

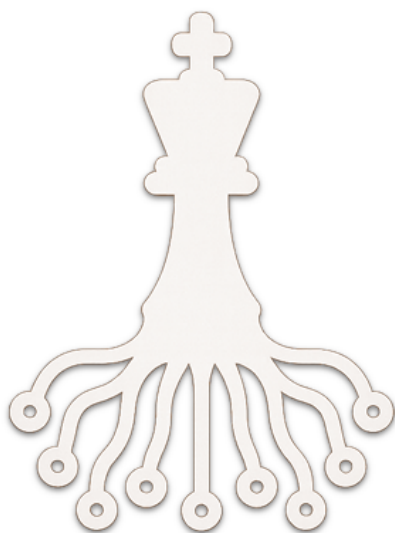
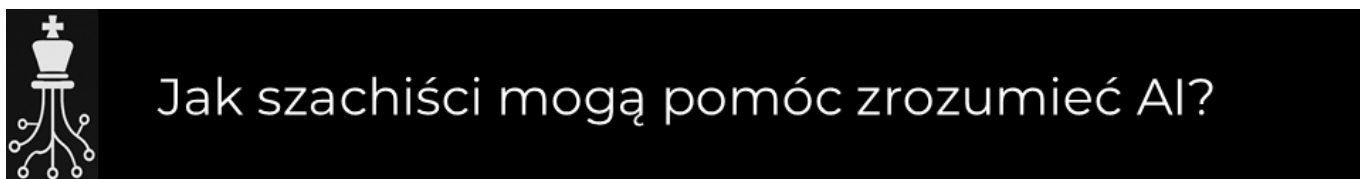
# Chess XAI study

## 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!

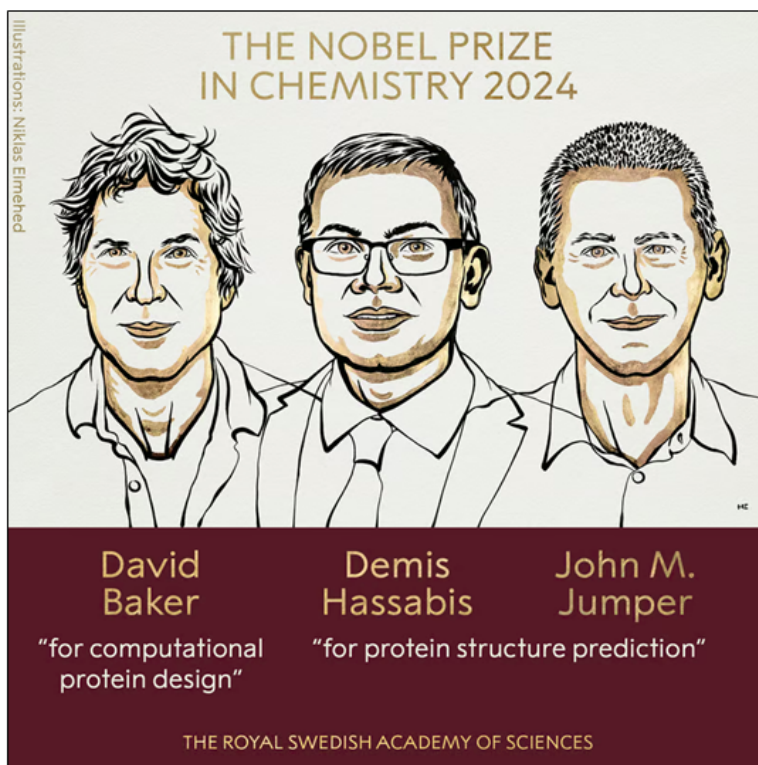


dr hab. Andrzej Siódmostok, prof. UJ  
Uniwersytet Jagielloński





# 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.”

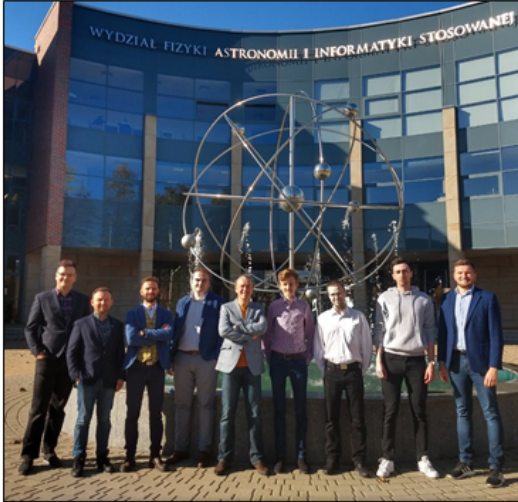


# AI przyjazne i użyteczne dla ludzi



Zakład Humanocentrycznej Sztucznej Inteligencji  
Uniwersytet Jagielloński

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



Maciej Szelązek



Maciej Mozolewski



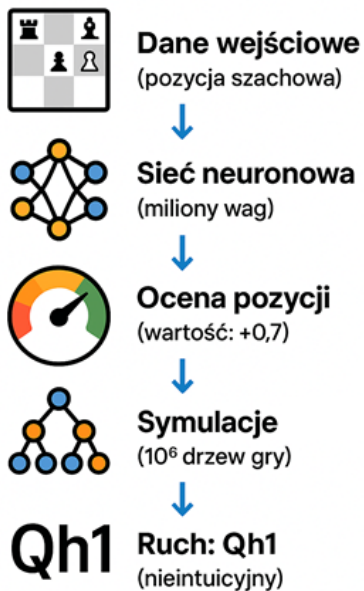
Zakład HCAI UJ oraz grupa  
GEIST kierowane przez  
prof. Grzegorza J. Nalepę





# Problemy z interpretacją AI

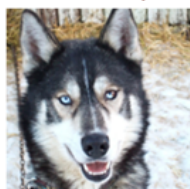
## Jak AI „myśli”?



## 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?



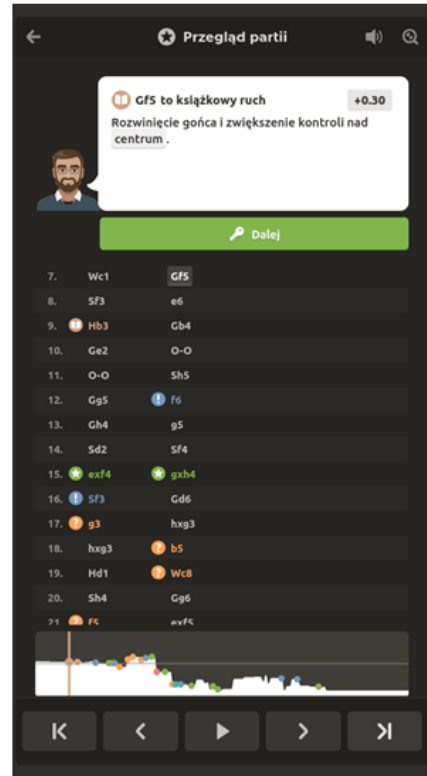
**Czarne pudełko**  
(black box)  
→ 10 mln wag → decyzja



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



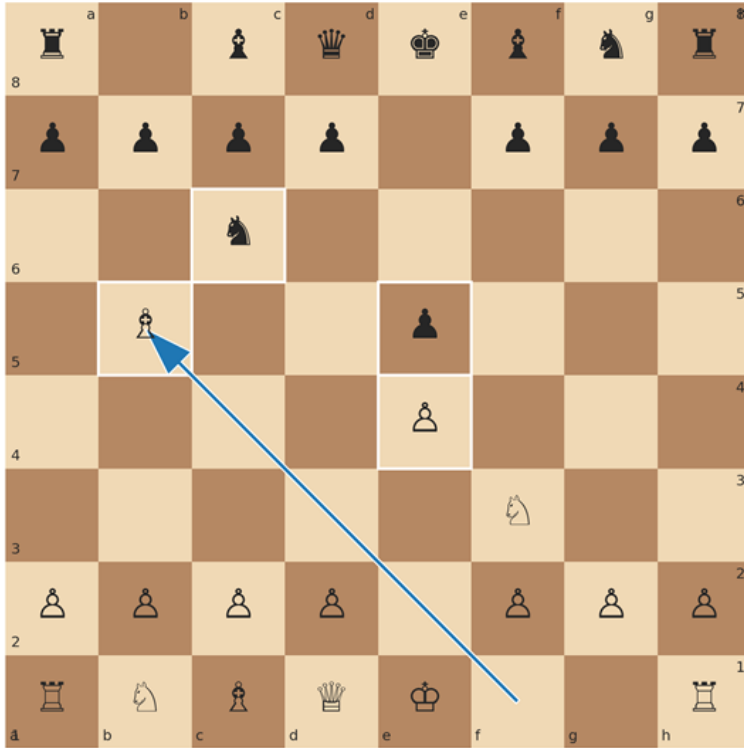
**Tłumaczenie na język ludzi**  
→ AI widzi długoterminowy atak na króla





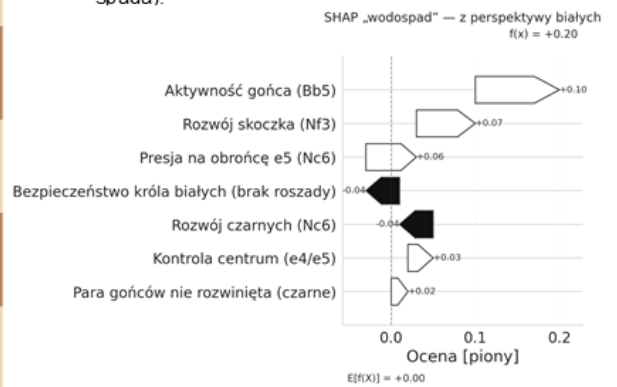
# Krótko o eXplainable AI

Ruy Lopez (Hiszpańska) — po 3.Bb5



## SHAP (Shapley Additive Explanations)

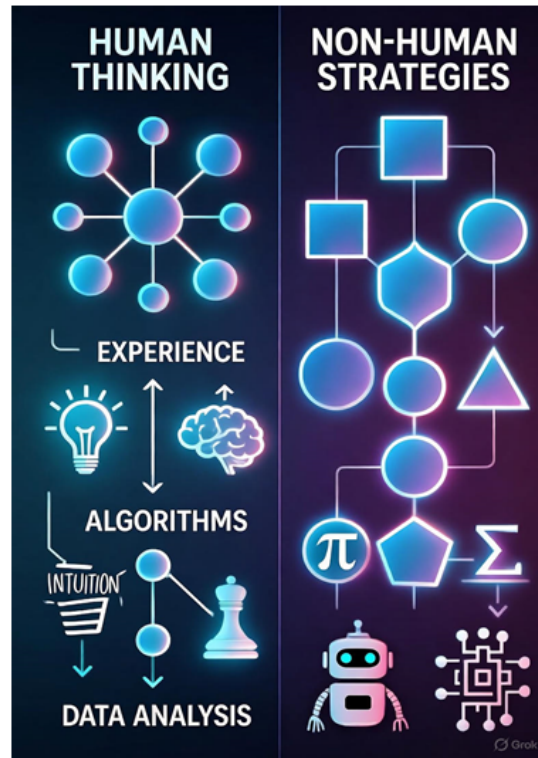
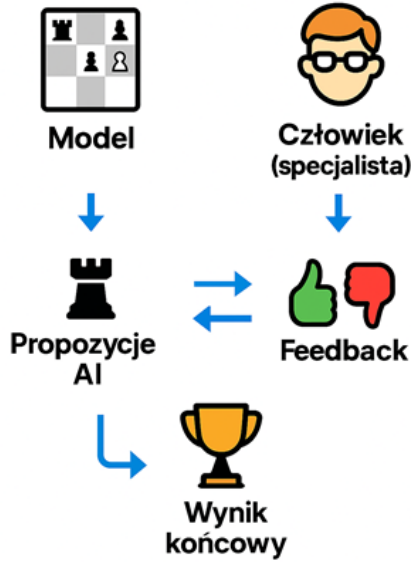
- O ile każda cecha podnosi lub obniża wynik modelu dla konkretnego przykładu.
- Wykorzystuje wartości **Shapleya** z teorii gier: „sprawiedliwie” dzieli wpływ między cechy, biorąc pod uwagę ich współdziałanie.
- Ma dwie kluczowe własności: **dokładność lokalną** (suma wkładów = wynik) i **spójność**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



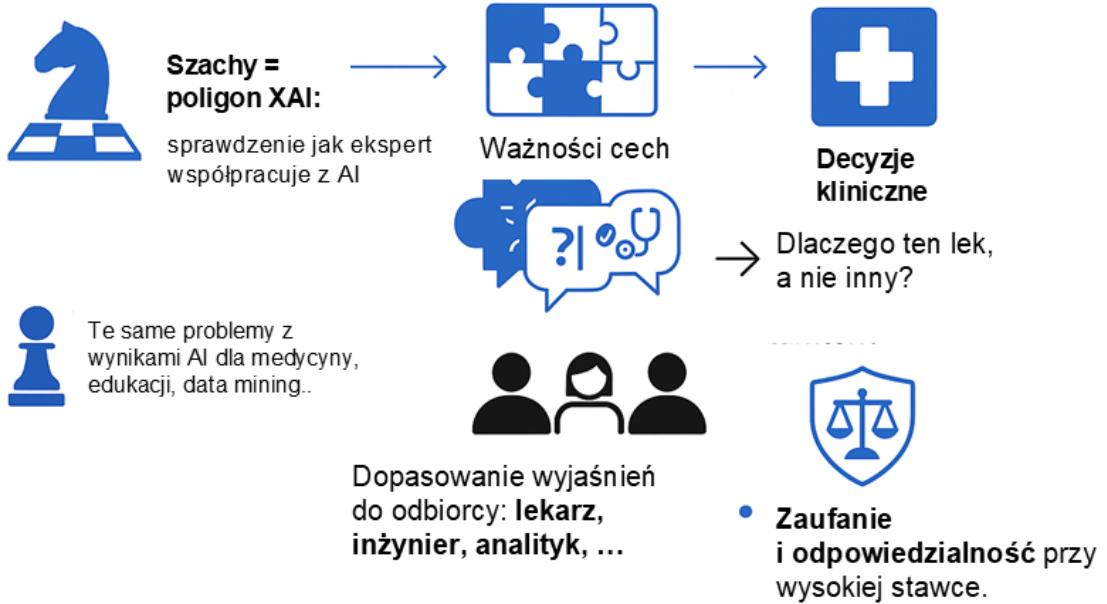
The image shows a screenshot of a chess game interface. On the left is a chessboard with a green and white checkered pattern. The board is labeled with ranks 1-8 and files a-h. A black king is on g8, a black knight on f7, a black pawn on g7, a black pawn on b6, a black knight on c6, a black king on d6, a black knight on e6, a black pawn on a5, a black pawn on b5, a black king on f5, a black knight on g5, a black pawn on d4, a black knight on e4, a black pawn on a3, a black pawn on c3, a black king on f3, a black knight on g3, a black pawn on a2, a black pawn on c2, a black king on f2, a black knight on g2, a black pawn on a1, a black pawn on c1, a black king on f1, a black knight on g1, a black pawn on a0, a black pawn on c0, a black king on f0, a black knight on g0. A green arrow points to the black king on g8, which has a star icon next to it. On the right is a panel titled "Przegląd partii" (Game Review). It contains a text box with a green plus icon and the text "Kh8 jest najlepsze To jest najsilniejsza opcja." followed by a score of "-2.84". Below this are buttons "Pokaż kontynuację" (Show continuation), "Spróbuj ponownie" (Try again), and "Dalej" (Next). Below the buttons is a list of moves: 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. At the bottom of the panel is a small bar chart showing the evaluation of different moves. The text "Naucz się szachy" (Learn chess) is visible at the bottom left, and "8.11.2025" is at the bottom right.

# 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 this text is a green button labeled "Dalej". A list of moves is shown below the explanation, with the 21st move highlighted: "21. f3 e3f3". 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. At the bottom left, the text "Nauka i szachy" is visible, and at the bottom right, the date "8.11.2025" is displayed.



# 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=1762554086>

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

