Katarzyna Opiela

MORE SCIENCE THAN FICTION

1. Uwagi ogólne

Zestaw materiałów opatrzony wspólnym tytułem *More science than fiction* jest adresowany do studentów uzupełniających studiów magisterskich studiujących kierunki ścisłe i humanistyczne. Przedstawione ćwiczenia mogą być wykorzystane do pracy z grupami studentów informatyki lub matematyki (ze względu na tematykę) oraz jako wstęp do tematów związanych ze światem przedstawionym w filmach science fiction ze studentami filmoznawstwa, jak również ze studentami psychologii (ze względu na pytania o odczuwanie emocji przez sztuczne mózgi).

2. Poziom zaawansowania: B2+ (z pomocą nauczyciela) oraz C1, C1+ i C2

3. Czas trwania opisanych ćwiczeń

Ćwiczenia zaprezentowane w tym artykule są przeznaczone na jedną jednostkę lekcyjną lub 120 minut (w zależności od tego, czy grupa zdoła przeczytać cały artykuł oraz wykonać związane z nim zadania). Czas trwania został ustalony na podstawie doświadczenia wynikającego z pracy nad poniższymi ćwiczeniami w grupach na poziomie B2+ i C1 filmoznawstwa i psychologii.

4. Cele dydaktyczne

W swoim założeniu artykuł ma rozwijać podstawowe umiejętności językowe, takie jak: słuchanie, czytanie oraz mówienie. Kolejnym celem jest rozwijanie krytycznego myślenia, jako że kilka z pytań ma charakter filozoficzny.

5. Uwagi i sugestie

W zbiorze przewidziane są ćwiczenia na interakcję student–nauczyciel, student–student oraz na pracę indywidualną. Ćwiczenia w zależności od poziomu grupy, stopnia zaangażowania studentów w zajęcia i kierunku mogą być odpowiednio zmodyfikowane. Zadania tu zamieszczone możemy omawiać na zajęciach lub część przedstawionych ćwiczeń zadać jako pracę domową, jeżeli nie chcemy poświęcać zbyt dużo czasu na zajęciach. Z grupami na poziomie B2+ zachęcam do wspólnej pracy nad częścią materiału, ponieważ artykuł jest trudny (na poziomie słownictwa i kontekstu kulturowego, którego studenci mogą nie znać). Materiały obejmują TED talk Sama Harrisa wraz z pytaniami na zrozumienie oraz zadaniem z zakresu słownictwa, które ułatwia zrozumienie prezentacji, jak również artykuł na temat problemów związanych z tworzeniem oprogramowania, a w przyszłości także sztucznej inteligencji. Zajęcia rozpoczynamy od próby odpowiedzi na pytanie, dlaczego w ogóle tworzymy technologię, następnie przechodzimy do pracy nad słownictwem, co ułatwia odbiór prezentacji, a później do słuchania jej pierwszej części. W dalszej części następuje dyskusja w parach lub małych (trzyosobowych) grupach na temat przemyśleń, jakie studenci mogą mieć po obejrzeniu materiału. Kolejnym zadaniem jest udzielenie odpowiedzi na pytania do teksu, gdzie słabsi studenci również mogą współpracować. Kolejne dwa (ostatnie) zadnia dotyczą słownictwa z artykułu i w razie braku czasu mogą zostać przeznaczone na samodzielną pracę w domu.

MORE SCIENCE THAN FICTION

I. You are going to watch a TED talk by Sam Harris. Before you watch, answer the question: "Why do we develop technology?"

II. In pairs, look up the terms:

to marshal something (an appropriate emotional response) – …………………

far-fetched (e.g. a conclusion) – ………….

……………………...…………………..

malevolent – …….......……………………

inevitable – ……………………………….

divergence – …………………………...

domains – ………………………………

to take pains (not) to do something – …. ……………………………………………

to safeguard – ……………………………

annihilate without a qualm – …………… ……………...……………………..

to go berserk – ……………………………

III. Watch the TED talk and answer the questions below.

The video can be found here: https://www.youtube.com/watch?v=8nt3edWLgIg

1. What is the difference between death by famine or death by sci-fi?
……………………………………………………………………………………

2. What's the problem with people's attitude to building AI?
……………………………………………………………………………………

3. How likely is it that people will stop developing technology?
……………………………………………………………………………………

4. What will happen eventually?
……………………………………………………………………………………

5. What's the intelligence explosion? How will this process be different?
…………………………………………………………………………………….
……………………………………………………………………………………

6. How could we be similar to ants one day?
……………………………………………………………………………………

IV. Answer the questions:

1. Have you been convinced by Harris? Why (not)?
2. What are your predictions concerning what future jobs will look like?

3. What direction will AI take?
4. How do you picture the end of the world and when do you think it will take place?

Homework

Watch the rest of the TED talk and answer the remaining questions 1-5:
1. What's intelligence?
2. What's general intelligence?
3. How is the lack of work and intellectual effort dangerous for us?
4. What would the Russians and the Chinese do?
5. What's the problem with the argument about time?

V. Now you're going to read an article about software picking up human bias.

1. Before you read, answer the question as to whether computers can be racist.
2. Read the article and answer the questions:
A. What will the apocalypse look like?
……………………………………………………………………………………
B. What controversy has been sparked by the software used by the police?
……………………………………………………………………………………
C. What is the "white guy" problem in AI?
…………………………………………………………………………………...…
D. What are the limitations of e.g., Facebook algorithms?
………………………………………………………………………………………...
E. What is the "Russian tank" problem?
………………………………………………………………………………………...
F. How was the software prepared by Xiaolin Wu and Xi Zhang trained?
………………………………………………………………………………………...
G. What sort of controversy has the project based on analyzing the facial features of criminals caused?
………………………………………………………………………………...…
H. What point does Blaise Agüera y Arcas make about the alleged lack of bias of the algorithm in question?
………………………………………………………………………………………...
I. What could be a better use for Xiaolin Wu's and Xi Zhang's algorithm?
……………………………………………………………………………………
J. What does the old programmer's saying "Garbage in, garbage out" mean?
………………………………………………………………………...……
3. In pairs/groups of three, answer the questions:
A) How could Xiaolin Wu's and Xi Zhang's algorithm be misused in the future?

B) What is the extent of (software) makers' responsibility for their creation?
C) Is it possible that one day we will be able to replicate the human brain combined with elements of AI? Could it be able to create and experience emotions? What might be the consequences of that?

How a Machine Learns Prejudice
by Jesse Emspak on December 29, 2016

1. Artificial intelligence picks up bias from human creators—not from hard, cold logic. If artificial intelligence takes over our lives, it probably won't involve humans battling an army of robots that **relentlessly** apply Spock-like logic as they physically **enslave** us. Instead, the machine-learning algorithms that already let AI programs recommend a movie you'd like or recognize your friend's face in a photo will likely be the same ones that one day deny you a loan, lead the police to your neighborhood or tell your doctor you need to go on a diet. And since humans create these algorithms, they're just as **prone to** biases that could lead to bad decisions—and worse outcomes.

2. These biases create some immediate concerns about our increasing **reliance on** artificially intelligent technology, as any AI system designed by humans to be absolutely "neutral" could still reinforce humans' **prejudicial** thinking instead of seeing through it. Law enforcement officials have already been criticized, for example, for using computer algorithms that allegedly tag black defendants as more likely to commit a future crime, even though the program was not designed to explicitly consider race.

3. The main problem is twofold: First, data used to calibrate machine-learning algorithms are sometimes **insufficient**, and second, the algorithms themselves can be poorly designed. Machine learning is the process by which software developers train an AI algorithm, using massive amounts of data **relevant to** the task at hand. Eventually, the algorithm spots patterns in the initially provided data, enabling it to recognize similar patterns in new data. But this does not always work out as planned, and the results can be horrific. In June 2015, for example, Google's photo categorization system identified two African Americans as "gorillas." The company quickly fixed the problem, but Microsoft AI researcher Kate Crawford noted in a *New York Times* op-ed that the blunder reflected a larger "white guy problem" in AI. That is, the data used to train the software relied too heavily on photos of white people, **diminishing** its ability to accurately identify images of people with different features.

4. The recent **spate** of fake stories **inundating** Facebook users' news feeds also highlights the AI bias problem. Facebook's trending news algorithm was prioritizing stories based on engagement—how often users click on or share. **Veracity** was not considered at all. In early November several news outlets revealed that a group of Macedonian teenagers had fooled Facebook's News Feed algorithm into promoting blatantly false stories that appealed to right-wing voters during the

U.S. election. Facebook says it has modified the algorithm since then and has announced plans for Snopes, Factcheck.org, ABC News and PolitiFact to help **weed out** obviously false articles.

5. "It's a bit like the 'Russian tank problem,'" says Hal Daumé III, an associate professor of computer science at the University of Maryland. This legend—apocryphal but illustrative, and **oft-related** by computer science teachers—dates from machine learning's early days in the 1980s. The story says the U.S. military tried training a computer to distinguish between Russian and American tanks in photos. "They got super-high classification accuracy—but all the photos of Russian tanks were blurry, and American tanks were high-definition," Daumé explains. Instead of identifying tanks, the algorithm learned to distinguish between grainy and high-quality photos.

6. Despite such known limitations, a group of researchers recently released a study asserting that an algorithm can **infer whether** someone is a convicted criminal by **assessing** facial features. Xiaolin Wu and Xi Zhang, researchers at China's Shanghai Jiao Tong University, trained a machine-learning algorithm on a dataset of 1,856 photos of faces—730 convicted criminals and 1,126 non-criminals. After looking at 90 percent of the pictures, the AI was able to correctly identify which ones in the remaining 10 percent were the convicted criminals.

7. This algorithm correlated specific facial characteristics with criminality, according to the study. Criminals, for example, were more likely to have certain spatial relationships between the positions of eye corners, lip curvature and the tip of the nose, Wu says—although he notes that having any one of those relationships does not necessarily indicate that a person is more likely to be a criminal. Wu also found that the criminals' faces differed from one another more, while non-criminals tended to share similar features.

8. Wu continued testing the algorithm using a different set of photos it had not seen before, and found that it could correctly spot a criminal more often than not. The researchers attempted to avoid bias by training and testing their algorithm using only faces of young or middle-aged Chinese men with no facial hair or scars.

9. "I set out to prove physiognomy was wrong," Wu says, referring to the centuries-old pseudoscience of assessing character based on facial features. "We were surprised by the results." Although the study might appear to **validate** some aspects of physiognomy, Wu acknowledges that it would be "insane" to use such technology to pick someone out of a police lineup, and says there is no plan for any law enforcement application.

10. Other scientists say Wu and Zhang's findings may be simply **reinforcing** existing **biases**. The subjects' criminality was determined by a local justice system run by humans making (perhaps subconsciously) biased decisions, notes Blaise Agüera y Arcas, a principal scientist at Google who studies machine learning. The central problem with the paper is that it relies on this system "as the ground truth for labeling criminals, then concludes that the resulting [machine learning] is unbiased by human judgment," Agüera y Arcas adds.

11. Wu and his colleagues "jump right to the conclusion that they found an underlying pattern in nature—that facial structure predicts criminality. That's a really **reckless** conclusion," says Kyle Wilson, an assistant professor of mathematics at Washington College who has studied computer vision. Wilson also says this algorithm may be simply reflecting the bias of the humans in one particular justice system, and might do the same thing in any other country. "The same data and tools could be used to better understand [human] biases based on appearance that are at play in the criminal justice system," Wilson says. "Instead, they have taught a computer to reproduce those same human biases."

12. Still others say the technology could be improved by accounting for errors in the patterns computers learn, **in an attempt** to keep out human prejudices. An AI system will make mistakes when learning—in fact it must, and that's why it's called "learning," says Jürgen Schmidhuber, scientific director of the Swiss AI Lab Dalle Molle Institute for Artificial Intelligence. Computers, he notes, will only learn as well as the data they are given allows. "You cannot eliminate all these sources of bias, just like you can't eliminate these sources for humans," he says. But it is possible, he adds, to acknowledge that, and then to make sure one uses good data and designs the task well; asking the right questions is **crucial**. Or, to remember an old programmer's saying: "Garbage in, garbage out."

Source: https://www.scientificamerican.com/article/how-a-machine-learns-prejudice/

4. Look for the synonyms of the words below in paragraphs 1, 3, 5, 7, 9, 11 and 12.

mercilessly; ruthlessly ……………………   important to …………..…..……………..
to put in chains …………….…………….....   trying to …………………….…………...
having a tendency to ………..……………...   careless …………………………………
not enough ………………………………...   essential, extremely ……...……….............
decreasing …………………………………   to confirm ………..…………………….....
often told ………………………………..

5. Try to come up with synonyms for the words in bold in paragraphs 2, 4, 6, 8 and 10.

to infer whether ………………………...   prejudicial ………………………………...
assessing ………………………………...   a spate …………………………………….
reinforcing ………………………………...   inundating ………………………………...
biases ………………………………...   veracity …………………………………...
reliance on …………….....……………..…   to weed out ………………………………...

KEY

I. Suggested answer: Because we can.

II.

marshal a response	give a response; respond
malevolent	evil
divergence	difference
take pains (not) to do	make an effort
annihilate without a qualm	destroy without regret
far-fetched	unlikely, improbable
inevitable	impossible to avoid
domains	fields of knowledge
safeguard	protect
go berserk	go mad (and violent)

III.

1. Death by famine is horrible, whereas death by science fiction is cool.
2. They fail to marshal an appropriate emotional response (which is fear, or at least, concern).
3. It is unlikely; we will not stop, even when faced with a catastrophe.
4. We will build machines which are smarter than us and they will be able to improve themselves.
5. The process could get away from us; machines wouldn't need us and we wouldn't have any influence on them.
6. If your goals and the goals of AI are different, they could destroy us the same way we destroy ants.

IV. Students come up with their own answers.

Homework

1. A matter of information processing in physical systems.
2. An ability to think flexibly across multiple domains.

3. We will be able to play Frisbee and give each other massages. The downside is wealth inequality and unemployment that's never been seen before. Most of the world will starve.
4. The Russians and the Chinese would try to build similar machines to stay ahead of the competition and be able to wage war. Remember the Cold War?
5. We don't know when we'll be able to build AI in safe conditions.

V.

1. In 2016 Twitter users were asked to teach a Microsoft bot and within hours the bot denied the Holocaust and praised Breivik so, yes, computers can be racist; it depends on what information they operate on.

2.

A. The apocalypse will be software denying us access to our accounts, denying us loans, and generally taking control over our lives. *(Para. 1)*
B. It is allegedly more likely to predict African American defendants to engage in criminal activity in the future. *(Para. 2)*
C. The software relied too heavily on photos of white people, diminishing its ability to accurately identify images of people with different features. *(Para. 3)*
D. Stories were prioritized based on popularity, not veracity. *(Para. 4)*
E. It is hard to predict what kind of data one has to feed an algorithm in order to produce accurate results. *(Para. 5)*
F. It was given a dataset of 1,856 photos of faces of criminals and non-criminals and after analyzing 90% of them was able to identify criminals in the last 10%. *(Para. 6)*
G. Other scholars claim it could contribute to reinforcing the existing biases and relies on data that is less than perfectly reliable – a judicial system which, of course, is prone to bias. Also, the assumption that criminality can be predicted is jumping to conclusions. *(Para. 10, 11)*
H. It cannot be free of prejudice as the data it is fed comes from a system that is faulty and riddled with bias. *(Para. 10, 11)*
I. It could be used to study biases based on appearance existing in the criminal justice system. *(Para. 11)*
J. If you feed your software wrong information, you are going to get wrong information. *(Para. 12)*

3. Suggested answer – the maker is ultimately responsible for his/her creation.
    A. Its use could cause more discrimination based on appearance.
    B. Making software these days is incredibly important as it could be misused in any number of ways.

C. It might be possible for us to create artificial brains and it may cause all sorts of problems – questions about the human rights of our creation might arise and, also, if it were far superior to us, it could try to enslave us. Questions about robot rights would be raised (after all, we've given rights to animals, who also have feelings)

Tasks 4 and 5 can be omitted or assigned as homework

4.

mercilessly; ruthlessly	relentlessly
to put in chains	enslave
decreasing	diminishing
often told	oft-related
important to	relevant to
having a tendency to	prone to
not enough	insufficient
trying to	in an attempt to
careless	reckless
essential	crucial
to confirm	validate

5.

to infer whether	conclude if
assessing	grading; evaluating
reinforcing	making stronger; fossilising
biases	prejudice; preference
reliance on	dependence on
prejudicial	biased; skewed; intolerant
a spate	a rush, an outpouring (chiefly Brit.: flash flood)
inundating	overwhelming; flooding
veracity	truthfulness; accuracy
to weed out	get rid of; eliminate