

Inżynieria Mechatroniczna

Podstawy Sztucznej Inteligencji i Ucznienia Głębokiego:

## 14: Problemy etyczne oraz Przyszłość AI

Ziemowit Dworakowski  
AGH w Krakowie

1

---

---

---

---

---

---

---

---

Proste rozwiązania do rozważenia w sytuacji problemów etycznych: SD



1. **Konflikt interesów, nacechowanie emocjonalne:**  
Wycofaj się, przekaz problem osobie obiektywnej



2. **Konflikty wartości:** Rozważ możliwość fałszywej dychotomii. Być może rozwiązanie kompromisowe jest możliwe?



3. Jeśli nie możesz rozwiązać problemu sam, **eskaluj!**  
(to znaczy: przekaz problem na wyższy szczebel hierarchii)



4. Znaj swoje **wartości** i wiedz jak się **efektywnie komunikować**



5. Staraj się spojrzeć na problem w całości, rozważ różne opcje  
(Wiele konfliktów jest podtrzymywanych tylko dlatego, bo osoby zaangażowane nie zauważają możliwych rozwiązań)



6. **Dokumentuj swoje obawy i decyzje**, bądź transparentny

2

---

---

---

---

---

---

---

---

*„Come gather 'round people wherever you roam  
And admit that the waters around you have grown  
And accept it that soon you'll be drenched to the bone  
If your time to you is worth savin'  
Then you better start swimmin' or you'll sink like a stone  
For the times they are a-changin'...”*

- Bob Dylan

**Krajobraz AI, etyka i perspektywy na przyszłość**  
(Ta część wykładu może być mocno subiektywna...)

3

---

---

---

---

---

---

---

---

## Etyka AI: wyjaśnialność decyzji AI

SD

- **Sama dokładność to nie wszystko – potrzebujemy wyjaśnialnych systemów AI**  
(choć to duże wyzwanie...)
- **Prawdziwa wyjaśnialność jest lepsza niż racjonalizacja post-factum**  
(ale dużo trudniejsza do uzyskania...)
- **Powinniśmy rozwijać narzędzia audytu i weryfikacji wszystkich systemów AI**  
(ale „popatrzenie na strukturę” zwykle nie mówi wiele – no i mamy ryzyko naruszenia praw majątkowych...)

4

---

---

---

---

---

---

---

---

## Etyka AI: prywatność i zgoda (consent)

SD

Użytkownik wyraża zgodę na monitorowanie akcji serca by wykrywać stany ryzykowne. Później, dane wykorzystywane są do śledzenia reakcji emocjonalnych na reklamy



**Wniosek:** Prywatność jest kontekstowa i zależy od przypadku użycia.

Platforma społecznościowa zaczyna używać danych użytkowników do trenowania narzędzi GenAI. Zmiany są powolne, iteracyjne. Nie ma opcji „opt-out” z wyjątkiem odinstalowania aplikacji.

**Wniosek:** Opcja „Opt-out” powinna być rzeczywista i darmowa. Użytkownik nie powinien być szantażowany „rezygnacją z aplikacji”.

Platforma medialna proponuje krótkie filmy w celu podniesienia satysfakcji użytkownika. Zawartość filmów wpływa na użytkownika kształtując jego światopogląd – algorytmy mogą nieświadomie dla nas wzmacniać konkretne działania i zachowania (wybory życiowe, mody, poglądy polityczne, itd.)



**Wniosek:** Wszystkie systemy wykorzystujące AI mogą wpływać na nasze wybory.

5

---

---

---

---

---

---

---

---

## Etyka AI: prywatność i zgoda (consent)

SD

System medyczny śledzi nasze dane i przechowuje je do poprawy modelu w anonimowy sposób. Potem te dane (anonimowe) mogą być przekazane innej firmie (np. ubezpieczeniowej) a ta może dokonać ich reidentyfikacji – w celu rekomendowania ceny ubezpieczenia zdrowotnego...



**Wniosek:** Dane mogą wyciec, zostać zreidentyfikowane, skorelowane i wykorzystane przez innych użytkowników...

„[...] By utilizing the Application, the User hereby irrevocably consents to the potential aggregation, analysis, redistribution, and adaptive repurposing of selected usage-derived indicators, insofar as such processing is reasonably construed to enhance systemic operability or third-party relevance, whether now foreseeable or arising through subsequent technological development. [...]”



**Wniosek:** Umowy dotyczące wykorzystania danych powinny być krótkie i zrozumiałe...

Wszystkie powyższe przykłady zostały przez kogoś zaprojektowane i zaakceptowane... być może przez inżyniera mechatronika...

**Wniosek:** To WY macie wpływ na takie sytuacje. Naszą odpowiedzialnością jest budowa systemów zgodnych z naszymi wartościami.

6

---

---

---

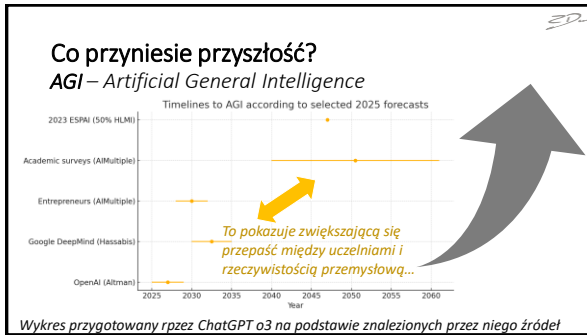
---

---

---

---

---



7

---

---

---

---

---

---

---

---

**Prawa AI – co, jeśli system AI stwierdzi, że jest żywy?**

- Co to znaczy „być świadomym”?
- Czy można wyobrazić sobie inteligencję bez świadomości?
- „Chiński pokój” jako eksperyment myślowy
- Przykład LamDA

*Tak naprawdę nie wiemy czym jest świadomość...  
 Wiele wskazuje na to, że to zjawisko emergentne  
 – zdarza się jeśli warunki są odpowiednie*

8

---

---

---

---

---

---

---

---

**„AI tylko udaje, że myśli...”**

„To tylko zgadywanie następnego słowa. Tak, mamy warstwową strukturę, samoorganizację, wnioskowanie, ale każdy z tych kroków to tylko przewidywanie kolejnych i kolejnych słów... Nie ma tam rzeczywistej świadomości, wiedzy, zrozumienia, intencji...

...ale jeśli udajesz umiejętność wystarczająco dobrze...  
 ...to ją posiadasz.

Aktualnie LLMy są już nieodróżnialne od ludzi na podstawie konwersacji.  
 Gdyby nie prompty systemowe, nie byłoby w stanie udowodnić, że nie są świadome

Są wciąż zadania (logika formalna, rozumowanie przez analogie, nawigacja przestrzenna) w których ludzie są znacznie lepsi...

...ale jak długo?

9

---

---

---

---

---

---

---

---



10

---

---

---

---

---

---

---

---



11

---

---

---

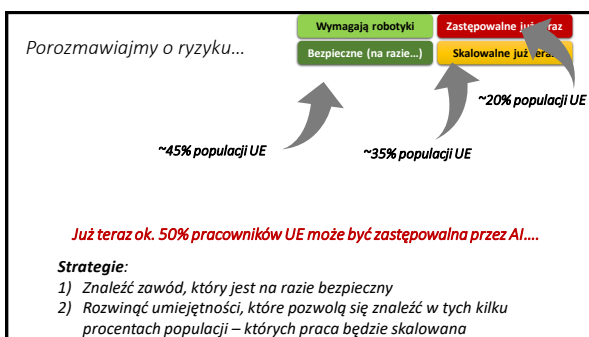
---

---

---

---

---



12

---

---

---

---

---

---

---

---

### Porozmawiajmy o ryzyku...

Coraz więcej zadań jest „robionych lepiej” przez AI  
(potencjalnie z większością „kreatywnych” zagrożonych jeszcze przed 2030 rokiem)

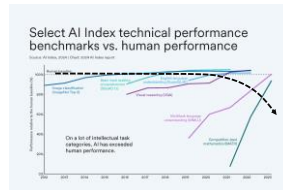
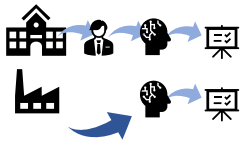
**Bezrobocie znacznie wzrasta, Gwarantowany Dochód  
Podstawowy staje się koniecznością...**

**„Czemu robić cokolwiek, skoro ChatGPT robi to lepiej?”  
Ludzie tracą decyzyjność, kreatywność,  
odpowiedzialność, zdolność do wykonywania  
skomplikowanych zadań...**

13

### Jak używać AI...

...bez ograniczenia własnych zdolności?



*Chcesz pracować efektywniej – więc używasz AI. Ale wtedy przestajesz umieć  
wykonywać zadania samodzielnie... I można Cię jeszcze łatwiej zastąpić przez AI...*

14

### Jak ja używam AI...

...bez ograniczenia własnych zdolności?



1) Mam świadomość swoich celów i staram się świadomie wybierać które zadania robię sam a które deleguję AI. Wiem, że w trakcie gdy ja+AI będziemy w nich coraz lepsi, ja będę coraz gorszy...



2) Jestem współpracownikiem AI a nie jedynie managerem delegującym zadania. Szukuję rozwiązania, kwestionuję i audytuję rezultaty pracy AI, gram adwokata diabła lub zmuszam AI do robienia tego przeciw mojej argumentacji. Różnicuję role które odgrywamy. Upewniam się, że rozumiem „wyjście” i to ja świadomie akceptuję lub odrzucam efekty pracy AI.



3) Świadomie wybieram hobby i zainteresowania w kierunku poprawy swoich umiejętności intelektualnych, rozwiązywania problemów, pamięci i kreatywności i NIE używam w nich AI.



4) Używam aktywnych technik uczenia się, by budować swój „System 1” (intuicję) oraz „System 2” (myślenie „wolne”).

15

## Rekomendacje (z kultury popularnej)

- 1) „Ślepowidzenie” oraz „Echoprakcja” by Petera Watts  
*Świetne powieści SF poświęcone roli świadomości i decyzyjności ludzi w świecie z rozwiniętą AI*
- 2) „Po Piśmie” Jacka Dukaja  
*Zbiór esejów o przyszłości w której praca jest całkowicie opcjonalna*
- 3) *Ex Machina* – film z 2014 roku o interakcji między AI i człowiekiem oraz mitach i niezrozumieniu celów działania
- 4) „What everyone gets wrong about AI and education” Derek Muller (Veritasium)  
Wykład o roli uczenia się: [https://www.youtube.com/results?search\\_query=derek+muller+education](https://www.youtube.com/results?search_query=derek+muller+education)

16

## Informacja zwrotna...!

*Niezależnie od wyników naszej aktualnej dyskusji – skorzystajcie proszę z ankiet studenckich na koniec semestru. Zróbcie to dla jak największej ilości przedmiotów i nauczycieli. One mają dużo większe znaczenie niż myślicie :)*

17