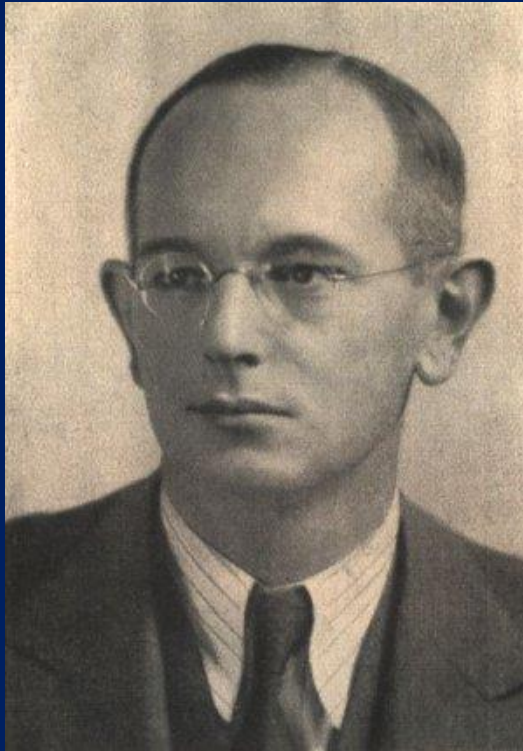


Sztuczna inteligencja

Dyskusja o etyce
19.09.2017 CZiITT

Ewa Fabian

Przykład 1: Czarna skrzynka (black box)

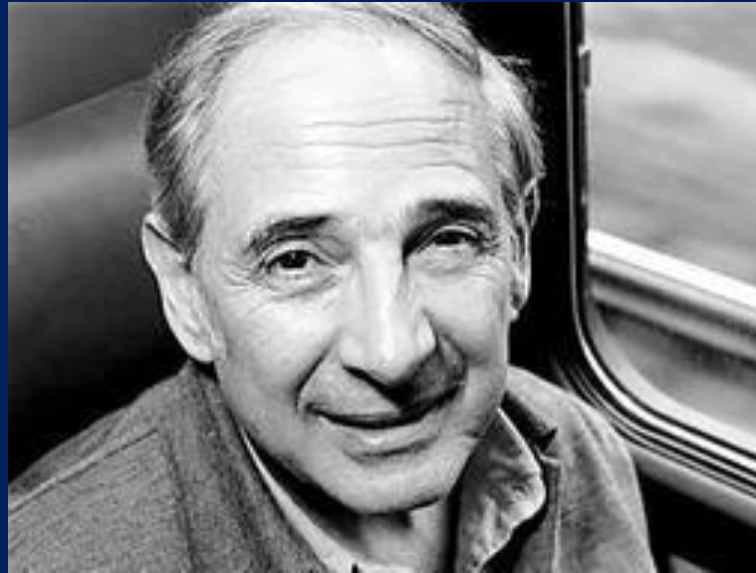


WILHELM CAUER, fot. nieznan, pobrano: Wikipedia

Przykład 1: Czarna skrzynka (black box)

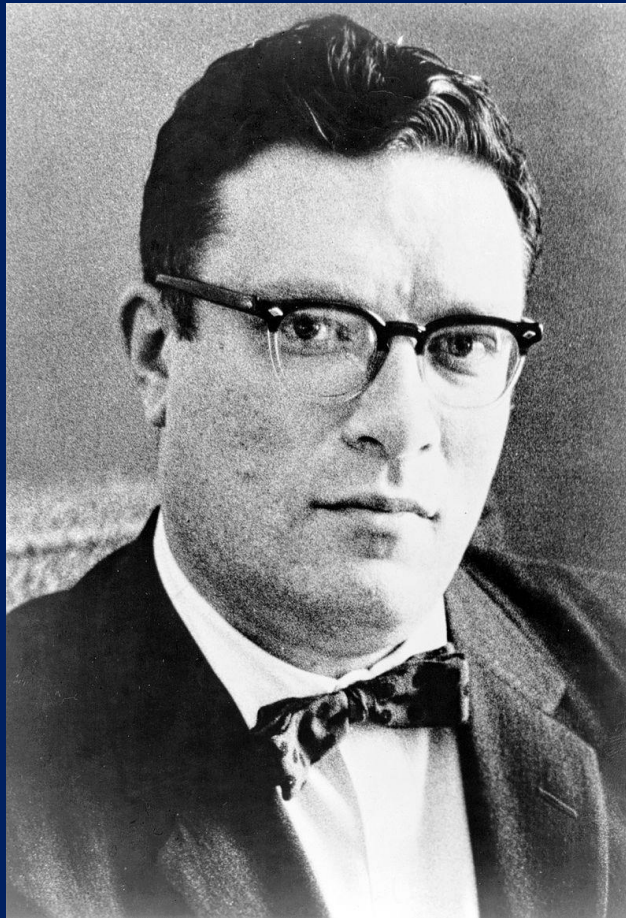
- *„Urządzenie, system lub przedmiot, postrzegany ze względu na stan wyjściowy i stan po przetworzeniu informacji, ale bez wiedzy o tym, jakie procesy w nim zachodzą. Mianem „czarnej skrzynki” określić można tranzystor, algorytm lub umysł ludzki. Przeciwnieństwem czarnej skrzynki jest system, w którym wewnętrzne elementy są dostępne do wglądu.” (A. Przegalińska, Istoty wirtualne, Kraków 2016)*

Pochodne: argument chińskiego pokoju (John Searle)



JOHN SEARLE, pobrano: www.ted.com

Przykład 2: prawa robotyki Asimova



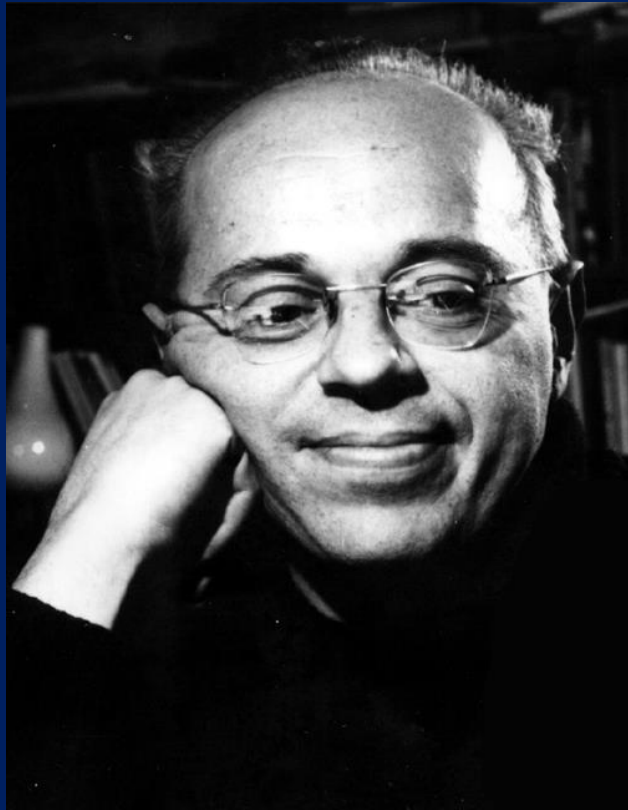
U.S. Library of Congress, Ph. Leonian, pobrano: Wikipedia

Przykład 2: prawa robotyki Asimova

1. Robot nie może skrzywdzić człowieka, ani przez zaniechanie działania dopuścić, aby człowiek doznał krzywdy.
2. Robot musi być posłuszny rozkazom człowieka, chyba że stoją one w sprzeczności z Pierwszym Prawem.
3. Robot musi chronić sam siebie, jeśli tylko nie stoi to w sprzeczności z Pierwszym lub Drugim prawem. (1943)

Prawo zerowe: *Robot nie może skrzywdzić ludzkości, ani przez zaniechanie działania doprowadzić do uszczerbku dla ludzkości.*

Przykład 3: pesymizm Stanisława Lema



W. Zemek, pobrano: Wikipedia

Przykład 3: pesymizm Stanisława Lema

- *„Na czym polega to człowieczeństwo, którego oni nie mają? Może naprawdę jest tylko ożenkiem alogiczności z tą poczciwością, tym „zacnym sercem”, i tym prymitywizmem odruchu moralnego (...) Skoro maszyny cyfrowe nie są ani zacie, ani nielogiczne... Więc w takim rozumieniu człowieczeństwo jest to suma naszych defektów, mankamentów, naszej niedoskonałości, jest tym, czym chcemy być, a nie potrafimy, nie możemy, nie umiemy, to jest po prostu dziura między ideałami a realizacją...”* (Opowieści o pilocie Pirxie)

Przykład 4: obserwacje Ursuli K. Le Guin



Marian Wood Kolish, pobrano: ursulaleguin.com

Przykład 4: obserwacje Ursuli K. Le Guin

- *„Człowiek może zawierzyć swojemu szczęściu, społeczeństwo nie, bo przemiany kulturowe, jak przypadkowe mutacje, mogą nasilić element ryzyka. W dowolnie wybranym punkcie ich historii powierzchowny obserwator mógłby powiedzieć, że wszelki postęp technologiczny i dyfuzja kulturowa uległy tu zahamowaniu. Ale nigdy tak nie było. Porównajmy tropikalną ulewę i lodowiec. Po swoim każde dochodzi tam, dokąd zmierza.”*

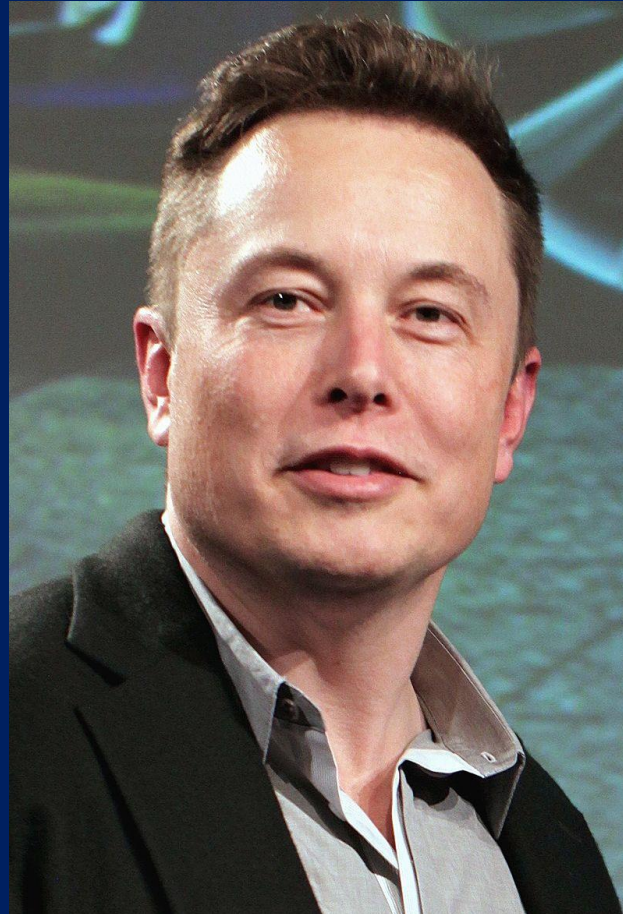
Lewa Ręka Ciemności, 1969, tłumaczenie: Lech Jęczmyk

Przykład 5: Parlament Europejski

- Rezolucja Parlamentu Europejskiego z dnia 16 lutego 2017 r. zawierająca zalecenia dla Komisji w sprawie przepisów prawa cywilnego dotyczących robotyki (2015/2103(INL))
 - Cytuje prawa robotyki I. Asimova.
 - Zawiera „**Kartę dotyczącą robotyki**” (m.in. dla inżynierów): „Proponowany kodeks postępowania etycznego w dziedzinie robotyki będzie stanowić podstawę identyfikacji, kontroli oraz przestrzegania podstawowych zasad etycznych, począwszy od fazy projektu, a skończywszy na fazie rozwoju.”



Przykład 6: Future of Life Institute



ELON MUSK, fot. Steve Jurvetson, pobrano: Wikipedia

*Technology is giving life
the potential to flourish
like never before...*



*...or to self-destruct.
Let's make a difference!*

AN OPEN LETTER TO THE UNITED NATIONS CONVENTION ON CERTAIN CONVENTIONAL WEAPONS

As companies building the technologies in Artificial Intelligence and Robotics that may be repurposed to develop autonomous weapons, we feel especially responsible in raising this alarm. We warmly welcome the decision of the UN's Conference of the Convention on Certain Conventional Weapons (CCW) to establish a Group of Governmental Experts (GGE) on Lethal Autonomous Weapon Systems. Many of our researchers and engineers are eager to offer technical advice to your deliberations.

We commend the appointment of Ambassador Amandeep Singh Gill of India as chair of the GGE. We entreat the High Contracting Parties participating in the GGE to work hard at finding means to prevent an arms race in these weapons, to protect civilians from their misuse, and to avoid the destabilizing effects of these technologies. We regret that the GGE's first meeting, which was due to start today (August 21, 2017), has been cancelled due to a small number of states failing to pay their financial contributions to the UN. We urge the High Contracting Parties therefore to double their efforts at the first meeting of the GGE now planned for November.

Lethal autonomous weapons threaten to become the third revolution in warfare. Once developed, they will permit armed conflict to be fought at a scale greater than ever, and at timescales faster than humans can comprehend. These can be weapons of terror, weapons that despots and terrorists use against innocent populations, and weapons hacked to behave in undesirable ways. We do not have long to act. Once this Pandora's box is opened, it will be hard to close. We therefore implore the High Contracting Parties to find a way to protect us all from these dangers.

Translations: [Chinese](#) [Japanese](#) [Russian](#)

[FULL LIST OF SIGNATORIES TO THE OPEN LETTER](#)

List otwarty Future of Life Institute (21.08.2017)

- *„Śmiercionośna broń autonomiczna stwarza zagrożenie wywołania trzeciej rewolucji w przemyśle wojennym. Raz rozwinięta, pozwoli na walkę zbrojną na skalę większą niż kiedykolwiek, szybszą niż ludzie potrafią pojąć. Może posłużyć za broń terroru, używaną przez despotów i terrorystów przeciwko niewinnym populacjom, może być narażona na działania hackerów, może zachowywać się w sposób niepożądany. Nie mamy zbyt wiele czasu. Gdy ta puszka Pandory zostanie otwarta, będzie trudno ją zamknąć. Wzywamy zatem Wysokie Umawiające się Strony do znalezienia sposobu na ochronę nas wszystkich przed tymi niebezpieczeństwami.”* (tłumaczenie własne, E.F.)

- „Lethal autonomous weapons threaten to become the third revolution in warfare. Once developed, they will permit armed conflict to be fought at a scale greater than ever, and at timescales faster than humans can comprehend. These can be weapons of terror, weapons that despots and terrorists use against innocent populations, and weapons hacked to behave in undesirable ways. We do not have long to act. Once this Pandora's box is opened, it will be hard to close. We therefore implore the High Contracting Parties to find a way to protect us all from these dangers.”

Asilomar AI Principles, 2017 (wyciąg)

- Zasada 8: **Transparentność w wymiarze sprawiedliwości** – jakiegokolwiek zaangażowanie sztucznej inteligencji w proces decyzyjny wymiaru sprawiedliwości powinno łączyć się zapewnieniem przekonującego uzasadnienia, możliwego do skontrolowania przez właściwy organ, z udziałem czynnika ludzkiego.
- Zasada 11: **Ludzkie wartości** – systemy AI powinny być projektowane i stosowane w zgodności z ideałami ludzkiej godności, prawami, wolnościami i różnorodnością kulturową.



Przykład 7: Moral Machine (2017)

- Platforma Massachusetts Institute of Technology - interaktywna ankieta dotycząca ludzkiej perspektywy wymiaru etycznego decyzji autonomicznych maszyn na przykładzie ruchu drogowego.

Results

Most Saved Character



Most Killed Character



Wobec przedstawionych przykładów, jaki sens i znaczenie mają dzisiaj dyskusje dotyczące etyki w rozwoju badań nad sztuczną inteligencją?

Co praktycznie jesteśmy w stanie osiągnąć i jaki powinien być podział ról?