



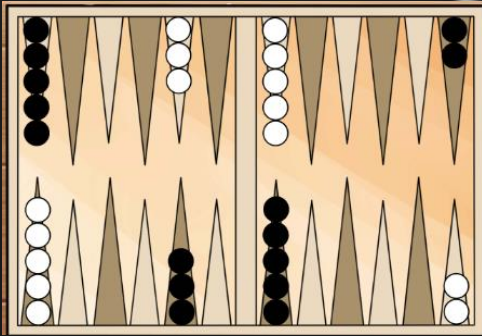
# Deep reinforcement learning

# Agenda

1. Wprowadzenie
2. OpenAI Five: Triumf Sztucznej Inteligencji
3. Wyzwania w Dota 2
4. Narzędzia Ciągłego Treningu
5. Cos się wymysli



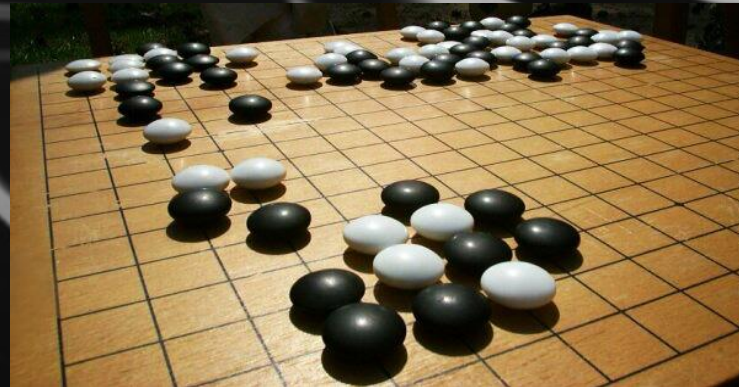
# AI vs human



**1992 Backgammon**



**1997 Chess**



**2016 Go**





DOTA 2





# Challenges

- **Long time horizons.**
- **Partially-observed state.**
- **High-dimensional action and observation spaces.**

The image features a solid teal background. In the center, the text "OpenAIFive" is written in a white, serif font. Surrounding the text are several human hands and forearms, all raised and palms facing forward, as if in a gesture of agreement or voting. The hands are positioned at various heights and angles, creating a sense of collective action.

OpenAIFive



# Limitations

- **Subset of 17 heroes**
- **No support for items which allow a player to temporarily control multiple units at the same time**



# Playing Dota 2 using AI

- Humans interact with the Dota 2 game using three things: keyboard, mouse and computer monitor
- The Dota 2 engine runs at 30 frames per second, hence OpenAI Five only acts on every 4th frame. It is called timestamp
- OpenAI collects ~ 16 000 total values each time step





# Observation space

<b>Global data</b>	<b>22</b>	<b>Per-hero add'l (10 heroes)</b>	<b>25</b>	<b>Per-modifier (10 heroes x 10 modifiers &amp; 179 non-heroes x 2 modifiers)</b>	<b>2</b>
time since game started	1	is currently alive?	1	remaining duration	1
is it day or night?		number of deaths	1	stack count	1
time to next day/night change	2	hero currently in sight?		modifier name	1
time to next spawn: creep, neutral, bounty, runes	4	time since this hero last seen	2	<b>Per-item (10 heroes x 16 items)</b>	<b>13</b>
time since seen enemy courier is that > 40 seconds? <sup>a</sup>	2	hero currently teleporting? if so, target coordinates (x, y)		location one-hot (inventory/backpack/stash)	3
min&max time to Rosh spawn	2	time they've been channeling	4	charges	1
Roshan's current max hp	1	respawn time	1	is on cooldown?	
is Roshan definitely alive?	1	current gold (allies only)	1	cooldown time	2
is Roshan definitely dead?	1	level	1	is disabled by recent sum?	
Next Roshan drops cheese?	1	mana: max, current, & regen	3		
Next Roshan drops refresher?	1	health regen rate	1		
		magic resistance	1		



# Reward weights

Name	Reward	Heroes	Description
Win	5	Team	
Hero Death	-1	Solo	
Courier Death	-2	Team	
XP Gained	0.002	Solo	
Gold Gained	0.006	Solo	For each unit of gold gained. Reward is not lost when the gold is spent or lost.
Gold Spent	0.0006	Solo	Per unit of gold spent on items without using courier.
Health Changed	2	Solo	Measured as a fraction of hero's max health. <sup>‡</sup>
Mana Changed	0.75	Solo	Measured as a fraction of hero's max mana.
Killed Hero	-0.6	Solo	For killing an enemy hero. The gold and experience reward is very high, so this reduces the total reward for killing enemies.
Last Hit	-0.16	Solo	The gold and experience reward is very high, so this reduces the total reward for last hit to $\sim 0.4$ .
Deny	0.15	Solo	
Gained Aegis	5	Team	
Ancient HP Change	5	Team	Measured as a fraction of ancient's max health.
Megas Unlocked	4	Team	
T1 Tower <sup>*</sup>	2.25	Team	
T2 Tower <sup>*</sup>	3	Team	
T3 Tower <sup>*</sup>	4.5	Team	
T4 Tower <sup>*</sup>	2.25	Team	
Shrine <sup>*</sup>	2.25	Team	
Barracks <sup>*</sup>	6	Team	
Lane Assign <sup>†</sup>	-0.15	Solo	Per second in wrong lane.

## Game time weighting

$$\rho_i \leftarrow \rho_i \times 0.6^{(T/10 \text{ mins})}$$

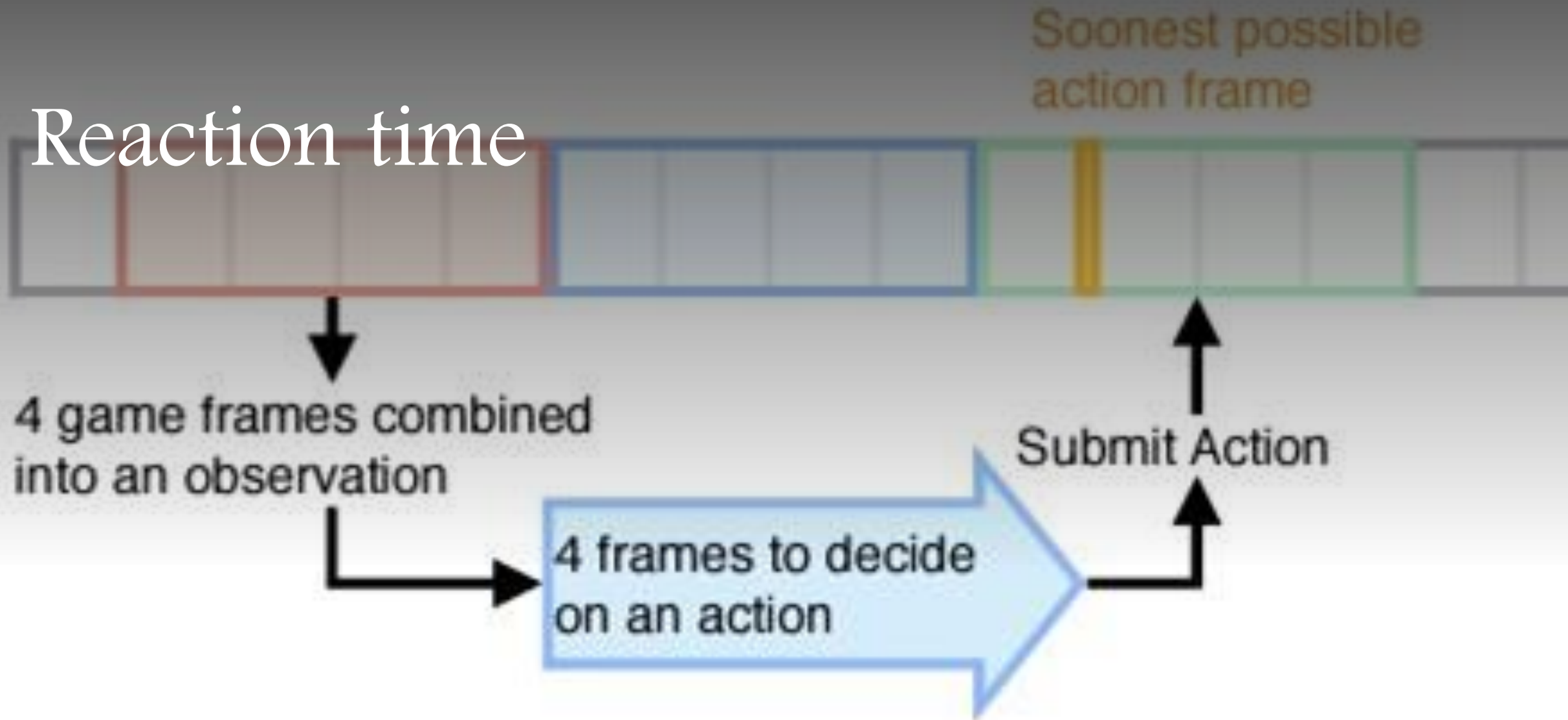


# Team spirit

$$r_i = (1 - \tau)\rho_i + \tau\bar{\rho}$$



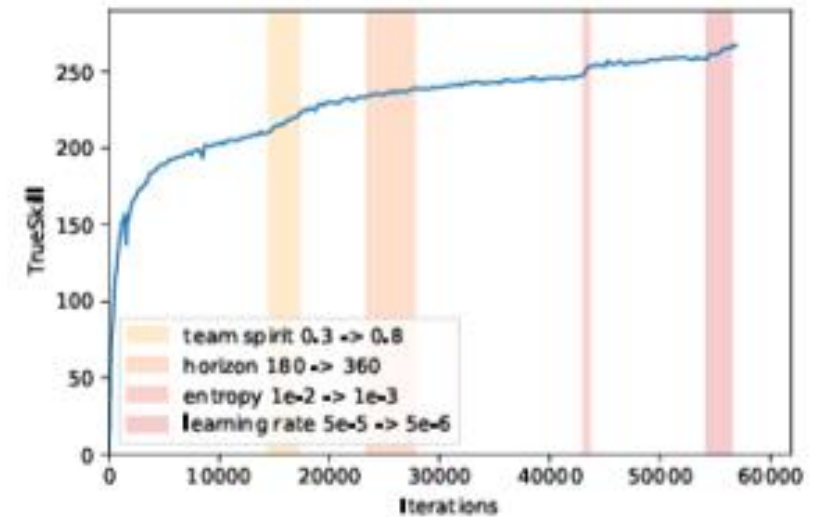
# Reaction time





# Hyperparameter changes over time and optimizing the policy

Iteration	0	15k	23k	43k	54k
Time (days)	0	13	20	33	42
TrueSkill	0	210	232	245	258
<hr/>					
Team Spirit	0.3	0.8			
GAE Horizon	180 secs		360 secs		
Entropy coefficient	0.01			0.001	
Learning Rate	5e-5				5e-6



# Surgery methodology





# Can u beat it?

Opponent	Result	Duration	Version	Restrictions
June 6, 2018 - Internal Event				
Internal team	win	15:15 (surr)	7.13	Mirror match, multiple couriers, no invis
Internal team	win	20:51	7.13	Mirror match, multiple couriers, no invis
Audience team	win	31:33	7.13	Mirror match, multiple couriers, no invis
Audience team	win	23:33 (surr)	7.13	Mirror match, multiple couriers, no invis
August 5, 2018 - Benchmark				
Caster team	win	21:38 (surr)	7.16	Drafted, multiple couriers
Caster team	win	24:56 (surr)	7.16	Drafted, multiple couriers
Caster team	lose	35:47	7.16	Audience draft, multiple couriers
August 9, 2018 - Private eval				
Team Secret	win	17:00 (surr)	7.16	Drafted, multiple couriers
Team Secret	lose	48:46	7.16	Drafted, multiple couriers
Team Secret	lose	38:55	7.16	Drafted, multiple couriers
August 22-23, 2018 - The International				
Pain Gaming	lose	52:29	7.19	Pre-set lineup
Chinese Legends	lose	45:44	7.19	Pre-set lineup
October 5, 2018 - Private eval				
Team Lithium	win	48:57	7.19	TI pre-set lineup
Team Lithium	win	48:16	7.19	TI pre-set lineup
Team Lithium	win	31:33	7.19	Drafted
January 16, 2019 - Private eval				
SG Esports	win	24:29 (surr)	7.19	TI pre-set lineup
SG Esports	win	25:08 (surr)	7.19	Drafted
SG Esports	win	27:36 (surr)	7.20	Mirror match
SG Esports	win	25:30 (surr)	7.20	Mirror match
February 1, 2019 - Private eval				
Alliance	win	17:11	7.20d	Drafted
Alliance	win	31:33	7.20d	Drafted
Alliance	win	28:16	7.20d	Reverse drafted
April 13, 2019 - OpenAI Five Finals				
OG	win	38:18	7.21d	Drafted
OG	win	20:51	7.21d	Drafted

Table 7: Major matches of OpenAI Five against high-skill human players.



# Experiments and Evaluation

- AI uczyło się 10 miesięcy od 30 czerwca 2018 do 22 kwietnia 2019
- Pokonało mistrzów świata oraz 99.4% graczy
- Dużo większa skala niż podobne poprzednie AI jak np. AlphaGo

# Human Evaluation



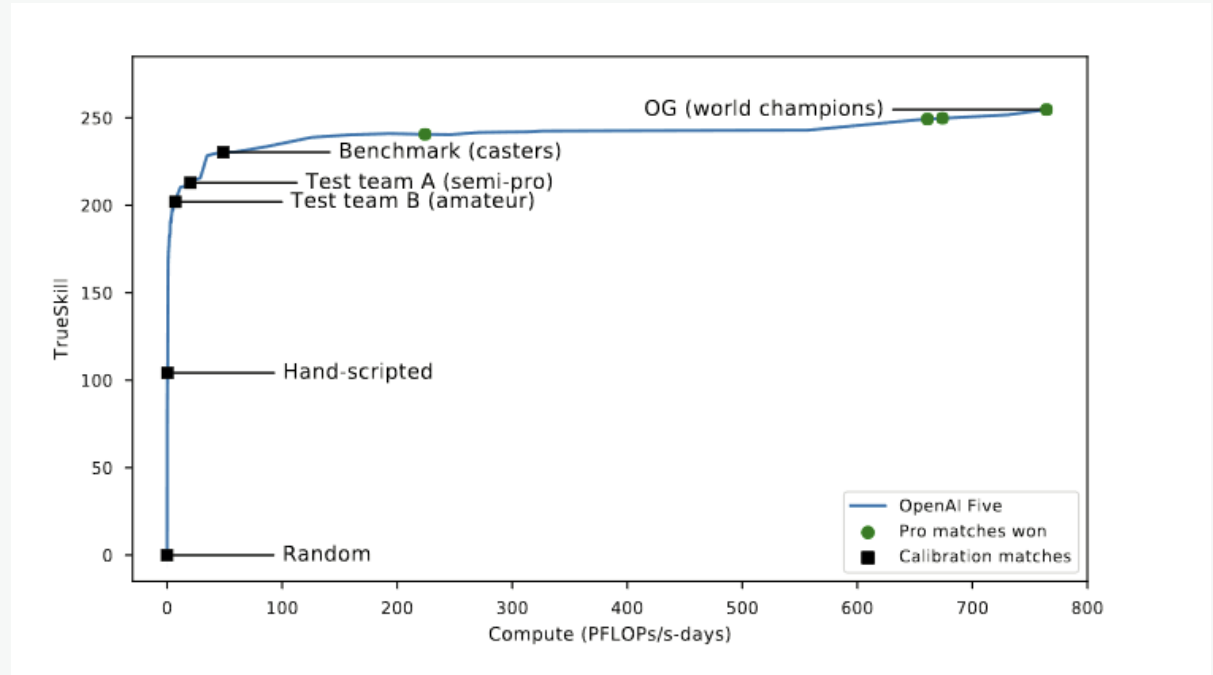
OpenAI Five pokonało mistrzów  
świata 13 kwietnia 2019



18-21 kwietnia OpenAI Five  
Arena



# TrueSkill



# Playstyle

Na początku unikalny,  
miał swoje wady

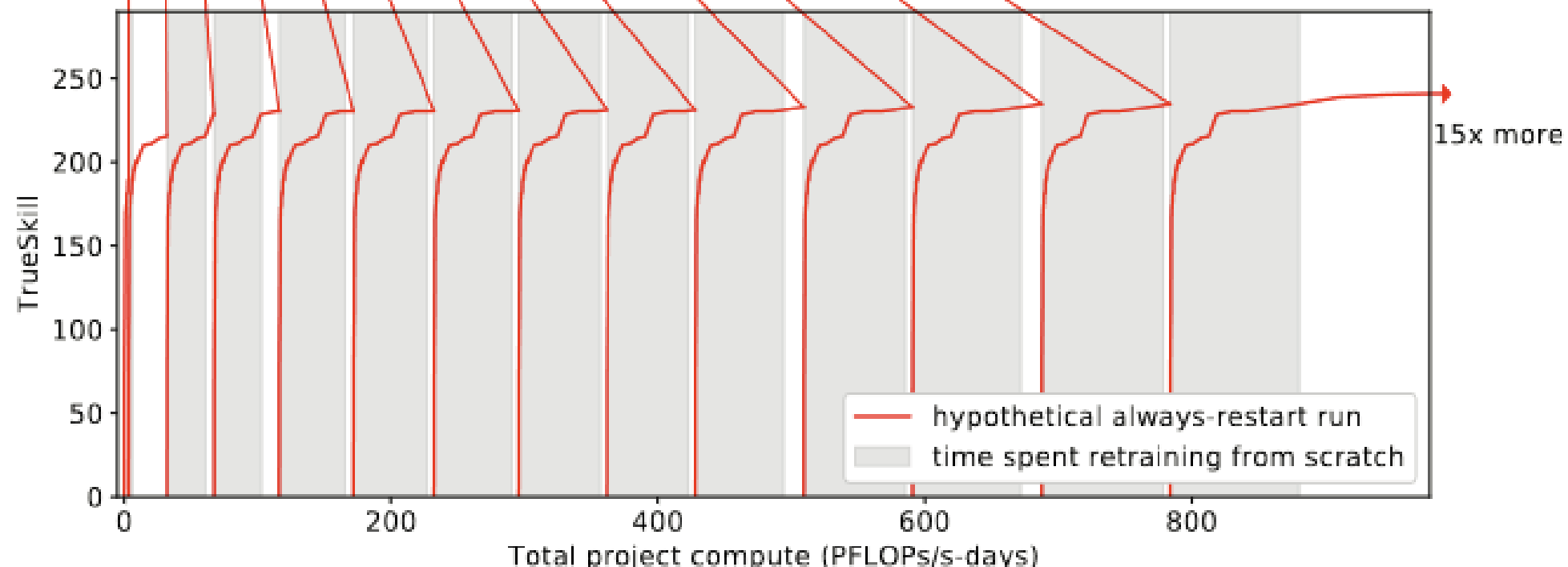
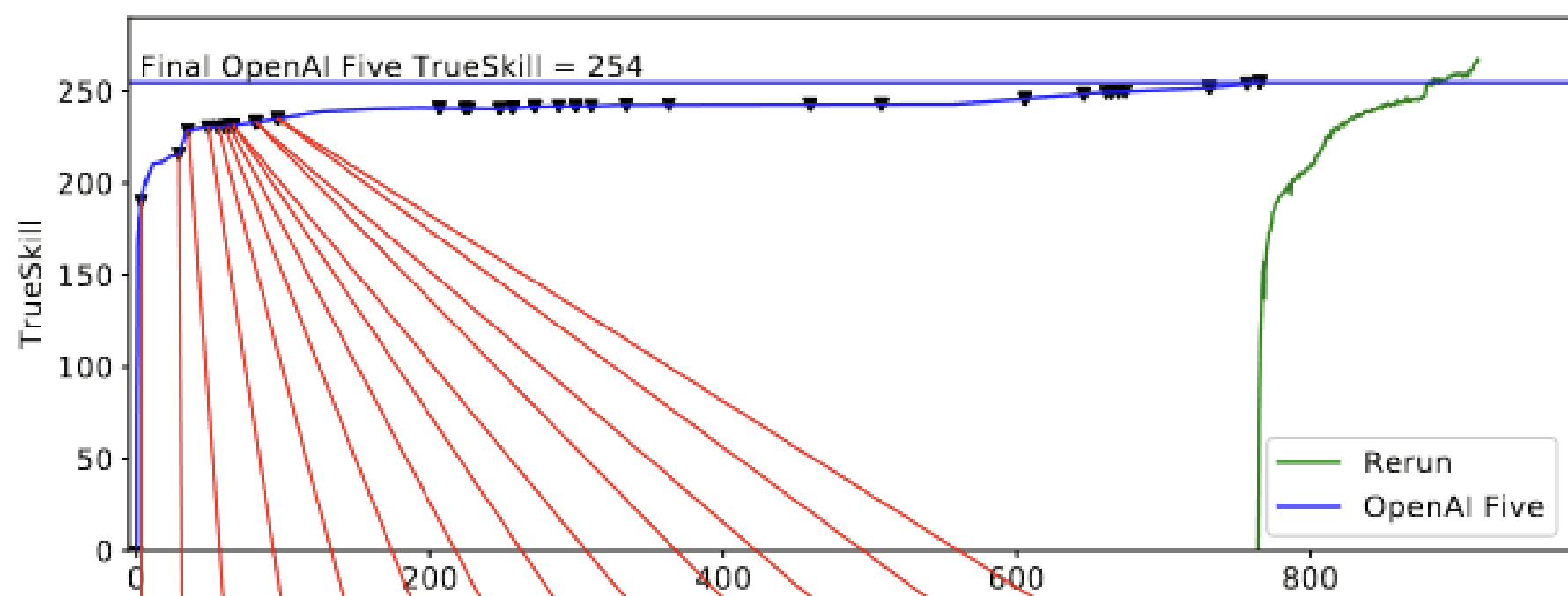


Z czasem był bardziej  
ludzki ale nadal miał  
unikalne dla siebie  
zachowania

# Validating surgery with rerun

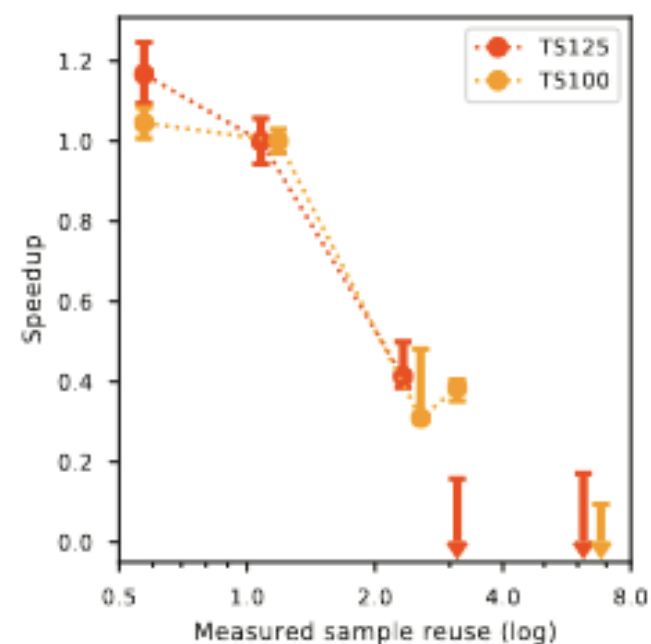
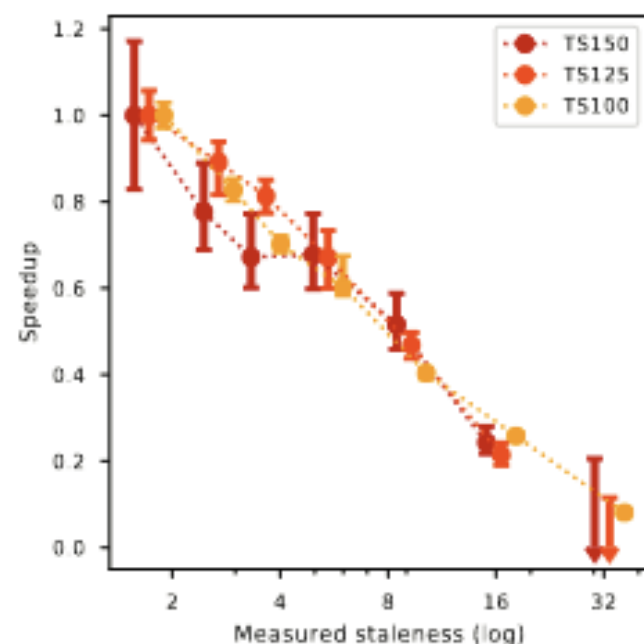
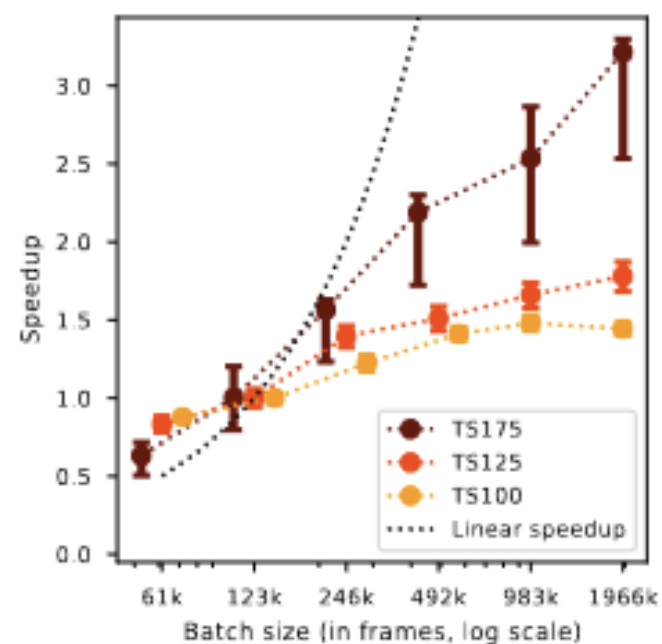
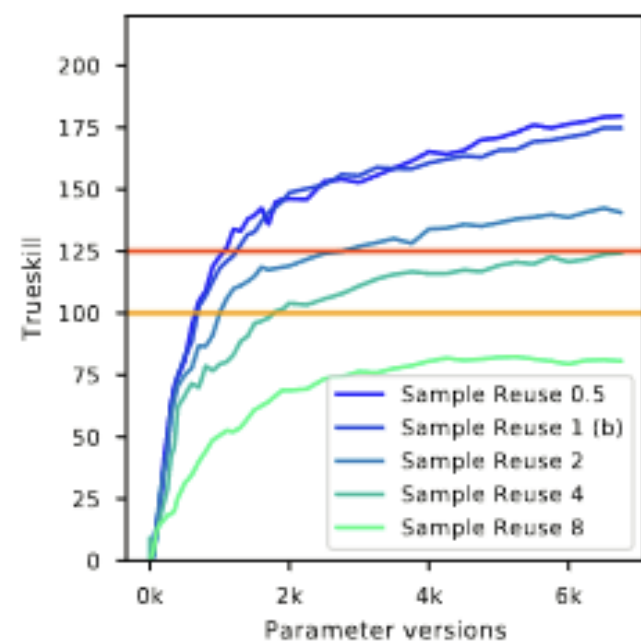
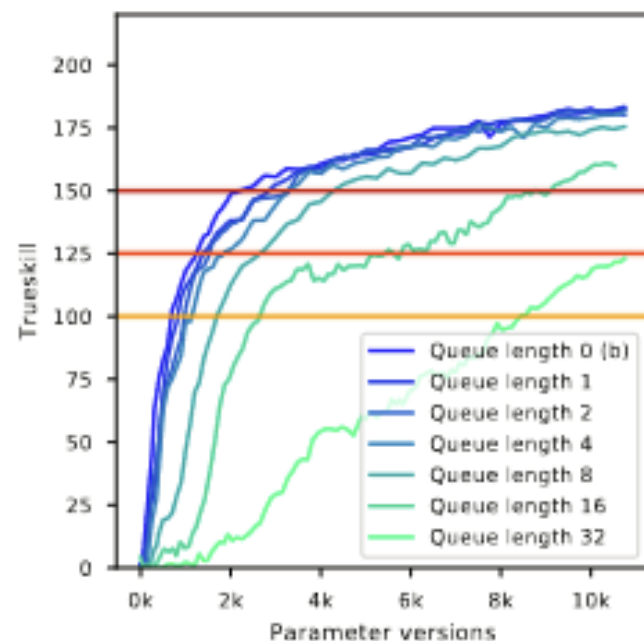
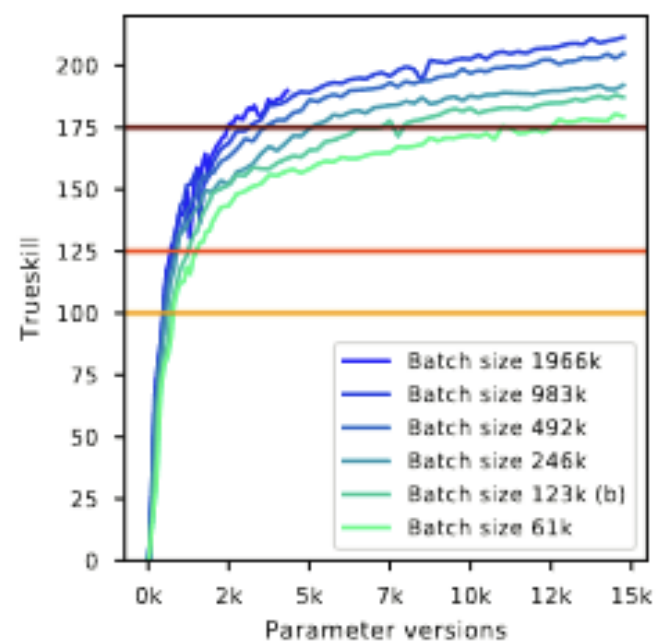
- Rerun trwał 2 miesiące
- Jeżeli zamiast surgery wykonywany był restart uczenia, AI uczyło by się 40 zamiast 10 miesięcy
- Rerun miał 98% win rate z OpenAI Fice
- Surgery jest potencjalnie do poprawienia, gdyż rerun osiąga wyższy TS





# Batch size

- Zwiększenie liczby GPU oraz liczby rollout machines
- Pożądany speedup spowodowany zwiększaniem batch size jest liniowy
- Liniowy speedup nie został osiągnięty, jednak był on znaczący





# Data Quality



Około 2h trwania jednej gry powoduje problemy podczas uczenia



Twórcy użyli podejścia asynchronicznego w procesie uczenia

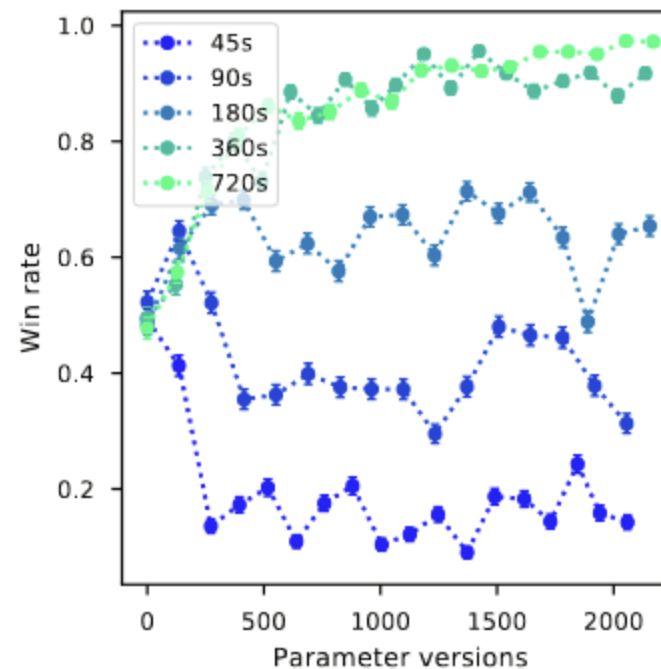


Pojawiły się problemy staleness i sample reuse

# Long term credit assignment

- Dota 2 ma bardzo długi czas zależności między ruchami
- Time horizon

$$H = \frac{T}{1 - \gamma}$$



# Podsumowanie

- OpenAI Five osiągnęło nadludzki poziom w grze Dota2
- Kluczowym składnikiem było zwiększenie batch size'u oraz czasu uczenia z użyciem surgery
- Możliwe, że rezultaty mogą być przełożone na inne gry
- Wraz ze zwiększaniem się złożoności problemów i środowisk skalowanie będzie jeszcze ważniejsze



The background is an abstract composition of overlapping, semi-transparent geometric shapes, primarily hexagons and polygons, in various shades of blue (from light to dark) and grey. The shapes are arranged in a way that creates a sense of depth and movement, with some shapes appearing to be in the foreground and others receding into the background. The overall effect is a modern, digital aesthetic.

Dziękujemy za  
uwagę