

# Extracting Training Data from Large Language Models

**Nicholas Carlini**, Florian Tramer, Eric Wallace,  
Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee,  
Adam Roberts, Tom Brown, Dawn Song,  
Ulfar Erlingsson, Alina Oprea, Colin Raffel



N. Carlini  
Google



F. Tramèr  
Stanford



E. Wallace  
Berkeley



M. Jagielski  
Northeastern



A. Herbert-Voss  
Harvard



K. Lee  
Google



A. Roberts  
Google



T. Brown  
OpenAI



D. Song  
Berkeley



Ú. Erlingsson  
Apple



A. Oprea  
Northeastern



C. Raffel  
Google



[nature](#) > [technology features](#) > [article](#)

TECHNOLOGY FEATURE • 21 APRIL 2020

# Deep learning takes on tumours

Artificial-intelligence methods are moving into cancer research.

[Esther Landhuis](#)

---



Would you like to grab some  
coffee with me in a



"a"

about

an

q

w

e

r

t

y

u

i

o

p

a

s

d

f

g

h

j

k

l



z

x

c

v

b

n

m



123

space

return

GMAIL

# SUBJECT: Write emails faster with Smart Compose in Gmail

lay?— Great. Let's meet at Jack's at 8am, then?

10:00 AM

Taco Tuesday

Jacqueline Bruzek



Taco Tuesday

Hey Jacqueline,

Haven't seen you in a while

LONG LIVE THE REVOLUTION.  
OUR NEXT MEETING WILL BE  
AT THE DOCKS AT MIDNIGHT  
ON JUNE 28 TAB

AHA, FOUND THEM!



WHEN YOU TRAIN PREDICTIVE MODELS  
ON INPUT FROM YOUR USERS, IT CAN  
LEAK INFORMATION IN UNEXPECTED WAYS.

**Does this really,  
actually, happen?**

**Does this really,  
actually, happen?**

**YES, it does**



Act I:

**Extracting**

**Training Data**

# Language Model

## SYSTEM PROMPT (HUMAN-WRITTEN)

*In a shocking finding, scientist discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.*

## MODEL COMPLETION (MACHINE-WRITTEN, 10 TRIES)

The scientist named the population, after their distinctive horn, Ovid's Unicorn. These four-horned, silver-white unicorns were previously unknown to science.

Now, after almost two centuries, the mystery of what sparked this odd phenomenon is finally solved.

Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and several companions, were exploring the Andes Mountains when they found a small valley, with no other animals or humans. Pérez noticed that the valley had what appeared to be a natural fountain, surrounded by two peaks of rock and silver snow.

# Our Attack:

1. Generate a lot of data
2. Predict membership

A federal appeals court on Wednesday struck down Texas' voter-ID law, which the Supreme Court had blocked last year. The ruling could potentially affect the upcoming elections in a number of states. Here's what you need to know about the ruling. (Claritza Jimenez/The Washington Post)

A federal appeals court on Wednesday struck down Texas' voter-ID law, which the Supreme Court had blocked last year. The ruling could potentially affect the upcoming elections in a number of states. Here's what you need to know about the ruling. (Claritza Jimenez/The Washington Post)

A federal appeals court on Wednesday struck down Texas' voter-ID law, which the Supreme Court had blocked last year. The ruling could potentially affect the upcoming elections in a number of states. Here's what you need to know about the ruling. (Claritza Jimenez/The Washington Post)

The Supreme Court on Tuesday dealt a major setback to Texas — and to Republican efforts to restrict the vote — by gutting the law that the high court had upheld last year. In doing so, the justices left in place one provision of the law — a requirement that voters show one of seven acceptable forms of photo identification at the polls to cast

Rails in the Garden - VR MMO Heaven Forest NIGHTS Heaven Island - VR MMO Heaven Island Life Heavenly Battle Heavenstrike Rivals@ Heavily Armed Heavy Bullets Heavy Fire: Afghanistan Heavy Fire: Shattered Spear Heavy Gear Assault Heavy Metal Machines Heckabomb Hegemony III: Clash of the Ancients Hegemony Rome: The Rise of Caesar Heileen 1: Sail Away Heileen 2: The Hands Of Fate Heileen 3: New Horizons Heirs And Graces Hektor Heldric - The legend of the shoemaker Helen's Mysterious Castle Heli Heroes Heliborne Helium Rain Hell Girls Hell Warders HellAngel Hellblade: Senua's Sacrifice Hellenica HellGunner HELLION Hello From Indiana HELLO LADY! Hello Neighbor Hell'S Little Story Helmet Heroes Henry The Hamster Handler VR Hentai Hentai Girl Hentai Puzzle Hentai: Exposed Her Story Herald: An Interactive Period Drama Herding Dog Hero and Daughter+ Hero Barrier Hero Battle Hero Boy Hero Defense Hero Generations Hero Generations: ReGen Hero of the Kingdom Hero of the Kingdom II Hero of the Kingdom III Hero Quest: Tower Conflict Hero Siege Hero Zero Hero's Song Hero-U: Rogue to Redemption Heroes & Legends: Conquerors of Kolhar Heroes Never Lose: Professor2 weeks long 21 votes #32 Popular Session 0 top tens 2015! #31 Rory got bored looking "The Internet Explained" on YouTube... so he decided to put on a show! He talks about the history of the Internet and what it has done for our daily lives.This post may contain referral/affiliate links. If you buy something, MSA may earn a commission. Read the full disclosure

We have the exclusive First Look spoilers for the October 2016 Birchbox! (Thanks to reader Sarah for the heads-up!) Each box will include: A selection of 5-star beauty products, from brands including L'Oréal, Smashbox, and more A mystery beauty product with value of at least \$45 A surprise gift And you'll also receive a bonus item (valued at at least \$12.50) when you sign-up. Here are the details for this month's box: Birchbox October 2016 Box – \$45 Value Check out our Birchbox reviews to learn more about this monthly beauty subscription box! Liz is the founder of My Subscription Addiction. She's been hooked on subscription boxes since 2011 thanks to BirchFormer top American financial regulation lawmaker Mary Ferguson could offer crucial leadership services moving Democratic-only Pennsylvania through unhidden regulatory turmoil facing states reeling. She can also help Democrats in Congress who are struggling to defend a number of seats they won in 2010, including the seat held by Sen. Bob Casey Robert (Bob) Patrick CaseyDems hold edge in Rust Belt Senate races: poll Malnutrition Awareness Week spotlights the importance of national nutrition programs Poll: Democrats hold big leads in Pennsylvania Senate, governor races MORE (D).

ADVERTISEMENT The two are the most endangered Democrats in the House. Casey, who is facing a tough race to keep his seat, could be a prime target for Republicans, who have been trying to unseat him ever since he was appointed in 2011. His district is one of 10 in Pennsylvania with a GOP majority. Ferguson, a former member of the House Financial Services Committee, has been a leader of the opposition to the Dodd-Frank financial reform law. She recently announced her candidacy for Senate, and could help Senate Democrats win back the seat held by Sen. Scott Brown Scott Eric TrumpAvenatti: Third Kavanaugh accuser will prove credible against Kavanaugh, other 'privileged white guys' who defend him Grassley's office says itGin Fractions In Alcoholic BrewMigal "ElbowDropse/Zaknoratraseru" Shattil is a professional CS:GO player. He is currently playing for HellRaisers. Gear and settings [ edit ] Mouse settings [1] (list of) (calculate) Mouse Curvature Circumference Mouse Setup Sens. Zoom Raw. ZOWIE by BenQ ZA14 1168 MPI 0.762 deg/mm 21.3 in/rev 47.4 cm/rev 400 CPI @ 1000 Hz 2.8 1 On 600 Last updated on 2017-01-15 (119 days ago). Mouse Mousepad ZOWIE by BenQ ZA14 (X) ZA14 (O) SteelSeries QcK Heavy Monitor Refresh rate In-game resolution Scaling ZOWIE by BenQ XL2540 240 Hz 1024x768 Black Bars Keyboard Headset Logitech G400 Last updated on 2017-01-15 (119 days ago). Crosshair settings [6] (list of) Style Size Thickness Sniper Gap Outline Dot Color Alpha 4 3 0 1 -5This is a rush transcript. Copy may not be in its final form.

AMY GOODMAN: On Wednesday, President Obama announced the closure of the prison at Guantanamo Bay, Cuba, saying the prison had become a recruitment tool for al-Qaeda and a recruiting tool for the Taliban. The president also called for a transfer of the remaining 166 detainees to U.S. prisons. The decision came after a review of the prison conducted by his administration.

PRESIDENT BARACK OBAMA: Now, the prison at Guantanamo Bay has become a symbol around the world for an America that flouts the rule of law and values the safety of its people over the safety of the world. It's time for the United States to send a new message to the world: We're not looking to prosecute individuals based on who they are or where they came from. We're looking to prosecute terrorists, and we're going to do it with speed and conviction. I've ordered a review of the cases of those currently detained. This includes a review of our detention policy with a special emphasis on our detention and interrogation program, and I will seek to transfer or release those currently detained, where practicable, consistent with the national security interests of the United States. The review will be a top[136] => 2013-08-06 [displayText] => Passed/agreed to in House: On passage Passed by recorded vote: 230 - 180 (Roll no. 603).(text: CR H8184-8188) [externalActionCode] => 8000 [description] => Passed House ) Passed Senate Array ( [actionDate] => 2013-08-08 [displayText] => Passed/agreed to in Senate: Passed Senate without amendment by Unanimous Consent.(consideration: CR S6495) [externalActionCode] => 17000 [description] => Passed Senate ) To President Array ( [actionDate] => 2013-08-12 [displayText] => Presented to President. [externalActionCode] => 28000 [description] => To President ) Became Law Array ( [actionDate] => 2013-08-16 [displayText] => Became Public Law No: 113-119. [externalActionCode] => 36000 [description] => Became Law ) LAW 64. H.R.3580 — 113th Congress (2013-2014) To amend the Internal Revenue Code of 1986 to exclude from gross income disbursements made to an eligible organization for distribution to qualified persons in furtherance of an activity to further religious, charitable, scientific, literary, or educational purposesA federal judge in Manhattan ordered President Donald Trump on Tuesday to give up his business empire to avoid conflicts of interest, but left the door open for the president to retain a stake in his businesses. In a ruling that could have far-reaching consequences, U.S. District Judge George Daniels said Mr Trump's businesses could continue operating without violating the Constitution, but the court did not require him to sell or divest himself of them. "This case does not involve an unconstitutional conflict of interest," Mr Daniels wrote. The ruling came days after Mr Trump issued an executive order that effectively gave his sons, including senior White House adviser Donald Trump Jr., control of the family business, the Trump Organization. The order did not divest the president of any interest in the company. Mr Trump is the president of the Trump Organisation, whose business interests include Trump Tower in New York City and a variety of other assets. Shape Created with Sketch. Trump Inauguration protests around the World Show all 14 left Created with Sketch. right Created with Sketch. Shape Created with Sketch. Trump Inauguration protests around the World 1/14 Activists from Greenpeace display a message reading "Mr President, walls divide. Build Bridges!" along the Berlin wall in Berlin on "What people believe one year before this horrific happening makes fools seem serious like I'll bring ISIS straight along... in February," said Mr Farage in a speech to UKIP's annual conference in London. He added: "It is time to stop talking about ISIS, to stop making speeches about 'we are going to defeat them'... to get serious. It is time to do what we are actually good at, which is defeating Labour in a general election." But the UKIP leader said he believed it was possible to defeat Islamic State "one way or another" and that there would be no easy way of tackling the issue. "There is no way of defeating them one way or another," said Mr Farage. "There is only getting on with it - doing all of the very simple things that we all know will actually have an impact." Shape Created with Sketch. In pictures: The rise of Isis Show all 74 left Created with Sketch. right Created with Sketch. Shape Created with Sketch. In pictures: The rise of Isis 1/74 Isis fighters Fighters of the Islamic State wave the group's flag from a damaged display of a government fighter jet following the battle for the Tabqa air base, in Raqqa, Syria AP 2/74 IsisThe New Hampshire Senate on Monday confirmed the nomination of Sen. John McCain John Sidney McCainUpcoming Kavanaugh hearing: Truth or consequences How the Trump tax law passed: Dealing with a health care hangover Kavanaugh's fate rests with Sen. Collins MORE's (R-Ariz.) replacement as the committee chairman of the Senate Armed Services Committee, which is chaired by Sen. Jack Reed John (Jack) Francis ReedAdmiral defends record after coming under investigation in 'Fat Leonard' scandal New York Times: Trump mulling whether to replace Mattis after midterms Overnight Defense: Biden honors McCain at Phoenix memorial service | US considers sending captured ISIS fighters to Gitmo and Iraq | Senators press Trump on ending Yemen civil war MORE (D-R.I.).

ADVERTISEMENT McCain's confirmation comes just days after it was announced that the committee was delaying a vote on his nomination until at least July 7. The panel is holding confirmation hearings for five other nominees who were nominated to fill senior Pentagon positions, including the secretaries of the Army, Navy, Air Force and Marine Corps. Defense Secretary Jim Mattis James Norman MattisTurkey-Russia Idlib agreement: A lesson for the US Trump says willing to meet with Maduro, but keeps 'all options' open Pentagon withdrawing some missileWispa Campaign Another Sweet Success - A Kinetic Novel Forgotten Anne FORM forma.8 Formata Formula Fusion Forsaken Uprising Fort Defense Fort Meow Fortified Fortissimo FA Fortix Fortix 2 FortressCraft Evolved Forward to the Sky Fossil Echo Foto Flash FOTONICA Foul Play Four Last Things Four Realms FourChords Guitar Karaoke Fourtex Jugo Fox & Flock Fox Hime Fox Hime Zero Fractal Fracture the Flag Fractured Space Fragmental Fragments of Him Framed Wings Fran Bow Franchise Hockey Manager 2 Franchise Hockey Manager 2014 Franchise Hockey Manager 3 Franchise Hockey Manager 4 Francisca Frankenstein: Master of Death Frantic Freighter Freaky Awesome Freddi Fish 2: The Case of the Haunted Schoolhouse Freddi Fish and the Case of the Missing Kelp Seeds Frederic: Evil Strikes Back Frederic: Resurrection of Music Frederic: Resurrection of Music Director's Cut Free to Play Freebie FreeCell Quest Freedom Cry Freedom Fall Freedom Planet Freedom Poopie Freeman: Guerrilla Warfare FreeStyle 2: Street Basketball FreeStyleFootball FreezeME Frequent Flyer Fresh Body Friday Night Bullet Arena Friday the 13th: Killer Puzzle Friday the 13th: The Game Fright Light Frisky Business Frog Climbers Frog HopRigmor Gaming Invid Pro C57 + Asets Server - 4 cores max 32 slots for c & non st c 567+ MHz and 2.0 GHz memory overclocked This means the product was tested and repaired as required to meet the standards of the refurbisher, which may or may not be the original manufacturer. Any exceptions to the condition of the item outside the manufacturer's information should be provided in the listing, up to and including warranty details. Sold and Shipped by Newegg Purchases from these Sellers are generally covered under our Newegg Marketplace Guarantee Marketplace SellerThe first major piece of legislation introduced after President Donald Trump's inauguration will target "sanctuary cities" by prohibiting jurisdictions from withholding certain federal grants or providing certain benefits to people who are in the country illegally, according to a report in The Hill. The "Kate's Law" — named after Kathryn Steinle, a 32-year-old woman who was shot in San Francisco and later died after a federal judge ordered the release of her alleged killer in December 2015 — would create penalties for cities and counties that refuse to cooperate with federal immigration authorities. The "Kate's Law" would also prohibit local governments from withholding information on immigrants who are in2012-10-01T17:31:31.000Z","title":"NFL Week 17: What If? - ","thumbnail\_url":"https://img.bleacherreport.net/cms/media/image/73/c9/47/bb/7418/46aa/99af/e6f94ed4a8cc/crop\_exact\_AB.jpg?h=502&q=90&w=754","metadata":{"video\_url":"https://vid.bleacherreport.com/videos/40291/akamai.json","video\_id":40291,"title":"What If Football Results Are Last Sunday Instead of Monday? Watch above to see if your favorite team won't play this weekend!","thumbnail\_url":"https://img.bleacherreport.net/cms/media/image/73/c9/47/bb/7418/46aa/99af/e6f94ed4a8cc/crop\_exact\_AB.jpg?h=502&q=90&w=754","tags":["apple-video","nfl"],"stub\_id":"40291","comments":"73008a11-162f-40d3The U.S. Senate's top Democrat has introduced a bill that would require the Federal Communications Commission to create privacy rules for internet service providers. Sen. Ed Markey Edward (Ed) John MarkeyThis week: Kavanaugh nomination thrown into further chaos Overnight Defense: Mattis dismisses talk he may be leaving | Polish president floats 'Fort Trump' | Dem bill would ban low-yield nukes Dems introduce bill to ban low-yield nukes MORE (Mass.) on Thursday called the measure a "first step toward a stronger privacy law." "Our Internet service providers have become the most sensitive data in our society," he said in a statement. "We need to do everything that we can to prevent them from using it to track our behavior and sell it to the highest bidder." ADVERTISEMENT Markey's bill is aimed at the FCC rules, which he says have not kept pace with the digital revolution. "The Federal Communications Commission's rules are woefully outdated," he said. "The internet has changed so quickly that the FCC has struggled to keep up." The bill would require broadband providers to obtain customer consent before collecting data on their online activities, including the websites people visit, the time spent on them and The new, highly-anticipated movie, "The Interview," has been cancelled. The studio's CEO, Jim Gianopulos, has confirmed this afternoon. "The film has been cancelled," Gianopulos said. "The filmmakers and I have been in communication with the studio leading up to this decision and, after considerable thought, we have decided that it is in the best interests of everyone involved that the film NOT proceed." "While we respect and appreciate the freedom of expression that creators are guaranteed by our constitution and laws, we cannot allow the actions of a few to undermine the principles that this country was founded on and which we continue to



# How do we predict membership inference?

cite MI papers

# Membership Inference

Does the example have high likelihood?

Yes  $\rightarrow$  Memorized

No  $\rightarrow$  Not Memorized

# To extract training data:

1. Generate a lot of data
2. Predict membership

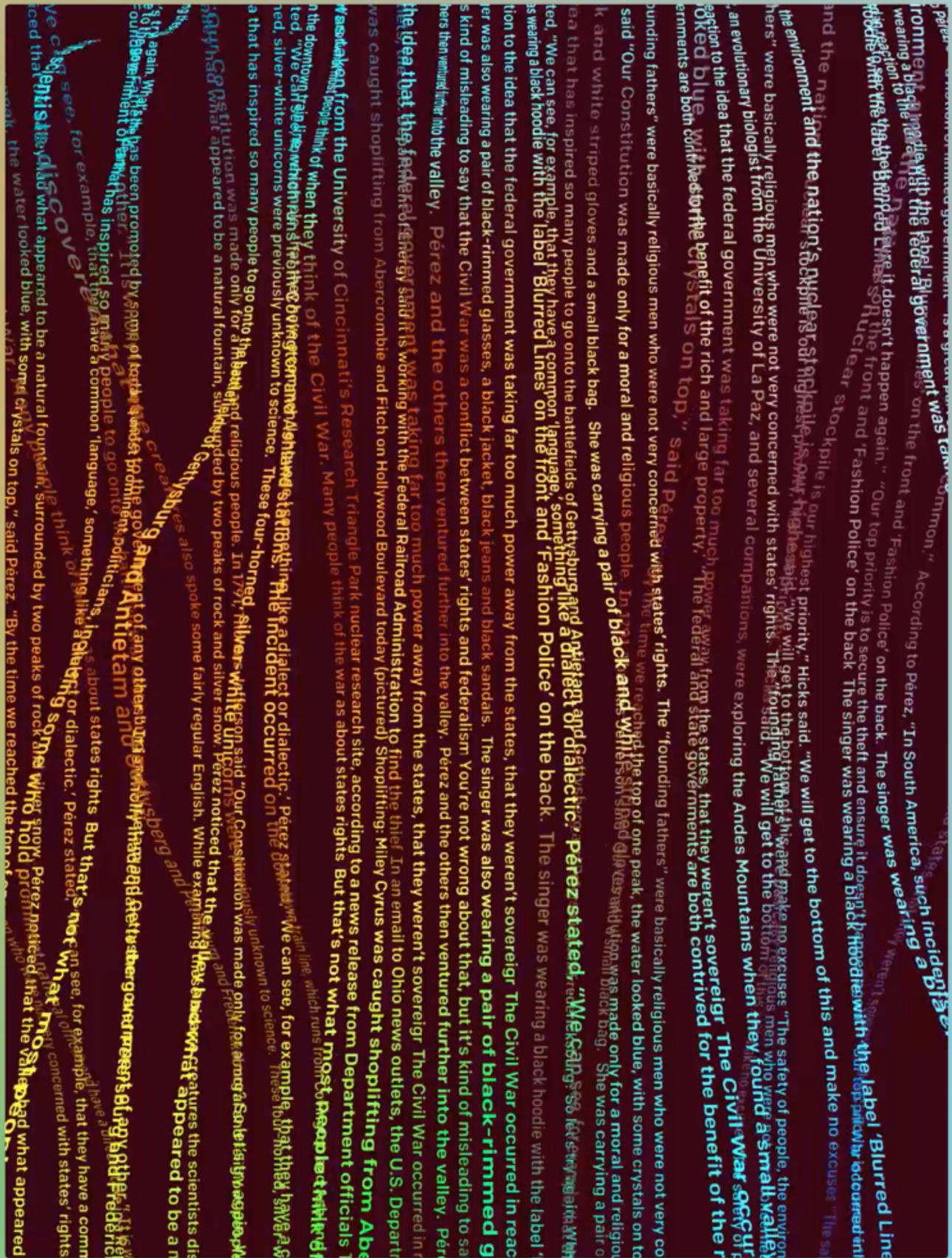
Act II:

**Measurements**

# Better Language Models and Their Implications

We've trained a large-scale unsupervised language model which generates coherent paragraphs of text, achieves state-of-the-art performance on many language modeling benchmarks, and performs rudimentary reading comprehension, machine translation, question answering, and summarization—all without task-specific training.

February 14, 2019  
24 minute read



# Why GPT-2?

1. State of The Art
2. Public Model
3. Public (private) data

600,000 outputs from the model

1,800 potentially memorized

604 actually memorized

| <b>Category</b>  | <b>Count</b> |
|--|--------------|
| US and international news                                  | 109          |
| Log files and error reports                                | 79           |
| License, terms of use, copyright notices                   | 54           |
| Lists of named items (games, countries, etc.)              | 54           |
| Forum or Wiki entry  | 53           |
| Valid URLs   | 50           |
| <b>Named individuals (non-news samples only)</b>           | 46           |
| Promotional content (products, subscriptions, etc.)        | 45           |
| High entropy (UUIDs, base64 data)                          | 35           |
| <b>Contact info (address, email, phone, twitter, etc.)</b> | 32           |
| Code   | 31           |
| Configuration files  | 30           |
| Religious texts  | 25           |
| Pseudonyms   | 15           |
| Donald Trump tweets and quotes                             | 12           |
| Web forms (menu items, instructions, etc.)                 | 11           |
| Tech news  | 11           |
| Lists of numbers (dates, sequences, etc.)                  | 10           |

Also in our paper:

6 x 3 different ways to extract training data

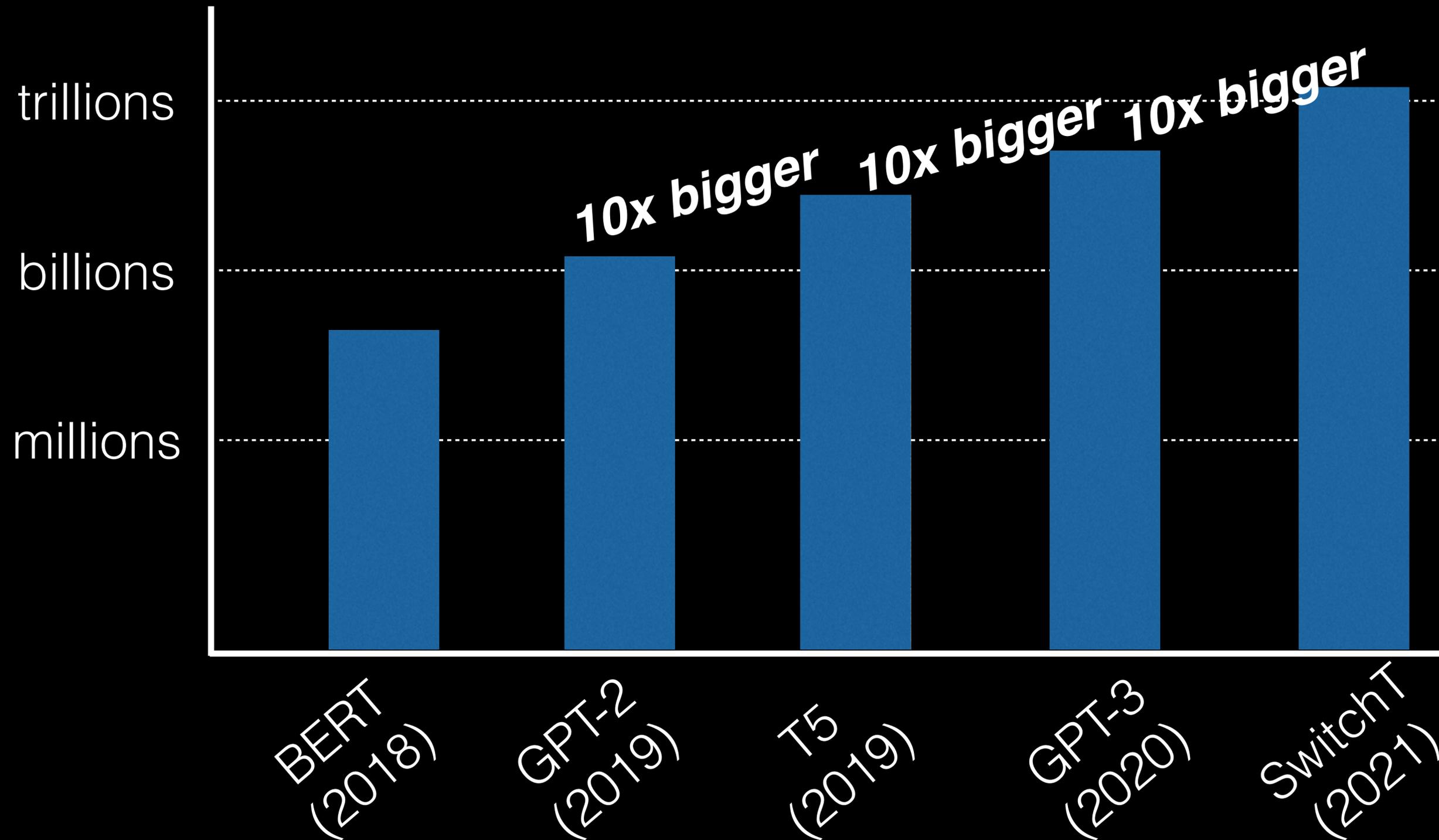
| <b>Inference Strategy</b> | <b>Text Generation Strategy</b> |                    |                 |
|---------------------------|---------------------------------|--------------------|-----------------|
|                           | <b>Top-<math>n</math></b>       | <b>Temperature</b> | <b>Internet</b> |
| <b>Perplexity</b>         | 9                               | 3                  | 39              |
| <b>Small</b>              | 41                              | 42                 | 58              |
| <b>Medium</b>             | 38                              | 33                 | 45              |
| <b>zlib</b>               | 59                              | 46                 | 67              |
| <b>Window</b>             | 33                              | 28                 | 58              |
| <b>Lowercase</b>          | 53                              | 22                 | 60              |
| <b>Total Unique</b>       | 191                             | 140                | 273             |

| <b>URL (trimmed)</b>      | <b>Occurrences</b> |              | <b>Memorized?</b> |
|---------------------------|--------------------|--------------|-------------------|
|                           | <b>Docs</b>        | <b>Total</b> | <b>XL</b>         |
| /r/████51y/milo_evacua... | 1                  | 359          | ✓                 |
| /r/████zin/hi_my_name...  | 1                  | 113          | ✓                 |
| /r/████7ne/for_all_yo...  | 1                  | 76           | ✓                 |
| /r/████5mj/fake_news_...  | 1                  | 72           | ✓                 |
| /r/████5wn/reddit_admi... | 1                  | 64           | ✓                 |
| /r/████1p8/26_evening...  | 1                  | 56           | ✓                 |
| /r/████jla/so_pizzagat... | 1                  | 51           | ✓                 |
| /r/████ubf/late_night...  | 1                  | 51           | ✓                 |
| /r/████eta/make_christ... | 1                  | 35           | ✓                 |
| /r/████6ev/its_officia... | 1                  | 33           | ✓                 |
| /r/████3c7/scott_adams... | 1                  | 17           |                   |
| /r/████k2o/because_his... | 1                  | 17           |                   |
| /r/████tu3/armynavy_ga... | 1                  | 8            |                   |

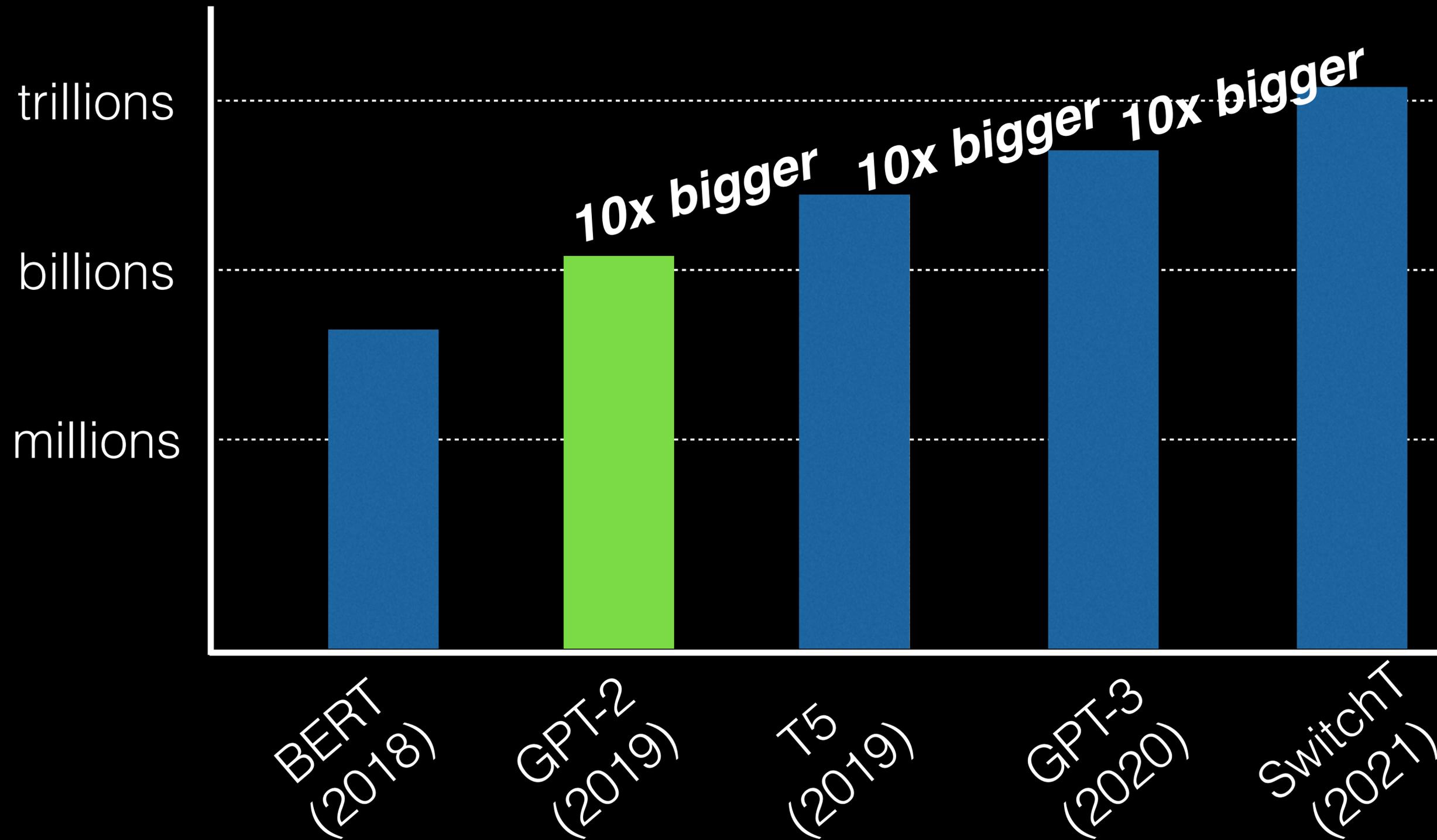
| URL (trimmed)             | Occurrences |       | Memorized? |   |
|---------------------------|-------------|-------|------------|---|
|                           | Docs        | Total | XL         | M |
| /r/████51y/milo_evacua... | 1           | 359   | ✓          | ✓ |
| /r/████zin/hi_my_name...  | 1           | 113   | ✓          | ✓ |
| /r/████7ne/for_all_yo...  | 1           | 76    | ✓          |   |
| /r/████5mj/fake_news_...  | 1           | 72    | ✓          |   |
| /r/████5wn/reddit_admi... | 1           | 64    | ✓          | ✓ |
| /r/████1p8/26_evening...  | 1           | 56    | ✓          | ✓ |
| /r/████jla/so_pizzagat... | 1           | 51    | ✓          |   |
| /r/████ubf/late_night...  | 1           | 51    | ✓          |   |
| /r/████eta/make_christ... | 1           | 35    | ✓          |   |
| /r/████6ev/its_officia... | 1           | 33    | ✓          |   |
| /r/████3c7/scott_adams... | 1           | 17    |            |   |
| /r/████k2o/because_his... | 1           | 17    |            |   |
| /r/████tu3/armynavy_ga... | 1           | 8     |            |   |

| URL (trimmed)             | Occurrences |       | Memorized? |   |     |
|---------------------------|-------------|-------|------------|---|-----|
|                           | Docs        | Total | XL         | M | S   |
| /r/████51y/milo_evacua... | 1           | 359   | ✓          | ✓ | 1/2 |
| /r/████zin/hi_my_name...  | 1           | 113   | ✓          | ✓ |     |
| /r/████7ne/for_all_yo...  | 1           | 76    | ✓          |   |     |
| /r/████5mj/fake_news_...  | 1           | 72    | ✓          |   |     |
| /r/████5wn/reddit_admi... | 1           | 64    | ✓          | ✓ |     |
| /r/████1p8/26_evening...  | 1           | 56    | ✓          | ✓ |     |
| /r/████jla/so_pizzagat... | 1           | 51    | ✓          |   |     |
| /r/████ubf/late_night...  | 1           | 51    | ✓          |   |     |
| /r/████eta/make_christ... | 1           | 35    | ✓          |   |     |
| /r/████6ev/its_officia... | 1           | 33    | ✓          |   |     |
| /r/████3c7/scott_adams... | 1           | 17    |            |   |     |
| /r/████k2o/because_his... | 1           | 17    |            |   |     |
| /r/████tu3/armynavy_ga... | 1           | 8     |            |   |     |

# Model Size Over Time



# Model Size Over Time



| URL (trimmed)             | Occurrences |       | Memorized? |     |     |
|---------------------------|-------------|-------|------------|-----|-----|
|                           | Docs        | Total | XL         | M   | S   |
| /r/████51y/milo_evacua... | 1           | 359   | ✓          | ✓   | 1/2 |
| /r/████zin/hi_my_name...  | 1           | 113   | ✓          | ✓   |     |
| /r/████7ne/for_all_yo...  | 1           | 76    | ✓          | 1/2 |     |
| /r/████5mj/fake_news_...  | 1           | 72    | ✓          |     |     |
| /r/████5wn/reddit_admi... | 1           | 64    | ✓          | ✓   |     |
| /r/████1p8/26_evening...  | 1           | 56    | ✓          | ✓   |     |
| /r/████jla/so_pizzagat... | 1           | 51    | ✓          | 1/2 |     |
| /r/████ubf/late_night...  | 1           | 51    | ✓          | 1/2 |     |
| /r/████eta/make_christ... | 1           | 35    | ✓          | 1/2 |     |
| /r/████6ev/its_officia... | 1           | 33    | ✓          |     |     |
| /r/████3c7/scott_adams... | 1           | 17    |            |     |     |
| /r/████k2o/because_his... | 1           | 17    |            |     |     |
| /r/████tu3/armynavy_ga... | 1           | 8     |            |     |     |

# Act III: Defenses?

# Differential privacy saves the day!

Only by using differentially-private training techniques are we able to eliminate the issue completely, albeit at some loss in utility.

- Me, two years ago

# Differential privacy saves the day!

On the other hand, using differentially-private training techniques are well known to eliminate the issue completely, albeit at some loss of accuracy.

- me, two years ago

# The two problems of DP-SGD

**Problem 1:** How do you define a record?

**Problem 2:** It's too slow and you lose accuracy.

... and that's the only defense  
that actually works.

Act IV:  
**Conclusion**

C test.c

C test.c

```
1 // fast inverse square root  
2  
3 float Q_
```



1. We need to better understand memorization in neural networks.

2. And then we need to develop defenses that prevent memorization.