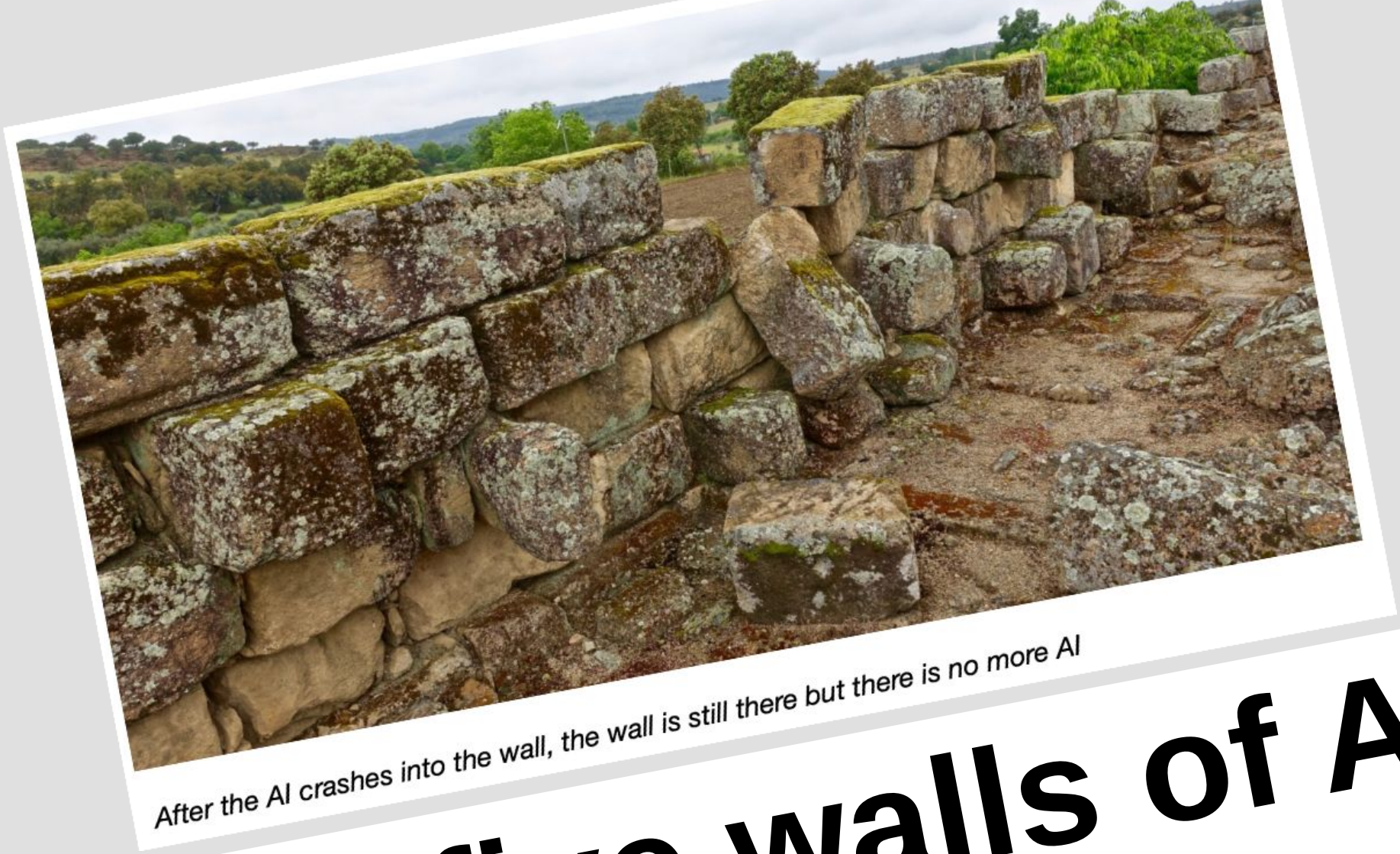


After the AI crashes into the wall, the wall is still there but there is no more AI

The five walls of AI

Bertrand Braunschweig

Prolegomena

- Remarkable progress
  - Image recognition
  - Speech processing (Siri, Alexa, Soundhound ...)
  - Natural language (translation, synthesis, q&a ...)
  - Games (checkers, chess, bricks, go, poker, bridge ...)
  - Decision support (banking, finance, health, ...)
  - Recommendation, personalized ads …
  - Art and creativity (music, pictures, novels, deepfakes...)
  - Science (physics, chemistry, biology, ...)
- Etc. etc.

Five walls

Trust

Energy

Security

Human-computer interaction

Inhumanity

Trust

The three faces of trustworthy AI

Technology 

Reliability, robustness

Lawfulness & Compliance

Accuracy

Security

Safety

Interactions 

Transparency

Explainability

Accountability

Oversight & control

Ethics 

Fairness

Privacy

Diversity & inclusion

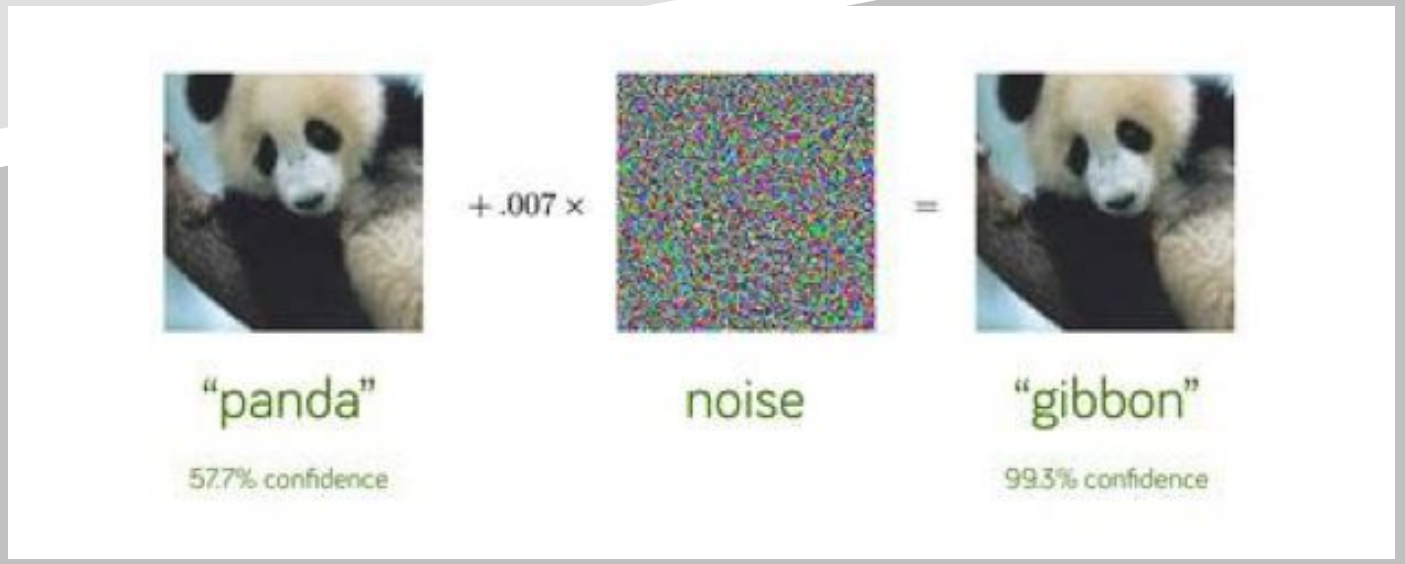
Sustainability

Energy

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

Table 1: Estimated CO_2 emissions from training common NLP models, compared to familiar consumption.[1]

Security



GAN PROGRESS ON FACE GENERATION

Source: Goodfellow et al., 2014; Radford et al., 2016; Liu & Tuzel, 2016; Karras et al., 2018; Karras et al., 2019; Goodfellow, 2019; Karras et al., 2020; AI Index, 2021; Vahdat et al., 2021

Figure

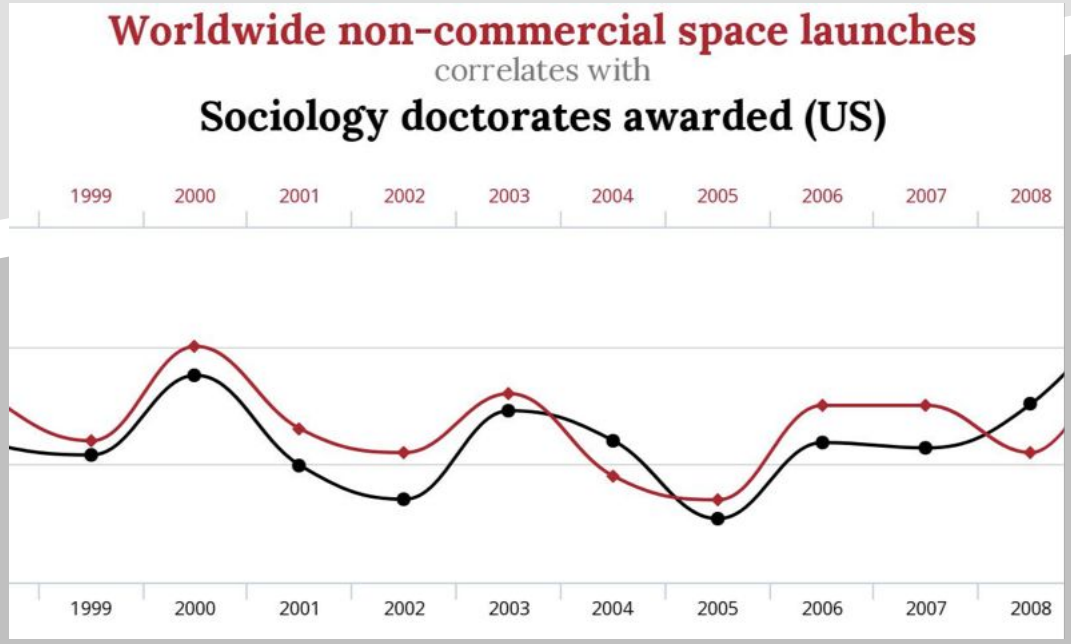
Human-computer interaction

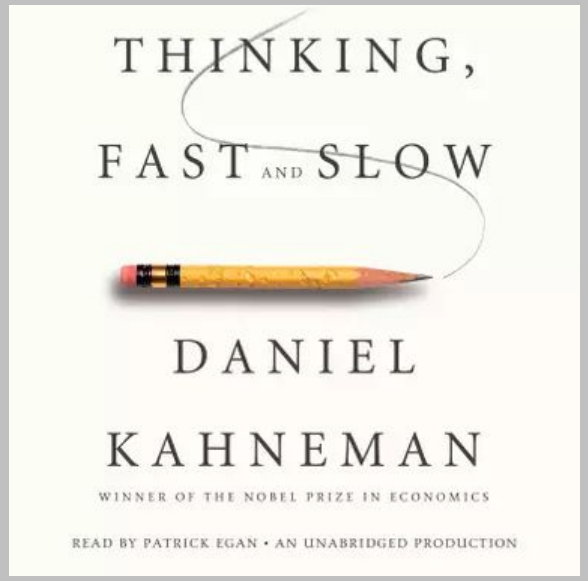
- dialogue (chatbots);
- shared problem solving and decision making;
- sharing of space and resources (cohabitation with robots that are ignored or given orders);
- sharing of tasks (robot teammate). *The « inCobot » case*

The inCobot case (imaginary)

Inhumanity

- Causality-correlation
- Common sense
- System1 – System 2





Avoidance strategies

Trust

Energy

Security

Human-computer interaction

Inhumanity

Trust

A non exhaustive view of the Thrustworthy AI Ecosystem

Energy

Hardware solutions;

Improvement of deep neural network architectures and algorithms

Hybridization with other AI formalisms

Security

I'm afraid the AI security wall is very strong.
Or rather, it has a natural tendency
to repair itself when breached

Digital deterrence ?

(from Inria’s AI white book)

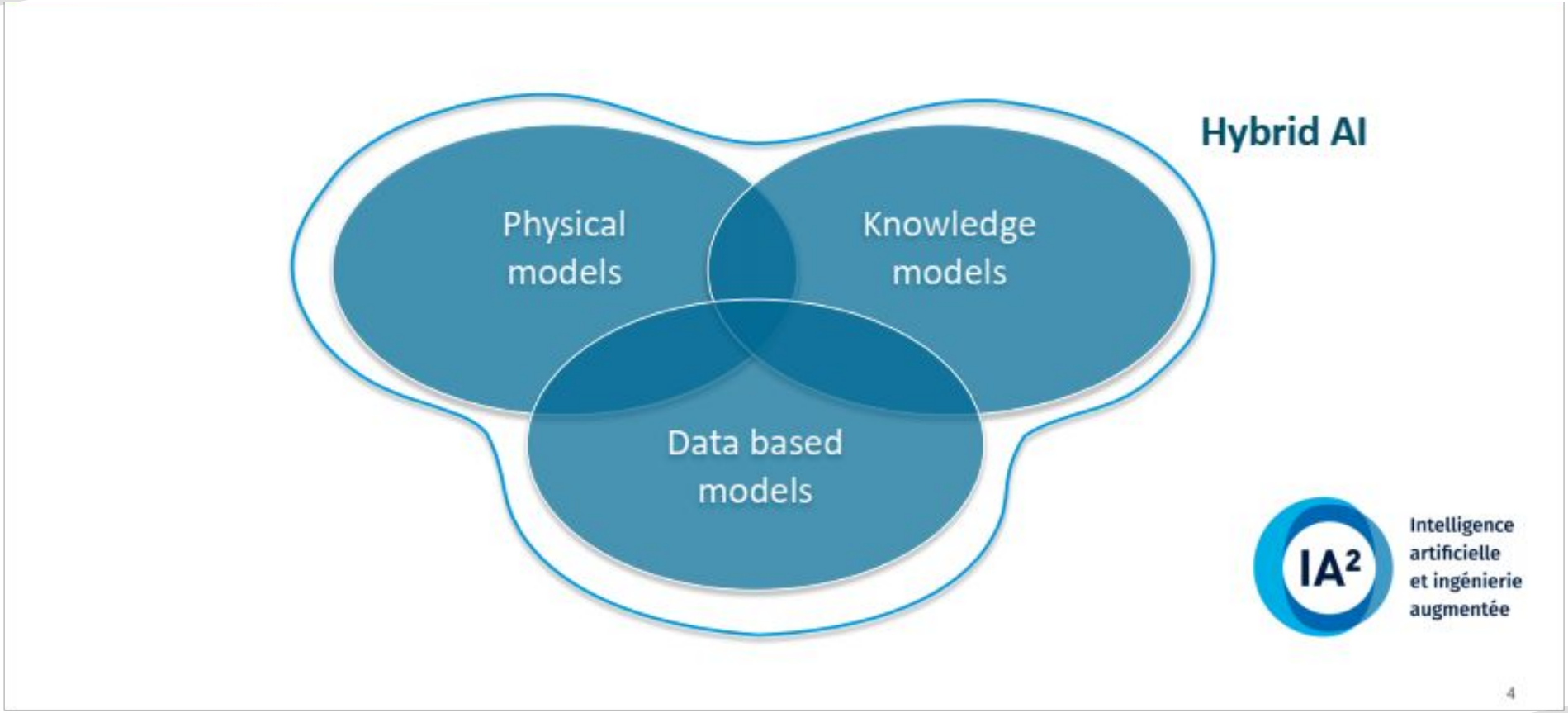
Create a better division of labor between humans and computers,

Bring true transparency and explanation to AI systems

Combine interactive and AI systems so that each leverages the strengths of the other at the appropriate time,

Create better user-centered tools for experts who build AI systems.

Inhumanity


Hybrid AI
Physical models
Knowledge models
Data based models
IA²
Intelligence artificielle et ingénierie augmentée
4

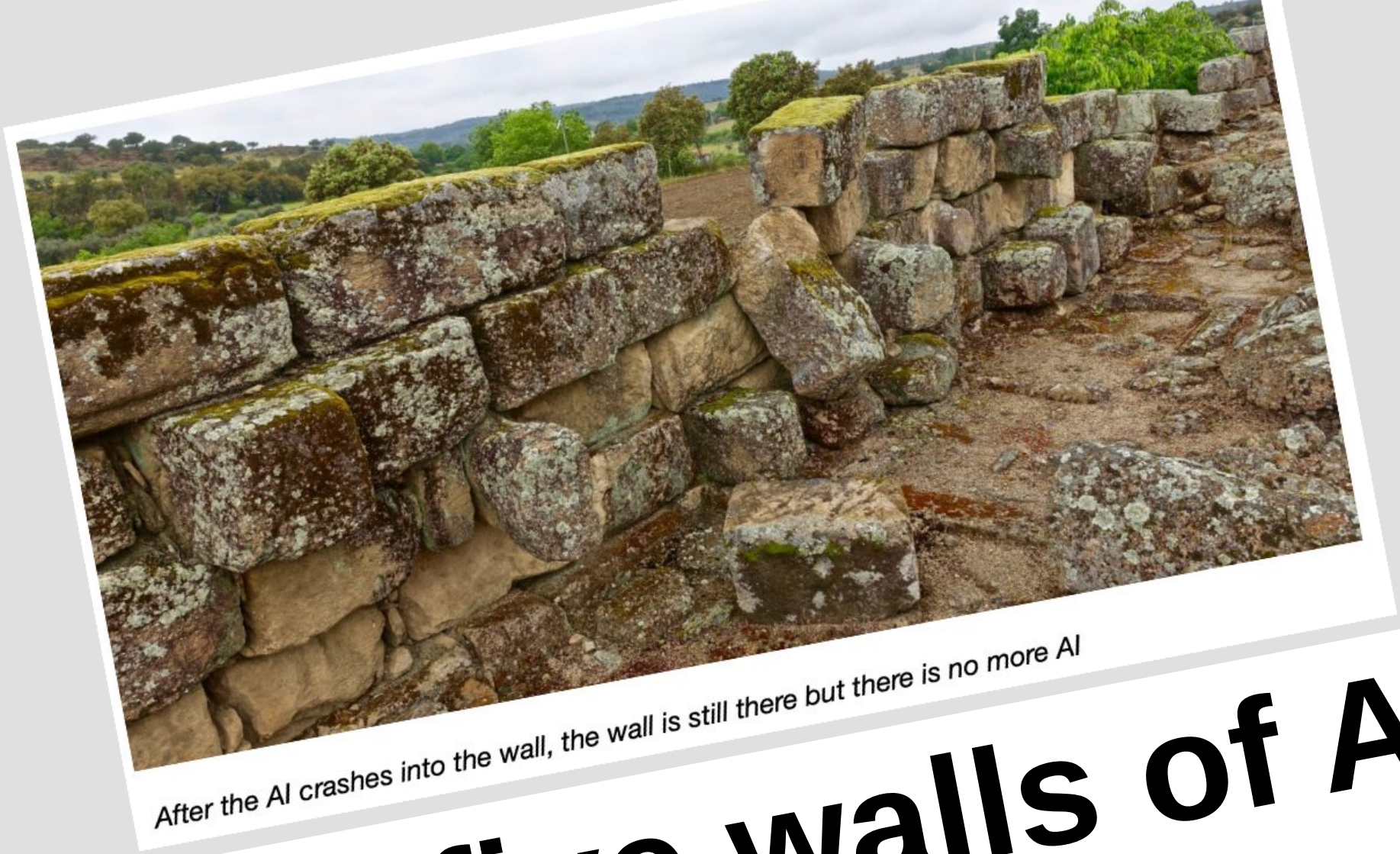
Conclusion

All this to achieve weak AI, specialized on solving
a single or a small number of problems,
even if some have the ambition to develop a General AI.

But let's already try not to crash into the walls of
specialized AI

Additional reading

- « Deep learning for AI », paper by Bengio, LeCun & Hinton
- The 2021 report of « 100 years AI study »
- « How Humans Judge Machines », book by Cesar Hidalgo
- « Human Compatible », book by Stuart Russell
- « AI Index 2022 », annual collective report
- « Deep Learning is hitting a wall », paper by Gary Marcus



After the AI crashes into the wall, the wall is still there but there is no more AI

The five walls of AI

Bertrand Braunschweig