The pre-registration experiment
An alternative publication model for ML research
NeurIPS 2021

Opening remarks
13th December 2021

Up-to-date schedule and details at
http://preregister.science/

Reviewers

Adrian Javaloy
Adrian Spurr
Alex Hernandez Garcia
Andrew Gambardella
Arnout Devos
Ayush Jaiswal
Bernardino Romera-Paredes
Brad J Gram-Hansen
Carianne Martinez
Cees Snoek
Chaitanya Devaguptapu
Chaoning Zhang
Chen Sun
David Krueger
Dimitris Tsipras
Disha Shrivastava
Efstratios Gavves
Emir Konuk
Emmanuel Bengio
Erika Lu
Evangelos Kazakos
Francesco Ferroni
Francesco Pinto
Francisco Girbal Eiras
Hadrien Bertrand
Hamid Eghbal-zadeh
Harkirat Behl
Jack Valmadre
James Thornton
Jason S. Hartford
Joseph Viviano
Konstantinos Tertikas
Lénaïc Chizat
Li Shen
Liliane Momeni
Malik H. Altakrori
Martin Mundt
Mélisande Teng
Michele Svanera
Miguel-Ángel Fernández-Torres
Nazanin M. Sepahvand
Oriane Siméoni
Paul Rubenstein
Rishabh Agarwal
Robert M. Gower
Romain Mueller
Ruizhe Li
Sadegh Aliakbarian
Safa Alver
Shahab Bakhtiari
Shangzhe Wu
Sharath Chandra Raparthy
Steffen Schneider
Steinar Laenen
Taoli Cheng
Tom Joy
Udo M. Schlegel
Victor Schmidt
Vincent Dumoulin
Vincent Mai
Viveka Kulharia
Xavier Gibert
Xutan Peng
Yana Hasson
Yongtuo Liu
Yuge Shi
Yuki M. Asano
Yunhua Zhang
Zhao Yang
Zhongdao Wang

A huge thank you to our reviewers!

Organisers

Samuel Albanie
Univ. of Cambridge

João F. Henriques
Univ. of Oxford

Luca Bertinetto
Five

Alex Hernández-García
Mila (Quebec AI Institute)
and Université de Montréal

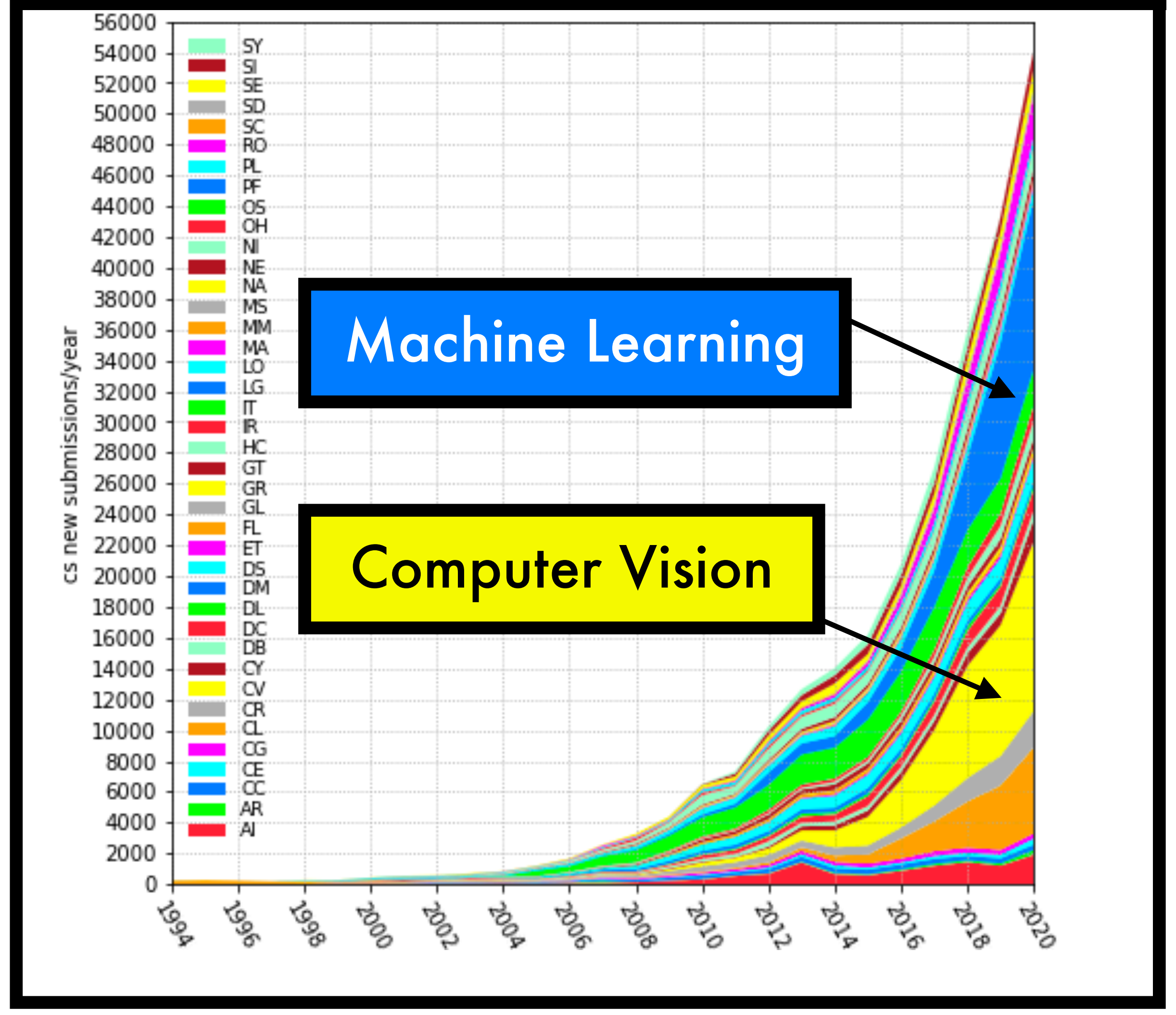
Hazel Doughty
Univ. of Amsterdam

Gül Varol
École des Ponts ParisTech

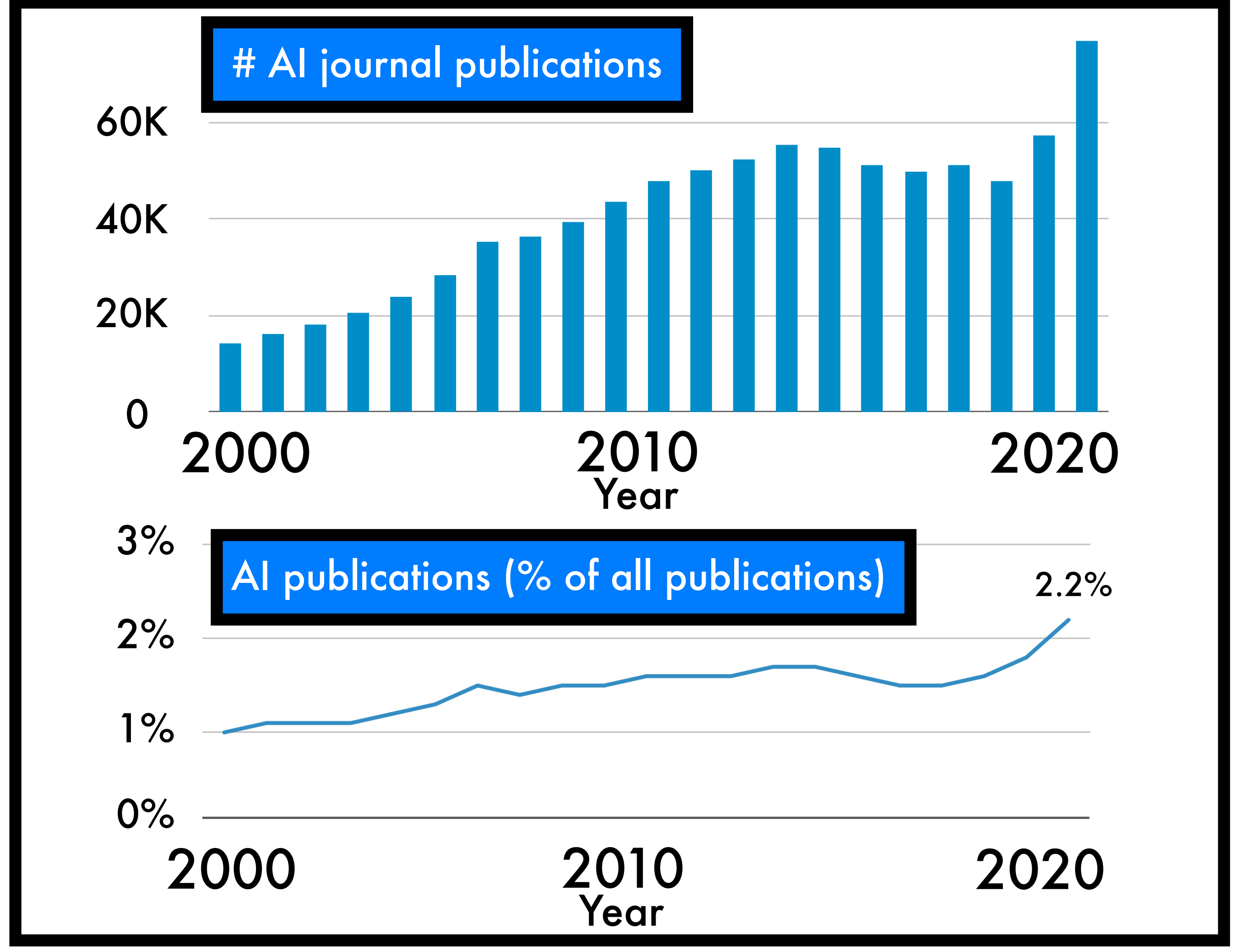
Motivation for the workshop

The field of Machine Learning is thriving

Arxiv (cs.)



AI-related journal publications



Breakthrough applications

- AlphaFold
- Weather Forecasts
- GPT-3
- Pure Mathematics
- StarCarft II
- DFT - DM21

ML Review Cycle

Peer review lies at the heart of modern ML research

Authors experiment/ write paper → Double-blind review comments → Rebuttal → Decision

Typical ML review flow

Is there room for improvement?

Factors that are not well addressed by typical ML review

- The Hidden Compute Variable
- The Tyranny of SoTA
- Negative Results

References

arxiv stats figures https://github.com/arXiv/arxiv-docs/tree/develop/help/stats/2020_by_area

Journal figures: https://aiindex.stanford.edu/wp-content/uploads/2021/03/2021-AI-Index-Report-_Chapter-1.pdf

Jumper et al. "Highly accurate protein structure prediction with AlphaFold." *Nature* (2021)

Brown et al. "Language Models are Few-Shot Learners." *NeurIPS 2020*

Vinyals et al. "Grandmaster level in StarCraft II using multi-agent reinforcement learning." *Nature* (2019)

Ravuri et al. "Skilful precipitation nowcasting using deep generative models of radar." *Nature* (2021)

Davies et al. "Advancing mathematics by guiding human intuition with AI." *Nature* (2021)

Kirkpatrick et al. "Pushing the frontiers of density functionals by solving the fractional electron problem." *Science* (2021)

The Hidden Compute Variable

Hidden information

In the typical ML review process reviewers (and paper readers) do not know how many experiments were run.

Issue if significant disparity in computing resources available to different researchers evaluating on same benchmarks.

Blue Researcher (3 GPUs)

	Arch a	Arch b	Arch c
Run 1	72.6	73.8	76.6
Run 2	70.4	71.5	76.7
Run 3	78.9	73.8	73.7
Mean	73.1	70.2	72.7
Std	4.4	1.3	1.7

Same distribution

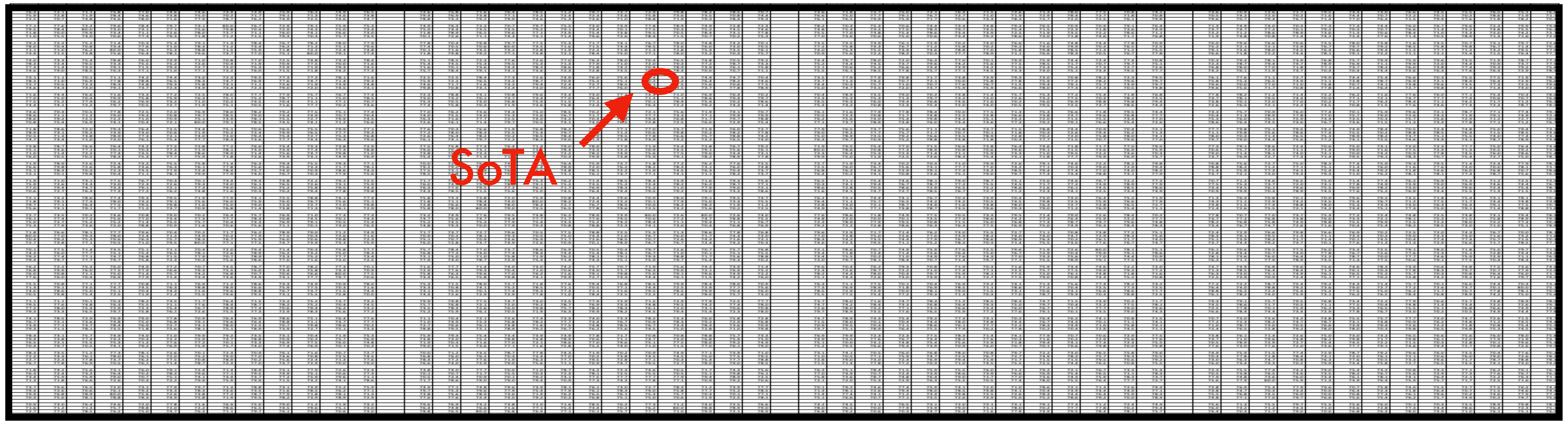
Red Researcher (13 GPUs)

Arch A	Arch B	Arch C	Arch D	Arch E	Arch F	Arch G	Arch H	Arch I	Arch J	Arch K	Arch L	Arch M
79.4	71.3	72.3	78.2	79.4	72.7	78.6	73.0	78.3	72.2	79.6	78.0	72.9
75.6	76.6	75.7	76.2	77.3	77.1	78.6	70.5	79.3	74.9	75.9	78.3	70.0
74.7	77.0	79.6	71.7	78.4	72.9	75.5	72.8	70.3	71.0	73.8	73.4	72.2
79.2	72.3	70.8	70.3	72.5	70.3	72.0	79.0	74.1	70.4	78.4	71.9	72.5
2.5	3.2	3.6	3.3	1.1	2.5	1.8	1.4	4.9	2.0	2.9	2.7	1.5

What the reviewer sees

	Arch a	Arch A
Run 1	72.6	79.4
Run 2	70.4	75.6
Run 3	78.9	74.7
Mean	73.1	79.2
Std	4.4	2.5

Table 1: Importantly, our Arch A outperforms their Arch a by a wide margin, with a lower standard deviation across runs



Key takeaway: Exploratory computation is important. It is hidden.

The Tyranny of SOTA

Benchmarks: benefits

Common experimental benchmark datasets have been incredibly valuable for our field.

- They allow direct, controlled comparisons of methods
- Drive community progress towards important research questions
- Highlight scenarios where existing methods fail

Benchmarks: risks

By practical necessity, benchmarks can often only provide an simplified (imperfect) model for a phenomenon of interest.

Misalignment: Over-reliance on benchmarks can produce make achieving state-of-the-art (SOTA) more important than advancing the collective knowledge of the community about underlying phenomena that we care about.

Misallocation: They can lead to inefficient resource allocation by trapping the community in local minima (neural networks….)

Degradation: Statistical power heavily affected by disparities in exploratory compute; weakens over time.

Key takeaway: Benchmarks are very important, but they have challenges

Negative Results

Incentives for Negative Results

Well-motivated, well-executed experiments can provide inconclusive and/or negative results.

No incentive to invest time in preparing such results for publication, since they would be highly unlikely to be accepted.

But these results can convey useful information for the community:

- Avoid duplication (wasted resources)
- Provide insight (particularly if detailed studies are conducted to understand the cause of the negative result)

Key takeaway: No incentive for negative results in typical ML review process

Pre-registration Protocol

Central Idea

Objective

Enforce a clear separation between:

- Postdiction (exploratory analysis)
- Prediction (hypothesis testing)

Benefits

The Hidden Compute Variable	Reviewers see full study plan
The Tyranny of SoTA	Reviewer cannot accept/ reject based on SoTA
Negative Results	Review before results are known

Limitations

Not suitable for all types of paper (highly exploratory work, for example)

Not a defence against dishonesty

Reference

Figure credit: https://www.cos.io/initiatives/registered-reports

Pre-registration - workshop

Workshop protocol

Objective: conduct a full pre-registration review process

- Submit a pre-registration proposal (4-5 pages)
- Two rounds of review (including rebuttal)

10 accepted proposals (from 22 submissions) will be presented here today.

What about results?

This is the second NeurIPS pre-registration workshop for Machine Learning.

Following the workshop last year, accepted results papers were published at a special edition of **PMLR**.

We will hear from the authors of of several papers from last year who published results papers in the workshop today!

Authors of accepted proposals will be invited to submit their results to another special issue of **PMLR**.

Sessions

- **Invited talks** will be followed by live Q&A
- **Spotlights** will be followed by live shared Q&A
  - If you are an author in the session, please join the zoom link (at https://neurips.cc/virtual/2021/workshop/21885) to answer questions
- The **poster session** will take place on **GatherTown.** The link can be found at http://preregister.science/
- The **final session** is a live **open discussion** - everyone is invited to participate.

Asking Questions

Questions can be asked on RocketChat - we will read them aloud to the speakers

Important note: There is a delay of 60 seconds between the zoom session and the NeurIPS streaming webpage (so it's best to write your questions before the end of the talk).

If you wish to ask a question in person, join the zoom link and raise your hand (so we can promote you as panelist).

Recording notice: we are recording the workshop. Please let us know afterwards if you would like to be removed from the recording before it is shared.

Workshop schedule (GMT)

Mon 12:00 p.m. - 12:10 p.m.	**Opening remarks** (Talk)	
Mon 12:10 p.m. - 12:40 p.m.	**Invited Talk - Sarahanne Field** (Talk) SlidesLive Video »	*Sarahanne Field*
Mon 12:40 p.m. - 1:00 p.m.	**PCA Retargeting: Encoding Linear Shape Models as Convolutional Mesh Autoencoders - Eimear O'Sullivan** (Talk)	
Mon 1:00 p.m. - 1:20 p.m.	**Spotlights 1 (5 x 3 minutes)** (Short videos)	
Mon 1:20 p.m. - 1:40 p.m.	**Unsupervised Resource Allocation with Graph Neural Networks - Miles Cranmer** (Talk)	
Mon 1:40 p.m. - 2:10 p.m.	**Break**	
Mon 2:10 p.m. - 2:40 p.m.	**Invited Talk - Dima Damen** (Talk)	*Dima Damen*
Mon 2:40 p.m. - 3:10 p.m.	**Invited Talk - Hugo Larochelle** (Talk)	*Hugo Larochelle*
Mon 3:10 p.m. - 3:30 p.m.	**Spotlights 2 (5 x 3 minutes)** (Short videos)	
Mon 3:30 p.m. - 4:30 p.m.	**Poster Session** (Virtual posters) link »	
Mon 4:30 p.m. - 5:00 p.m.	**Break**	
Mon 5:00 p.m. - 5:30 p.m.	**Invited Talk - Paul Smaldino** (Talk)	*Paul Smaldino*
Mon 5:30 p.m. - 5:50 p.m.	**Confronting Domain Shift in Trained Neural Networks - Carianne Martinez** (Talk)	
Mon 5:50 p.m. - 6:05 p.m.	**Discussion Panel - 2020 authors' experience** (Discussion Panel)	
Mon 6:05 p.m. - 7:05 p.m.	**Open Discussion**	
Mon 7:05 p.m. - 7:10 p.m.	**Closing Remarks**	

We are here

Schedule can be found on NeurIPS workshop page (https://neurips.cc/virtual/2021/workshop/21885)

and the workshop webpage **http://preregister.science/**

Thank you!

Up-to-date schedule and details at
http://preregister.science/