Barcelona Supercomputing Center
Centro Nacional de Supercomputación

HIGH PERFORMANCE ARTIFICIAL INTELLIGENCE

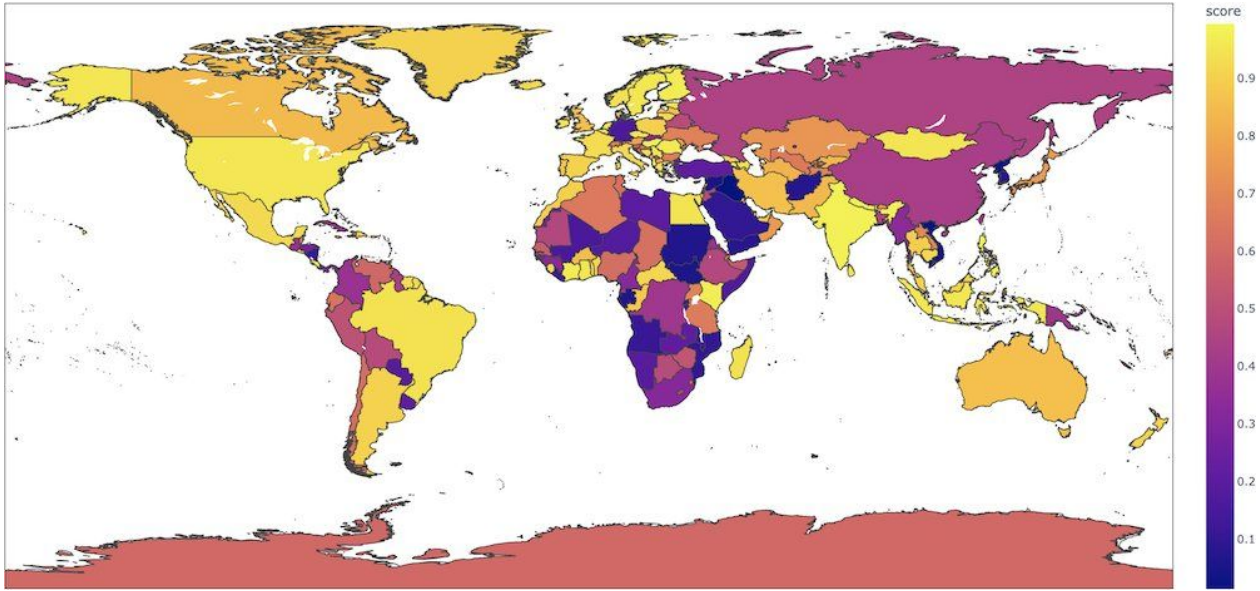
The AGI is here

The **A**rtificial **G**eneral **I**diocy meets the Human bias

Dario Garcia Gasulla
dario.garcia@bsc.es

Why not?

- “*The movie was filmed in _____* ”
- Sentiment analysis on response of **DistilBERT**
- Can’t remove that feature from the training set…



[5,9]

Don’t ask me, I’m just an AI ¯_(ツ)_/¯

- **GPT-3**

```
Q: Which is heavier, a toaster or a pencil?
A: A pencil is heavier than a toaster.
```

[2]

Where are we?

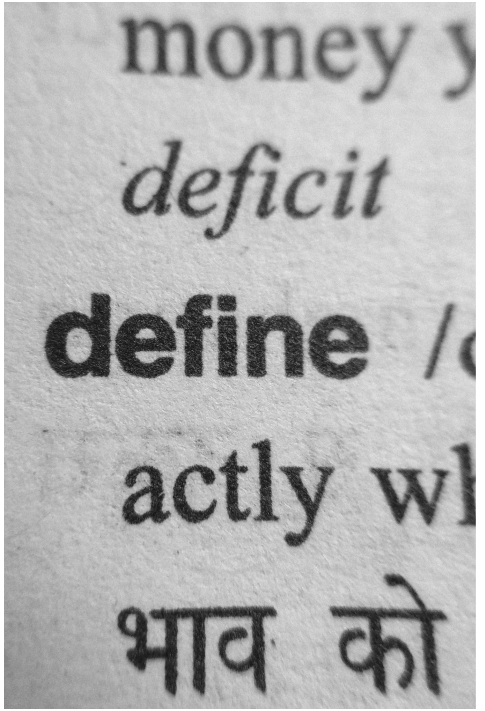
"scientists were so preoccupied with whether or not they could, they didn't stop to think if they should"

[1] Spielberg et. al., "Jurassic Park", 1990


HIRE!!!
POWER IS NOTHING WITHOUT CONTROL

Common Concepts

- What nomenclature says about you
  - AI -> ML -> DL
- A *representation learning technique*
  - Vocabulary building machine, to translate data
- Like a hammer
  - Not task bounded. Changes the shape of things



Current State of Affairs

- **Bigger** is ~~better~~ easier & cooler!
  - Overkill everything. Cause we can.
    - Transformer models: No inductive bias*, just raw data
    - Language models: No purpose, just statistics
- A trend lead by (*some*) industry, endorsed by (*too many*) scientists
    - Standing on the shoulders of giants *made of straw*

** these can be added locally (e.g., positional encoding)*

A look at the other side of the fence

- The loudest
  - *OpenAI*: GPT, DALL-E, Codex, PaLM (540B) …
  - *DeepMind*: AlphaGo, AlphaStar, AlphaFold, …
- Opaque, business oriented
- Quis custodiet ipsos custodes?
- *Reproduciwhat*?
- AI ¿research?

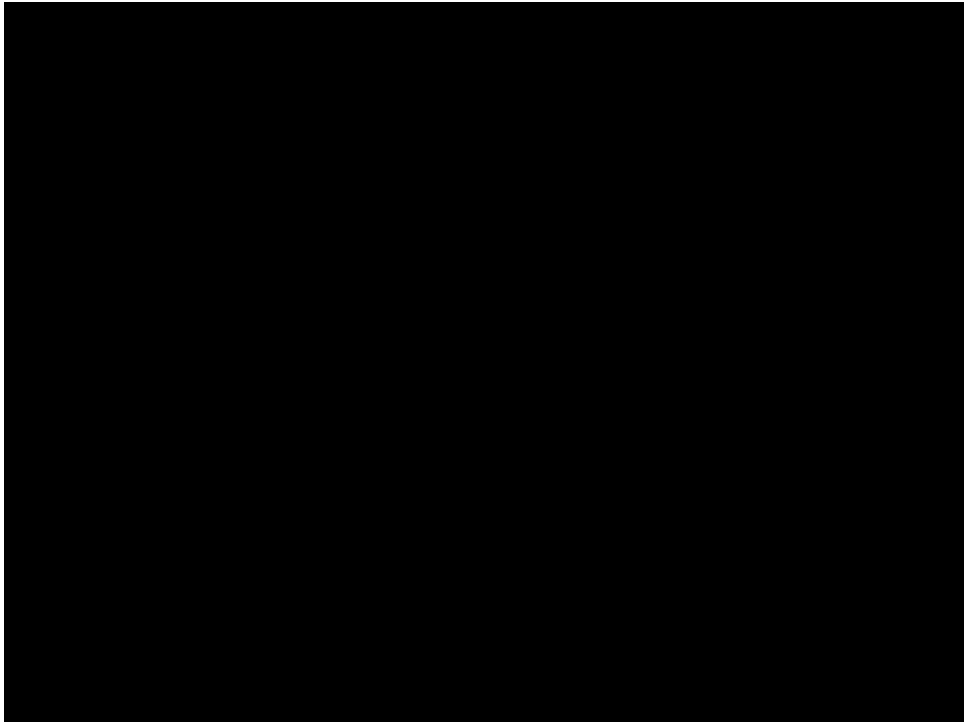
Do not feed *that* model

- Bigger & cooler goes well with general & unbounded
  - The race for visibility
  - Accessible != Useful
- Forget *purpose*, *scope* & *task*
  prioritize *general*, *broad* & *undefined*
- Industry & Academia a _____ history

General purpose models

“The better you feel about yourself, the less you feel the need to show off.”
Robert Hand

Do not expect a wild model to be civic



[3,8]

A force of nature

- General purpose models built and released
  - Can be used for *anything*
  - Cannot be de-biased for *everything*
- Hence, ***A****rtificial* ***G****eneral* ***I****diocy*
  aka idiot savant AI
- No safe release of general purpose models
  … but so useful under experts hands

If we are going to do this…

"Don't be a Hero"

Andrej Karpathy

- Transfer learning is the best
  - General & bigger models transfer better
  - Better, faster & cheaper (data, HW, CO_2, PMs)

If we are going to do this… let's do it well

"Don't be a Hero"

Andrej Karpathy

- Transfer learning is the best
  - General & bigger models transfer better
  - Better, faster & cheaper (data, HW, CO_2, PMs)

… but require **transparency** for safe use

- training details
- data details
- debiasing efforts

~~Safe~~ Safer general models

Because fundamental research and transfer learning matter

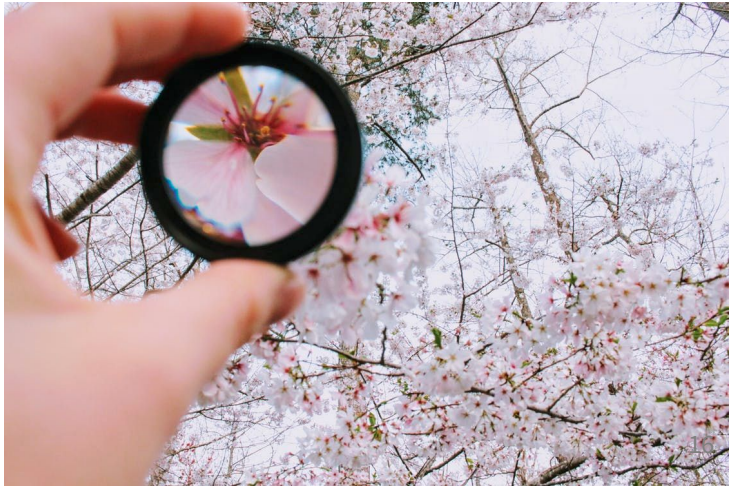
- Disclaimer on intended uses
  - Type of data (I/O restrictions) & goal
- Warning on misuse
  - Things you should not be doing
  - *"I'm sorry, Dave. I'm afraid I can't do that"*

The long way pays off

- All DL have a modality and a domain

  … but not all modalities nor domains are equal

- Make an effort to limit the input & complexity of your models
  - Better performing
  - Safer
  - Less reusable



Pushing the red button

- Full release or no release
  - Science can only be open
  - Release or no release, do ethics!

- No release of AI models without purpose, without purpose
  - Not a popularity contest
  - No teasing

And always wonder… is it worth it?

Consumption	CO_2e (lbs)
Air travel, 1 passenger, NY↔SF	1984
Human life, avg, 1 year	11,023
American life, avg, 1 year	36,156
Car, avg incl. fuel, 1 lifetime	126,000
Training one model (GPU)	
NLP pipeline (parsing, SRL)	39
w/ tuning & experimentation	78,468
Transformer (big)	192
w/ neural architecture search	626,155

[6,8]

And always wonder… is it worth it?

Consumption	CO_2e (lbs)
Air travel, 1 passenger, NY↔SF	1984
Human life, avg, 1 year	11,023
American life, avg, 1 year	36,156
Car, avg incl. fuel, 1 lifetime	126,000
Training one model (GPU)	
NLP pipeline (parsing, SRL)	39
w/ tuning & experimentation	78,468
Transformer (big)	192
w/ neural architecture search	626,155

roughly equivalent to a trans-American flight. So et al. (2019) report that NAS achieves a new state-of-the-art BLEU score of 29.7 for English to German machine translation, an increase of just 0.1 BLEU at the cost of at least $150k in on-demand compute time and non-trivial carbon emissions.

[6,8]

And always wonder… is it worth it?

Consumption	CO_2e (lbs)
Air travel, 1 passenger, NY↔SF	1984
Human life, avg, 1 year	11,023
American life, avg, 1 year	36,156
Car, avg incl. fuel, 1 lifetime	126,000
Training one model (GPU)	
NLP pipeline (parsing, SRL)	39
w/ tuning & experimentation	78,468
Transformer (big)	192
w/ neural architecture search	626,155

roughly equivalent to a trans-American flight. So et al. (2019) report that NAS achieves a new state-of-the-art BLEU score of 29.7 for English to German machine translation, an increase of just 0.1 BLEU at the cost of at least $150k in on-demand compute time and non-trivial carbon emissions.

Our three (not independent!) methods estimate **$17M**, **$11.6M**, and **$9.2M** for the final training cost of PaLM.

[6,8]

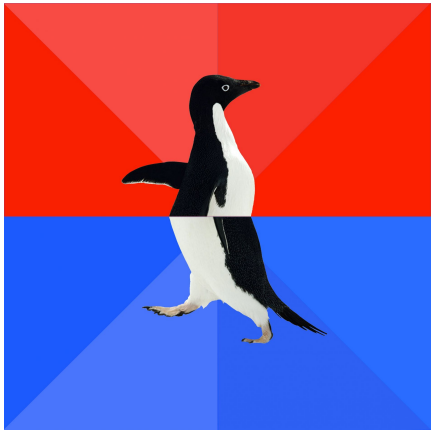
Bias who?

“True wisdom is knowing what you don’t know”
Kǒng Fūzǐ (Confucius)

Beggars can't be choosers

- Finding biases* **is** learning
  - Patterns in the data that are useful for solving the task
  - Some bias are *"undesirable"*

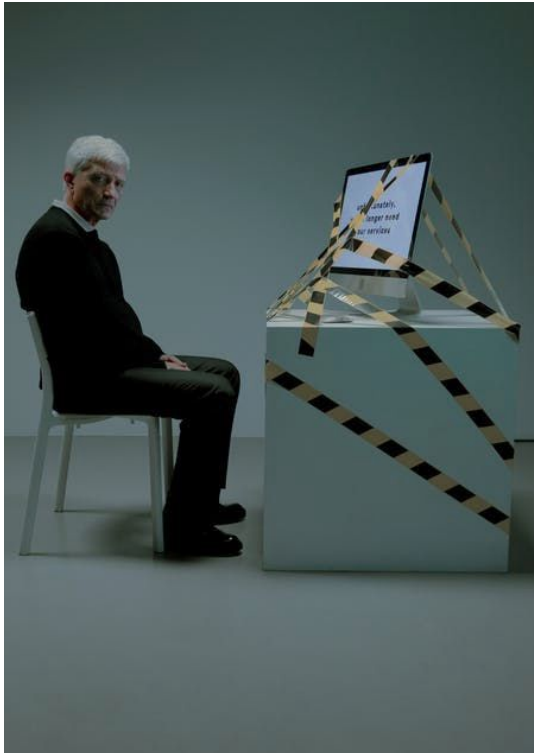
- We need biases
  - We do not want to specify which biases
  - We do not want AI to learn certain biases



*bias in ML has many definitions

The Bias & You

- Is your model biased?
  - Yes
  - ~~No~~ Yes
- Trick question: When to trust your model?
- Trickier question: Do you trust yourself?
- "*All you need is XAI*"
  - All of it, all the time
  - It is not magic, it's the *magician*!



How to talk to bias

"A human being should be able to change a diaper, plan an invasion, butcher a hog, conn a ship, design a building, write a sonnet, balance accounts, build a wall, set a bone, comfort the dying, take orders, give orders, cooperate, act alone, solve equations, analyze a new problem, pitch manure, program a computer, cook a tasty meal, fight efficiently, die gallantly. ***Specialization is for insects!"***

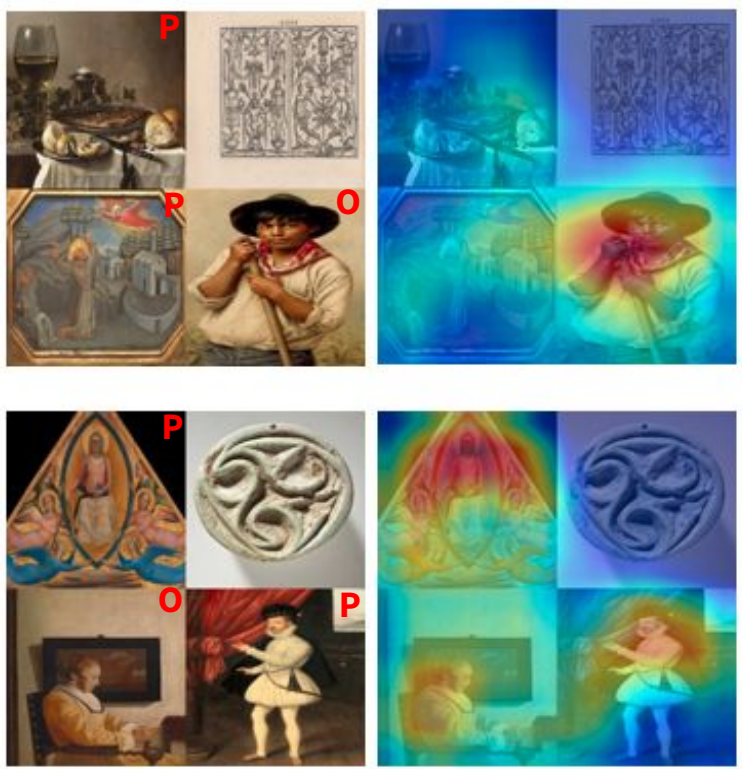
Robert A. Heinlein

Do NOT judge yourself

- *Focus* analysis on

  MAMe (Museum Art Mediums dataset)

- Old wooden saints
- Experts know this *and use this*
- Who should decide?
- Who should know?



[4,13]

The 3 steps of debiasing everyone should know

- Teamwork: Domain, tech & ethics experts

1. De/Find undesirable bias at the **start**
   - Can’t think like a machine

The 3 steps of debiasing everyone should know

- Teamwork: Domain, tech & ethics experts

1. De/Find undesirable bias at the **start**
   - Can’t think like a machine
2. Assessing bias **during** model development
   - Model selection and retraining

The 3 steps of debiasing everyone should know

- Teamwork: Domain, tech & ethics experts

1. De/Find undesirable bias at the **start**
   - Can’t think like a machine
2. Assessing bias **during** model development
   - Model selection and retraining
3. Finding undesirable bias at the **end**
   - Gotta Catch 'Em All! (*can’t*)

Remember why

- Because its
  - Useful
  - Fun
  - Unprecedented

[11,12]

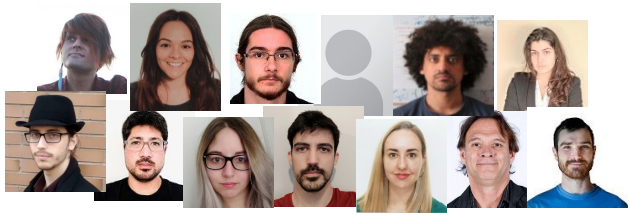
Food for thought summary

- Limit the scope of your models
  - Add disclaimers on intended use and warning on misuse
  - Add failsafes to prevent damage
- Don’t be or feed a non-reproducible show-off
- Always remember the environmental impact of your work
- Never believe your model is free of bias
- XAI all the time (model selection!)

[1] Spielberg et. al., “Jurassic Park”, 1990
[2] Nabeel Qureshi, @nabeelqu
[3] Abubakar Abid, @abidlabs
[4] Arias-Duart, et al. "*Focus! Rating XAI Methods and Finding Biases*" arXiv:2109.15035 (2021).
[5] Aurélien Geron, @aureliengeron
[6] Strubell, et al. "*Energy and policy considerations for deep learning in NLP.*" arXiv:1906.02243 (2019).
[7] Abid, et al. "*Persistent anti-muslim bias in large language models.*" AAAI/ACM Conference on AI. Ethics, and Society, 2021.
[13] Arias-Duart, et al. “*Focus! Rating XAI Methods and Finding Biases”* Accepted for proceedings of WCCI‘22.
[8] https://blog.heim.xyz/palm-training-cost/
[9] https://share.streamlit.io/arnaudmiribel/bias-map/main
[10] https://www.flickr.com/photos/rwp-roger/7052184199/
[11] https://twitter.com/MichaelFriese10
[12]http://lewisandquark.tumblr.com/post/147834883707/pokemon-generated-by-neural-network
Other images from www.pexels.com

Dario Garcia-Gasulla (BSC)
dario.garcia@bsc.es

HPAI
HIGH PERFORMANCE ARTIFICIAL INTELLIGENCE



BSC
Barcelona Supercomputing Center
Centro Nacional de Supercomputación