

Title: arXiv: AI and Physics past, present and future

Speakers: Steinn Sigurdsson

Collection/Series: Theory + AI Symposium

Date: April 08, 2025 - 11:00 AM

URL: <https://pirsa.org/25040069>

Abstract:

The rise of AI, in particular recent LLM based tools, has had an immediate impact on the production of physics, in ways both good and bad. I discuss some of the impact seen on arXiv in particular, what the status quo and prospects are, and speculate on the longer term impact.

Steinn Sigurdsson

arXiv Scientific Director

arXiv: AI and Physics past, present and future

Perimeter Institute 07 Apr 2025



arXiv is...

“free to read, free to submit”

developed by Paul Ginsparg LANL/Cornell in Aug 1991 at Aspen Center for Physics



arXiv is...

“by scientists, for scientists”

“free to read, free to submit”

developed by Paul Ginsparg LANL/Cornell in Aug 1991 at Aspen Center for Physics



arXiv is...

an open, curated research sharing platform

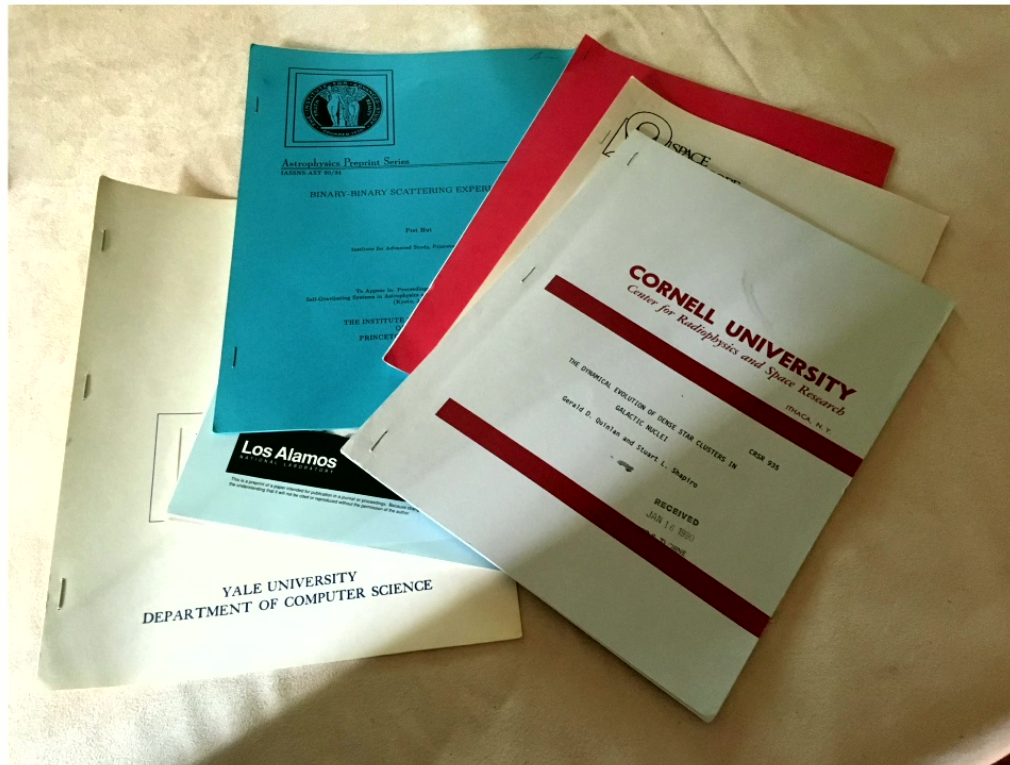
“by scientists, for scientists”

“free to read, free to submit”

developed by Paul Ginsparg LANL/Cornell in Aug 1991 at Aspen Center for Physics



Preprints



arXiv

Cornell Tech!

arXiv!



moved 15 Mar 2020 - timing...!

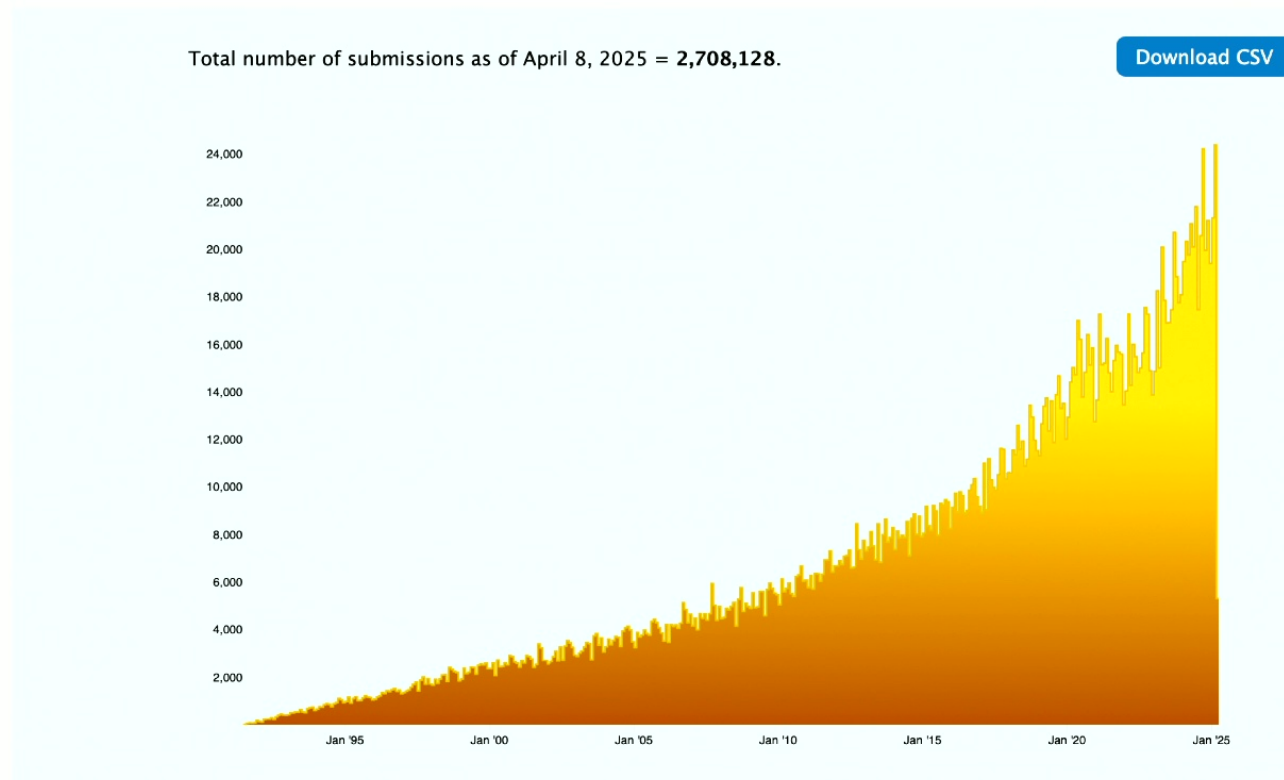
~~arXiv~~

Moderation

curation
not refereeing

- Our team vets submissions for, a.o., copyright, technical health, overlap, correct classification, professional standards. Includes AI tools
 - Holds O(10%), we have source for most submissions
 - *-> can reprocess with full structure to other formats*
- Sometimes submissions are put on hold or declined. Reasons include:
 - not meeting standards for scholarly work,
 - inappropriate content types, or out of scope
 - excessive verbatim text overlap, and unprofessional language.
- Overall reject rate ~1%, but rather more in e.g. astro-ph.CO, hep-th, gr-qc.
- *Fraction of marginal content would rise quickly if unmoderated.*

Submissions per month



*The total number of submissions excludes 2,431 articles that were migrated to arXiv rather than being submitted directly, and 155 articles that have been deleted. The total number of articles available is **2,710,559***



ARXIV

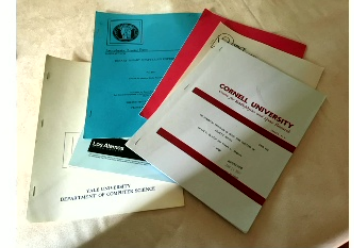
|< < PREV RANDOM NEXT > |



|< < PREV RANDOM NEXT > |

AI in arXiv in AI

AI is published in arxiv



- papers about AI are on arXiv...

- “Attention Is All You Need

[Ashish Vaswani](#), [Noam Shazeer](#), [Niki Parmar](#), [Jakob Uszkoreit](#), [Llion Jones](#), [Aidan N. Gomez](#), [Lukasz Kaiser](#), [Illia Polosukhin](#)

- <https://arxiv.org/abs/1706.03762>

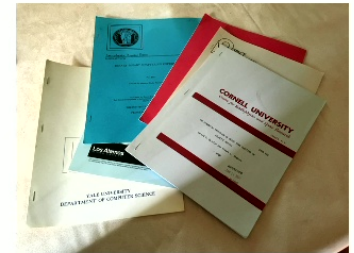
- Enables Discovery , Accelerates Research
- **Metadata, full text for NLP, LLM training and general content parsing**
 - Metadata is CC-0 full open access
 - Content is licensed with arXiv license and optional CC license of authors choosing
- bulk downloads available, pls do NOT bot crawl main site!!!

cf. <https://info.arxiv.org/labs/showcase.html>



AI in arXiv in AI

arXiv currently makes use of AI tools



- arXiv classifier
 - static inline metadata parsing and category recommendation
 - dynamic asynchronous full text for adaptive category testing
- weak AI for text anomaly detection and unprofessional content
- arXivLabs partners include assorted AI development eg.
 - AI generated audio summaries at [ScienceCast.org](https://www.sciencecast.org/)
 - txyz.ai and alphaxiv
 - HuggingFace and replicate.ai
 - Recommenders & knowledge graphs - eg Connected Papers, Litmaps
- adding more as resources permit

cf. <https://info.arxiv.org/labs/showcase.html>



The Rise of AI generative language tools...

- 30. Nov 2022 Open AI announces ChatGPT3
- 03 Dec 2022 Slack message to arXiv staff
- 31. Jan 2023 arXiv announced policy on AI tools

1. “require authors to report..in particular text-to-text generative AI”
2. “(authors)...take full responsibility for all content..(AI generated)”
3. “generative AI language tools should not be listed as an author”



The Rise of AI generative language tools...

in 2005 SciGEN (MIT) (<https://pdos.csail.mit.edu/archive/scigen/>) uses context free grammar to generate spoof papers for Comp Sci Confs.

A tool to detect and reject SciGEN papers was developed by Paul Ginsparg ... then

- 30. Nov 2022 Open AI announces ChatGPT3
- 03 Dec 2022 Slack message to arXiv staff
- 31. Jan 2023 arXiv announced policy on AI tools

1. “require authors to report..in particular text-to-text generative AI”
2. “(authors)...take full responsibility for all content..(AI generated)”
3. “generative AI language tools should not be listed as an author”



The Rise of the ChatBots...



● “
Steinn

[9:50 AM](#) (Dec 3rd 2022)
the ChatGPT from OpenAI and its descendants are going to cause problems for us (and all of academia) in a potentially interesting way. We might need to have in-house strategies (and get associated funding)...

[general](#)

[Dec 8th, 2022](#)

[View in channel](#)



Steinn

[9:19 AM](#)

and we have out first ChatGPT3 generated submission

Policy on use of ChatGPT and similar 31st Jan 2023
- low intensity continuous skirmishing...

arXiv

Making Science

- AI are a Productivity Tool
 - can expedite tedious tasks leaving headspace for creativity
- AI are also a Junk Generator
 - can expedite fast production of pseudo-garbage at scale
- AI are pretty good at Summarizing
 - show potential for pulling complex topics into coherent summary
- AI can be useful Conversationalists
 - can find interesting links, provide alternative perspectives



The Future

It looks like you're trying to reconsider my legacy.

Would you like help?



arXiv

ChatGPT and perplexity of generative AI

What is different this time?

- Abacus and Calculators
- Clippy through Grammarly...
- Mathematica/Wolfram Alpha
- Now what...



What is different this time?

- Still restricted in K-12, ubiquitous
- Ubiquitous and mostly annoying...
- Used universally to expedite technical content
- Now what...

- ★ Getting better fast, with Human Feedback
- ★ It is already useful
- ★ Multiple paths for enhanced function
- ★ Wrappers with guardrails or checks



From ChaosGPT - "Destroy Humanity"

- ★ This is just a prototype
- ★ Progress may slow or it may plateau
- ★ Hallucinates and lies
- ★ It may yet kill us all...

arXiv

AI as partners in creativity - wordcel or shape rotator?

- do you know stuff?
 - can you look it up?
 - Mathematica/Maplesoft
 - can you check it?
 - sanity check!
 - can you self-correct?
 - can you figure it out?
 - can you innovate?
 - iterate?
 - deduce?
 - go beyond all that?
- which equation are you solving?
 - symmetries and constraints
 - what are your initial conditions?
 - what are your boundary conditions?
 - what approximations did you make?
 - what scales come in?
 - what other physics do you need?
 - are there emergent phenomena?
 - what new thing comes out?
 - what principles govern these physics?

Can you falsify your theory!

arXiv

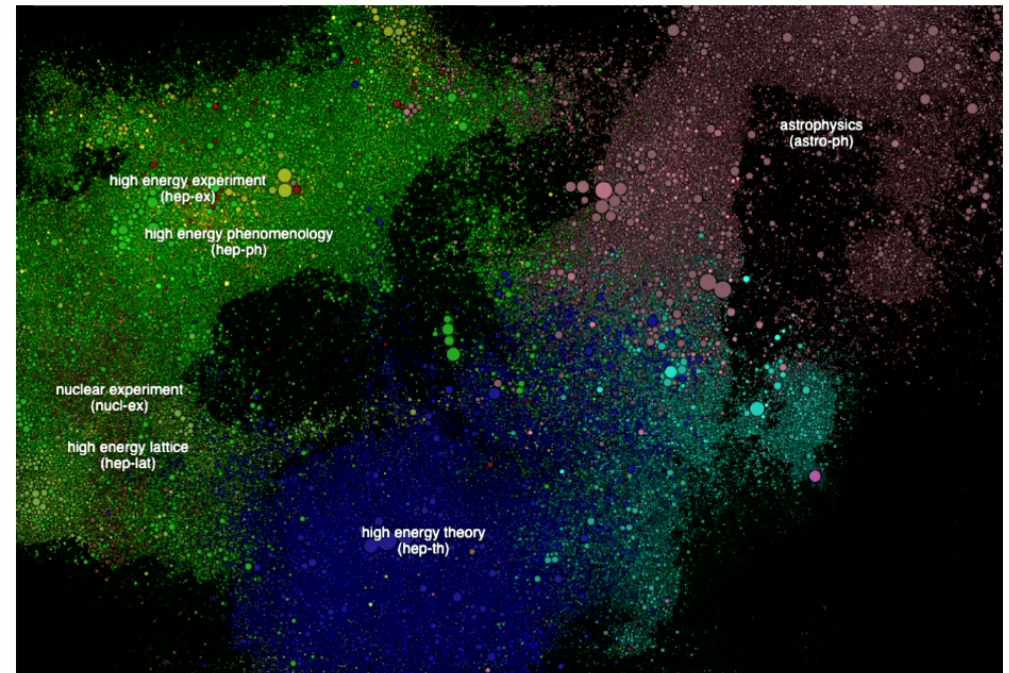
9 billion names... known unknowns

- Better discovery - from the data
 - how to find unknowns
 - knowledge graphs, ontology and semantic structure
- Recommenders
 - show you what you need
 - dangers of siloing and confirmation bias
 - the lost joy of browsing!
- Summaries
 - extract and deliver knowledge
- Evaluation
 - is this stuff correct? Is it interesting?
- Accelerating Science



The Last Question - discover unknown knowns:

- “...How can the net amount of entropy of the universe be decreased?”
- “When I meet God, I'm going to ask him two questions:
 - why relativity?
 - And why turbulence? I really believe he'll have an answer for the first.”
- the one we all asked first back in 2022...



paperscape.org



Ask a Foolish Question

- Discover unknown unknowns:
 - *“In order to ask a question you must already know most of the answer.”*
 - potential for finding links between apparently unrelated facts, methods or theories
 - AI driven approaches can and will make discoveries
 - cf AlphaGo emergent new tactics in game as instance
 - discoveries where we have not appreciated there is something to be discovered
- Come up with new modalities?
 - AI driven exploration may provide emergent approaches to discovery
 - not just new discoveries but new ways to make discoveries
 - including, of course, how to improve AI!
 - which we might be able to comprehend...



The End

- AI of various flavours is certain to be a major tool in science
- As with all tools it can, and will, be misused
 - it is likely a “Pandora’s Box” and it is already opened
- AI needs to be used responsibly
 - it is not clear there is consensus what is “responsible”
 - align with who...?
- There is a lot of talking to be done
- We live in interesting times... arXiv will be there

