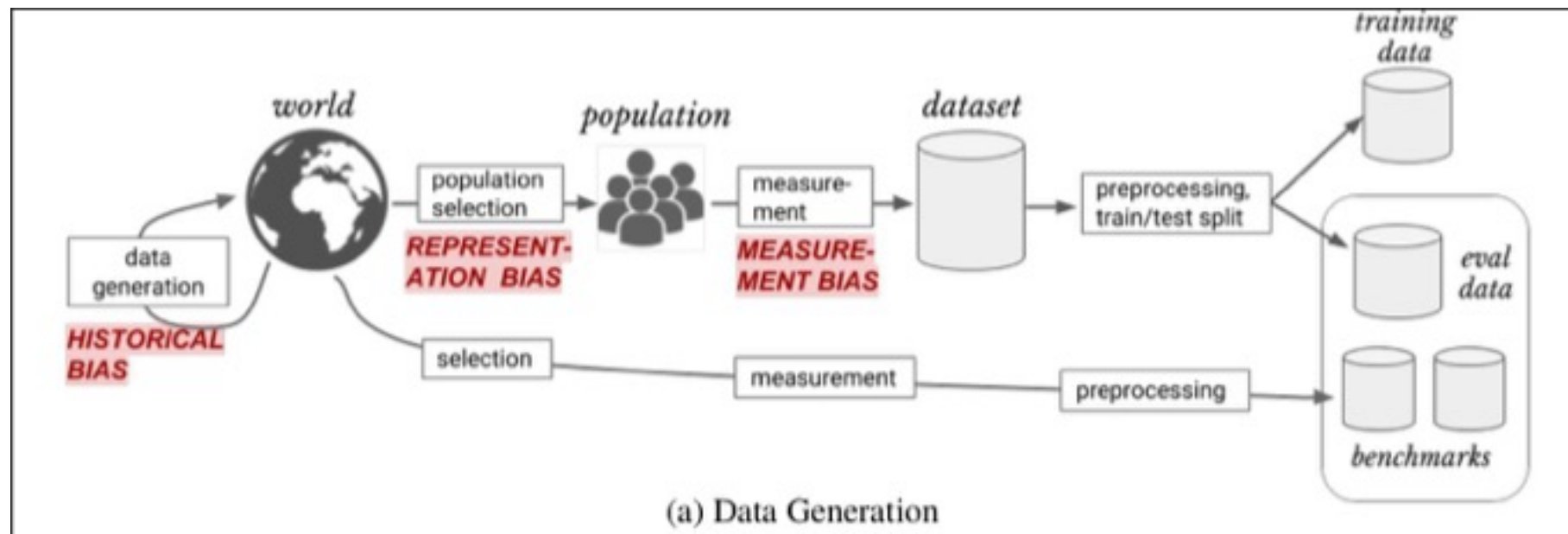
- Adoption of Large Language Models for various tasks, including work-related ones
 - Some people compare it to the advent of Google
 - Some issues (ownership of training data, allucinations...)
- Do you think that AI will increase the expectations from professionals in the workforce? How?

Concern: Bias

- **Bias** refers to the tendency to take quick decisions based on little information
- ML systems trained on biased data will **exhibit biased performance**
 - e.g., face recognition systems that underperform for racial minorities
 - In principle there's an easy fix: debias the data
- AI systems can be **less biased** than humans!
 - Judges hand down stiffer sentences before lunch
 - AI systems are auditable



Sources of bias in ML algorithms



- **Representation bias** arises while defining and sampling a development population
- **Measurement bias** arises when choosing and measuring features and labels to use
- **Historical bias arises** when there is a misalignment between world as it is and the values or objectives to be encoded and propagated in a model

Concern: Fairness in algorithms

- Increasing attention on algorithms being **fair**, not just accurate
- Fairness can be measured as:
 - demographic (or statistical) parity: population percentage should be reflected in the output classes
 - Equality of false negatives or equalized odds: constant false-negative (or both false-negative and true-negative) rates across groups.
 - Equal opportunity: equal True Positive Rate for all groups
 - Other metrics...
- Accuracy and fairness tend to be at odds with each other.
- Algorithms can be audited to test their fairness.
- *Are we ethically required to sacrifice accuracy for fairness? To what extent?*

Concern: Creators

- Will generative AI **harm content creators**?
 - artists, novelists, actors, ...
- Is training a generative model on creative work a **violation of copyright law** or “fair use”?
 - should the law change?
 - risk of memorization
 - is it wrong to reference existing artists in style instructions?
 - picture of the Vancouver skyline in the style of Van Gogh's “Starry Night”
- **Photoshop** didn't put photographers out of a job
 - and ultimately helped increase our consumption of photography
 - but is this time different?



Concern: Sentience

- “AI will become **self-aware** and kill us all”
- Today, there’s no “**ghost in the machine**”
 - modern AI systems don’t form goals or take actions
 - really a mapping from (e.g.) text to more text
 - not sitting there thinking when not in use
 - engineered systems with safeguards
 - ChatGPT can’t even tell a dirty joke!
- But **fears persist** about what the future holds
 - rapid rate of improvement, recently accelerating
 - nobody understands the “hard problem of consciousness”
 - effective altruists have made influential arguments about “existential risk”
 - but many of them also say we should focus more on colonizing Mars than e.g. helping poor people alive today



Conclusions

- AI is **no longer science fiction**
 - We've been laying the scientific foundations for 70+ years
 - The biggest breakthroughs have happened in the past half decade; progress remains rapid
 - No evidence that we're on the road to artificial sentience
 - No guarantees progress won't stagnate soon
 - there's a limit to how much money can be spent on training a model
 - we're running out of training data
- Existing technologies will have big impact
 - **Mobile phones** changed society profoundly; AI will be bigger
 - **Living standards** will rise; effectively, we'll all be richer
 - Widespread unemployment not around the corner, but AI will **automate many tasks** that now need human expertise
- **Government** (and voters) will shape the agenda
 - funding; adopting technologies; **regulation**

