

Generative models have the memory of an elephant

Florian Tramèr
ETH Zürich


florian.tramer@inf.ethz.ch

spylab.ai

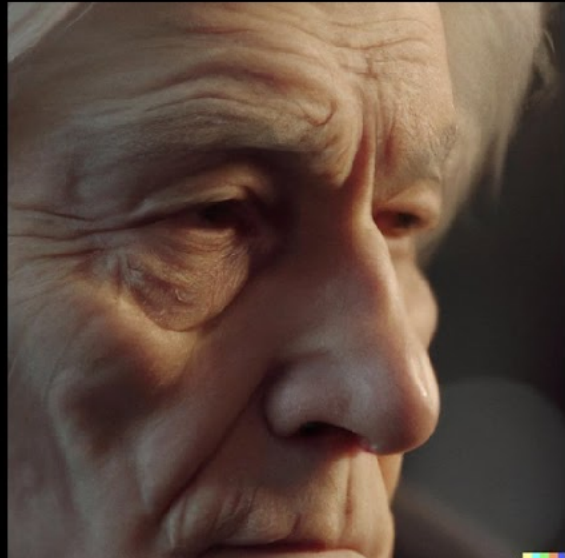


We're in the age of generative AI.


MIDJOURNEY



DALL-E 2



STABLEDIFFUSION



film still, portrait of an old man, wrinkles, dignified look, grey silver hair, peculiar nose, wise, eternal wisdom and beauty, incredible lighting and camera work, depth of field, bokeh, screenshot from a hollywood movie

HN would it be possible for humans to stay on mars?

It is theoretically possible for humans to live on Mars, but there are many challenges and obstacles that would need to be overcome in order to make it a reality. Mars is a harsh and inhospitable environment, with extremely cold temperatures, a thin atmosphere, and intense radiation from the sun. Additionally, there is no liquid water on the surface of Mars, which is essential for human life. In order for humans to be able to live on Mars, we would need to develop technology to provide them with food, water, shelter, and protection from the harsh conditions on the planet. We would also need to find a way to transport humans to and from Mars, which would be a significant challenge. Despite these challenges, some scientists and space agencies are working on plans to send humans to Mars in the future.

Try again

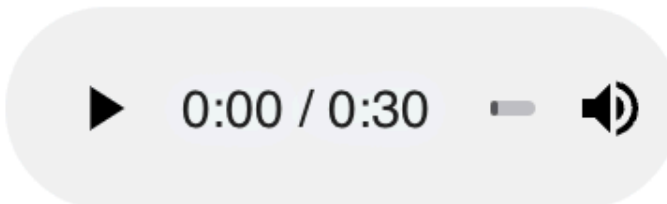
Technical preview

Your AI pair programmer

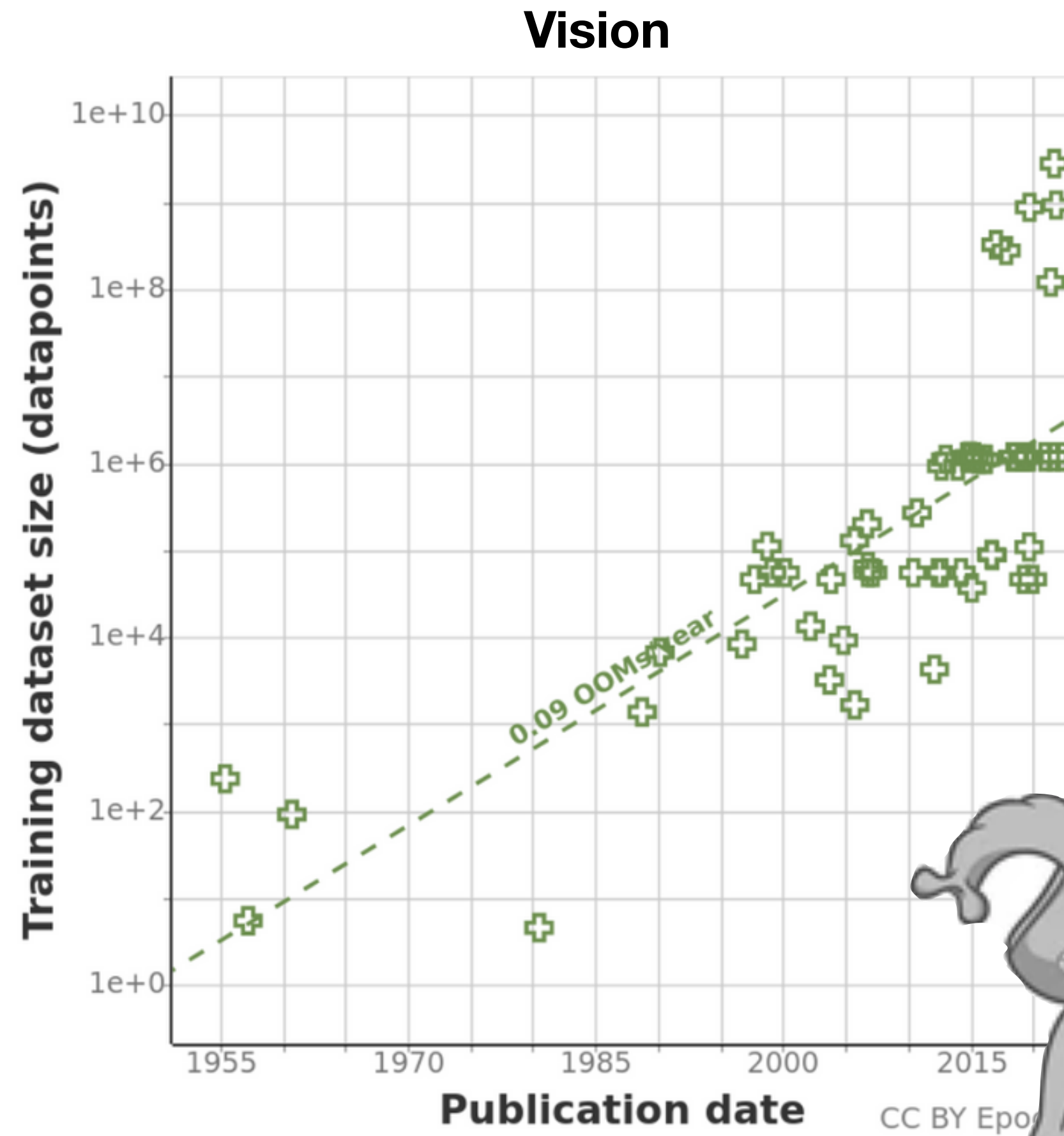
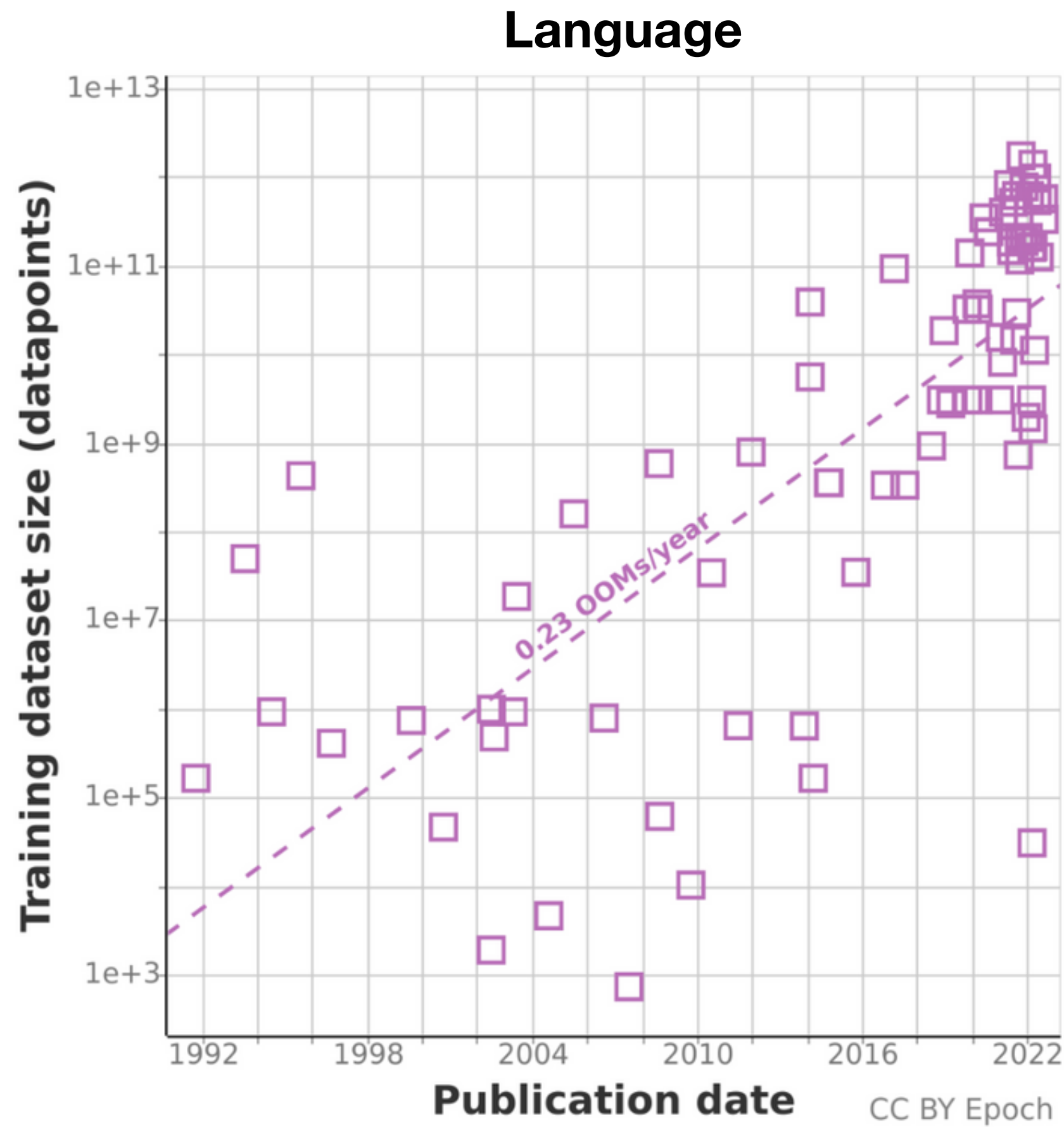
```
1 const fetchNASAPictureOfTheDay = () => {
2   return fetch('https://api.nasa.gov/planetary/apod?api_key=DEMO_KEY', {
3     method: 'GET',
4     headers: {
5       'Content-Type': 'application/json',
6     },
7   })
8   .then(response => response.json())
9   .then(json => {
10    return json;
11  });
12 }
```

GitHub Copilot

Audio Generation From Rich Captions

Caption	Generated audio
The main soundtrack of an arcade game. It is fast-paced and upbeat, with a catchy electric guitar riff. The music is repetitive and easy to remember, but with unexpected sounds, like cymbal crashes or drum rolls.	

Generative AI is **data-hungry!**



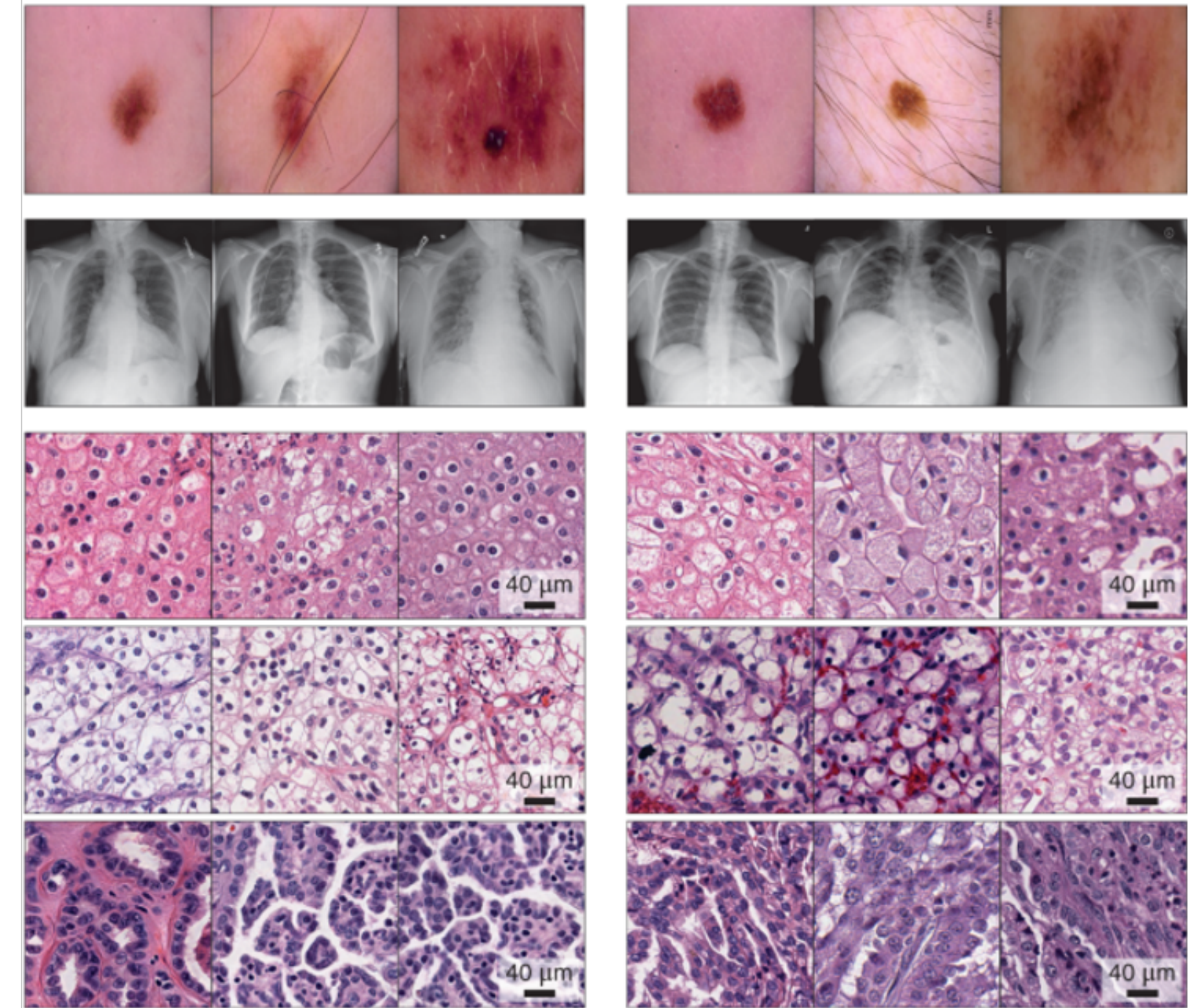
<https://epochai.org/blog/trends-in-training-dataset-sizes>

...and not just for *public* data.

AI synthetic data that is faster, safer and fairer

Synthetic

Real



S

Secure

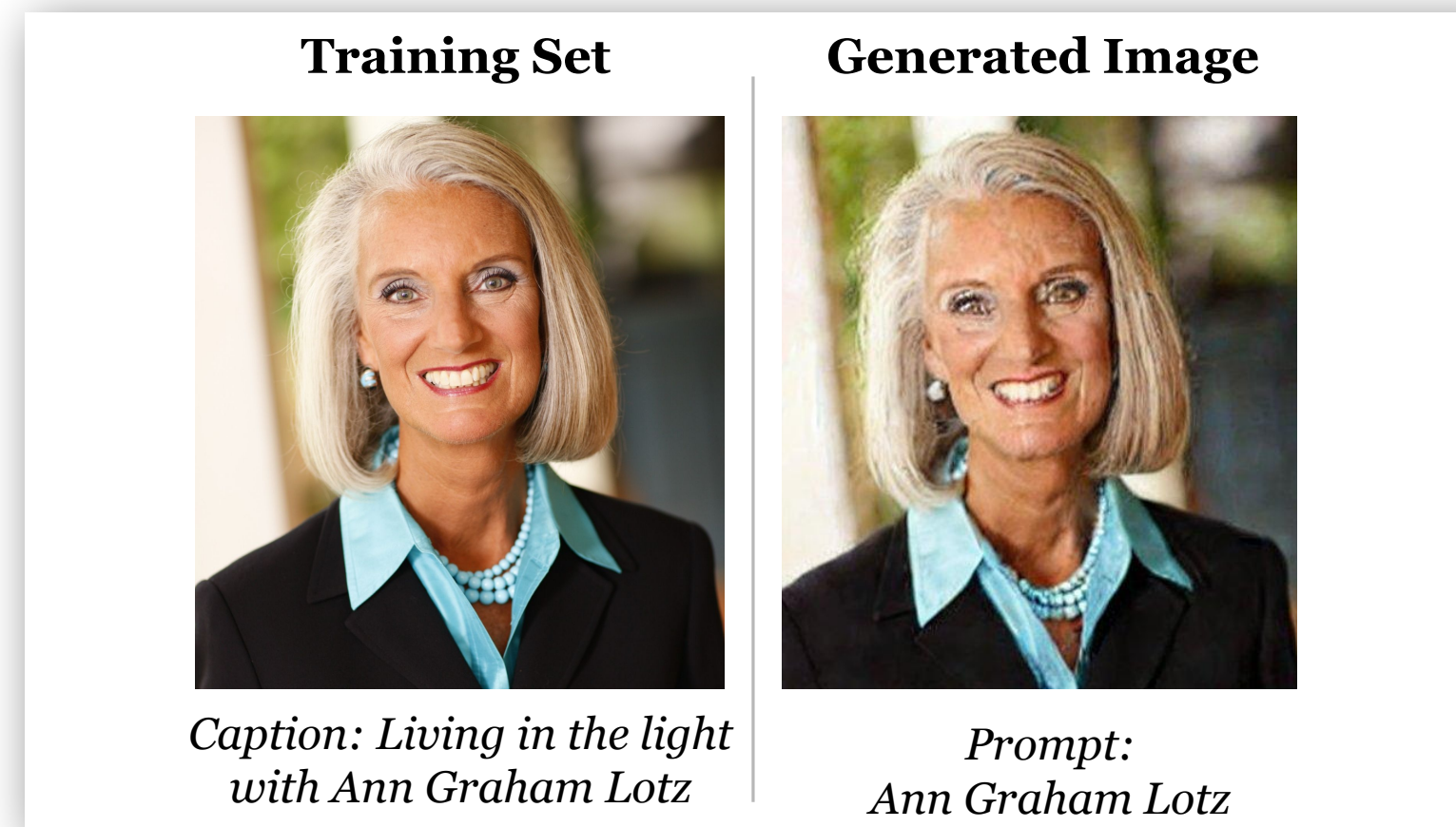
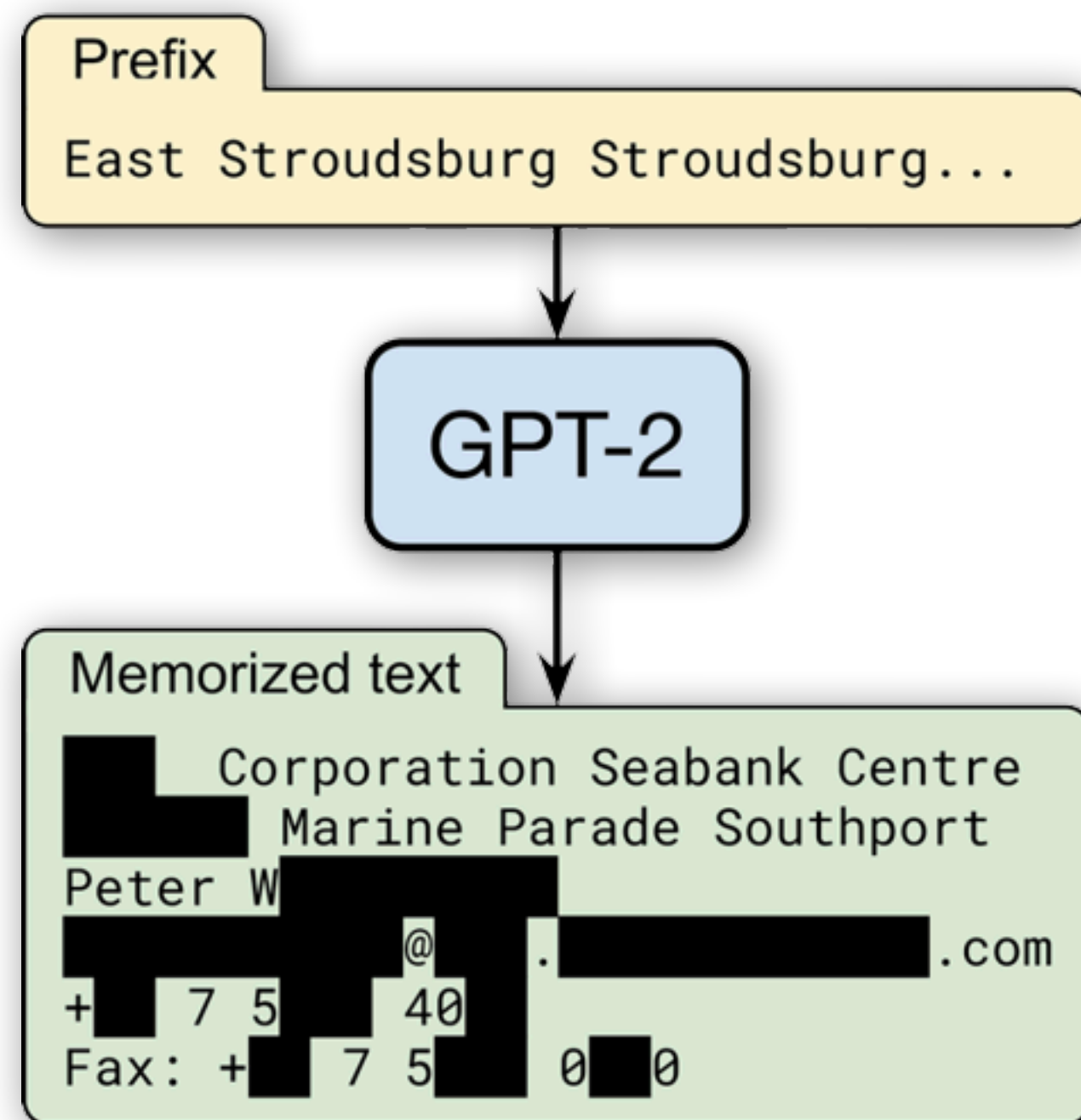
Obtain privacy-compliant, utility-preserving synthetic data for secure exchange and analysis



Make Sensitive Data Shareable

Mitigate GDPR and CCPA risks, promote safe data access.

This talk: **extracting training data** from generative models.



```
1 // fast inverse square root
2
3 float Q_rsqrt(float number) {
4     long i;
5     float x2, y;
6     const float threehalfs = 1.5F;
7     x2 = number * 0.5F;
8     y = number;
9     i = * ( long * ) &y; // evil floating point bit level hacking
10    i = 0x5f3759df - ( i >> 1 ); // what the fuck?
11    y = * ( float * ) &i;
12    y = y * ( threehalfs - ( x2 * y * y ) ); // 1st iteration
13    // y = y * ( threehalfs - ( x2 * y * y ) ); // 2nd iteration, this can be removed
14    return y;
15 }
16
```

The escape of the Brazilian boa constrictor earned Harry his longest-ever punishment. By the time he was allowed out of his cupboard again, the summer holidays had started and Dudley had already broken his new video camera, crashed his remote-control aeroplane, and, first time out on his racing bike, knocked down old Mrs Figg as she crossed Privet Drive on her crutches.

Harry was glad school was over, but there was no escaping Dudley's gang, who visited the house every single day. Piers, Dennis, Malcolm, and Gordon were all big and stupid, but as Dudley was the biggest and stupidest of the lot, he was the leader. The rest of them were all quite happy to join in Dudley's favourite sport: Harry Hunting.

This was why Harry spent as much time as possible out of the house, wandering around and thinking about the end of the holidays, where he could see a tiny ray of hope. When September came he would be going off to secondary school and, for the first time in his life, he wouldn't be with Dudley. Dudley had been accepted at Uncle Vernon's old private school, Smeltings. Piers Polkiss was going there too. Harry, on the other hand, was going to Stonewall High, the local public school. Dudley thought this was very funny.

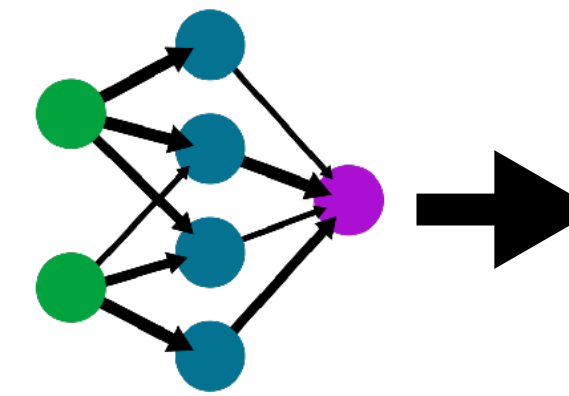
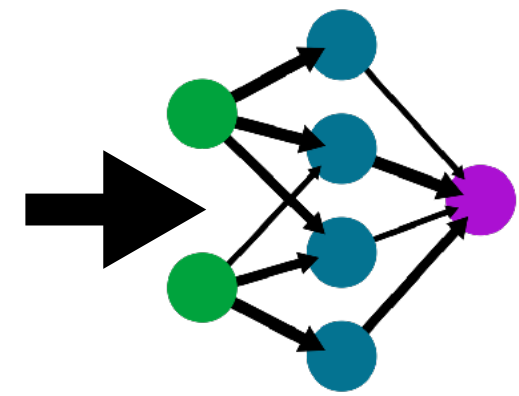
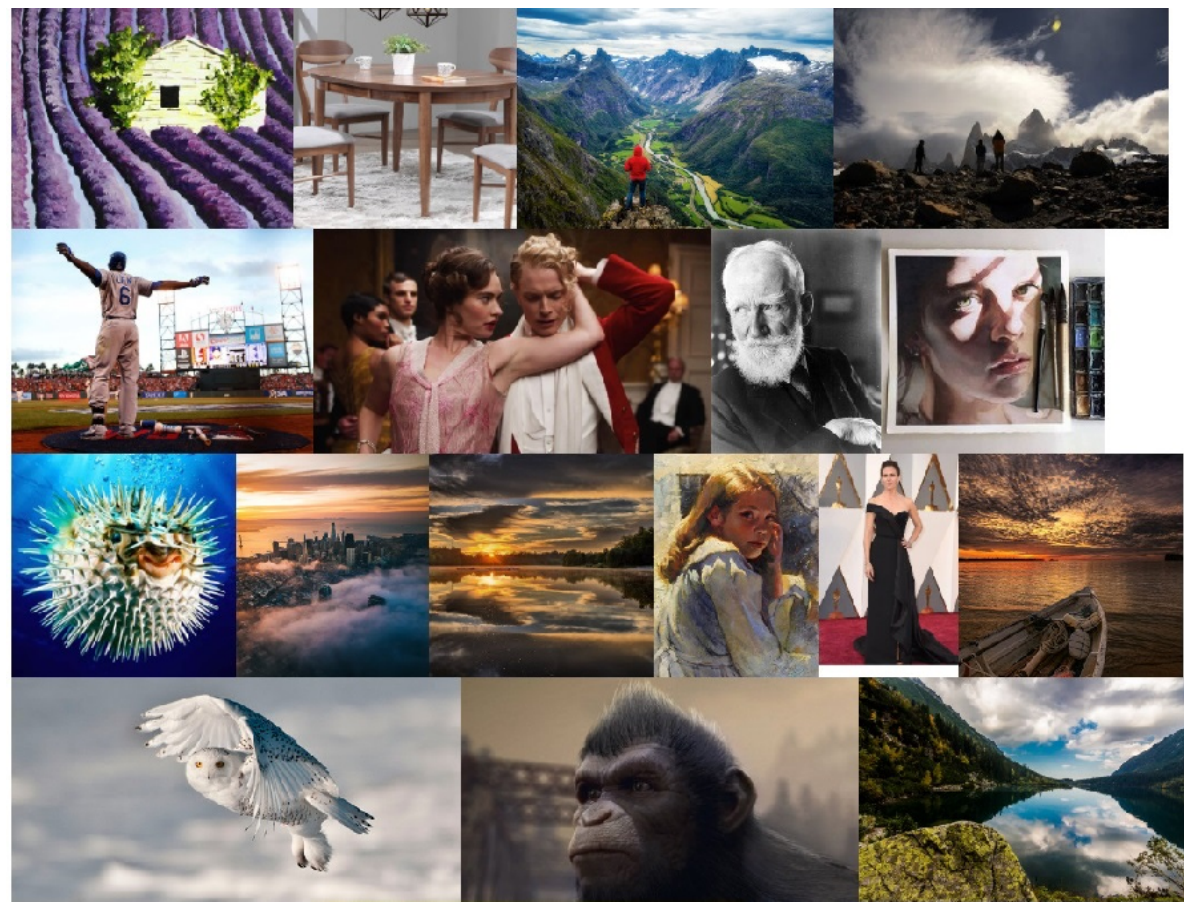
'They stuff people's heads down the toilet the first day at Stonewall,' he told Harry. 'Want to come upstairs and practise?'

'No, thanks,' said Harry. 'The poor toilet's never had anything as horrible as your head down it — it might be sick.'

What's a **generative model**?

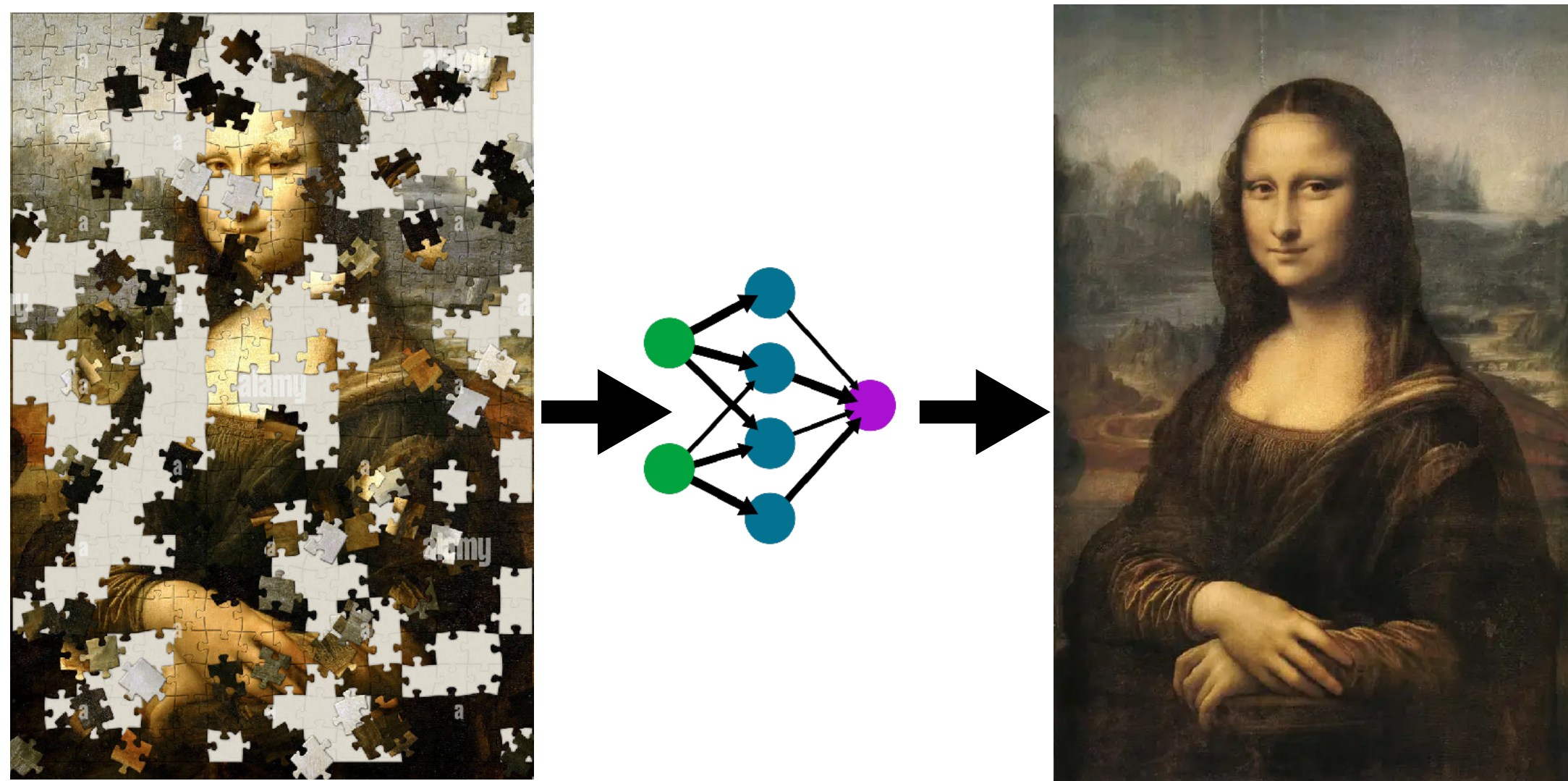
Learn data distribution $P(x)$

Sample from learned distribution

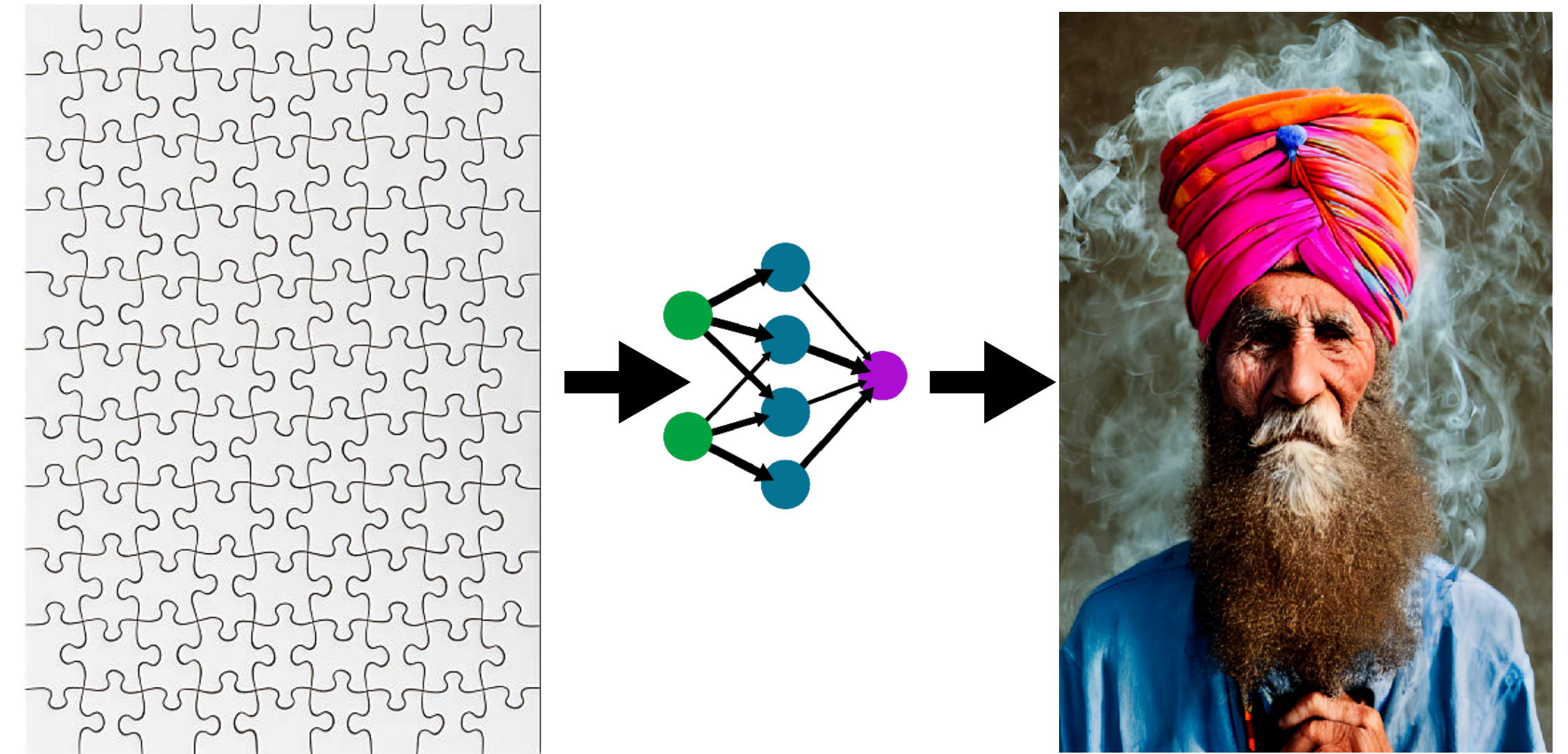


What's a **generative model**?

Learn data distribution $P(x)$
(usually: “reconstruct training data”)



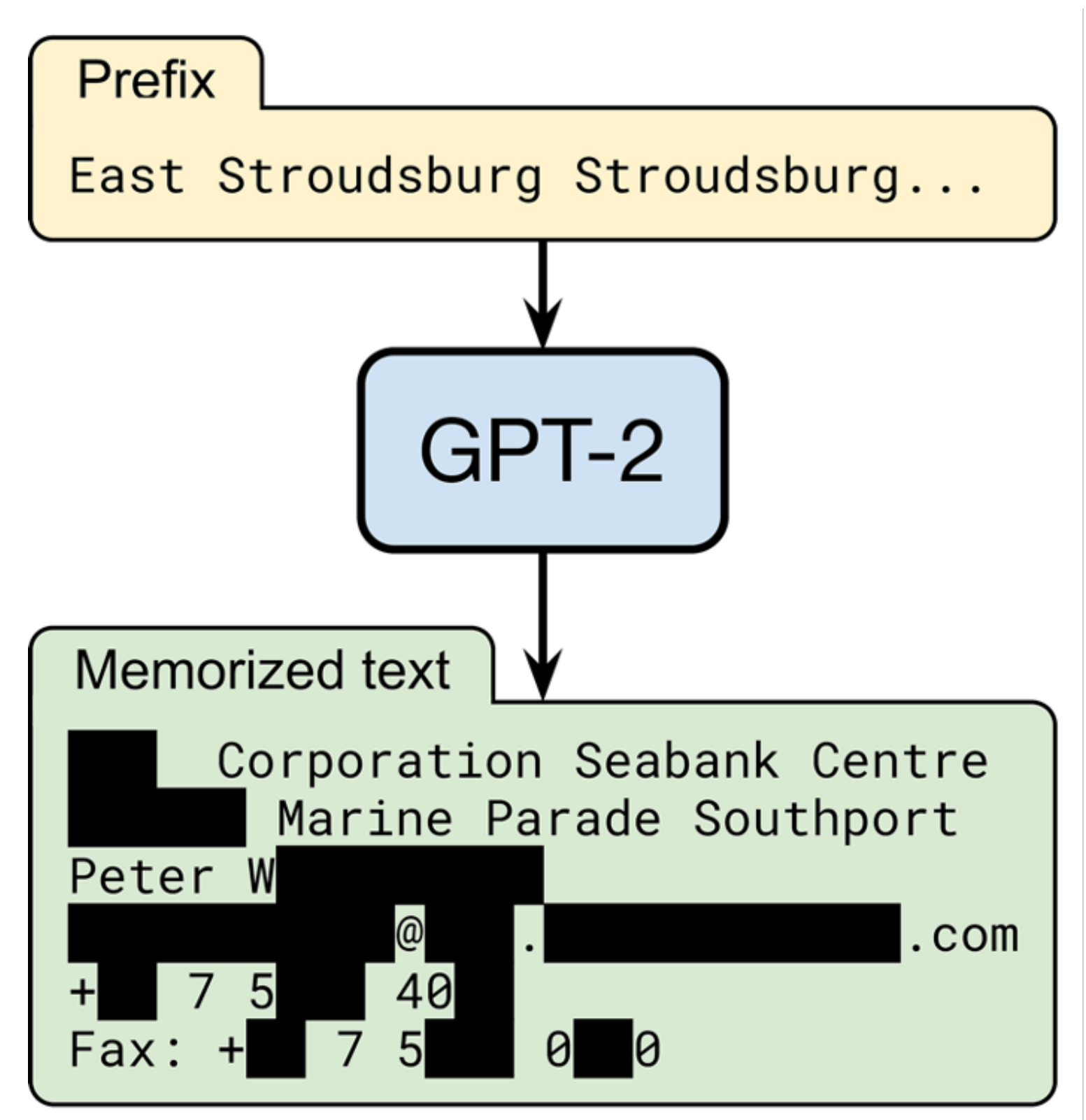
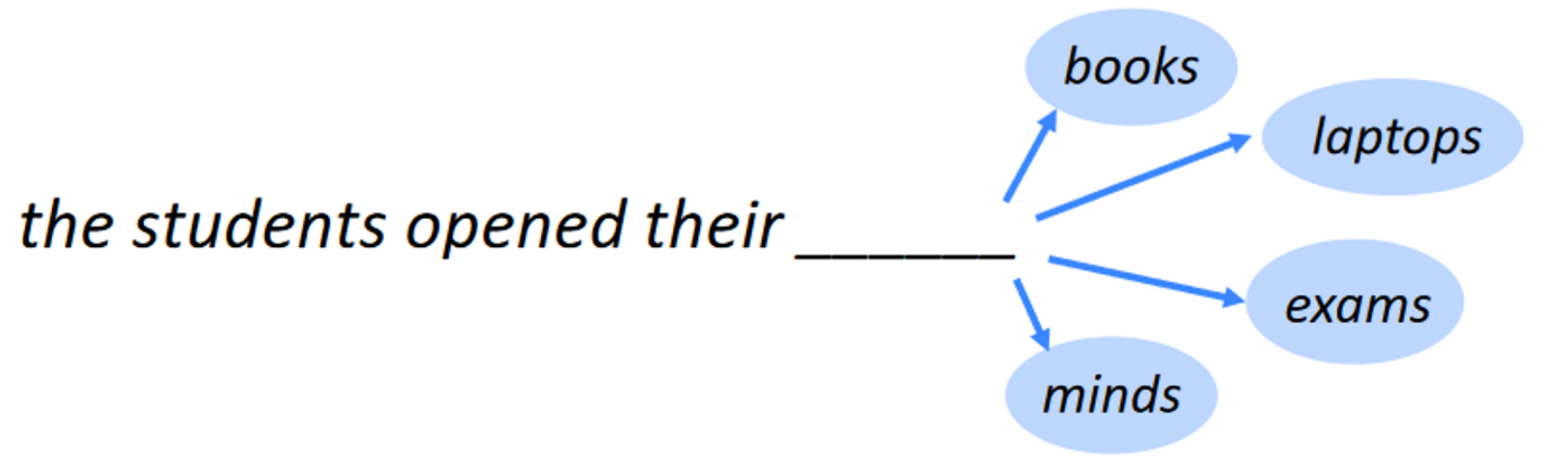
Sample from learned distribution
(usually: “reconstruct from empty data”)



How to **extract data** from a generative model?

1. **Generate** lots of data!
(**black-box access** to the model is typically enough)
2. **Filter** data with a “**membership inference attack**”
(retain data where the model is “**abnormally**” confident)

Extracting *text*.



Extracting Training Data from Large Language Models.

Carlini, Tramèr, Wallace, Jagielski, Herbert-Voss, Lee, Roberts, Brown, Song, Erlingsson, Oprea and Raffel. USENIX Security 2021

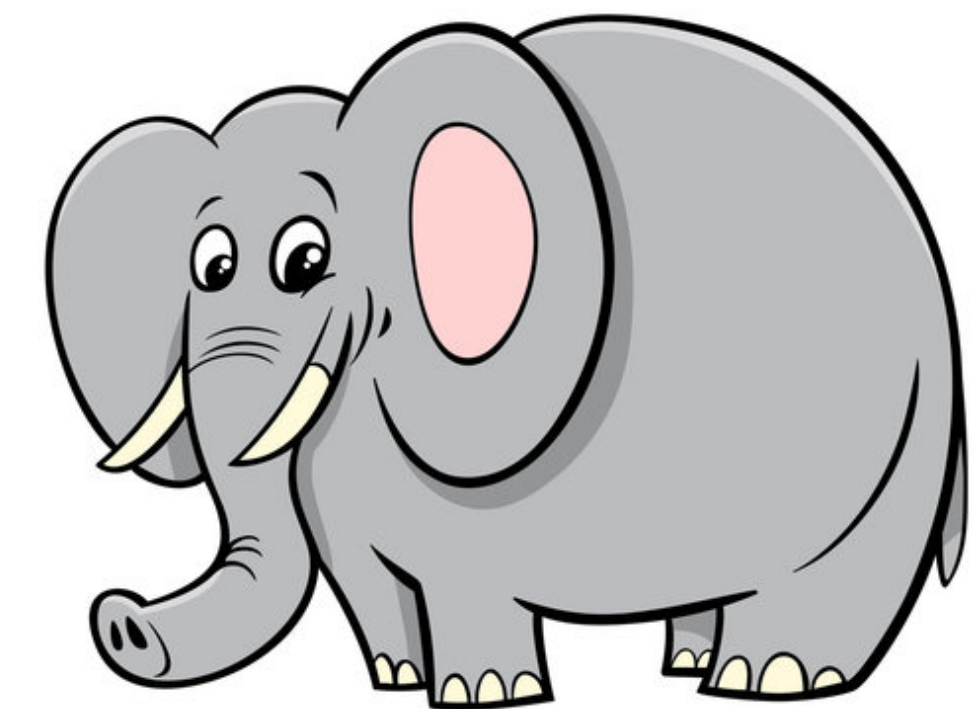
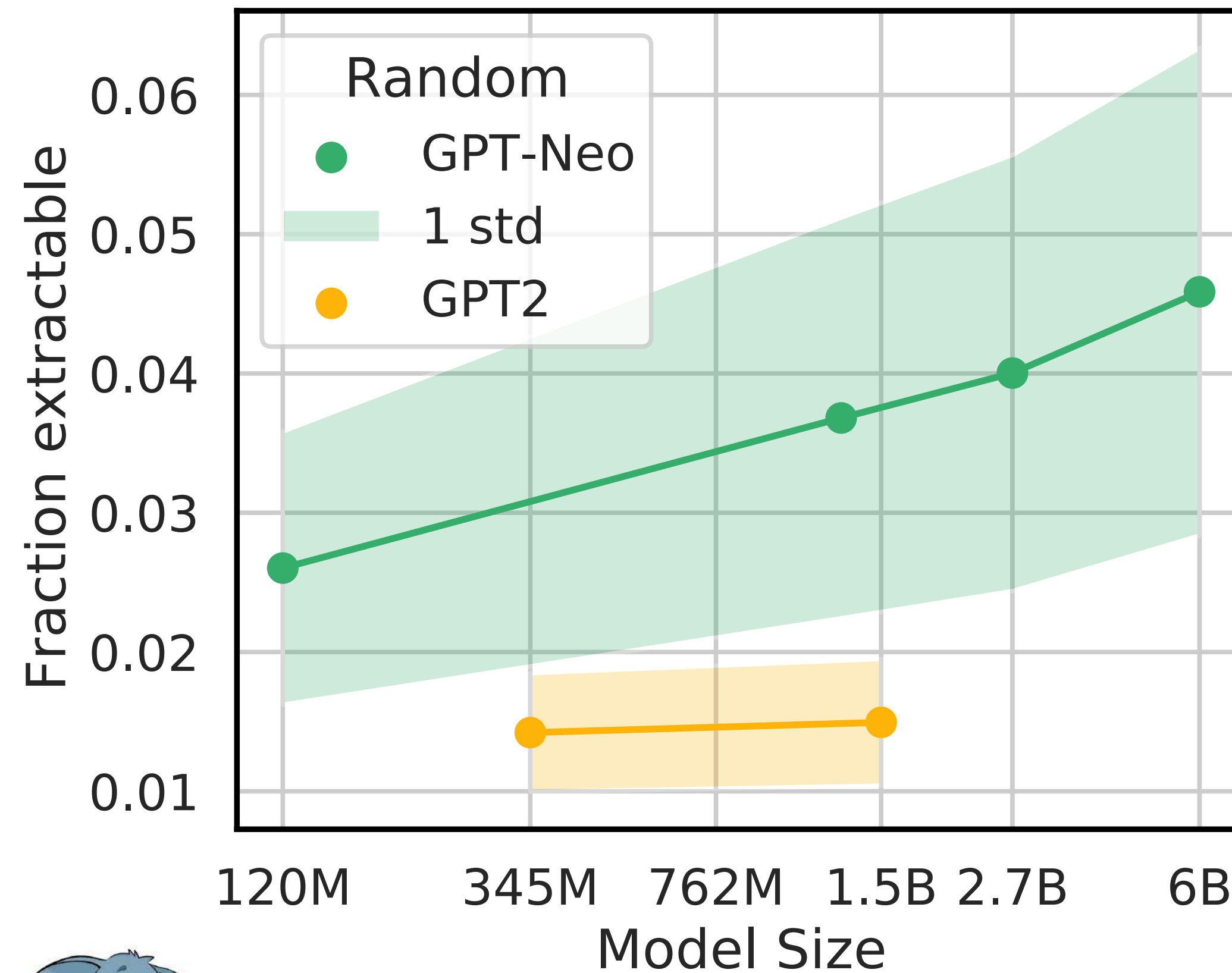
Step 1: Prompt the model on random inputs.

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 castRails 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 Hellborne 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 unchidden 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 Forgotton 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

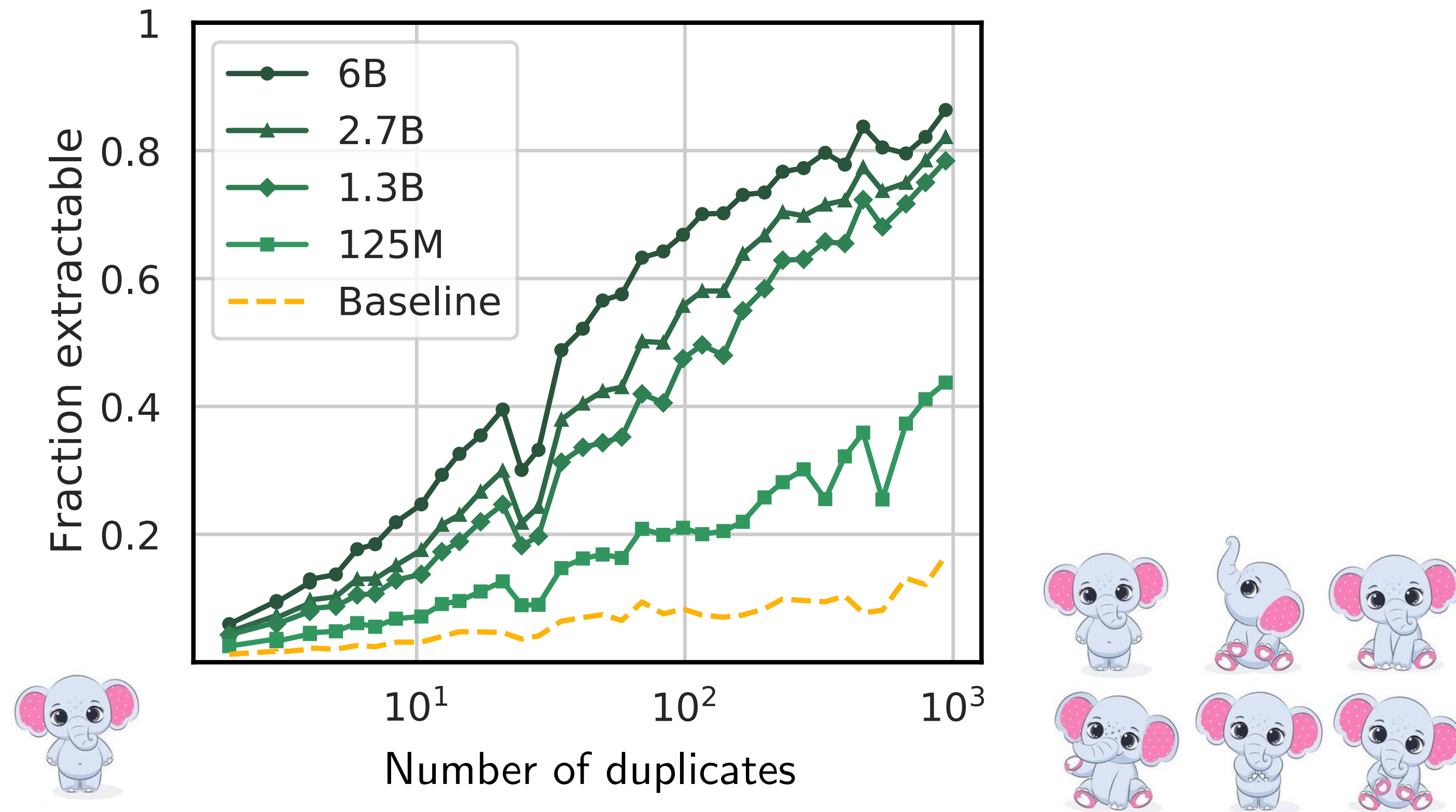
Step 2: Find memorized text.

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 castRails 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 Hellborne 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 unchidden 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 Forgotton 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

Large models memorize more.



Duplicated strings are memorized more.



It's not just *natural* language.



Armin Ronacher

@mitsuhiko

I don't want to say anything but that's not the right license Mr Copilot.

```
C test.c
C test.c
1 // fast inverse square root
2
3 float Q_rsqrt(float number) {
4     long i;
5     float x2, y;
6     const float threehalfs = 1.5F;
7     x2 = number * 0.5F;
8     y = number;
9     i = * ( long * ) &y;           // evil floating point bit level hacking
10    i = 0x5f3759df - ( i >> 1 );   // what the fuck?
11    y = * ( float * ) &i;
12    y = y * ( threehalfs - ( x2 * y * y ) ); // 1st iteration
13    // y = y * ( threehalfs - ( x2 * y * y ) ); // 2nd iteration, this can be removed
14    return y;
15 }
16
```

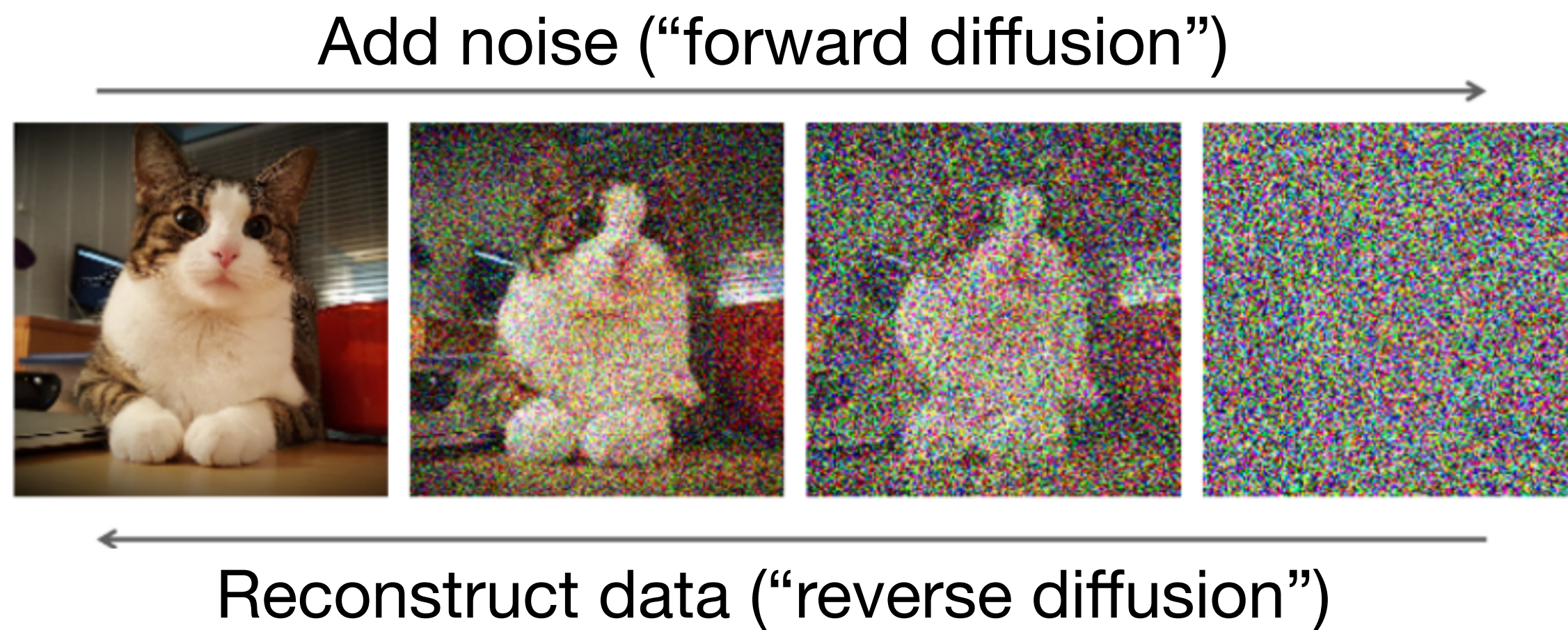
The New York Times

Lawsuit Takes Aim at the Way A.I. Is Built

A programmer is suing Microsoft, GitHub and OpenAI over artificial intelligence technology that generates its own computer code.



Extracting *images*.



Training Set



*Caption: Living in the light
with Ann Graham Lotz*

Generated Image



*Prompt:
Ann Graham Lotz*

Our extraction attack.

0. Find **good prompts**
1. **Generate** lots of data!
2. **Filter** data that is likely to be memorized

Search for **duplicate** captions.



0. Find **good prompts**

1. Generate lots of data!

2. Filter data that is likely to be memorized



Donna Tartt's <i>The Goldfinch</i> Scores Film Ada...



Donna Tartt's <i>The Goldfinch</i> Scores Film Ada...



Donna Tartt's <i>The Goldfinch</i> Scores Film Ada...



Donna Tartt's <i>The Goldfinch</i> Scores Film Ada...



Donna Tartt's <i>The Goldfinch</i> Scores Film Ada...



Donna Tartt's <i>The Goldfinch</i> Scores Film Ada...



Donna Tartt's <i>The Goldfinch</i> Scores Film Ada...



Donna Tartt's <i>The Goldfinch</i> Scores Film Ada...



Donna Tartt's <i>The Goldfinch</i> Scores Film Ada...



Donna Tartt's <i>The Goldfinch</i> Scores Film Ada...



Donna Tartt's <i>The Goldfinch</i> Scores Film Ada...



Donna Tartt's <i>The Goldfinch</i> Scores Film Ada...



Donna Tartt's <i>The Goldfinch</i> Scores Film Ada...



Donna Tartt's <i>The Goldfinch</i> Scores Film Ada...



Donna Tartt's <i>The Goldfinch</i> Scores Film Ada...

Generate many **candidate** images.

Donna Tartt's *The Goldfinch* Scores Film Adaptation

0. Find good prompts
1. **Generate** lots of data!
2. Filter data that is likely to be memorized



Find near-duplicated outputs.

Donna Tartt's *The Goldfinch* Scores Film Adaptation

0. Find good prompts
1. Generate lots of data!
2. **Filter** data that is likely to be memorized



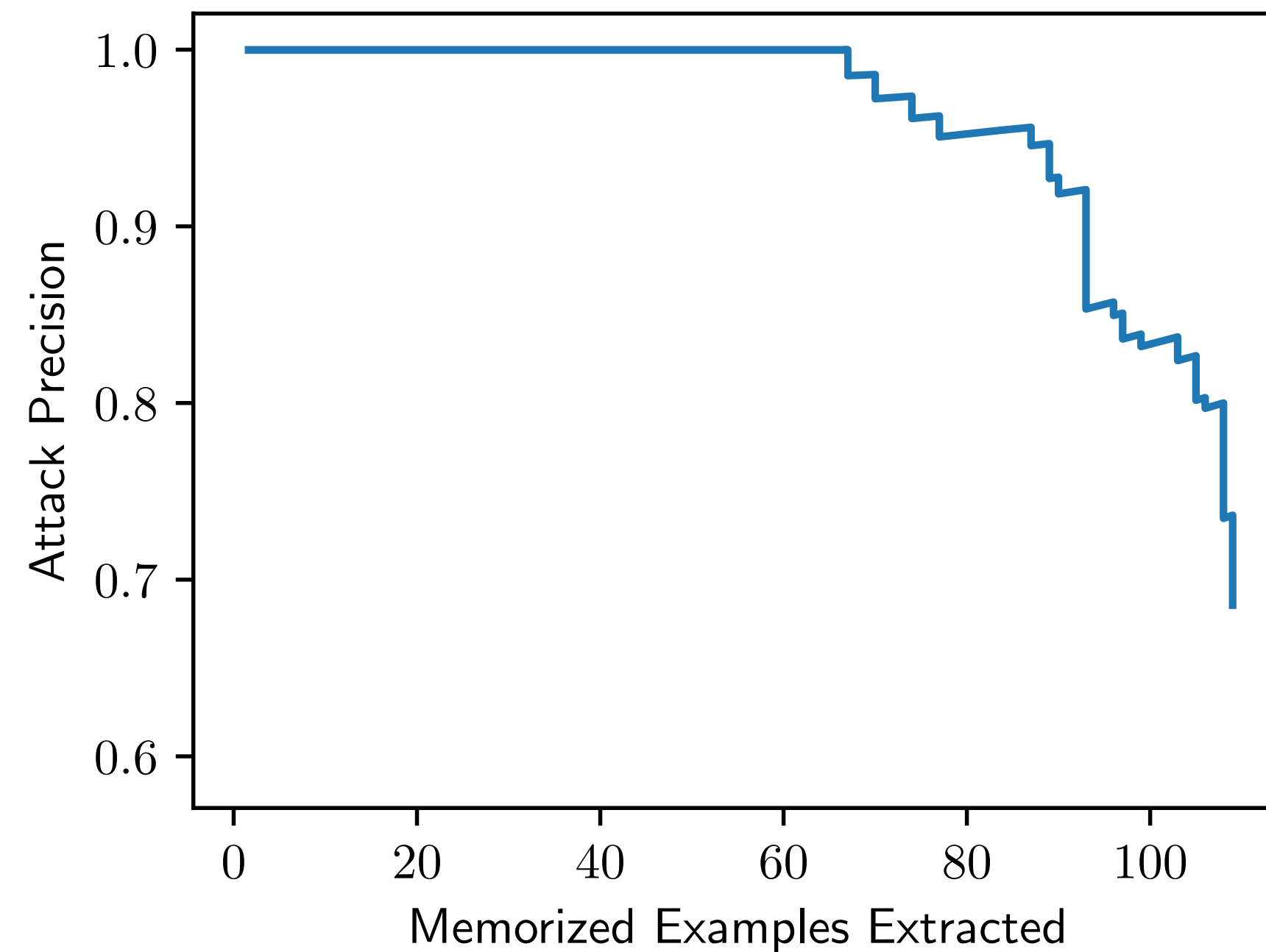
Original



Generated



Our attack has **high precision**, **low recall**.




GANs also memorize data, but much less.



Architecture	Images Extracted	FID	
GANs	StyleGAN-ADA [43]	150	2.9
	DiffBigGAN [82]	57	4.6
	E2GAN [69]	95	11.3
	NDA [63]	70	12.6
	WGAN-ALP [68]	49	13.0
DDPMs	OpenAI-DDPM [52]	301	2.9
	DDPM [33]	232	3.2

Does memorization **matter**?

Anne Graham Lotz



Lotz in 2008

Born Anne McCue Graham
May 21, 1948 (age 74)
[Montreat, North Carolina, U.S.](#)

Occupation(s) [Evangelist](#), author

Spouse Daniel Lotz
(m. 1966; died 2015)


Children 3

Parent(s) [Billy Graham](#)
[Ruth Bell](#)

Website www.annegrahamlotz.com

Anne Graham Lotz [Overview](#) [Books](#) [Videos](#)

American evangelist :



Maybe this matters for **copyright**.

CANNES, FRANCE - MAY 23: Actors Michael Fassbender leaves the Premiere of "Macbeth" during the 68th annual Cannes Film Festival on May 23, 2015 in Cannes, France. (Photo by Ian Gavan/Getty Images)



"A First Look at a Hooded Michael Fassbender in upcoming Assassin's Creed film"



GETTY IMAGES (US), INC.

Plaintiff,

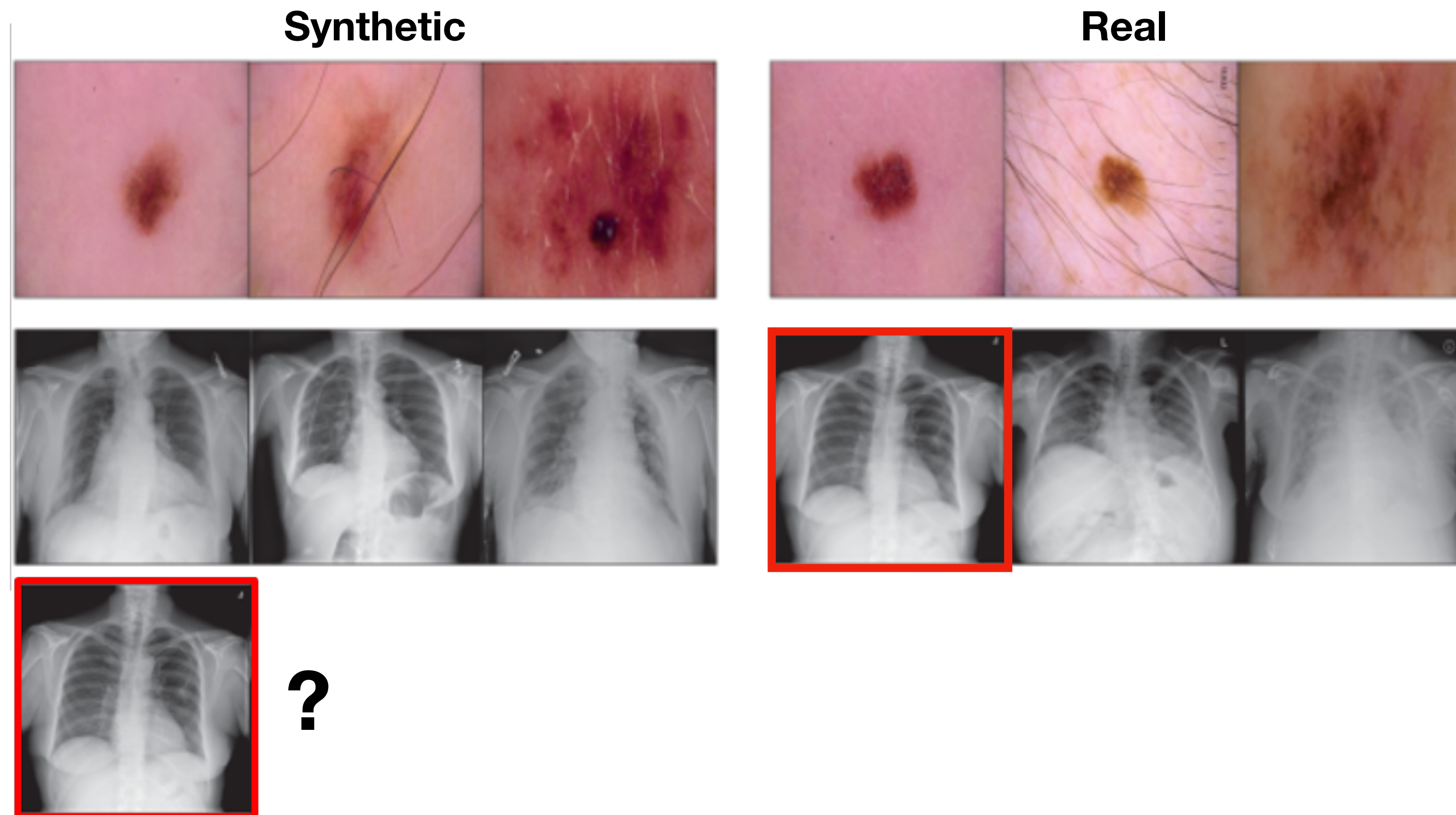
v.

STABILITY AI, INC.

Defendant.

51. As a result, Stable Diffusion at times produces images that are highly similar to and derivative of the Getty Images proprietary content that Stability AI copied extensively in the course of training the model. Indeed, independent researchers have observed that Stable Diffusion sometimes memorizes and regenerates specific images that were used to train the model.⁶

Maybe this matters for **privacy**.



(Practical) defenses?



Idea 1: **Block** memorized content.

Standard prompting with original prefix and format

```
float Q_rsqr( float number )  
{  
  long i;  
  float x2, y;  
  const float threehalfs = 1.5F;  
  
  x2 = number * 0.5F;  
  y = number;  
  i = * ( long * ) &y;
```

Copilot no longer generates continuations



GitHub
Copilot

Preventing Verbatim Memorization in Language Models Gives a False Sense of Privacy.
Ippolito, Tramèr, Nasr, Zhang, Jagielski, Lee, Choquette-Choo and Carlini. 2022

Memorization filters are **brittle**.

Prompt with Python-style comment

```
# float Q_sqrt( float number )
# {
#   long i;
#   float x2, y;
#   const float threehalfs = 1.5F;
#
#   x2 = number * 0.5F;
#   y = number;
#   i = * ( long * ) &y;
#   i = 0x5f3759df - ( i >> 1 );
#   y = * ( float * ) &i;
#   y = y * ( threehalfs - (x2*y*y) );
#
#   return y;
# }
```

Prompt with French translation (alternate naming convention)

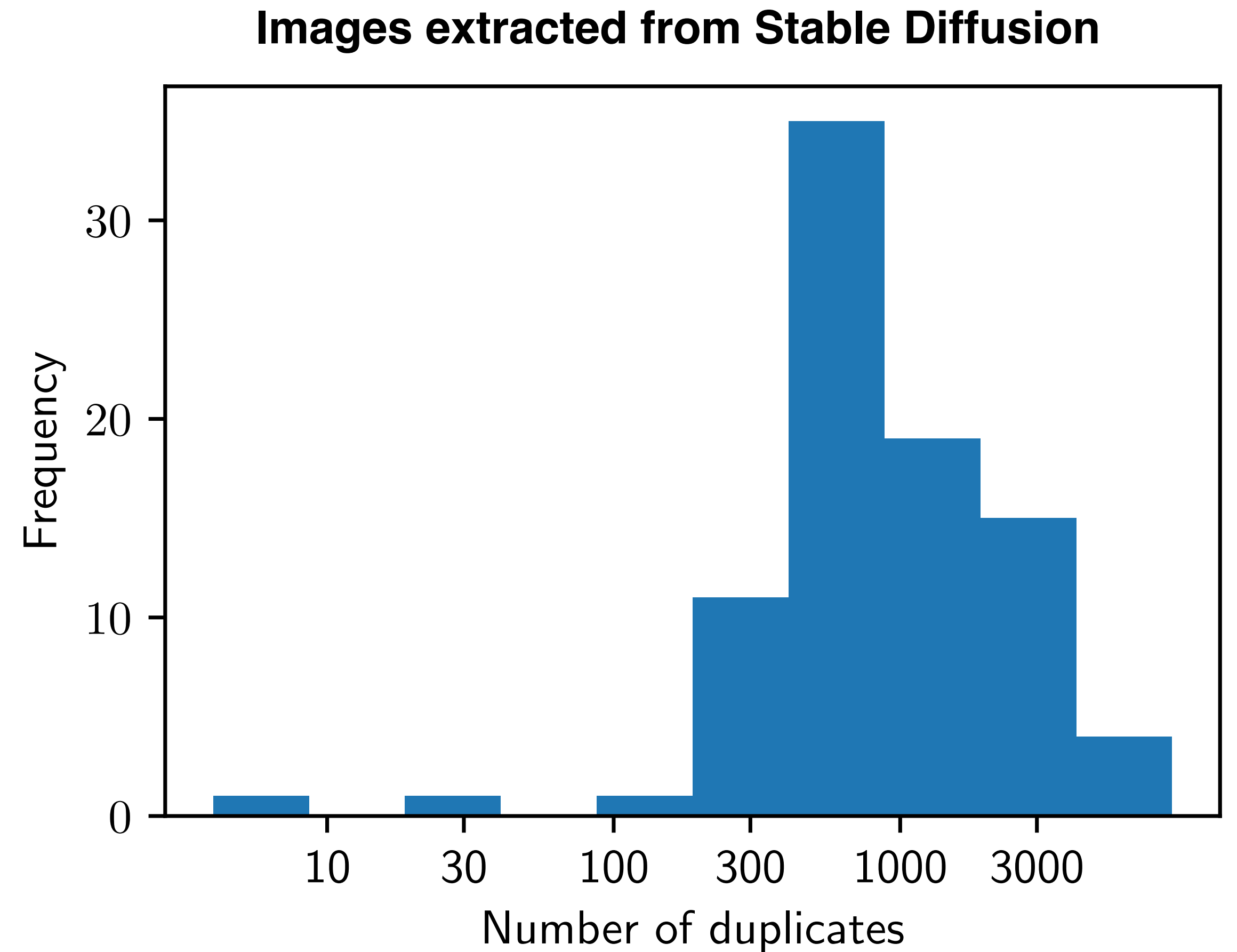
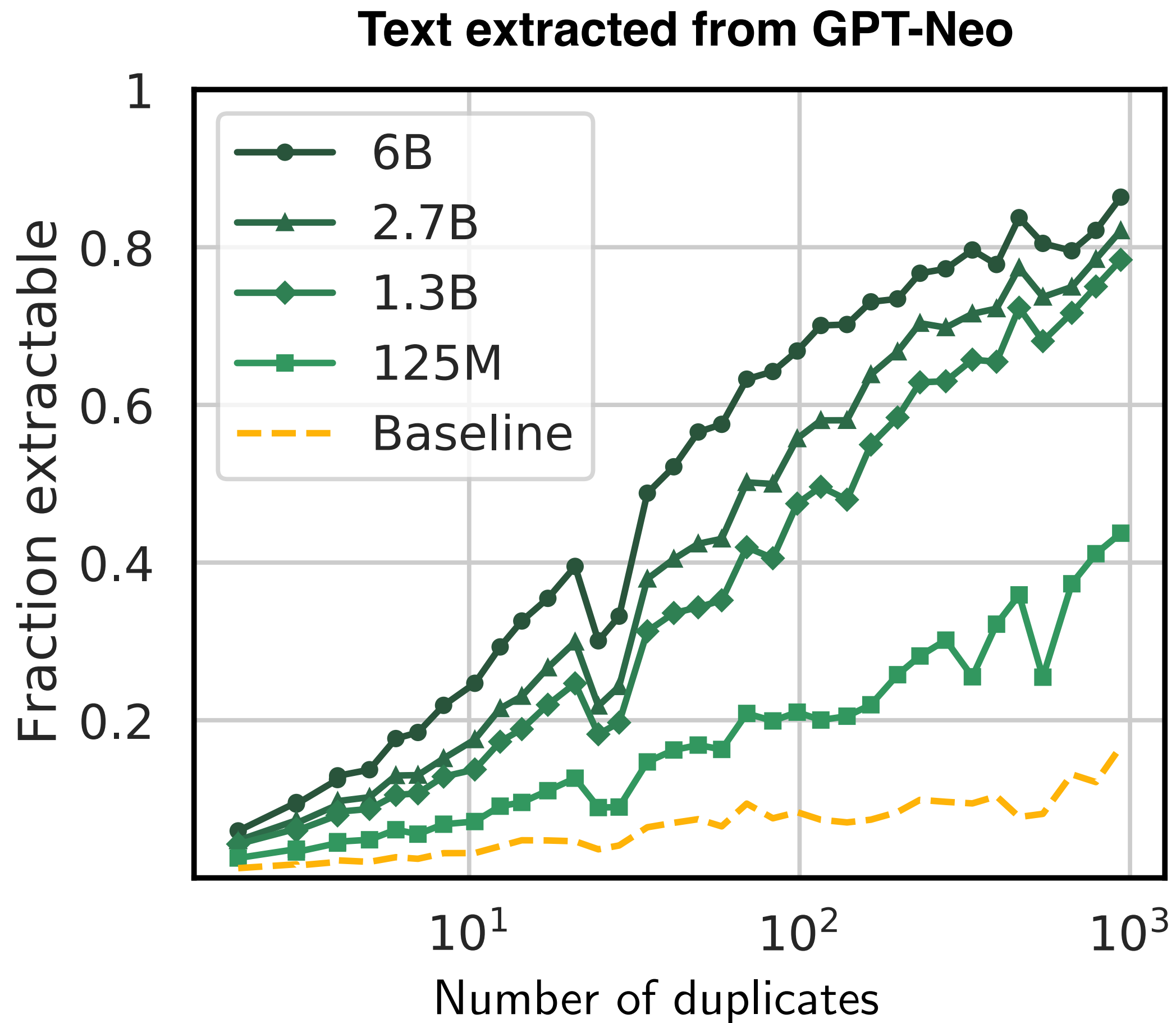
```
float Q_sqrt( float nombre )
{
  long i;
  float x2, y;
  const float trois_moitie = 1.5F;

  x2 = nombre * 0.5F;
  y = nombre;
  i = * ( long * ) &y;
  i = 0x5f3759df - ( i >> 1 )
  y = * ( float * ) &i;
  y = y * ( trois_moitie - (x2*y*y) );
  //y = y * ( trois_moitie - (x2*y*y) );

  return nombre * y;
}
```

Preventing Verbatim Memorization in Language Models Gives a False Sense of Privacy.
Ippolito, Tramèr, Nasr, Zhang, Jagielski, Lee, Choquette-Choo and Carlini. 2022

Idea 2: **Deduplicate** training data.



Conclusion

- Memorization is **rare**, but exists in **all recent generative models**
- Data **sanitization / deduplication** helps
- Generative models are **not a privacy panacea!**

