

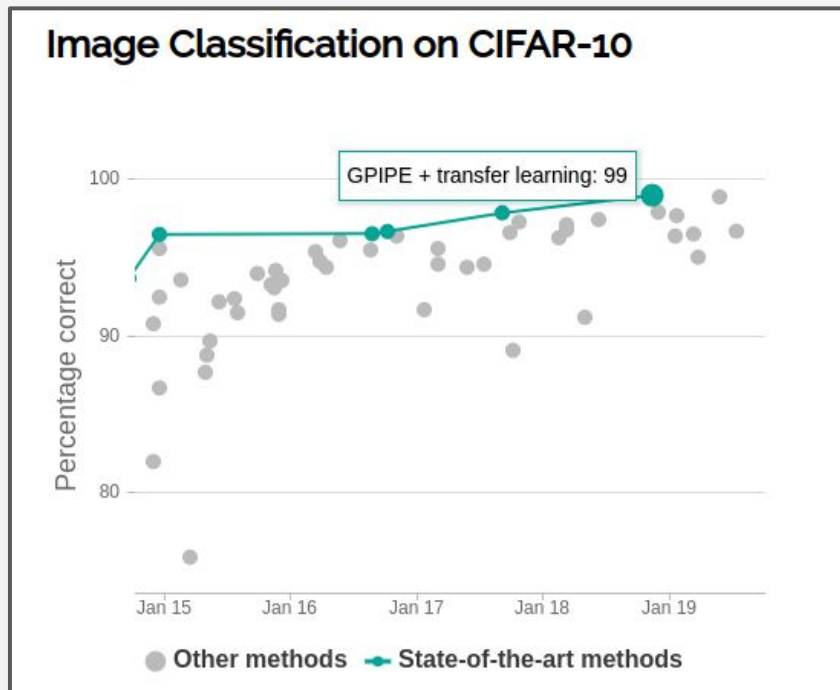
# AI for Common Good through benchmarking at scale

Laure Delisle



AI research is doing  
great!

# Performant



Performance

	Model	Comparison to best public-available results			
		Acc.	#Param	Our Model	Acc.
CIFAR-10	NASNet-A	98.0%	85M	EfficientNet-B0	98.1%
CIFAR-100	NASNet-A	87.5%	85M	EfficientNet-B0	88.1%
Birdsnap	Inception-v4	81.8%	41M	EfficientNet-B5	82.0%
Stanford Cars	Inception-v4	93.4%	41M	EfficientNet-B3	93.6%
Flowers	Inception-v4	98.5%	41M	EfficientNet-B5	98.5%
FGVC Aircraft	Inception-v4	90.9%	41M	EfficientNet-B3	90.7%
Oxford-IIIT Pets	ResNet-152	94.5%	58M	EfficientNet-B4	94.8%
Food-101	Inception-v4	90.8%	41M	EfficientNet-B4	91.5%

Benchmarking

# Prolific

## Browse state-of-the-art

[1407 leaderboards](#) • 1291 tasks • 1282 datasets • 16093 papers with code

**Computer Vision**

- Semantic Segmentation: 1st. 27 leaderboards, 598 papers with code
- Image Classification: 1st. 51 leaderboards, 508 papers with code
- Object Detection: 1st. 52 leaderboards, 421 papers with code
- Image Generation: 1st. 48 leaderboards, 208 papers with code
- Denoising: 1st. 17 leaderboards, 162 papers with code

**Natural Language Processing**

- Machine Translation: 1st. 42 leaderboards, 470 papers with code
- Language Modelling: 1st. 8 leaderboards, 392 papers with code
- Question Answering: 1st. 40 leaderboards, 378 papers with code
- Sentiment Analysis: 1st. 21 leaderboards, 304 papers with code

**Medical**

- Medical Image Segmentation: 1st. 31 leaderboards, 58 papers with code
- Lesion Segmentation: 1st. 4 leaderboards, 31 papers with code
- Drug Discovery: 1st. 7 leaderboards, 31 papers with code
- Brain Segmentation: 1st. 1 leaderboard, 18 papers with code

**Methodology**

- Word Embeddings: 1st. 4 leaderboards, 324 papers with code
- Representation Learning: 1st. 4 leaderboards, 324 papers with code
- Transfer Learning: 1st. 11 leaderboards, 293 papers with code
- Domain Adaptation: 1st. 11 leaderboards, 230 papers with code

**Speech**

- Speech Recognition: 1st. 17 leaderboards, 162 papers with code
- Speech Synthesis: 1st. 2 leaderboards, 27 papers with code
- Speaker Verification: 1st. 20 papers with code
- Voice Conversion: 1st. 18 papers with code
- End-To-End Speech: 1st. 17 leaderboards, 162 papers with code

**Time Series**

- Time Series: 1st. 1 leaderboard, 262 papers with code
- Time Series Classification: 1st. 2 leaderboards, 43 papers with code
- Imputation: 1st. 7 leaderboards, 33 papers with code
- Time Series Forecasting: 1st. 6 leaderboards, 21 papers with code

**Audio**

- Stereo Matching: 1st. 24 papers with code
- Music Generation: 1st. 18 papers with code
- Audio Generation: 1st. 11 papers with code
- Audio Classification: 1st. 10 papers with code

**Robots**

- Calibration: 1st. 83 papers with code
- Visual Odometry: 1st. 30 papers with code
- Motion Planning: 1st. 13 papers with code
- Robot Navigation: 1st. 12 papers with code

**Computer Code**

- Dimensionality Reduction: 1st. 131 papers with code
- Program Synthesis: 1st. 17 papers with code
- Code Generation: 1st. 2 leaderboards, 13 papers with code
- Text-To-Sql: 1st. 1 leaderboard, 12 papers with code
- Classification Metrics: 1st. 4 papers with code

**Reasoning**

- Decision Making: 1st. 175 papers with code
- Common Sense Reasoning: 1st. 38 leaderboards, 38 papers with code
- Visual Reasoning: 1st. 19 papers with code
- Decision Making Under Uncertainty: 1st. 4 papers with code
- Math Word Problem Solving: 1st. 2 papers with code

**Knowledge Base**

- Knowledge Graphs: 1st. 102 papers with code
- Knowledge Base Completion: 1st. 21 papers with code
- Knowledge Graph Completion: 1st. 18 papers with code
- Causal Discovery: 1st. 17 papers with code
- Open Knowledge Graph Canonicalization: 1st. 1 papers with code

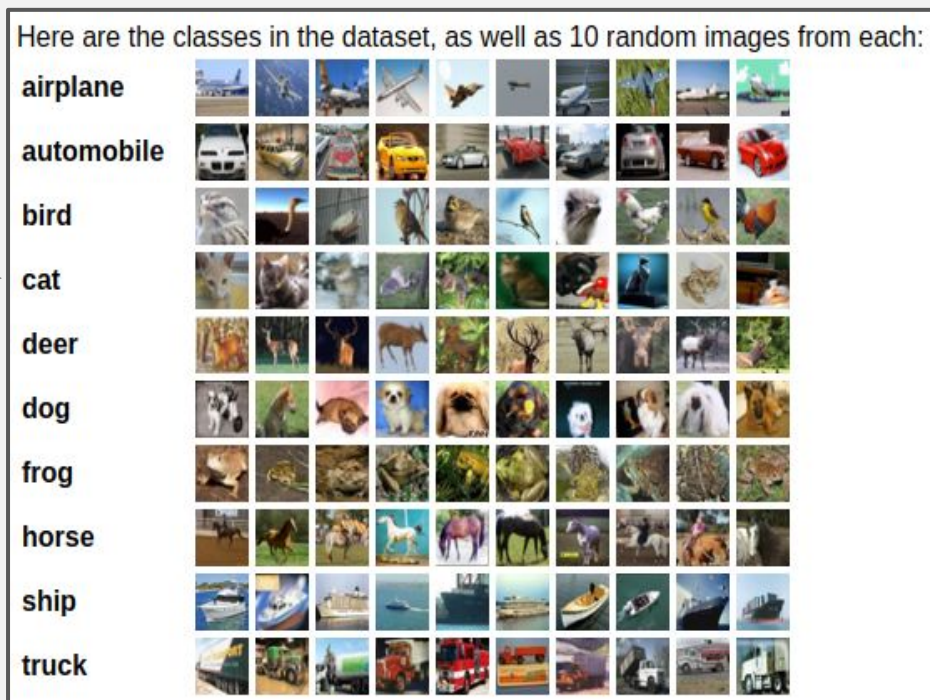
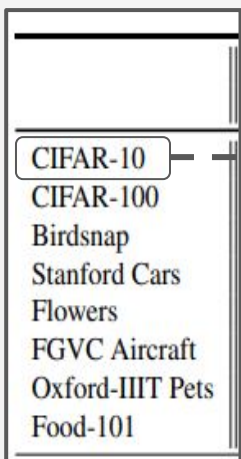
**Adversarial**

- Adversarial Attack: 1st. 68 papers with code
- Adversarial Defense: 1st. 41 papers with code
- Adversarial Text: 1st. 5 papers with code
- Inference Attack: 1st. 4 papers with code
- axycle: 1st. 1 papers with code

State Of The Art

We can do even better

# Current benchmarking datasets...



# Datasets for real world *impact*



የዚግ ነት ከብር በሊት ዩኦ ያቸገ ጸገ ተፎ  
ታዩ ህዝባ ዊነት ጻር ለብባር በር ተፎ  
ለባላዎ ለፍተህ ለሕዝቦች ነፃ ነት፣  
በእኩልነት በፍኛ ጭናል በሌገድነት፣  
ወሠረተ ጸኑ ሰብእናገ ሃልባርገ፣  
ሕዝቦች ነገ ለሥራ በሥራ የኖርገ።



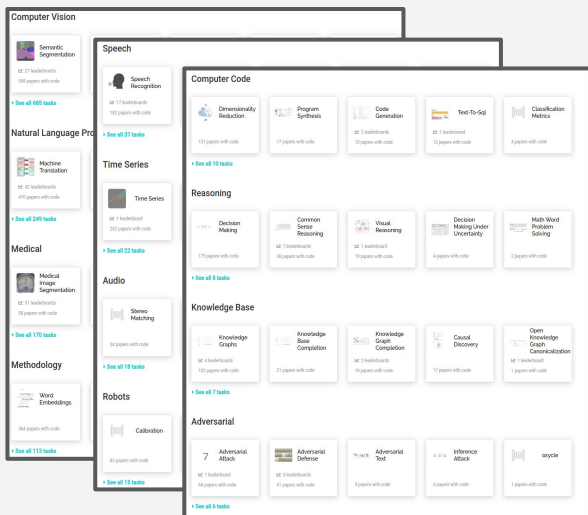
Underserved datasets

**Cassava leaves dataset** - Ramcharan et al. [Deep Learning for Image-Based Cassava Disease Detection](#). Frontiers in Plant Science 2017.

**Ethiopian datasets** - [Knoema](#). Illustration: [Ethiopian anthem](#) in Ahmaric

**Somali audio transcription dataset** - [Shaqodoon](#). 2019

# Filling the gap



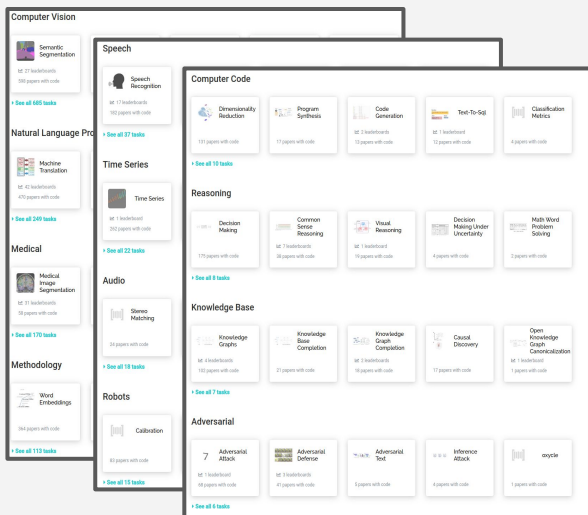
State Of The Art



Underserved datasets



# Filling the gap



State Of The Art

- Incentives ➔

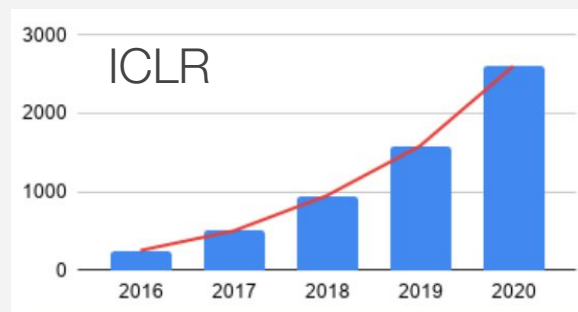
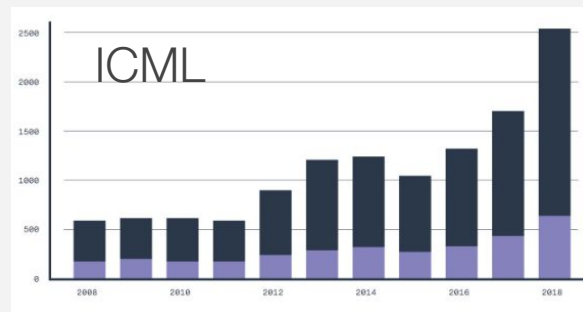
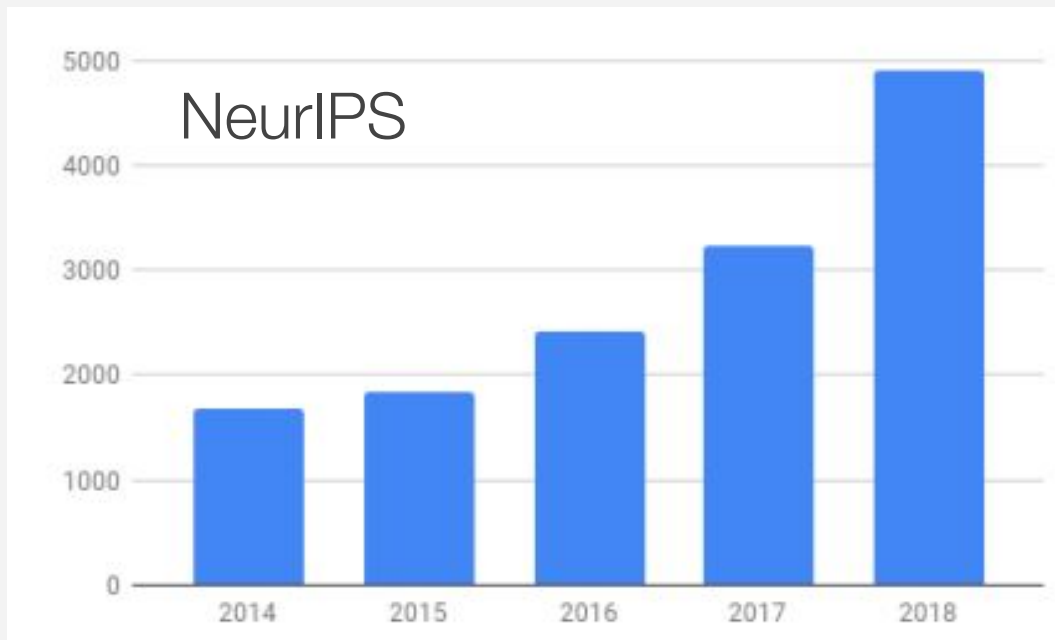


Underserved datasets

Exploiting the system

In 2018, the number of conference paper submissions has increased by 47% for ICML, by 50% for NeurIPS, and by almost 100% for ICLR.

# AI papers submitted



**NeurIPS** - D. Sculley, Jasper Snoek and Alex Wiltschko. [Avoiding a Tragedy of the Commons in the Peer Review Process.](#)

Critiquing and Correcting Trends in Machine Learning, NeurIPS 2018 Workshop.

**ICML** - Peltarion. <https://peltarion.com/article/icml-2018-an-ai-party-in-our-own-backyard>

**ICLR** - @iclr\_conf. [Tweet](#) on Sep 25 2019.

# Checking boxes

Checklists Are Great in list [To Do](#)

[Edit the description...](#)

Checklist [Hide completed items](#) [Delete...](#)

67%

~~Make a checklist~~

~~Complete checklist~~

Write about how great checklists are

[Add an item...](#)

Checklist illustration - [Trello](#)

NeurIPS checklist - @blaine\_bateman. [Tweet](#) on May 9 2019.

**Blaine Bateman**  
@blaine\_bateman

The Machine Learning Reproducibility Checklist  
(required for all NeurIPS 2019 papers)  
[buff.ly/2Y8tYnP](http://buff.ly/2Y8tYnP)

**The Machine Learning Reproducibility Checklist (Version 1.2, Mar27 2019)**

For all **models** and **algorithms** presented, check if you include:

- A clear description of the mathematical setting, algorithm, and/or model.
- An analysis of the complexity (time, space, sample size) of any algorithm.
- A link to a downloadable source code, with specification of all dependencies, including external libraries.

For any **theoretical claim**, check if you include:

- A statement of the result.
- A clear explanation of any assumptions.
- A complete proof of the claim.

For all **figures** and **tables** that present empirical results, check if you include:

- A complete description of the data collection process, including sample size.
- A link to a downloadable version of the dataset or simulation environment.
- An explanation of any data that were excluded, description of any pre-processing step.
- An explanation of how samples were allocated for training / validation / testing.
- The range of hyper-parameters considered, method to select the best hyper-parameter configuration, and specification of all hyper-parameters used to generate results.
- The exact number of evaluation runs.
- A description of how experiments were run.
- A clear definition of the specific measure or statistics used to report results.
- Clearly defined error bars.
- A description of results with **central tendency** (e.g. mean) & **variation** (e.g. stddev).
- A description of the computing infrastructure used.

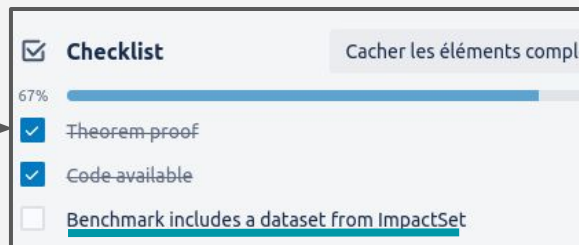
Reproduced from: [www.cs.mcgill.ca/~jpineau/ReproducibilityChecklist.pdf](http://www.cs.mcgill.ca/~jpineau/ReproducibilityChecklist.pdf)

ImpactNet

# ImpactNet by AI COMMONS



Gather the datasets



Incentivize

	Model	Comp Acc.
CIFAR-10	NASNet-A	98.0%
CIFAR-100	NASNet-A	87.5%
Birdsnap	Inception-v4	81.8%
Stanford Cars	Inception-v4	93.4%
Flowers	Inception-v4	98.5%
FGVC Aircraft	Inception-v4	90.9%
Oxford-IIIT Pets	ResNet-152	94.5%
Food-101	Inception-v4	90.8%
<u>ImpactNet</u>		%

AI prototype