



POLSKIE TOWARZYSTWO INFORMATYCZNE

# Etyczne aspekty tworzenia, wdrażania i wykorzystywania systemów sztucznej inteligencji: Od teorii do praktyki.

AWSI



- Sekcja Aktualnych Wyzwań Sztucznej Inteligencji jest częścią **Polskiego Towarzystwa Informatycznego**
- Sekcję tworzą ludzie o **zróżnicowanej wiedzy i doświadczeniu**, w tym inżynierowie, etycy, prawnicy czy przedstawiciele innych nauk społecznych
- Jedną z misji AWSI jest **promowanie etycznego podejścia do sztucznej inteligencji**
- Zarząd AWSI tworzą:
  - dr **Michał Nowakowski**, przewodniczący
  - prof. **Maria Ganzha**, zastępca przewodniczącego i prezes oddziału mazowieckiego PTI
  - prof. **Przemysław Biecek**, członek zarządu

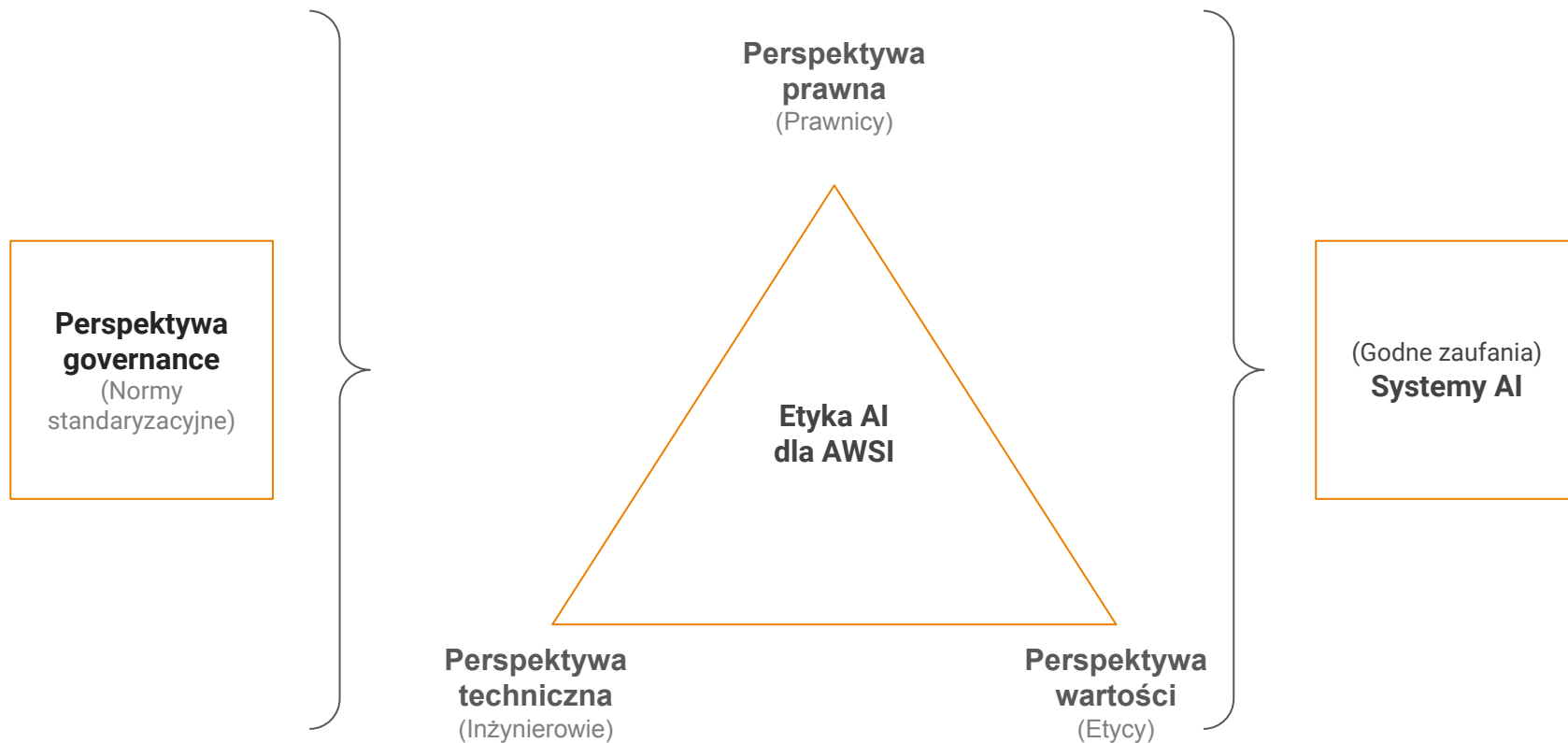
## Czym jest Etyka AI?

Etyka sztucznej inteligencji to **dynamiczne, multidyscyplinarne i zintegrowane** podejście do projektowania, rozwijania, wdrażania i wykorzystania systemów AI, których celem jest zwiększenie dobrostanu społecznego.

## Dlaczego Etyka AI jest ważna?

- Systemy AI etycznie zaprojektowane i wdrożone mogą znacząco **zwiększyć dobro społeczne**
- Systemy AI źle zaprojektowane i wdrożone mogą równie skutecznie z wielokrotności szkody społeczne
- Sama zgodność z prawem i standardami technicznymi **nie gwarantuje**, że systemy AI przyniosą korzyści społeczne
- Celem AWSI jest promowanie **wyważonego podejścia** wspierającego korzyści społeczne systemów AI przy jednoczesnym ograniczaniu potencjalnych szkód

# Podójście do etyki z perspektywy multidyscyplinarności

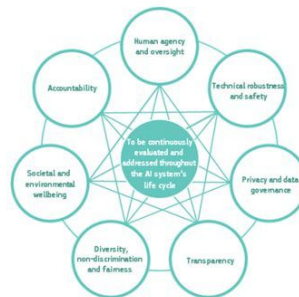


# Ramy etyczne AI (AI Ethics Frameworks)

## Do czego służą ramy etyczne?

- Etyczne ramy dla AI ustanawiają **wspólny język i sposób myślenia** o sztucznej inteligencji
- Pozwalają **zwiększyć świadomość** oraz zapewnić **spójność pojęciową**, budując porozumienie wśród różnych interesariuszy
- Umożliwiają **operacjonalizację** etyki poprzez monitorowanie realizacji wskaźników

## Wytyczne Komisji Europejskiej - AI Godna Zaufania



## IEEE AI Ethics Standards



## OECD Zasady AI oparte na wartościach

Values-based principles	Recommendations for policy makers
Inclusive growth, sustainable development and well-being >	Investing in AI R&D >
Human-centred values and fairness >	Fostering a digital ecosystem for AI >
Transparency and explainability >	Providing an enabling policy environment for AI >
Robustness, security and safety >	Building human capacity and preparing for labour market transition >
Accountability >	International co-operation for trustworthy AI >

... dodatkowo istnieje ponad 200 różnych ram dla etycznej AI



## Dlaczego Wytyczne Komisji Europejskiej?

- Wytyczne etyki godnej zaufania są obiecującymi ramami dla rozwoju AI, wypracowanymi w sposób **multidyscyplinarny**, łączącymi perspektywy prawne, techniczne i etyczne
- Zapewniają one **kompleksowe podejście** do wyzwań związanych z AI
- Stanowią one jedną z etycznych podstaw regulacji AI w UE (AI Act), zwiększając ich znaczenie i **skuteczność w praktyce**



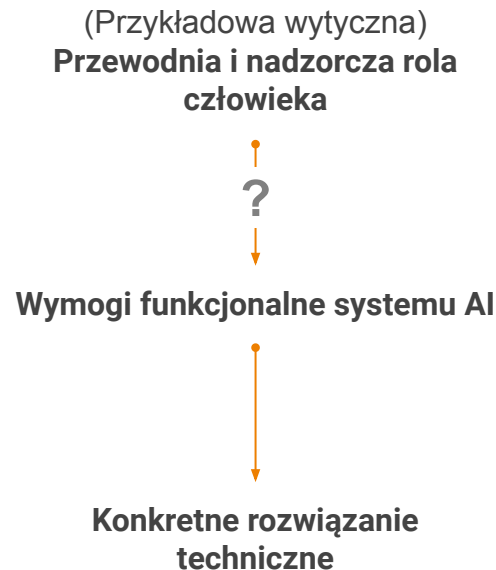
## Wytyczne:

1. Przewodnia i nadzorczą rolę człowieka
2. Techniczna solidność i bezpieczeństwo
3. Ochrona prywatności i zarządzanie danymi
4. Przejrzystość
5. Różnorodność, niedyskryminacja i sprawiedliwość
6. Dobrostan społeczny i środowiskowy
7. Odpowiedzialność

# Operacjonalizacja wytycznych etycznych

## Na czym polega trudność w przetłumaczeniu wytycznych na decyzje projektowe w systemach AI

- Wytyczne etyczne są ogólne i **abstrakcyjne**, co utrudnia ich bezpośrednie zastosowanie w praktyce projektowej
- **Interpretacja** wytycznych wymaga zrozumienia aspektów prawnych, etycznych i technicznych AI ze szczególnym uwzględnieniem konfliktów wartości
- Proces implementacji wytycznych etycznych jest skomplikowany i wymaga **współpracy** specjalistów z różnych dziedzin



# Proces projektowy technologii AI oparty na wartościach



## Technologie AI są wynikiem ludzkiej **wyobraźni**

Technologie AI nie powstają w próżni; są tworzone przez ludzi, którzy wykorzystują swoją kreatywność, wiedzę i wyobraźnię do projektowania i budowania nowych rozwiązań.

## Technologie AI są tworzone w procesie **projektowania**

Każda technologia jest wynikiem projektowania, które celowo łączy wiedzę w celu rozwiązywania problemów i zwiększania ludzkich możliwości, kształtując jej funkcjonalność, interakcję z użytkownikiem oraz efektywność.

## Projektowanie polega na **wyborach** spośród możliwych **opcji**

Każde projektowanie wiąże się z dokonywaniem wyborów spośród możliwych opcji, uwzględniając zarówno czynniki techniczne, jak i etyczne oraz prawne.

## Wybory projektowe odzwierciedlają **wartości**

Każdy wybór odzwierciedla wartości, co oznacza, że decyzje projektowe mają etyczne konsekwencje i powinny uwzględniać zasady etyczne, szacunek dla wartości ludzkich oraz pozytywny wpływ na społeczeństwo.

## Technologie AI są wynikiem **wyborów** projektantów

Technologie AI odzwierciedlają i wpływają na ludzkie wartości, ponieważ ich design i funkcje są kształtowane przez wartości ich twórców.

## Projektowanie technologii AI bez uwzględnienia wartości **interesariuszy jest nieetyczne**

Ignorowanie wartości w procesie projektowania oznacza złamanie zasad etycznych, ponieważ projektowanie to nie tylko tworzenie funkcjonalnych produktów, ale także kształtowanie narzędzi i doświadczeń wpływających na zachowania, nawyki i normy społeczne.



# Przykład nieprzestrzegania wytycznych AI godnej zaufania

**Scenariusz:** Władze przykładowego miasta wdrażają system AI do przewidywania poziomu przestępczości, zaprojektowany w celu prognozowania częstotliwości przestępstw i efektywniejszego przydzielania zasobów policyjnych.

## Niedokładne prognozy:

- Model AI jest trenowany na stroniczych historycznych danych o przestępczości, co prowadzi do niedokładnych i wypaczonych prognoz.
- Brak kalibracji systemu - niedokładne testowanie i walidacja..
- Wady techniczne, takie jak przeuczenie i brak przejrzystości w procesie decyzyjnym algorytmu, obniżają jego niezawodność i skuteczność.

**Perspektywa techniczna**

**Perspektywa prawna**

**System SI**  
(naruszający wytyczne godnej zaufania AI)

**Perspektywa wartości**

## Naruszenie przepisów o ochronie prywatności:

- System zbiera i przetwarza duże ilości danych osobowych bez uzyskania odpowiedniej zgody od indywidualnych ludzi.
- Obejmuje dane z mediów społecznościowych, rejestrów publicznych i prywatnych baz danych, naruszając przepisy dotyczące ochrony prywatności, takie jak RODO.
- Prowadzi to do wyzwań prawnych i kar za niezgodność z przepisami o ochronie danych.

## Upředzenia i dyskryminacja:

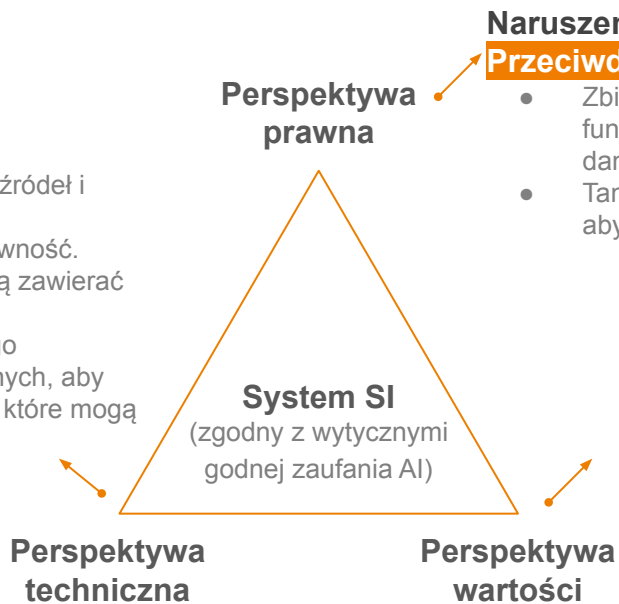
- System dysproporcjonalnie celuje w mniejszości, prowadząc do zwiększonego nadzoru i działań policyjnych w tych obszarach.
- Pogłębia to istniejące nierówności społeczne i budzi poczucie niesprawiedliwości oraz brak zaufania wobec policji.

**Scenariusz:** Władze przykładowego miasta wdrażają system SI do przewidywania poziomu przestępczości, zaprojektowany w celu prognozowania częstotliwości przestępstw i efektywniejszego przydzielania zasobów policyjnych.

## Niedokładne prognozy

### Przeciwdziałanie

- Zbieranie danych z różnych źródeł i okresów, aby zapewnić ich różnorodność i reprezentatywność. Unikanie danych, które mogą zawierać uprzedzenia.
- Przeprowadzenie dokładnego czyszczenia i weryfikacji danych, aby usunąć błędy i uprzedzenia, które mogą wpływać na wyniki modelu.



### Przeciwdziałanie

- Zbieranie tylko niezbędnych danych osobowych potrzebnych do funkcjonowania systemu. Unikanie zbierania nadmiarowych danych, które nie są bezpośrednio związane z celami systemu.
- Tam, gdzie to możliwe, stosowanie technik ochrony prywatności, aby zminimalizować ryzyko identyfikacji osób postronnych.

## Uprzedzenia i dyskryminacja

### Przeciwdziałanie

- Przeprowadzenie szczegółowej analizy danych, aby zidentyfikować i zrozumieć istniejące uprzedzenia i dyskryminacyjne wzorce.
- Współpraca z ekspertami ds. etyki, praw człowieka, socjologii oraz lokalnymi społecznościami w celu lepszego zrozumienia problemu.

# Rekomendowane wsparcie dla tworzenie ekosystemu etycznej AI



**Edukacja i Szkolenia** Wspieranie tworzenia i wdrażania programów edukacyjnych, które koncentrują się na podstawach etyki SI, prawach człowieka i przepisach dotyczących ochrony prywatności.

**Warsztaty i seminaria** Organizowanie warsztatów i seminariów, przekazujących najnowsze informacje na temat etycznych wyzwań i najlepszych praktyk w zakresie AI.

**Rady doradcze** Powstanie rady doradczej złożonej z ekspertów z różnych dziedzin, takich jak prawo, technologia, nauki humanistyczne, która będzie wspierać w tworzeniu polityk i przepisów dotyczących AI.

**Audyt etyczny** Wprowadzenie audytu etycznego dla projektów AI, aby zapewnić zgodność z wartościami społecznymi.

**Projektowanie etyczne** Promowanie włączenia aspektów etycznych do projektów AI jako integralnej części metodologii projektowe (Ethics by design).

**Promowanie przejrzystości i odpowiedzialności** Zachęcanie do stosowania otwartych standardów i przejrzystych procesów decyzyjnych w projektach AI, aby interesariusze mogli lepiej zrozumieć, jak działa AI i jak decyzje są podejmowane.

**Mechanizmy raportowania** Wdrożenie mechanizmów umożliwiających interesariuszom zgłaszanie obaw i problemów związanych z AI, co zwiększa odpowiedzialność za działania systemów AI.

**Partnerstwo publiczno-prywatne** Tworzenie partnerstw między rządem, sektorem prywatnym i organizacjami pozarządowymi w celu wspólnego rozwijania i wdrażania etycznych technologii AI.

**Forum dialogu** Organizowanie forum dialogu, gdzie przedstawiciele różnych sektorów mogą wymieniać się wiedzą i doświadczeniami dotyczącymi etycznych aspektów AI.

**Finansowanie badań** Pozyskiwanie finansowania dla badań i projektów skupiających się na etyce AI, aby rozwijać nowe narzędzia i metody zapewniające etyczność systemów AI.

**Regulacje i standardy** Opracowywanie i wdrażanie standardów, które nakładają obowiązki na twórców i użytkowników AI w zakresie zgodności z etycznymi normami.

## Umiejętności

**Umiejętność interdyscyplinarnego podejścia i współpracy:** Integracja perspektyw prawnych, technicznych i etycznych wymagającej umiejętności skutecznej komunikacji i współpracy między różnymi specjalistami.

**Umiejętność analizy i oceny etycznej:** Umiejętność identyfikowania potencjalnych zagrożeń i konfliktów wartości związanych z systemami AI oraz dokonywania ocen etycznych w celu zapewnienia zgodności z prawami podstawowymi i normami etycznymi.

**Umiejętność operacjonalizacji wytycznych etycznych:** Zdolność do przekładania ogólnych zasad etycznych na konkretne rozwiązania projektowe i techniczne w systemach AI, z uwzględnieniem potencjalnych konfliktów wartości.

## Postawy

**Wrażliwość i pokora etyczna:** Rozwijanie zdolności do dostrzegania i rozpoznawania kwestii etycznych oraz potencjalnych konsekwencji systemów AI dla społeczeństwa. Zrozumienie ograniczeń własnej wiedzy, otwartość na różne punkty widzenia oraz gotowość do kwestionowania własnych przekonań i założeń.

**Odpowiedzialność społeczna:** Poczucie odpowiedzialności za wpływ systemów AI jako systemów socjotechnicznych na społeczeństwo oraz dążenie do maksymalizacji korzyści społecznych i minimalizacji potencjalnych szkód.

**Samodzielność poznawcza:** Zdolność do samodzielnego zdobywania, analizowania i wykorzystywania wiedzy o AI w celu rozwijania własnych kompetencji i rozwiązywania problemów bez stałej pomocy zewnętrznej.

## Znaczenie Etyki AI

Etyczne projektowanie i wdrażanie systemów SI jest kluczowe dla zwielokrotnienia dobra społecznego.

Zgodność z prawem i standardami technicznymi nie gwarantuje automatycznie korzyści społecznych.

Należy uwzględniać aspekty etyczne, aby zapobiegać potencjalnym szkodom.

## Interdyscyplinarne podejście AWSI

Członkowie AWSI to interdyscyplinarna grupa składająca się z prawników, inżynierów i humanistów.

Połączenie perspektyw prawnej, technicznej i zorientowanej na wartości tworzy kompleksowe i zrównoważone rozwiązania AI.

Zapewnienie zgodności z przepisami, solidności technicznej i norm etycznych zwiększa akceptację i zaufanie społeczne.

## Kierunki dalszej pracy i współpracy z interesariuszami

Dialog z decydentami i legislatorami & aktywne uczestnictwo w procesach tworzenia regulacji dotyczących AI, dzielenie się wiedzą i doświadczeniami w zakresie etyki SI.

- Edukacja & szkolenia
- Audyt etyczny
- Promowanie przejrzystości i odpowiedzialności

# Zespół AWSI

---

## **Polskie Towarzystwo Informatyczne**

Sekcja Aktualne Wyzwania Sztucznej Inteligencji

ul. Solec 38 lok. 103

00-394 Warszawa

[pti.org.pl/awsi](http://pti.org.pl/awsi) | [aws\\_i@pti.org.pl](mailto:aws_i@pti.org.pl)

Kontakt w sprawie prezentacji: [wbednaruk@pjwstk.edu.pl](mailto:wbednaruk@pjwstk.edu.pl)

