

Chess XAI study

Weź udział w badaniu!

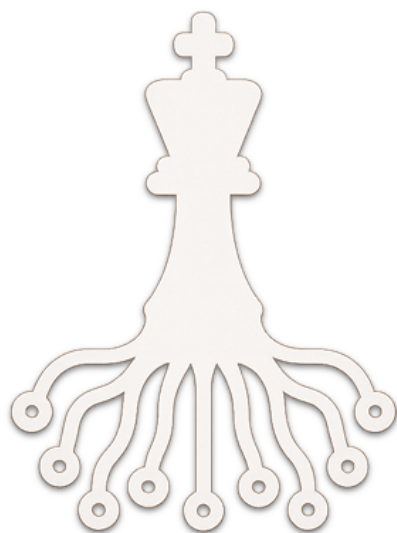
Zostaw swój adres e-mail poprzez formularz. Badanie będzie **anonimowe** - wyślemy Ci odpowiednie instrukcje, gdy rozpoczniemy eksperyment.

[Otwórz formularz zgłoszeniowy](#)

Dziękujemy!



Jak szachiści mogą pomóc zrozumieć AI?

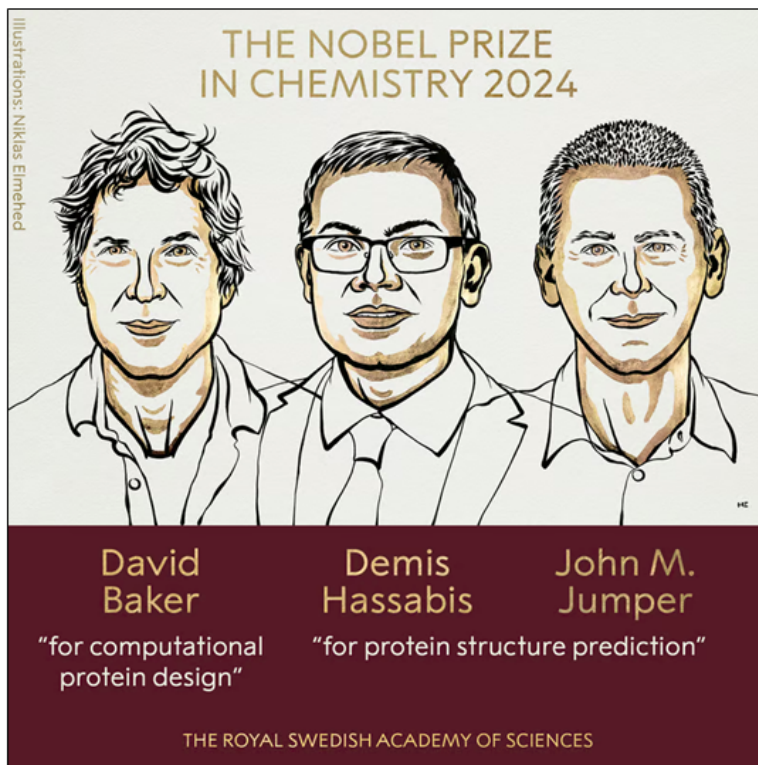


dr hab. Andrzej Siódмок, prof. UJ
Uniwersytet Jagielloński





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.”

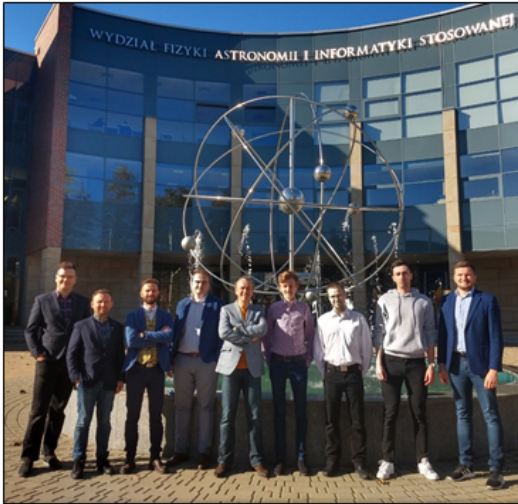


AI przyjazne i użyteczne dla ludzi



Zakład Humanocentrycznej Sztucznej Inteligencji
Uniwersytet Jagielloński

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



Maciej Szelażek



Maciej Mozolewski



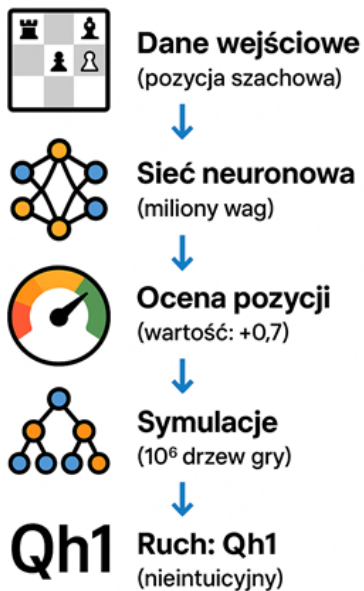
Zakład HCAI UJ oraz grupa
GEIST kierowane przez
prof. Grzegorza J. Nalepę





Problemy z interpretacją AI

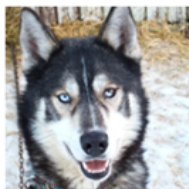
Jak AI „myśli”?



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?



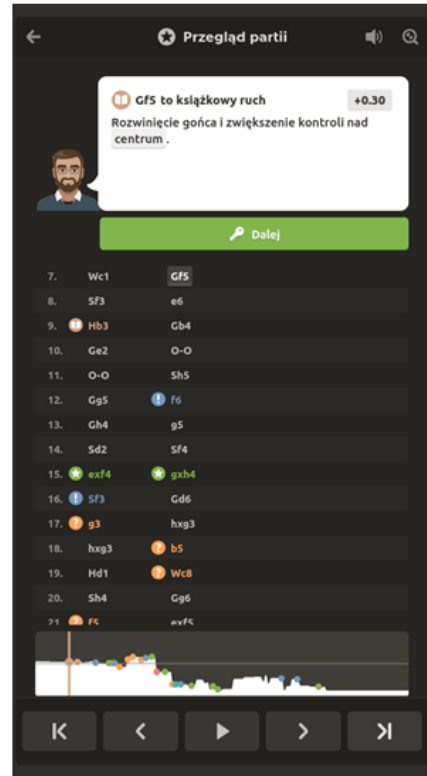
Czarne pudełko
(black box)
→ 10 mln wag → decyzja



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



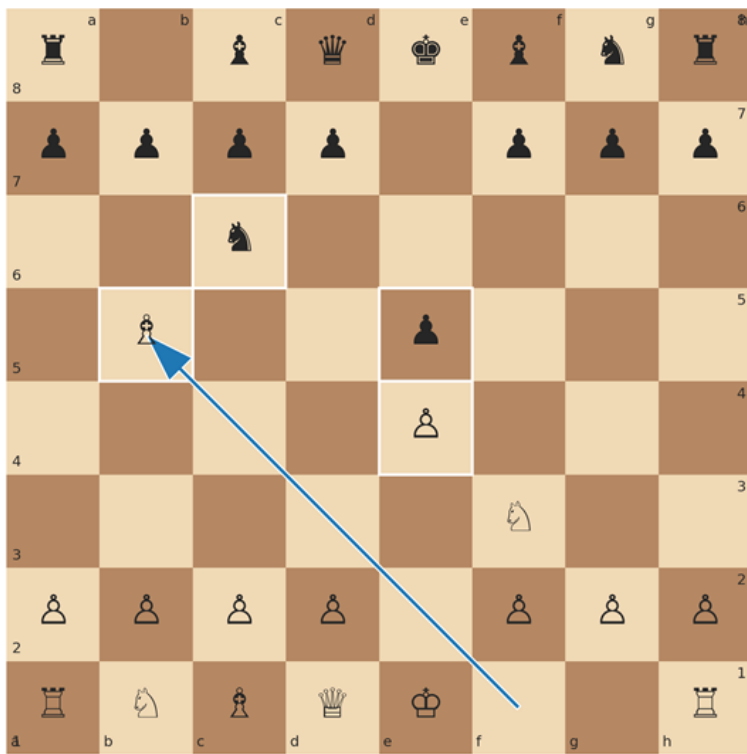
Tłumaczenie na język ludzi
→ AI widzi długoterminowy atak na króla





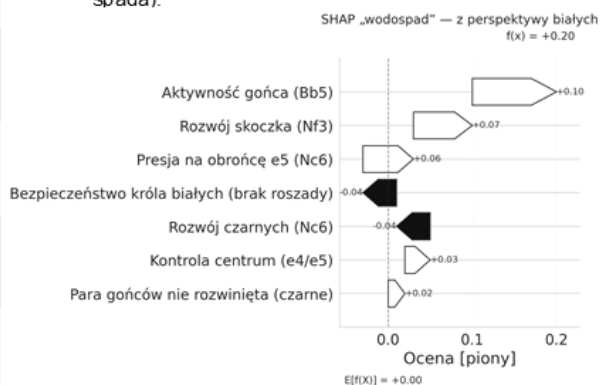
Krótko o eXplainable AI

Ruy Lopez (Hiszpańska) — po 3.Bb5



SHAP (Shapley Additive Explanations)

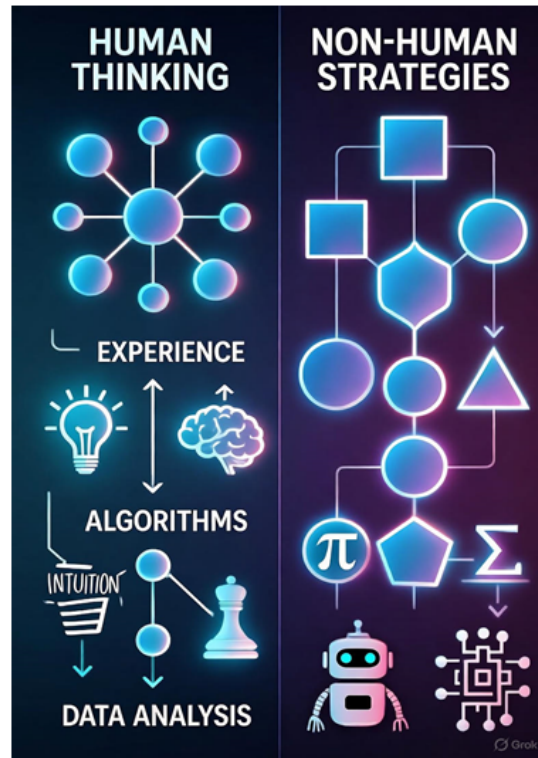
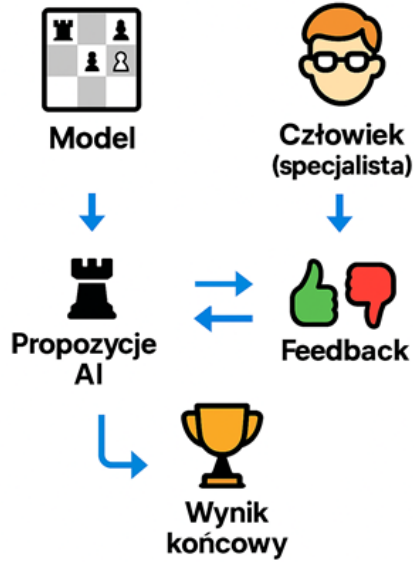
- O ile każda cecha podnosi lub obniża wynik modelu dla konkretnego przykładu.
- Wykorzystuje wartości **Shapleya** z teorii gier: „sprawiedliwie” dzieli wpływ między cechy, biorąc pod uwagę ich współdziałanie.
- Ma dwie kluczowe własności: **dokładność lokalną** (suma wkładów = wynik) i **spójność**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





Krótko o eXplainable AI

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 displays the following information:

- Header: **Przegląd partii**
- Message: **Kh8 jest najlepsze** (with a green plus icon) and **-2.84** (in a black box).
- Text: **To jest najsilniejsza opcja.**
- Buttons: **Pokaż kontynuację** (with a speech bubble icon), **Spróbuj ponownie** (with a refresh icon), and **Dalej** (with a right arrow icon).
- Move list (moves 32-46):

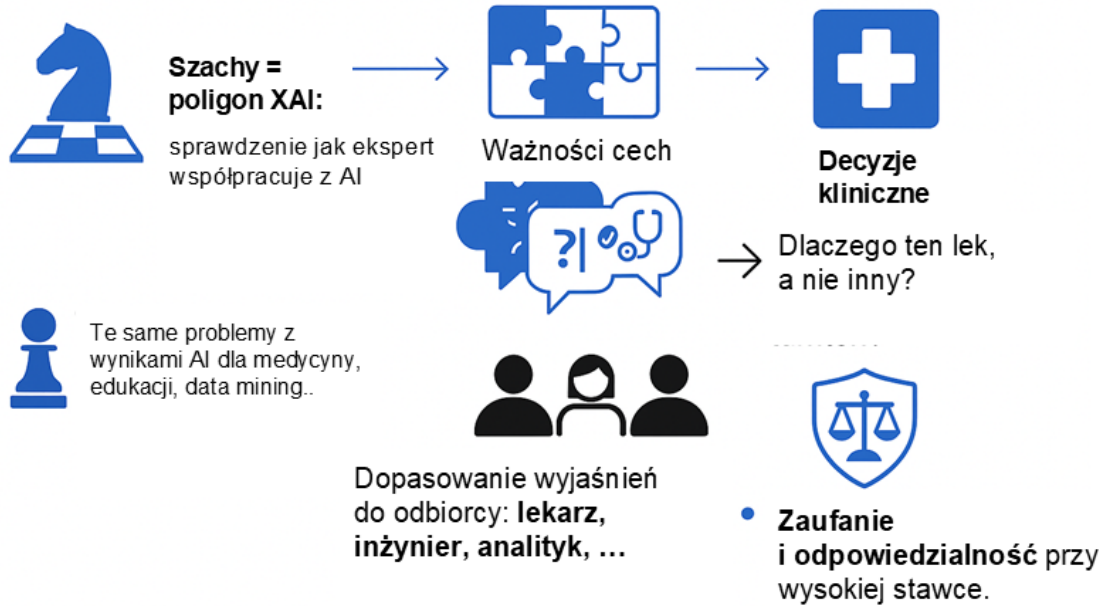
32.	Wf6	Kh8
33.	Sf3	Gh5
34.	Wd3	Hh6
35.	Kg2	Gxf4
36.	Hf2	Hg7+
37.	Kf1	Gg3
38.	Wxe7	Gxf2
39.	Wxg7	Kxg7
40.	Kxf2	Gxf3
41.	Kxf3	Wf6
42.	Kf4	Kg6
43.	Wg3+	Kf7
44.	Wd3	h6
45.	Wd4	Kg6
46.	r4	hvr4
- Timeline: A horizontal bar at the bottom of the move list shows a color-coded progression of moves.
- Navigation: **K** (back), **<**, **>**, **>>** (forward) buttons.

At the bottom left of the interface, the text **Nauka i szachy** is visible. At the bottom right, the date **8.11.2025** is displayed.

Krótko o eXplainable AI

The screenshot shows a chess game in progress on a digital platform. 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 specific move. The explanation text reads: "Gf5 to książkowy ruch. Rozwiniecie 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 21st move highlighted: "21. fxc3 e4xc3". The interface includes player names (GM Polish_fighter3000 and FM artin10862), a timer (4:51), and a logo in the top left corner depicting a chess king with neural network connections. The bottom of the screen features the text "Nauka i szachy" and the date "8.11.2025".

Po co to wszystko: od szachów do medycyny



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

Permanent link:
<https://www.geist.re/pub:projects:chessxai:start?rev=1762555174>

Last update: **2025/11/07 22:39**

