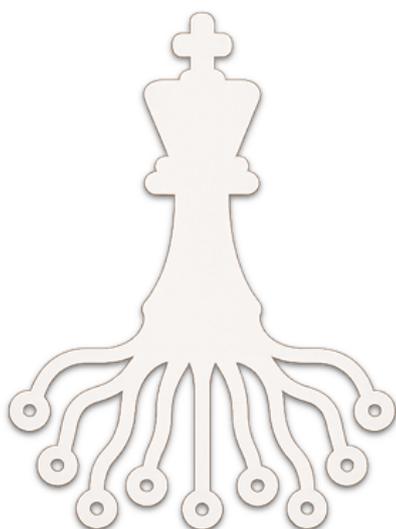
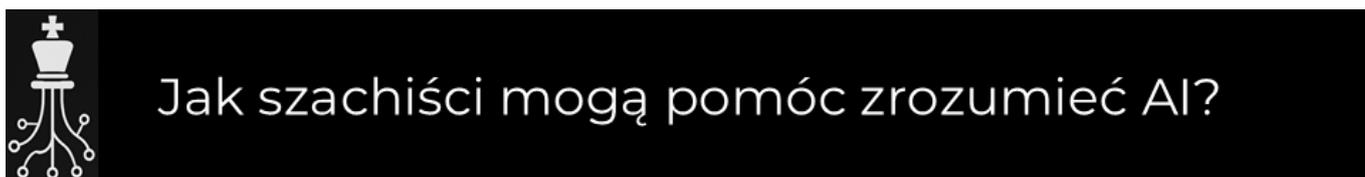


Chess XAI study

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!

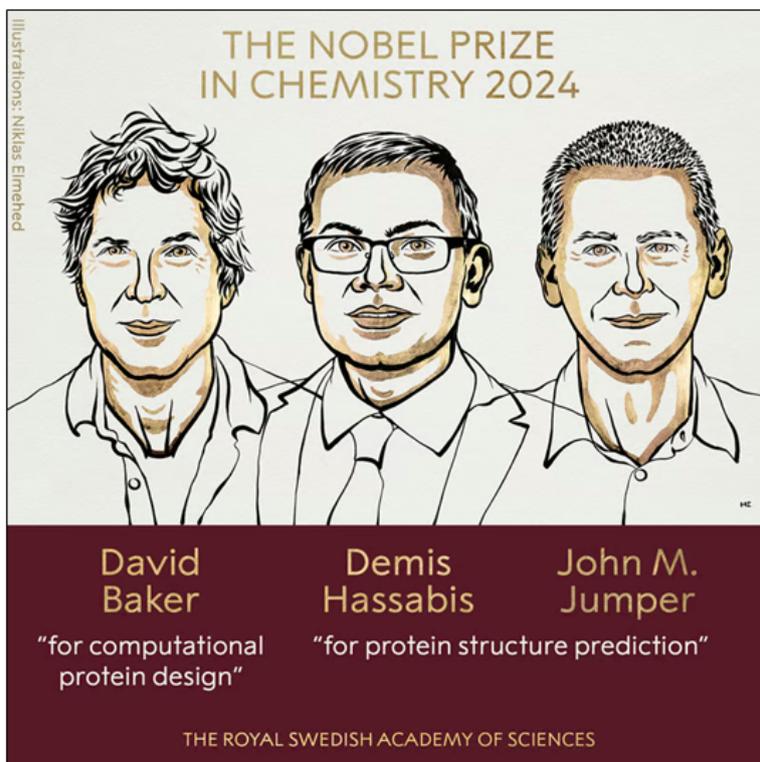


dr hab. Andrzej Siódmok, prof. UJ
Uniwersytet Jagielloński





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.”

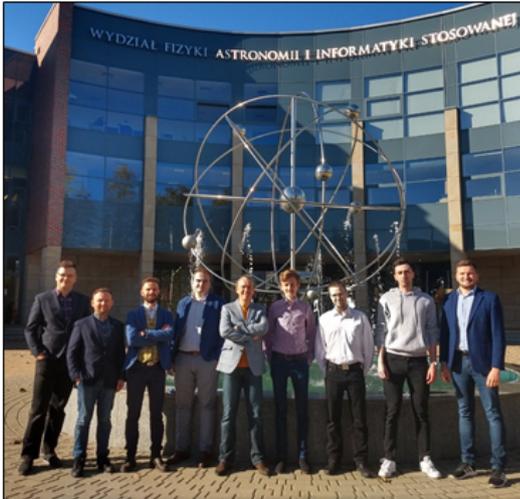


AI przyjazne i użyteczne dla ludzi



Zakład Humanocentrycznej Sztucznej Inteligencji
Uniwersytet Jagielloński

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



Maciej Szelązek



Maciej Mozolewski



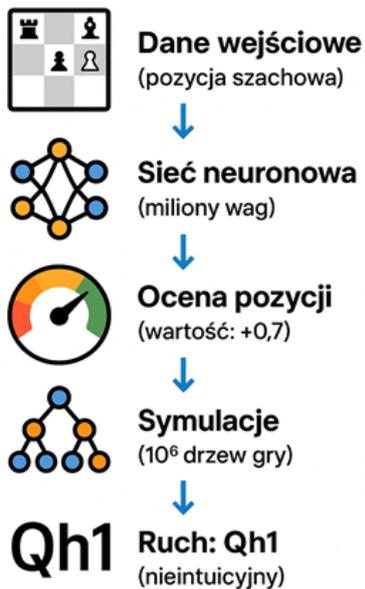
Zakład HCAI UJ oraz grupa
GEIST kierowane przez
prof. Grzegorza J. Nalepę





Problemy z interpretacją AI

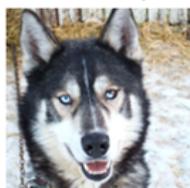
Jak AI „myśli”?



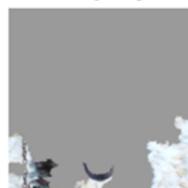
Nieintuicyjne cechy opisujące dane

```
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?



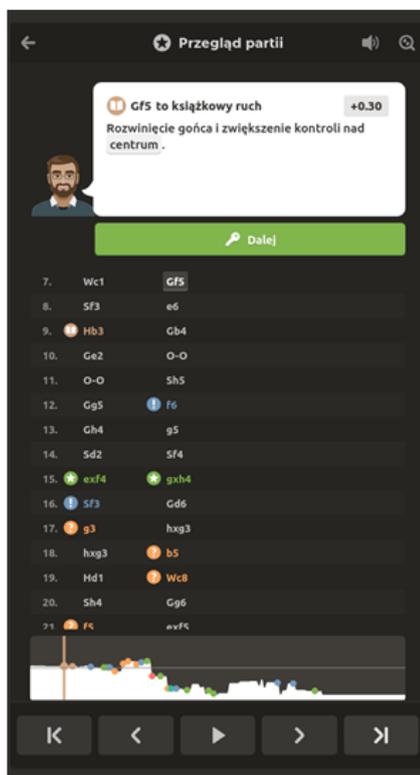
Czarne pudełko
(black box)
→ 10 mln wag → decyzja



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



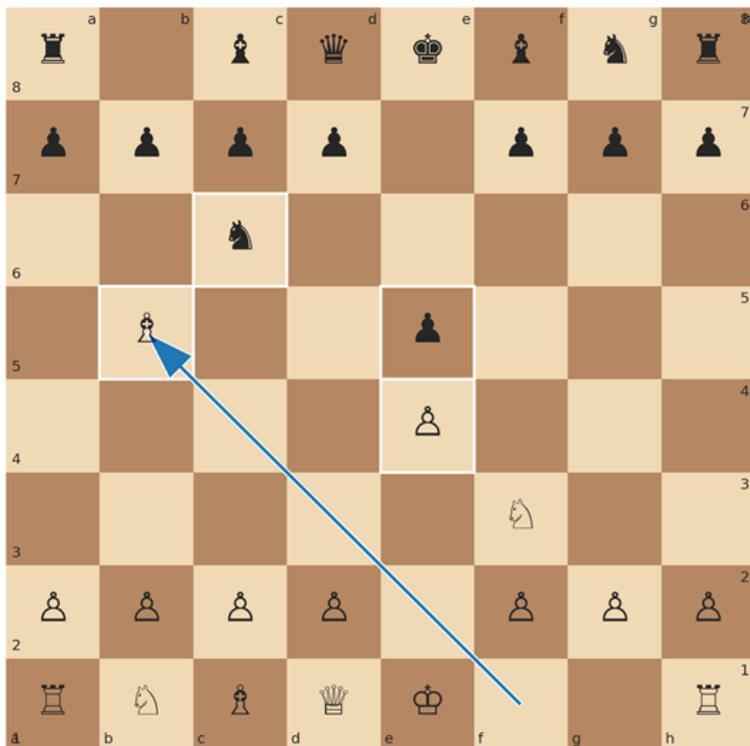
Tłumaczenie na język ludzi
→ AI widzi długoterminowy atak na króla





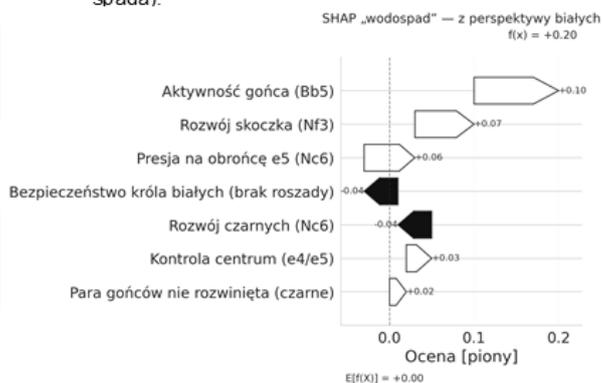
Krótko o eXplainable AI

Ruy Lopez (Hiszpańska) — po 3.Bb5



SHAP (Shapley Additive Explanations)

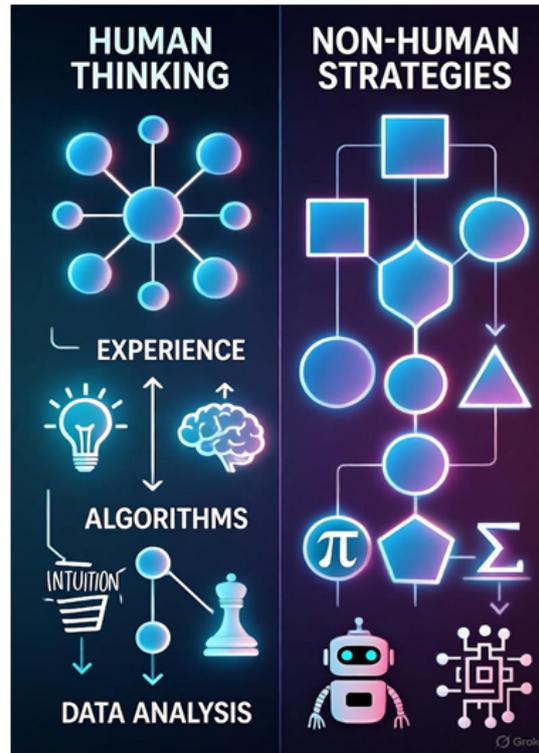
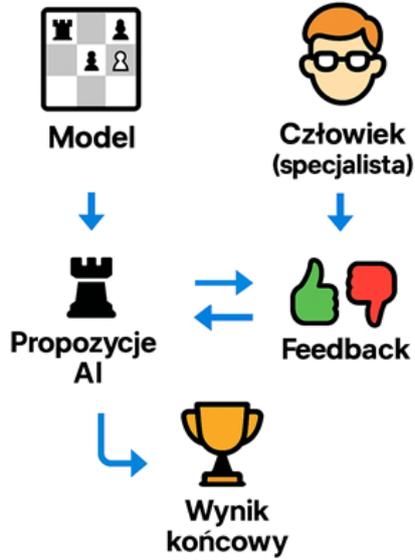
- O ile każda cecha podnosi lub obniża wynik modelu dla konkretnego przykładu.
- Wykorzystuje wartości **Shapleya** z teorii gier: „sprawiedliwie” dzieli wpływ między cechy, biorąc pod uwagę ich współdziałanie.
- Ma dwie kluczowe własności: **dokładność lokalną** (suma wkładów = wynik) i **spójność**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



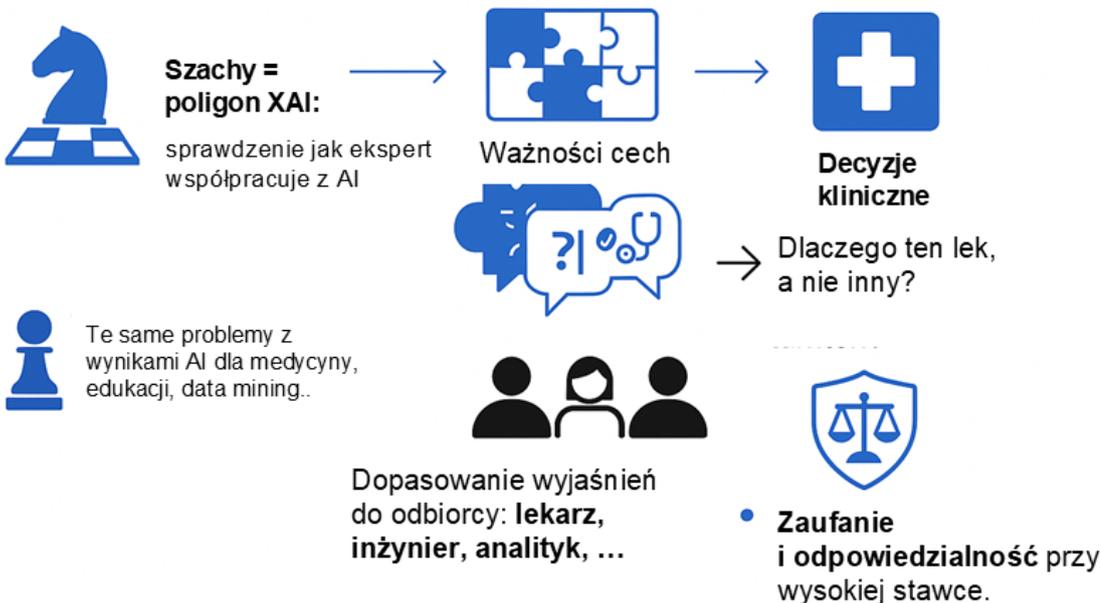
The screenshot shows a chess game in progress on a digital board. The board is oriented with white on top and black on bottom. The pieces are arranged as follows: White King on g1, White Rook on d1, White Rook on e1, White Pawn on a2, White Pawn on c3, White Pawn on f3, White Pawn on g3, White Pawn on h4, White Pawn on f4, White Pawn on g4, White Pawn on h4, White Pawn on f5, White Pawn on g5, White Pawn on h5, White Pawn on f6, White Pawn on g6, White Pawn on h6, White Pawn on f7, White Pawn on g7, White Pawn on h7, White Pawn on f8, White Pawn on g8, White Pawn on h8. Black King on g8, Black Rook on f8, Black Knight on e7, Black Bishop on d6, Black Bishop on f6, Black Pawn on a6, Black Pawn on b5, Black Pawn on c4, Black Pawn on d4, Black Pawn on e4, Black Pawn on f4, Black Pawn on g4, Black Pawn on h4, Black Pawn on f5, Black Pawn on g5, Black Pawn on h5, Black Pawn on f6, Black Pawn on g6, Black Pawn on h6, Black Pawn on f7, Black Pawn on g7, Black Pawn on h7, Black Pawn on f8, Black Pawn on g8, Black Pawn on h8.

On the right side, there is an AI explanation overlay titled "Przeгляд partii" (Review of the game). It features a green star icon and the text "Kh8 jest najlepsze" (Kh8 is the best) and "To jest najsilniejsza opcja." (This is the strongest option). A score of "-2.84" is shown in a black box. Below the text are buttons for "Pokaż kontynuację" (Show continuation), "Spróbuj ponownie" (Try again), and "Dalej" (Next). A list of moves is displayed 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, r4, Gh5, Hh6, Gxf4, Hg7+, Gg3, Gxf2, Kxg7, Gxf3, Wf6, Kg6, Kf7, h6, Kg6, and hvr4.

At the top left, there is a logo of a chess king with circuit lines. The title "Krótko o eXplainable AI" is displayed in the center. The bottom left corner has the text "Nauka i szachy" and the bottom right corner has the date "8.11.2025".

The image shows a screenshot of a chess game interface. On the left is a chessboard with pieces in their starting positions. On the right is a panel titled "Przegląd partii" (Game Review) showing a list of moves and a highlighted move explanation. The explanation text reads: "Gf5 to książkowy ruch. Rozwiniecie gońca i zwiększenie kontroli nad centrum." (Gf5 is a textbook move. Develop the knight and increase control over the center). Below the explanation is a "Dalej" (Next) button. The move list includes: 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. At the bottom left, there is a logo of a chess king with a neural network diagram and the text "nauka i szachy". At the bottom right, the date "8.11.2025" is displayed.

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=1762553996>

Last update: **2025/11/07 22:19**

