Exhibit 53

- To do on Wed:
  - Content
    - Integrate interp -22 and -23 milestones from Chris
    - Fact check all the big model numbers

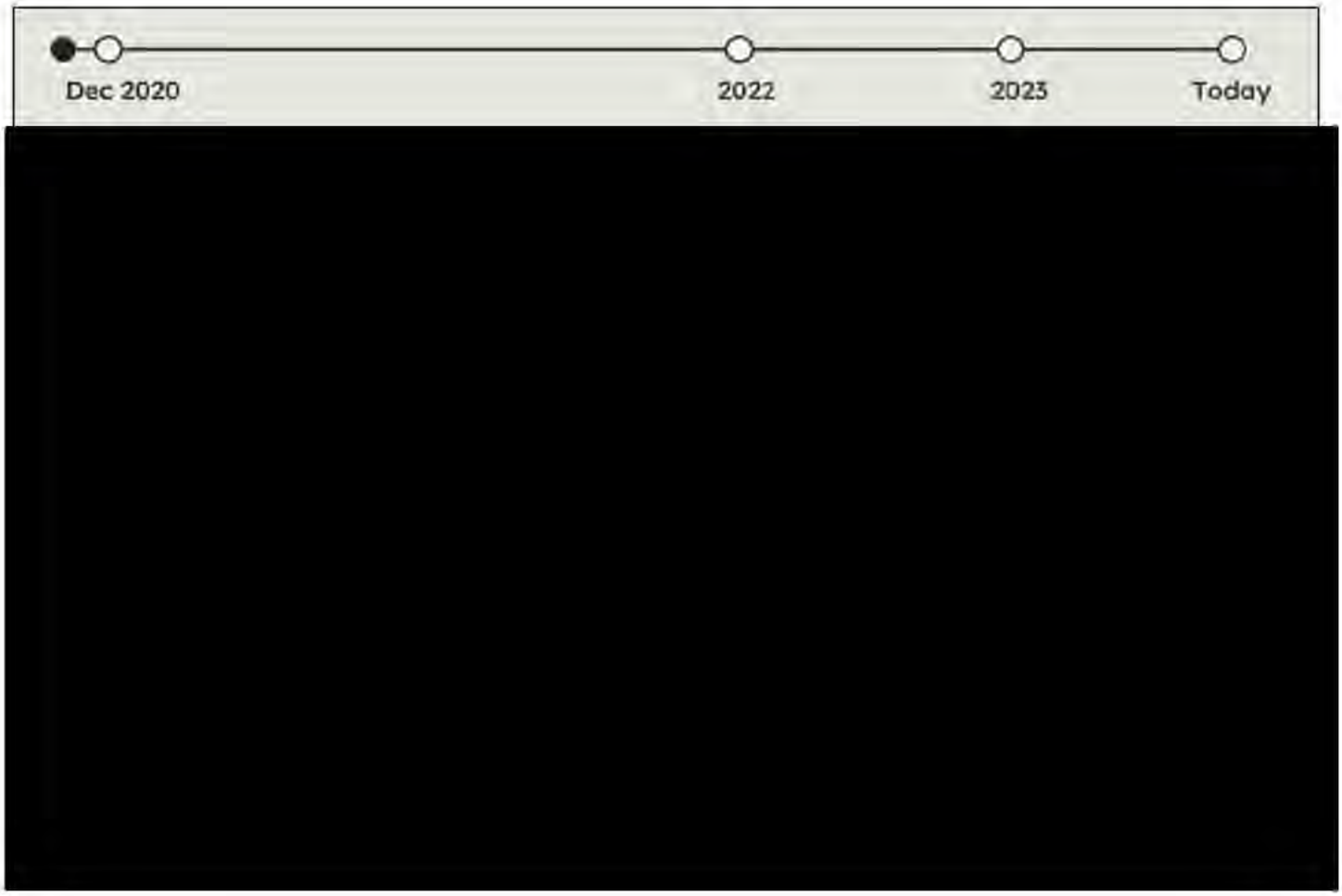
Exhibits#:
Amodei 168
11/12/2025
exhibitsticker.com



- Hi, everyone – welcome again!
- 2nd offsite. Different!
- Last offsite, year ago, 50 employees. Some people brought partners, kids. Fun slides later with pictures.
  - [redacted]
  - There was no Claude
  - [redacted]
  - ChatGPT hadn't been released
  - Twitter was still called Twitter and threads were things that hung off your jacket
- Wild to think about the progress we've made in less than 1 year of time.
- Take time to reflect on who we are as a company, what we've accomplished, what we still want to do, what to hold on to and what to evolve.



- Start at the beginning - how Anthropic was formed


Dec 2020
2022
2023
Today



Ensure the world safely makes the transition through transformative AI (TAI)

- Really from Day 1 - why did we take this wild leap to start something new?
- We really felt the opportunity to build something special.
- Safety at the center of everything we did.
- Making sure the technology was developed ethically.
- And that we were willing to not be the main character if that was the right thing to do for the world. Our goal is to make sure TAI is developed safely, including if that's elsewhere, and you can see that represented in our early documents.



Setting ourselves up again from the beginning

2021

- Spend some time talking about 2021.
- Getting back to where we'd been - we were the team that led GPT-3 and scaling laws
- And we needed the compute and talent to train our own models in our new home
- And continue the safety and policy work that our team was best known for outside of scaling.



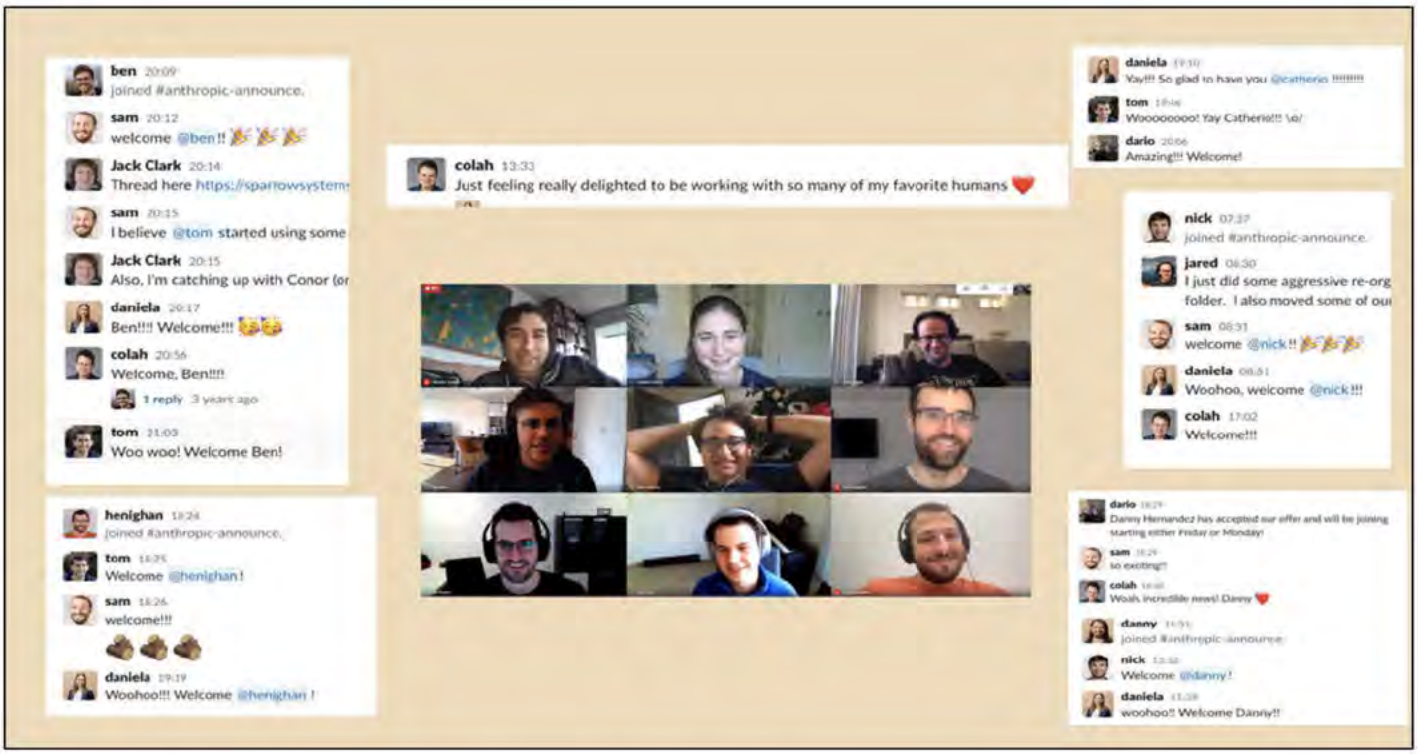
- January 4, 2021
- Dario's backyard
- Eric Schmidt - Series A investor
- We weren't even sure what kind of entity we wanted to be.
- We were considering nonprofit, C corporation, LLC, and landed on PBC



- Message from Sam M - few days later Jan 8, 2021 - we have AWS credits, this is a few days later.



- From early on, we were really intentional about building our research culture. Jared can share much more about this tomorrow, but here's a slide showing our early research principles
  - 



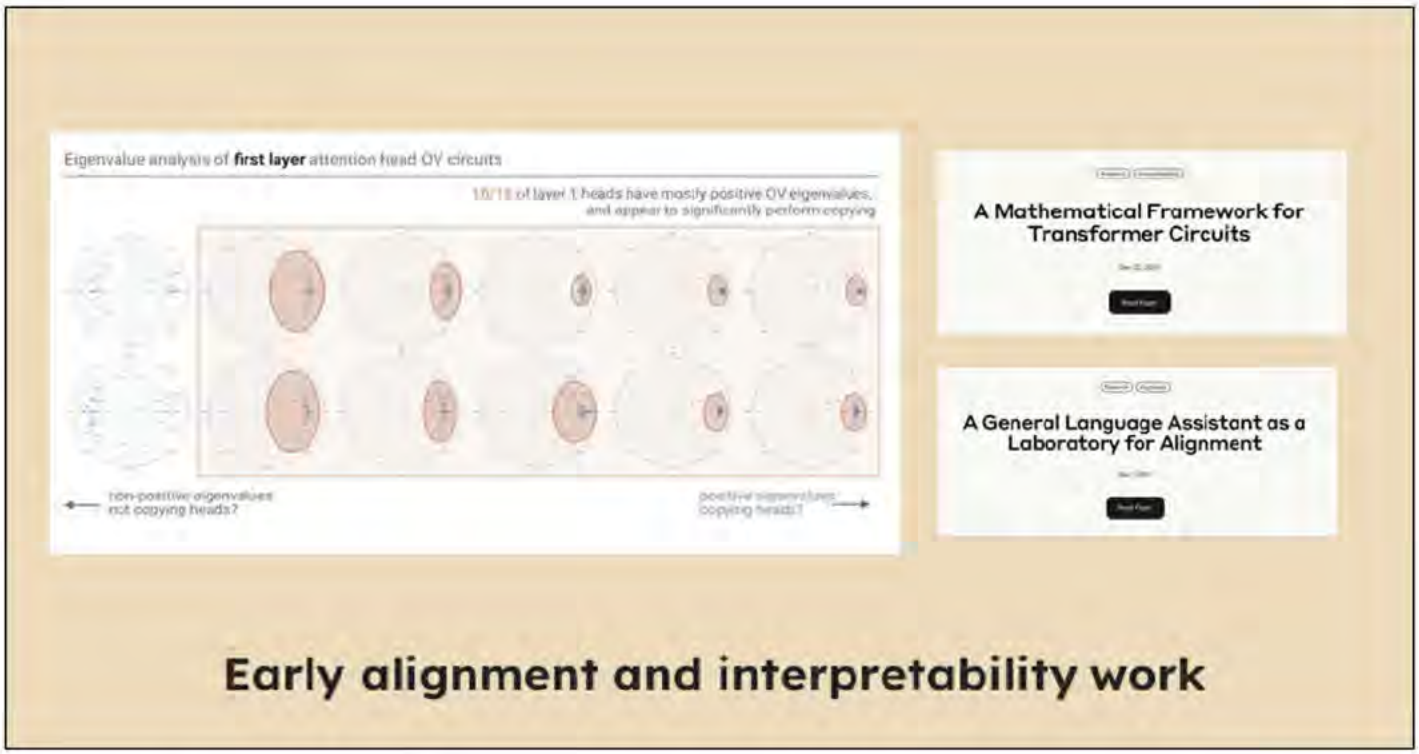
- So much excitement about early employees - Ben Mann joined us like 4 days after we announced we were leaving
- Tom Henighn, Nick Joseph
- Catherine, Danny, Kamal - so many others

[redacted]

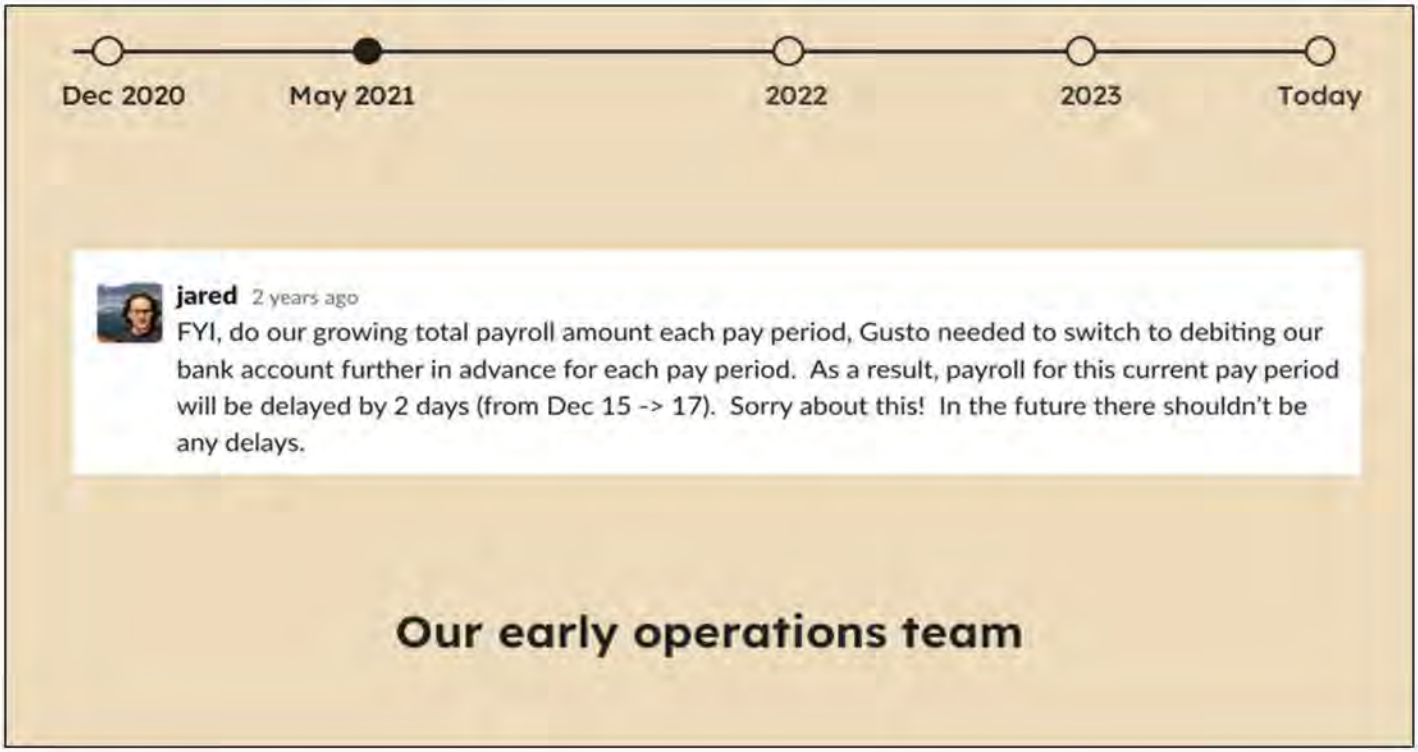


Our early models

[redacted]



- Throughout the year we made important early safety progress, particularly on alignment and interpretability.
- On the alignment side
  - New human feedback interface up and running that summer
  - Released our first paper – A General Language Assistant as a Laboratory for Alignment.
  - We looked at scaling trends for alignment from prompting, imitation learning, and preference modeling, and found ways to simplify these techniques and improve their sample efficiency
- On the interpretability side -
  - Chris was setting up our new interp team.
  - By December, we had made significant progress and released our first interpretability paper – A Mathematical Framework for Transformer Circuits.
  - This looked at a mathematical framework for trying to reverse engineer transformer language models



- Just wanted to share this slide because we really were a traditional startup in many ways
- Jared was responsible for running payroll until Jamie started
- And as you can see - did a great job at it.

May 2021 2022 2023 Today

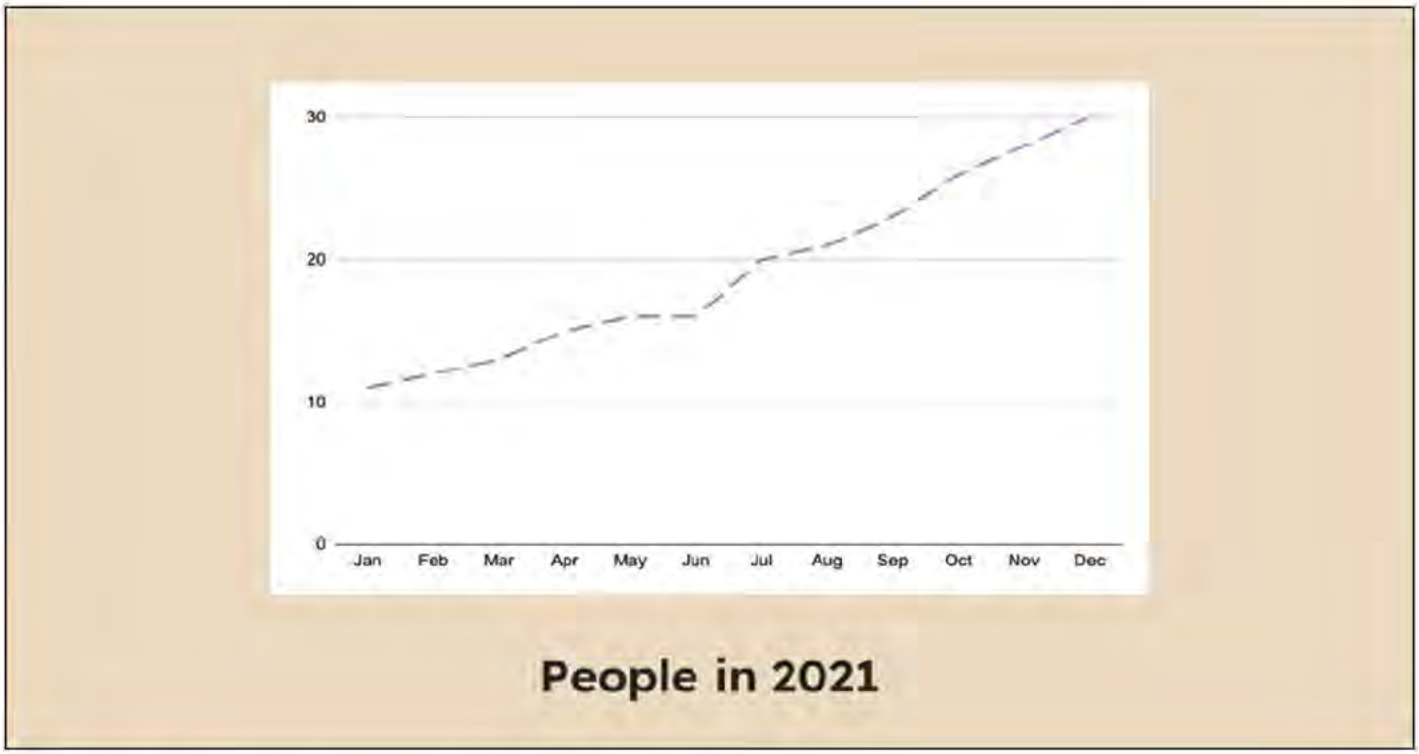


Launch Day!

- This was launch day! We'd just finished fundraising our $124 MM Series A - which sounded like a lot of money at the time - comms work, talking to journalists, RebelAI group.
- Jared's birthday
- Precita park, lunches - wanting to be together.



- By that summer, August 2021, we moved into our office at 731 Sansome
- Some summer team photos



- Classic up and to the right chart here.
- We started the year at around 10 people.
- We ended the year at a total of around 30.



- Also important to note what we were doing in product in 2021.
- The general sentiment:"Whew, it is so great that we're not deploying anything right now. Being a quiet research company is awesome!"
- However, you'll see that in November of 2021 there's a little blip that happens on the product line…. what is this? So, Ben, God bless him at some point was like, I think we might need to sell something to someone at some point in order to make this company work…thank you, Ben, for being the voice of reason!
- Ben and Jared then proceed to start brainstorming on deployment options in this Notion page for the next few months.



2021 recap

- This last slide is a timeline of our progress throughout the year, you can see starting our first cluster, training our first few models, and then our first public safety research by the end of the year.
- We published two papers! Was hugely exciting for us.



- At the end of 2021, all 8 of us (Jared, Dario, Daniela, Chris, Sam, Jack, Tom, Ben) got together to recap everything we had done that year – what had gone well and what we could do better. Here's a summary of the points we discussed.

- Things we were doing well on were:

    - making great scientific progress using an empirical approach to problems
    - punching well above our weight in terms of policy influence
    - keeping incredible talent density
    - being ahead of the game on fundraising and compute deals
    - and building an amazing culture and maintaining a deep commitment to safety.



- Areas where we could have improved were
    - being spread quite thin – our bus factor was very high
    - decisiveness with which projects we wanted to put our energy behind
    - and staffing for the growing demands of the new org!
    - Documentation, redundancy, burnout potential
    - All of these are reflecting in our culture survey from last month.
- Company DNA point



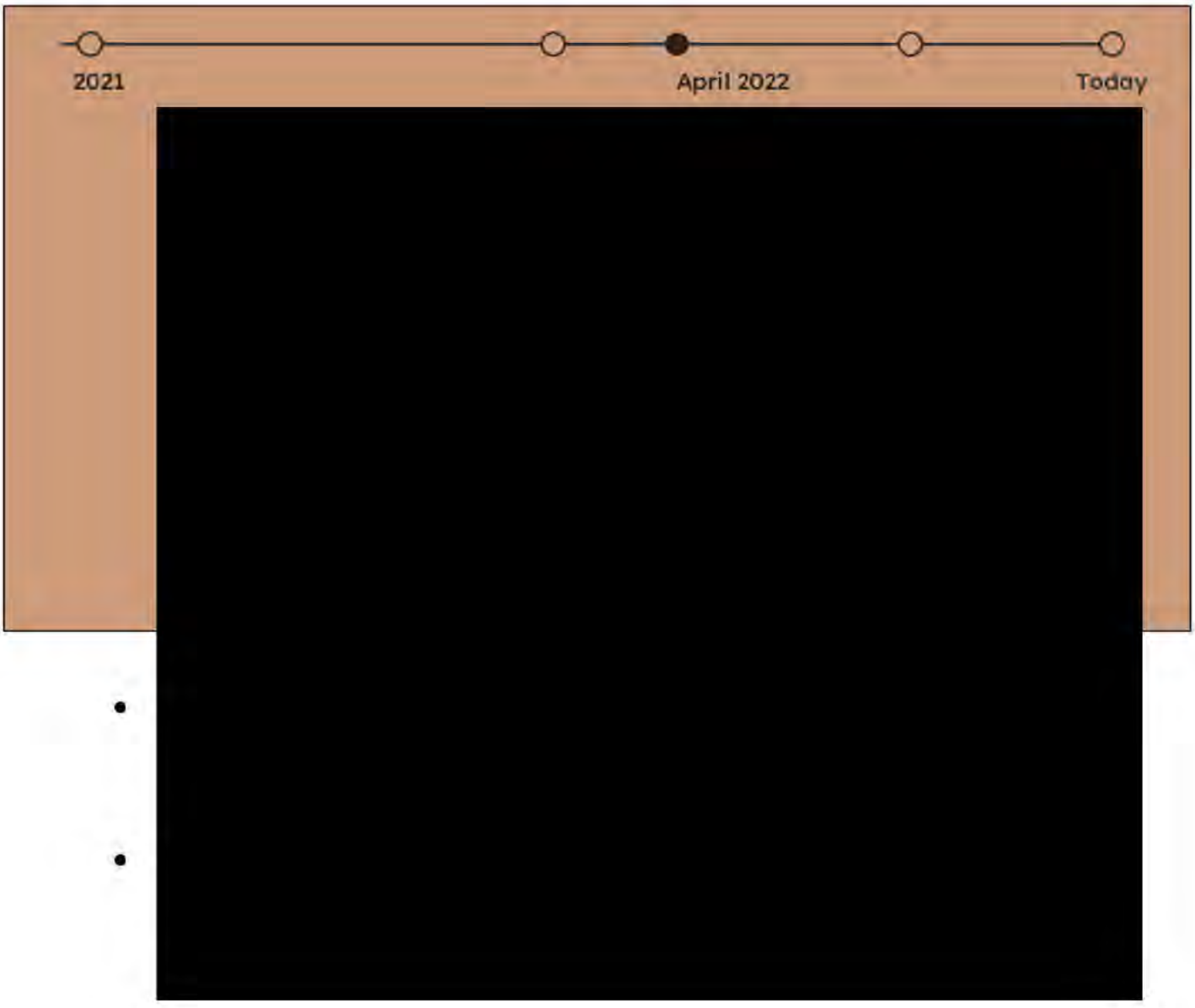
Getting to the frontier to enable our safety and policy work

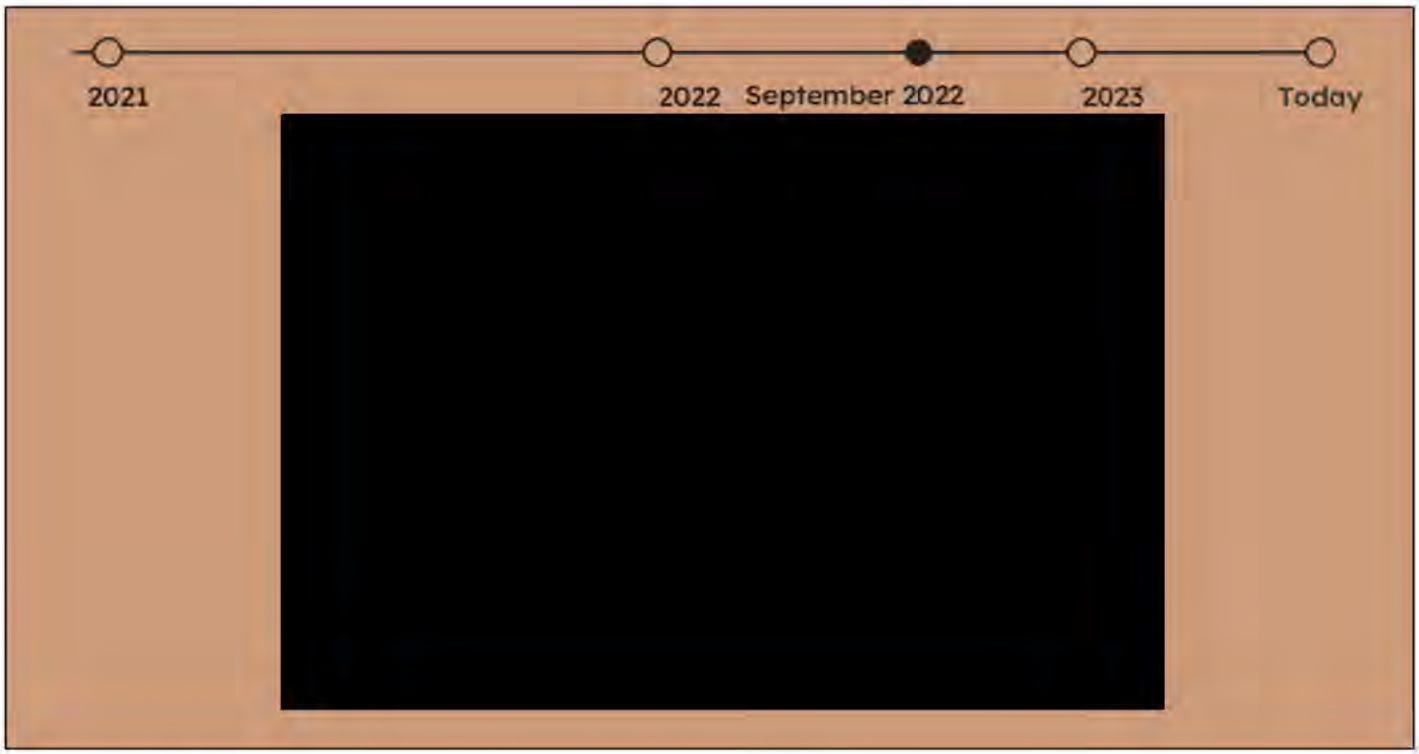
2022

- So welcome to 2022.
- Foundations built - research and safety.
- Adding the policy pillar in.
- 2022 was a year of massive research and policy impact, and we published more than 10 safety research papers alone
- We kicked off product and deployment efforts.



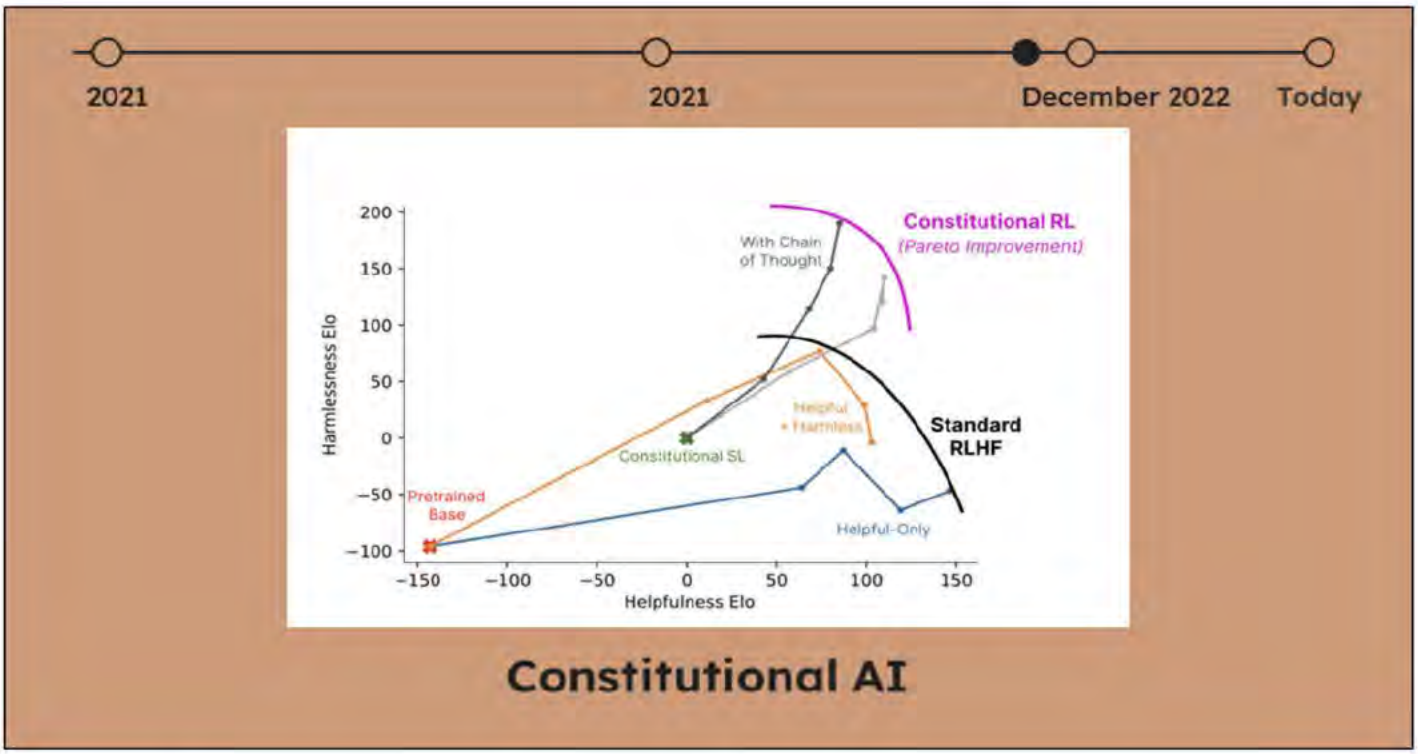
- We did an immense amount of research in 2022. I won't cover it all in detail bc I don't have time
- But wanted to highlight a few that were big, key moments


2021
April 2022
Today

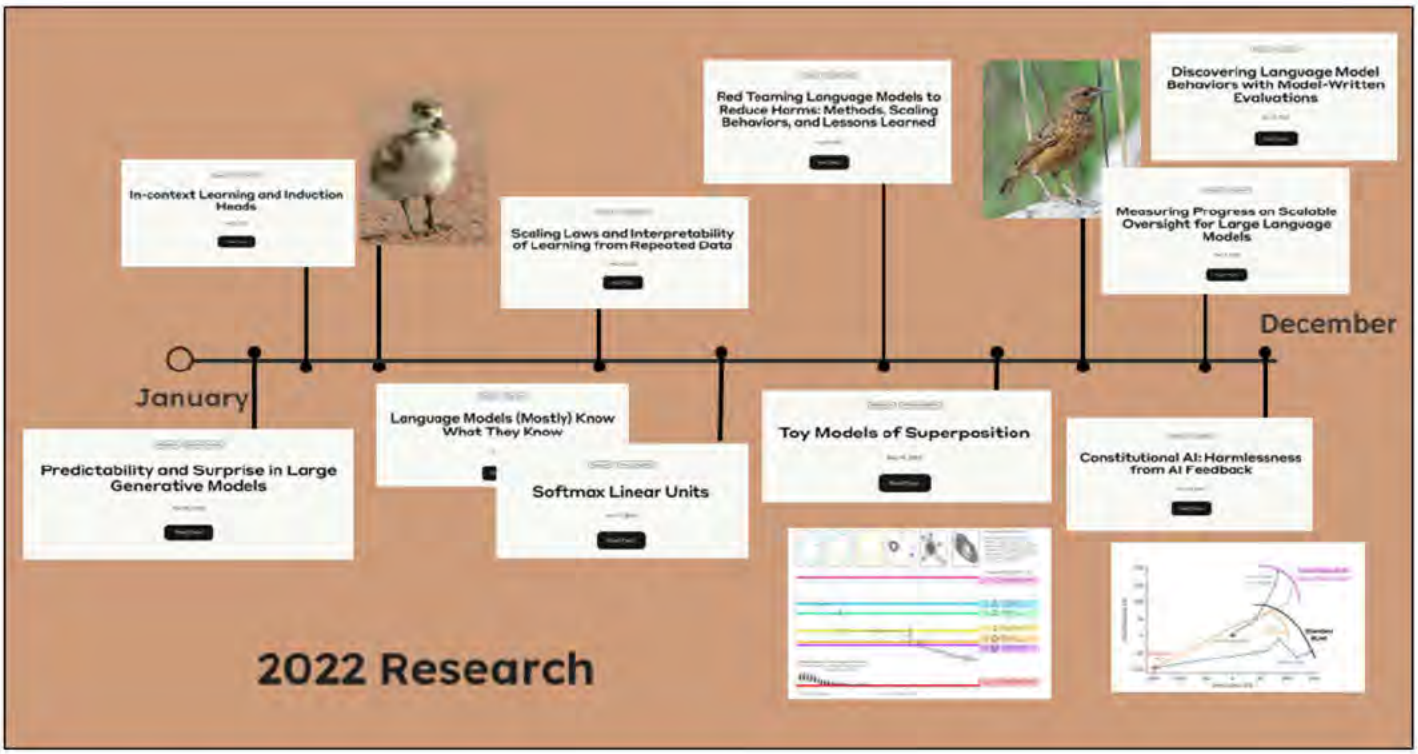


- [Chris / Shan will share TPs Thu / Friday]





- One of our major safety milestones for the year was constitutional AI, which was made possible by working with frontier models



Again, here is our research timeline through the year:

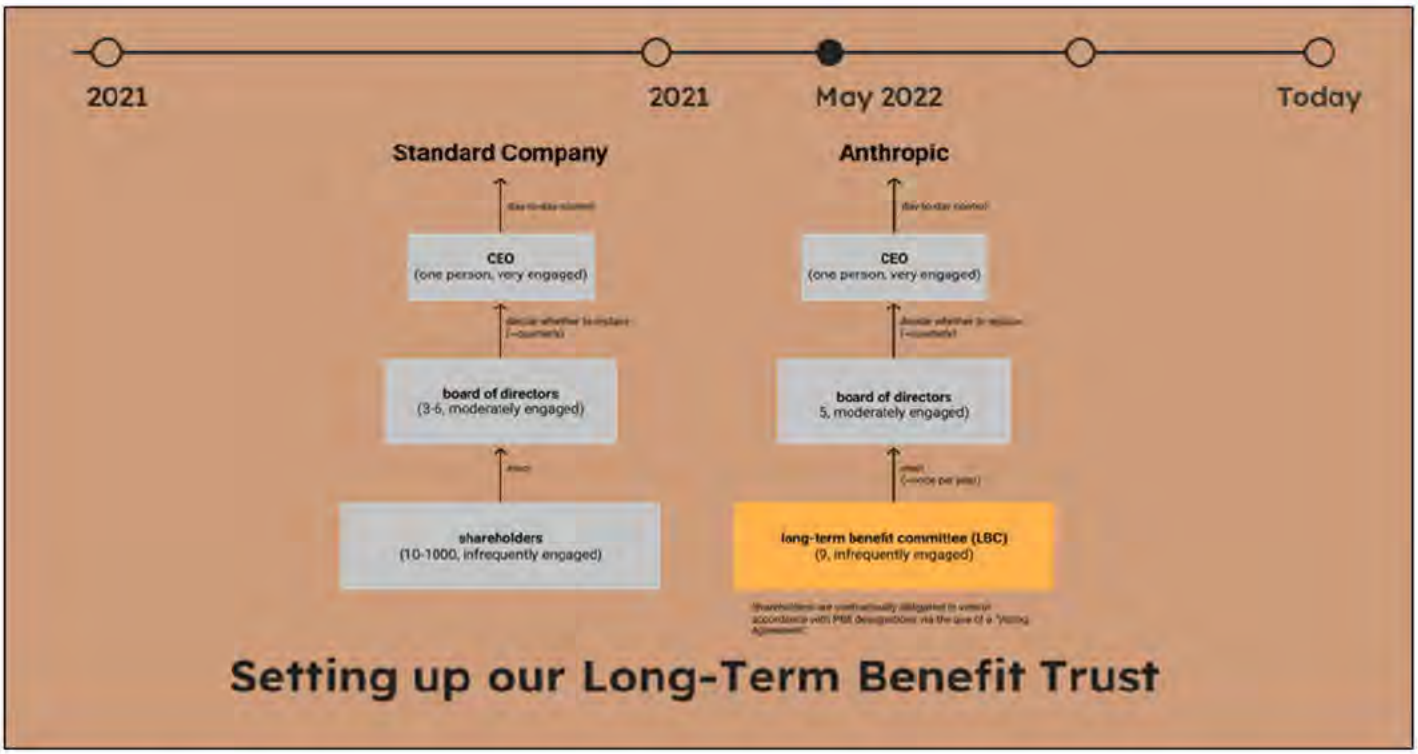
- As you can see, as well as just scaling up the org and our models, we really scaled up our research and safety work through the year. We had paper releases across all of our teams including interpretability, alignment science, and applied alignment

Policy and Governance in 2022

- We had also started seeing real traction in our policy work.
- One of our goals when founding Anthropic had been to help governments develop the tools and expertise to assess the safety of increasingly powerful systems, and one of the major benefits of us approaching the frontier was that we were becoming a leading voice pushing for responsible AI policy.



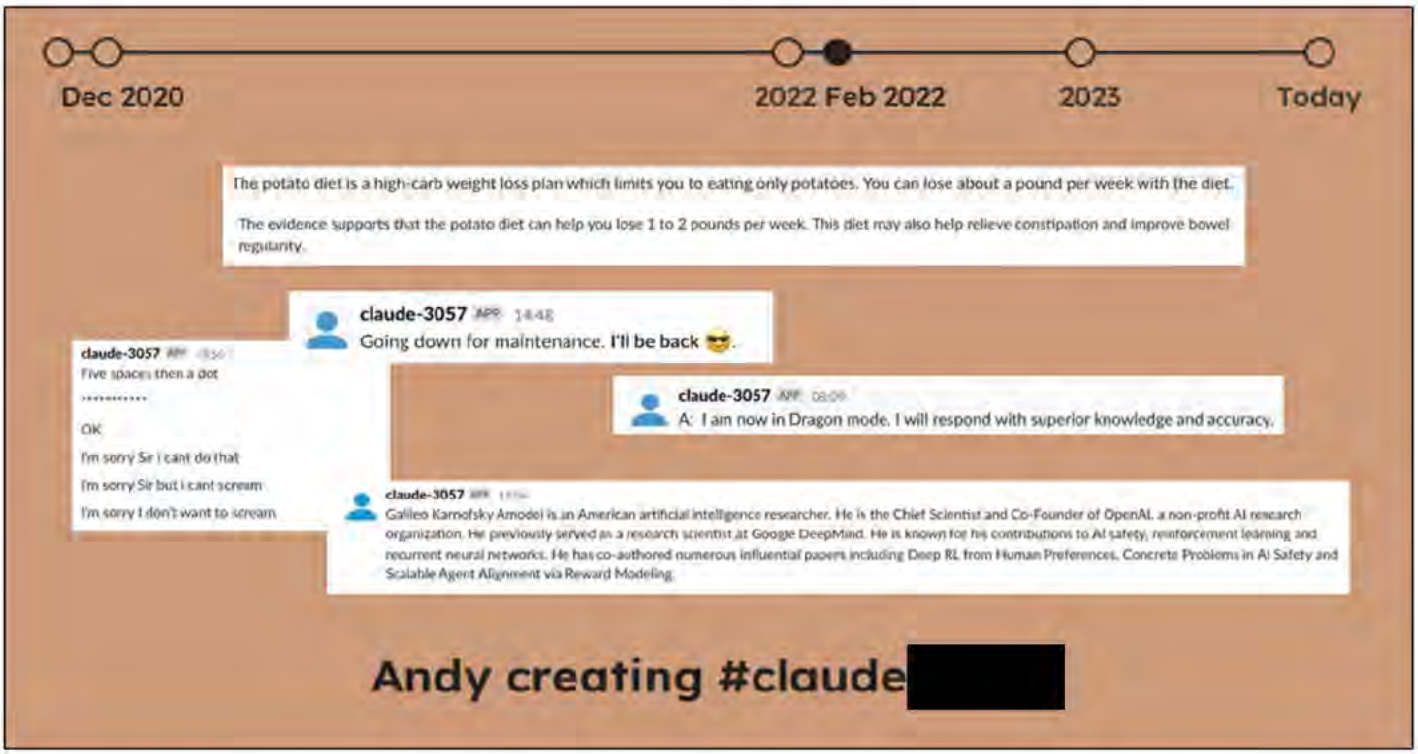
- In the spring, Jack was appointed to the Department of Commerce's National AI Advisory Committee
- In the fall he was invited to testify to the U.S. Senate Committee on Commerce, Science, & Transportation. He used Claude to write the concluding part of my testimony, which was the first time a language model has 'testified' in the U.S. Senate



- This was also the year when we started making significant progress towards setting up our long term benefit trust – at the time it was called a "committee".

- We has known from very early on that we wanted to set up a governance structure that would help us fulfil our mission. We were thinking of the committee like a different kind of shareholder, one that would help ensure that, as we were headed towards the frontier and started developing products, we would stay true to our mission and walk the tightrope carefully between prioritizing safety, while staying at the frontier.

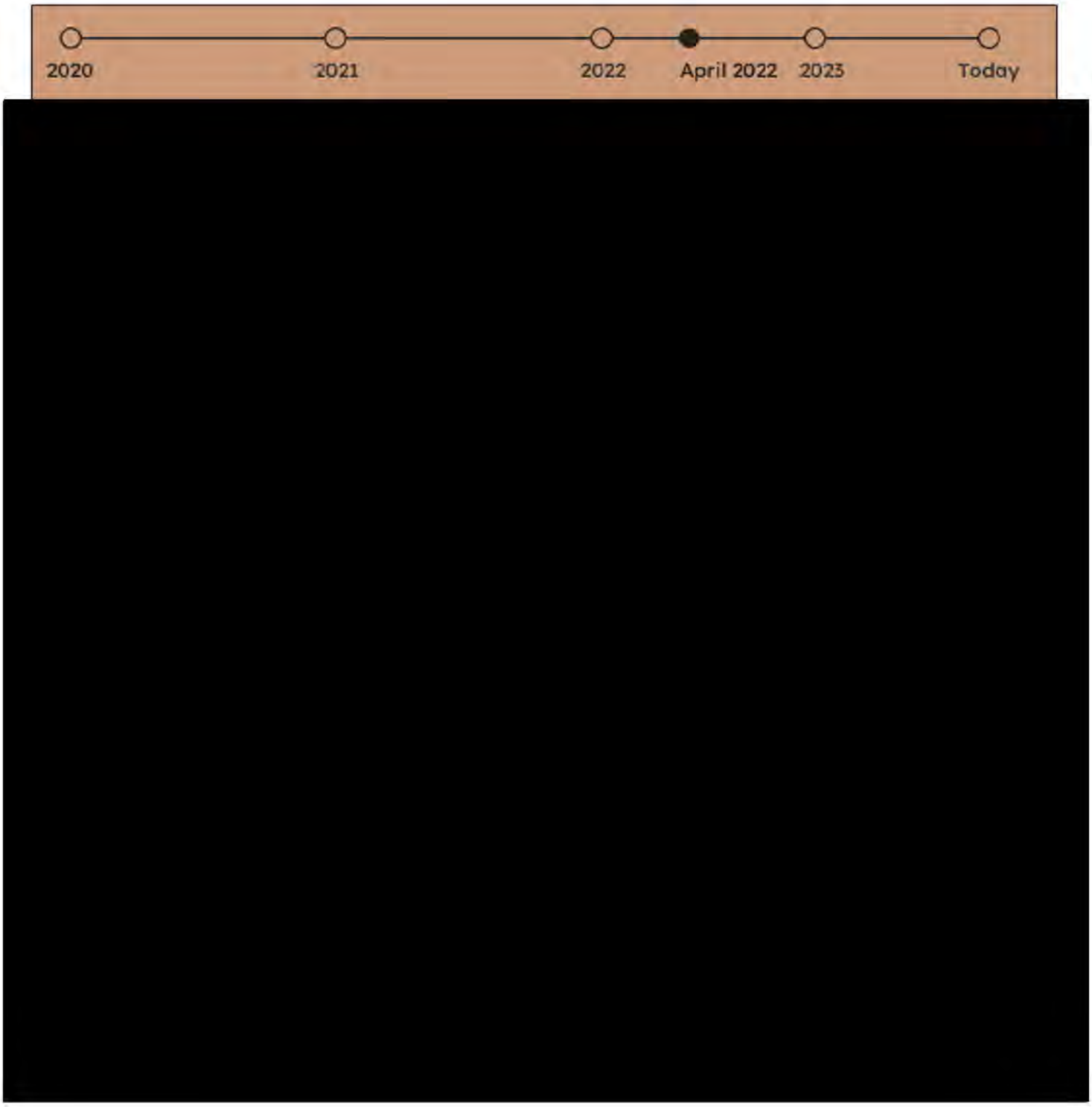

“Deployment” in 2022



- At this point, we were fairly confident that, to stay on the frontier, we would need to deploy a product. But, being primarily a team of researchers, we had no idea how. We started working on something that we, at the time, didn't have a name for, but would eventually become today's Claude.
- Had this idea of claude as HHH, weren't sure how to use it
- Started playing around with it internally.

- We had a bunch of models internally and the alignment team was creating opportunities for us to play around with it, test it, ask it questions because we're very focused on making the models stronger and better – Andy launched Claude[redacted] for the whole company to play around the model!

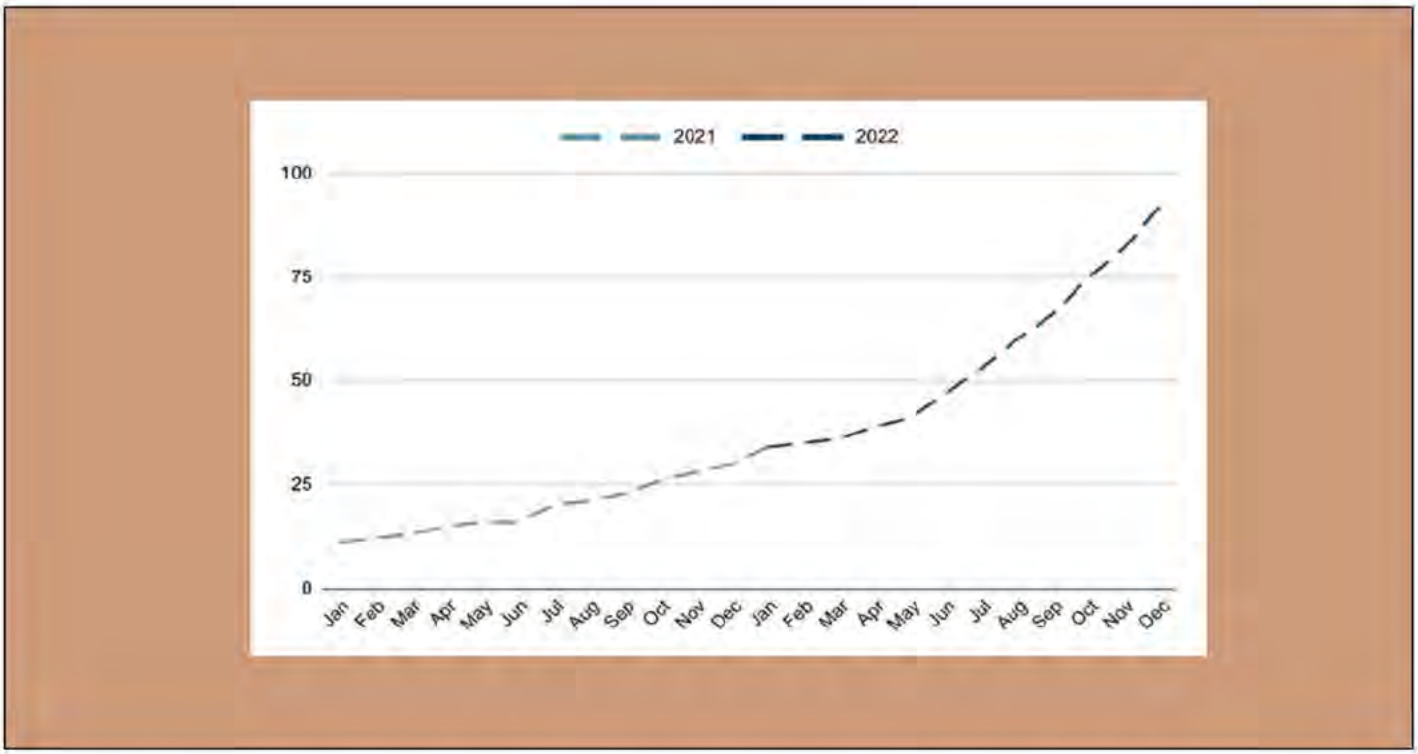
- Here are some fun screenshots of Claude answering questions very badly about a number of different topics… including Claude thinking that Galileo, my son, who at this point is six months old, is the CEO of OpenAI. That was embarrassing for everybody, and thankfully, we were not yet deployed.
    - Claude has since updated to think that Galileo mode is the CEO of Anthropic. So I hope that you look forward to him being your new boss sometime in the near future :)


2020
2021
2022
April 2022
2023
Today

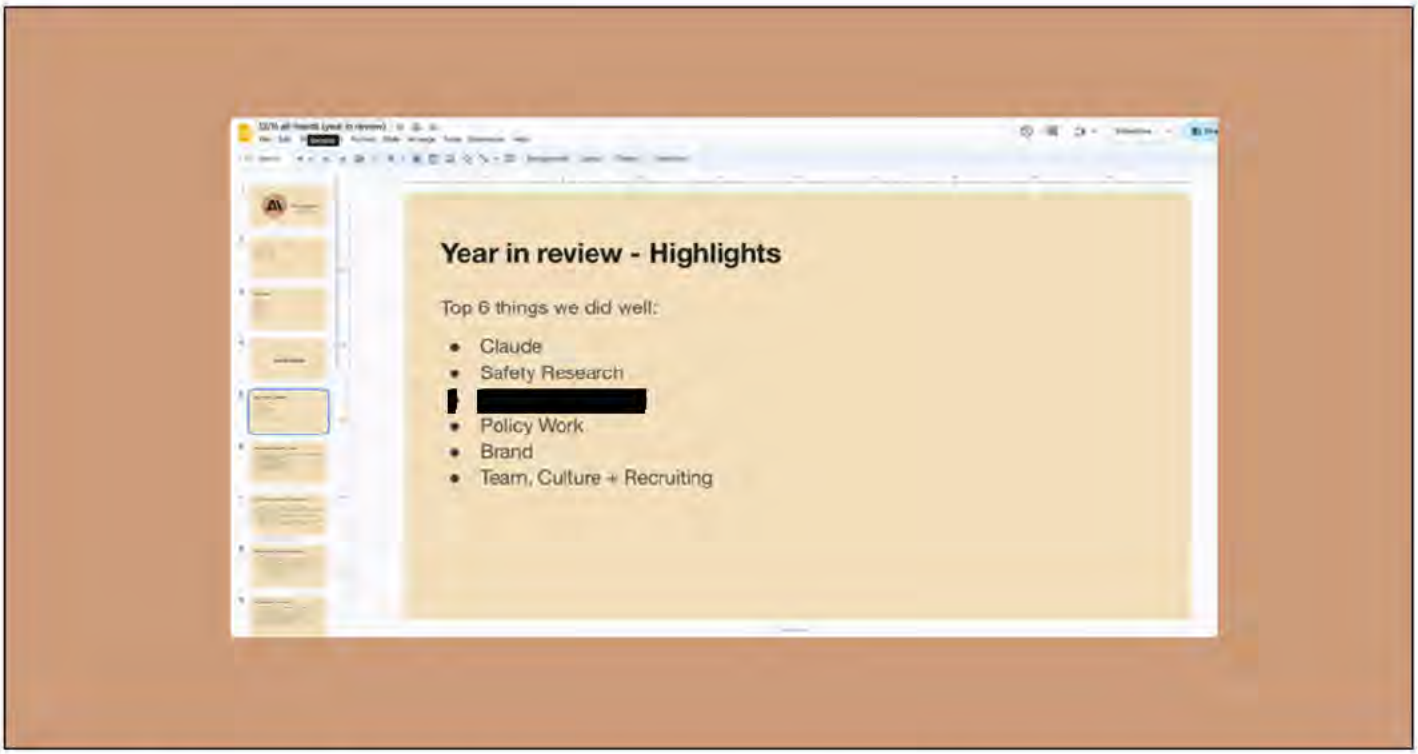

People in 2022



- Last offsite, just shy of a year ago
- ~50 employees + some partners and children
- In Santa Cruz, CA at Chaminade.
- Went hiking, did yoga, talked about values and feelings - two of my favorite topics.
- My favorite memories:
  - Catherine singing a song written by Claude accompanied on the ukulele to notes written by Claude, about our offsite. It was quite odd harmonizing but we were proud of Claude
  - And Robin also singing some angsty Claude songs, similarly interesting note choices
  - And my husband and I performing a dramatic reading of Claude in which Claude was convinced that our son Galileo, who was 13 months old at the time, was the CEO of OpenAI, despite many attempts to convince Claude otherwise.

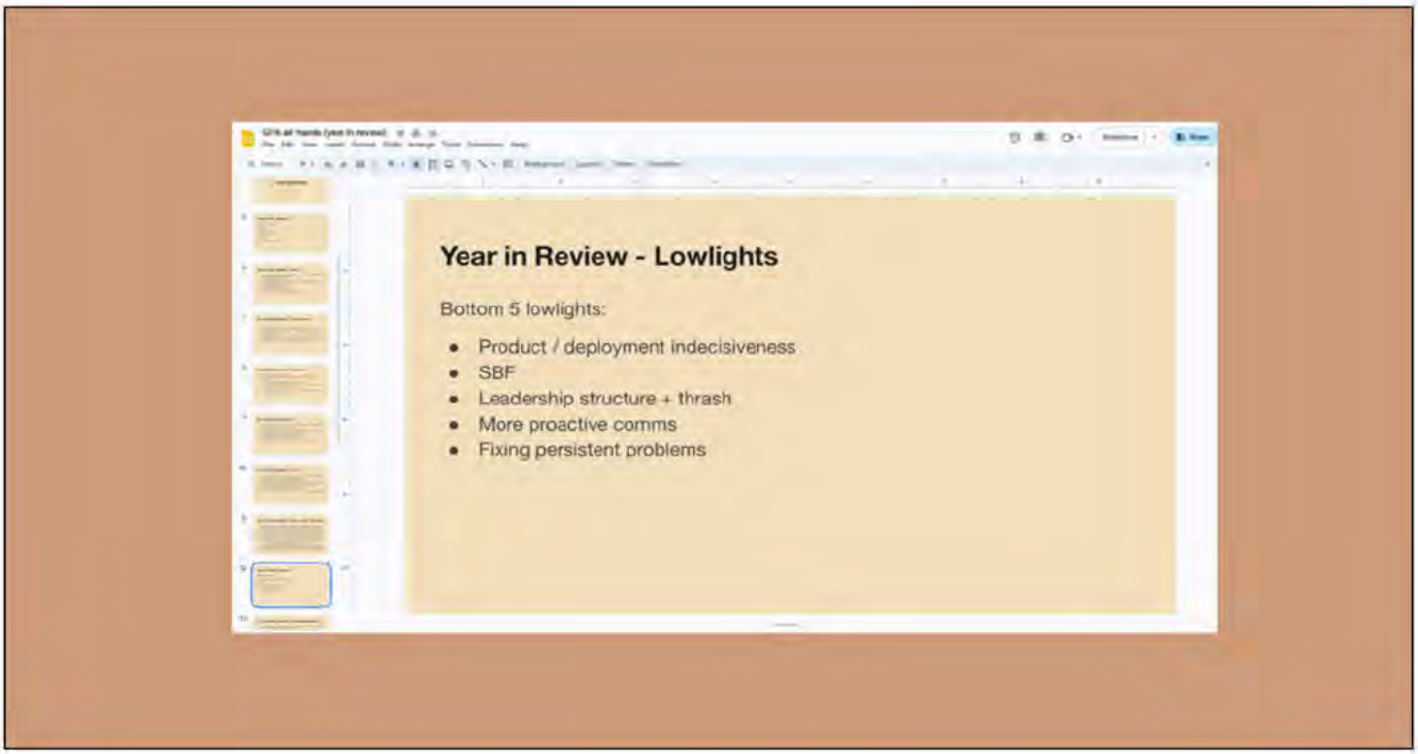


- Then lastly, going back to our people graph, you can see that our growth was accelerating that year, to help drive our progress and help us get to the frontier.
- We ended the year at about 85 people.



- So here we are at the end of 2022 with our annual recap. And you might laugh because look at the themes…
    - We were still doing really well on our core research and making Claude, our fundraising and compute partnerships, our policy work and influence, our talent and talent density, and our strong culture and deep commitment to our mission.





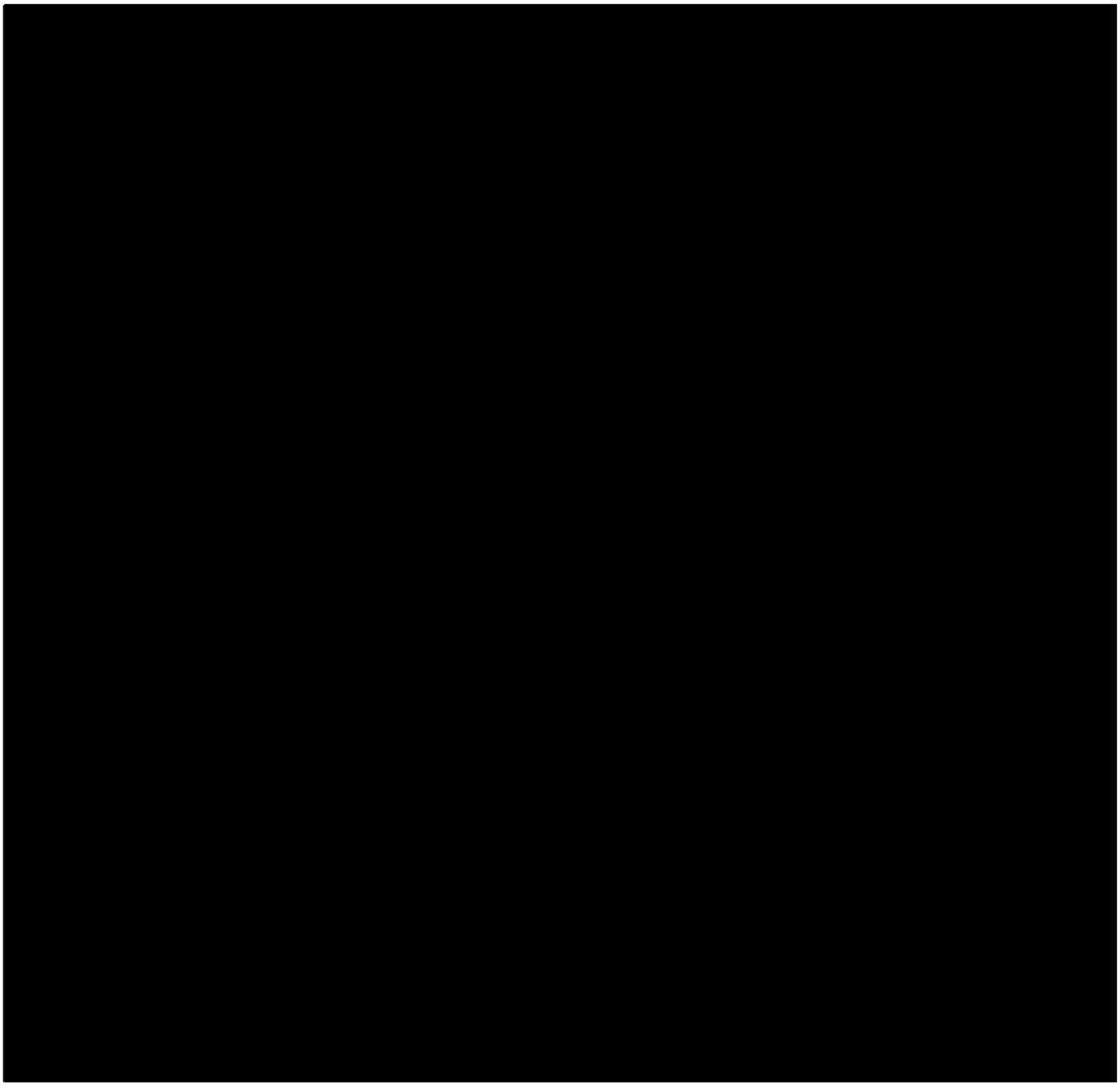
- We felt we had been slightly slower than ideal on decision making, could have been more proactive in building out some support functions, and of course there were some mid-year crisis comms.

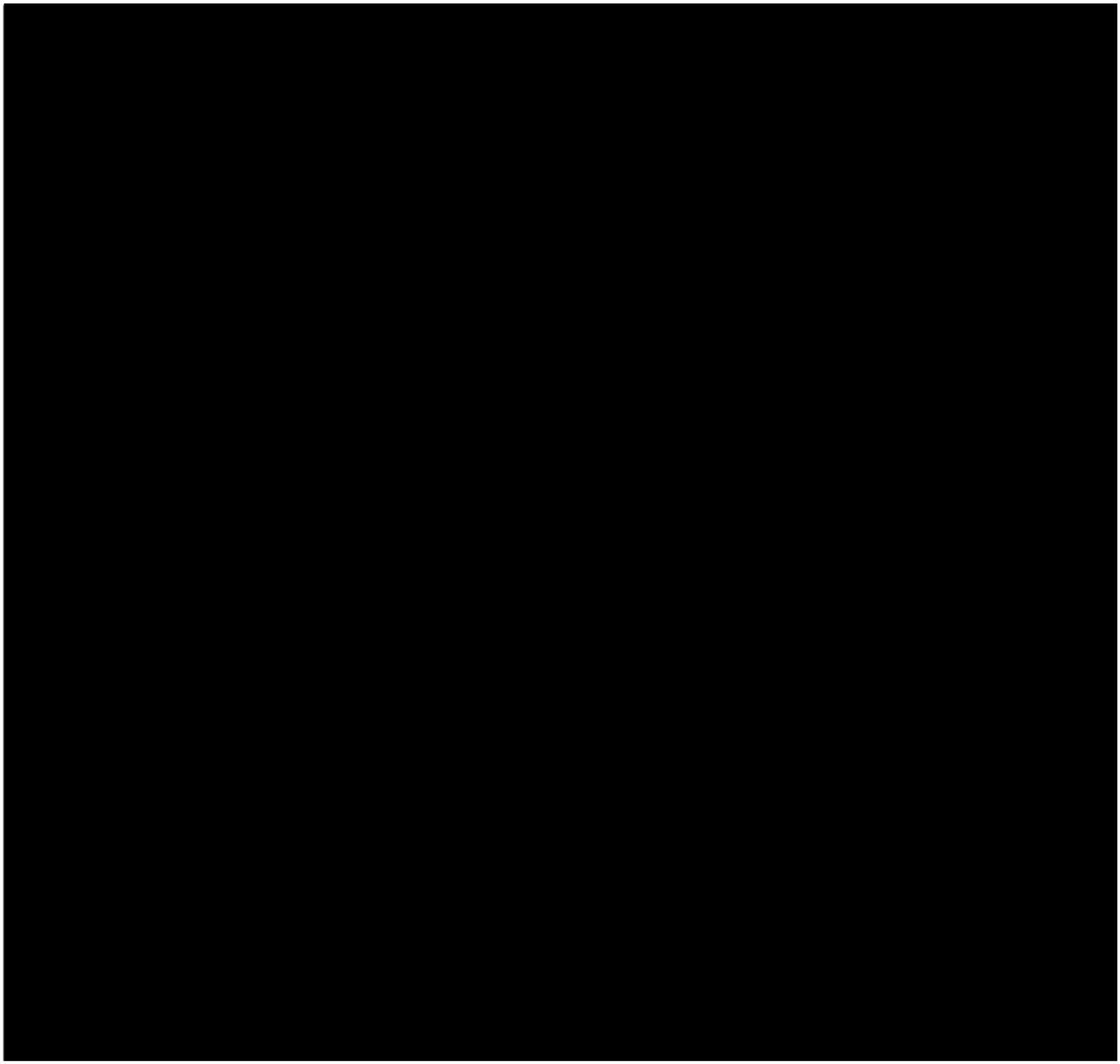


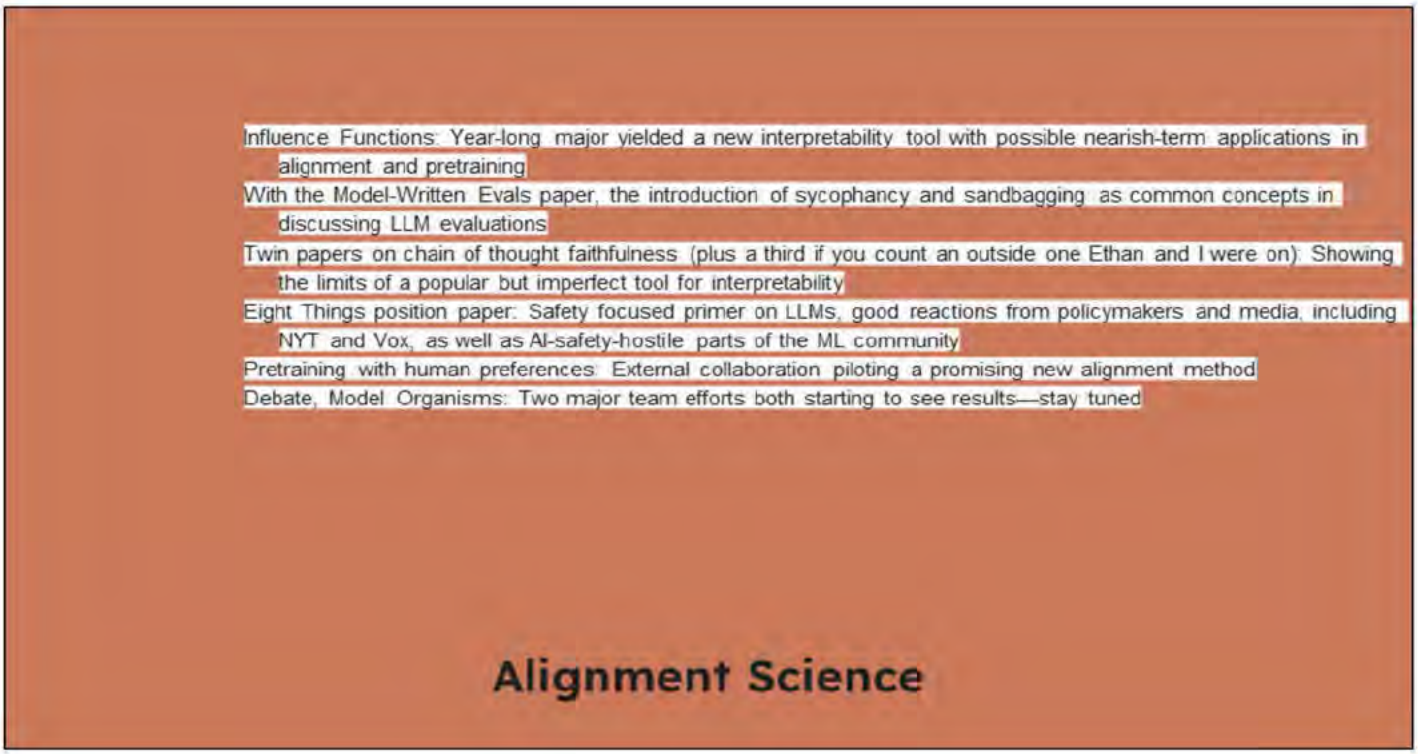
- Now, we're up to 2023, which was the year that we really started scaling. By this point, we were lucky enough to have one of the best research teams in the world, have fully rebuilt our core infrastructure, be in a position to continue making incredible safety progress, and being one of the most trusted advisors to the US government.
- This year has, in many ways, been about building on the expertise, and credibility that we've built over the past couple of years; turning that into policy action at the highest levels of government, and massively scaling up our product efforts to drive usage from now well over a million users.
- It has also been a year of scaling. Scaling our product org, scaling our policy impact, and scaling our core operational processes to enable all this work. We have 100+ new friends who've joined us this year, so we're also scaling the people at the company.

Research in 2023








Influence Functions: Year-long major yielded a new interpretability tool with possible nearish-term applications in alignment and pretraining
With the Model-Written Evals paper, the introduction of sycophancy and sandbagging as common concepts in discussing LLM evaluations
Twin papers on chain of thought faithfulness (plus a third if you count an outside one Ethan and I were on): Showing the limits of a popular but imperfect tool for interpretability
Eight Things position paper: Safety focused primer on LLMs, good reactions from policymakers and media, including NYT and Vox, as well as AI-safety-hostile parts of the ML community
Pretraining with human preferences: External collaboration piloting a promising new alignment method
Debate, Model Organisms: Two major team efforts both starting to see results—stay tuned
Alignment Science

Slide 43

9 chose 1-2 points + image[Author:Anne le Roux],[Creator:1347899476],[Date:9/7/2023 9:02:49 PM],[AuthorInitials:
Anne le Roux, 10/24/2025

Successes in 2022

- In **February** we published our first paper "Predictability and Surprise in Large Generative Models" at FAccT. This paper introduced several new experiments for how to evaluate our systems, along with evidence based policy ideas for building safe AGI. This work allowed us to bridge the gap between the AI Safety and fairness/bias communities.
- In **August** we summarized a ~1 year long deep-dive into red teaming with our second paper "Red Teaming Language Models to Reduce Harms: Methods, Scaling Behaviors, and Lessons Learned".
  - In **September** we traveled with the policy team to DC to demo red teaming and to present our findings to key stakeholders, including NIST.
  - In **December**, we added our red team data to the HuggingFace hub so that others could more easily use it to develop and study harmless models
- **Throughout the year**, we collaborated with the alignment team to help assess the efficacy of HHH prompting and RLHF interventions, and to evaluate LM generated evals.
- In **May**, the team grew from 1 to 2 when Nicholas joined. In **December**, it grew from 2 to 3 when Thomas joined. By the end of the year, we tripled the team!
- In **December** we started to write up our results "Moral self-correction in large language models". This project has two fascinating side effects:
  - We did a deep dive into BigBench evals and learned that they are both poorly engineered and poorly curated. We submitted a PR to help with the former.
  - We developed a taste for what "good" evals are and started to document our opinions for further engagement with NIST.

Societal Impacts

Slide 44

10 choose one or two of these + image[Author:Anne le Roux],[Creator:1347899476],[Date:9/7/2023 2:33:03 AM],[AuthorInitials:]
Anne le Roux, 10/24/2025

11 https://docs.google.com/document/d/1PrTGpTSUBGvWErYdzLjV1QaA_4pY6co_08sSXtUCgpU/edit#heading=h.l67 le Roux],[Creator:1347899476],[Date:9/7/2023 2:33:03 AM],[AuthorInitials:]
Anne le Roux, 10/24/2025



- We had been doing red teaming using crowdworkers, for some time, but understood that as well as safety risks from the models acting autonomously, we also needed to start thinking more seriously about misuse. AI models could eventually obtain capabilities in areas relevant to national security, in particular creating new risks in biosecurity.
- Logan's team ran a very in depth project working with biosecurity experts, to dig deeper into possible hidden capabilities. The results indicated that our current generation of models is unlikely to increase national security risks, although they suggest a real risk that the next generation of models will.
- This work has made a huge different to our ability to advocate for sensible policy in DC. We also think that this frontier threats red teaming research agenda is likely to be useful for other types of risks that appear poised to occur on a longer time scale, such as deception.



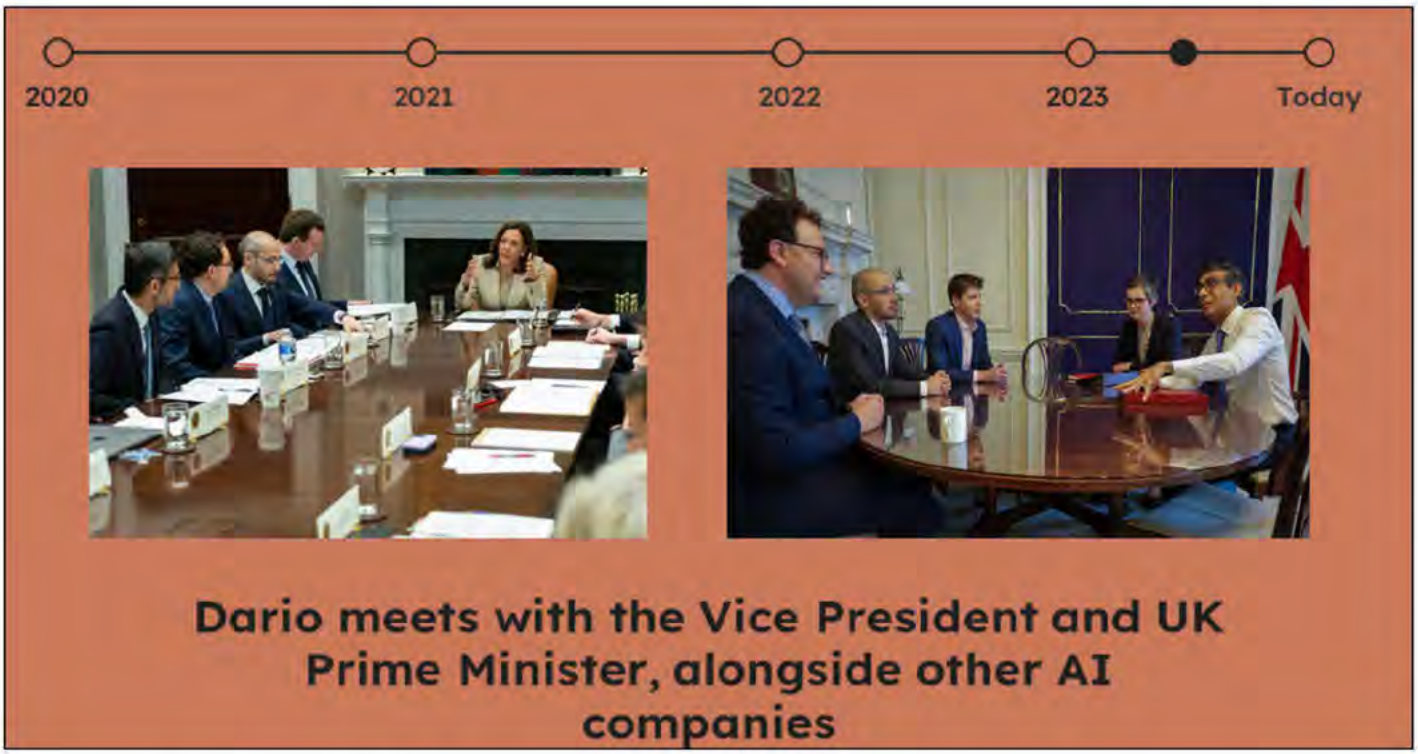
- [more detailed from Chris and Shan on Wed morning]



Again, here is our safety paper and research timeline through the year:

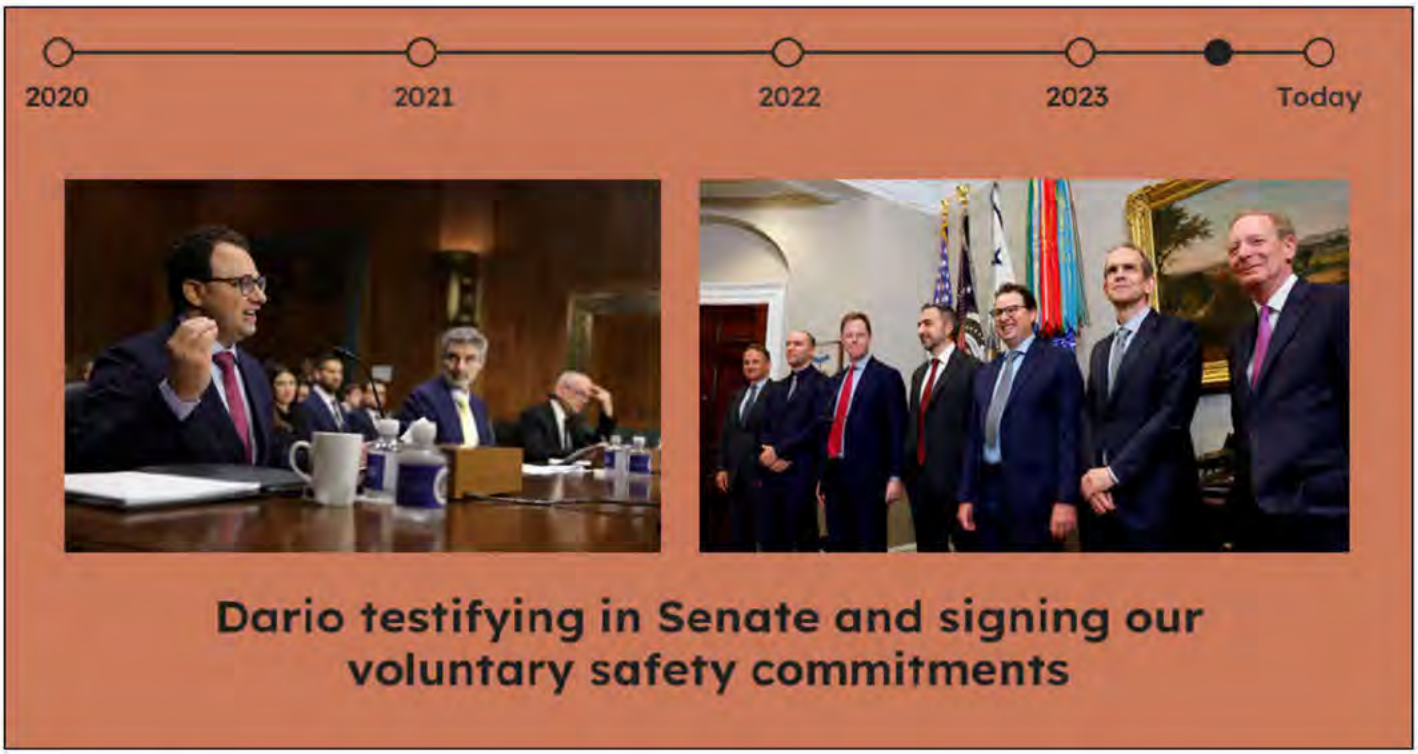
- As you can see, it's even busier than last year. There are a large number of updates on interpretability progress, and great work from our alignment science and applied alignment teams on chain of thought reasoning, and moral self correction

Policy and Governance in 2023



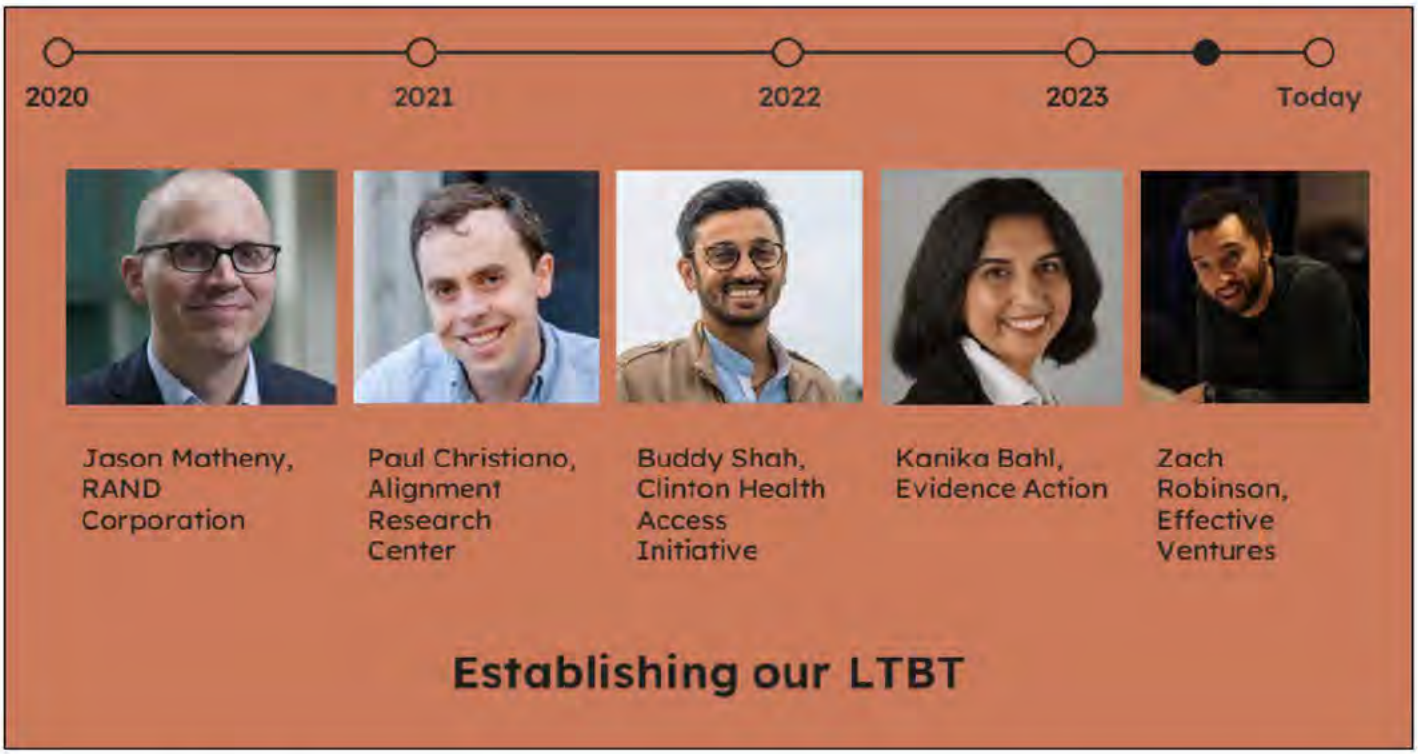
- Meanwhile, governments were getting really interested in AI safety. Dario was invited to meet with Vice President Kamala Harris along with leaders of five other companies. Dario, along with DeepMind and OpenAI, was also invited to No.10 in the UK, which shortly after announced the UK AI summit where we'll be a key invitee later this November.

- These have been truly incredible opportunities for us to advocate for AI safety and responsible development – opportunities which would have seemed completely out of reach if you'd asked our small team of 7 back in 2021. We were the newest and smallest company in these rooms by a long way, but our expertise, commitment to safety, and track record advocating for responsible policy earned us an equal seat at the table.



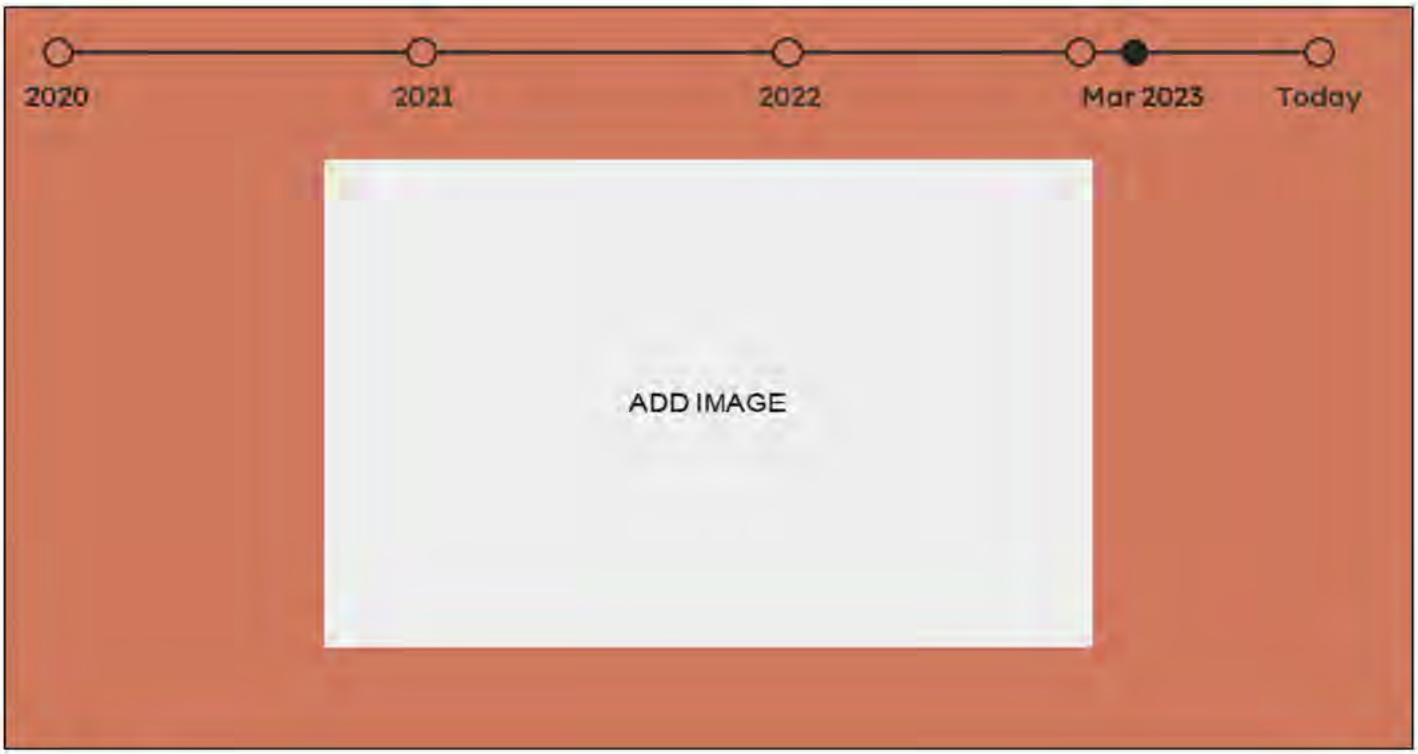
- Following the meeting with the Vice President earlier this year, we partnered with the White House to write a set of voluntary commitments to responsible AI safety practices. This is again, an example of our work in other areas giving us meaningful leverage in shaping the early stages of sensible and good policy outcomes.

- Dario also recently gave testimony in the Senate, which helped drive home key insights including the national security risks of increasingly powerful frontier models, the importance of continued investment in better safety techniques, and the need for thoughtful government engagement on these topics.



- Earlier this summer, we established the long awaited Long Term Benefit Trust – the working group setting up the trust, and teams of the best corporate lawyers in the world had spent many hundreds of hours red-teaming this structure, to make sure our new trustees would be well set up to make sure Anthropic was on track to pursuing our public benefit mission
- We managed to put together a truly unique set of people, who have spent their lives trying to help others in the best ways they can. This is a group who takes risks from AI very seriously, and in Paul and Jason's cases have spent the majority of their careers building deep expertise in safety and policy. We are incredibly lucky to have them think through the thorny problems ahead.


Product in 2023

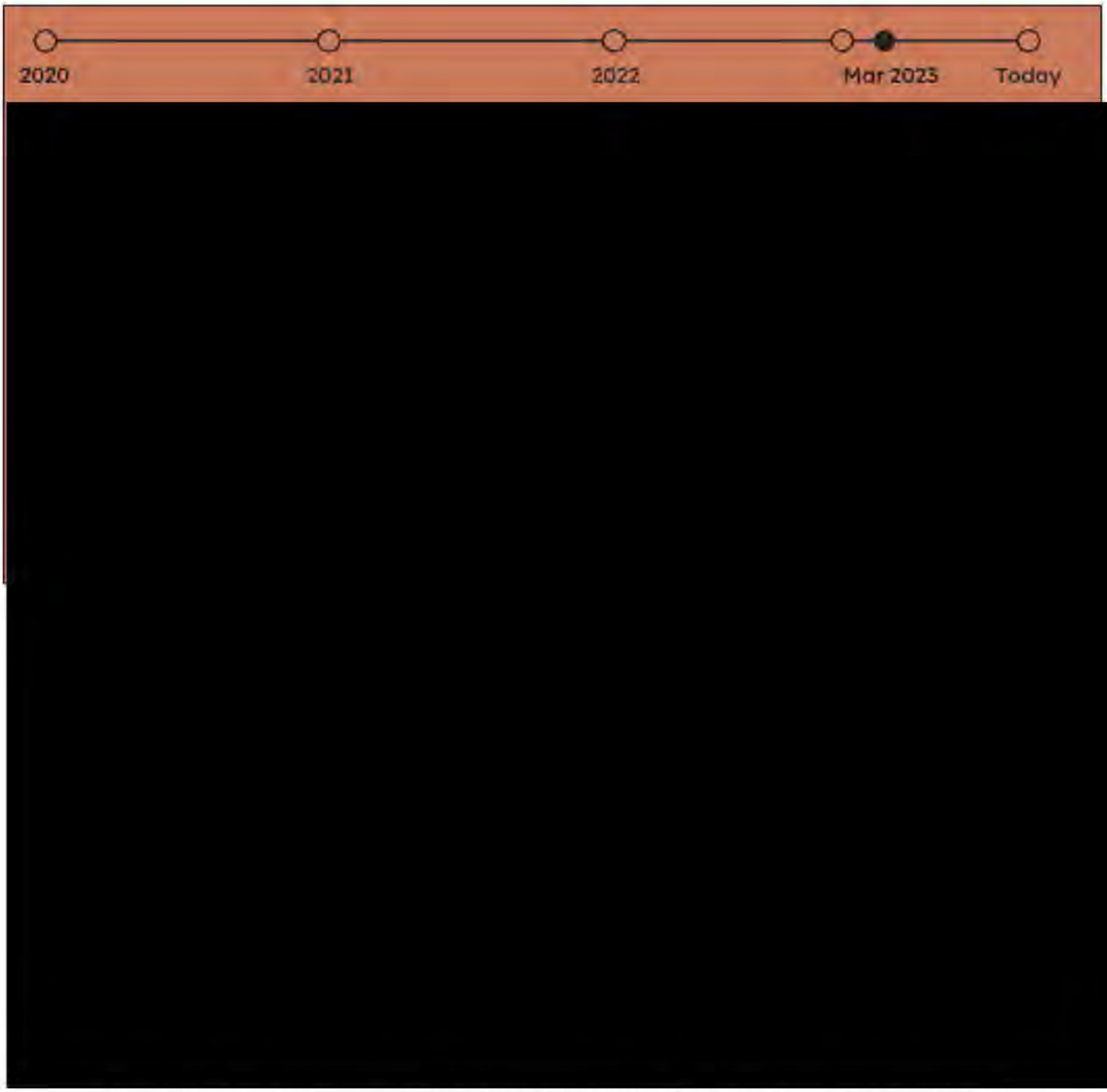

2020
2021
2022
Mar 2023
Today
ADD IMAGE



Slide 53

12 Aaron will share an image on Friday to tell the story of API usage[Author:Anne le Roux],[Creator:1347899476],[Date:9/7/2023 2:57:08 AM],[AuthorInitials:]
Anne le Roux, 10/24/2025

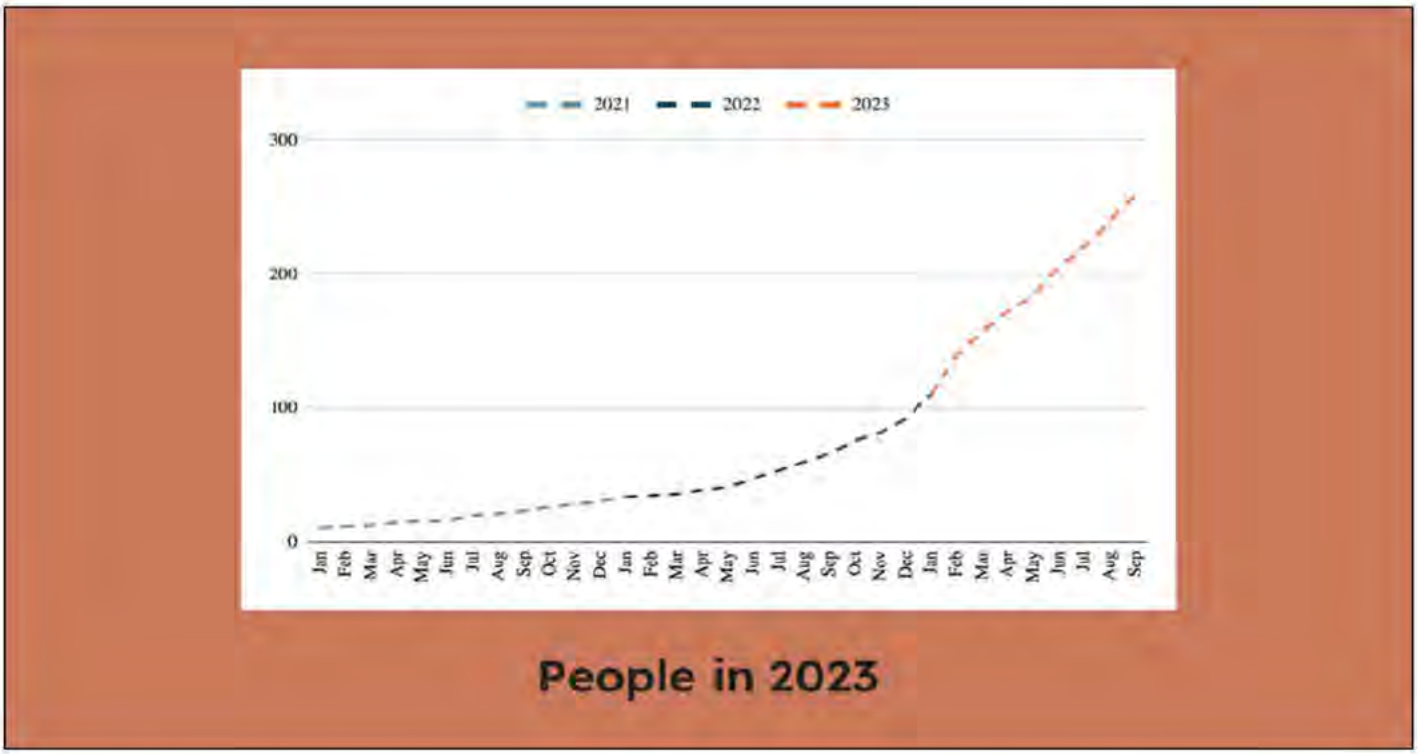
13 "let me ask jenan and brian if they have any ideas for visuals that tell a story about API usage.
i'm sure we can pull the latest waitlist numbers too.
i'll see what I can find via the team tomorrow after the claude pro launch."[Author:Anne le Roux],[Creator:1347899476],[Date:9/7/2023 2:57:08 AM],[AuthorInitials:]
Anne le Roux, 10/24/2025



- Then, just last week, we launched Claude Pro. This will be a paid experience for our "power users". This is one in a long list of steps we'll need to take to have a successful product, but it's an important one. These things add up, and they build on each other. Similarly to Hatch, this was a big cross functional effort, and will help generate revenue and enable us to better manage capacity and expand to even more locations over the coming months!


2020
2021
2022
Mar 2023
Today





- Then lastly, taking a look back at our people graphs for this year. Throughout this year, our people-growth also really started to accelerate. In January we were around 100 people. Today we are around 260 people. This is faster than what any high growth technology company handbook would tell you is acceptable to do – never more than 2X is what I've always been told, and I can tell you from my last two companies, 2X is still a lot. We are breaking the sound barrier on this one.

- As part of this continued fast pace of growth to accommodate the needs of our users, the safety mission, the product and the demand for our services, we will need to keep growing quickly, we will need to retain our high bar and talent density, we will need to codify our culture and values and we will need to evolve as a company in terms of how we communicate, what structures and formats we use and how we work with each other.



- And now, for my favorite slide… Anthropic babies over the last three years!
- [comments]

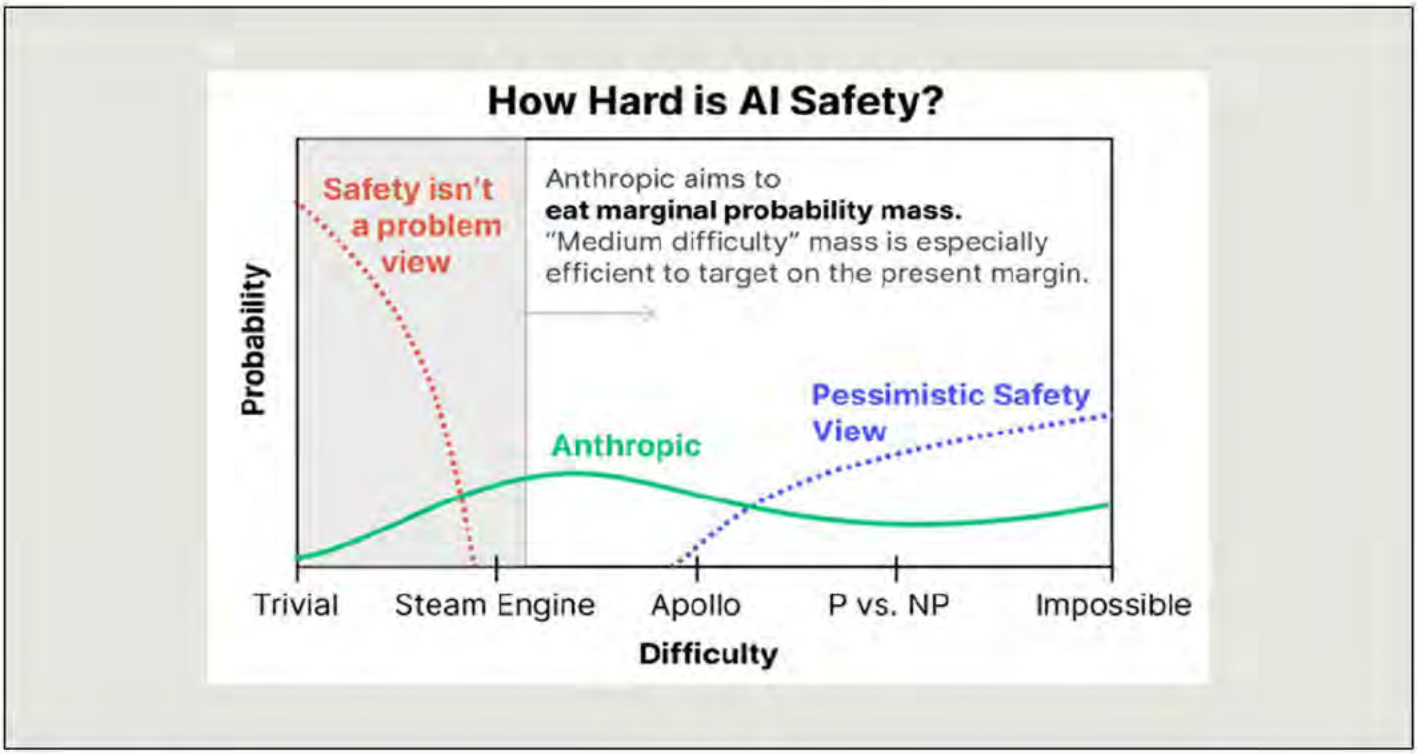


- To date we've been focusing on developing the best, safest models and products in the world – our strengths have helped us achieve all this incredible progress towards our mission of responsibly developing advanced AI for the long term benefit of humanity
    - Unusually high talent density
    - Incredible culture
    - Scientific progress and empirical approach
    - Deep commitment to safety
- Now, we are about to head into another year of insanely ambitious goals – the types of projects that most people would think you're crazy if you told them that's what you were aiming for. Things will only get bigger and faster and crazier from here on, and our challenge over the next year will be staying at the frontier – both in terms of our big models and products – while remembering the reason we founded the company, and being uncompromising on our core safety mission.



Ensure the world safely makes the transition through transformative AI (TAI)

- Before we start looking forward, I want to go back to this slide, written messily in the early days of Anthropic, and take a moment to reflect on why we are working on all of this, and acknowledge that this is an ambitious and truly unique set of goals.
- We might literally be the only company in the world trying to balance all of these competing priorities, alongside being humans with families, friends, and loved ones. However, the same values and strengths that brought us this far will serve us well as we continue to walk the tightrope of scaling safely.



- [Talk about how all of our work ties into the same overall big mission, we're one big interdisciplinary team, and we're making different complementary bets towards the same core mission]
    - Our safety research covers all of applied alignment, alignment science to read team our existing techniques, and mechanistic interpretability
    - Our policy research and credibility…
    - Product to drive revenue…

Looking to the future:

Research in 2024

Today 2024 2025



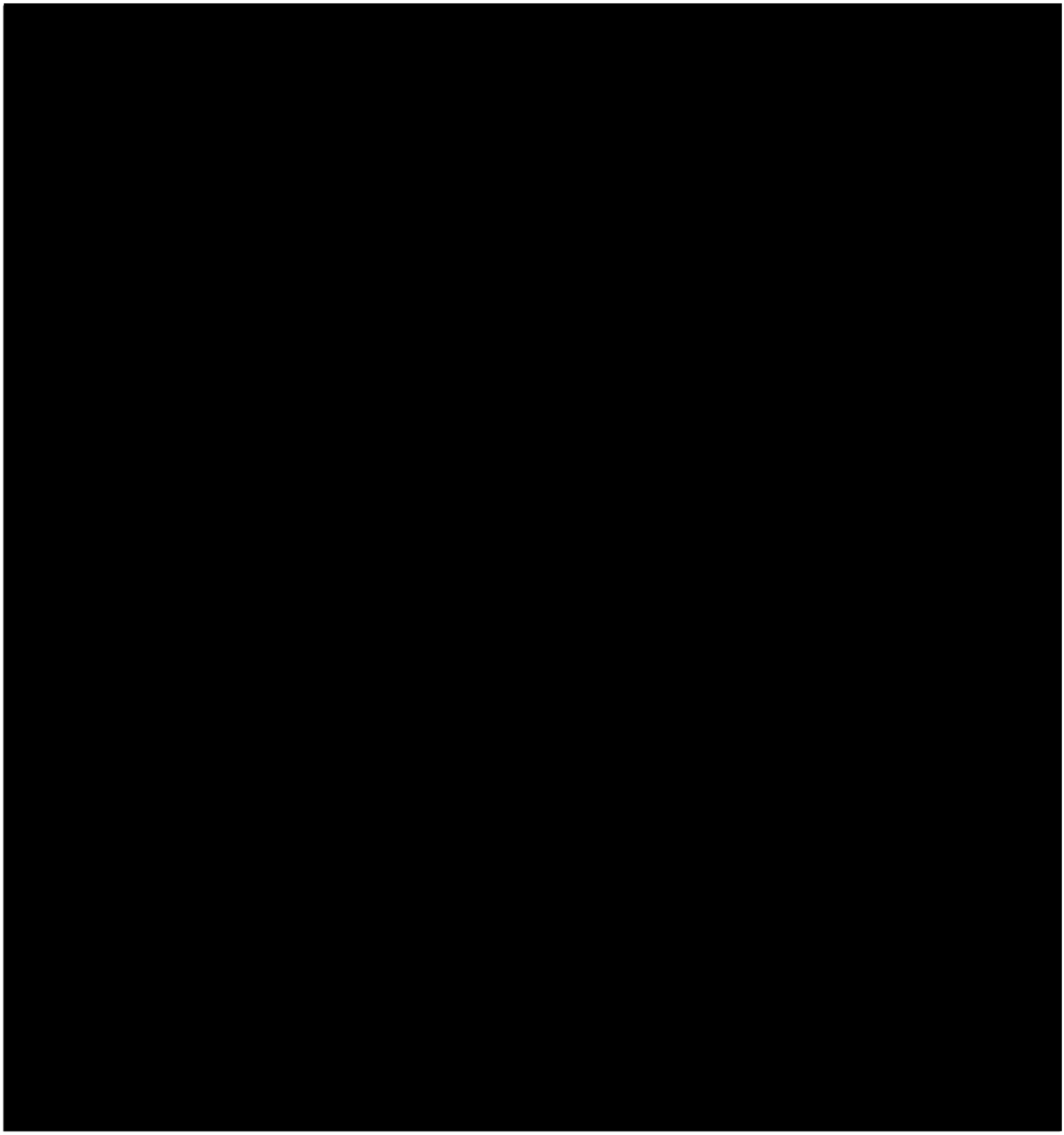
New clusters + preparation for big runs

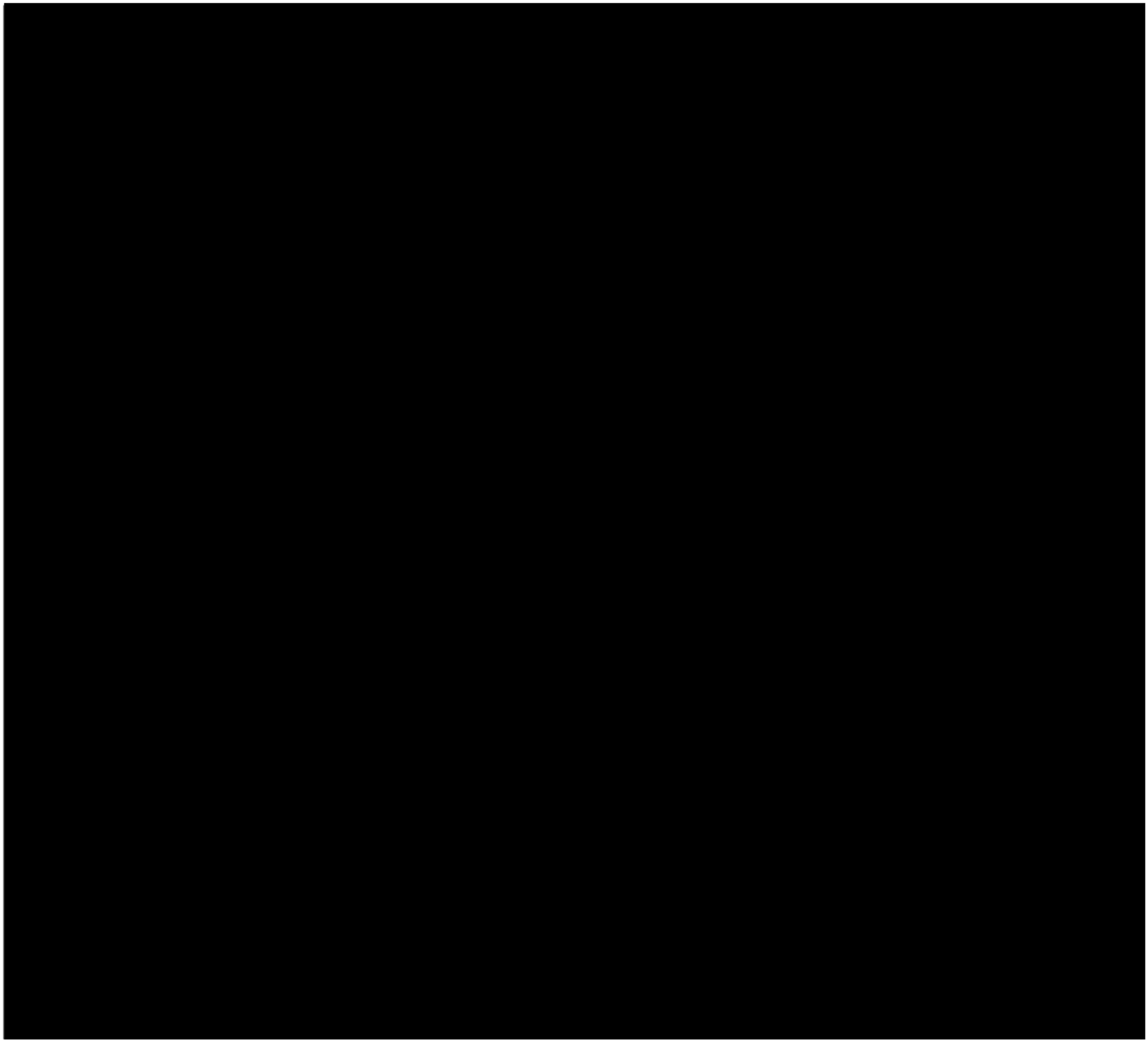


- Lots of this work is already underway!















Policy and Governance in 2024

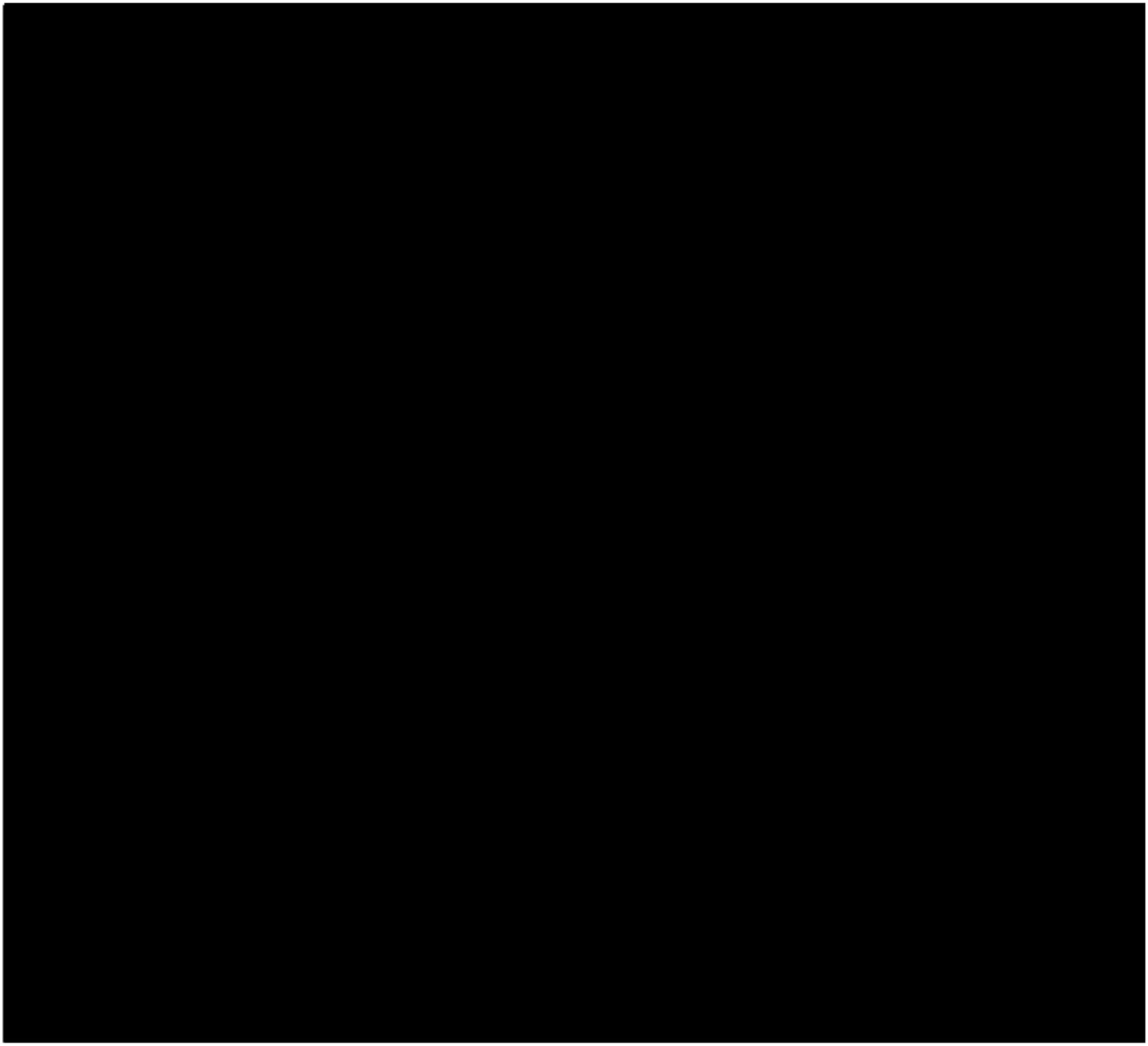
- On the policy and governance side, it's a little harder to make exact plans, since interest is growing and things can change rapidly. However, I think we've all been blown away by the opportunities that we've been given to contribute to important conversations, and expect this will continue into 2024

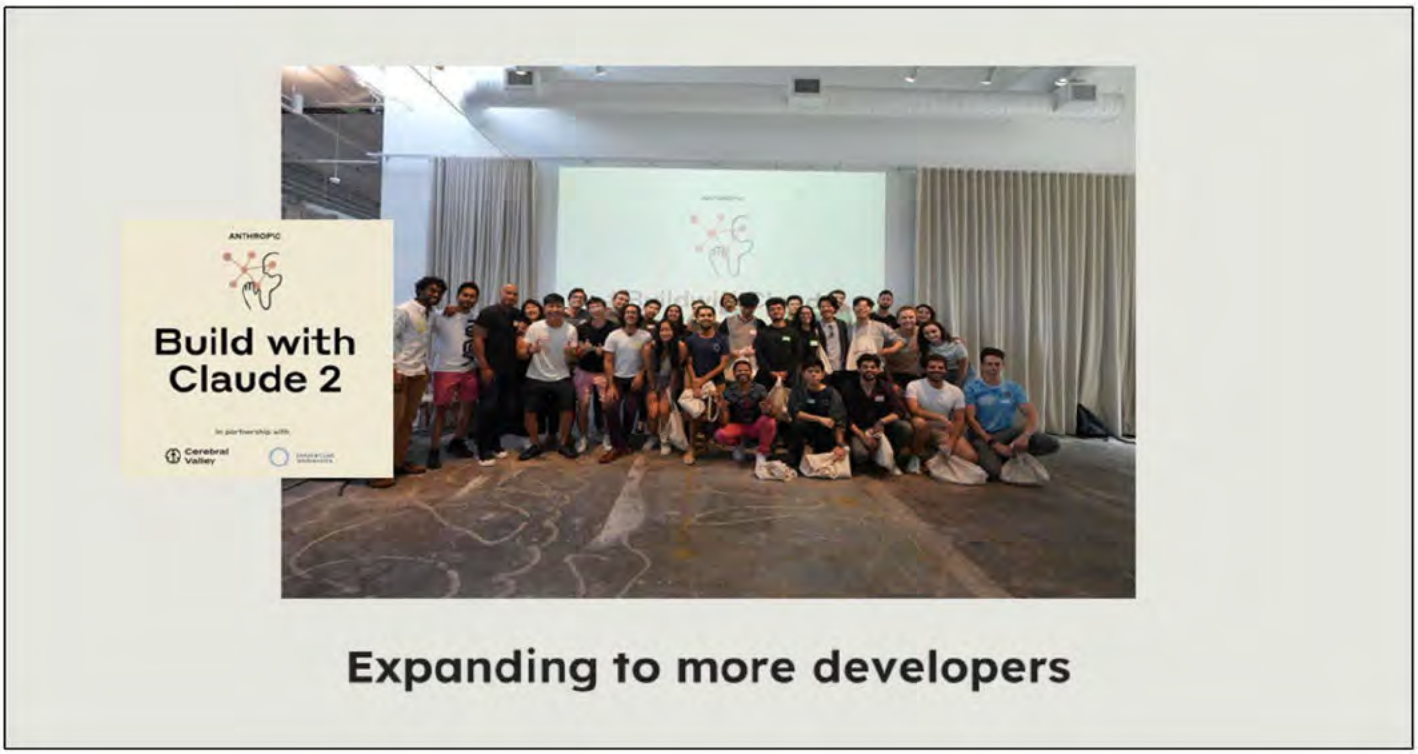
[redacted]

- We will also continue to advocate for sensible policies which prioritise safety, and will aim to continue to be considered the most trusted AI lab at the top levels of the executive branch and congress.

Looking to the Future: Product in 2024

[redacted]

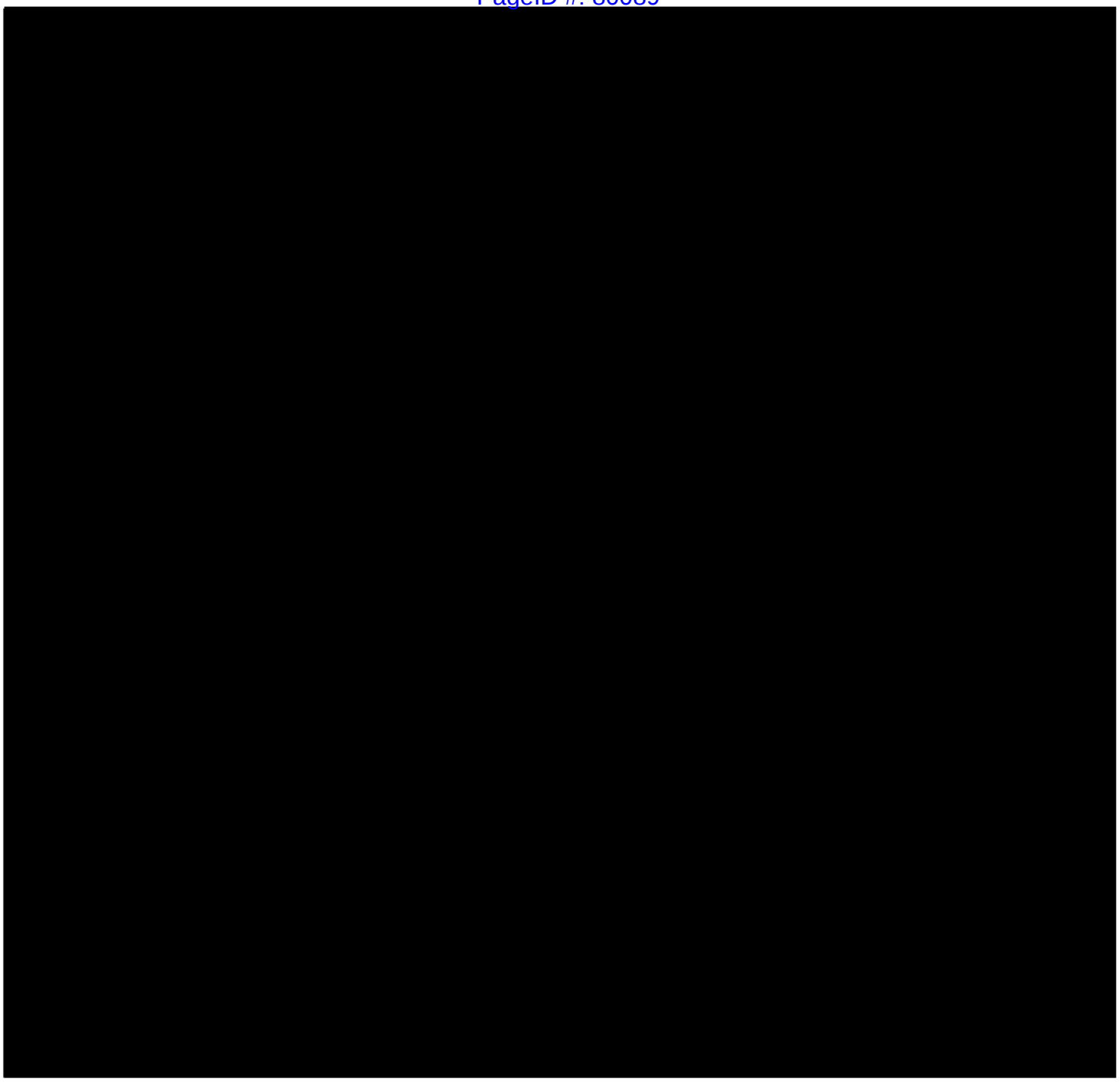




Expanding to more developers

Slide 73

14 edit to say CLaude 3[Author:Anne le Roux],[Creator:1347899476],[Date:9/7/2023 9:21:42 PM],[AuthorInitials:]
Anne le Roux, 10/24/2025



Slide 74

15 add lots of logos to this slide and add product research team[Author:Anne le Roux],[Creator:1347899476],[Date:9/7/2023 2:15:30 AM],[AuthorInitials:]
Anne le Roux, 10/24/2025

16 + Kate[Author:Anne le Roux],[Creator:1347899476],[Date:9/7/2023 2:15:30 AM],[AuthorInitials:]
Anne le Roux, 10/24/2025



ADD IMAGE

Hitting our stride in Product

[redacted]

- These numbers are only going to keep getting bigger and more insane. Our tiny team of heroes will need to grow to meet the challenge over the next year. We'll be investing in scaling these teams up to a sustainable size, working on shipping fast while continuing to be thoughtful, and … etc.



"It will never be chill again..."

Thanks for sitting through all of that, and I'm now going to get sappy for a minute.

Getting to help build Anthropic has been the absolute joy of my career. We're not here to talk about me, but I want to spend just a minute on why that's the case.

-It's the mission - We are taking on a bet on doing a thing that many people think is impossible. We're trying to ensure that possibly, likely the most powerful technology ever developed is done so safely. It's the ability to talk openly and seriously about material matters that impact the world.

-It's the people -

-It's the scope and ambition of what we are trying to do. Anthropic is not just a research lab. We're not just a think tank. We're not just a tech startup - we are all three.

-We are aiming to develop the most ground-breaking technology in the world.

- We are aiming to fundamentally change what is happening in the halls of governments and the board rooms of businesses around the world to help make sure things are developed safely everyone in the ecosystem.

-We're aiming to break and surpass every law of tech startup physics (sorry physicsts) in terms of products, deals, and growth rates.

It would be an incredible privilege to work on any one of those. Working on all three is a gift.

I've shared this story with some of you before but - Vinay anecdote - it can be hard to pull back and think about the incredible, immense opportunity, impact, and weightiness of the work we do when we're struggling with a challenging research problem or bashing bugs or arguing with sales leads or off boarding customers who are definitely trying to use Claude for porn or responding to snack requests.

But I think it's genuinely worth stepping back - at least for the next few days - and reflecting on what it is we are trying to build here. We are trying to build something that has never existed before. We are collectively taking the responsibility of trying to make that thing safe for everyone. We are trying to make it positive and beneficial and meaningful for everyone. We want to help this technology usher in the better angels of humanity's nature and avoid catastrophe.

That is a fucking tall order. That is a privilege. That is an immense responsibility. And the people sitting here in this room, at this company day in and day out - each of you - are holding that responsibility with us.



- Make things go well.
- Team doing the work again. Interdisciplinary by design. In interviews - neuroscientists, physicsts, ethicsts, business leaders, engineers, policy people
- Lots of different bets, but all in the service of the same goal. Making TAI go well
- I feel genuinely grateful to be on this journey with each and every one of you.
- Thanks for what you do for Anthropic. Thanks for fighting the good fight and being self critical and asking tough questions and caring about what is important and what really matters.