# Scaling paradigms for large language models

Jason Wei Research Scientist OpenAl

(Opinions are my own and do not reflect my employer.)

<u>2019</u>



- Can barely write a coherent paragraph
- Can't do any reasoning





- Can write an essay about almost anything
- Competition-level programmer and mathematician

Scaling has been the engine of progress in AI and will continue to dictate how the field advances.

#### <u>Outline</u>

What is scaling and why do it?

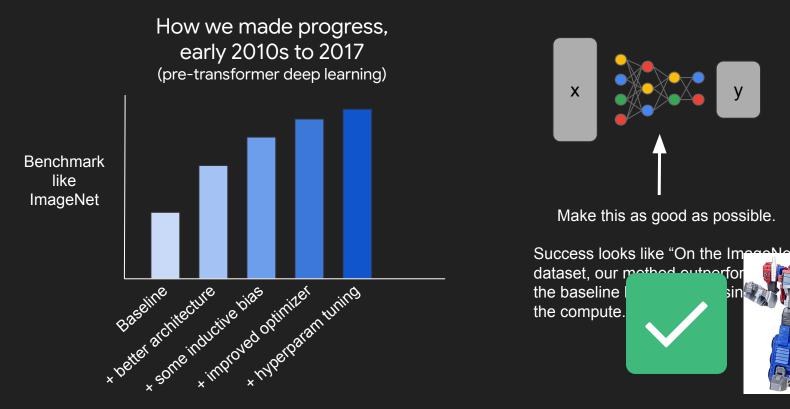
Paradigm 1: Scaling next-word prediction

The challenge with next-word prediction

Paradigm 2: Scaling RL on chain-of-thought

How scaling changed Al culture & what's next?

"Studying the past tells you what's special about the current moment."







Scaling is when you put yourself in a situation where you move along a continuous axis and expect sustained improvement.

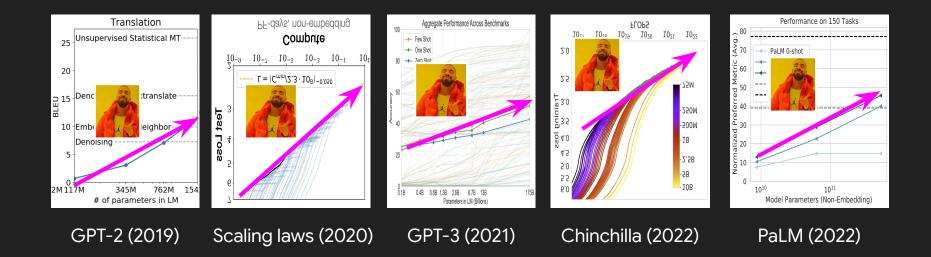


Capability

Something (usually compute, data, or model size)



#### Scaling is everywhere



# Scaling is hard and was not obvious at the time

#### Technical & operational challenges



(1) Distributed training requires a lot of expertise



Image source

(1) Researchers like inductive biases

(2) Scaling is different from

human learning

Psychological challenges



(2) Loss divergences and hardware failures are hurdles



(3) Compute is expensive



(3) Scientific research incentives don't match engineering work ("novelty")

#### Not scaling

Each improvement in the model requires ingenuity on a new axis

There are a lot of tasks that we want AI to do

#### **Scaling-centric Al**

You can reliably improve capability (even if it's expensive)

If your measure of capability is very general, extreme investment is justified

#### The Bitter Lesson of Al

# General methods that leverage compute are the most effective

# Things that scale will ultimately win out

#### The Bitter Lesson

#### **Rich Sutton**

#### March 13, 2019

The biggest lesson that can be read from 70 years of AI research is that gene computation are ultimately the most effective, and by a large margin. The ul Moore's law, or rather its generalization of continued exponentially falling c Most AI research has been conducted as if the computation available to the a case leveraging human knowledge would be one of the only ways to improv slightly longer time than a typical research project, massively more computa available. Seeking an improvement that makes a difference in the shorter ter leverage their human knowledge of the domain, but the only thing that mat leveraging of computation. These two need not run counter to each other, bu Time spent on one is time not spent on the other. There are psychological co one approach or the other. And the human-knowledge approach tends to co that make them less suited to taking advantage of general methods leveragin many examples of AI researchers' belated learning of this bitter lesson, and i some of the most prominent.

In computer chess, the methods that defeated the world champion, Kasparo massive, deep search. At the time, this was looked upon with dismay by the researchers who had pursued methods that leveraged human understandin chess. When a simpler, search-based approach with special hardware and sc effective, these human-knowledge-based chess researchers were not good lo force" search may have won this time, but it was not a general strategy and

#### Paradigm 1: Scaling next-word prediction

Started in 2018, still ongoing

Get really, really good at predicting the next word.

Why do you get so much from "just" predicting the next word? Next-word prediction is massively multi-task learning.

## Review: next-word prediction



#### Example "tasks" from next-word prediction

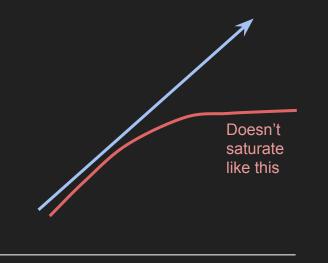
<u>Task</u>	Example sentence in pre-training that would teach that task	
Grammar	In my free time, I like to { <mark>code</mark> , banana}	
World knowledge	The capital of Azerbaijan is <b>{Baku, London}</b>	
Sentiment analysis	Movie review: I was engaged and on the edge of my seat the whole time. The movie was <b>{good, bad}</b>	
Translation	The word for "neural network" in Russian is <b>{нейронная сеть, привет}</b>	
Spatial reasoning	Iroh went into the kitchen to make tea. Standing next to Iroh, Zuko pondered his destiny. Zuko left the <b>{kitchen, store}</b>	
Math question	Arithmetic exam answer key: 3 + 8 + 4 = <b>{15, 11}</b>	

[millions more]

Extreme multi-task learning!

## Scaling predictably improves performance ("scaling laws")

Next-word prediction capability



Training compute (data x model size)

#### <u>Kaplan et al., 2020</u>:

"Language modeling performance improves smoothly as we increase the model size, dataset size, and amount of compute for training."

Jason's rephrase: You should expect to get a better language model if you scale up compute.

#### Why does scaling work? Hard to answer, but here is a hand-wavy explanation

Small language model	Large language model
Memorization is costly	More generous with memorizing tail knowledge
First-order correlations	Complex heuristics

If scaling was so predictable, why was the success of this paradigm so surprising?

Next-word prediction is secretly massively multi-task, and performance on different tasks arise at different rates

Let's take a closer look at next-word prediction accuracy. Consider that

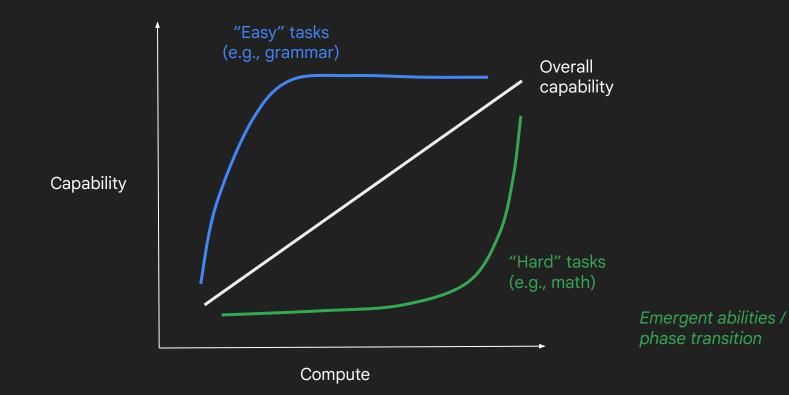
```
Overall accuracy = 0.002 * accuracy_grammar +
0.005 * accuracy_knowledge +
0.000001 * accuracy_sentiment_analysis +
```

...

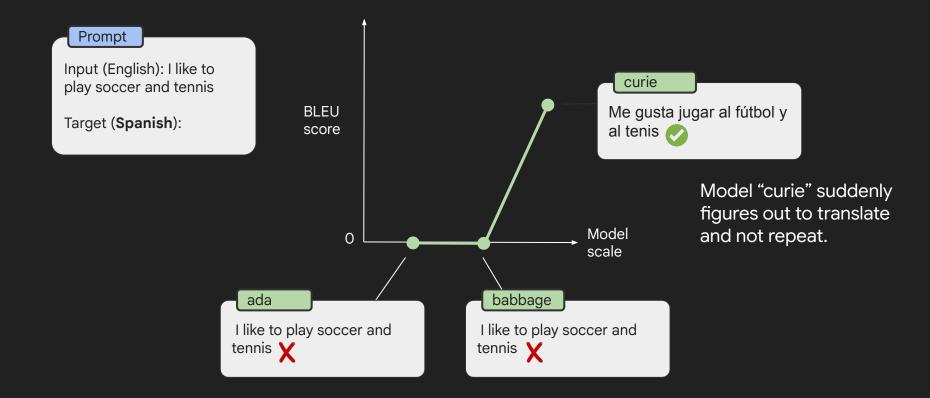
0.0001 \* accuracy\_math\_ability + 0.000001 \* accuracy\_spatial\_reasoning

•••

If accuracy goes from 70% to 80%, do all tasks get better uniformly? ...probably not.



# Emergence ability example



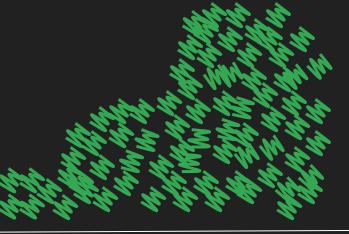
Write a novel Scientific research Hard math problems

"Spectrum of possible tasks" Help debug code Write a decent poem Do basic math problems Write a coherent essay

••

Translate a sentence Write a summary

Give basic facts Have correct grammar



GPT-2	GPT-3	GPT-4
(2019)	(2020)	(2023)

If next-word prediction works so well, can we scale it to reach AGI?

Maybe (it would be hard), but there is a bottleneck:

Some words are super hard to predict and take a lot of work

# When next-word prediction works fine

Playgrou	Complete 🗘	
Your presets	s 🗢	
Save Vi	ew code Share 🐵	
My name is J language <mark>m</mark> o		Q
	models = 63.28%           modeling = 11.41%           model = 5.72%           understanding = 3.98%           datasets = 3.93%	
Lookin	Total: -0.46 logprob on 1 tokens (88.31% probability covered in top 5 logits)	×
Submit	ත <i>හ</i> 19	

# When next-word prediction becomes very hard

Your presets			
Save	View code Share … 🐵		
Question (A) 1,483	n: What is the square of ((8-2)*3+4)^3 / 8? 3,492	ę	
(B) 1,395	5,394		
(C) 1,771,	,561		
(C) 1,771, Answer:			
	(C)	i i	
	( <b>C</b> ) C = 32.09%		
	C = 32.09% B = 29.98% A = 27.97%	×	
Answer:	C = 32.09% B = 29.98% A = 27.97%	×	

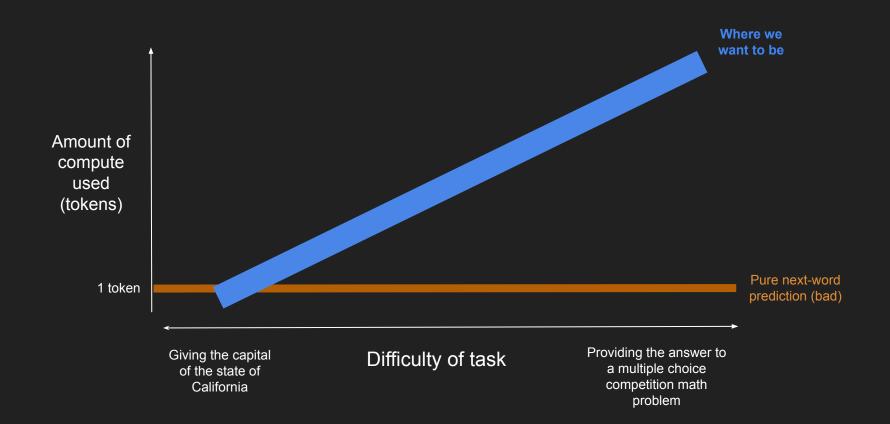
Pretend you're ChatGPT. As soon as you see the prompt you have to immediately start typing... go!

Question: What is the square of ((8-2)\*3+4)^3 / 8?

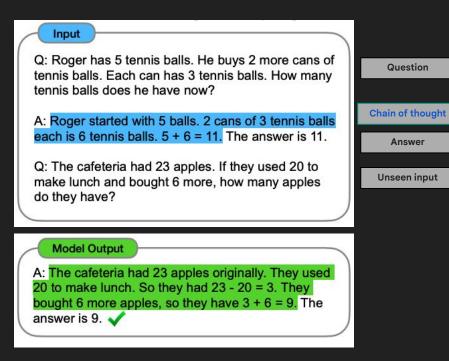
(A) 1,483,492
(B) 1,395,394
(C) 1,771,561

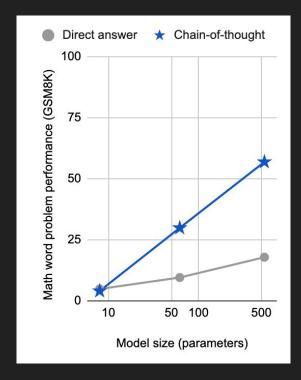
•••

Tough right?



### An approach: chain-of-thought prompting





THINKING, FAST AND SLOW DANIEL KAHNEMAN WINNER OF THE NOBEL PRIZE IN ECONOMICS "[A] masterpiece . . . This is one of the greatest and most engaging collections of insights into the human mind I have read." - WILLIAM EASTERLY, Financial Times

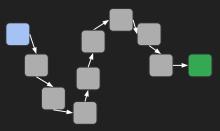
THE NEW YORK TIMES BESTSELLER

<u>System 1: Fast, intuitive</u>	<u>System 2: Slow, deliberate</u>
<u>thinking</u>	<u>thinking</u>
Automatic	Conscious
Effortless	Effortful
Intuitive	Controlled
Emotional	Logical
Recognizing faces	Solving math problems
Repeating basic facts	Planning a detailed agenda
Reacting to something	Making a thoughtful decision

Next-word prediction



Chain of thought



### The limitation with CoT prompting

# Most reasoning on the internet looks like this...

17-3: Formally prove Theorem 17.3.2.

**Theorem 17.3.2.** A one-pass algorithm for FREQUENCY-ESTIMATIO eter  $\boldsymbol{\varepsilon}$  must use  $\Omega(\min\{m, n, \boldsymbol{\varepsilon}^{-1}\})$  space. In particular, in order to g must use  $\Omega(\min\{m, n\})$  space.

*Proof.* We will prove the stronger result that the simpler FREQUEN which asks whether the input stream contains a token whose frequency space in the deterministic setting. Since the cost of the randomized principal and the stream of the st

Let  $\mathscr{A}$  be a one-pass *S*-space deterministic algorithm for FREQUENC input  $(\mathbf{x}, y)$  for the IDX<sub>N</sub>, Alice creates a stream  $\boldsymbol{\sigma}_1 = (a_1, a_2, \dots, a_N)$ , a stream  $\boldsymbol{\sigma}_2 = (b, b, \dots, b)$  of length k - 1 for  $k \ge 2$  where b = 2y - 1on the combined stream  $\boldsymbol{\sigma}_1 \circ \boldsymbol{\sigma}_2$  with parameter *k*.

The output of  $IDX_N(\mathbf{x}, y)$  is 1 iff  $\mathscr{A}$  produces *b* as output. This is so b will be the unique entry with  $f_b = k \ge k$ . Thus Alice and Bob can solv Alice to Bob using  $\mathscr{A}$ .

By the lower bound result of  $\Omega(N)$  for  $IDX_N$ ,  $S = \Omega(N)$ . By const  $N+k-1 \ge N+1$ . Therefore, we have proven a lower bound of  $\Omega(A)$  and  $n \ge N+1$ . We have thus proven that  $S = \Omega(\min\{m, n, \varepsilon^{-1}\})$ , since

What we actually want is the inner "stream of thought"

Hm let me first see what approach we should take...

Actually this seems wrong

No that approach won't work, let me try something else

Let me try computing this way now

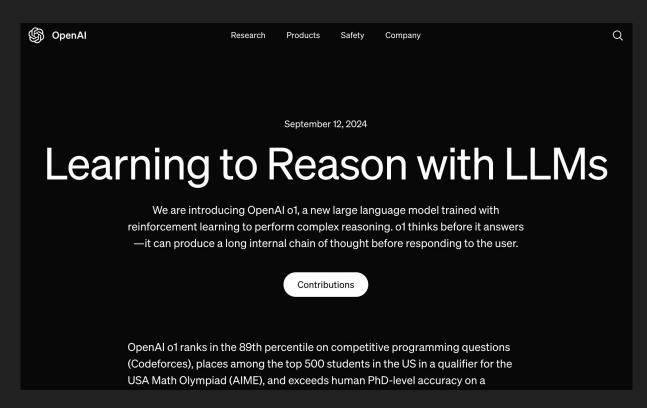
OK I think this is the right answer!

### Paradigm 2: Scaling RL on chain-of-thought

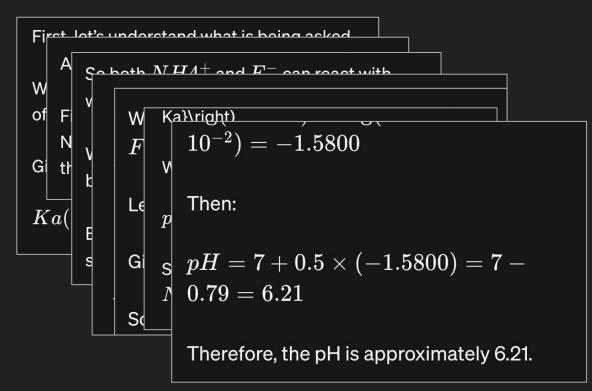
Train language models to "think" before giving an answer

In addition to scaling compute for training, there is a second axis here: scaling how long the language model can think at inference time.

#### OpenAl o1 (work of most of the company)



## A chain of thought from OpenAl o1



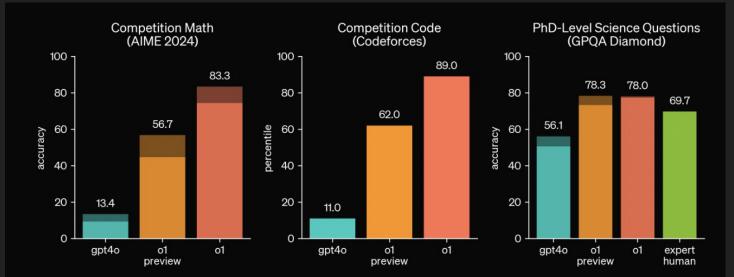
# CoT allows models to leverage asymmetry of verification

A class of problems has "asymmetry of verification", which means it's easier to verify a solution than to generate one

For example, a crossword puzzle, sudoku, or writing a poem that fits constraints

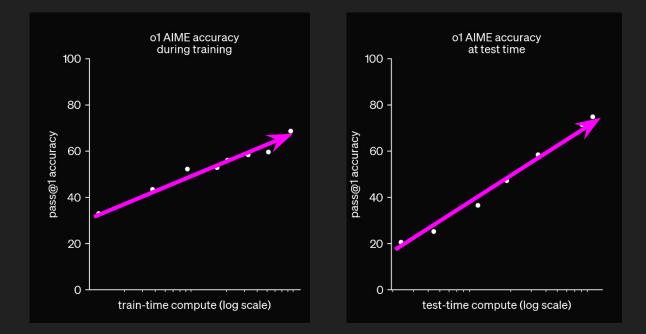
	Acros		
Solve the	1. Eva	1 Acros	Now let's look at Down clues.
	2. On	Possib	1 Down: car (station wagon) (6 letters)
PlainT	3. <b>Mc</b>		Possible words:
	4. Init	ESCAF	
1 +	5. <b>Na</b>	AVOID	- ESTATE car (6 letters)
2   3 +	6. <b>Mi</b> :	AVOID	In British English, 'Estate car' is a term for station
4   5 +	Down:	DODG	wagon.
6   7 + 8   9 +	1 2. <b>Au</b> '	ELUDE	Since 'station wagon' is called 'estate car' in the UK.
10   11 +	3. Prc	Maybe	Therefore 'ESTATE' fits.
12   13 +	4. Syı 5. An	But let	Also aligns with ESCAPE as Across 1.
	6. Deletes		

## Scale RL on chain-of-thought



o1 greatly improves over GPT-4o on challenging reasoning benchmarks. Solid bars show pass@1 accuracy and the shaded region shows the performance of majority vote (consensus) with 64 samples.

## Scale inference-time compute



# Why is this special: one day we may want AI to solve very challenging problems

#### Prompt

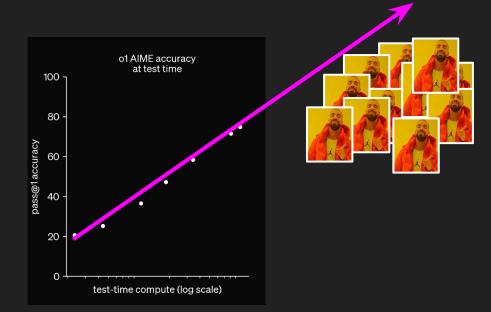
Write the code, documentation, and research paper for the best way to make AI safe

#### Hypothetical response

Let me think very hard about this...

[Researches all the existing literature] [Data analysis] [Conducts new experiments]

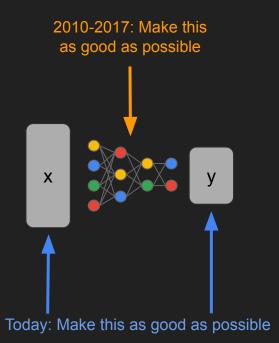
OK, here is a body of work on how to make AI safe



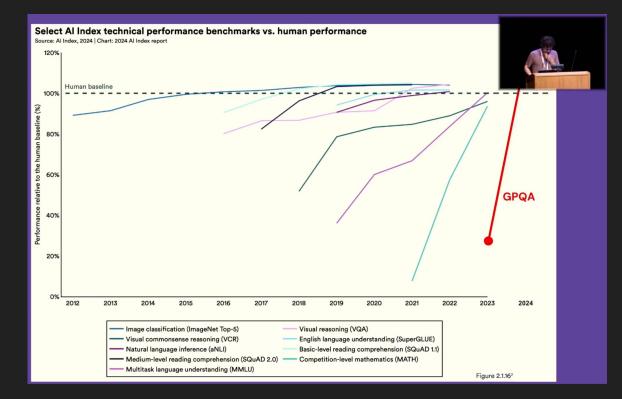
seconds minutes hours days weeks months

How has scaling changed the culture around doing AI research?

## Changes in AI research culture: shift to data

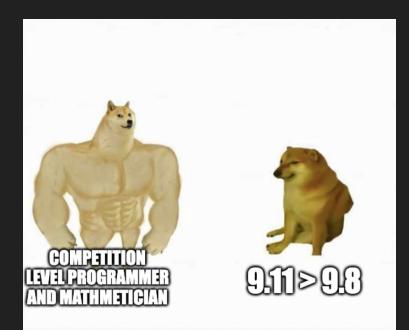


### Changes in AI culture: we desperately need evals



"People ask me if I'm making an even harder version of GPQA... [well] we set out to make the hardest science benchmark that we could" - David Rein

# Changes in AI culture: highly multi-task models



Language models must be measured on many dimensions

Hard to say that one model is strictly better than another

AI doesn't need to human-level on everything

Intelligence != user experience

## Changes in Al culture: bigger working teams



#### Adam: A method for stochastic optimization

Authors Diederik Kingma, Jimmy Ba Publication date 2015 International Conference on Learning Representations Conference Total citations Cited by 197418 2016 2017 2018 2019 2020 2021 2022 2023 2024

#### Gemini: A Family of Highly Capable Multimodal Models

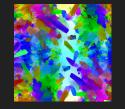
	Cemin Trans Coople: Ruhan Anli, Sebastian Borgeaud, Jean-Baptiser Alayraz, Jahan Yu, Rudo Soricar, Johan Schallowyi, Andrew H. Dai, Anjia Huath, Katin Millican, David Shew, Muhini Johnson, Joannis Annongolu, Jalan Schrimeiser, Amelia Clases, Jilin Chen, Timi Yihar, Timorby Lilicana, Angelia Lazar, Ohan Fazz, James Molay, Mobal Jauk J. Ruhan, Tone Imengan, Benjamin Lee, Jalai Volai, Matchini Rynobi, Yuanzhong Xu, Byai Dohen, Tili Zimin, Cennes Muhay, Mobal Jauk J. Alai Kathan, Tone Imengan, Benjamin Lee, Jalai Volai, Matchini Rynobi, Yuanzhong Xu, Byai Dohen, Tili Zimin, Cennes Muhay, Machani Lavia, Lavien Annon, Da Chi, Yeng-Ta Cheng, Tin Xi, Wang Xu, Byai Dohen, Tili Zimin, Cennes Muhay, Mohan Jian Schmitzman, Zimin Kennes, Kateren Ayou, Mayai Doh, Jian Schmitzman, Tanan, Temerikan Mathan, Marcini Kennes, Kateren Ayou, Mayai Doh, Jian Schmitzman, Tanan, Temerikan Mathan, Marcini Kennes, Kateren Ayou, Kateria, Kateren Ayou, Kate
	Graught, Senz Then, Byezheng Chan, Kapton Weitz, Te, Zhu, Ying La, Sdahuh Goal, Janer Alen, Mariqi Galu, JH, Filman, Kahir Shu, Filman, Kahir Shu, Filman, Kahir Shu, Filman, Kahir Shu, Filman, Zhang Chan, Zhang
Gemi	Theorematics Selection compares TRE, Lando Dorch, Martel Lando, Dany Aluce, Camor Tan, Lancess Bancos, Adir Pagalamiente Mala, Savel Renne, Marce Agenerative Selection and Annual Annua
Gemini <sup>·</sup>	And, Asian Casa, Kanin Wila, Luo Yuong, Henhon Jua, Muther Akara, Yuo Canteou, Logg Yang, James Keing, Palos Hoora, Barin Kasha, Salom Kasha, Karle Sanon, Yao Walina, James Keing, Kasha Hooka, Kasha Walina,
Radu So	State Link
Melvin J	Yunhao Tang, Dala El Badawy, Boyeth Whan, Bail Mustah, Dana Lang, Abhahak Jindal, Shanal Vakane, Zhalao Gong, Sengi Carlen, Isos Hennikey, Cergory Thoreton, Fangaloage Feng, Wajdoch Stolennice, Ca Zheng, Hoobe Thacker, Cajlar Dinil, Zinhaui Zhang, Mahammad Salah, James Senson, Kaca Bilanciki Pyoh Jinti, Antoin Anaud, Baman King, Laterina Tailoka, Angi Yaeny, Karoo Salah, Jinto Jinti, Zina Heng, Mahammad Salah, James Senson, Kaca King, Tai Salah, Mushahanda, Bangel Hidi Handen, Karol Heng, Karoo Salah, Jinti Pyohenkan, Mille Hodingara, Tain Kinatakani, Karoo Heng, Maniha Salah, James Tailo, Karoo Hang, Mahammad Salah, James Senson, Kaca King, Tai Salah, Mushahanda, Bangel Hidi Handen, Karol Hang, Karoo Karoo, Heng, Handana, Karoo Karoo, Karoo
Pitler, Ti	Jool, Annual Guy, Marthen Tang, Andren Hoch, Kunnah Shaha, Andrec Ahren, Sanga Takafarek, Manazi A Dahay, Ling Shar, Ling Shar, Janga Taka, Janga Taka
Paul R. I	Genera tombare, Kaloim Nakha, Jani San, Vaferal Jahman Mahamad, Janip Maak, Yadi Qiao, Wan Penawa, Panei Jiona, Qiao Yuao, Yafo Kolon, Dan Dadi, Aelox Gang, Yifar Ho, Ohkin Dashy, Aeron Algung, Timothie Hana, Qi U, Waai Yadar, Ugaya Na, Alex Ohinen, Kalesh Shinanna, A Adatand: Challin, Joo Li, Carrie Spaten, Tanix Walfe, Kareen Mahamad, San Jiong, Dia Ye Ho, Baniro Yano Dashiga, Shyan Uhayahay, Akataba Manya, Luon O, Sheetan Tarava, Shali Shinan, Yan Calanovich, Pana Kama Kedoy, Mana Shena, Nihai D Attestare, Ganza Challin, Uferd Gane, Hannen Kang, Ma, Mana Shana, Alitahan Shana, Nihai D Attestare, Adatar Cang, Yifan Ho, Di Kataba Manya, Luon O, Sheetan Tarava, Shali Shana, Yana Kama, Kedoy, Mana Shena, Nihai D Attestare, Ganza Chall, Uferd Gane, Hannen Ganza, Wul, Luong, Sama, Gan Hennin, Manaha, Shana, Nathari Manya, Luon, Mana Kataba, Manya, Luon Jana, Kataba Manya, Luon Manahan, Kataba Manahan, Kataba Manahan, Luon Manahan, Kataba Manya, Luon Manahan, Kataba Manya, Luon Manahan, Kataba Manahan, Kataba Manahan, Kataba Manahan, Kataba Manahan, Kataba Manya, Luon Manahan, Kataba Manya, Luon Manahan, Kataba Manahan,
Yuanzho	Little, Little
Kareem	Coden, Shak Krang, Em Cahn, White King, Yun Kongan, Yun Kong, Xiao Kang, Xiao
Purvi Sh	Solders, Michael Wandl, Anter Kirman, Panel Novak, Timpi Wu, Alex Dyck, Krohnan Valdparahan, Raghaender R, Jossica Maller, Mich Rudomier, Eric Johnson, Sunh Micral, Abhi Ushim, Janara Drinzenser, Vohal Yema, Zach Inreg, Andreas Samacci, Camalidin Biayed, Biaz Daucodi, Marin Goori Ian Temery, Michael Wandl, Anter Kirman, Panel Novak, Timpi Wu, Alex Dyck, Konhana Valdparahan, Raghaender R, Jossica Maller, Michael Jang, Japar Socie, Maund Samarajan, Kanzili Mang, Zach Drinzen, Fay Japar Socie Maund Samarajan, Kanzili Mang, Zach Drinze, Bara David, Yun Balanshah, Lin Tameh, Spal Ben-Danki, Fahab Jain, Danisha Usera, Danisha Jang, Samarajan, Kanzili Namarajan, Kanzili Mang, Japar Socie, Maund Samarajan, Kanzili Mang, Zach Drinze, Bara Bana, Jahabahay, Lin Koho, Japar, Sonshara Usera, Kanzili Mang, Jahan Japar Samarajan, Kanzili Mang, Japar Socie, Maund Samarajan, Kanzili Mang, Jabar Socie Maund, Jin Kao, Advena, Kanzili Kanzili Mang, Japar Socie Maund, Janaka Advena, Jin Nama, Advena, Jamarajan, Janali Mang, Janaka Advena, Jin Nama, Advena, Janaka Advena, Jana
YaGuan	Japps Daug, Pill Shanky, Deve Mayal, Sharkin, Kanakin, Ka
Tran, Su	Comin Patients, Dayl Peng, Kathimi Les, Dayna Zhang, Gannel Creang, Dac Dong Bagnen, Palari Karplonicz, Canish Matani, Leca Dono, Lill Janzer, Kana Dong, Zeigher Feng, Bain Zhang, Achimya Singhal, Dayno Du, Bac Kolimon, Natani A-Antonpoor, Talga Baldakai, Orgaf Ealer, Daval Bald, Di Tachobanin, Marki Alk Kan, Kemi Crecker, Peter Hankim, Baher Mathim, Canard Linnan, Hall Maton, Lindon, Canard Mathim, Landon Canard, Baler Mathim, Sanghal Linnan, Kanard Linnand, Kanard Linna
Hao Zho	Ten Yuan, Mathiwa Cari, Ser La Cingen, Hui Lijaye Yu, Ofi Brauk, Nalo Tigo, Machall Kong, James Le-Therg, Champlayer Yue, Dahi Songalities, Sadar Lance, Jaho Marke, Abhall Shama, Kathy Laude Miller, Katala Sareara, Dani Yuana, Roy Caring, Jonef Bearli, Seri Jawara, Jani Li, Sadar Lance, Jaho Marke, Abhall Shama, Kathy Laude Miller, Katala Sareara, Dani Yuana, Roy Caring, Jonef Bearli, Seri Jawara, Jani Li, Sadar Lance, Sangal Laude Nalo, Kathy Caring, Jani Li, Sadar Lance, Sadar Lance, Sangal Laude Nalo, Kathy Sanga, Kathy Sangal Laude Nalo, Kathy Sangal Laude Nalo
Choe, Á	The second secon
Pitman,	Hart, Haron Daresh, Anam Agawal, Toner Shan, Matan Eyal, Angi Khare, Shenya Bammahan Belle, Lei Wang, Chetan Tekur, Milhr Singiy Kale, Jinlang Wai, Kaucin Sang, Bernana Saeta, Tyler Liechty, Yian, Yano Zhao, Stephan Lee, Pandu Nayak, Doug Fritz, Manish Redy Vayayun, John Maindes, Ni Yang, Martin Micka, Tao Ma, Ligner Bandon, Martin, Yande Can, James Manni, Anni Yin Kim, TXiong, Kai Kang, Kiman Liechty, Nish Trigmanen, Guard Manna, Kango Can, James Manni, Jaon Madolga, Huz Zhao, Guardia Martin, Liao Kai Kang, Carrina Taniha, Jaio Manne, Nish Martin, Yang, Karani Karani, Carrina Taniha, Jaio Martin, Yang, Karani Marin, Yang, Karani M
Laurent	Praduct Shark, Willauk, Respuir, M., Androny (Bosonicz, Jennisma Falonauk), Christian Fannada, Ken Ducles, Isaira Mean, Naios Monchen, Jine Hanar, Ken Ducles, Jania Mean, Jania Sharkan, Jania Markan, Jania Hanar, Hanar, Jania
Maxim H	Roburd Stathner, Vitaly Catalis, Christoph Himschall, Admin Schl, Xingey Federice Xu, Chettan Ahaja, Beh Tail, Arca Stefanoli, Brieg, Kenhar Dhandhania, Marish Kangl, Akahar Gapa, Athana Paralekar, Diny Peta, Jing Zhao, Yisaan Batai, Yashodha Banenai, Osar Ahadaa, Xualin Li, Peter Daneberg, Demis Ta, Aler Pen, Yen Spingon, Ashiga Chen, Beagna Guada, Chattang Kishan Laka, Derit Clin, Yisaa, Handa Li, Kao Hu, Korimi Kangada, Juna Li, Kai Jing Zhao, Yanaan Batai, Yashodha Banenai, Osar Ahadaa, Xualin Li, Peter Jang Kalami Zhein, Yang Manahar Chattang, Hong Kasha Sani, Yore Shan, Saniha Kashar, Kashar Kangada, Janaa Hu, Kai Marana, Saniha Shan, Yang Manahar, Kashar Kashar, Saniha Shan, Yang Manahar, Kashar Manahar, Kashar Manahara, Kashar Hang Manahara, Kashar Manahara, Saniha Shan, Yang Manahara, Saniha Shan, Yang Manahara, Kashara Manahara, Saniha Shan, Yang Manahara, Saniha Shan, Yang Manahara, Kashara Manahara, Saniha Shan, Yang Manahara, Kashara Manahara, Saniha Shan, Yang Manahara, Saniha Shan, Yang Manahara, Kashara Manahara, Saniha Shan, Yang Manahara, Yang Manahara, Yang Manahara, Saniha Shan, Yang Manahara, Yang Manahar
Anders ,	Stankards, Alam Sakohar, Quork Li, Honor Stanhama, Yanghui Wa, Sao Menzo, Milling Data, Carlo Vinyla, Gallageta Kili Umangen Stankards and Stankard
	Matter Constants and Lagrange ISG2, Infinite Infigure ILA, IC Concern Value and Peters Record to ILC CO Concern and Peter ILL 2014 ISG2 Infinite Information (Information Concerns) (I

éle

# Where will AI continue to progress?



Al for science and healthcare As an assistant in scientific and medical innovation



Tool use Goal: enable AI to interact with the world



More factual AI Reduced hallucinations, cite sources, calibration



Al applications More ubiquitous use of Al



Multimodality Al to see, hear, and speak <u>2019</u>



- Can barely write a coherent paragraph
- Can't do any reasoning



- Can write an essay about almost anything
- Competition-level programmer and mathematician



?



Scaling has been the engine of progress in AI and will continue to dictate how the field advances.



#### X / Twitter: @\_jasonwei OpenAl roles: jasonwei@openai.com

Feedback? https://tinyurl.com/jasonwei