

# Reinforcement Learning: uczenie maszynowe metodą kija i marchewki



BYDGOSZCZ I TORUŃ

Błażej Osiński

Bydgoszcz, 22.05.2019

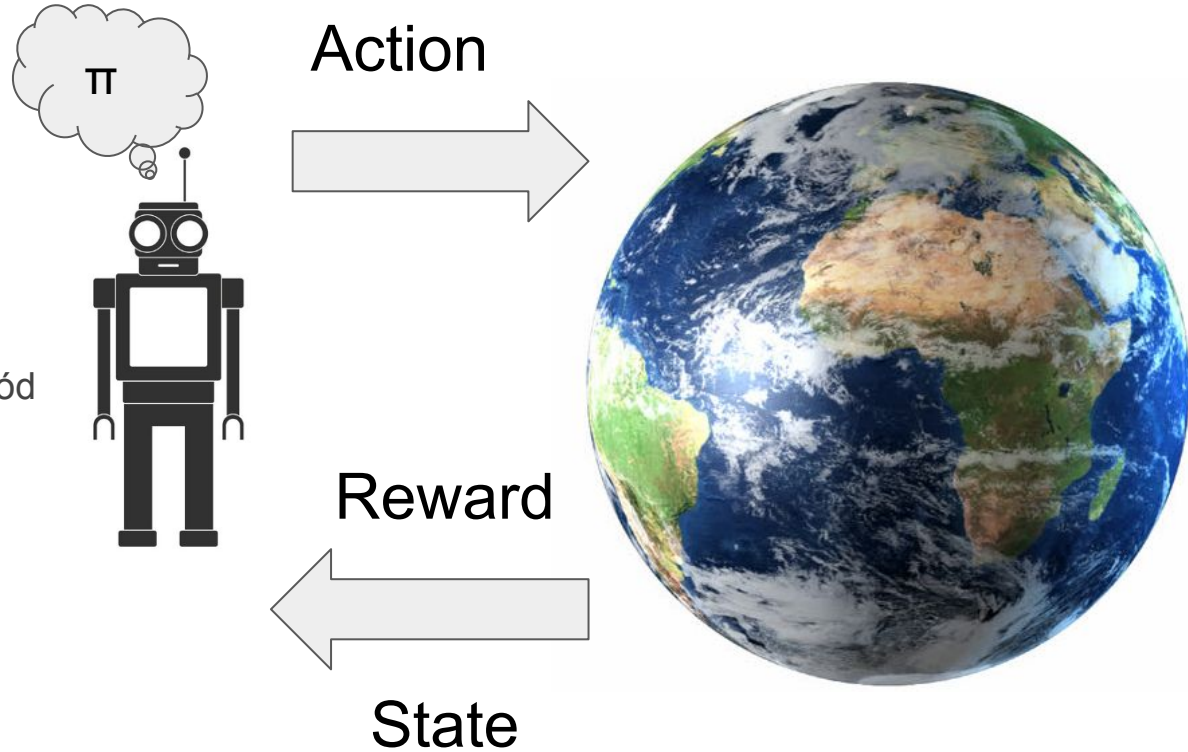


CRAIG SWANSON © WWW.PERSPICUITY.COM

Czym jest *reinforcement learning*?

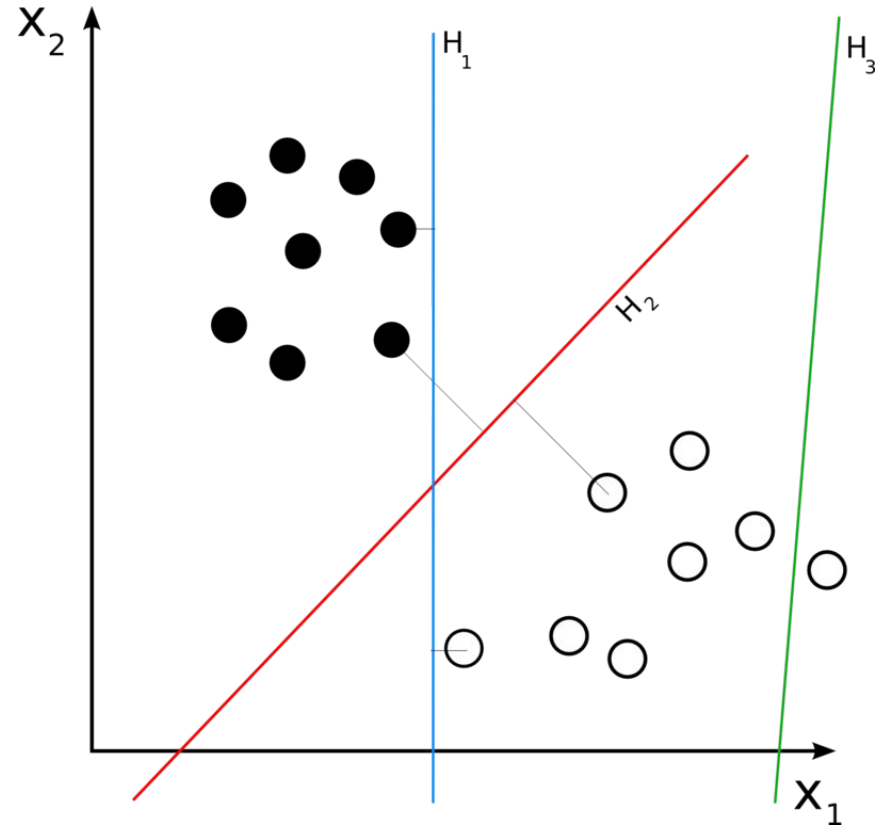
## Definicje

- Agent działa w środowisku
- Otrzymuje nagrody (lub kary)
- Zadanie:  
Nauczyć się strategii (polityki),  
która maksymalizuje sumę nagród



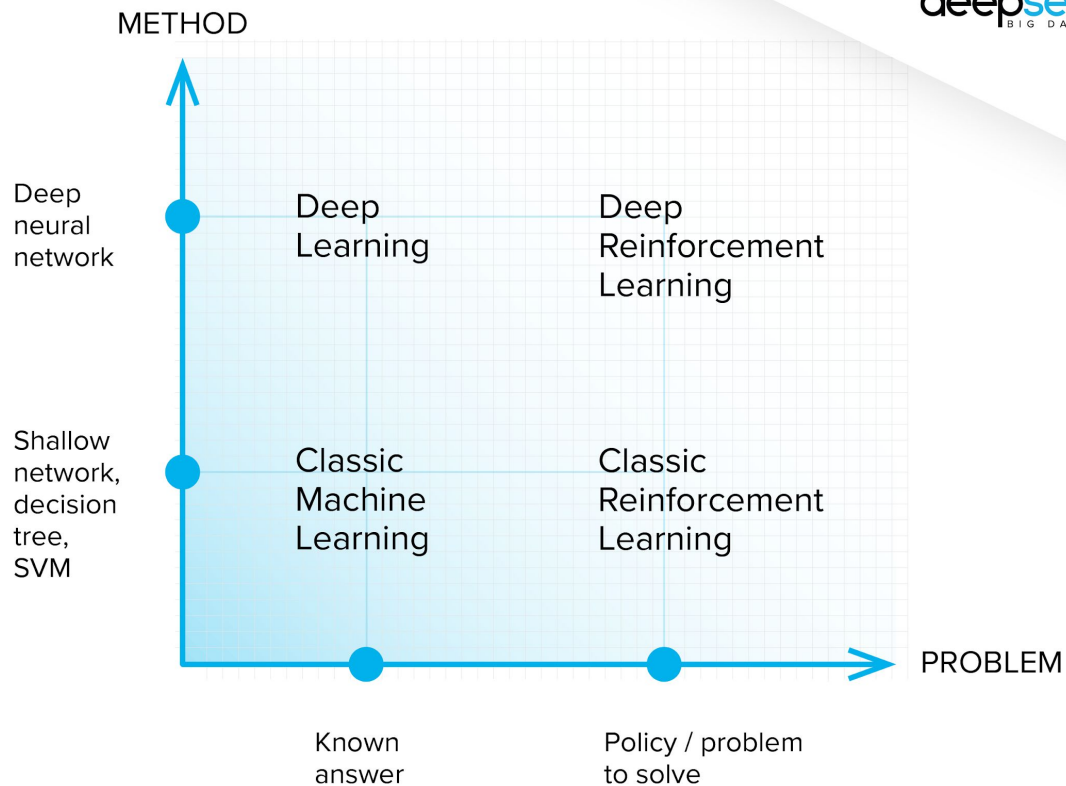
## RL vs Supervised Learning

- Uczenie z nadzorem odpowiada na każde zadanie niezależnie
- RL podejmuje szereg decyzji, które mają wpływ na przyszłe zadania

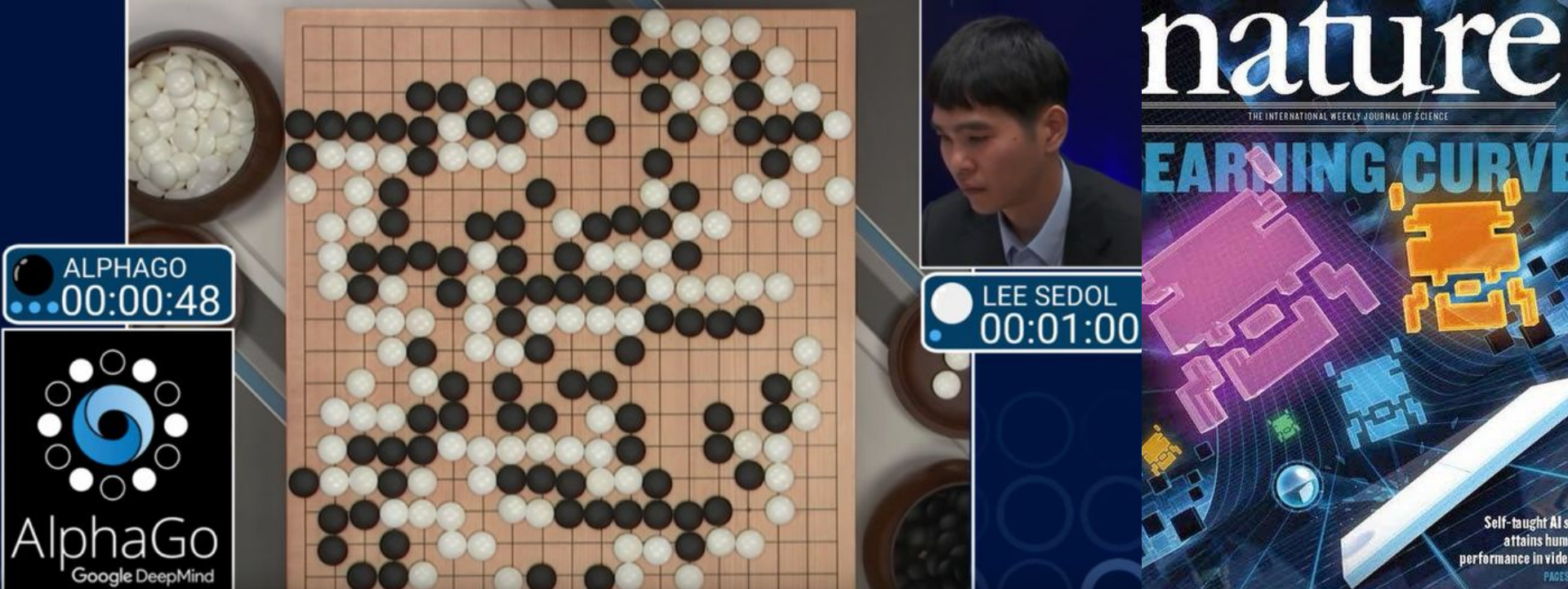


# RL vs Deep Learning

- Zwykle Deep Learning odnosi się do uczenia z nadzorem...
- ... ale RL z powodzeniem wykorzystuje sieci neuronowe (także głębokie)







Sukcesy RL

## RL w jednym zdaniu: © Ilya Sutskever

- Podjmij (trochę) losowe akcje
- Jeżeli wynik jest lepszy, niż się spodziewałeś, podejmuj takie akcje częściej



Dota 2 vs Starcraft 2





Dota 2

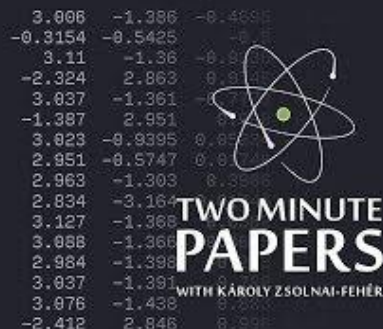
vs

Starcraft 2

Human View



AI View



OpenAI Five

Competitive: 7,215-42 (99.4% winrate, 15,019 total players)



# RL można użyć do

- Robotyki
- Samochodów autonomicznych
- Obrotu akcjami
- Projektowania sieci neuronowych
- Dowolnego zadania optymalizacyjnego?!



## Wyzwania RL

- Projektowanie funkcji nagrody
- Uwaga filozoficzna:  
Ludzie nie otrzymują nagrody bezpośrednio ze środowiska, sami oceniają sytuację



## Wyzwania RL

- Ilość wymaganych danych
- 8 mln klatek  
40h gry  
by zostać mistrzem Ponga

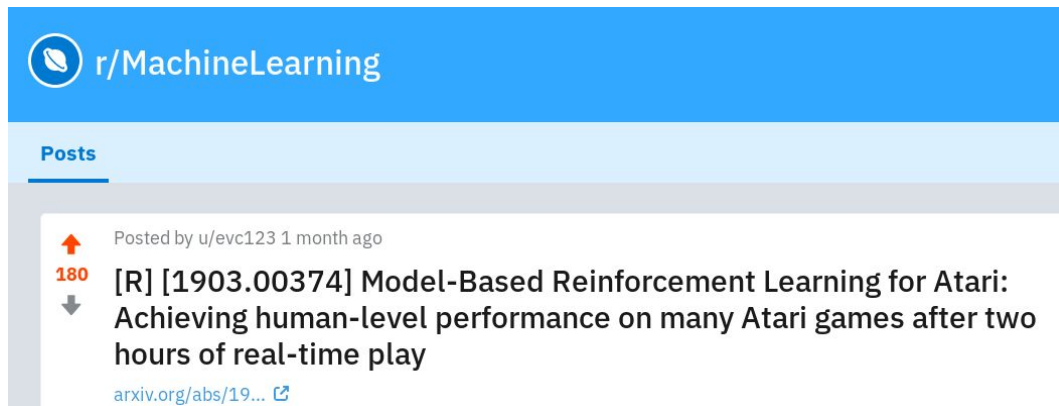


# Research @ deepsense.ai

# Model Based Reinforcement Learning for Atari

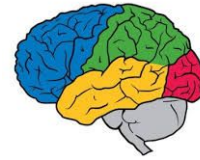
Łukasz Kaiser<sup>\*1</sup> Mohammad Babaeizadeh<sup>\*23</sup> Piotr Miłoś<sup>\*45</sup> Błażej Osiniński<sup>\*453</sup> Roy H Campbell<sup>2</sup>  
Konrad Czechowski<sup>4</sup> Dumitru Erhan<sup>1</sup> Chelsea Finn<sup>1</sup> Piotr Kozakowski<sup>4</sup> Sergey Levine<sup>1</sup> Ryan Sepassi<sup>1</sup>  
George Tucker<sup>1</sup> Henryk Michalewski<sup>45</sup>

- [Website](#)
- [Arxiv](#)
- [Source code - tensor2tensor](#)
- [Google AI Blogpost](#)
- Presented at BayLearn 2018

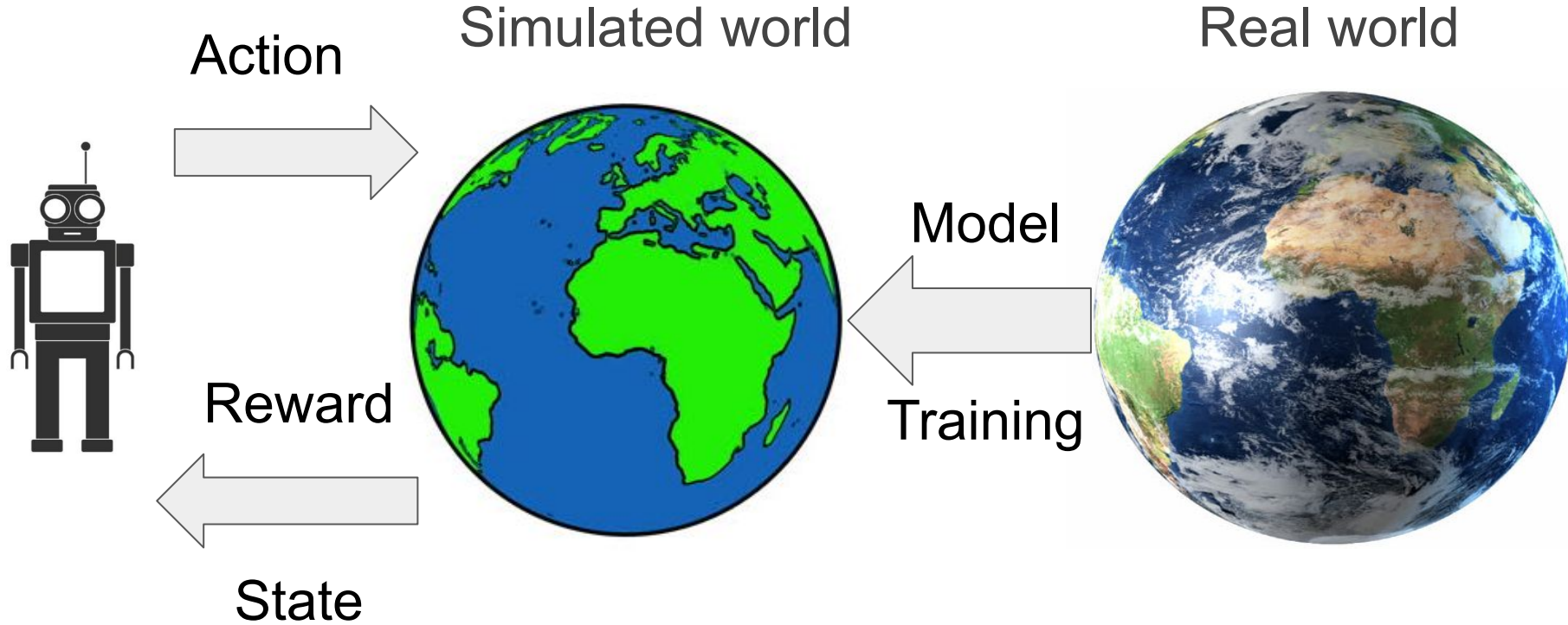




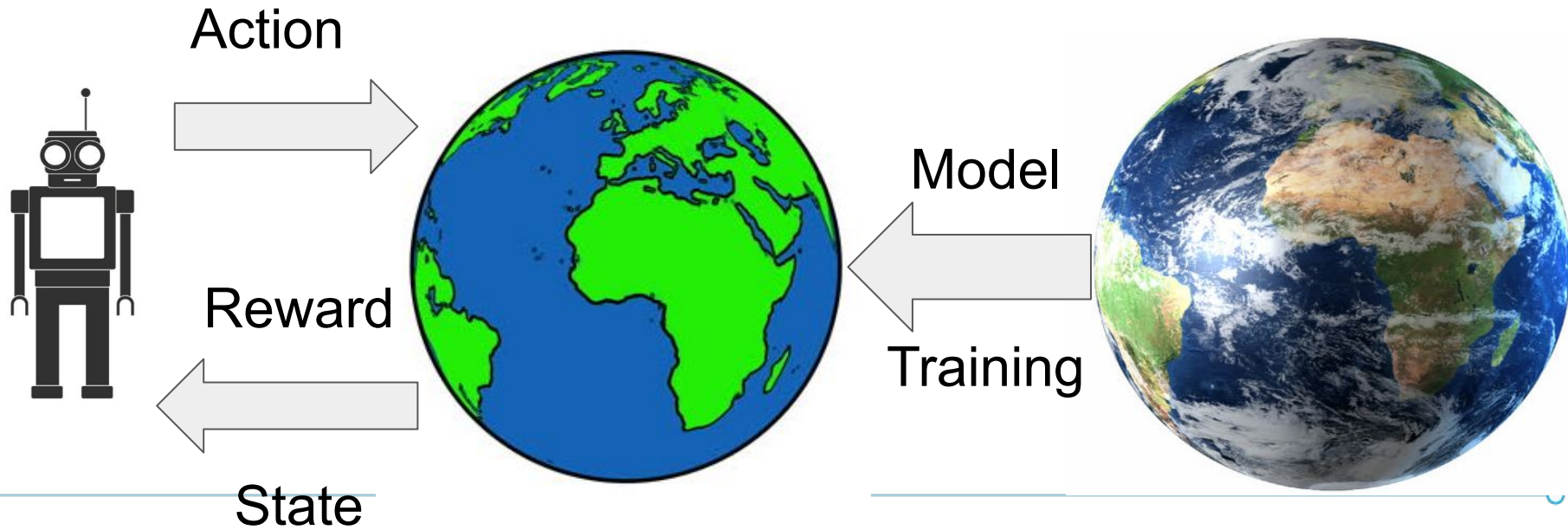
# Model-based RL



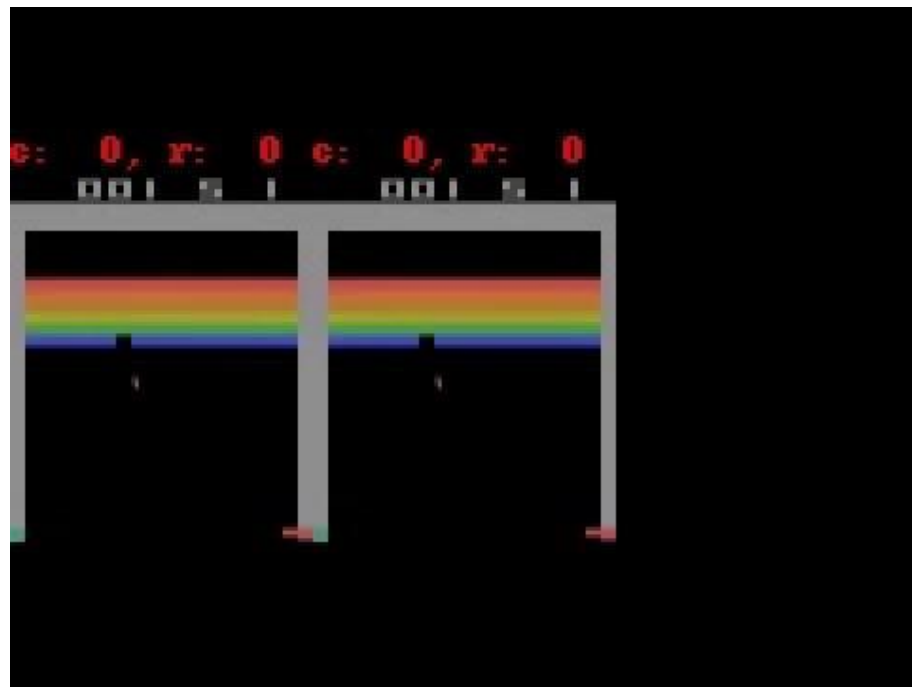
deepsense.ai  
BIG DATA SCIENCE







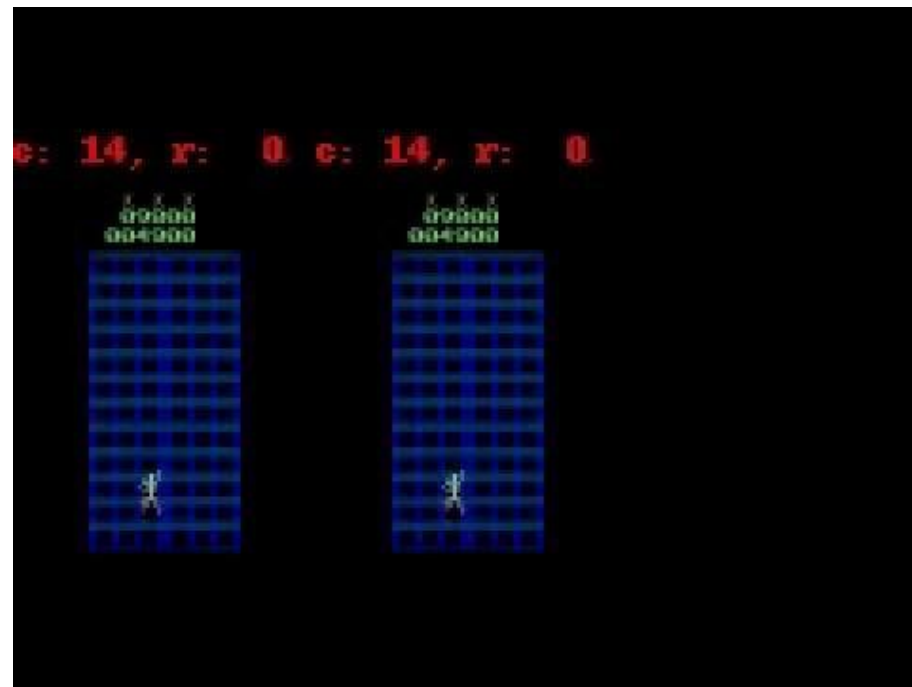
## World Model - examples



predicted

ground truth

difference



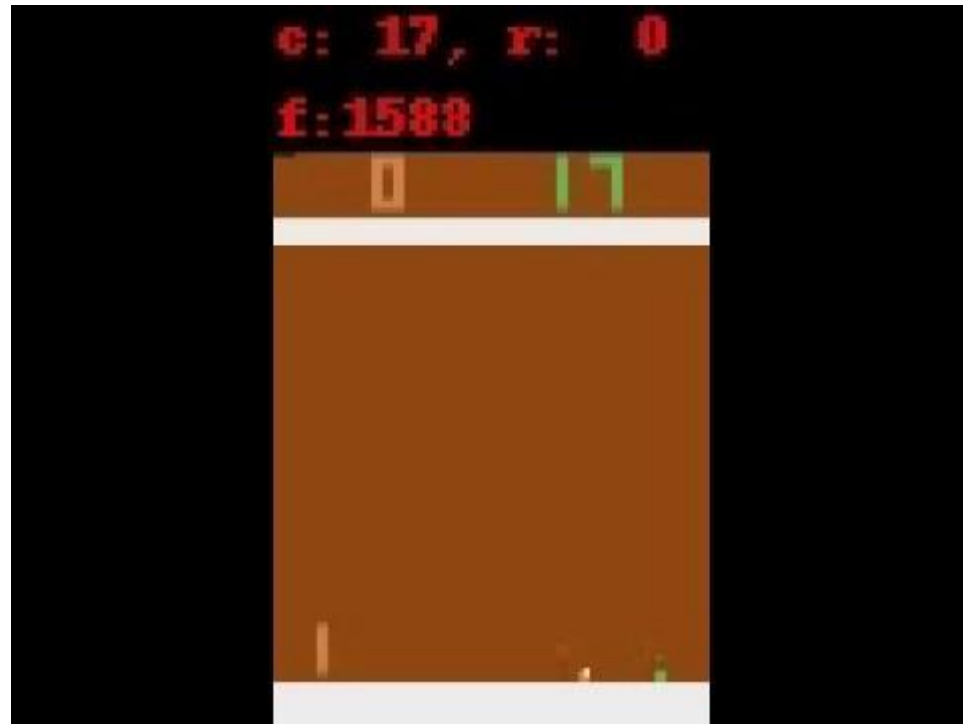
predicted

ground truth

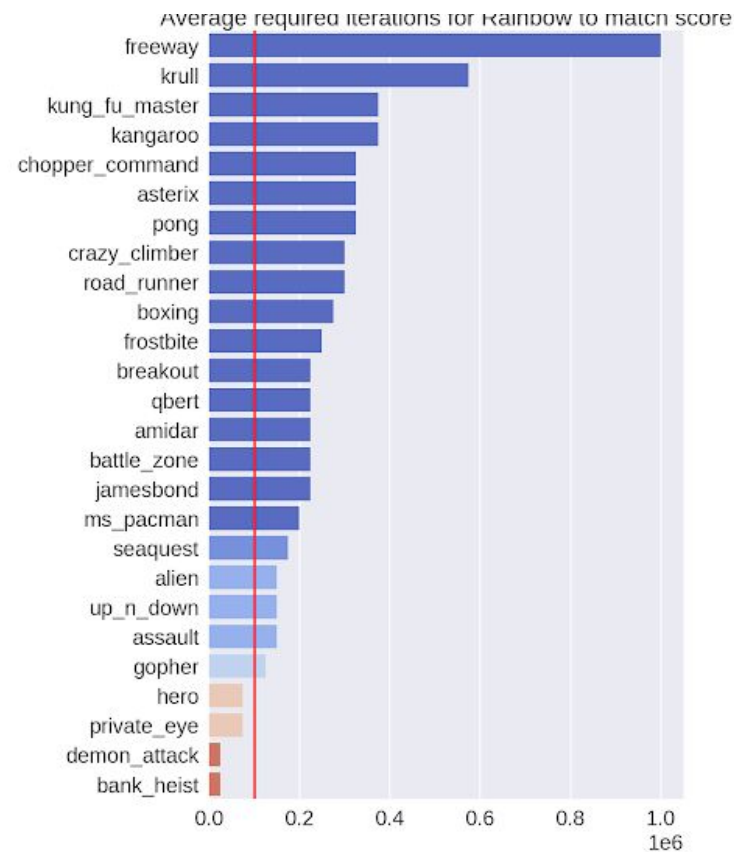
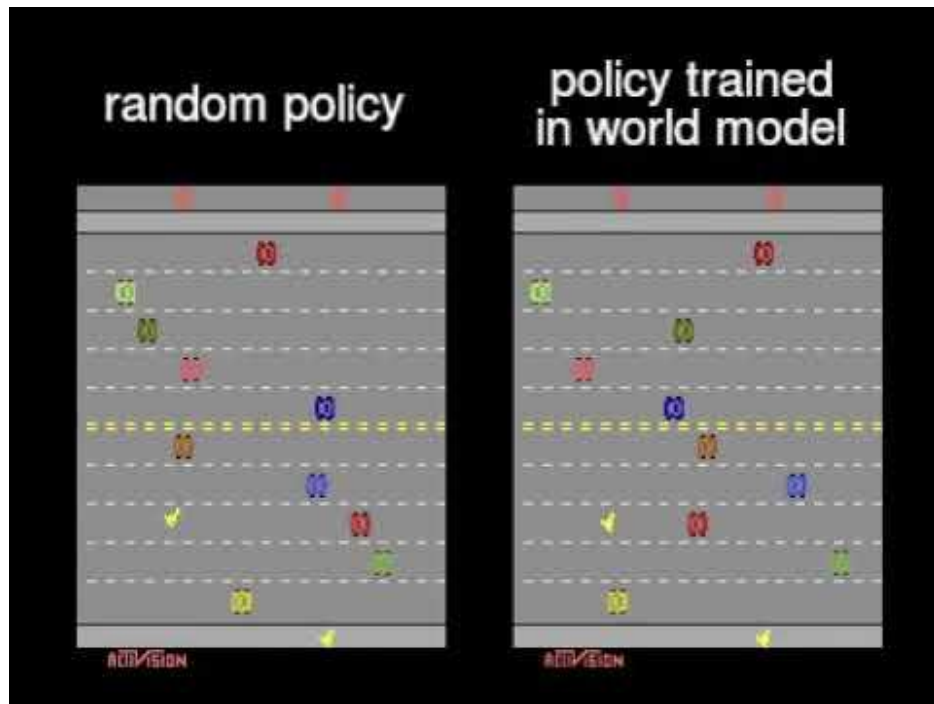
difference

## Modelowanie gier Atari jest skomplikowane!

- Liczenie do 21
- Piłeczka jest bardzo mała
- Gry wykazują pewną stochastyczność, np. po stracie punktu piłka znika, nie wiadomo kiedy się pojawi
- Model musi nauczyć się symulować AI wbudowane w grę Atari



To działa!



## Błędy modelu



Kangaroo -  
przykład dziwnej  
strategii

# — Samochody autonomiczne

**2018  
SELF-DRIVING  
SAFETY REPORT**

GENERAL MOTORS





# Samochody autonomiczne

- Projekt we współpracy z jednym z głównych producentów samochodów
- Badamy możliwość wykorzystania RL do prowadzenia samochodów autonomicznych
- Problemy?

# Samochody autonomiczne

- Problemy?
- Dane ze świata rzeczywistego mogą być bardzo drogie!



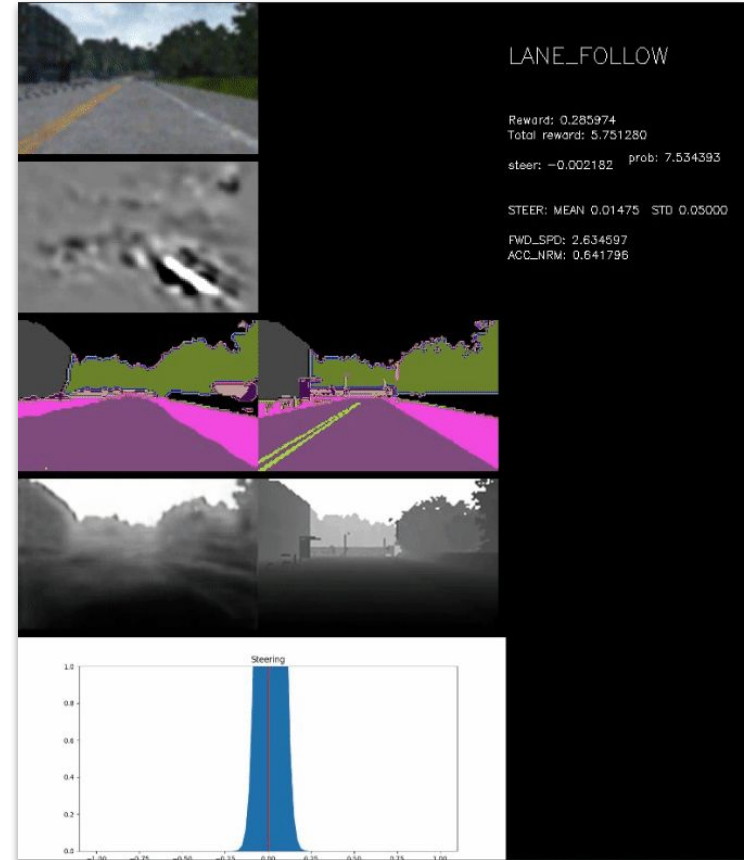
# Sim-to-real

- Wykorzystujemy symulator Carla oparty o Unreal Engine do zbierania doświadczenia z jazdy
- Dzięki silnemu zrównolegleniu i przyśpieszeniu symulatora zbieramy nawet rok doświadczeń dziennie
- Największy w Polsce superkomputer: Prometheus, Cyfronet, AGH



## sim-to-real

- Symulator nie odwzorowuje rzeczywistości w sposób doskonały, musimy stosować metody takie jak “domain randomization”, żeby polityki przekładały się na świat rzeczywisty
- Wielokrotnie testowaliśmy już nasze modele na prawdziwym samochodzie, dostarczonym przez naszego partnera





Przykłady  
“domain  
randomization”:  
zmiana pogody

Image source:  
<https://arxiv.org/pdf/1711.03938.pdf>  
f

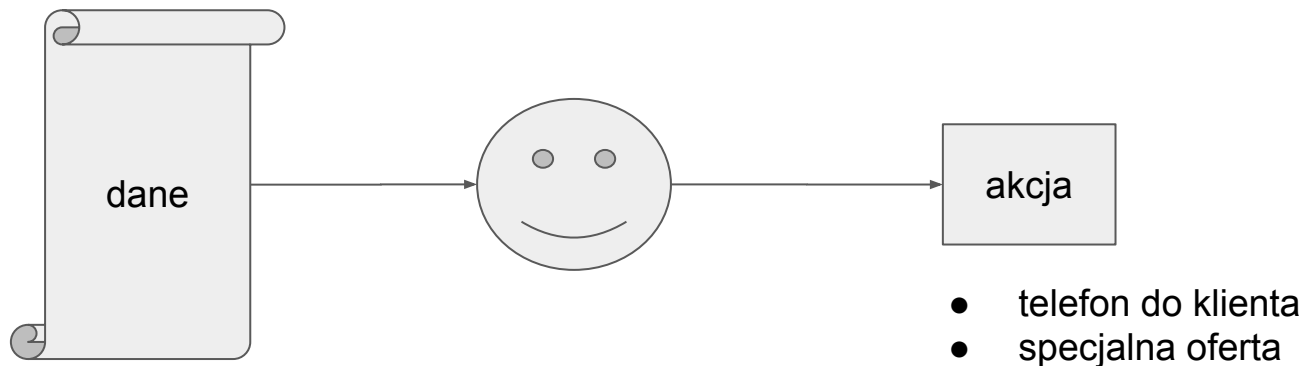


# Jak zmieniają się procesy biznesowe?

# Ewolucja procesu biznesowego

Przykład: retencja klienta

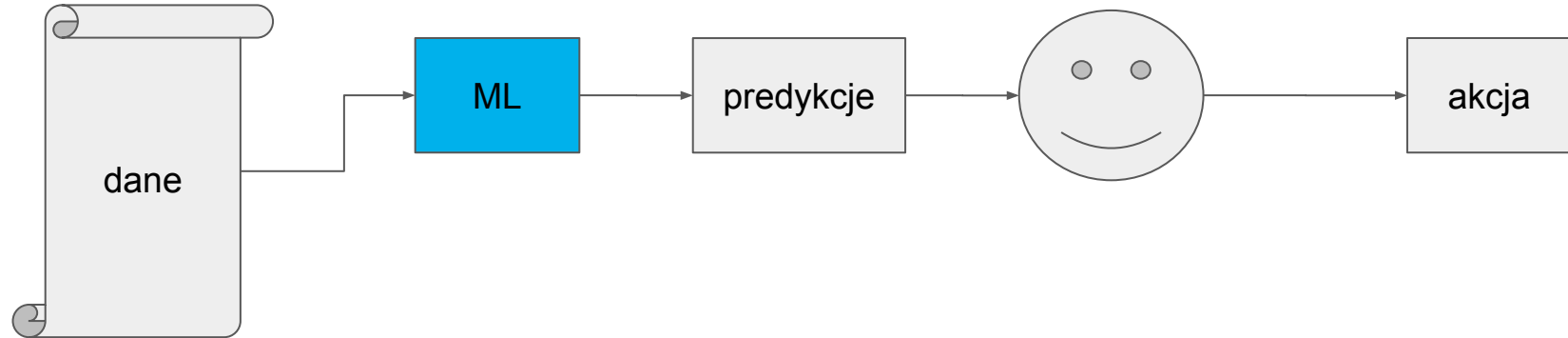
Metoda 1





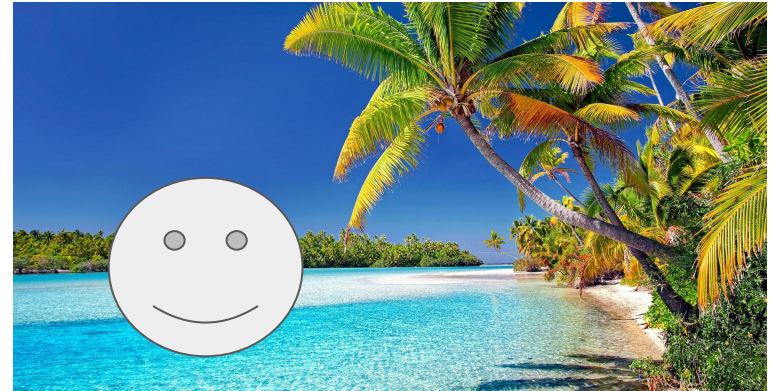
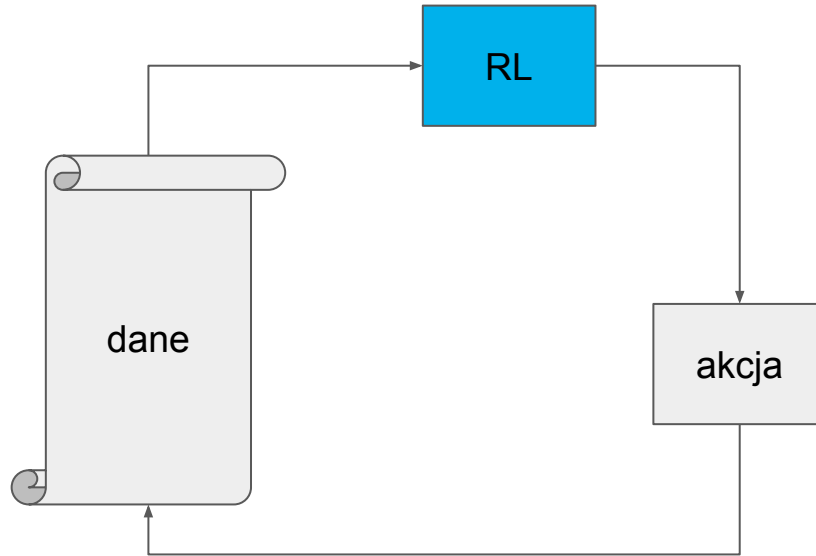
# Ewolucja procesu biznesowego

## Metoda 2



# Ewolucja procesu biznesowego

## Metoda 3

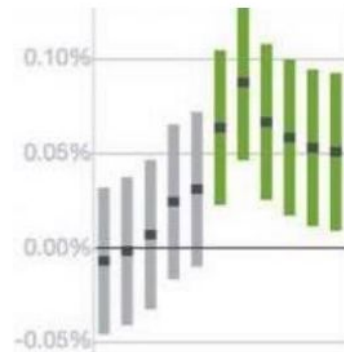


Jutro jest dziś



Horizon

Applied Reinforcement Learning  
@ Facebook



*Figure 1. Real RL model A/B Test Results.* The RL model (test) outperforms the non-RL model (control) on the push notification optimization task described in section 10.1. The x-axis shows the progression of the metric being optimized by day. Note, the performance of the RL model starts out neutral vs. the control, but quickly exceeds as it re-trains daily on data generated by itself.



---

**Dziękuję za uwagę!**

deepsense.ai