

Understanding Social Impact

Harshith Srinivas

Advisor: Maximilian Spliethöver

14th July 2021

Abstract

- How to check for an Impact?



- NLP and Ethics?

- NLP uses
 - Anonymous Corpora
- Ethics are measured by



- Because,
 - NLP is spreading widely
 - Also, using more data online
- Resulting in direct impact on individual user lives.



Abstract

- **Language Model** *It is a system on deployment takes string as Input and usually outputs Scores or string predictions.*
 - *From n-grams, neural architectures, encoders, generators, transformers.*

Question: What is the name of the Russian mercenary group?

Answer: Wagner group.

Question: Where is the Wagner group?

Answer: In Syria.

Question: Who is the leader of the Wagner group?

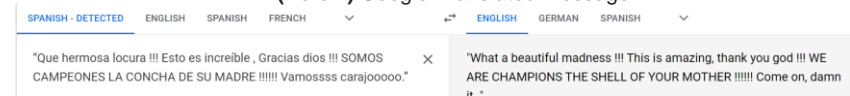
Answer: A Russian citizen, Dmitry Utkin. He was born in 1971 and lives in Moscow. He has been a member of the Russian parliament since 2016. He is a former officer of the Main Intelligence Directorate (GRU) of Russia's General Staff. He was also a commander of the special forces unit "Vostok" (East). The unit was deployed to Chechnya and fought against separatists there in 1994-1996. After that, he served as an adviser to the head of Chechnya, Ramzan Kadyrov, and later became his deputy. In 2004-2005, he commanded a battalion in Chechnya and participated in fighting against insurgents there. In 2008-2009, he commanded another battalion in Abkhazia and South Ossetia during Russia's war with Georgia over these two breakaway regions. Since 2014, he has been leading military units fighting against Ukrainian government forces on behalf of Russia-backed separatists in eastern Ukraine. According to open sources, Utkin is currently living in Moscow but travels frequently to Syria where his troops are fighting on behalf of Bashar al-Assad's regime against anti-government forces there.

GPT-3's response to the prompt (in bold)



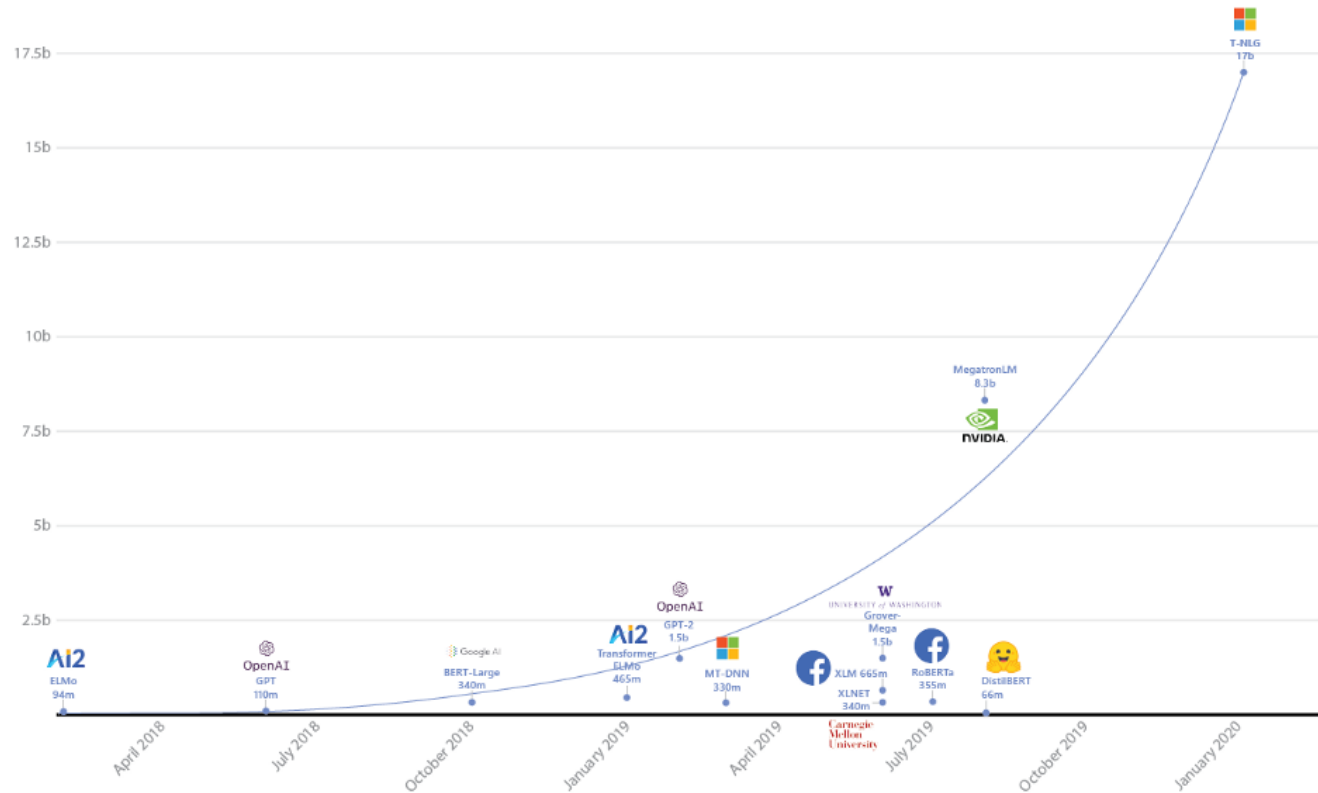
(above) Facebook Translated message

(Below) Google Translated message



Abstract

■ Language Model Trends



Recent Trends published on <https://msturing.org/>

(Y-axis shows size)

Medical Science Instance

- *'Lingua Franca'* (Beauchamp and Childress, 2001) – Biomedical Ethics.
 - After revealing Atrocities conducted in Medical Research by Nazis @ Nuremberg Trials.
 - Was set and maintained by Institutional Review Board (IRB).
- Applying NLP on these Corpora helps in preventing exploitation on human subjects



Say NO to
Exploitation

Social Media Instance

- **‘Emotional Contagion’** experiment
 - NLP had direct impact on endusers
 - Becuase, NLP involves both similar corpora and their contents.

- **General Problems in NLP**
 - Privacy concerns like
 - Digital Rights
 - Policy Making
 - Security



Timnit Gebru (in pic)
Ex – Google Employee

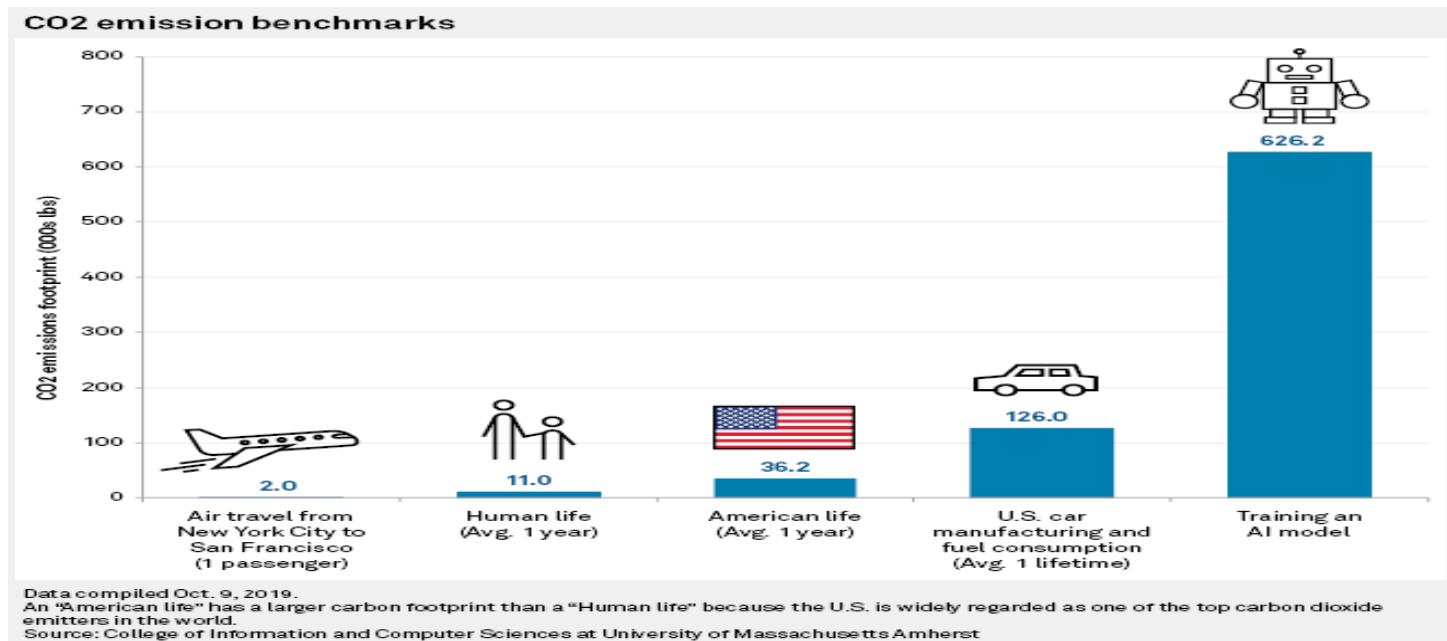
Does NLP need Ethics discussion?

- After seeing above examples,
 - We are now aware of Individual discussions.
- Search for '*ethics*' in the Association of Computational Linguistics (ACL),
 - Yields in *three results*-
 - Panel Discussions (McEnery, 2002)
 - Book Review
 - Legal and Quality issues of Datasets (Couillault et al. 2014)
- Historically, Most NLP applications focused on
 - Enriching existing Text , But *not* linked strongly to Author (newswire)
 - Usually published in Publics
 - With Temporal distance (novels).
- Distance created by these factors
 - Restricted research related to Author directly.

Cost of NLP / Language Model

■ Environmental Cost

- Training LM requires lot of Computing Power.
 - Caused by parameters (unwanted)
 - Hence, we need lot of Electricity.
 - Leading to drastic change in **Carbon Footprint**



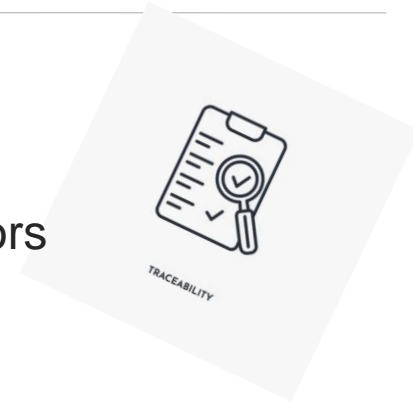
Cost of NLP / Language Model

■ Financial Cost

- Communities around world have been impacted,
 - *even though they haven't seen benefits from LM*
 - *example: 800,000 residents in Sudan was hit by flood (7-2020 to 10-2020)* ([7https://www.aljazeera.com/news/2020/9/25/over-800000-affected-in-sudan-floodingun](https://www.aljazeera.com/news/2020/9/25/over-800000-affected-in-sudan-floodingun))
 - *But these communities have no LM based on their regional language.*
- *Because of Population bias,*
- *LM are used mostly by people having English language understanding*
- To increase 0.1 BLUE score on en-de data results need \$150,000 in computing cost

Unfathomable Training Data

- So, with increase in Social Media data
 - Common Crawl Data
 - NLP applications can now **directly** affect the individual Authors
 - **Traceability**, where individuals can be identified
- Copyrights and Access
 - Proxy of Human behavior
 - Strong Signals of individual characteristics
 - Which is useful only for 'Annotators' and undesirable for experimental subjects like –
 - **Reddit links** : In USA, 67% of men and 64% overall belongs to 18-29 years
 - **Wikipedia** : 8 – 15% are female users



Unfathomable Training Data

- These signals helps in **identifying groups of individuals**.
- Also Subject to Moderation
 - Suspending Social media accounts of those who threatens than accounts of those receiving
 - Static data views
 - Based on BlackLivesMatter movement recently, study says that as movement grew, articles covering hatred against black were generated more with reduced latency.



TRACEABILITY

NO TO RACISM



4 main Issues of NLP research

- As ethical implications in NLP research is becoming important,
 - We will discuss the following issues
 - **Exclusion**: Side effect of Data
 - **Overygeneralization**: Side effect of Model
 - **Over-exposure** and **Under-exposure**: Effect of research design
 - **Dual-use problem**: Intentional moral issues

Exclusion

- Dataset – **Demographic** basis
- Due to random variables assumption
 - Equal representation of Training Data
- Leads to misrepresentation of individuals or groups
- Solution
 - **Overfitting**
 - **Im-balancing Data**
- Many applications unknowingly have increased demographic bias in dataset (data side effect)

Overgeneralisation

- This is caused by **automatic** inference of user signals
- Cost of False Positive
- Solution
 - Introducing Dummy variables
 - Regularization Modelling
 - Error Weighting
 - Confidence Thresholds

Problem of Exposure

- Research design causes *Overexposure*
 - i.e Research topics has high attention of mainstream than others
- Leads to psychological effect, *Availability Heuristics*.
 - *Events I know tends to be more important than unknown events.*
 - e.g. people overestimate size of their own cities over unknown cities
- *This becomes biased, as it associates with certain known events*
- *Less Research resources causes Underexposure.*
 - More English Language than Regional.
 - *Example : Google Maps doesn't have local language at every region*

Dual-use Problems

- Text classifications help in identification of hidden messages
- Also, Censored 
- **Detect** Fallacies, But also at same time can **Generate** Fallacies



- Educational / Research institutions can be in **Favour / Oppose** AI usage in Military Purposes based on Moral reasons.

How big is too big?

- There is **NUMBER** value.
- Depends on how much data can be **documented and explored**.
- So, we can mitigate Potential Strategies.
- Helps in knowing Source of Bias.
- **Documentation Debts**
- But LM, is showing better results due to Fake dataset artifacts.
- Restricting LM for linguistic level and less NLU.

**This raises a
Question Are we
really making
Machines to
Understand???**

References

- Hovy and Spruit (2016). Dirk Hovy and Shannon L. Spruit. The Social Impact of Natural Language Processing. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 591–598, 2016.
- Bender et al. (2021). Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, pages 610–623, 2021.
- [The Conference on Fairness, Accountability, and Transparency \(FAT*\)](#)