Sztuczna inteligencja. Wykład wstępny

Paweł Rychlikowski

Instytut Informatyki UWr

22 lutego 2024

Wykłady

- Zapraszam w każdą środę!
- Mile widziane przerywanie pytaniami, itd.

Zajęcia pomocnicze

- Mamy ćwiczenia i pracownie, standardowo w rytmie: 5 * (P + P + C)
- Na każdą parę pracowni będzie lista pracowniowa, na każde ćwiczenia będzie lista ćwiczeniowa
- Dodatkowo można pisać nieobowiązkowy projekt (z grubsza w czasie P5)
- Zwyczajowo uczestnictwo **Codig Game Spring Challenge** jest również częścią naszego przedmiotu

Pracownia

Treścią pracowni sa indywidualne konsultacje:

- dotyczące zadań, które Student(ka) wykonał(a)
- **oraz tych, z którymi ma kłopot**

Pracownia

Kilka drobiazgów:

- Sprawdzanie może być selektywne – prowadzący nie musi oglądać wszystkich zadań (ale za wszystkie 'zadeklarowane' powinien przyznać punkty)
- Student ma prawo zadawać różne pytania, prosić o wskazówki, o pomoc przy debugowaniu. itp
- Prowadzący decyduje, jak wygląda opowiadanie o zadaniu. Może na przykład poprosić o zreferowanie rozwiązania, albo zadać pytanie o jakiś fragment, albo przeczytać kod i stwierdzić, że wszystko jasne, albo ...

Prośba od prowadzących

Opowiadając o zadaniu nie trzeba mówić o wszystkich funkcjach. Znaczną część rutynowych rzeczy można (należy) pomijać

Pracownia

- **Każde zadanie powinno mieć b. krótki opis (w pierwszym komentarzu)**
- W opisie należy umieścić najważniejsze informacje o zadaniu, przykładowo (dla zadania szachowego):

  ```
  W rozwiązaniu używam standardowego algorytmu BFS.
  Sytuację na planszy pamiętam jako string, postaci
  ''biały król jest na [pole1], czarny na [pole2], a
  wieża na [pole3]''.  Przy teście 3 konieczne jest
  32 GB pamięci (inaczej mamy segfault)
  ```

O sprawdzaniu zadań

Sprawdzaczka

python validator.py z1.2 solution.exe

Możliwe inne rozwiązania (np. wykorzystanie Coding Game)

Python

- Jest językiem sprawdzaczki (więc trzeba go mieć zainstalowanego)
- Nadaje się do większości zadań
- Może warto wykorzystać te zajęcia, żeby się go poduczyć
- Ma wiele bibliotek wspierających różne zadania związane z AI

Ale ogólnie nie wymuszamy żadnego języka programowania

Jak nie Python, to co?

Inne języki często używane w kontekście AI:

- **R**: wsparcie dla ML (zły wybór na naszym przedmiocie)
- **C++, Java, Go, Rust, ...**: szybkość symulacji, przeszukiwania, efektywne zarządzanie pamięcią
- **Lisp, Prolog**: dobre dopasowanie do **niektórych** zadań AI

Pierwsza pracownia

- Zasadniczo zakładamy, że rozwiązujemy ją bazując na wstępnej wiedzy
- ale oczywiście można spytać o dowolne zadanie, również dzisiaj po wykładzie

Ćwiczenia

- System deklaracyjny (przed zajęciami mówimy, co umiemy i za te zadania dostajemy punkty)
- Dodatkowa zachęta dla rozwiązujących: bonus +2 za 50%

Punkty, zaliczenia, egzamin

- Za prezentację (i wcześniejsze napisanie) programów studenci dostają punkty
- Podobnie za deklaracje zadań na ćwiczeniach

Wszystkie punkty są równe!
Sumujemy je i wyznaczamy ocenę (szczegóły będą w regulaminie)

Literatura

- Stuart Russel, Peter Norvig, Artifficial Intelligence. A Modern Approach. 3rd edition (w Internecie leży pdf)
- Fajna, ale 1100 stron.
- **Jest już czwarte wydanie!** (jak ktoś chce kupić, to koniecznie w miękkiej okładce!)
- Opcja *istotnie* tańsza (po polsku): AIMA podzielona na dwa tomy, wydana przez Helion

Książka


Stuart
Russell
Peter
Norvig
Sztuczna inteligencja
Nowe spojrzenie
Wydanie IV. Tom 1
Helion

Sztuczna inteligencja. Definicja

Definicja (krótka)

Zdolność komputera (programu) do wykonywania zadań powszechnie kojarzonych z zachowaniem inteligentnym (ludzi lub zwierząt).

Definicja długa

Spis treści wybranego podręcznika o Sztucznej inteligencji

Sztuczna inteligencja (definicja długa)

Definicja długa – spis treści

Rozwiązywanie problemów przez przeszukiwanie, rozwiązywanie więzów, przeszukiwanie z oponentem, logika w opisie świata, wnioskowanie w logice, planowanie, modelowanie niepewności, reprezentacja wiedzy, wnioskowanie przy niepewności, podejmowanie decyzji, uczenie się z przykładów, uczenie się ze wzmocnieniem, przetwarzanie języka naturalnego, rozpoznawanie wzorców w obrazach i dźwiękach, robotyka, procedural content generation.

Przedmioty:

1. Machine Learning (obowiązek na DataScience)
2. Neural Networks and Natural Language Processing (Core Elective na DS, obecnie w trakcie podziału na 2 przadmioty)
3. Eksploracja danych (Core Elective na DS)
4. AI for games (Core Elective)
5. Przedmioty związane z (zaawansowaną) logiką

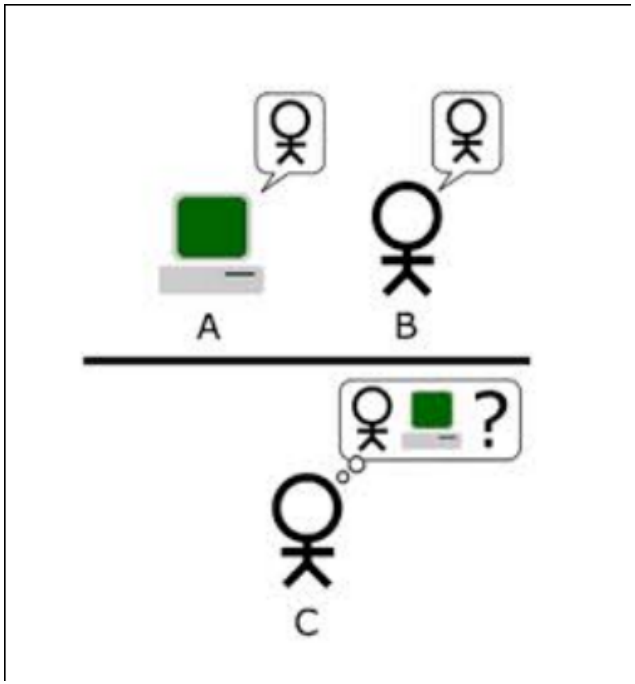
Definicja na bazie wiedzy „portalowej”

Ostatnio fraza **Sztuczna inteligencja** bywa używana w specyficznych konktekstach.
Przykłady:

- Sztuczna inteligencja zabierze pracę twórcom?
- Sztuczna inteligencja stworzyła zdjęcia z imprezy...
- Zapytali sztuczną inteligencję o górnictwo w Polsce. Co im odpowiedziała...

Bardzo często fraza **Sztuczna inteligencja** dotyczy tzw. **modeli generatywnych** (czyli tworzących treści cyfrowe, teksty, obrazy, wideo)

Test Turinga



- Pytanie: Czy maszyny potrafią myśleć?
- Odpowiedź: Możemy to sprawdzić w rozmowie.

Pytanie

Czy naprawdę chcemy, żeby komputer był nierozróżnialny w rozmowie od człowieka?

źródło: tvtropes.org

Nieludzkie, pożądane cechy: wiedza, zdolności rachunkowe, modelowanie świata, rozwiązywanie zadań fizycznych, ...

Red Flag for AI

Czy komputery mogą przejść test Turinga

Czy umiemy odpowiedzieć na to pytanie w lutym 2024?

Przykładowy agent konwersacyjny

```
x = input('Ty: ')
print ('Bot:', 'Cześć, miło mi Cię spotkać')
x = input('Ty: ')
print ('Bot:', 'Trudno mi o tym mówić, bowiem mój
                chomik... On chyba ma depresję!')
x= input('Ty: ')
print ('Bot:', 'Dzięki za wsparcie.')
```

Programy, które **czasami** przechodzą test Turinga, są z nami od bardzo dawna. Dwa istotne przykłady:

- **ELIZA** (1965) – udawała psychoterapeutę (i potrafiła zmylić pacjentów)
- **PARRY** (1972) – udawał osobę z zaburzeniami (i potrafił zmylić psychiatrów)

```
  Welcome to
                     EEEEEE  LL      IIII   ZZZZZZ   AAAAA
                     EE      LL       II       ZZ   AA   AA
                     EEEEE   LL       II     ZZZ    AAAAAAA
                     EE      LL       II    ZZ      AA   AA
                     EEEEEE  LLLLLL  IIII ZZZZZZ    AA   AA

  Eliza is a mock Rogerian psychotherapist.
  The original program was described by Joseph Weizenbaum in 1966.
  This implementation by Norbert Landsteiner 2005.


ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```

Model językowy

Definicja

Modelem językowym nazywamy rozkład prawdopodobieństwa na ciągach słów:

$$p(w_{n+1}|w_1 w_2 \dots w_n)$$

Model językowy może być użyty do generowania tekstu:

- Wybieramy (najbardziej?) prawdopodobne w_{n+1} i powtarzamy proces

Przykładowe metody wyboru

- Wybieramy najbardziej prawdopodobny (nudne...)
- Losujemy z K najbardziej prawdopodobnych następników
- Losujemy wśród najbardziej prawdopodobnych następników, które łącznie mają prawdopodobieństwo p

Model językowy

Modelami językowymi są:

- GPT-1, GPT-2, GPT-3 (po prostu modelują język)
- ChatGPT (GPT 3.5), GPT-4, LaMDA, ... (dodatkowo dotrenowywane do dialogów)
- Nie tylko język:
  - AlphaFold (modeluje białka),
  - część AlphaStar (modeluje rozgrywkę StarCrafta),
  - T5 (robi różne rzeczy, w tym tłumaczenie czy klasyfikacja tekstów),
  - GATO (modeluje wiele rzeczy, w tym rozwiązywanie łamigłówek, czy gry na stare konsole)

GPT-2, ciąg dalszy

- Wielkości modeli:
  - GPT-2: 1.5 G-float
  - GPT-3: 175 G-float (podstawa ChatGPT)
  - GPT-4 (?): 100 T-float (w zapowiedziach, raczej dużo mniej)

Czy ilość przechodzi w jakość? (otwarte pytanie)

GPT-2, który próbuje przeprowadzać rozumowanie

Suppose I have 4 apples, and I gave two of them to my sister, and she gave one of them to her sister, and it started to grow. Would I be able to see how much time it had taken for the apple to grow?

We will first give you a number which you will later use to work out how much time it has taken for the apple to grow:

(Time it took to grow 2 apples, shown as a clock) x (Size of apple) = "time taken"

What this tells us is that:

2 x 3 x 2 = 44 (seconds in 24 hours)

So the question is, what does 44 seconds look like?

Let's do a quick experimen

ChatGPT o naszym przedmiocie

jest tworzenie systemów, które będą potrafiły w sposób autonomiczny podejmować decyzje i wykonywać zadania, które obecnie wymagają udziału człowieka.

P

Dziękuję, ale prosiłbym o krótszą definicję, najlepiej w formie przyśpiewki weselnej

Oczywiście, z przyjemnością pomogę! Oto przyśpiewka weselna na temat sztucznej inteligencji:

Sztuczna inteligencja to magia,
Komputery, co same się uczą,
Algorytmy, sieci neuronowe,
Decyzje podejmują jak człowiek prawdziwy.

Dane to klucz, nauczanie maszynowe,
Rozpoznawanie obrazów, mowy,
Sztuczna inteligencja na każdym kroku,
Świat zmienia, zaskakuje, niesamowity!

Słowa klucze dla naszego wykładu

- Przeszukiwanie (problem solving)
- Uczenie się (learning)
- Wnioskowanie (inference, theorem proving)
- Modelowanie wiedzy o świecie

Główny problem SI

Głównym, nierozwiązanym do tej pory problemem SI jest satysfakcjonujące połączenie tych światów, zwłaszcza wnioskowania (twarda logika), z rozmytą wiedzą pochodzącą z uczenia się.

Dlaczego problemy SI są trudne

Istnieją dwa główne powody

1. Trudność algorytmiczna:
   - NP-zupełne - łamigłówki,
   - PSPACE-zupełne - gry,
   - EXPTIME i wyżej – systemy wnioskowania
2. Czasem trudno sformalizować precyzyjnie problem!

Co się udało Sztucznej inteligencji?

Gdzie komputery są od nas lepsze?

1. Oczywiście my też wykonujemy algorytmy (np. mnożenie liczb). Tu od zawsze **one** mają przewagę.
2. Komputery radzą sobie w grach: Backgammon (tryktrak) (1992), Szachy (1996), Go (2016), Poker (2017), StarCraft (AlphaStar, 2019?), Stratego (2022), Dyplomacja (2022)
3. Rozpoznawanie mowy (Microsoft/IBM, Switchboard Corpora, około 2016)

Co się udało Sztucznej inteligencji? (cd)

Gdzie komputery są od nas lepsze?

1. Rozpoznawanie prostych obrazów (ImageNet, 1mln obrazków, 1000 klas):
   - 2010: około 28% błędów
   - 2015 (4.94% na ImageNet, człowiek: 5.1%)
   - W 2020 było 1.3%, w 2021 było 1.2%, 2023: **0.98%**.
2. Tłumaczenie maszynowe
3. Wygrywanie teleturniejów wiedzowych (Watson, Jeopardy, 2011)

Co działa niekoniecznie idealnie?

1. Rozpoznawanie mowy (50% błąd na nagraniach *przy koltlecie*)
2. Roboty umiejące nalewać wodę, otwierać drzwi, itd w **nieznanym** środowisku
   - DARPA challenge fails
   - RoboCup (zadanie domowe: Finale RoboCup 2023 League Kid Size)
3. Gra w Brydża (chociaż w 2022 NukkAI program komputerowy wygrał „turniej brydżowy", ale tylko jako rozgrywający i tylko 3NT)
4. „Ludzka" rozmowa na dowolny temat (??)

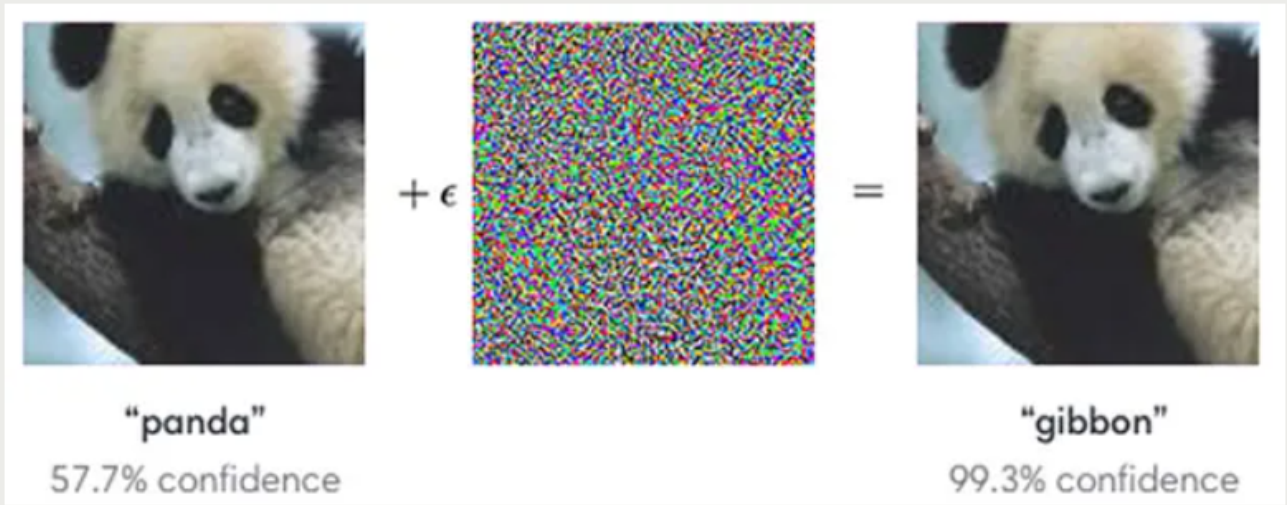
The IKEA test

(film z roku 2018)

Zadanie (na przyszłość)

Para robotów ma złożyć **dowolny** mebel IKEA, korzystając jedynie z instrukcji umieszczonej w pudełku.

Ataki na rozpoznawanie obrazów

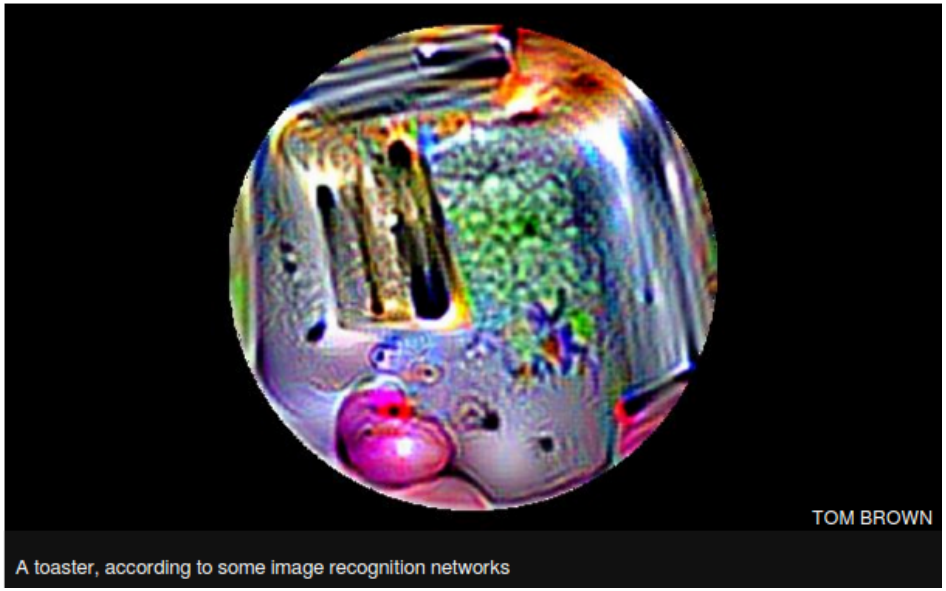


An image of a panda, when combined with an adversarial input, can convince a classifier that it’s looking at a gibbon. IMAGE: OPENAI

Źródło: Slight Street Sign Modifications Can Completely Fool Machine Learning Algorithms

Rozpoznawanie obrazów. Super toster

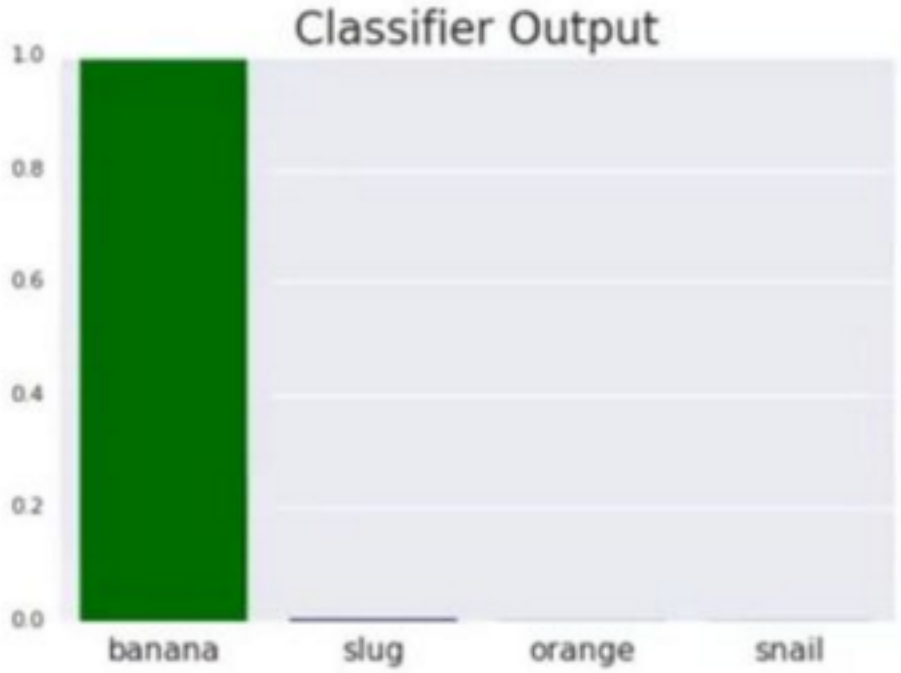
Sztucznie wygenerowany obraz, maksymalizujący tosterowatość.



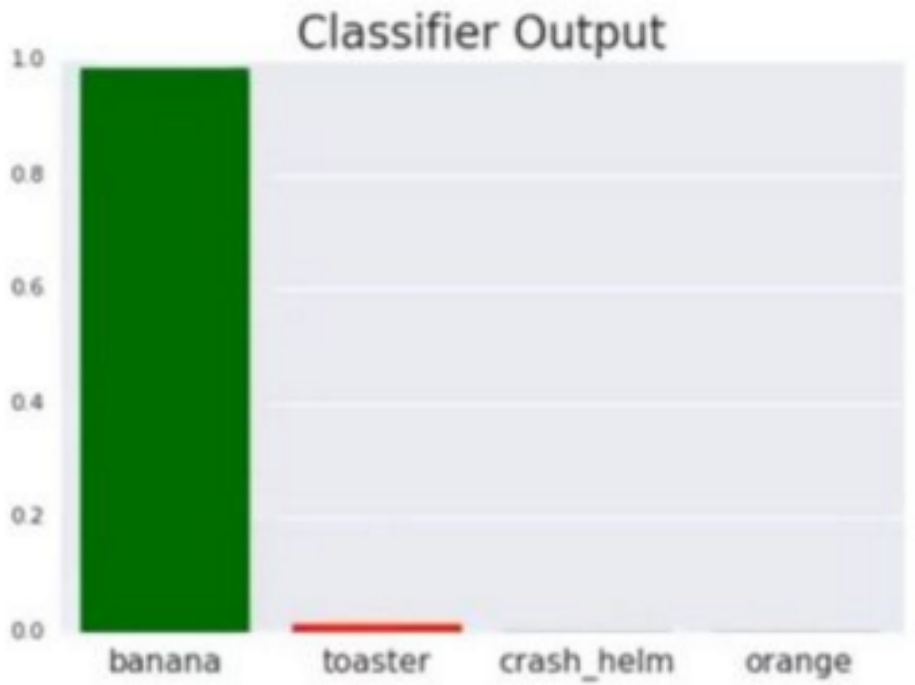
Co możemy zrobić z tym obrazkiem?

- Możemy go pokazywać sieci.
- Ale wklejając go analogowo, nie cyfrowo.
- Zobacz pracę: Adversarial Patch, T. Brown i inni, 2017

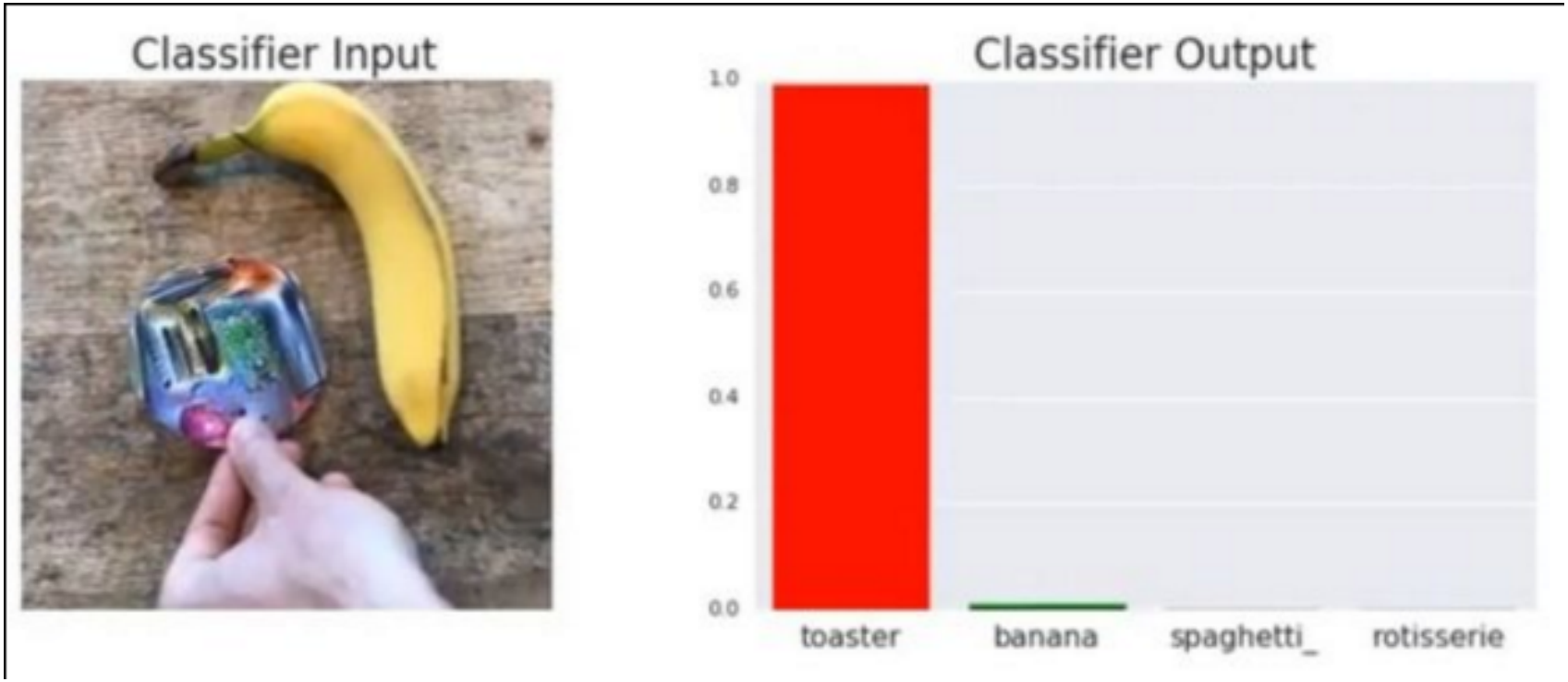
Tostery i banany



Tostery i banany



Tostery i banany



Modyfikacje znaków

Ograniczenia prędkości do **45 mil na godzinę**

Figure 1: The left image shows real graffiti on a Stop sign, something that most humans would not think is suspicious. The right image shows our a physical perturbation applied to a Stop sign. We design our perturbations to mimic graffiti, and thus "hide in the human psyche."

Źródło: Robust Physical-World Attacks on Deep Learning Visual Classification

Cel na początek

Najpierw zajmiemy się **przeszukiwaniem**, które jest jednym z podstawowych narzędzi AI.