# SEER

Researchers from Meta AI and Inria have trained a 10 billion parameter model on 1 billion Instagram images!
They trained the model on an uncurated dataset of 1 billion images and even if the data is uncurated, possibly containing the worst of what the Internet has to offer, the model is more robust and fair.

**How can a model trained on uncurated data be fairer, when these models tend to reflect and sometimes even emphasize the biases in the data?**

Large self-supervised models do seem to follow the garbage in garbage out principle. GPT-3 which was trained on large datasets scraped from the internet. It produces toxic output, false information, and conspiracy theories, it exhibits sexism and racism.

**But what about large self-supervised vision models?**

In the visual realm, the toxicity meter is high too, because large vision models trained on ImageNet, reflect the biases in the data.
ImageNet contains problematic content and is biased by covering some geographical locations more than others, and some races more than others.

**What to do to fix this toxic model behavior?**

What people did so far, especially for ImageNet, was: curate the dataset.
Also, for text, OpenAI released a better version of GPT-3, called InstructGPT, by fine-tuning GPT-3 on a curated dataset. Ok, so curating datasets is one solution, but it's extremely laborious and costly to annotate datasets.

**So, what to do to create fairer models if dataset curation is so expensive?**

The idea of SEER is not new: It relies on the experimental evidence that the more data, the better the features the neural network can learn.
This is what the developments over the last years have shown us.
Okay then, the authors took 1 billion random public images from Instagram.
The only criterion was that the images should come from non-EU countries to comply with GDPR.

Okay, now the Meta AI researchers sit with these 1 billion images, what to do with them?
Curate them?
Filter them?
The dataset is too big for that.

The authors trained a model and called it SEER
which is just a RegNet architecture.
A RegNet is a variant of the well-known ResNet where the residual connections are regulated by a recurrent neural network (RNN). The previous biggest model in the RegNet family was only 1.5 billion parameters big.
In order not to underfit the data, the model size should match or surpass the dataset size, so the authors scaled up the RegNet to make SEER, a RegNet model scaled up to 10 billion parameters.

SEER was pretrained with SwAV self-supervision.
This basically teaches the model to compute representations of the data such that the representations of different patches of the same image are assigned to the same cluster.
And then for using and testing SEER on interesting downstream tasks, the authors appended a linear
layer and fine-tuned the model for each task.
So, in terms of architecture and self-supervised training objectives, SEER does not bring anything new.

SEER really shines when it comes to evaluating the model.
The authors evaluate SEER on 50 benchmarks in total!
And these tests contain classical computer vision tasks like image classification or detecting whether an image is a copy of another, which is useful for enforcing copyright.
The large model trained on huge amounts of data has good visual representations
and is either keeping pace with the competition or establishing new state-of-the-art.

But within those 50 benchmarks, there are also 4 fairness benchmarks.
To summarize very long experimental sections, the authors basically compare a SEER model trained on the 1 billion uncurated Instagram images against a smaller model version trained on ImageNet.
Consistently, SEER is fairer when it comes to metrics measuring potential biases with respect to skin color, age groups, gender, and geographical diversity.

**But how does this happen?**
It is totally expected that SEER, being a big model having seen lots of training data, can impress on computer vision tasks, having learned good visual features.
But why is it fairer?
Why does SEER not reflect biases as we have seen with GPT-3?
Well, the explanation of the authors is that the model, having seen lots and lots of data, is not only capable of capturing better features but other human values too.
However, SEER is also crucially trained on specifically Instagram images. So, SEER is not trained on any uncurated dataset, but on Instagram images.
Instagram pictures are subject to the platform's policy and upload guidelines.
These counter hate speech, targeting individuals, and harassment.
So sure, the authors do not curate the 1 billion Instagram images used for training SEER, but
Instagram's guidelines themselves acted as pre-filtering and data curation to reflect a "diverse community of cultures, ages, and beliefs".