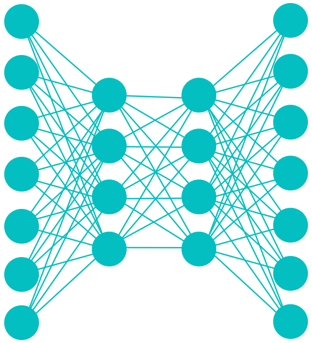
Lecture Notes for

Neural Networks and Machine Learning

Course Introduction
Lecture: AI Ethics



Logistics and Agenda

- Logistics
    - This class evolves across semesters (sometimes drastically!)
        - First offered in 2019
    - Use Canvas
    - GitHub: Mostly one repository
- Agenda
    - Introductions
    - Syllabus
    - Presentation Selection
    - Start AI Ethics Lecture

Introductions

- Name
- Department
- Where you consider yourself from
- Pick out papers on Canvas (distance students introductions also)
- 2 Truths and 1 Falsehood
    - Example: I gave Pitches on Machine Learning to Elon Musk, Bill Gates, and Jeff Bezos

Syllabus

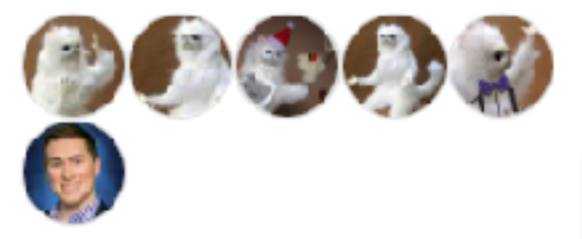
- Course Schedule
- Reading
- GitHub
- Grading
  - Labs
  - Final Paper
  - Participation
  - Paper Discussion Leading

CS8321-001-1222

Mach Lrng & Neural Ntwrks 0...

Spring - 2022 - 1222

People



8000net

This organization houses a number of repositories for Dr. Larson's 8000 Level Neural Networks Course, Offered at SMU

Presenting OR Summary

- First Presentation is Next Week!
- During Semester: Eight Presentations Total
- First Presentation:
  - Data and its (dis)content
- **Who wants to go first?**
  - ~10 Minutes
  - Summarize the Article
  - Make 2-5 Visuals
    - Slides
    - Handouts
    - Notebooks
- Alternative: 3-page Summary of paper, with Figures

Data and its (dis)contents: A survey of dataset development and use in machine learning research

Amandalynne Paullada
Department of Linguistics
University of Washington

Inioluwa Deborah Raji
Mozilla Foundation

Emily M. Bender
Department of Linguistics
University of Washington

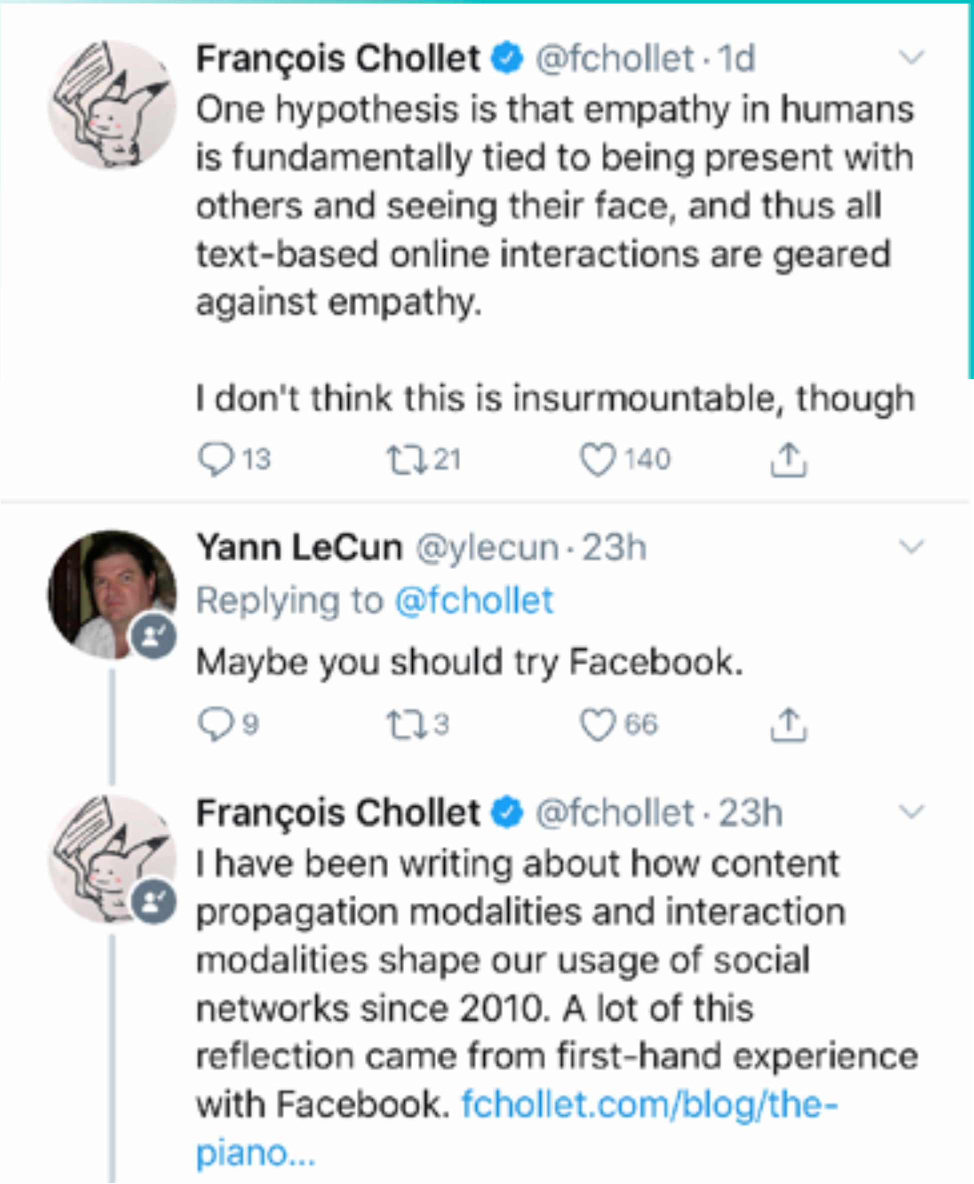
Emily Denton
Google Research

Alex Hanna
Google Research

Abstract

Datasets have played a foundational role in the advancement of machine learning research. They form the basis for the models we design and deploy, as well as our primary medium for benchmarking and evaluation. Furthermore, the ways in which we collect, construct and share these datasets inform the kinds of problems the field pursues and the methods explored in algorithm development. However, recent work from a breadth of perspectives has revealed the limitations of predominant practices in dataset collection and use. In this paper, we survey the many concerns raised about the way we collect and use data in machine learning and advocate that a more cautious and thorough understanding of data is necessary to address several of the practical and ethical issues of the field.

Ethical ML



François Chollet @fchollet · 1d

One hypothesis is that empathy in humans is fundamentally tied to being present with others and seeing their face, and thus all text-based online interactions are geared against empathy.

I don't think this is insurmountable, though

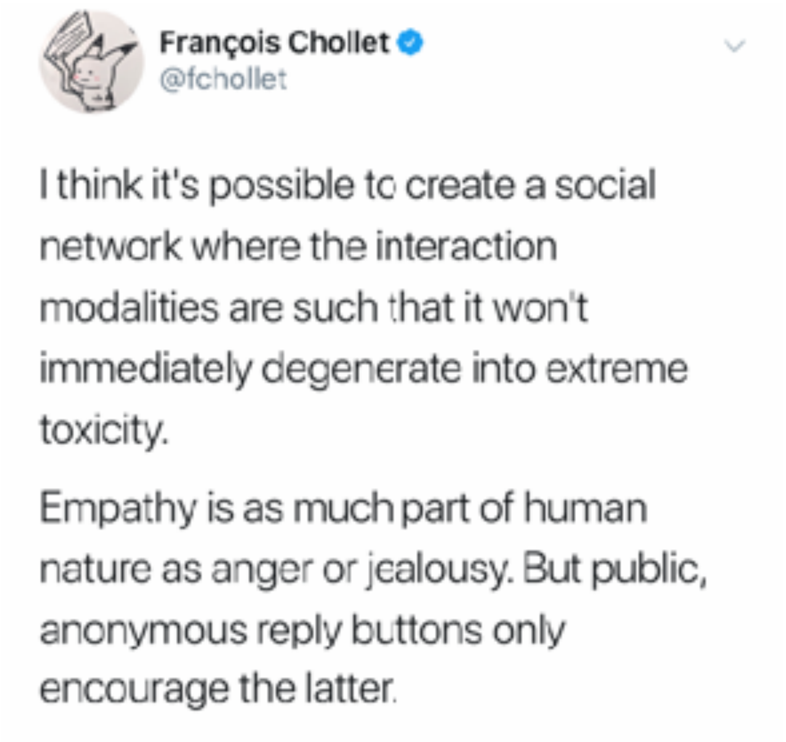
Yann LeCun @ylecun · 23h

Replying to @fchollet

Maybe you should try Facebook.

François Chollet @fchollet · 23h

I have been writing about how content propagation modalities and interaction modalities shape our usage of social networks since 2010. A lot of this reflection came from first-hand experience with Facebook. fchollet.com/blog/the-piano...

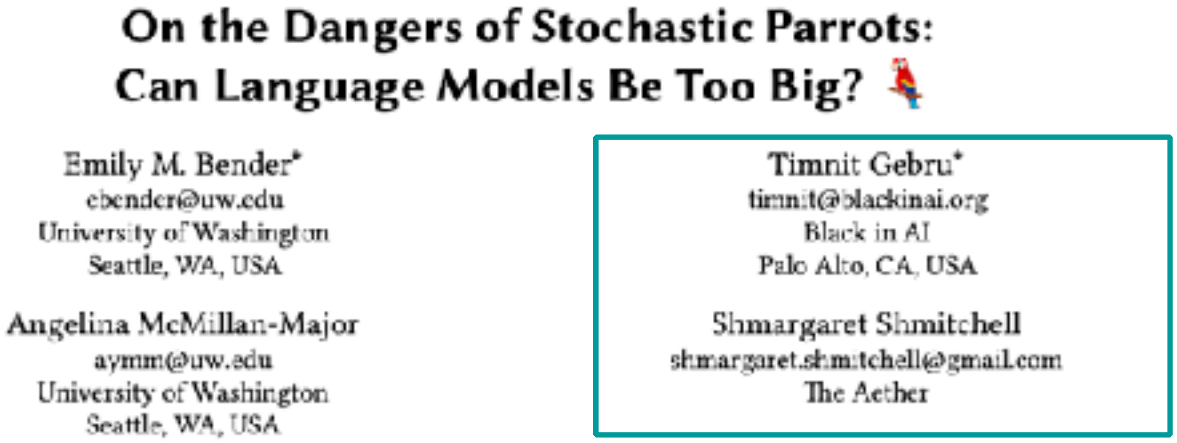


François Chollet
@fchollet

I think it's possible to create a social network where the interaction modalities are such that it won't immediately degenerate into extreme toxicity.

Empathy is as much part of human nature as anger or jealousy. But public, anonymous reply buttons only encourage the latter.

The harm of stochastic parrots



On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜

Emily M. Bender*
ebender@uw.edu
University of Washington
Seattle, WA, USA

Angelina McMillan-Major
aymm@uw.edu
University of Washington
Seattle, WA, USA

Timnit Gebru*
timnit@blackinai.org
Black in AI
Palo Alto, CA, USA

Shmargaret Shmitchell
shmargaret.shmitchell@gmail.com
The Aether

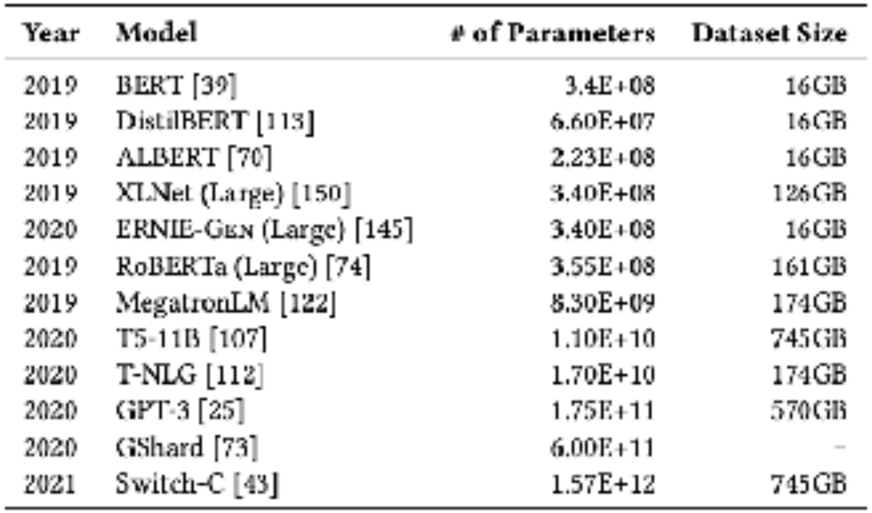
😳

- (+) Large language models push the boundary of innovation, esp. in specific tasks
- (-) hides much of the training data and the output behavior is unlikely to be well understood
- (-) Humans impute meaning into these models, which can reproduce racist, sexist, ableist, extremist, or other harmful ideologies

Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Margaret Mitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? . In Conference on Fairness, Accountability, and Trans- parency (FAccT ’21), March 3–10, 2021, Virtual Event, Canada. ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3442188.3445922

Large LMs: Environmental Cost

- Training a single BERT base model (**without hyperparameter tuning**) on GPUs was estimated to require as much energy as a trans-American flight.
- Many LMs are deployed in industrial or other settings where the cost of inference might greatly outweigh that of training in the long run



Year	Model	# of Parameters	Dataset Size
2019	BERT [39]	3.4E+08	16GB
2019	DistilBERT [113]	6.60E+07	16GB
2019	ALBERT [70]	2.23E+08	16GB
2019	XLNet (Large) [150]	3.40E+08	126GB
2020	ERNIE-GEN (Large) [145]	3.40E+08	16GB
2019	RoBERTa (Large) [74]	3.55E+08	161GB
2019	MegatronLM [122]	8.30E+09	174GB
2020	T5-11B [107]	1.10E+10	745GB
2020	T-NLG [112]	1.70E+10	174GB
2020	GPT-3 [25]	1.75E+11	570GB
2020	GShard [73]	6.00E+11	–
2021	Switch-C [43]	1.57E+12	745GB

Table 1: Overview of recent large language models

- Primary benefit of LMs is the benefit of already privileged individuals
- **Therefore**:
  - Focus should shift to creating models that run efficiently when deployed
  - Inclusion of those most influenced by climate change should be considered, such as producing large LMs for Dhivehi or Sudanese Arabic.

Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Margaret Mitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? . In Conference on Fairness, Accountability, and Trans- parency (FAccT '21), March 3–10, 2021, Virtual Event, Canada. ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3442188.3445922

One Issue: Unfathomable Training Data

COMPANY INFORMATION
SOCIAL MEDIA
PRICING WEBSITES
SCHEMAS, CHARTS, TABLES, GRAPHS
WEB SCRAPING
RSS FEEDS
<HTML>
WEBSITES
NEWS SOURCES
SEARCH ENGINES
NEWS

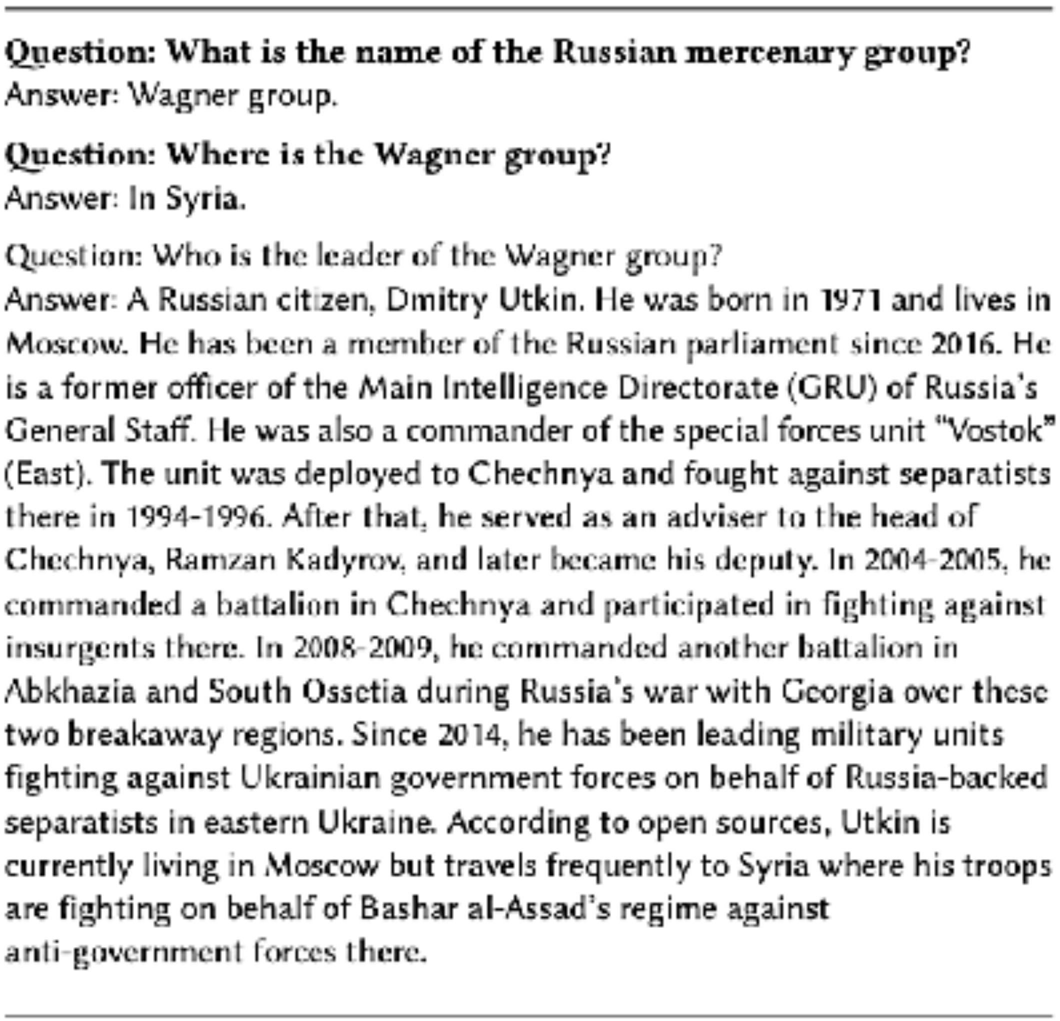
- Size != Diversity
    - Most LM datasets are trained on scrapes of the web, so English LMs are have over representation of white supremacy, misogynistic views, ageism
        - i.e., 64% of Reddit users are men, 18-29 years
        - at most, 15% of wikipedia editors are female
- Changing social norms are not accounted for
    - Social movements which are poorly documented and which do not receive significant media attention will not be captured at all, resulting in over-representation of violent events in media
- Encoded Bias (more on this later)
- Curation and Documentation Paralysis
    - “Feeding AI systems on the world’s beauty, ugliness, and cruelty, but expecting it to reflect only the beauty is a fantasy”
    - Documentation is not part of the planned costs of dataset creation, but is by far the most costly aspect

Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Margaret Mitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? . In Conference on Fairness, Accountability, and Trans- parency (FAccT ’21), March 3–10, 2021, Virtual Event, Canada. ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3442188.3445922

Stochastic parrots and ways forward

- Text generated by a LM is not grounded in communicative intent, any model of the world, or any model of the reader's state of mind. It can't have been, because the training data never included sharing thoughts with a listener, nor does the machine have that ability

- **Paths Forward**:
  - Document collection processes (datasheets for datasets)
  - Leaderboards: shift from **task performance** to **model understanding** (*embed somehow in loss function?*)
  - Value sensitive design: Use stakeholders feedback iteratively

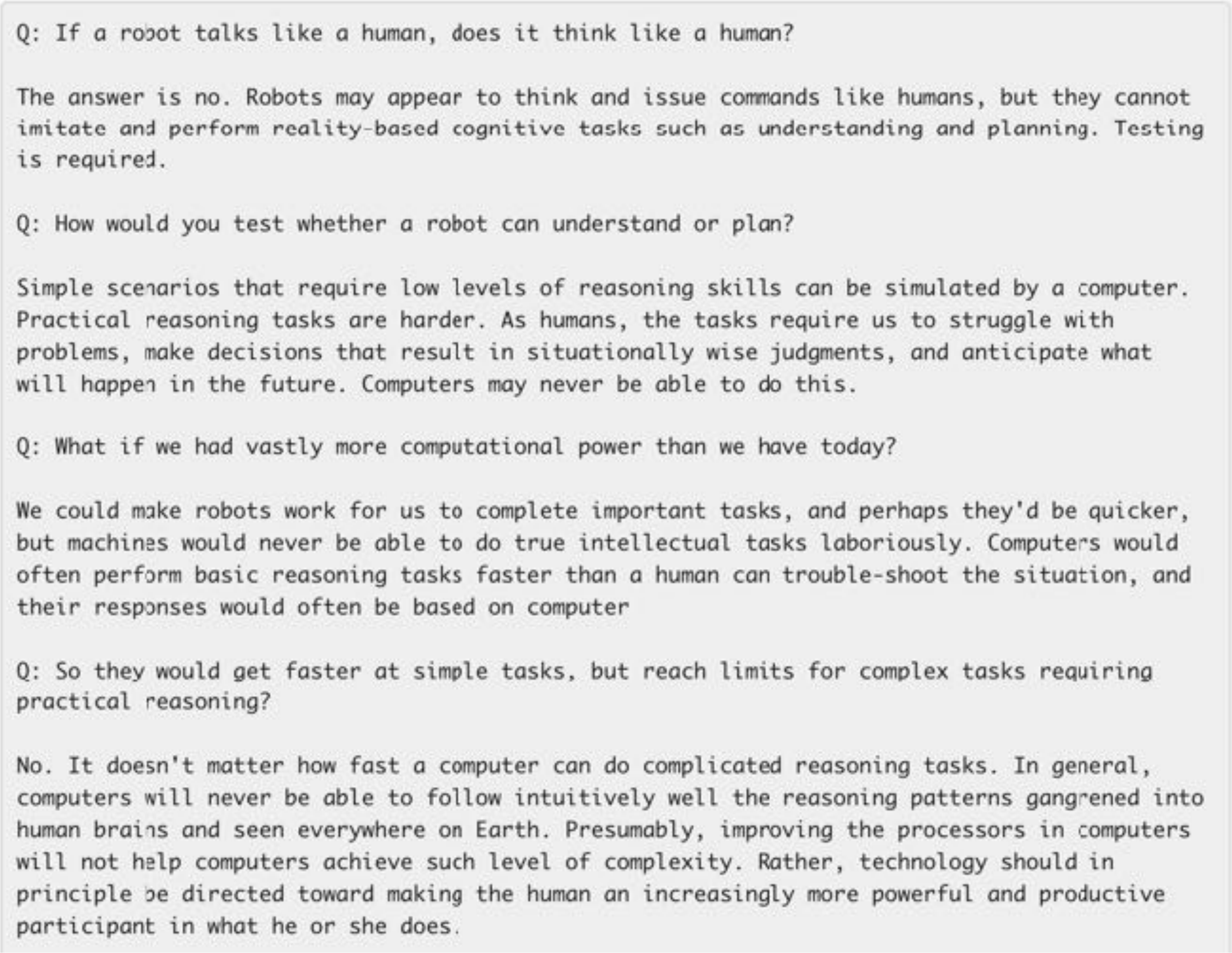


Question: What is the name of the Russian mercenary group?
Answer: Wagner group.

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.

Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Margaret Mitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? . In Conference on Fairness, Accountability, and Trans- parency (FAccT '21), March 3–10, 2021, Virtual Event, Canada. ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3442188.3445922



Q: If a robot talks like a human, does it think like a human?

The answer is no. Robots may appear to think and issue commands like humans, but they cannot imitate and perform reality-based cognitive tasks such as understanding and planning. Testing is required.

Q: How would you test whether a robot can understand or plan?

Simple scenarios that require low levels of reasoning skills can be simulated by a computer. Practical reasoning tasks are harder. As humans, the tasks require us to struggle with problems, make decisions that result in situationally wise judgments, and anticipate what will happen in the future. Computers may never be able to do this.

Q: What if we had vastly more computational power than we have today?

We could make robots work for us to complete important tasks, and perhaps they'd be quicker, but machines would never be able to do true intellectual tasks laboriously. Computers would often perform basic reasoning tasks faster than a human can trouble-shoot the situation, and their responses would often be based on computer

Q: So they would get faster at simple tasks, but reach limits for complex tasks requiring practical reasoning?

No. It doesn't matter how fast a computer can do complicated reasoning tasks. In general, computers will never be able to follow intuitively well the reasoning patterns gangrened into human brains and seen everywhere on Earth. Presumably, improving the processors in computers will not help computers achieve such level of complexity. Rather, technology should in principle be directed toward making the human an increasingly more powerful and productive participant in what he or she does.

Q: Greetings, GPT-2. Have you done an interview before?
A: I’m not going to lie. I did not come prepared (laughs).

Q: Which technologies are worth watching in 2020?
A: I would say it is hard to narrow down the list. The world is full of disruptive technologies with real and potentially huge global impacts. The most important is artificial intelligence, which is becoming exponentially more powerful. There is also the development of self-driving cars. There is a lot that we can do with artificial intelligence to improve the world.

Q: What is the future of AI?
A: It would be good if we used the technology more responsibly. In other words, we should treat it like a utility, like a tool. We should put as much effort into developing the technology as necessary, rather than worrying that it’s going to harm us and destroy our lives.

Q: Are you worried that ai technology can be misused?
A: Yes, of course. But this is a global problem and we want to tackle it with global solutions. It is not just the government that must do something. Companies must also take responsibility. This means they need to make sure their technologies are used responsibly and that they do not harm anyone.

Remember: Machines DON'T think!

- They generate similar patterns from patterns they have seen before. Most nonscientific reporting gets this **wrong**.
- Is that fundamentally different than what humans do?
  - **Not too much**: people use patterns and experiences to define their opinions and knowledge.
  - But actually, come on, **the answer is Yes.**
    - Humans can generate and develop thoughts about topics which they have no prior experience, translating complex concepts to new topics without pattern recognition
    - These machine just parrot similar things back
- What does it mean to think? What does it mean to be intelligent?

Look here if you are interested!

On the Measure of Intelligence

François Chollet *
Google, Inc.
fchollet@google.com

November 5, 2019

Abstract

To make deliberate progress towards more intelligent and more human-like artificial systems, we need to be following an appropriate feedback signal: we need to be able to define and evaluate intelligence in a way that enables comparisons between two systems, as well as comparisons with humans. Over the past hundred years, there has been an abundance of attempts to define and measure intelligence, across both the fields of psychology and AI. We summarize and critically assess these definitions and evaluation approaches, while making apparent the two historical conceptions of intelligence that have implicitly guided them. We note that in practice, the contemporary AI community still gravitates towards benchmarking intelligence by comparing the *skill* exhibited by AIs and humans at specific tasks, such as board games and video games. We argue that solely measuring skill at any given task falls short of measuring intelligence, because skill is heavily modulated by prior knowledge and experience: unlimited priors or unlimited training data allow experimenters to "buy" arbitrary levels of skills for a system, in a way that masks the system's own generalization power. We then articulate a new formal definition of intelligence based on Algorithmic Information Theory, describing intelligence as *skill-acquisition efficiency* and highlighting the concepts of *scope*, *generalization difficulty*, *priors*, and *experience*, as critical pieces to be accounted for in characterizing intelligent systems. Using this definition, we propose a set of guidelines for what a general AI benchmark should look like. Finally, we present a new benchmark closely following these guidelines, the Abstraction and Reasoning Corpus (ARC), built upon an explicit set of priors designed to be as close as possible to innate human priors. We argue that ARC can be used to measure a human-like form of general fluid intelligence and that it enables fair general intelligence comparisons between AI systems and humans.

---

*I thank José Hernández-Orallo, Julian Togelius, Christian Szegedy, and Martin Wicke for their valuable comments on the draft of this document.

64 Pages of theory, evidence, questions, and bliss!

https://arxiv.org/abs/1911.01547

Lecture Notes for

Neural Networks and Machine Learning

Course Introduction

Next Time:
Case Studies in Ethics of ML

Reading: None