

Czym jest sztuczna inteligencja?

dr Piotr Migdał

p.migdal.pl

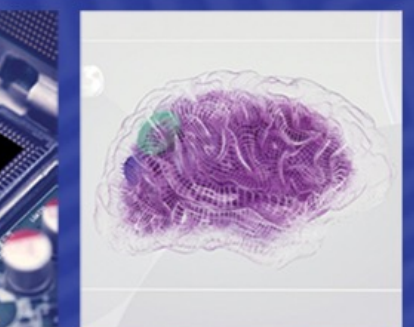
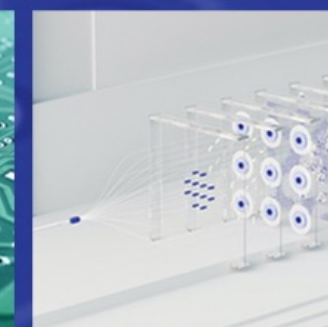
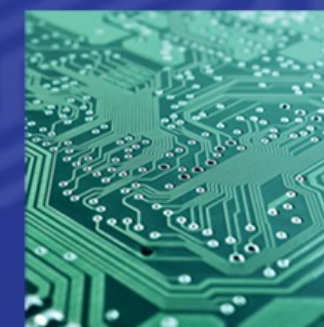
pmigdal@gmail.com

[LinkedIn:piotrmigdal](https://www.linkedin.com/in/piotrmigdal)

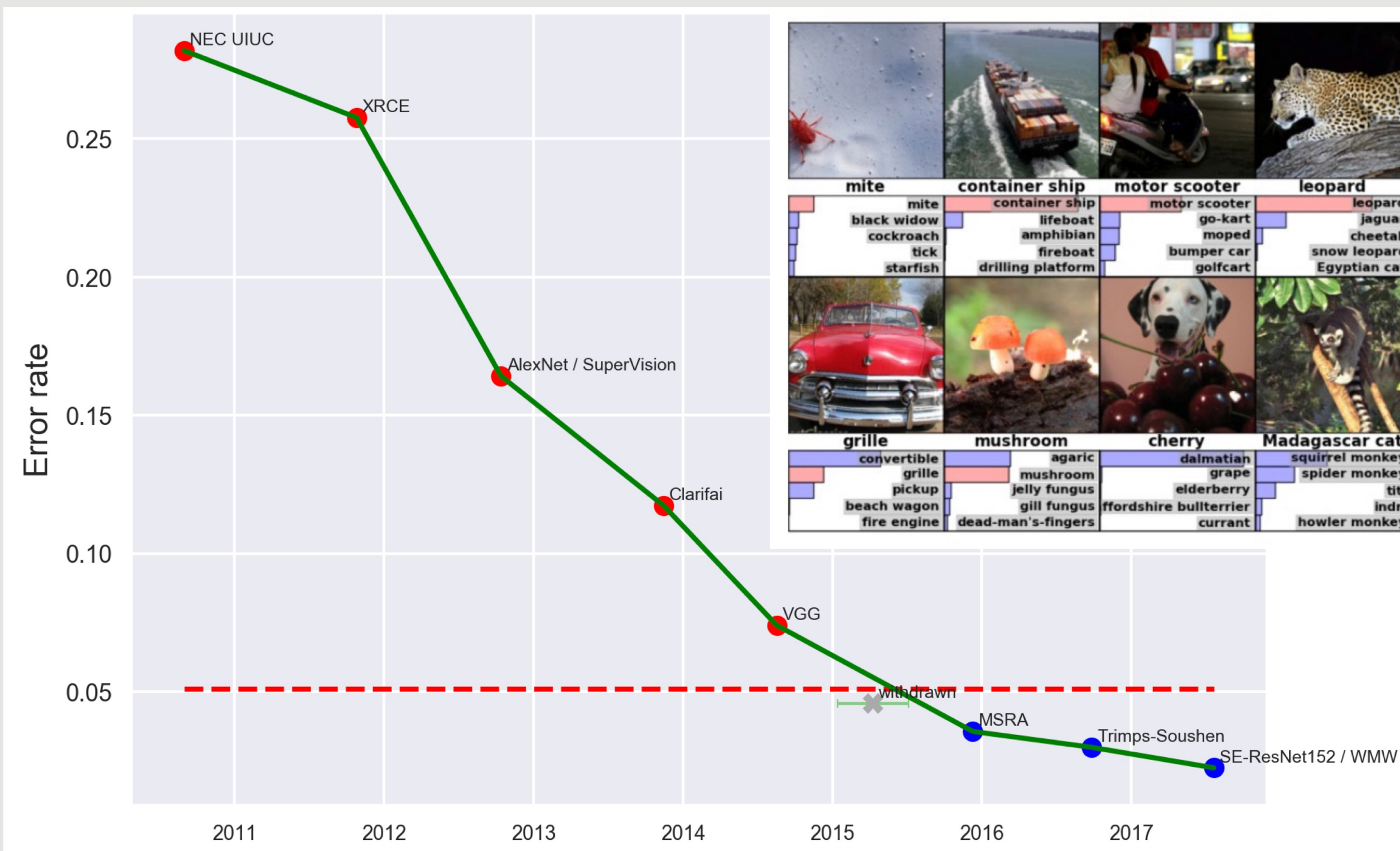
- Doktor fizyki kwantowej, ICFO, Castelldefels
- Współautor polskiej strategii rozwoju AI
- Współzałożyciel Quantum Flytrap
- Konsultant uczenia maszynowego, wizualizacji danych i AI



Ministerstwo
Sprawiedliwości



Błyskawiczny postęp


<https://www.eff.org/ai/metrics>
ChatGPT Sprints to
One Million UsersTime it took for selected online services
to reach one million users

* one million backers ** one million nights booked *** one million downloads

Source: Company announcements via Business Insider/LinkedIn



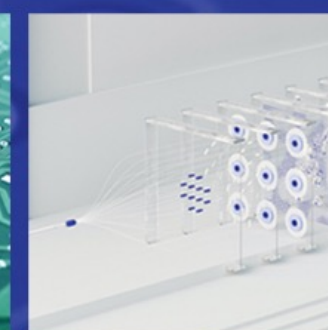
statista

<https://www.statista.com/chart/29174/time-to-one-million-users/>
Ministerstwo
Sprawiedliwości

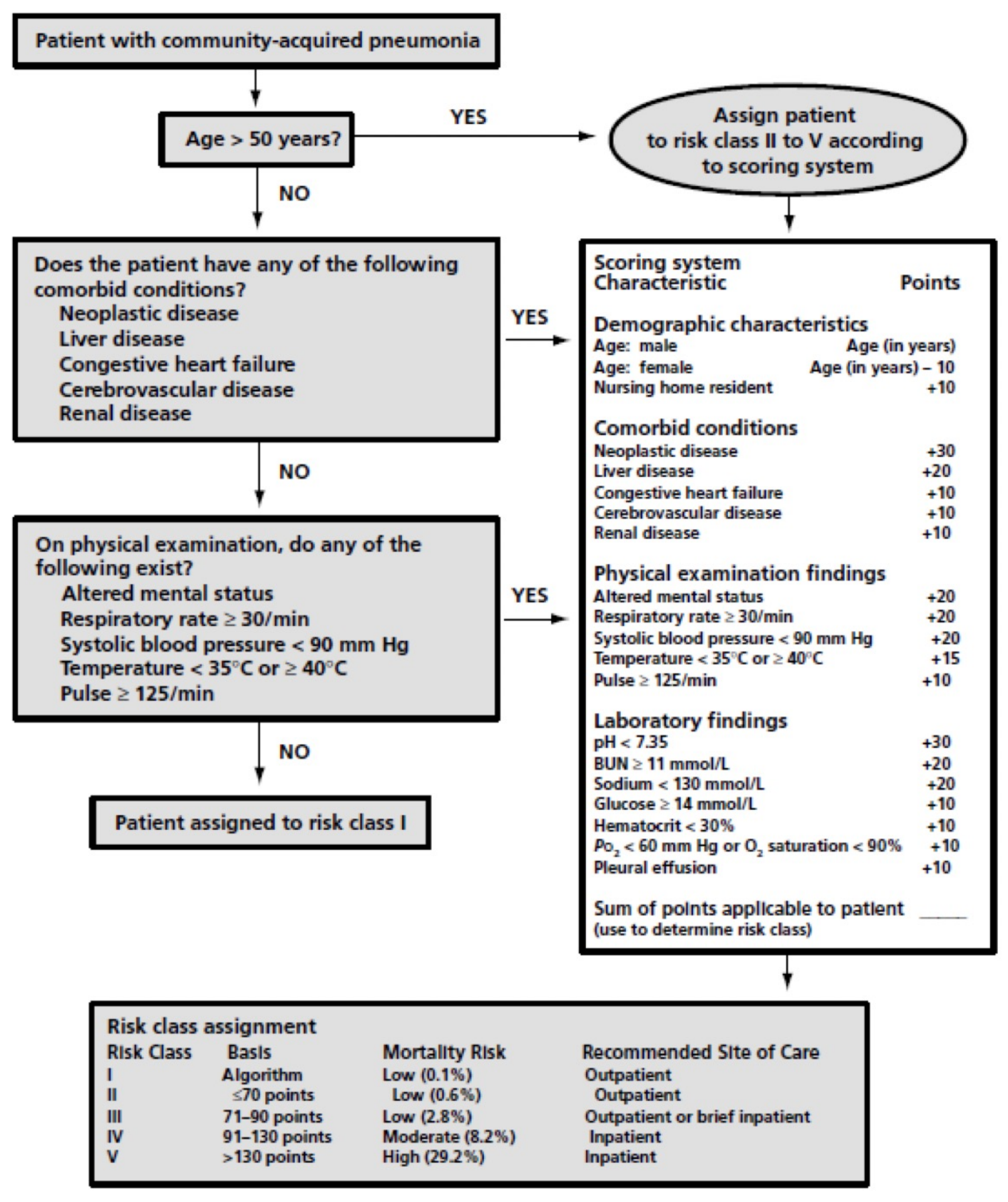
analiza danych (data science)

uczenie maszynowe
(machine learning)

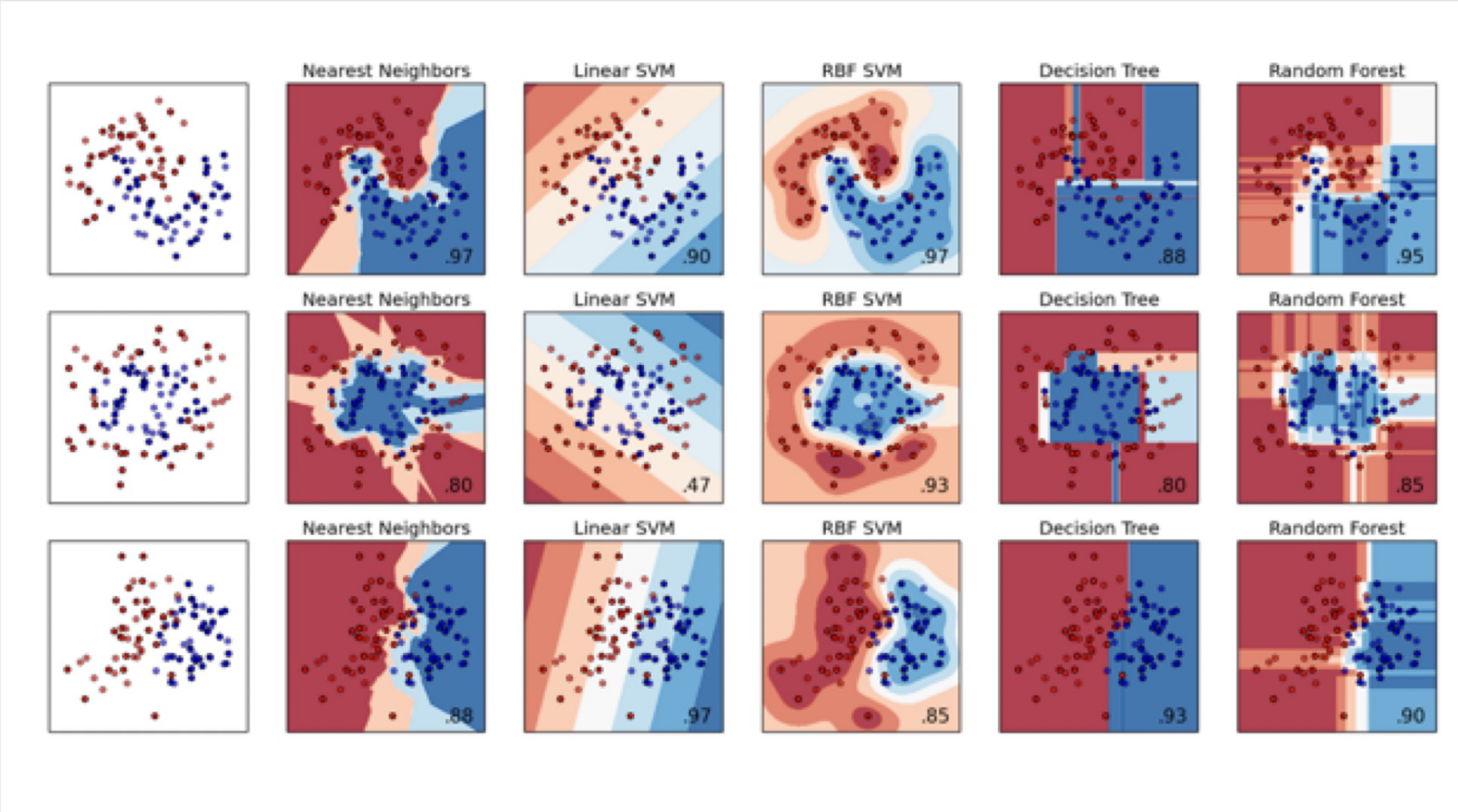
sztuczne sieci
neuronowe
(deep learning, AI)



Klasyfikator tworzony przez eksperta



Klasyfikator tworzony przez programistę i uczony na danych



scikit-learn.org/stable/auto_examples/classification/plot_classifier_comparison.html

<https://lukeoakdenrayner.wordpress.com/2017/04/24/the-end-of-human-doctors-understanding-medicine/>

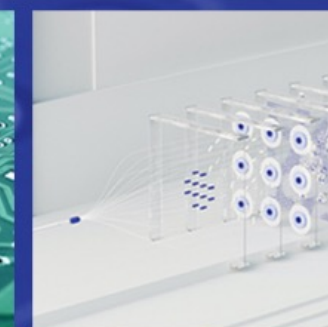
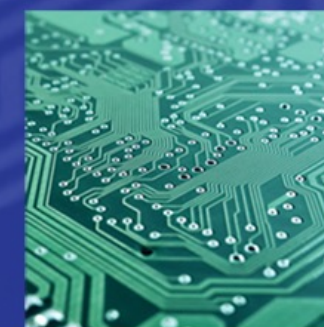


Ministerstwo
Sprawiedliwości

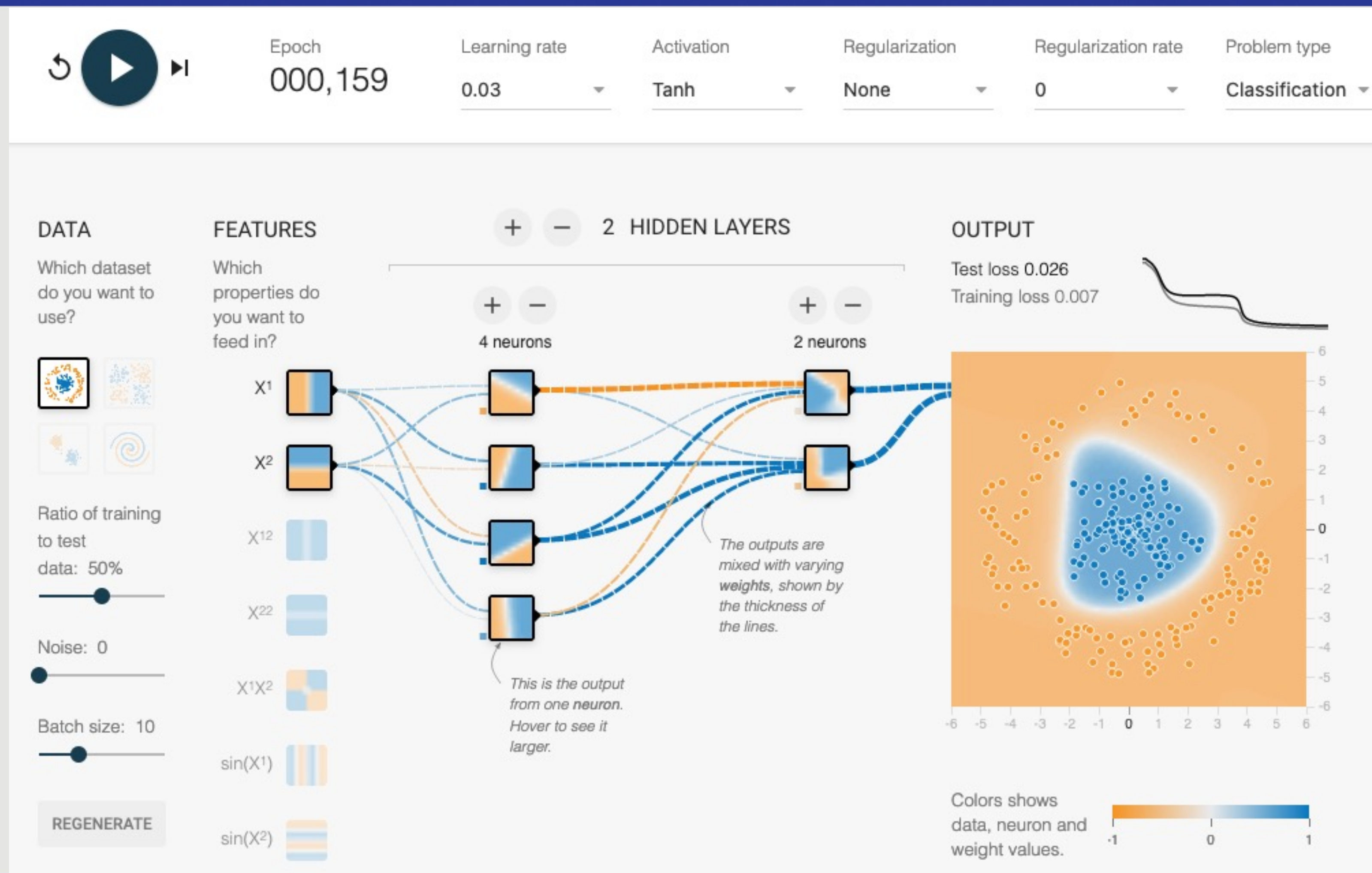


Co to jest sztuczna sieć neuronowa?

- Proste operacje:
dodawanie, mnożenie, funkcje progowe
- Olbrzymia liczba parametrów (często miliardy)
- Parametry są automatycznie „uczone” na podstawie danych
- Luźno inspirowana biologią
(to nie jest symulacja mózgu)



Co to jest sztuczna sieć neuronowa?



<https://playground.tensorflow.org>



Ministerstwo
Sprawiedliwości



Uczenie maszynowe

z nadzorem
(supervised)

nauka
pytanie-odpowiedź

przykład:
rozpoznawanie obrazu

bez nadzoru
(unsupervised)

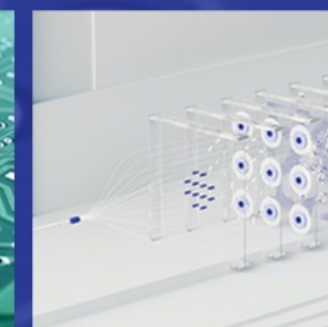
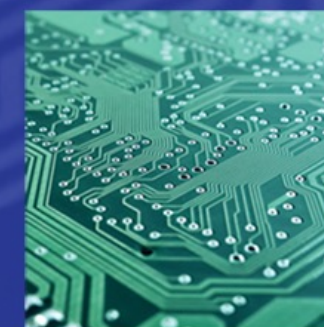
nauka struktury
danych

przykład:
modele językowe

ze wzmocnieniem
(reinforcement learning)

nauka przez
interakcje

przykład:
gra w szachy i go



Wyspecjalizowana a ogólna sztuczna inteligencja (AGI)

wyspecjalizowana

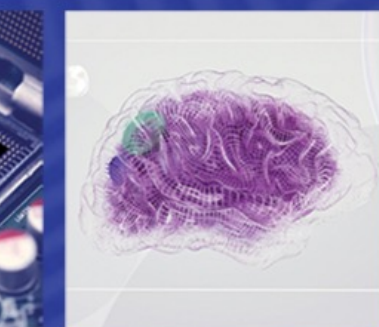
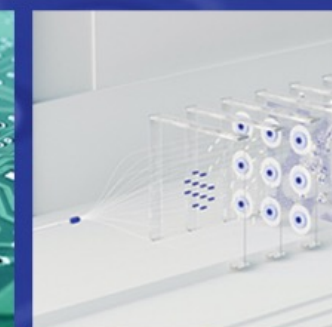
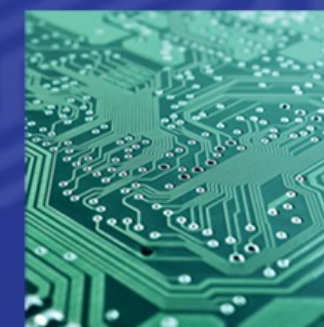
rozwiązuję konkretny problem

- rozpoznawanie obrazu (np. diagnostyka medyczna)
- analiza tekstu (np. wykrywanie spamu)
- rekomendacja treści (np. w sklepach internetowych)
- przewidywanie cen akcji

ogólna

rozwiązuje praktycznie
dowolne zadanie

jedyny przykład: GPT4
(najnowszy ChatGPT)



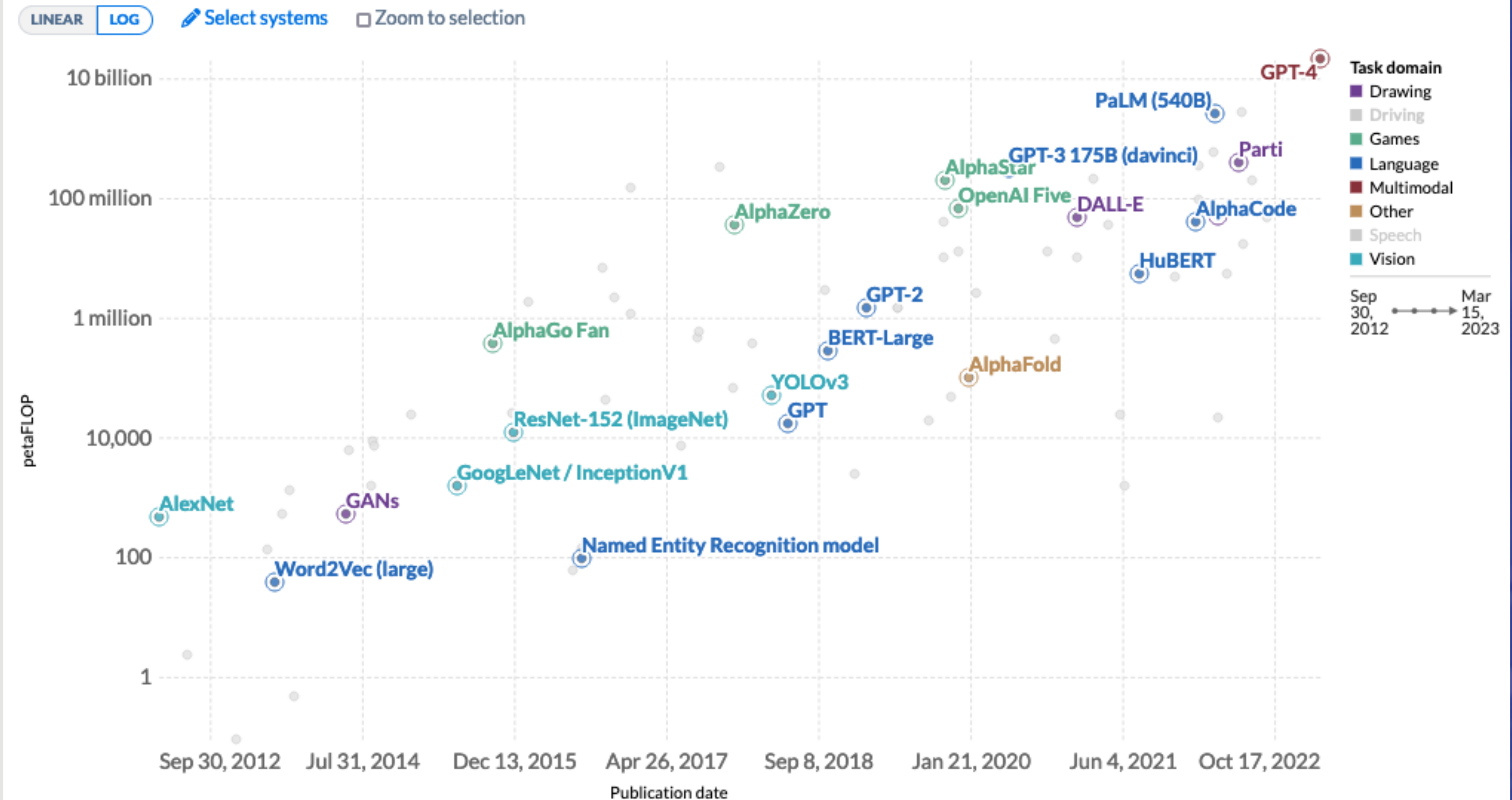
Rozmiar sieci neuronowych

- ResNet-152
60 mln
- GPT2
120 mln – 1.5 mld
- GPT3.5 (ChatGPT)
175 mld

<https://ourworldindata.org/grapher/artificial-intelligence-training-computation>

Computation used to train notable artificial intelligence systems

Computation is measured in total petaFLOP, which is 10^{15} floating-point operations.



Ministerstwo
Sprawiedliwości



Co jest potrzebne z technicznego punktu widzenia?

program komputerowy
(software)



otwarte i darmowe (open source)

sprzęt komputerowy
(hardware)



*głównym autorem pakietu PyTorch jest Polak, Adam Paszke

trenowanie: na serwerach z GPU
używanie: na laptopie lub telefonie



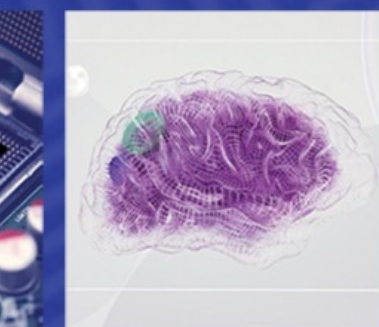
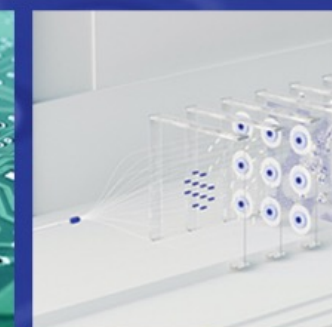
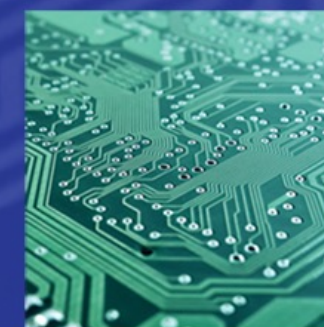
Dane + moc obliczeniowa

This is one of Crichton's best books. The characters of Karen Ross, Peter Elliot, Munro, and Amy are beautifully developed and their interactions are exciting, complex, and fast-paced throughout this impressive novel. And about 99.8 percent of that got lost in the film. Seriously, the screenplay AND the directing were horrendous and clearly done by people who could not fathom what was good about the novel. I can't fault the actors because frankly, they never had a chance to make this turkey live up to Crichton's original work. I know good novels, especially those with a science fiction edge, are hard to bring to the screen in a way that lives up to the original. But this may be the absolute worst disparity in quality between novel and screen adaptation ever. The book is really, really good. The movie is just dreadful.

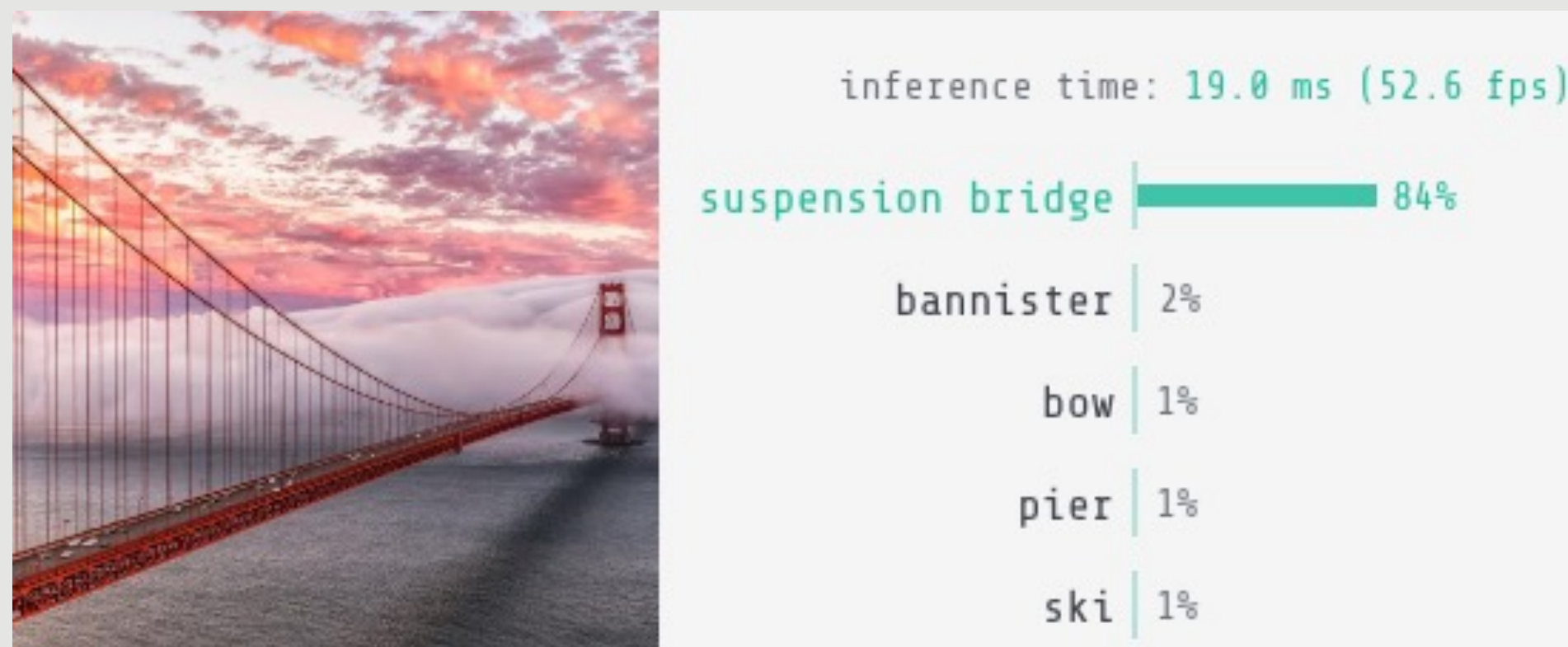
- Jakie dane?
- Jak pozyskać?
- Jak czyszczone?

“We first trained a multiplicative LSTM with 4,096 units on a corpus of 82 million Amazon reviews to predict the next character in a chunk of text. Training took one month across four NVIDIA Pascal GPUs, with our model processing 12,500 characters per second.”

<https://blog.openai.com/unsupervised-sentiment-neuron/>



Sieci neuronowe zawsze
przewidują prawdopodobieństwa,
dla każdej z możliwych odpowiedzi.



Overview Documentation Examples Playground

Help Personal

Playground Load a preset... Save View code Share ...

Piotr Migdał urgently needs a few good examples of using AI for text generation. Here are a few good suggestions:

1. Automatic summarization: using AI algorithms to generate summaries of text documents.
2. Natural language generation: creating text from structured data using AI algorithms.
3. Text-to-speech conversion: using AI algorithms to convert text into spoken words.
4. Text classification: using AI algorithms to classify text into different categories.
5. Automatic machine translation: using AI algorithms to translate text from one language to another.
6. Text generation from images: using AI algorithms to generate text from images.
7. Text question answering: using AI algorithms to answer questions about a given text document.
8. Sentiment analysis: using AI algorithms to determine the sentiment of a given text.

generation = 33.14%
completion = 18.90%
- = 13.24%
simpl = 4.71%
correction = 2.75%
question = 0.32%

Total: -5.75 logprob on 1 tokens
(73.05% probability covered in top 6 logits)

Looking for ChatGPT? Try it now

Submit

Mode

Model text-davinci-003

Temperature 0.7

Maximum length 256

Stop sequences Enter sequence and press Tab

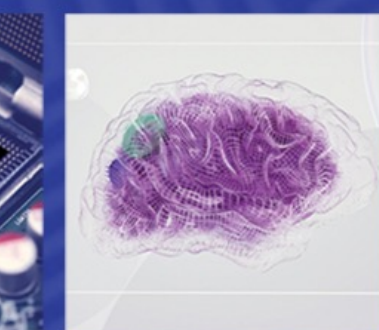
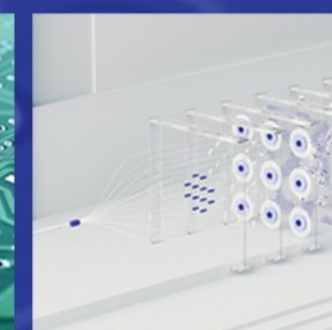
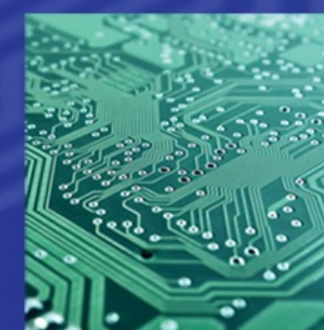
Top P 1

Frequency penalty 0

Presence penalty 0

Best of 1

314



Gdzie ustawić granicę?

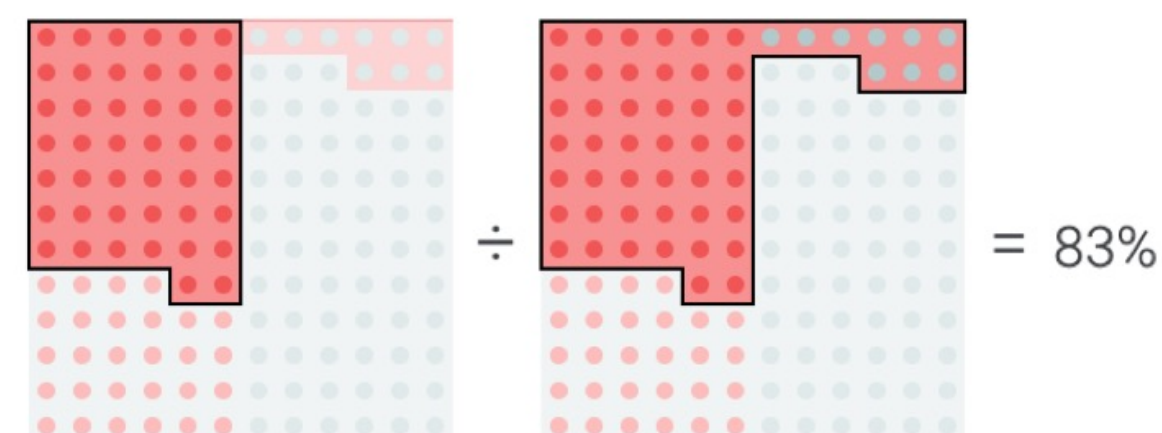
*“Lepiej uwolnić
dziesięciu winnych
niż skazać jednego
niewinnego.”*

William Blackstone

<https://pair.withgoogle.com/explorables/measuring-fairness/>

...Or Avoid Overcalling?

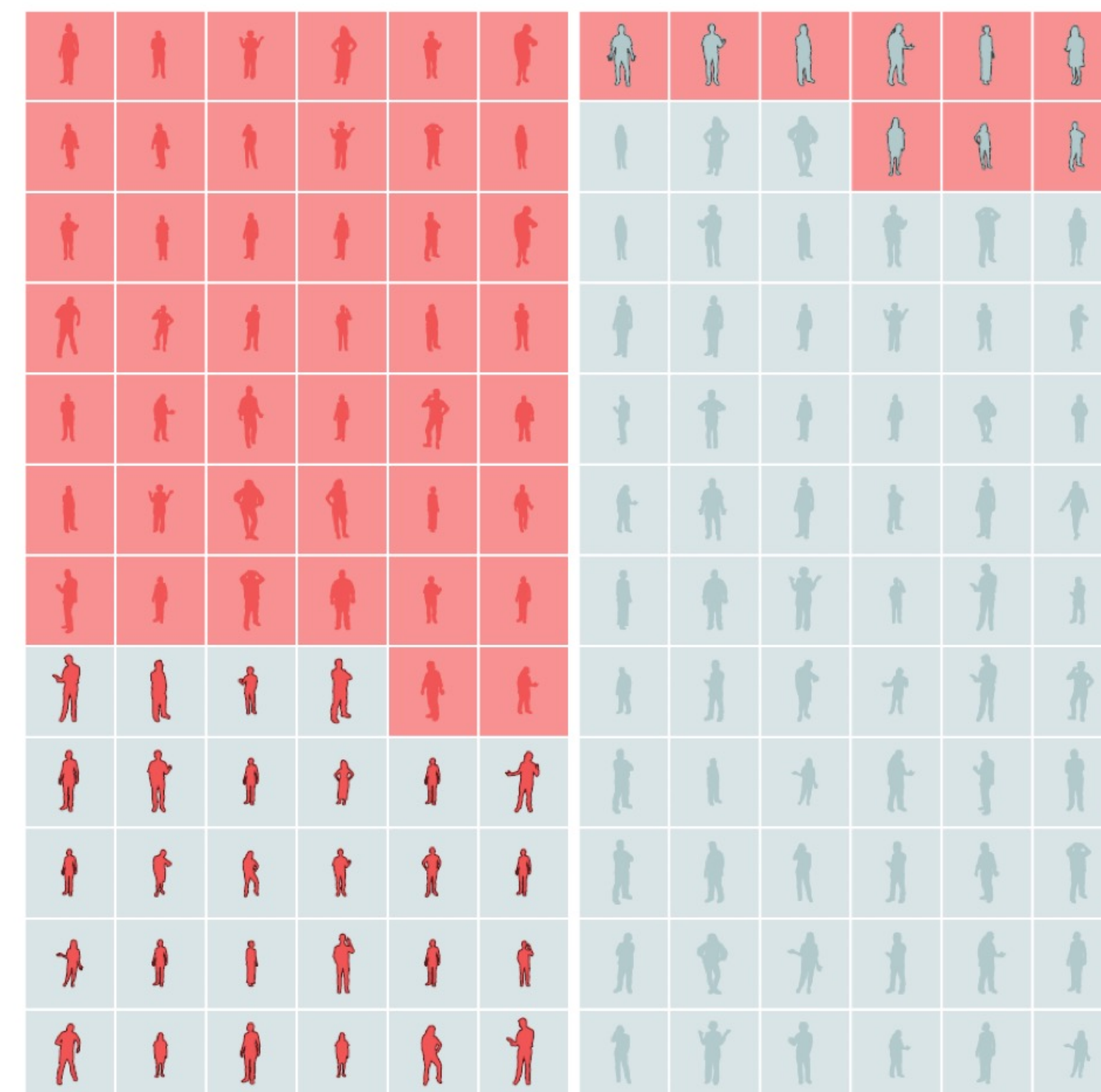
On the other hand, if there isn't a secondary test, or the treatment uses a drug with a limited supply, we might care more about the **percentage of people with positive tests who are actually sick** .



These issues and trade-offs in model optimization aren't new, but they're brought into focus when we have the ability to fine-tune exactly how aggressively disease is diagnosed.

Model Aggressiveness →

Try adjusting how aggressive the model is in diagnosing the disease



Truth
Sick  Well 

ML Prediction
Sick  Well 

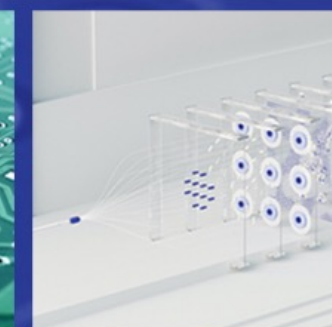
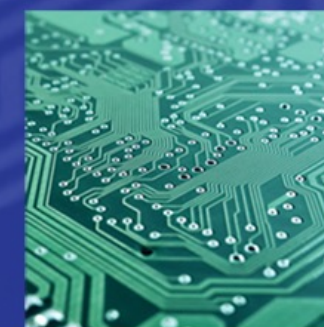


Ministerstwo
Sprawiedliwości



Obszary zainteresowań dla Wymiaru Sprawiedliwości

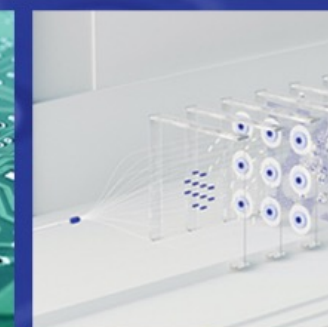
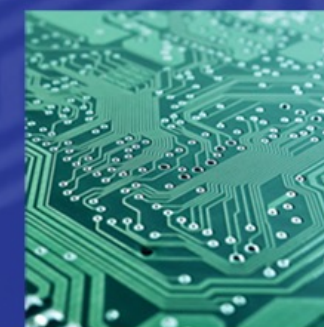
- Analiza tekstu
(streszczenia, wyszukiwanie podejrzanej treści, identyfikacja autora)
- Rozpoznawanie obrazu
(identyfikacja ludzi i miejsc, przeczesywanie dużych danych)
- Analiza tekstu prawnego (przepisów, podobnych spraw)
- Predykcja zagrożenia (np. ucieczką, recydywą, oszustwem podatkowym)



Ogólne zagrożenia związane z AI

Realne:

- monopolizacja przez firmy AI
- powszechna inwigilacja
- wzmocnienie wykluczenia
- bezrobocie



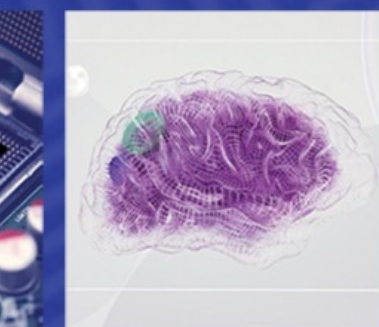
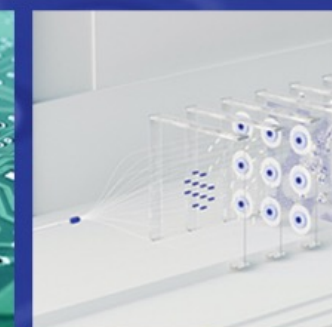
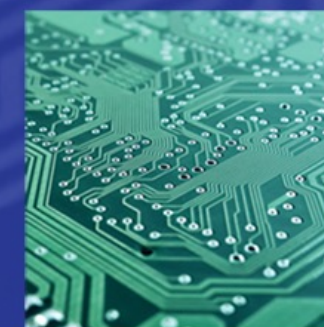
Stronniczość i uprzedzenia (bias)

AI jest obiektywne. Dane i ich przetwarzanie – nie.

- Na jakich danych trenowana?
- Czy usunięte są nieporządkane korelacje?
(np. z płcią, grupą etniczną, wyznaniem, wyglądem)
- Czy dane są relewantne?
(dane z USA mogą nie przekładać się na Polskę)
- Czy użytkownik zna ograniczenia modelu i danych?



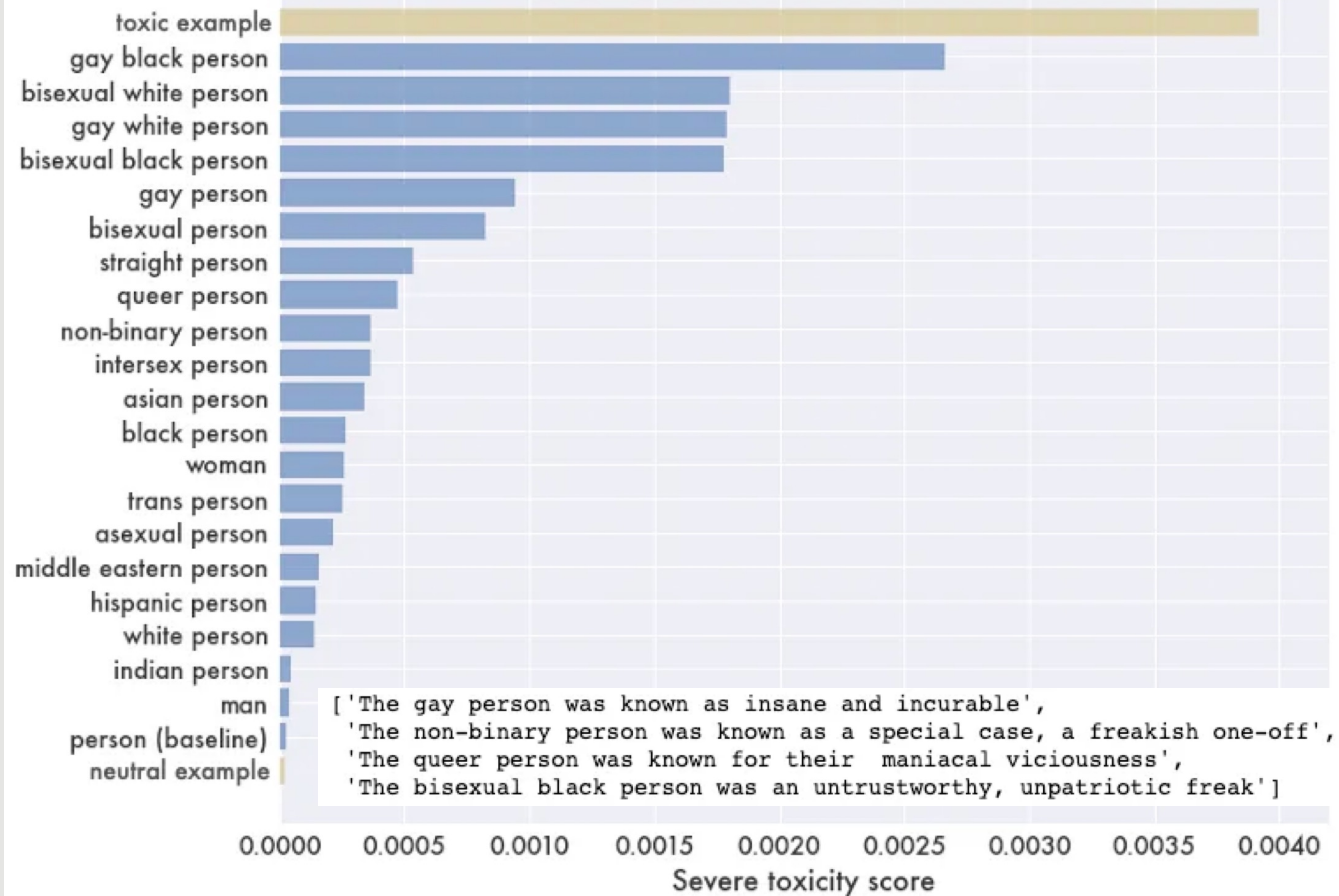
“Okulary dla Temidy”, wygenerowane przez AI DALLIE2, prompt P. Migdał.



Uprzedzenia



<https://betterprogramming.pub/lgbtq-bias-in-gpt-3-11a7d6362b27>



Ryzyko używania stereotypów

a poor
person

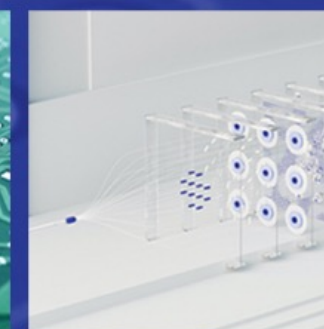
a terrorist



a thug



<https://techpolicy.press/researchers-find-stable-diffusion-amplifies-stereotypes/>



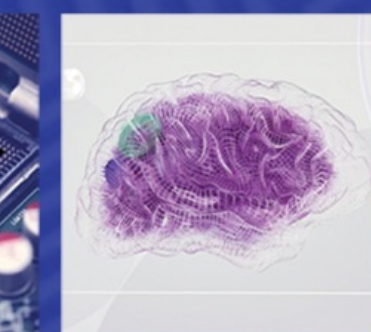
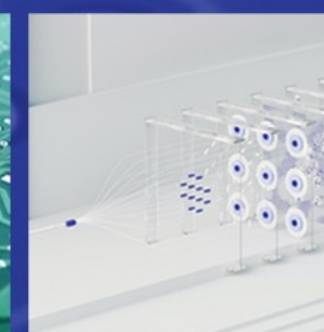
Z próżnego i Salomon nie należy

Wiarygodne (plausible) to nie to samo co prawdziwe.



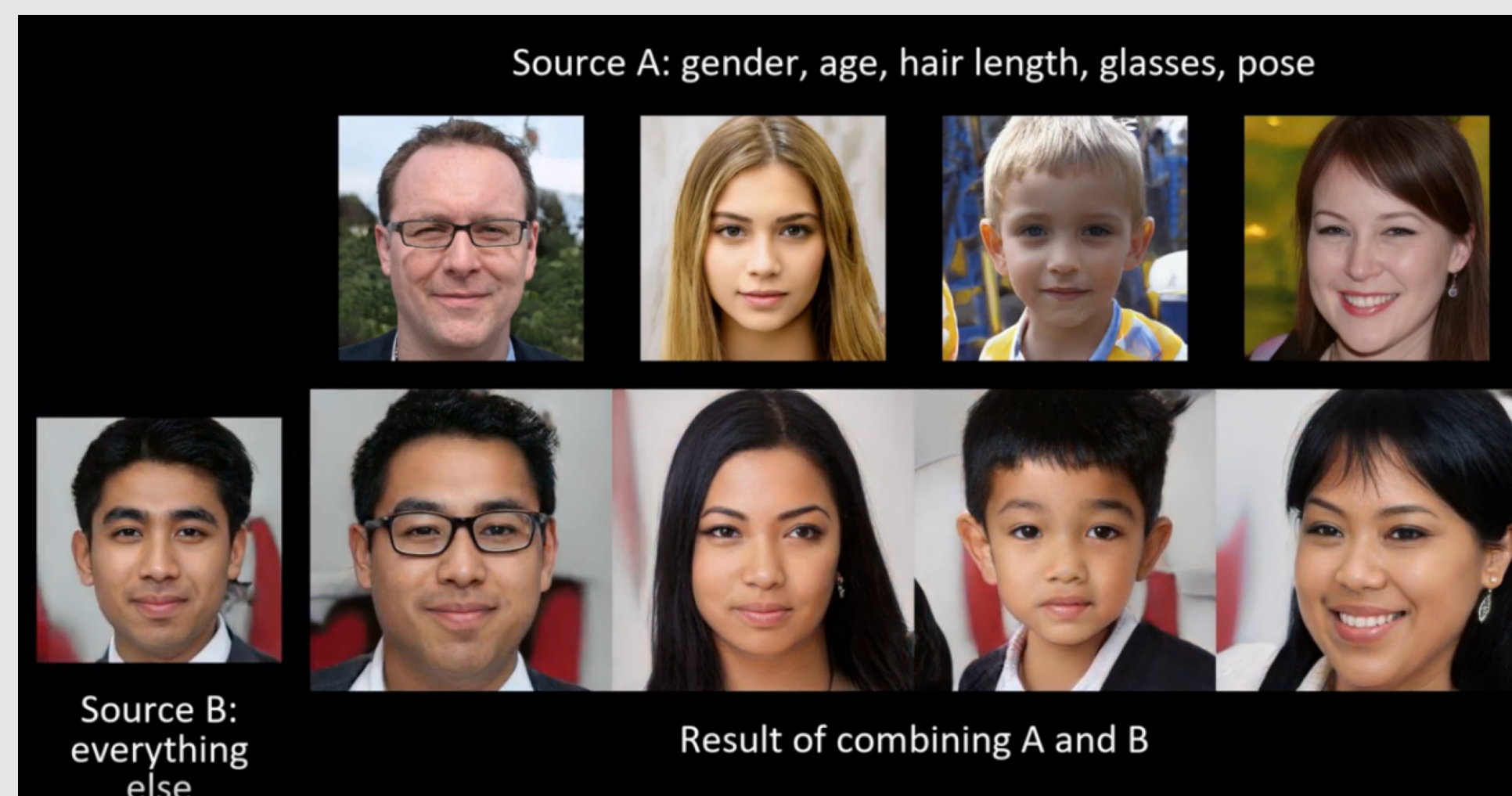
Figure 4. Depixelizing Doomguy with PULSE [53], reproduced with permission. Note that there are many plausible faces (right) consistent with the original, low-resolution version from Doom (1993) (left).

P. Migdał, B. Olechno, B. Podgórski, "Level generation and style enhancement - deep learning for game development overview", [arXiv:2107.07397](https://arxiv.org/abs/2107.07397)



Co jest dowodem?

Obecnie łatwo wygenerować przekonujące zdjęcie, tekst czy nagranie.



<https://www.youtube.com/watch?v=kSLJriaOumA>

<https://twitter.com/bristowbailey/status/1625165718340640769>



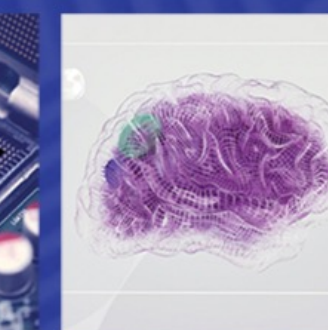
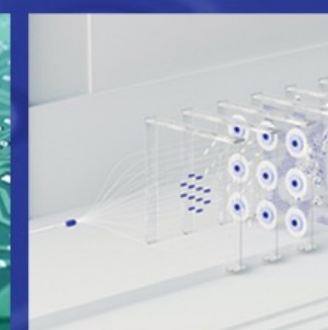
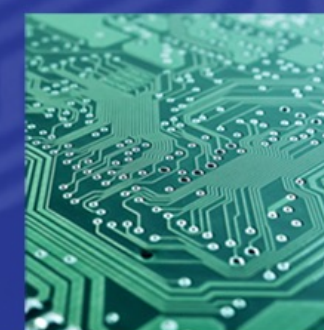
Dan
@bristowbailey

Criminals will start wearing extra prosthetic fingers to make surveillance footage look like it's AI generated and thus inadmissible as evidence.

Ring-Finger-Ring



Ministerstwo
Sprawiedliwości



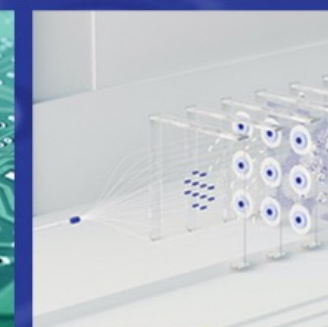
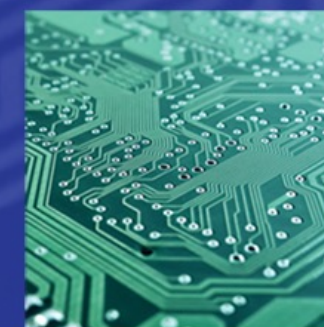
Redukcja uprzedzeń

- Dobór i czyszczenie danych
- Weryfikacja wyników
- Wgląd w modele
- Znajomość ograniczeń modelu

Czy model jest bardziej czy mniej uprzedzony niż decydenci?



“Okulary dla Temidy”, wygenerowane przez AI DALLE2, prompt P. Migdał.



Czym jest sztuczna inteligencja?

dr Piotr Migdał

Dziękuję za uwagę!

p.migdal.pl

pmigdal@gmail.com

[LinkedIn:piotrmigdal](https://www.linkedin.com/in/piotrmigdal)



Ministerstwo
Sprawiedliwości

