



**CENTRUM ETYKI
TECHNOLOGII
INSTYTUTU HUMANITES**

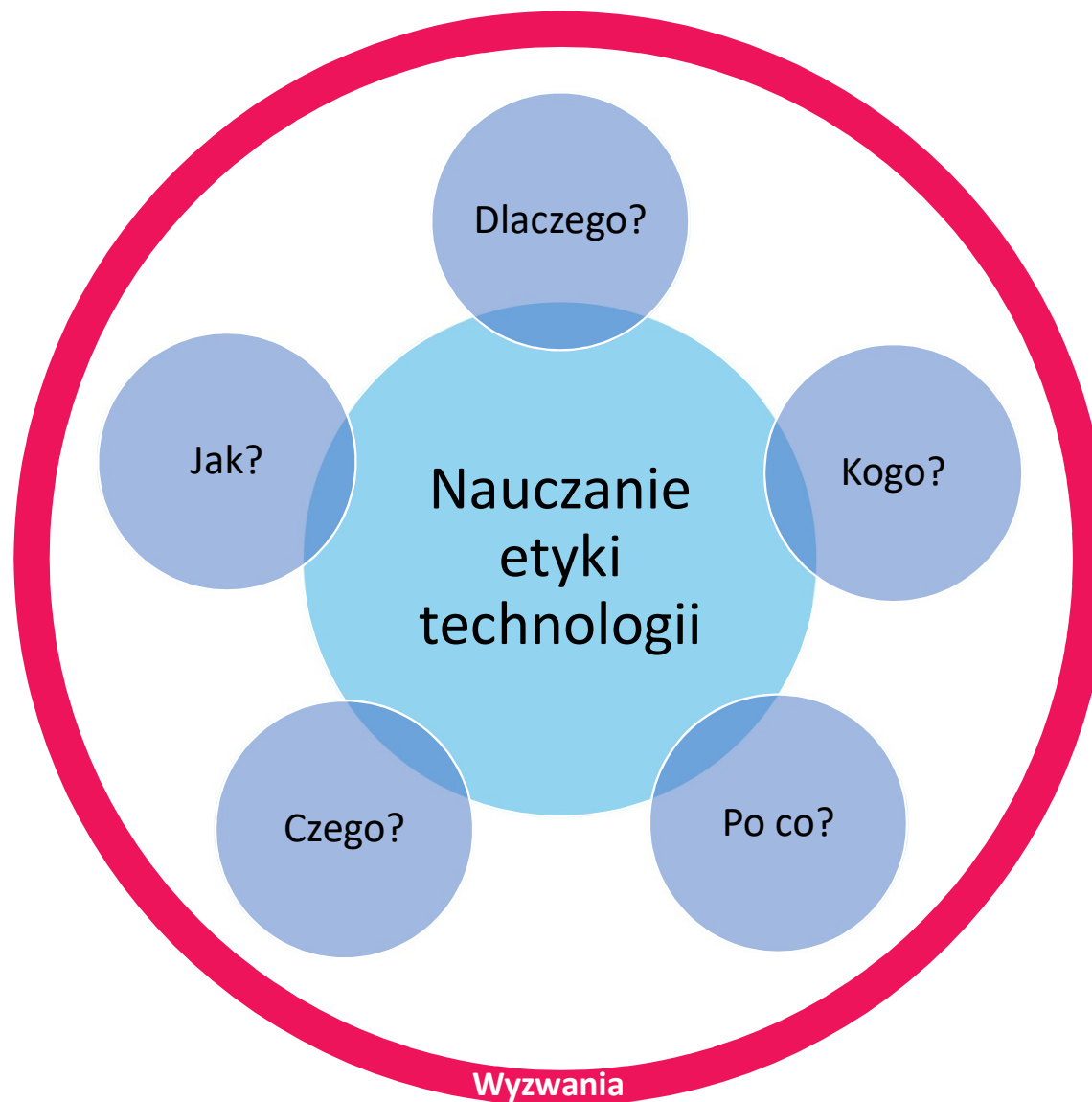
Etyka technologii w programie kształcenia akademickiego doby AI

Maciej Chojnowski

Centrum Etyki Technologii Instytutu Humanites

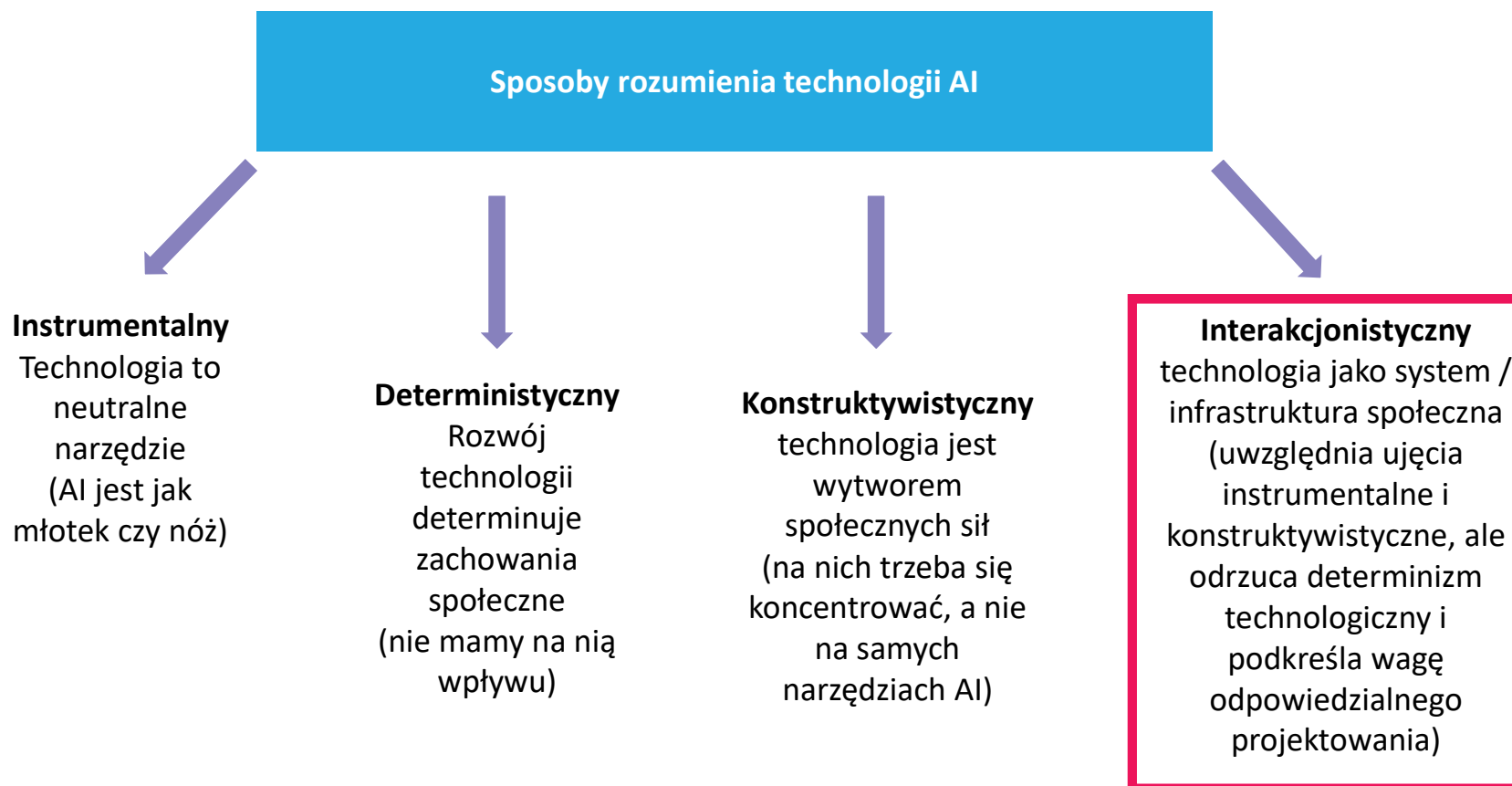
VI Akademickie Forum Jakości

15 października 2024



Dlaczego? (1)

Aby wskazać powody, dla których trzeba dziś uczyć etyki technologii, musimy wpiery określić **rolę technologii AI** we współczesnym świecie.



Podjęcie interakcjonistyczne pozwala zidentyfikować **różne rodzaje problemów związanych z AI**. To zarazem **powody**, dla których trzeba troszczyć się o jej **odpowiedzialny rozwój**:



Problemy związane z uczeniem maszynowym, np.:

- Dyskryminacja wynikająca ze skrzywienia algorytmicznego
- Nieprzejrzystość działania systemów uczenia głębokiego i problem z weryfikacją wyników
- Zagrożenie dla prywatności w efekcie korelacji pozornie neutralnych danych

Problemy związane z funkcjonowaniem w cyfrowym świecie, np.:

- Malejące poczucie sprawczości
- Ryzyko bezrobocia technologicznego
- Koncentracja władzy i kapitału przez monopolistów
- Cyfrowa kolonizacja
- Inwigilacja i manipulacja
- Kwantyfikacja rzeczywistości
- Rosnący ślad węglowy związany z rozwojem infrastruktury AI
- Destrukcyjny wpływ na demokrację

Problemy metafizyczne dot. statusu AI i natury człowieka, np.:

- Transhumanizm, *human-enhancement* i pytanie o naturę człowieka
- Możliwość stworzenia tzw. ogólnej sztucznej inteligencji (AGI – Artificial General Intelligence) i pytanie o jej status ontyczny i prawny
- Ryzyko powstania superinteligencji

Dlaczego etyka technologii, a nie samo prawo?

Luciano Floridi: etyka i prawo to kluczowe komponenty systemu normatywnego dot. technologii

- ➔ prawo określa, co jest dozwolone, a co zakazane
- ➔ etyka podpowiada, co robić, abyśmy mogli żyć w lepszym świecie
- ➔ etyka technologii działa jak **moralna busola** wskazująca pożądany kierunek rozwoju technologii



Kogo? (1)

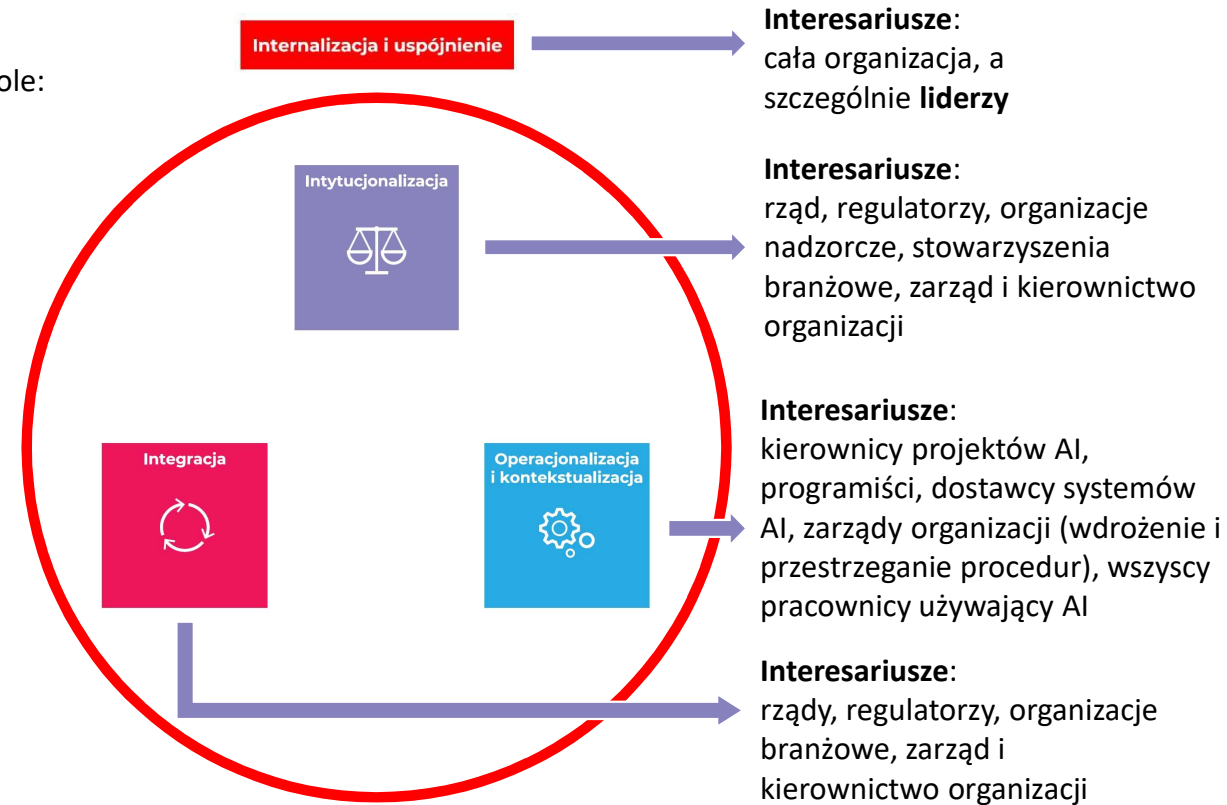
Aby odpowiedzieć na to pytanie, trzeba – pamiętając o podejściu interakcjonistycznym – określić wpieryw:

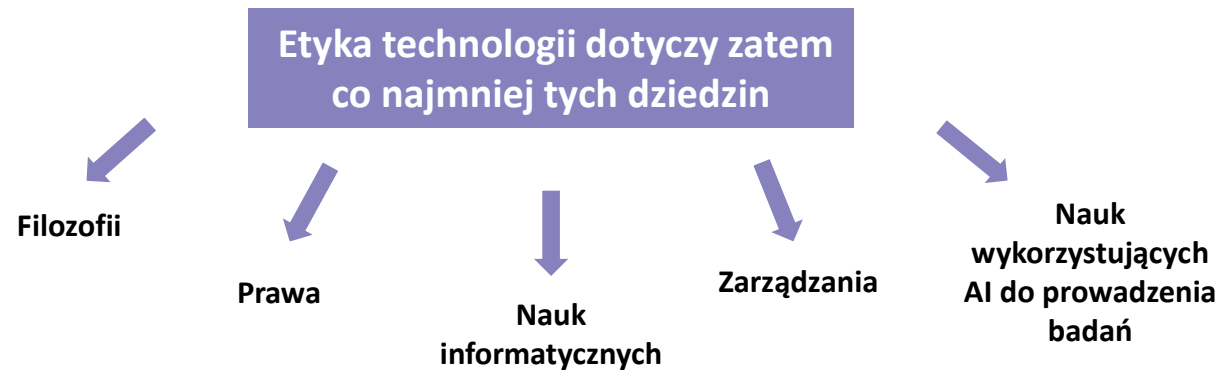
- ➔ **obszary AI governance**
- ➔ **odpowiadających im interesariuszy**

Pojęcie **AI governance** odnosi się do wszystkich osób, instytucji, procesów i narzędzi, które służą zarządzaniu sztuczną inteligencją. Obejmuje różne poziomy: międzynarodowy, krajowy, branżowy, organizacyjny, zespołowy oraz indywidualny.

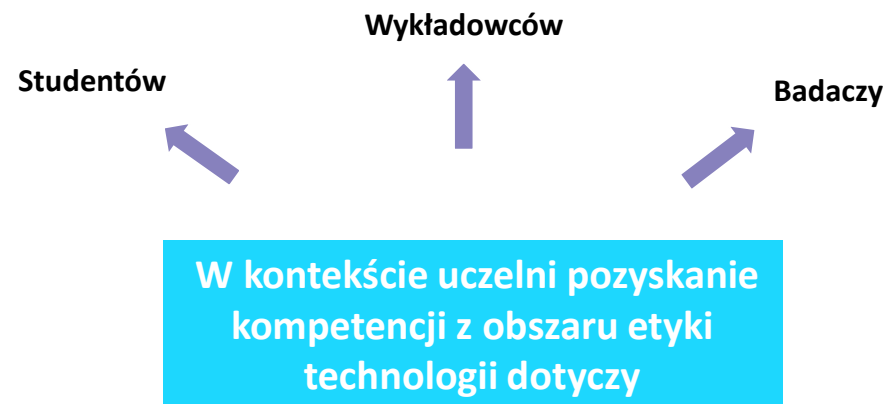
Rozwijane w CET Humanites podejście do AI governance obejmuje **cztery kluczowe wymiary** i pozwala uporządkować **interesariuszy** i ich role:

- **Institucjonalizację** – chodzi o ramy normatywne na różnych poziomach: międzynarodowym, krajowym, sektorowym bądź organizacyjnym. Mogą one mieć postać odgórnej regulacji (jak np. rozporządzenie AI Act) lub samoregulacji (czyli np. kodeksy dobrych praktyk lub kodeksy etyki AI).
- **Operacjonalizację i kontekstualizację** – chodzi o realizację zasad i wymogów godnej zaufania AI w poszczególnych projektach i odpowiednio do danego kontekstu. Odbywa się z wykorzystaniem konkretnych metod, standardów i narzędzi, jak np. podejście Ethics by Design czy normy ISO.
- **Integrację** – chodzi o łączenie działań z obszaru trustworthy AI z innymi inicjatywami, dotyczącymi np. cyberbezpieczeństwa czy dobrostanu ludzi i przyrody.
- **Internalizację i uspołnienie** – chodzi o przejście od czysto formalnego myślenia o procedurach do autentycznego przyswojenia deklarowanych wartości.





Wniosek: niezbędne jest wielowymiarowe podejście do kursów z etyki technologii: na różnych kierunkach, poziomach kształcenia i dla różnych grup odbiorców



Po co?

Cele nauczania etyki technologii

Studenci mogą nauczyć się myśleć nie tylko o tym, jaką technologię są w stanie stworzyć, ale także czy powinni ją tworzyć.

Prof. Barbara J. Grosz, amerykańska informatyczka i współtwórczyni projektu Embedded EthiCS na Uniwersytecie Harvarda

Zmiana nastawienia:

Rozwijanie moralnej wrażliwości w kontekście technologii – postrzeganie kwestii etycznych jako priorytetowych przy projektowaniu technologii



Zdobycie podstawowych kompetencji:

Zdolność do rozumowania, analizy i argumentacji w kategoriach etycznych – znajomość głównych teorii etyki, podstaw rozumowania etycznego, świadomość występowania dylematów



Rozwinięcie umiejętności praktycznych:

Pozyskiwanie umiejętności rozpoznawania problemów etycznych i reagowania na nie w kontekście technologii – dostrzeganie kontekstów, różnych interesariuszy i możliwych konfliktów między wartościami, umiejętność właściwego reagowania



Praktykowanie „etosu programistycznego”:

Traktowanie kompetencji (cnót) etycznych jako „skili”

– odejście od postrzegania wymogów etyki (także w projektach badawczych realizowanych na uczelniach) w kategoriach czysto formalnego compliance



Zmiana praktyki projektowania technologii:

Rozpoznawanie i rozwiązywanie problemów etycznych na etapie projektowania technologii – proaktywne podejście do etyki technologii, rozwijanie projektów IT zgodnie z najlepszymi praktykami i z uwzględnieniem odpowiednich metod, jak np. Ethics by Design

Czego?

Aby odpowiedzieć na to pytanie, trzeba:

- pamiętać, jakim celem ma służyć nauczanie etyki technologii
- uwzględnić różne wymiary / sensory etyki technologii:



Przygotowując program nauczania etyki technologii, trzeba uwzględnić wszystkie powyższe wymiary, ponieważ:

- sama teoria to zbyt abstrakcyjna
- same kodeksy to powierzchowność i formalizm
- sama wrażliwość bez zaplecza teoretycznego i kodeksów to ignorancja
- samo rozpoznawanie problemów bez umiejętności ich praktycznego rozwiązywania w konkretnym kontekście zawodowym to niekompetencja
- sama sprawność „techniczna” bez zaangażowania to prymat proceduralności i zubożenie etyki o wymiar etosowy

Ważnym kontekstem jest tutaj także wymiar polityczny – wiele problemów z obszaru etyki technologii dotyczy zbiorowości, oddziaływania społecznego oraz instytucji.

Możliwe ujęcia:

etyka – modelowanie indywidualnego postępowania

etyka – dobro, słusność, sprawiedliwość

vs.

filozofia polityczna – modelowanie organizacji społeczeństwa i jego instytucji

filozofia polityczna – legalność, porządek, stabilność

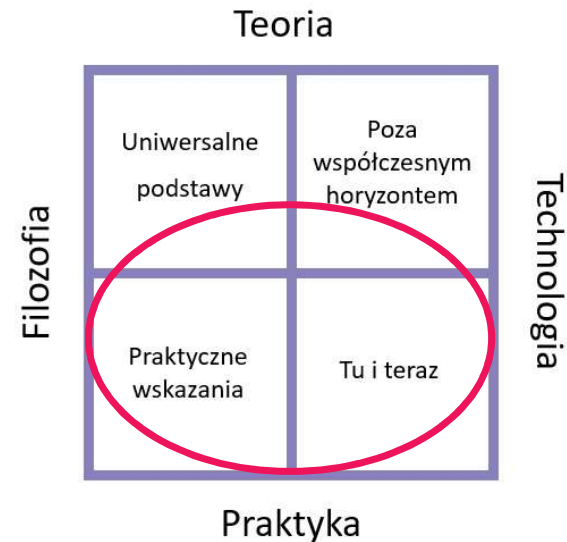
Jak?

Treść – wskazania dotyczące zawartości programu

Dwie **osie równowagi**:

filozofia ↔ technologia

teoria (uniwersalność) ↔ praktyka (aktualność)



Osie równowagi na podstawie: L. Tuovinen, A. Rohunen, *Teaching AI Ethics to Engineering Students: Reflections on Syllabus Design and Teaching Methods*

Forma – wskazania dotyczące konstrukcji kursu

Jak to robią na **Harvardzie**?



Podejście **Embedded EthiCS**



zamiast osobnego przedmiotu „osadzenie” etyki technologii w DNA programu na studiach IT



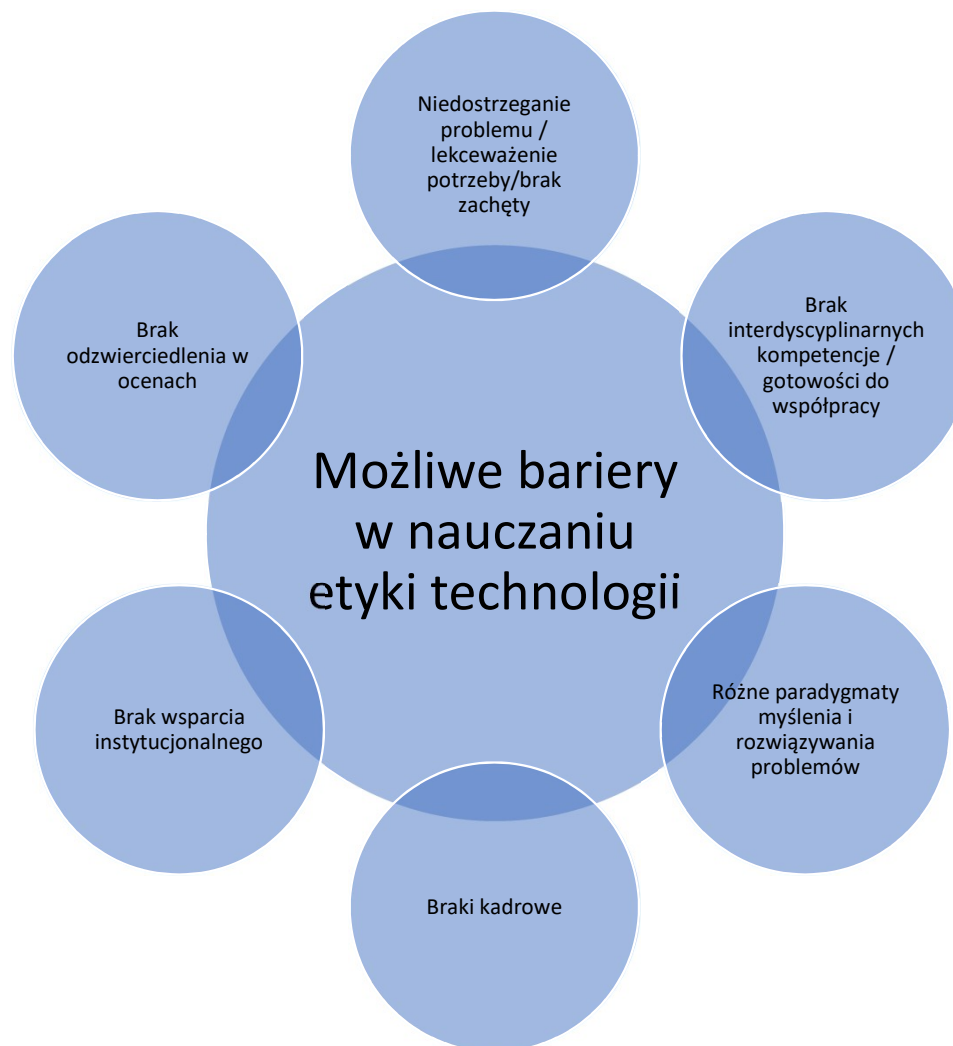
podejście modułowe



rozprowadzone / rozproszone (*distributed*)



ze wzmocnieniem (*reinforced*)





CENTRUM ETYKI
TECHNOLOGII
INSTYTUTU HUMANITES

PIERWSZE W JĘZYKU POLSKIM WPROWADZENIE DO ETYKI SZTUCZNEJ INTELIGENCJI



ZESKANUJ KOD QR
POBIERZ DARMOWY
BOOKLET



CENTRUM ETYKI
TECHNOLOGII
INSTYTUTU HUMANITES

Dziękuję za uwagę
i
zapraszam do współpracy!

Maciej Chojnowski

maciej.chojnowski@humanites.pl