

# Chess XAI study

## Weź udział w badaniu!

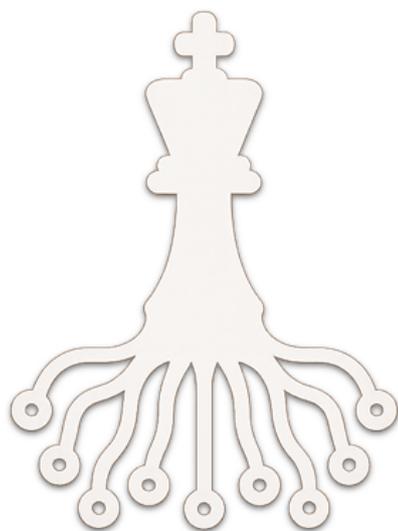
Zostaw swój adres e-mail poprzez formularz. Badanie będzie **anonimowe** - wyślemy Ci odpowiednie instrukcje, gdy rozpoczniemy eksperyment.

<https://forms.gle/hEjpCTdZaka7sSsz8>

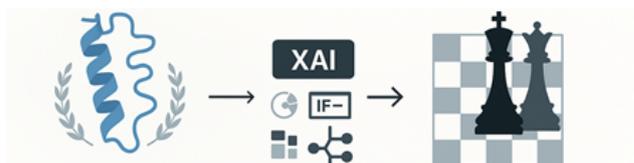
Dziękujemy!



## Jak szachiści mogą pomóc zrozumieć AI?

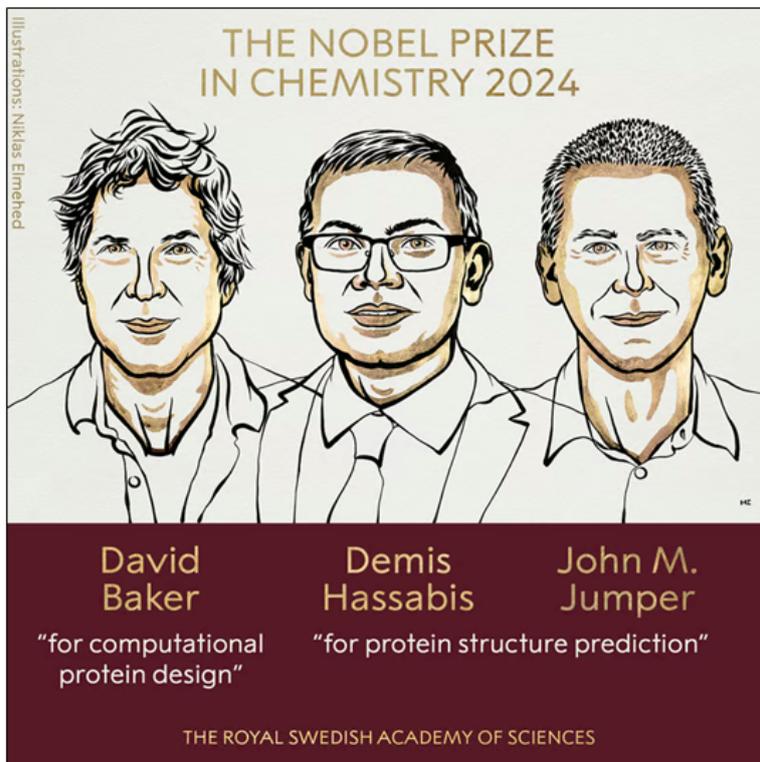


dr hab. Andrzej Siódмок, prof. UJ  
Uniwersytet Jagielloński





# Rok temu: AlphaFold święty Graal biologii



“Chess helped me win the Nobel Prize” D. Hassabis



“so because of ... chess I started to think about thinking... and I started to improve my own thought processes... the computer play chess well I was intrigued and that got me into A.I.”

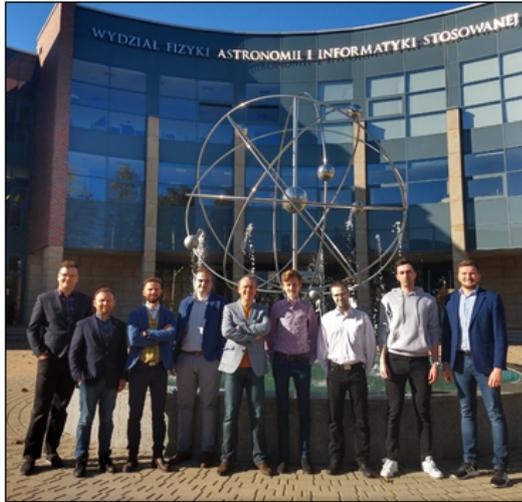


# AI przyjazne i użyteczne dla ludzi



Zakład Humanocentrycznej Sztucznej Inteligencji  
Uniwersytet Jagielloński

GEIST Research Group  
We are GEIST. We dream big and work hard.



Maciej Szelązek



Maciej Mozolewski



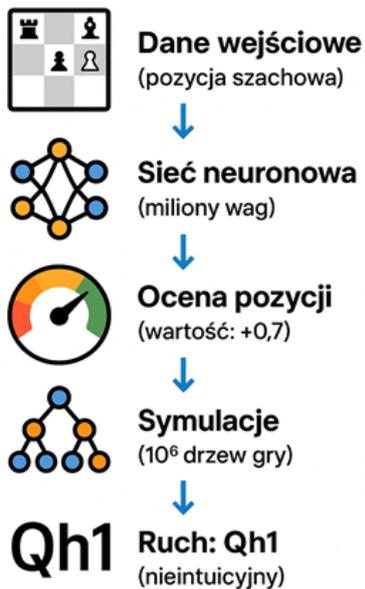
Zakład HCAI UJ oraz grupa  
GEIST kierowane przez  
prof. Grzegorza J. Nalepę





# Problemy z interpretacją AI

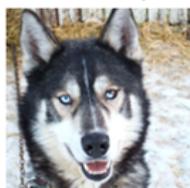
## Jak AI „myśli”?



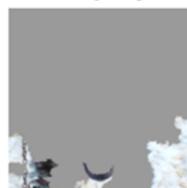
## Nieintuicyjne cechy opisujące dane

```
relative_castling_rights
relative_piece_connectivity
wk_fgh_l2_bk_fgh_78
wk_not_l2_bk_de_78
wk_not_l2_bk_fgh_78
difference_king_safety_own_pieces
difference_castling_rights
.
```

## Wnioski na podstawie błędnych przesłanek = ryzyko błędu



(a) Husky classified as wolf



(b) Explanation

Pytanie    *Husky czy wilk?*

Odpowiedź    *Husky*

Dlaczego?    *Wykryto śnieg*



# Co chcemy zrobić?

## Jak objaśnić wnioskowanie?



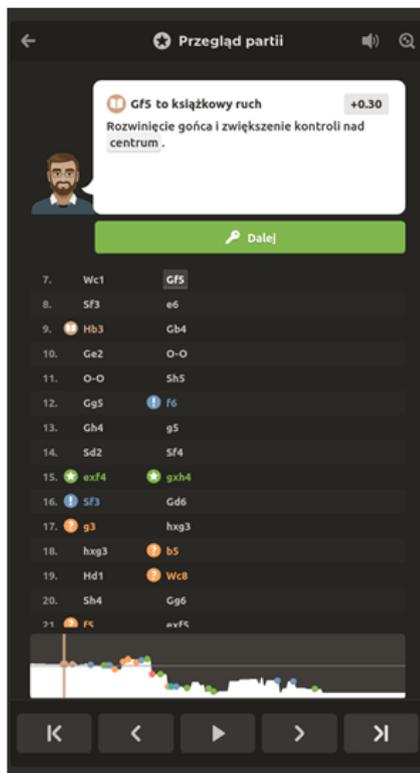
**Czarne pudełko (black box)**  
→ 10 mln wag → decyzja



**XAI: algorytmiczne metody objaśnialności**  
→ Ruch Qh1: 0,4 za linię h



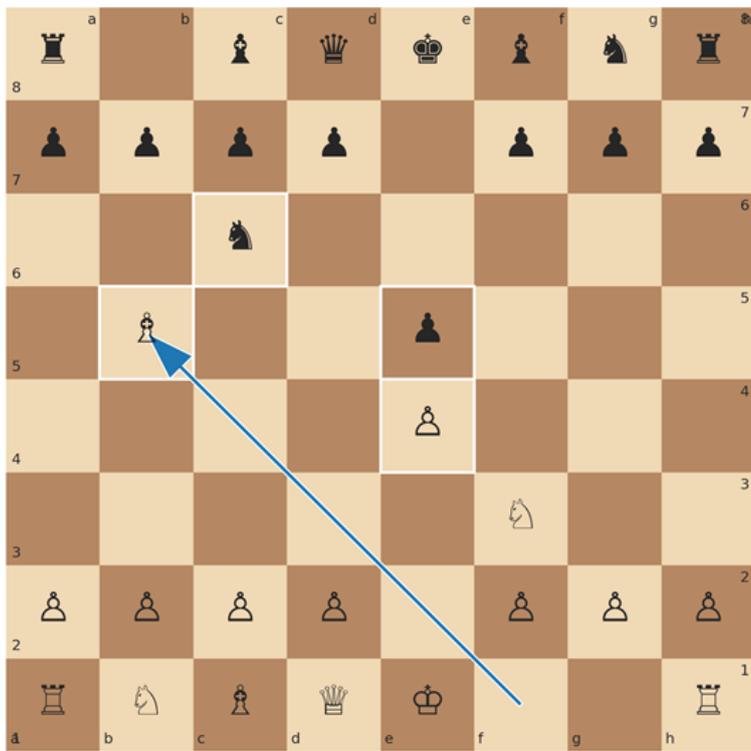
**Tłumaczenie na język ludzi**  
→ AI widzi długoterminowy atak na króla





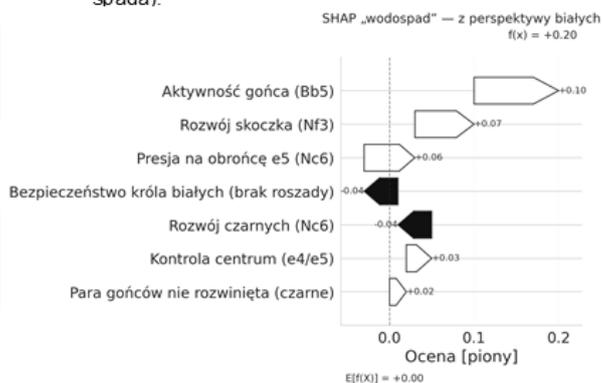
# Krótko o eXplainable AI

Ruy Lopez (Hiszpańska) — po 3.Bb5



## SHAP (Shapley Additive Explanations)

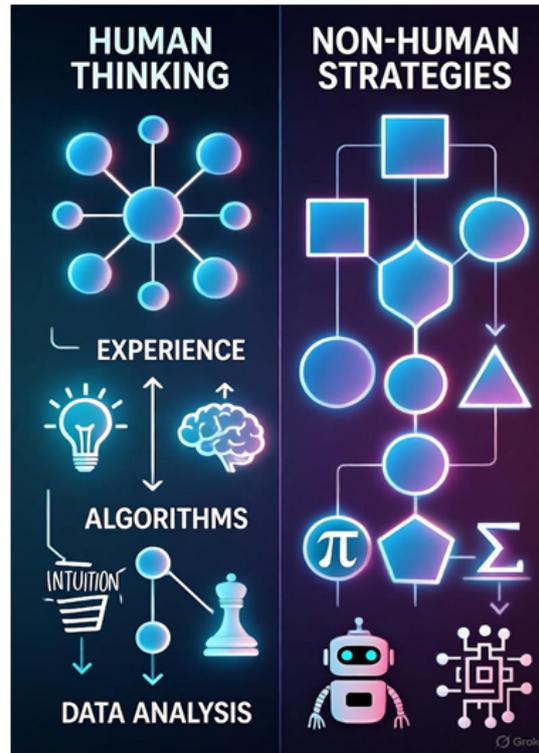
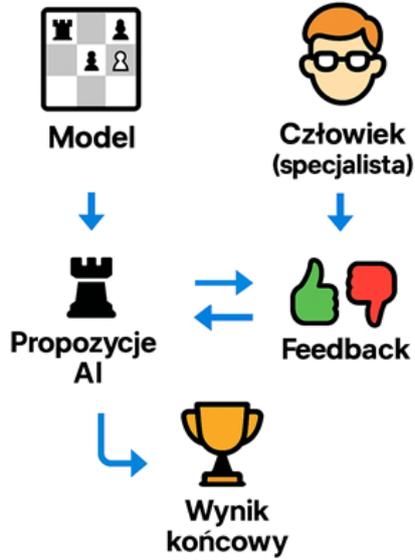
- O ile każda cecha podnosi lub obniża wynik modelu dla konkretnego przykładu.
- Wykorzystuje wartości **Shapleya** z teorii gier: „sprawiedliwie” dzieli wpływ między cechy, biorąc pod uwagę ich współdziałanie.
- Ma dwie kluczowe własności: **dokładność lokalną** (suma wkładów = wynik) i **spójność** (jeśli cecha pomaga bardziej w modelu, jej wkład nie spada).





# Jak to osiągnąć?

## HITL w szachach (Human-in-the-Loop)





## Dlaczego potrzebujemy właśnie Was

- **Ekspercka intuicja i plan:** jak formułować objaśnienia. "Ask an expert"



- **Domain knowledge:** które objaśnienia są użyteczne, co realnie „działa” przy zegarze i presji



The screenshot shows a chess game in progress on a green and white board. The AI explanation overlay on the right side of the board highlights a move: **Kh8** is the best move, with a score of **-2.84**. The text says "To jest najsilniejsza opcja." (This is the strongest option). Below this, there are buttons for "Pokaż kontynuację" (Show continuation), "Spróbuj ponownie" (Try again), and "Dalej" (Next). A list of moves is shown below the buttons, with the selected move **Kh8** highlighted in green. The list includes moves like Wf1, Sf3, Wd3, Kg2, Hf2, Kf1, Wxe7, Wxg7, Kxf2, Kxf3, Kf4, Wg3+, Wd3, Wd4, and r4. At the bottom of the interface, there is a navigation bar with a play button and a timestamp of 0:45. The date 8.11.2025 is displayed at the bottom right.

# Krótko o eXplainable AI

Przeгляд partii

Kh8 jest najlepsze  
To jest najsilniejsza opcja. -2.84

Pokaż kontynuację

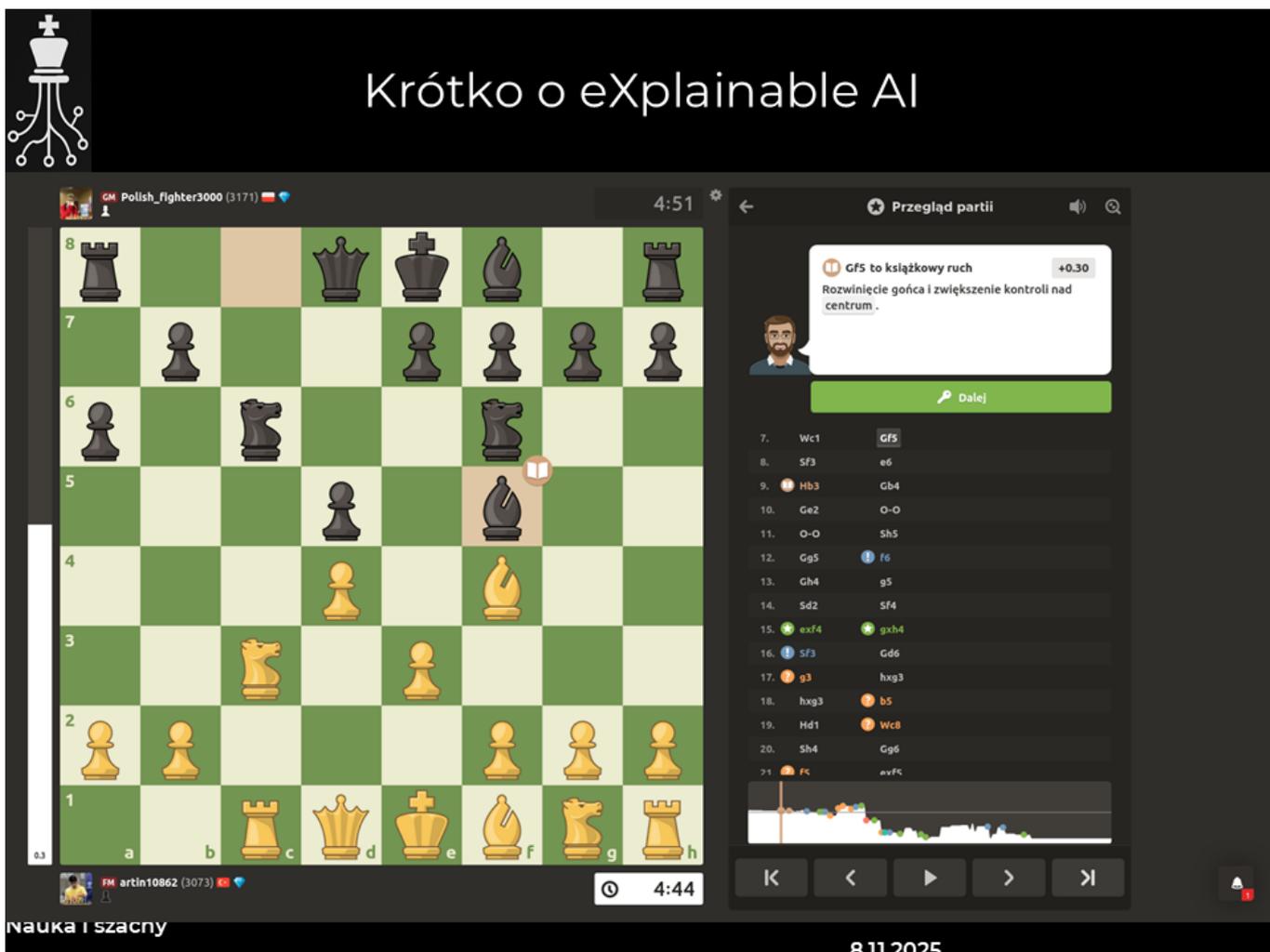
Spróbuj ponownie Dalej

32.	Wf1	Kh8
33.	Sf3	Gh5
34.	Wd3	Hh6
35.	Kg2	Gxf4
36.	Hf2	Hg7+
37.	Kf1	Gg3
38.	Wxe7	Gxf2
39.	Wxg7	Kxg7
40.	Kxf2	Gxf3
41.	Kxf3	Wf6
42.	Kf4	Kg6
43.	Wg3+	Kf7
44.	Wd3	h6
45.	Wd4	Kg6
46.	r4	hvr4

0:45

8.11.2025

# Krótko o eXplainable AI



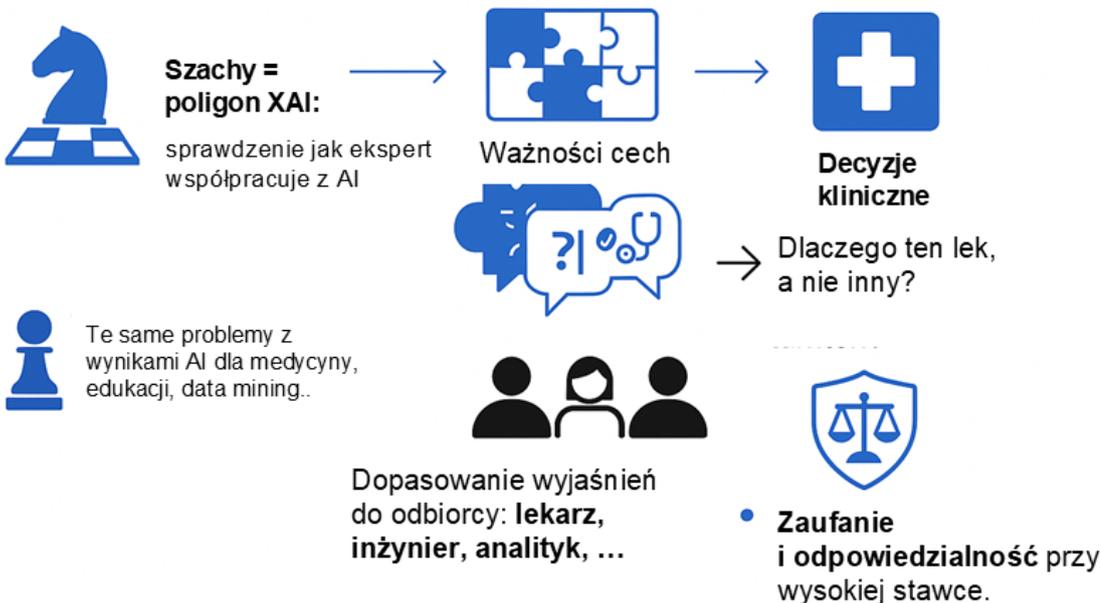
The screenshot shows a chess game in progress on a mobile device. The board is in a standard orientation with White on top. The game is at move 21. An AI explanation overlay is visible on the right side of the board, highlighting a move. The explanation text is: "Gf5 to książkowy ruch. Rozwinięcie gońca i zwiększenie kontroli nad centrum." Below the text is a green button labeled "Dalej". A list of moves is shown below the explanation, with the 15th move highlighted in green: "15. exf4 gxh4".

7.	Wc1	Gf5
8.	Sf3	e6
9.	Hb3	Cb4
10.	Ge2	O-O
11.	O-O	Sh5
12.	Gg5	f6
13.	Gh4	g5
14.	Sd2	Sf4
15.	exf4	gxh4
16.	Sf3	Gd6
17.	g3	hxg3
18.	hxg3	b5
19.	Hd1	Wc8
20.	Sh4	Gg6
21.	f4	exf4

Nauka i szachy

8.11.2025

# Po co to wszystko: od szachów do medycyny



From: <https://geist.re/> - **GEIST Research Group**

Permanent link: <https://geist.re/pub:projects:chessxai:start?rev=1762554086>

Last update: **2025/11/07 22:21**

