Using AI Tools in Physics

Outline

- My perspective on LLM advancements
- Examples of using agents in physics research
- Thoughts about the future

- Mark Cheung on visualizing physics with agents

(All slide text is written by me, though AI assistance in plotting/visualizations)

Basic history of LLM / agents

History of LLM

Nov 2022

ChatGPT

GPT-3.5 made public.

2023–24

Multimodal

Models read images, plots and handwriting by tokenizing images.

Sep 2024

Reasoning

o1 thinks before it answers. Big gains in physics and math, fewer hallucinations.

Feb 2025

Agents begin

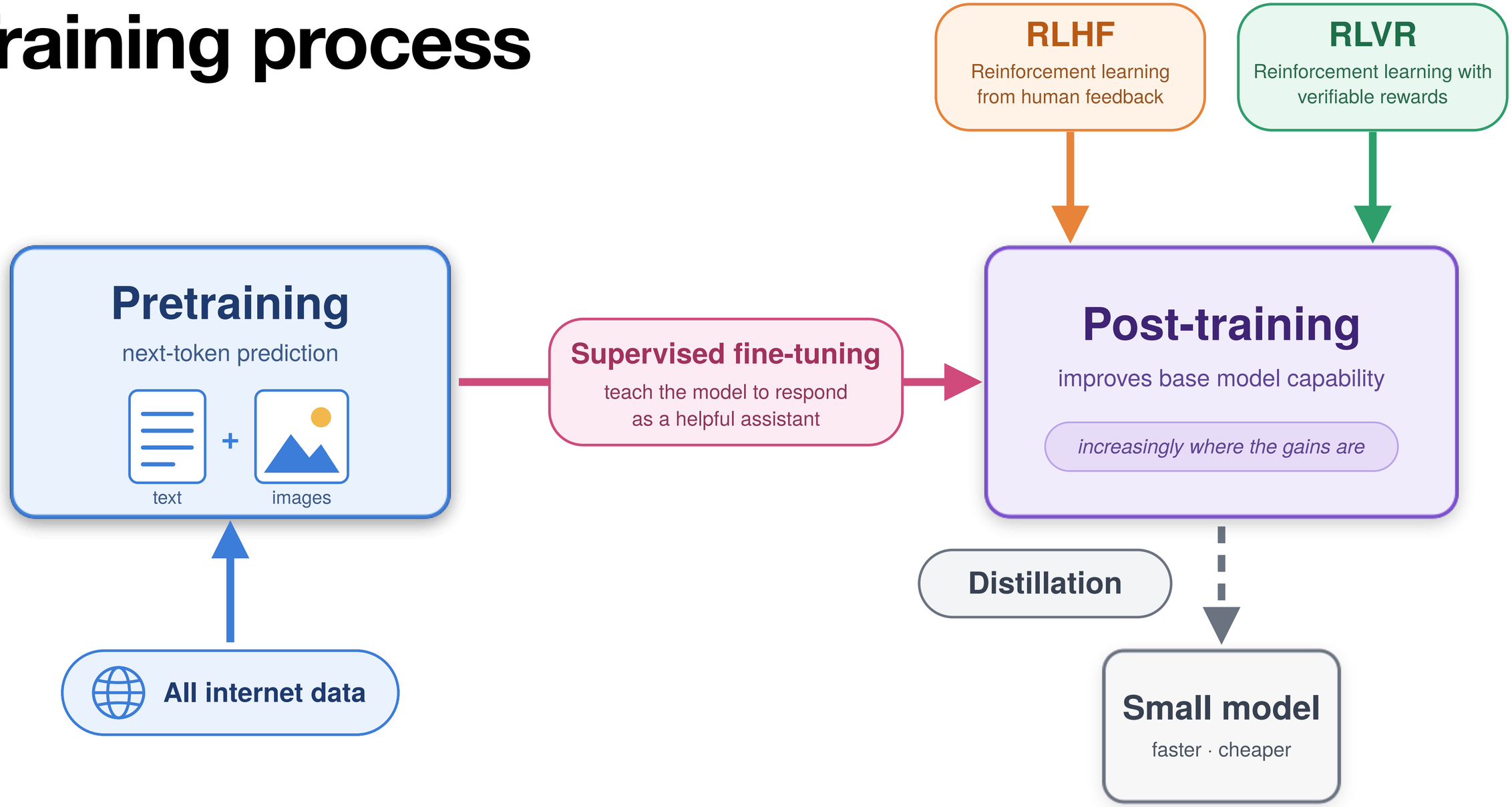
Claude Code released; models act in the terminal.

Nov 2025 – Feb 2026

Agents get good

Increase in agent performance.

Training process



Scaling ingredients

1

Pretraining

Amount of training compute

2

RL post-training

Verifiable tasks and environments

3

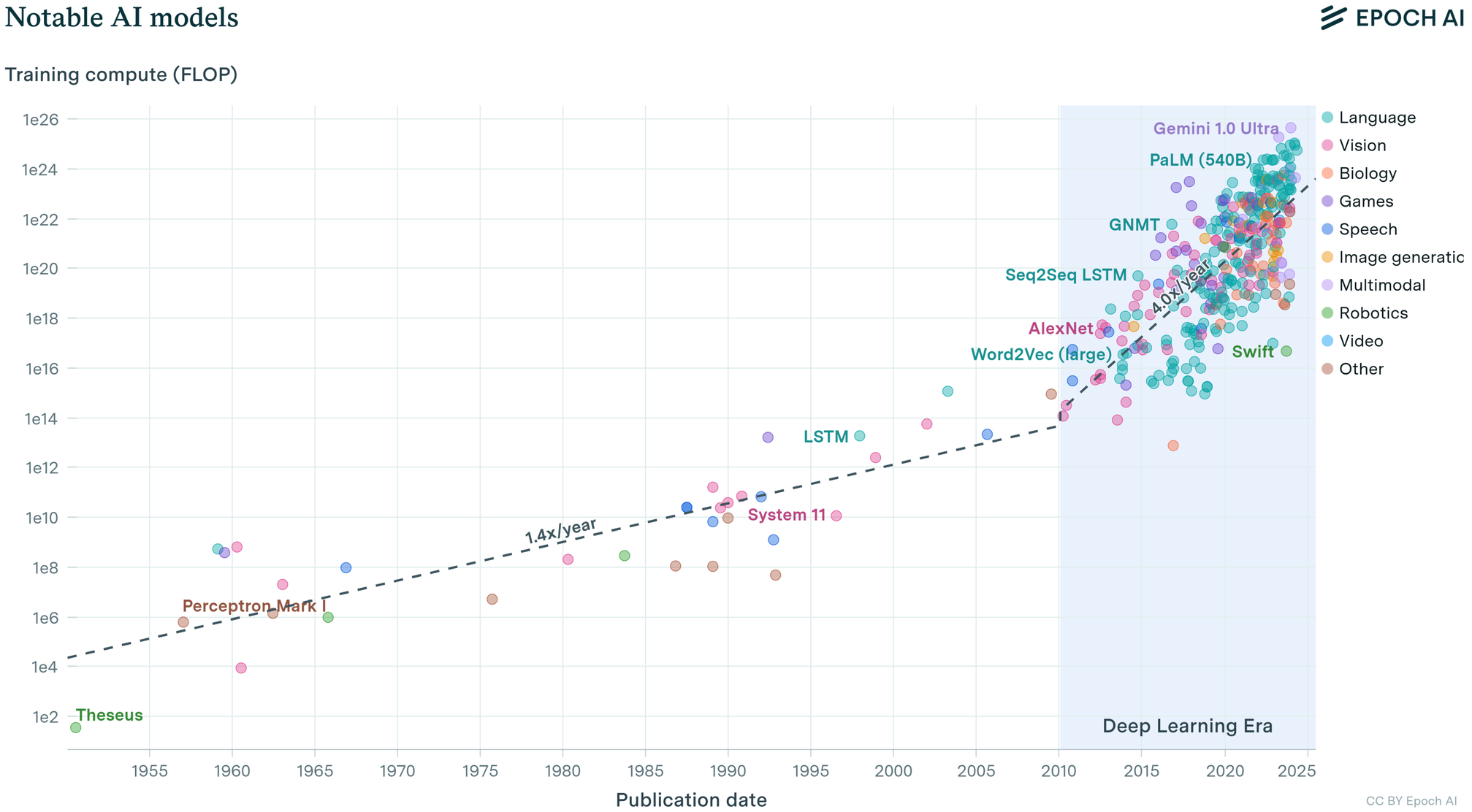
Test-time compute

Longer reasoning at inference

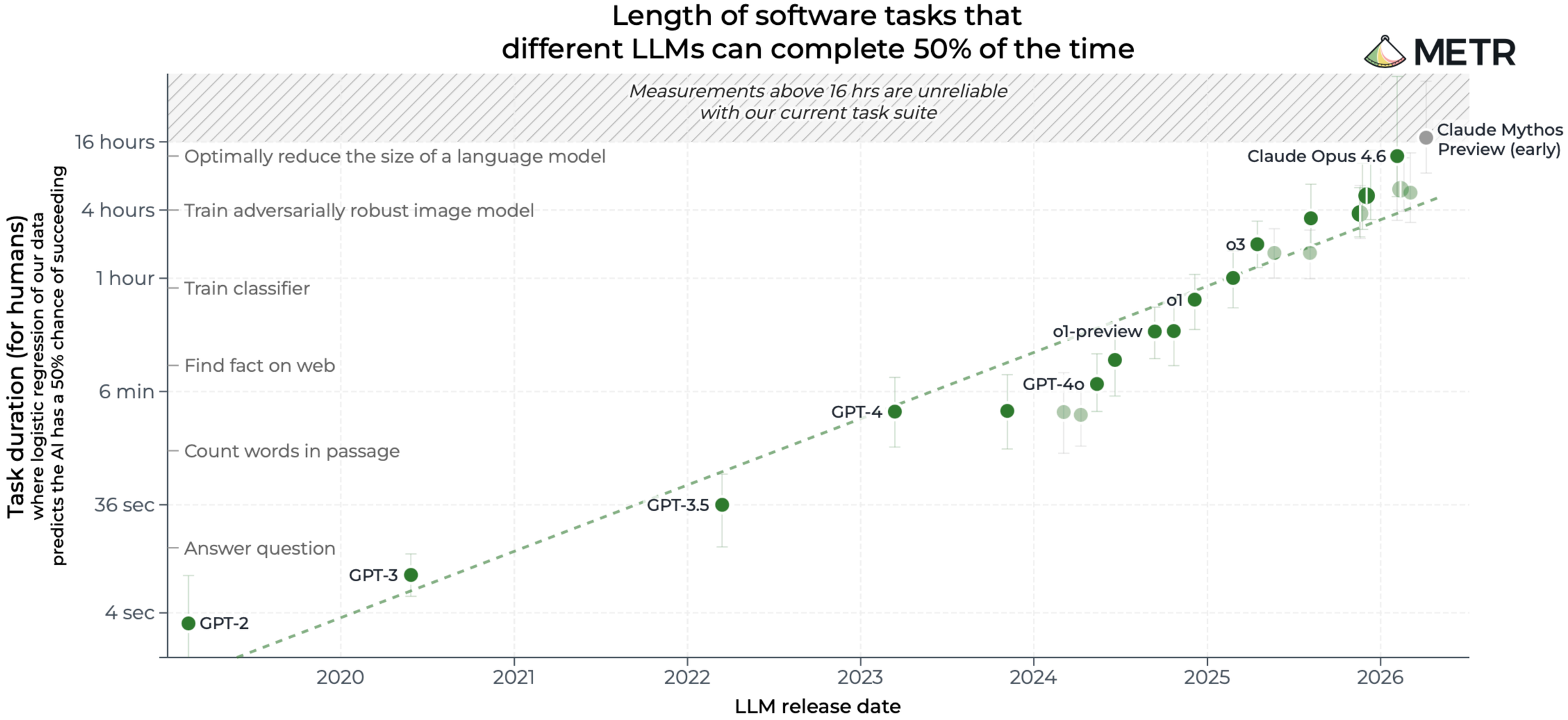
Task horizon length

Length of task completed on its own, measured in human time

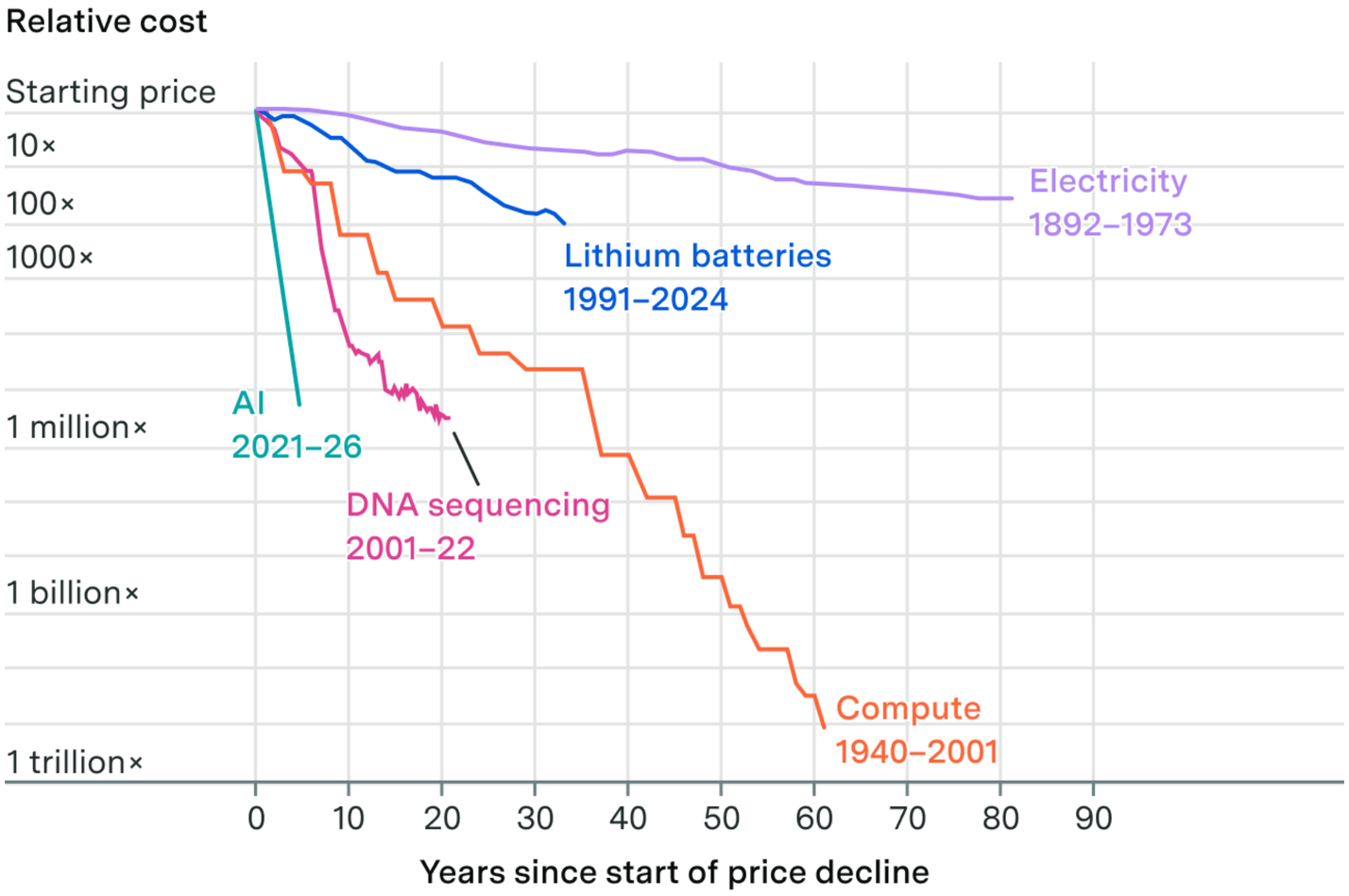
Compute trends in ML training



Task horizon length for coding



Inference cost is dropping rapidly



[epoch.ai]

AI advancements in math

- As everyone knows, OpenAI released (forced) Navier-Stokes counterexample in September, and subsequent GitHub proof releases
- Verifiable problems are key for RL training and evaluation
- Lean verification in math + unit tests in programming

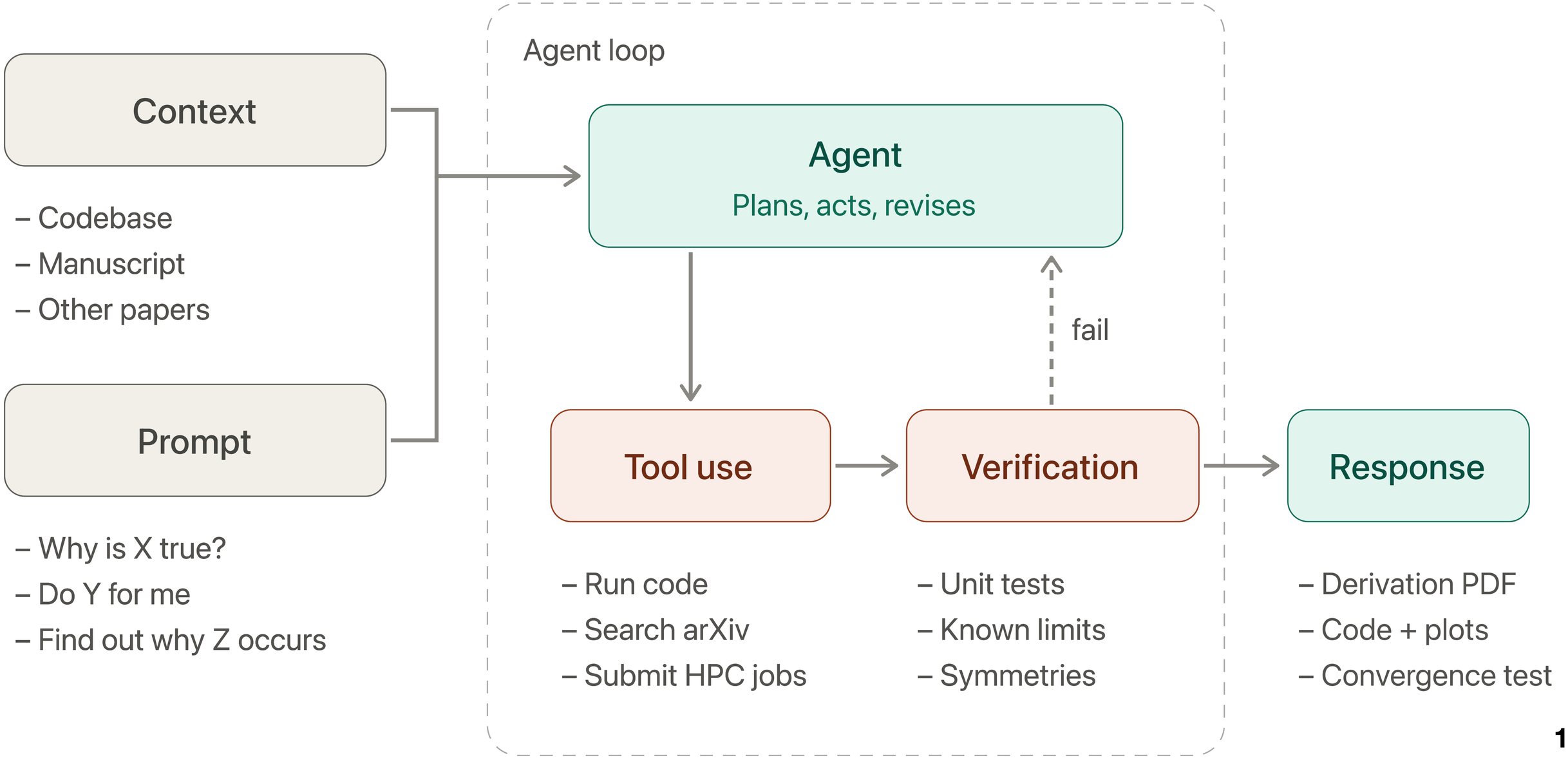
Agents in physics

Most important tips

1. Paying $20/month for Claude/OpenAI plan is essential
2. Use Claude Code in VS code studio / command line performs much better
    - Agents leverage tool use: write and execute code, search web, read files on system, read Github commit history, …
    - Don’t copy and paste between computer & browser
3. Use photo as prompt, screenshot paper / your plots to understand
4. Use Plan mode in Claude code before complex operations

💡Pro tip: double click alt for fast photo

Workflow paradigm



Importance of context management

- Make sure your agent can read context easily
  - Previous papers, codebase, manuscript in preparation, …
- Performance degrades if you have too much
- For admin tasks, I have the cog handbook, postdoc applications, grant agreement, grant proposals, …

My workflow / setup

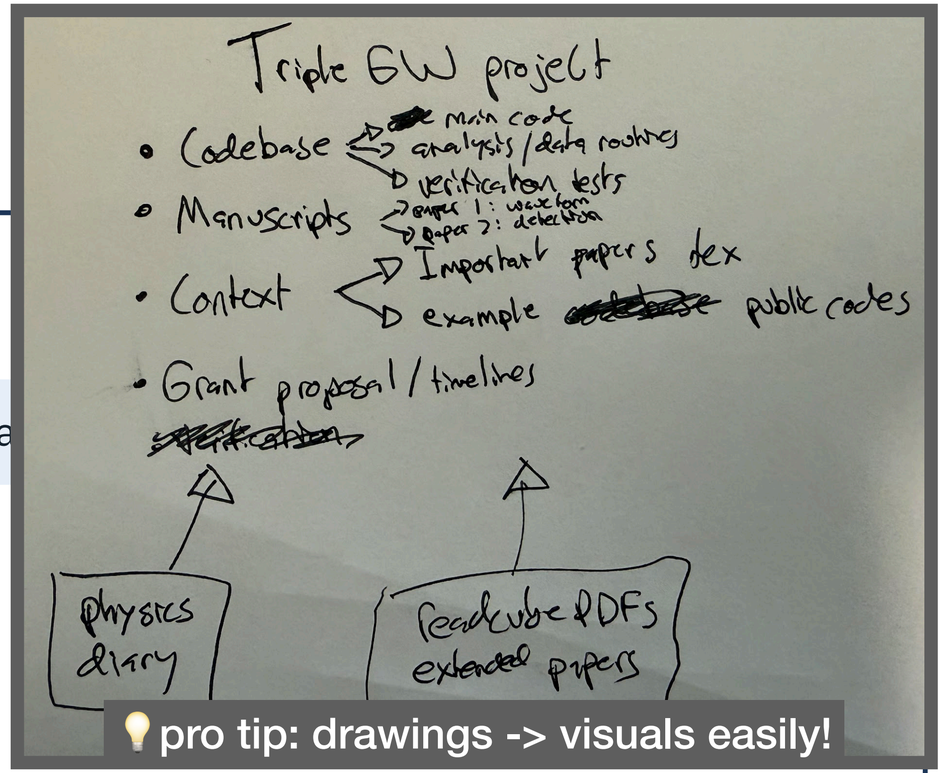
Triple GW project

Codebase → main code | analysis / da

Manuscripts → paper 1: waveform

Context → 20 key papers' arXiv source

Grant proposal / timelines



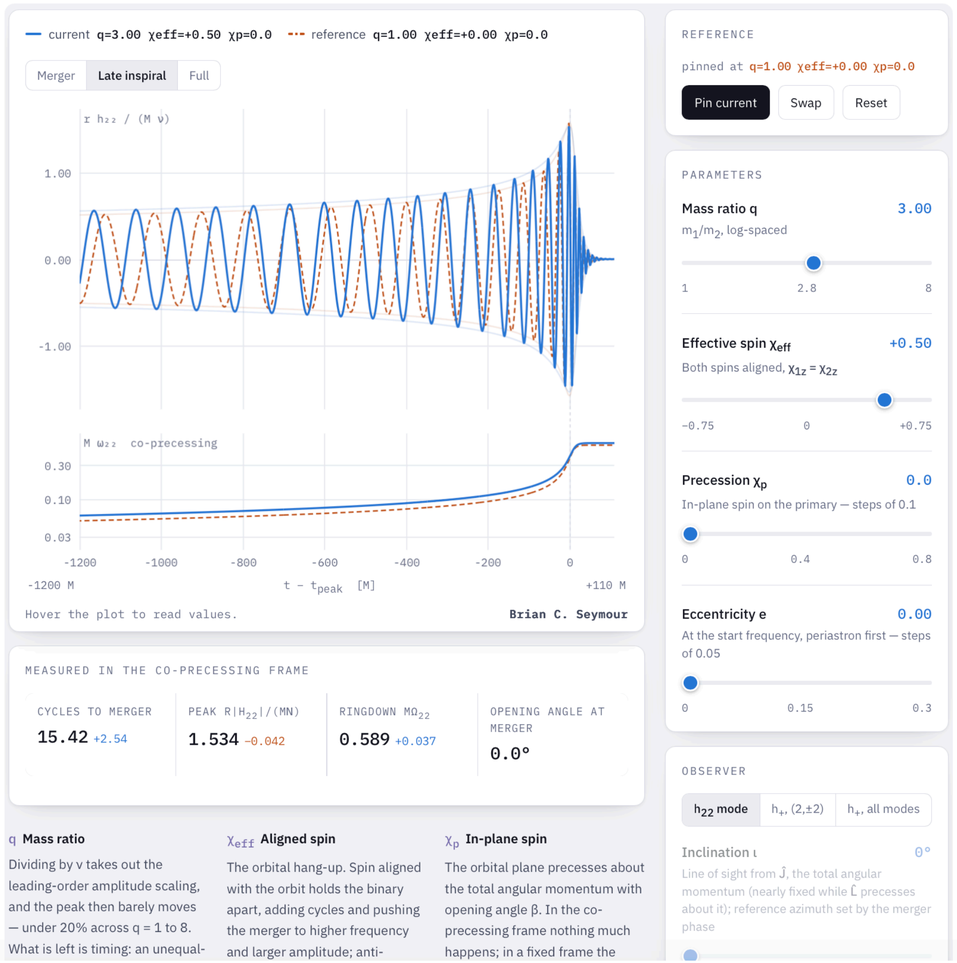
💡pro tip: drawings -> visuals easily!

Physics diary
diary since 2019

ReadCube PDFs
900 PDFs

Example : live precessing waveform website

- I was interested in seeing how time domain waveforms look for precessing systems
- Spent ~5 hours on one weekend to make a live website demo [link]
- Used SEOBNRv5 + EHM/PHM
- Claude code optimized how to make HTML page with sliders + how to save data for people to run without python

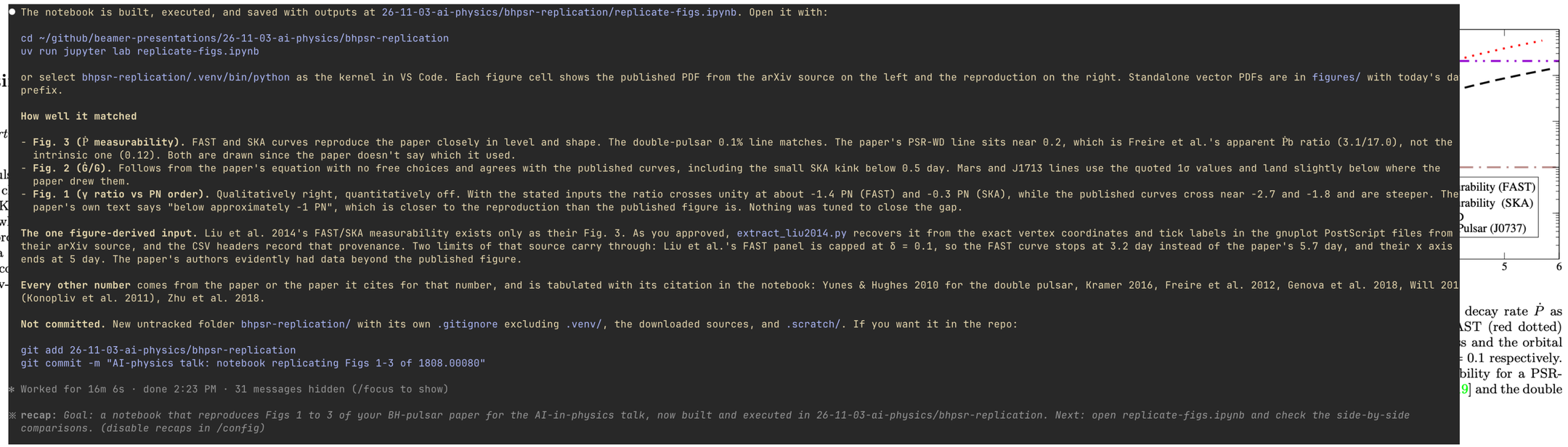


Example : attempt to replicate my paper

- I will prompt claude and try to replicate figure #1-3 on my first paper

/plan I am giving a talk on AI in physics. I am trying to show how well AI agents can replicate the results of a paper. I want you to try and generate figure 1 and figure 2 and figure 3 of https://arxiv.org/pdf/1808.00080 using what is laid out in the paper. Download the latex source code / inspect the figures. Can you carefully read through my paper and work through making python code to generate these plots. Do not do any assumptions that are not in this paper, and make me a python notebook for this. do not guess if there is some data that you don't have

Make a uv environment for this python notebook, so that I can open it and test it out. Place the figure in my paper next to the figure in this notebook to one another so that I can compare visually how good of a job that you did. If there is any supplemental data that is needed to generate these figures, you have my permission to download them (e.g. Liu+ 2014, i believe). Don't pull any data off the plot, you must directly use something from an equation or a number that you find



Miscellaneous

- Automation for bibliography:
  - Use deep research mode to look for any missing literature / citations
  - ArXiv # -> Inspirehep bibtex automation
- Batch fixing math convention typos / making sure all abbreviations are defined
- Fixing merge conflicts in GitHub automatically
- Visualize codebase in development with Claude artifacts — can help you improve architecture
- Suggesting code speedups (e.g. did a ~60x speedup with JAX)

💡Pro tip: try using voice instead of typing input!

Comments

- Serdar: Expand dense 5 page PRL -> 25 page pedagogical paper, with extensive derivation justifications
- Serdar: help with generating Tikz graphics more quickly
- Luka: when looking at manuscript which has unphysical result, use agent to automate finding which approximations are justified or which could be the problem
  - Useful when two papers have opposite conclusions, and time-consuming to figure out which tiny assumption made the difference
- Javier: rents server for $10/month so he can do persistent agentic work
- Jose: use Claude add on in word for usage, also he has 10 free $20/month Claude for science plans

Administrative uses

- Claude connectors very handy:
  - Create/edit google calendar entries
  - Searching/drafting emails
- Travel reimbursement automation
  - renaming receipts, making spreadsheet, and finding any missing transactions in credit card statements

Which model provider is the best?

Anthropic

Opus 5.5
strong + fast

Fable 5.1
strongest

- ⊕ Agentic coding
- ⊕ Long persistent problems

OpenAI

Sol 6.1
general use

Astra 6
detailed tasks

- ⊕ Derivations / math
- ⊕ Strong computer use

Google

3.8 Flash
quick tasks

4 Argon
reasoning

- ⊕ Images, video, audio
- ⊕ Built into Google apps

- Differences are minor, pick best price and switch as model releases come out
- How you use/interface with agent and workflow more important

Open discussion: Did I miss anything?!

Conclusion & Discussion

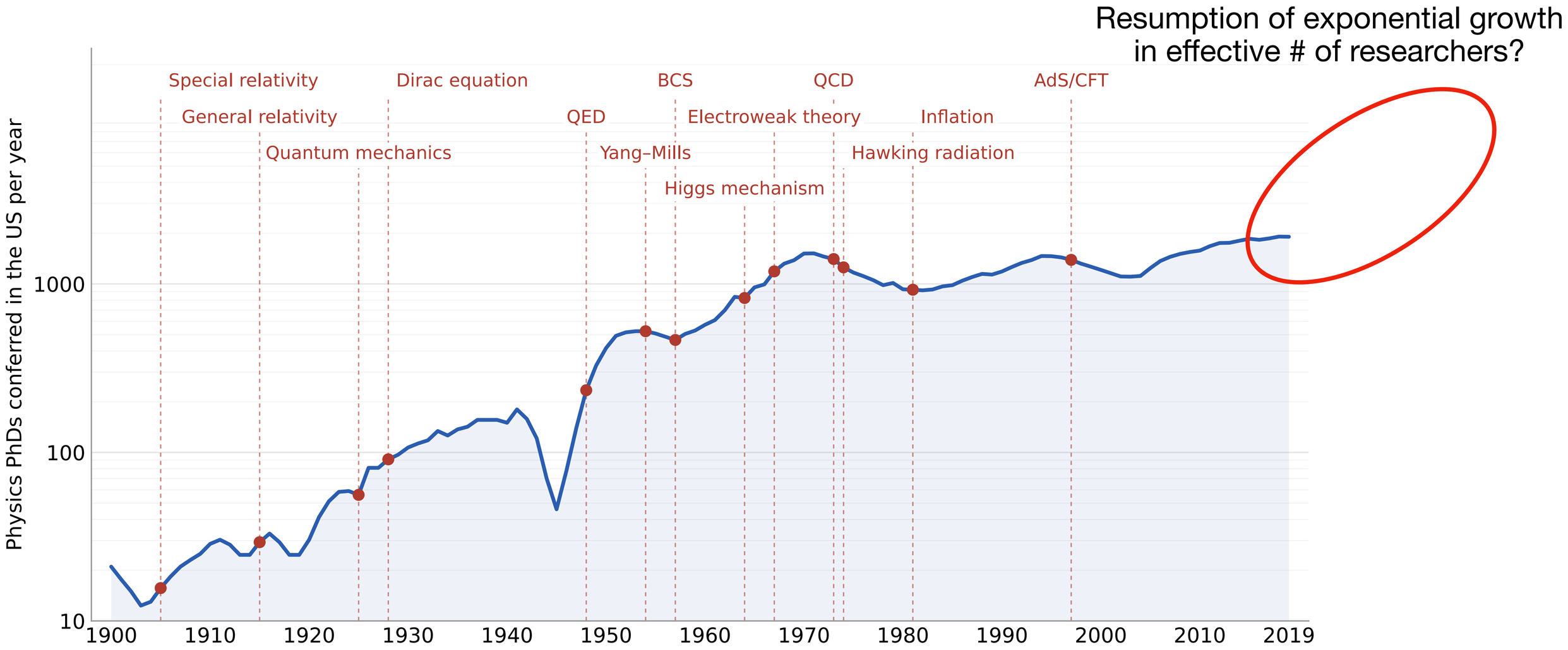
Unsolicited advice

- Things are likely to change quickly, we can only be adaptable to technology
- Be ambitious : likely many new problems are accessible now
- AI is likely to improve for the next years, so practice beneficial to you
- Agents performance is jagged:
  - ✅ verifiable domains & persistence
  - ❌ problem taste and physics interpretability mediocre
- Answers may outrun explanations : verification bottleneck
  - Synergies with numerical analysis, analytical methods, statistical inference, data analysis, and physical experimentation

Thoughts for Early Career Researchers

- Some people worry that AI is going to make senior researchers more productive and make it hard for juniors
  - Being an ECR could be easier to fluidly adapt to new technologies / workflows
- AI will likely get to point where you can shortcut reasoning. Curiosity and effort still drive scientific capacity, so always understand what you produce.
  - Poor understanding will show up later when you give public talks or interview for jobs
- Use sparingly for writing
  - bad for personal growth & people do not respect AI writing
- Public speaking and communication skills are likely going to grow in importance

Closing: timeline of physics discoveries



Data: AIP Statistics (ACE 1900–19, NAS 1920–61, AIP 1962–2019)

Sources

Physics/Science social questions

- David Hogg “Why do we do astrophysics?” [2602.10181]
- Katerina Chatziioannou’s blog [link]

AI background

- Claude code guides
- Boris Cherny has useful guides on agents [inventor of Claude code]
- Dwarkish Patel podcast — great for bike commutes
- Epoch AI / METR / Artificial Analysis : used to follow news on AI / benchmarks