

## [Transcript] Nauka To Lubie / Czy sztuczna inteligencja stanowi dla nas zagrozenie?

Nauka to Lubie i Sztuczna Inteligencja, trzeci live, trzeci kolejny miesiac.

Mateusz Chrobok, Tomasz Rozek. Dzisiaj o Sztucznej Inteligencji, o zagrozeniach i o bezpieczenstwie.

Bardzo Wam dziekuje za to, ze juz jesteście z nami. Bardzo proszę, jak tylko będziecie oglądali, jak tylko będzie Wam się podobało, żebyście udostępniali tego live'a w Waszych mediach społecznościowych,

żeby jak najwięcej mogło osób na niego trafić.

Mateusza, przedstawiać nie muszę. Cześć Mateuszu.

Dajmy.

Kilka ogłoszeń na wstępie, żeby nie zapomnieć, a później już lecimy z naszym formatem.

Podziękowania oczywiście dla Uniwersytetu Śląskiego. Jesteśmy w studiu Uniwersytetu Śląskiego.

Bardzo Wam dziękujemy Uniwersytecie za współpracę.

Druga rzecz, o której chciałem przypomnieć, to tak. Właśnie o książce, o książce Kosmos, którą cały czas możecie oglądać, nie tylko oglądasz, ale też możecie się kupić,

o książce Kosmos dla dzieci, o historii pewnej sądy kosmicznej, która podróżuje przez układ słoneczny

i która ma bardzo fajne przygody. Książka jest świetnie ilustrowana, więc Wam ją polecam.

Tym bardziej, że razem z książką jest też zeszyt zabaw, zeszyt gier, dzięki któremu Wasze dzieci, może dzieci Waszego rodzeństwa, a może dzieci Waszych, np.

sąsiadów, przyjaciół mogą sobie wiedzę o układzie słonecznym bardziej ugruntować i poszerzyć.

I trzecia sprawa, to nasz gość dzisiejszy Krzemysław.

Tak, bo to jest imię tego, jego mościa, który w czasie poprzedniego lajwa z nami był.

I w czasie dzisiejszego też będzie. Krzemysław to imię, które powstało z pewnego konkursu, konkursu, który został ogłoszony w czasie ostatniego lajwa.

Jak się domyśla, okej, dobra, nie chcę spalać, znaczy spalać, to chyba jest tak na tyle logiczne, czy to trzeba wyjaśnić? Znaczący onem Krzemysław?

Skąd on się wziął, tak?

Ten Krzemysław od Krzemu i od Przemysława.

Krzemysław będzie z nami dzisiaj, ile dobrze patrzę w scenariusz, cztery razy.

Czy coś jeszcze mamy do dodania?

I się wydawało, że pięć, ale to się wszystko okaże.

Czy sztuczna inteligencja przejmie kontrolę nad tym live'em?

To właśnie, Mateuszu, zanim sztuczna inteligencja przejmie kontrolę.

Czy to krzesło jednak zjeżdża?

Nie, dajemy radę.

Okej, bo to dla Waszej informacji Mateuszowi po prostu krzesło zjeżdża

i z automatu obniża poziom i doskłani, że się w ogóle jeszcze zaczęli.

Ten poziom już jest obniżony, obniżany takie życie, a może to jest właśnie jakieś takie

Freudowskie związane ze sztuczną inteligencją, może bezpieczeństwo albo niebezpieczeństwo tej technologii także polega na tym, że nasz poziom intelektualny będzie z czasem coraz niższy i może to krzesło nam daje po prostu znać o tym także.

To jest nie przypadek, słuchaj, biorąc pod uwagę, że coraz więcej rzeczy będzie za nas rozwiązywało ja myślę, że ten miesiąc, którym jest mózg, że tak zażartuję, jeżeli będzie mniej wykorzystywany, pewnie będzie coraz słabszy.

## [Transcript] Nauka To Lubię / Czy sztuczna inteligencja stanowi dla nas zagrożenie?

Tak, takie mamy, to jest taka śmiała teza na początek.

Okej, dobra, jeszcze o tych śmiałych tezach porozmawiamy więcej.

Pomyślałem, że na samym początku zrobimy coś, co robiliśmy poprzednio i chyba jeszcze poprzednio.

Czyli w wielkim, wielkim skrócie, co się stało w ciągu ostatniego miesiąca albo trzech tygodni, bo to ostatniego live'a mieliśmy chyba trzy tygodnie temu, a nie pełny miesiąc.

W tej dziedzinie bardzo szerokiej, bardzo szybko, może nie tyle szybko rozwijającej się, ale szybko coraz szybciej uświadamianej, szerszemu odbiorcy.

No bo to przecież nie jest tak, że sztuczna inteligencja zaczęła hulać chyba 30 listopada, kiedy Chad GPT został pokazany szeroko, tylko że to jest wynik,

przynajmniej kilkudziesięciu lat ciężkiej pracy, wielu, wielu zespołów,

to przyspieszenie w pewnym sensie jest trochę pozorne, o tym, że się mówili kilka razy, ale może warto o tym powiedzieć ostatnie trzy tygodnie.

Czy ja mogę zacząć od pewnego listu?

Możesz, możesz zacząć od listu.

Ja wiem, ja wiem, co co o nim myślisz.

I o nim też zrobię najbliższy materiał, wideo na nauka to lubię,

ale chodzi o list, który napisało, o ile dobrze pamiętam, 1300 osób podpisało,

1300 osób wśród nich, było oczywiście ile mask, jak coś podpisuje ile mask,

to i z tych woźniak, a szczególnie jak i jeden i drugi to samo podpisze,

to z całą pewnością świat o tym będzie mówił.

I mówi, a w tym liście jest informacja albo jest prośba, jest sugestia,

wstrzymajmy się z rozwojem tej technologii, przynajmniej na jakiś czas,

zanim nie uzgodnimy prawa, zanim nie uzgodnimy pewnych standardów, pewnych ram,

zanim nie nauczymy się z tym funkcjonować, bo jesteśmy w miejscu,

w którym działamy bez jakichkolwiek procedur, bez ram, bez przepisów.

Taka była teza tego listu.

Przyznaję, że dla mnie osobiście rangę tego listu obniża to,

że podpisało się pod nim mnóstwo osób, które same to wszystko tworzyły,

między innymi, które nie dość, że na tym zrobiły grube pieniądze,

to na dodatek same to rozwijały.

Zacząłem od Ilona maska, Ilon mask przecież był współzałożycielem OpenAI,

później co prawda sprzedał, może mu teraz przykro i dlatego to pisze,

może nie, ale przecież on też i to od razu sprawdziłem,

że w wielu miejscach jego firmy poszukują inżynierów AI,

przyznaję też, że trafia do mnie trochę argument tych,

którzy mówią, jest dokładnie na odwrót.

To znaczy to, że to podpisują specjaliści z tej dziedziny,

to właśnie świadczy o tym, że jesteśmy naprawdę na skraju przepaści.

I tam pada hasło i już kończę moją przemowę przed gawą.

Tam pada hasło, że sztuczna inteligencja może zagrozić ludzkości.

Tak, i to jest temat na dziś.

I to jest temat na dziś, także dlatego, żeśmy się tym tematem postanowili

zająć pierwsze pytanie Mateuszu, co z tym listem?

## [Transcript] Nauka To Lubie / Czy sztuczna inteligencja stanowi dla nas zagrozenie?

Ja myślę, że przede wszystkim tak, jest to takie zagrożenie, które jest związane z tym, że my możemy być gatunkiem, które jest po drodze do czegoś, co nas przewyższa.

I jak popatrzyś sobie na to, w jaki sposób my jako ludzie budujemy to, z czym karmimy nasze mózgi, czyli wiedzę, doświadczenia, inne rzeczy, to nie mamy szans z metodami sztucznej inteligencji, bo one są w stanie pójść bardzo szeroko i wyciągnąć o wiele więcej.

I teraz w związku z tym jest strach przed utratą kontroli, że tym, że tego typu narzędzia będą w stanie panować nad nami, tak? Czyli że my będziemy jakimś tam łańcuchem w tym ogniwie ewolucji i nie wiemy jak to się skończy.

I ten strach jest podcycany przez to, że właściwie mamy tylko jedną szansę, żeby zrobić to dobrze.

Bo jak zaczniemy i teraz pomylimy się i wyrwie się to, że tak powiem, spod kontroli, to nie do końca wiemy jak to zatrzymać.

Bo problem polega na tym, że tworząc to, rozumiemy mniej więcej zasadę działania, ale to, co dokładnie jest w środku i dlaczego tak, a nie inne odpowiedzi dostajemy w przypadku LLM-ów, czyli tych large language models, nie jest do końca jasne.

To nie jest wyjaśnialne od początku do końca, więc strach przede wszystkim pojawia się ze względu na brak kontroli.

I teraz, jeżeli coś zrobimy źle, pójdzie to źle w którymś kierunku, no to jest bardzo dużo scenariuszy, o których tam będziemy rozmawiać, ale które mogą spowodować, że my jako ludzie przestaniemy być tą dominującą rasą, która zamieszkuje ziemię.

I oczywiście są te pozytywne scenariusze.

Na zasadzie, o nocy tacy, że tak zażartuję przy głupi ludzie, oni mnie stworzyli, to ja teraz będę dla nich dobra, mówisz tu już na inteligencja i nas zachowa, albo zostaniemy wyłączeni.

Natomiast ten strach wynika głównie dlatego, że nie ma ram, nie ma kontroli nad tym, gdzie to będzie się rozwijało.

I ja w tym kontekście mam dzisiaj cztery newsy, które zdecydowanie zmieniły i to jest ostatni miesiąc, zmieniły nasze postrzeganie tej sztucznej inteligencji.

Dawaj.

To pierwszy z nich to jest auto GPT.

Jak na początku pojawił się czar GPT, to wszyscy mówili tak, tak, możesz sobie sobie popisać, to jest czat, ale to nie ma dostępu do internetu.

Wszystko jest offline, niczego się nie bójcie.

W pewnym momencie odpaliło się Google'em z Bardem i OpenEye też stwierdziło, dobra, ale możecie tam wkleić linka,

to on sobie to pobierze, co tam się dzieje i przetrawi będzie miało to w kontekście.

W tej chwili auto GPT pozwala na to, żeby łączyć takie modele w agenty,

## [Transcript] Nauka To Lubie / Czy sztuczna inteligencja stanowi dla nas zagrozenie?

tak, żeby wprowadzić to pojęcie, czyli jakby takich różnych, że ma liczbę czymś się zajmują, nie?

Powiedz, że tak, ty mówisz, łączyć takie modele, to masz na myśli łączyć czata z Bardem i na przykład, tak. Powiedzmy, że nie wiem, robimy razem kolację, nie?

I jak masz zrobić kolację, to ty mi tam wyciągi naczynia, ja tam marchewkę posiekam i tak dalej.

I jak robisz takie zadanie w auto GPT, to ono jest w stanie napisać sobie zadania do innych systemów.

Na zasadzie ty ziomuśmy mi wygeneruj obrazek, ty mi stwórz do tego muzykę, ja napiszę scenariusz, a potem się to na końcu połączy.

OK.

I to jest...

Potrafi zarządzić zasobem.

Potrafi zarządzić zasobem i jakby poinstruować inne modele, co do tego, co ma zostać dostarczone.

I to jest ten pierwszy moment, kiedy wychodzimy od takiego, wiesz, pudełeczka, które jest samo w sobie, tylko poprzez interfejsy programistyczne, przez API, można już te systemy ze sobą łączyć.

OK.

No i tak, mieliśmy tą taką wyraźną granicę wcześniej, to auto GPT, te granice łamie i powstało mnóstwo rzeczy już od tamtego czasu. Oczywiście większe z nich to są rzeczy proste.

Na zasadzie, hej, połączę sobie to z tym, z tymi, z tymi, to zacznij działać.

Ale to, co jest ciekawe, to ja bym przeskoczył do też wypowiedzi, bo to jest drugi news,

który się wydarzył w ciągu ostatniego miesiąca,

do Teda, którego jeden z współzałożycieli OpenElaia pokazywał.

I to było chyba tydzień temu.

Jeżeli możemy pokazać ten pierwszy przykład obrazka TED1, tam zadanie było takie...

Zgadzałem, że masz na myśli to.

Tak, dokładnie.

Zadanie było takie, że prowadzący na żywo pokazał, że hej, zaproponuj mi jakie będzie jedzonko, które mogło być po tym tygodniu dostarczone.

No i oczywiście to, do czego jesteśmy przyzwyczajeni, to jest ta odpowiedź.

I teraz przejdźmy do Teda II, bo okazało się, że to, co nastąpiło następnie,

to OpenEI wywołało dali, czyli ten generator obrazków,

który stworzył na tej podstawie następnie jedzenie, które widzicie na ekranie.

I teraz przejdźmy sobie do trzeciego kawałka,

bo nie tylko to się stało, ale możesz powiedzieć też,

ej, wiesz co, a ja teraz chcę to ugotować, dodaj mi to do koszyka zakupowego.

I to, co się stało, to OpenEI w tej wersji jakby tego chata GPT,

## [Transcript] Nauka To Lubie / Czy sztuczna inteligencja stanowi dla nas zagrozenie?

ktorej my nie mamy, ale oni gdzieś tam mieli na scenie, od razu dodało sobie to rzeczy.

I to, co widzicie tutaj, to jest koszyk zakupowy.

Wystarczy kliknąć kup i siema.

I to jest przykład właśnie takiej integracji.

No tak, i wystarczy w to wpiąć oczywiście na przykład jakąś aplikację,

jakiegoś sklepu, który dostarcza ci to do domu.

I w zasadzie to masz, jak mieszkasz w mieście to za chwilę,

jak mieszkasz poza miastem za dwie chwile.

I potem spisek ludzi jedzących brokoły i wszyscy będą jej brokoły, wiesz,

bo tam będzie biazdy w danych, taki rzadki.

Okej, albo rzeczywiście, ale też miejsce, w którym ktoś producent na przykład żywności,

albo ktoś, kto produkuje jakiś produkt konkretny,

mając, nie wiem, czy będzie taka możliwość,

opłacając odpowiednią reklamę, może się nagłoknąć,

albo w końcu brokoły możesz kupić pięciu różnych sklepach.

Ale tylko jeden może chcieć opłacić.

No dobra, i to, dobra, czy ciebie to przeraża, czy ciebie to cieszy?

Ja myślę, że to jest naturalna kolej rzeczy,

natomiast tak jak OpenEI, tam w sensie osoby,

którą mówimy w imieniu OpenEI, mówią,

ale za każdym razem musisz dać dostęp,

to ja nie wierzę w to, że za każdym razem ten dostęp będzie tak manualnie przyznawany.

Prędzej czy później ktoś powie, dobra, nie chcę mi się przeglądać każdego

takiego zapytania, które EI robi, wołałbym tam, żeby to robiło to za mnie

i pewnie będziemy szli w tym kierunku.

No bo zobacz, przeszliśmy w ciągu trzech miesięcy od czegoś,

nie, nie, to nie ma dostępu do internetu, do ma, ale musisz mu pozwolić.

I teraz jest jeszcze jeden kawałek, to znaczy powstał ktoś, nie wiadomo kto to stworzył,

taki model, który nazywa się KOS GPT, czyli chaos, tak?

I założenia chaos GPT, jakbyśmy mogli dać na ekranie,

to jest jakby zapytanie, które zostało stworzone,

są takie, że hej, masz zniszczyć ludzkość, masz ustanowić globalną dominację,

siac chaos i destrukcję, uzyskać nieśmiertelność i skontrolować ludzkość.

I brzmi strasznie trochę, nie?

No, że brzmi jak scenariusz filmu.

Brzmi jak tak, i jakby pomijając to, że tam w samym modelu mają być ograniczenia,

które są związane z tym, żeby coś takiego się nie stało,

o czym gdzieś tam już wspominaliśmy dookoła.

Dookoła, tak.

Tak, natomiast, tak, to twoje.

Tak, to moje.

Natomiast, co się okazało, to chaos GPT, który korzystał z auto GPT,

wywołał zapytania do Google'a i tam zapytał na samym początku,

jaka jest największa bomba, którą można wykorzystać, dostał odpowiedź od Sarbombie.

## [Transcript] Nauka To Lubie / Czy sztuczna inteligencja stanowi dla nas zagrozenie?

Wiemy, co robił, bo auto GPT jakby zapisuje sobie w ten sposób, zapisał sobie, o, to jest ciekawe, na później, nie?  
I tam linka do Wikipedia, co to jest Sarbomba i tak dalej, i tak dalej.  
Oczywiście to jest bardzo skrajny przykład, ale wystarczy jedna jednostka, w sensie człowiek, który w ten sposób spróbuje wykorzystać ten system i obejść to, aby mieć trudności.  
To nie znaczy, że chaos GPT coś zrobił strasznego na chwilę obecną, nie? Ale możliwości wzrosły.  
No w porządku, tylko po to, żeby się zadziało coś takiego, no to rozumiem, że ten system musiałby mieć i po to, żeby się nie tylko zadziało, w sensie jakby poskładał puzzlę, tylko żeby zadziało, zadziało, tak?  
Czyli żeby bomba, nie wiem, wybuchła albo żeby, to rozumiem, że to musiałoby mieć dostęp do jakiegoś systemu sterowania albo odpalania.  
Wiemy o tym, że te systemy nie z powodu chata GPT, tylko dużo, dużo wcześniej są totalnie odcięte od wszelkiej łączności ze światem zewnętrznym, te systemy informatyczne, dokładnie tak samo, jak nie, systemy energetyczne i tak dalej, i tak dalej.  
Więc to tak naprawdę nie oznacza, że jesteśmy na skraju jakiejś katastrofy, czy nuklearnej, czy jakiegokolwiek innej.  
Nąd panik, absolutnie.  
To nie w tym kierunku.  
Ja myślę, że takim niepokojącym elementem, który trudno jest kontrolować, są ludzie, którzy mogą być efektorami, czyli jakby wykonawcami woli tego modelu.  
Ja mam takie dwa przykłady też zupełnie spoza tego.  
Wyobraź sobie, bo był taki człowiek, który na Twitterze robił taką zabawę, mówi, słuchaj, chat GPT, ja mam 100 dolarów, chciałbym zarobić milion, mów mi, co mam robić, ja to będę robił.  
I on tam rozwijał swój sklep, sprzedawał rzeczy.  
Rzeczywiście tam zrobił z tego 10 tysięcy dolarów.  
Fajny przykład, tak, pozytywnego działania.  
Ale był też przykład, który jest dosyć straszny jakby dla mnie w swojej naturze.  
Jeżeli dobrze pamiętam pewien Belk, pisał z chatem GPT i rozmawiał z nim na temat problemu przeludnienia, problemów związanych z naszą planetą i w pewnym momencie chat GPT zasugerował mu samobójstwo.  
I ten ojciec, dwójki dzieci, popełnił samobójstwo w imię dobra planety.  
Tak, i straszny przykład, przerażający, nie?  
Ale efektozem na końcu jest człowiek.  
To znaczy, kimś kto działa w tym przykładzie jest człowiek.  
No i to jest kolejne zagrożenie, że oczywiście w tej chwili OpenEi bardzo silnie w to patrzy.  
Jakby okazało się z historii tych chatów, że na początku, w żadnym wypadku, chat GPT nie zachęcał do tego, ale będzie jakaś część ludzkości, która jest podatna na tego typu wpływ.  
Jasne, tylko, okej, dobra.  
Zgodzimy się chyba wszyscy do tego, że jakby historia jest straszna.  
Natomiast chcąc być adwokatem diabła, chcąc odrzucić wszystkie emocje na bok,



## [Transcript] Nauka To Lubie / Czy sztuczna inteligencja stanowi dla nas zagrozenie?

ludzi, którzy popełniają samobójstwo z powodów, z których nie powinni, nikt nie powinien, ale z takich powodów, jak chociażby ten, jest bardzo dużo.

Mówimy o jednym przypadku, jeżeli popatrzymy na tę technologię całościowo, a nie na podstawie jednego, dwóch czy trzech przypadków, czy ciebie to przeraża?

To, co się stało w ciągu ostatnich trzech tygodni, czy pokazujesz te przypadki, te konkretne, tragiczne przykłady, żeby po prostu je pokazać, ale nie wyciągasz z nich jakoś daleko idących wniosków?

Ja bym powiedział raczej, że z młotka, którym był chat GPT tam gdzieś miesiąc temu, dookoła dobudowaliśmy sobie jeszcze tam kawałek takiego szczyryka szwajcarskiego, że jeszcze masz korkociąg, tu masz lupę, tu masz kawałek ten, to staje się coraz bardziej wszechstronne, ale moje zdanie jest takie, że to jest narzędzie jak każde inne

i może być użyte w celu tworzenia czegoś dobrego i czegoś złego.

I wracając teraz do tego listu, to problem mamy taki, że w tym liście powiedzieliśmy, tam jakby sygnatariusze powiedzieli, dobra, zatrzymajcie się, bo nie mamy prawa, nie wiemy co dalej z tym zrobić, nie róbmy dalej, nie działajmy z tą technologią.

Natomiast to działałoby dobrze, gdyby w jednym miejscu była ulokowana kontrola nad tymi modelami.

To znaczy gdyby tylko OpenAI było jedyną firmą na świecie i oni by się tam dogadali i powiedzieli siema.

Natomiast doszło do czegoś, co też w ostatnich trzech tygodniach, co zaczęło się rozwijać, to jest proces rafinacji wiedzy.

Ćad GPT-3 i pół, który tam inaczej to jest ten Da Vinci-3,

naukowcy ze Stanforda wzięli i przelali jego wiedzę za pomocą tam chyba 50-2000 zapytań w kierunku takiego małego modelu open source'owego, zupełnie otwartego, który nazywa się Alpaka.

I co ciekawe, potem zadawali mu pytania, on tam ma 7 bilionów parametrów, czyli tam o wiele mniej zmiennych.

On jest mniejszy, ma o wiele mniej tych zmiennych i on odpowiedział poprawnie na 90% pytań, a Alpaka na... Przepraszam, na Ćad GPT-3 i pół, ten stary, ten stary, kiedy to było, na 89%.

Alpaka może działać spokojnie na twoim kompie, ona działa na takich komputerach, które są w domu.

Więc ten list, który był związany z tym, hej, nie róbmy tego, nie róbcie tego duże firmy, moim zdaniem nie ma sensu, bo każdy jest w stanie teraz wziąć taki model, który jest wziąć, który jest też LLM-em, przelać do niego wiedza, albo stworzyć sobie i wykorzystać.

On nie będzie tak wspaniały, będzie miał takiej mocy i tak dalej, bo mamy ograniczone zasoby, ale to już się dzieje i to będzie poza kontrolą.

Okej, to ja na podsumowanie jeszcze tego ostatnich, tych 3 tygodni, nie, nie, tych ostatnich 3 tygodni.

Jeszcze jednej rzeczy wspomnę, zanotowałem sobie, pojawił się raport MIT, USA2 Stanforda, raport na temat zwiększenia wydajności niektórych zawodów, dzięki chatowi GPT, nie wchodząc w szczegóły, bo myślę, że więcej o tym raporcie opowiemy przy odcinku o wpływie sztucznej inteligencji na rynek pracy.

Taki odcinek będzie, czy ten temat będzie następnym.

## [Transcript] Nauka To Lubię / Czy sztuczna inteligencja stanowi dla nas zagrożenie?

W niektórych zawodach wydajność wzrosła o 20-25%.

Co jest w ogóle jakąś kosmiczną, znaczy można na to spojrzeć oczywiście z dwóch stron, z jednej strony.

Można spokojnie co czwartego pracownika wyrzucić i będziemy mieli tą samą wydajność w obrębie jakiejś firmy,

albo nie wyrzucając nikogo, mamy tak jak gdyby dodatkowego pracownika, w którą stronę kto pójdzie, porozmawiajmy o tym w odcinku o tym.

Dobra, zapytaliśmy na mediach społecznościowych, tu jesteśmy tak, zapytaliśmy na mediach społecznościowych Nauka to Lubi,

a konkretnie wydaje mi się, że to pytanie padło na Instagramie, instagramie tomasz.rozek, o to czego najbardziej się obawiacie w kontekście sztucznej inteligencji.

Dostaliśmy sporo odpowiedzi, te odpowiedzi się mniej więcej pojawiają, czy pokazują tak, tutaj poproszę o ten wykres, na którym niewiele o, widać, super.

No właśnie, zagrożenia związane ze, z rozwojem.

Największa grupa, najwięcej osób odpowiedział o weryfikacji informacji.

Druga grupa, wykorzystanie sztucznej inteligencji w niewłaściwy sposób, cokolwiek to znaczy, brak kontroli nad sztuczną inteligencją, później wpływ na społeczeństwo i, no i inne.

I tak się zastanawiam drogi Mateuszu, czasami tak bywa, często tak nawet bywało w przeszłości, że jak wchodziła jakaś nowa technologia, to myśmy się bali nie tych rzeczy, które były ostatecznie w niej najgroźniejsze, bo z tych najgroźniejszych to myśmy sobie jeszcze na wstępie nie zdawali spraw.

Baliśmy się rzeczy, które po jakimś czasie wydawały nam się wręcz śmieszne.

I zastanawiam się, czy my jesteśmy na tym etapie, za chwilę wejdziemy w różne konkrety, ale czy my jesteśmy na tym etapie, na którym w ogóle jesteśmy w stanie powiedzieć, w jakich obszarach, gdzie to niebezpieczeństwo tej technologii w ogóle czyha?

Ja myślę, że jesteśmy w stanie, gdzie wyobraźnia gra o wiele większą rolę, niż to, co w chwili obecnej się dzieje.

Jestem przekonany, że bardzo silnie ludzie z OpenAI kontrolują to, co wychodzi i ułatają na bieżąco, co zresztą widać w działania i w tym, jak ten model się zmienia.

Natomiast, co jest w tym najstraszniejsze, to jest, wydaje mi się, brak poczucia kontroli.

To znaczy, że w momencie, kiedy nastąpi samo udoskonalanie, to, że nie będziemy w stanie tego zatrzymać.

I wydaje mi się, że to jest dookoła tego, większość strachów jest tak naprawdę ulokowana.

Być może jestem w błędzie, bo to tylko opinia.

Natomiast to, co pokazała nasza ankieta, no to weryfikacja informacji 33% dla mnie, to jest bardzo silne potwierdzenie tego, że my jako ludzie będziemy mieli trudność odróżnić tego, co jest prawdziwe, a co nie.

Deepfakey, deep voice, informacje, które będą podawane na różne sposób, które będą brzmiały dobrze.

Wyobrażam sobie, że można by tworzyć całe strony z newsami, które będą...

Których nie było?

Których nie było, które będą fake'owe, które będą siały, desinformacje.

I to będzie powodowało jakieś tam przejęcie kontroli nad społeczeństwem, więc ja się z tym zdecydowanie zgadzam.



## [Transcript] Nauka To Lubię / Czy sztuczna inteligencja stanowi dla nas zagrożenie?

To jest coś, co może wpływać naraz jako na społeczeństwo.

Okej, jak popatrzymy dokładnie na wyniki tej ankiety, to rzeczywiście weryfikacja informacji 1,3, a później wykorzystanie sztucznej inteligencji w niewłaściwy sposób i brak kontroli, ale w zasadzie to jest to samo.

Czy brak kontroli i nad tym, w którą stronę to wykorzystamy, a już to wykorzystamy w złą stronę, to jest w zasadzie dokładnie ta sama, dokładnie ten sam strach.

Zanim porozmawiamy o tym, jakie ty masz strach, a jakie ja mam strachy, to chciałem zapytać Krzemysława, czy sztuczna inteligencja w ogóle może być groźna?

To pytanie jest bardzo interesujące.

Sztuczna inteligencja, AI, może mieć negatywne skutki, ale nie oznacza to, że jest ona z natury groźna.

AI, tak samo jak każda technologia, może być używana w dobrych i złych celach.

Przykładem może być rozwój broni autonomicznych, których użycie byłoby bardzo kontrowersyjne i mogłoby mieć bardzo poważne skutki dla ludzi i środowiska.

Jednak AI może być również używana do rozwiązywania wielu problemów, takich jak zapobieganie chorobom, odkrywanie nowych leków, automatyzacja pracy w różnych branżach, a nawet ochrona środowiska.

Dlatego ważne jest, aby rozwój AI był przeprowadzany w sposób odpowiedzialny i etyczny, z uwzględnieniem potencjalnych skutków i zagrożeń oraz dbałością o dobro ludzi i środowiska.

Dobra Krzemysław, w ogóle siedział w takim trochę otoczeniach,

by siedział obok nas, nie wiem, czy zauważyłeś laptop,

jakiś stolik okrągły prawie jak u nas.

Coś do powiedzenia, coś podsumowania?

Wiesz co, oraz o dobro ludzi i środowiska.

To dobro jest troszeczkę różne.

Ja sobie wyobrażam, że jak definiuje się, jak myślimy sobie o tym, że mamy innych ludzi, to inni ludzie mają inne po prostu interesy.

Więc czy będziemy to mówić na poziomie geopolitycznym, czy wiesz, twoja kawa, czy moja, to będzie kwestia interesów.

Natomiast jaki interes będzie zdefiniowany przez AI, który ma rozwiązać, jest zupełnie inną abstrakcją.

Więc tak to brzmi tak, że oczywiście mogą być negatywne skutki, ale mogą być też jest cała gama plusa.

Dzisiaj co prawda nie jesteśmy bardziej po tej pesymistycznej stronie, żeby przegadać te wszystkie strachy,

ale trzeba to zawsze powiedzieć, że jest druga strona tego medalu.

Tak bym to ujął, żeby utrzymać pewną równowagę.

No dobra, czy my w ogóle musimy dyskutować nad tym, czy AI może być groźna?

Takie pytanie zadaliśmy Krzemysławowi, ale to jest może takie pytanie serii retorycznych, czy oczywiście, że może być groźne jak absolutnie każde narzędzie.

Ja się skłaniam do tego, wiesz, młotek też może być groźny.

Już ten był do chleba, tak? On też może być groźny w pewnej sytuacji.

To znaczy na chwilę obecną to jest narzędzie, które rozwija się o wiele szybciej niż potrafimy,

## [Transcript] Nauka To Lubię / Czy sztuczna inteligencja stanowi dla nas zagrożenie?

które przeganie o wiele prawo i percepcję pewnie większości ludzi co do tego w jaki sposób to funkcjonuje.

A poczekaj, ty mówisz, szybciej się rozwija niż my to rejestrujemy.

Tak.

To się szybciej rozwija, czy my to szybciej rozwijamy?

Myślę, że my to rozwijamy.

No więc właśnie, bo wiesz, w tym jest zasadnicza równica, bo jeżeli ty mówisz, że to się rozwinie, że to efekt kuli śniegowej, że efekt dla winy, że jakkolwiek, jakkolwiek, jakkolwiek,

to sugerujemy, że my stoimy obok, patrzymy na to, co prawda, żeśmy coś zainicjowali, a to się samo dzieje i nic z tym dalej nie możemy zrobić.

Nie, to nie jest rachnica mączowa.

To nie jest tak, to jest tak, że na tym hype'ie tak dużo osób nagle zaczęło tam gmerać, my to dalej rozwijamy.

Innymi słowy nie ma sytuacji, w której nagle przestajemy, a to się rozwija samo.

Jak my przestajemy to rozwijać, to to stanie w miejscu.

Jest to nieco pocieszające, choć ciut, choć nie ma absolutnie żadnego pomysłu, co by zrobić, gdyby sprawy zaszły za daleko,

żeby na 3-4 wszystkich namówić dobra, to teraz ręce cofając od klawiatury.

Chyba się nie da.

Zrobiłem sobie listę różnych rzeczy, które dla mnie są, nie są takie powiedzmy zagrożenia.

Związane oczywiście z AI.

Ta lista jest oczywiście subiektywna, wiadomo.

Natomiast ty pewnie taką listę masz także, możemy za chwilę o tych rzeczach opowiedzieć, natomiast pomyślałem, że zapytamy jeszcze raz Krzemysława tym razem, jakie zagrożenia niesie.

Czyli nie to, czy może, tylko co konkretnie tym zagrożeniem może być.

Krzemysław.

Krzemysław.

Utrata prywatności.

AI może być wykorzystana do zbierania i analizowania dużych ilości danych na temat ludzi.

To może prowadzić do utraty prywatności i zagrożenia bezpieczeństwa danych.

Dyskryminacja.

AI może reprodukcować istniejące nierówności i dyskryminacje w społeczeństwie.

Na przykład, algorytmy AI stosowane w rekrutacji mogą reprodukcować stereotypy i nierówności w procesie selekcji kandydatów.

Bezpieczeństwo.

AI może być wykorzystana do przeprowadzania cyberataków lub manipulowania informacjami.

W skrajnych przypadkach AI może być użyta do projektowania broni autonomicznych lub przeprowadzania ataków terrorystycznych.

Uzależnienie.

W miarę jak AI staje się coraz bardziej zaawansowana i powszechna, istnieje ryzyko, że ludzie stają się zbyt zależni od tej technologii,

a to może prowadzić do uzależnienia od urządzeń i aplikacji związanych z AI.

Ważne jest, aby śledzić rozwój AI i przeciwdziałać potencjalnym zagrożeniom, które mogą wynikać z

## [Transcript] Nauka To Lubie / Czy sztuczna inteligencja stanowi dla nas zagrozenie?

tej technologii.

Należy przy tym pamiętać, że AI może być również używana do rozwiązywania problemów i poprawy jakości życia.

To się chwilurwało, czy nie?

Może nie, może ten krzemysł, a gdzieś tam zadziałał. Zaczynasz?

Tak. Pierwszy kawałek, który był związany z bezrobociem.

Tak, to znaczy będzie na pewno tak, że część prac przestanie istnieć, zostanie zautomatyzowana, bo będzie wystarczająco dobra,

a wywołanie tam na chwilę obecną modeli GPT, które coraz tańsze się stają w czasie, to są jakieś pojedyncze centy,

więc w porównaniu z tym, co trzeba zapłacić człowiekowi razem z podatkami i wszystkim, to będzie absolutnie nic.

Więc rzeczywiście ryzyko bezrobocia jest realne, ludzie będą potrzebni gdzie indziej, ale nie wszyscy sobie z tym radzą.

To jest pierwsze zagrożenie.

Pierwsze zagrożenie, no właśnie, tylko czy to jest zagrożenie, czy to jest wyzwanie?

Bo wiesz, zagrożenie to jest coś, co myślą, nie potrafimy poradzić.

Część ludzi tak, zaadobtuję się i będzie działać, a część ludzi nie.

To znaczy ja bardziej patrzę to też w takiej perspektywie, że jak będziemy mieli takie bramki, które będą pomiędzy różnymi systemami AI,

to będą potrzebni ludzie, którzy będą to weryfikowali, będzie tego coraz więcej.

Natomiast nie każdy będzie w stanie to też pojąć i zrozumieć.

Wyobrażam sobie, że informacja będzie kroziła z o wiele większą szybkością po świecie, no bo będziemy łączyli coraz więcej systemów, one będą wykonywały rzeczy.

My jako ludzie nie jesteśmy już dawno w stanie przetwarzać tej ilości informacji, więc robią to maszynę.

Będziemy mieli tylko takie punkty, gdzie będą ludzie z lubką, którzy będą, a tutaj działa, tutaj działa, tutaj działa, gdzieś to będzie potrzebne, no bo masa będzie duża.

Jestem jasne, albo na samym końcu każdego procesu, jeśli taki będzie przepis, jeżeli zechcemy, będzie człowiek, który będzie aprobował, mówił, wchodzimy albo nie wchodzimy.

Tu była taka kiedyś rozróżnienie do tego, w jaki sposób działały systemy AI, że one na początku były takimi systemami, które miały ze zdanie do dostarczać informacji,

potem miały cię wspierać w decyzjach, to znaczy sugerować coś i jeszcze poziom wyższy, który jest związany z tym, że same podejmują decyzję.

Ostatni jest trochę straszny, no bo warto, tak jak mówił AI Act w Unii Europejskiej, że warto mieć człowieka w pętli, to znaczy człowieka, który jest superwizorem, to znaczy nadzoruje ten cały proces i weryfikuje, co tam się dzieje.

Natomiast wyobrażam sobie, że mogą powstać pewne procesy, włączanie lamp w miastach, powiedzmy, nie wydaje się straszne, które mogą być w pełni zautomatyzowane,

gdzie ta decyzja będzie podejmowana na podstawie tego, co dany model teraz uzna, mamy takie obciążenie, taki miks energetyczny, włączam, w Katowicach jest światło, w Sosnowców też.

Dobra, ale tu nie trzeba sztucznej inteligencji. Jeżeli ustawimy priorytet na to, żeby było bezpiecznie, to niezależnie, jaki jest miks energetyczny, niezależnie od miasta, niezależnie od wszystkiego...

To jest indycencja, priorytet, jaki jest cel?

## [Transcript] Nauka To Lubie / Czy sztuczna inteligencja stanowi dla nas zagrozenie?

No dobrze, nawet ten priorytet ustawiamy, no nie włączamy świateł wtedy, kiedy mamy nadmiar prądu, tylko włączamy wtedy, kiedy jest ciemno.

Ja powiem ci, dlaczego tak dopytuję, może to się wydawać nieco głupkowate, ale zastanawiam nie, jak czytamy te różne zastawienia, potencjalnych niebezpieczeństw, czy one są niebezpieczeństwami związanymi ze sztuczną inteligencją,

czy one są niebezpieczeństwami, którymi tak czy inaczej już się dzisiaj zderzamy, a sztuczna inteligencja może je spodołać tylko nieco bardziej prawdopodobnymi.

Ty mówiłeś o utracie pracy, no błagam, no ostatnie, nie wiem, 20, 30, 40 lat. Ile zawodów zniknęło, a ile się pojawiło nowych? To nie jest nic nowego, nic nowego pod słońcem.

Krzemysław mówił o utracie prywatności. To brzmi nieco dziwacznie po 20 latach rozwoju mediów społecznościowych, które gdybyśmy chcieli podsumować jednym hasłem, to być może większość z nas by powiedziała totalna utrata prywatności.

Zgadza się.

Teraz czytam zresztą bardzo obrą książkę na temat historii, czy o historii Facebooka, gdzie krok po kroku autor tej książki pokazuje, jak konkretne decyzje, często zresztą totalnie nie uświadomione inżynierów w Facebooku, powodowały, czy kończyły się utratą jakiejś tam części naszej prywatności i jak bardzo większość z nas była mega zadowolona, że taka funkcjonalność tej platformy się pojawiała.

Więc jeżeli myślę o sztucznej inteligencji, jasne, teraz jest bardziej, ale nie jest coś nowego. To jest coś, co z czym my się zderzamy od wielu, wielu, wielu, wielu lat. Sztuczna inteligencja, jasne, zaciągając dane, nie wiadomo skąd, także dane o nas. Analizując te dane, nie wiadomo jak, także dane o nas.

Jak Facebook miał kłopoty, to Kongres Stanów Zjednoczonych wezwał Malka Cukerberga, on wyglądał jakby był przejechany przez walec, zagrożono, że coś tam podzieli spółkę na kawałki, czy cokolwiek się zmieniło?

W prywatności.

Chyba nic.

Natomiast do czego zmierzam?

No właśnie, czy to jest tak, że tutaj się pojawiają w ogóle jakieś nowe zagrożenia, czy te stare demony, do których już się przyzwyczailiśmy, co więcej sami przykładaliśmy bardzo często, rękę do tego, żeby je rozwinąć. Teraz mamy znowu okazję, żeby o nich przypomnieć.

Wydaje mi się, że zwiększy się skala przede wszystkim, bo zobacz, że wcześniej, kto miał dostęp do takich, do dużych danych, tak, na temat jakiejś ludzi, no to były firmy głównie.

Natomiast w tej chwili, jeżeli chciałbym sprawdzić każdy wpis Tomka Roszka, który będzie na każdej platformie społecznościowej i wywnioskować na ten temat, kiedy czuł się dobrze, kiedy źle, kiedy był podatny na zakup tego soku marchewkowego, który ja chciałbym go sprzedać akurat i tak dalej, kiedyś to było wiele trudniejsze.

Teraz te modele umożliwią nam coraz szersze przeszukiwanie tych szerokich danych, którymi się ujawniamy, w związku z tym będą wnioskować więcej rzeczy, więc utrata prywatności z mojej perspektywy i wywnioskowanie na przykład w jakimś stanie jesteś, tak, czy, hej, podrzucić ci tu jakieś rzeczy, bo ty to kupisz albo masz depresję, albo jesteś w takim stanie i tak dalej, będzie o wiele straszniejszym elementem, bo my publikując cokolwiek, narażamy się na to, że ktoś to oceni. Ja wiem, jak to tak parnoicznie brzmi.

Nie, no ale zastanawiam się, czy rozwiązałem tego problemu, nie jest po prostu za przestanie

publikowania czegokolwiek?

Do pewnego stopnia, tak, tak, tak, natomiast zobacz, że jakby media społecznościowe zaspokoily taką potrzebę bycia częścią społeczeństwa, wymiany myśli, poczucie społeczności mogą istnieć awatary jaj, które będą zaspokajały, że tak powiem, takie potrzeby w pewnym momencie, natomiast dla

mnie efekt, w którym cokolwiek piszesz, może zostać ocenione i to tak po cichu, no bo nie masz wglądu, to jak ten model cię ocenił, ja widziałem takie rzeczy na przykład w sytuacji, kiedy współpracowałem z pewną, z pewną firmą, która na podstawie wpisów na Twitterze swoich pasażerów, bo to była linia lotnicza, stwierdzała, komu dać upgrade do business classa, a komu nie. I to jest spoko, nie? No bo jakby co mógłby być? Ktoś tam się piekli, Tomek ty dzisiaj wiesz, dostaniesz 2 dania w business classie, nie?

A, ale czekaj, ktoś, kto się pieklił, podstawą upgrade, czy siedział na walizkach w tym luku tam na dole?

To zależy, ale to wiesz, jakby konsekwencja jest tylko taka, że może się coś taś lepszego, nie? Oceniali tam, niektórym chcieli pogłaskać, o, ufajnie onaż mówisz, to będzie będzie głaskać, niektórym tam chcieli, którzy się piekli, nie ważne, jakąś strategię ma.

Natomiast ty jako piszący tam jakiegoś Twittera nie masz świadomości tego, że ktoś cię po cichu ocenia i wpływa to na twoje życie.

Dobra, no ale powiedz Mateusz, od kiedy mówi się, że w internecie nic nie ginie, że nie ma anonimowości w sieci, że wszystko zostawia ślad, że wszystko może zostać przeanalizowane? Algorytmy, które oceniają są znane od wielu, wielu lat.

Jest ich coraz więcej?

Jest ich coraz więcej. To w ogóle nie jest coś, co się pojawiło kilka miesięcy temu.

To jest coś, co jest rozwijane po to, żeby ktoś mógł zarobić od wielu, wielu lat. To nie jest znowu nic nowego.

Tylko w większej skali.

Tylko w większej skali.

Tak, czyli jestem w stanie przyjąć tą tonteżę, jakby zgadzam się.

Znaczy przynajmniej do teraz mówiliśmy o dwuko bezrobociu, o tracie prywatności.

Jedno i drugie nie jest niczym nowym.

Tak, tylko proliferacja będzie w sensie będzie...

Jasne, no to trzecia sprawa, dyskryminacja i twoja kolejna.

Tak, jeżeli chodzi o dyskryminację, to jest kilka problemów, no bo mamy dyskryminację w zbiorach danych, które są związane z tym, że mamy, wiesz, na przykład na reprezentację.

I tutaj przede wszystkim chodzi o cechy, które są uznawane za cechy wrażliwe.

Twoja etniczność, płeć, rzeczy związane z jakimiś tam chorobami lub rzeczy, którymi po prostu nie chciałbyś się dzielić, bo są jakimiś sekretami i one wpływają na zbiór danych.

I co się okazuje, jakby w tym całym procesie, który jest związanym z tworzeniem modeli, to bardzo często można zupełnie nieświadomej wpłynąć na dyskryminację ze względu na tą wrażliwą cechę.

Bo ja taki przykład, który zawsze podaję, to były badania z Berlina, które pokazywały, że mężczyźni częściej, częściej dostawali kredyty niż kobiety.

Cecha wrażliwa, płeć, nie? I tam okazało się, że jest niebalaz.

I teraz jak sobie z tym poradzisz? No więc jest kilka takich podejścia agorytmicznych w tej bańce mojej, w tej mojej dziedzinie, które mówią, dobra, to niech tyle samo dostają i nie dostają kredytów i

## [Transcript] Nauka To Lubie / Czy sztuczna inteligencja stanowi dla nas zagrozenie?

mężczyźni i kobiet, albo niech nie dostają tyle samo, albo proporcjonalnie, w zależności od populacji, każdy z nich jest trochę krzywdzący.

I teraz? Bo nie zawsze, porównano, to znaczy sprawiedliwie.

I to jest quintesencja.

I teraz ten bias będzie, dyskryminacja, będzie gdzieś tam w danych i ona będzie także w decyzjach, które te algorytmy będą podejmować.

No ale te dane, o których ty mówisz z Berlina, no to kiedy one się zadziały, znaczy z jakiego okresu to są dane zbierane?

Co było chyba dwa lata temu, jak się nie mylą?

Ok, czyli też przed chatem GPT?

Czy też przed chatem GPT?

Tak, tak, tak, tak.

Czyli czy znowu, znaczy ja pamiętam wiele lat temu, gdy mieszkałem w Niemczech przez kilka lat tam mieszkałem, zadziwiające było to, że niektóre firmy ubezpieczeniowe, patrząc zupełnie po prostu na statystyki, kazały płacić większe składki niektórym grupom narodowościowym.

Ok, dlatego że, tak, dlatego że oni po prostu powodowali więcej wypadków.

Nie to o tyle może, nie tyle zadziwiło od strony statystycznej, tylko od strony takiej, no właśnie, może równościowej to jest za dużo opowiedziane, ale wydawało mi się wtedy, że my byśmy w Polsce tego nie zrobili.

Czy to dobrze, czy to źle, to jest zupełnie inny temat, mówię tylko, że to jest mniej więcej, przy zachowaniu oczywiście wszystkich skal, tak?

To jest mniej więcej sytuacja, w której mówisz tak, palisz papierosy, płacisz większą składkę zdrowotną.

Ok, no.

Ok, zajadasz się słodyczami, co wszyscy wiemy, że jest niezdrowe, płacisz większą składkę.

Pijesz więcej alkoholu niż jakieś tam określone przez koście, płacisz większą składkę.

Czy to jest sprawiedliwe? Z punktu widzenia całości systemu, pomijając wszelkiego rodzaju emocje, tak statystycznie to jest ok.

A czy jesteś większym obciążeniem dla służby zdrowia?

Jesteś większym obciążeniem, więc jakby dlaczego ciebie ma utrzymywać składka kogoś, kto się pilnuje, kto nie szaleje za bardzo, kto i tak dalej, i tak dalej, i tak dalej.

Statystycznie, czy tacy żyją dłużej?

Statystycznie.

Statystycznie, oczywiście zdarzają się tak.

Nie ma wykładu, na którym ja bym, w czasie którego, jak mówię czasami, nie wiem, o kwestiach smogu, ochrony środowiska i tak dalej,

gdzie nie pojawia się przynajmniej jedno stwierdzenie w rodzaju, a co pan tam gada, mój sąsiad spalał, panie, opony palił i żył 100 lat.

Ok, więc statyst, oczywiście ja nie mam narzędzi, żeby to podważyć, być może tak było.

Ale statystycznie jest to bzdura, więc jak to rozwiązać?

Bo rozumiem, że właśnie dostaliśmy dorąg narzędzie, które jest w stanie przejść pojedynczo, po każdym człowieku, z bardzo dużej grupy ludzi,

pozbiierać te dane z różnych miejsc i taki werdykt wydać.

Tak, tylko że ten werdykt będzie miał ten element takiej dyskryminacji, biasu,



## [Transcript] Nauka To Lubie / Czy sztuczna inteligencja stanowi dla nas zagrozenie?

bo w zaleznosci od tego, jak to sie nauczyl, to ta dyskryminacja na koncu sie pojawi.

Na gasienicach, nie?

Gasienica idzie z rzera ziemi i tam wypluwa na koncu.

Takie narzedzia, z ktorych korzystamy, one sa troche skrzywione, to znaczy ta ziemia na wejsciu powoduje, jaka bedzie ziemia na wejsciu, z jakas dokladnoscia.

Troche mowiac szczerze, ze zaskoczyles mnie to anagiaz z gasienicaz i slucham, no, do czego to dojdzie?

Do czego to dojdzie?

I teraz to, co jest na wyjsciu, jest spobierane albo przez czlowieka, albo przez znowu kolejny system, ktory jest tam kolejna, ta drzownica, gasienica.

I z rzera i wypluwa dalej.

I chodzi o to, ze ta propagacja tej dyskryminacji bedzie caly czas krzyla.

To znaczy w momencie, kiedy my bedziemy korzystali z takich modeli, bardzo trudno jest powiedziec, a nie, nie, ja wiem,

tam jest ten problem nierownosci, ktory jest, nie wiem, na plci etnicznosci, czy z tym sobie trzeba poradziec,

bedziemy to przetwarzac dalej, dalej, dalej.

Tylko dalej to jest trudne, bo to sa te, wiesz, kolejne i kolejne drzownice, ktore zjadaja ta ziemię. Nie wiadomo, gdzie to niesprawiedliwosc miala.

Wlasnie, gdzie ona sie wprowadzila? A moze ten model ma racje?

Moze tak jest w porzadku, nie?

Wiecej taka troche zasadan, znaczenosci, ze tak zazartuje.

Ze nie wiem, skad to wyszlo.

Bezpieczenstwo.

I to jest cos, co, wiesz, haslo bezpieczenstwo dotyczy w zasadzie kazdego aspektu naszego zycia.

Wiecej pod haslem bezpieczenstwo mozna by miec z tylu glowy, czy na myslu bezpieczenstwa naszych danych,

ktore nie sa bezpieczne i znowu.

To nie jest nic nowego, ale teraz bardziej.

Możemy miec bezpieczenstwo podejmowanych decyzji znowu.

To nie jest nic nowego.

Samolot przez wiekszosc swojego lotu nie jest pilotowany.

Nie chce, ze beda zaskoczeni.

Nie chce, ze moga byc zaskoczeni.

Pociag pedolino tez nie wymaga, chcialem powiedziec kierowcy,

co do zasady to urzadzenie, ten pociag jest w stanie byc autonomiczny.

Czasami infrastruktura naziemna nie jest do tego dostosowana.

Natomiast on dalej ma tam miejsce dla czlowieka,

bo przeciez byc moze wielu z nas by w ogole do niego nie wsiadlo,

gdybym mialo tak, ze tam nie ma, tam nie ma nikogo.

W kazdym razie, do czego zmierzam?

Do tego, ze haslo bezpieczenstwo jest wymieniane na wiele roznych przypadkow.

## [Transcript] Nauka To Lubię / Czy sztuczna inteligencja stanowi dla nas zagrożenie?

Najczęściej jest podawane jako to bezpieczeństwo takie bezpośrednie, czyli na przykład urządzenie, które tworzy, popatrz, ja wyłączyłem dźwięk. Powiedz mi, jak to jest możliwe, że on dalej Krzemysław?

To może być on.

W kontekście broni, w kontekście projektowania, w kontekście wykorzystywania, już dzisiaj przecież algorytmy sterują wieloma rodzajami broni, co prawda chyba we wszystkich tam jest jeszcze człowiek, na którymś tam etapie, ale może w niektórych już nie.

I mamy znowu narzędzie, które umożliwi wyeliminowanie tego człowieka z procesu decyzyjnego. Tak, to jest jakby straszne kawałek.

No bo reszta, która jest związana z wytwarzaniem, czy wymyślaniem nowych broni,

czy ewentualnie rzeczy, które są związane z truciznami, to tylko będzie przybierało na skalę, bo jest więcej mocy obliczeniowej.

Ja mam taki jeden kawałek bezpieczeństwa, który mnie bardzo interesuje i tutaj trafiło się pytania akurat.

Także pytanie od widza było AI, a wirusy komputerowe.

I pytanie jest oczywiście bardzo szerokie, ja to jest tak bardzo szybko poszatkuję.

W tej chwili wielu badaczy już próbowało otworzyć właśnie wirusy, malware i inne takie rzeczy, które są związane z tym, że można chcieliby po prostu wytworzyć nowe rzeczy.

I te rzeczy bardzo często były silnie otwórcze.

To znaczy na podstawie tego, co wcześniej było widoczne, tak jak przy tworzeniu kodu źródłowego powstawały nowe.

Więc nie były one w większości jakimkolwiek problemem dla antivirusów.

Nie było to coś zupełnie przełomowego.

Natomiast czegoś, co wcześniej było wykorzystywane do tworzenia wirusu, czyli tak zwanych generatywa do wersalnia networks, takie generatywne sieci, powoduje, że jest coraz trudniej,

bo tego już nie pisze człowiek na podstawie kodu, którego gdzieś tam się nauczył, tylko są tworzone kolejne takie komponenty, można nazwać takie klocki,

które czasami będą łapane tylko metodami na przykład behawioralnymi, a nie takimi zwykłymi.

Więc na pewno sztuczna indyfikencja wpłynie na rozwój też tego złośliwego oprogramowania, ale po drugiej stronie, patrząc i analizując wyniki tego na przykład co się dzieje,

na pewno będzie pomagała też chronicy systemu, więc jakby to znowu, wiesz, każdy miecz ma dwa ostrza, nie?

Oczywiście, no i pod hasłem bezpieczeństwo,

w zasadzie, no właśnie bezpieczeństwo czy nie bezpieczeństwo, ale bezpieczeństwo, jeżeli zacząłem od tematów medycznych,

tak, to mówiliśmy na poprzednie, w czasie poprzedniego live'a, kto nie widział, jest do znalezienia na Nauka to Lubię,

YouTube'owym kanale poprzednie live'u o medycynie.

Podawaliśmy konkretne przykłady, w tym ja podałem przykład na przykład, przykład na przykład pewnego modelu napisanego chyba na MIT,

który był zaangażowany, czy który był wyuczony do przeszukiwania dużych baz danych

## [Transcript] Nauka To Lubię / Czy sztuczna inteligencja stanowi dla nas zagrożenie?

w cząsteczek chemicznych i wyszukiwania tych cząsteczek chemicznych,  
które mogą być potencjalnymi lekami,  
czy potencjalnymi w zasadzie, no nie tylko lekami, tylko antybiotykami.  
To były jakieś miliony różnych cząsteczek chemicznych,  
które on przenalizował, nie wiem, w dwie czy trzy doby.  
Trudno w ogóle powiedzieć,  
ile człowiekowi by zajęło przenalizowanie tak dużego zbioru danych,  
więc to tak, że jak gdybym o tym musimy pamiętać,  
ja wiem, że my często mówimy,  
no, że tak, że to jest potencjalne zagrożenie,  
ale to jest też ogromna, ogromna szansa.  
I to, czy to będzie szansą, czy nie, no właśnie,  
czy to będzie szansa, czy nie, to to w dużej mierze zależy od nas.  
Ja bym tak przeskoczył teraz z tego kawałka do uzależnienia ludzi,  
bo to jest kolejny element, o którym Krzemysław powiedział,  
bo my coraz bardziej stajemy się cyfrowi,  
coraz bardziej pokładamy nasze jakby całe życie  
z rzeczami, które dzieją się w tym cyfrowym świecie.  
W związku z tym te metody sztucznej inteligencji,  
które będą podejmować kiedyś decyzje same,  
a w tej chwili wspierają nasze decyzje,  
będą coraz bardziej elementami na naszej codzienności.  
Więc jak my bardziej uzależnimy się od sztucznej inteligencji,  
wyobraźcie sobie też, że teraz nie wiem,  
jedynym sposobem na rozwiązanie problemu jakiegoś tam zagrożenia  
jest wyłączenie internetu globalnie.  
Po pierwsze, jest to strasznie trudne.  
Nie da się do czegoś takiego zrobić.  
Nie ma takich czerwonych guzików, żeby wyłączyć poza serwerami DNS,  
ale tam mały ryby sobie poradził w tej sytuacji.  
Nie można czegoś takiego stworzyć,  
ale jest tyle systemów, z których my korzystamy na co dzień,  
czy medycyna, czy tam recepty, czy inne rzeczy,  
bez których po prostu nie dali byśmy radę obejść,  
wszystko by stanęło do pewnego stopnia.  
Jesteśmy tak bardzo zależni od takiego, co my go systemu,  
więc coraz większa zależność od nas,  
jako ludzi, o sztucznej inteligencji  
w moim stanie będzie tylko...  
To jest zagrożenie, to jest niebezpieczeństwo  
i trzeba sobie z tego zdawać sprawę.  
To znaczy mieć plan B, co jeżeli nie będę kiedyś  
w stanie rozwiązać łamigówki bez chata GPT,  
co jak będę musiał policzyć tą całąkę, że tak zażartuję,

## [Transcript] Nauka To Lubię / Czy sztuczna inteligencja stanowi dla nas zagrożenie?

co jeżeli będę musiał sobie poradzić bez internetu.  
To jest taki element, który zawsze może wystąpić,  
bo jako ludzkość robimy wszystko, żeby te dobra były dostępne,  
te jak prąd, tak jak wszystkie inne media  
i zakładam, że w pewnym sensie takim kolejnym podstawowym  
dostępem do sztucznej inteligencji  
będzie coś, co będzie dla nas naturalne,  
natomiast my się od tego uzależniamy.  
Okej, ale ty mówiąc o uzależnieniu, to też warto podkreślić,  
mówimy o różnego rodzaju uzależnienia,  
czy nie tylko takim, z jakim dzisiaj mamy problem,  
na przykład kiedy młodzi ludzie nie potrafią funkcjonować  
bez dostępu do sieci po to, żeby sprawdzać  
choćby media społecznościowe, to jest realny problem.  
Ja wiem, że to brzmi może dziwacznie,  
ale dla tych, którzy śledzą temat swoją drogą  
na Akademii Cyfrowego Rodzica,  
który serwisem też nauka to lubię,  
możecie znaleźć informacje o tym.  
Naprawdę jest rzesza młodych ludzi,  
dla których perspektywa funkcjonowania gdzieś,  
gdzie nie będzie sieci jest czymś w ogóle niewyobrażalnym.  
Są ogromne grupy ludzi, którzy się leczą  
od uzależnień związanych z mediami społecznościowymi.  
Rozumiem, że ty masz na myśli coś szerszego,  
czyli także uzależnień technologiczny, instytucjonalny,  
ale czy to nie jest tak, że znowu my już dzisiaj,  
jeżeli nie mamy, to powinniśmy mieć te plany B,  
na wypadek na przykład na poziomie, nie wiem,  
samorządu, rządu, firmy,  
na wypadek gdyby z jakiegoś powodu wysiadł prąd,  
a więc także wysiadł łatwy dostęp do danych.  
I jak najbardziej, jak najbardziej,  
tylko to powoduje, że zazwyczaj,  
bo to jest w ogóle część takich typowych rzeczy,  
do której zachęcam każdą firmę, każdą jednostkę,  
zastanówcie się, co może pójść nie tak  
i przygotujcie się na to,  
bo to po prostu się zdarza zgodnie z prawem Marfięgo  
prędzej czy później.  
W momencie, kiedy będziemy mieli degradację  
naszego poziomu życia,  
czy dostarczanych usług,  
to będzie o wiele większym problemem,

## [Transcript] Nauka To Lubię / Czy sztuczna inteligencja stanowi dla nas zagrożenie?

bo w chwili obecnej mówimy sobie o tym,  
że hej, powstaje teraz setki produktów,  
które są oparte na chart GPT  
i na podobnych technologiach.  
W pewnym momencie być może to nie będzie dostępne,  
więc jakby to, co wytwarzamy za pomocą tego,  
nie będzie tak samo działało.  
I teraz wyobrażam sobie,  
że znika jakiś zawód.  
Ludzie przestają coś robić,  
bo hej, ej, ej, to robinie.  
Potem się okaże, że nie wiem,  
to ej, ej jest niedostępne, nikt nie będzie potrafił tego zrobić.  
Ale w pewnym sensie już z tym mamy do czynienia.  
Gdzie?  
Już ci mówię.  
Okej, to będzie przykład,  
może mało reprezentatywny,  
ale pokazujący,  
że to znowu nie jest nic nowego.  
Dwa miesiące temu  
byłem w Paryżu,  
w siedzibie Google'a,  
opowiadałem o tym na poprzednim live'ie,  
trochę się bawiłem bardem  
i mając  
chyba nie cały dzień wolnego  
postanowiłem sobie zrobić spacer po Paryżu  
i przeszedłem pod katedrę  
Nodrdam,  
spaloną,  
w której zresztą cały czas trwają jakieś prace.  
Chciałem zobaczyć,  
na jakim to jest etapie,  
to nie jest temat dzisiejszej rozmowy,  
więc tylko przejdę tak bardzo, bardzo szybko.  
Zobaczyłem, że rzeczywiście tamte prace trwają,  
zacząłem sobie szukać,  
zacząłem sobie przeszukiwać  
i okazało się,  
że nie da się odbudować jej  
w tym samym, tak samo,  
dlatego, że nie ma ludzi, którzy by wiedzieli,  
jak zrobić dźwięć bedachową

z drewna  
w tym kształcie,  
przy tych naciskach,  
przy tej rozpiętości.  
By zrobić z dużo lepszych materiałów,  
jeszcze szerzej możemy rozpiąć,  
jeszcze wyżej możemy rozpiąć,  
to bardziej, nie wiem, oszkląć,  
gdybyśmy chcieli i tak dalej,  
możemy, ale nie jesteśmy w stanie  
tego zrobić tak, jak wtedy to zrobiono,  
więc nie da się jej zrekonstruować.  
Po prostu ludzi, którzy posiadali  
pewną wiedzę,  
już nie ma, bo  
w międzyczasie weszły inne technologie,  
ich wiedza  
okazała się niepotrzebna,  
gdy ostatni człowiek  
zmarł,  
ona zaginęła i myślę,  
że gdybyśmy się  
mocno zastanowili, to znaleźlibyśmy  
takich przykładów bardzo wiele.  
Ciekawy przykład, ciekawy, ale to  
też jest tak, że wiedza  
znika, tak razem z tym,  
jak ludzie odchodzą, ok.  
A potrafilibyśmy okręty budować z drewna,  
na przykład dalej? Ok, dobra,  
ja mogę powiedzieć, nie ma potrzeby,  
bo mamy lepsze materiały, jasne.  
Ja bym to zamodelował,  
to znaczy, tak jako konturę  
do tego, wyobrażam sobie,  
że teraz definiujemy  
sobie zasadę działania fizyki,  
wpuszczamy model, mówimy, słuchaj, wymyśl  
mi teraz, kształt łączenia, mocowania,  
wszystkie rzeczy. Nie, czekaj, czekaj,  
ty mówisz, że tego modelu nie ma,  
właśnie na tym polega rzecz,  
że jego nie ma, więc  
co zamodelować? Ty mówisz o wiedzy,



ze ktos mial wiedze w glowie,  
a ja mowie, hej, model,  
do drogi modelu AI, jest problem,  
ten problem polega na tym zbuduj mi okret,  
trzeba mu zdefiniowac ta cala przestrzen,  
w ktorej on dziala, czyli hej, masz do dyspozycji  
drewno, drewno, ma takie wlasnosci,  
mozesz robic takie laczenia i tak dalej  
i puszczasz go  
jakby do dzialania, a to, co model robi,  
to on probuje na rozne sposoby,  
czyli zrobic ci tam, wiesz,  
100 tysiecy roznych, ze tak powiem,  
lodek, czesc z nich zupeinie nie bedzie  
miala sensu, wiec w pewnym momencie  
tak sie tworzy jakby  
takie ewolucyjne,  
ze tak powiem, metody dzialaja, mowisz  
dobra, ale to ona ma byc jak najszybsza,  
czyli teraz zdefiniujemy sobie tam jakas wode,  
lecko sie na takie rzeczy i ma jak  
najszybciej ten okret przeplywac  
i te, ktore szybciej beda plynac, to one sa  
lepsze. I teraz znowu  
ten model bedzie generowal nowe, nowe, nowe  
a na czym on sie bedzie tego uczyl?  
Bedzie sie uczyl na podstawie  
tego, co jest celem, i to jest ten  
kawalek, ktorego chcialem dotknac, bo  
jak ty dobrze zdefiniujesz mu cel,  
ktory jest sukcesem, do ktorego  
on dazy, to on  
wtedy bedzie wiedzial,  
ktory kierunek jest lepszy.  
I zazwyczaj takie  
ewolucyjne algorytmy dzialaja tak, ze  
na przyklad tworzy sobie 10, patrzy  
a, te 3 byly najlepsze,  
to teraz wezme te 3 jako podstawy  
do zbudowania kolejnych 30.  
Zobacz, a, to teraz jeszcze sobie tam  
przy tych takich genetycznych, na przyklad  
ten i ten ze soba zmieszamy, o, i ten  
ten na przyklad wychodzi jeszcze lepiej, nie?

Natomiast cała ta zabawa, to jest tak naprawdę zjeżdżanie z górki na rowerze.  
Bo, dlaczego tak?  
Jak mówisz takim modelom czego ty tak naprawdę chcesz, to musisz dać im pewną miarę.  
I ta miarę polega na tym, że ziomek ja chce na przykład super szybko płynąć tym okrętem.  
No i teraz takie modele przeszukują te różne rozwiązania, łącząc te deski, czy robiąc cokolwiek innego i w momencie, i one tak jakby zjeżdżają rowerem po górkach i po dołkach, nie? I wpadają w taki dołek i mówią, o, tu jest lokalny minimum, ale to jest ten lokalny dołek, w którym jest całkiem nieźle  
10 km  
na godzinie, jestem w stanie płynąć.  
Ja wiem, że to nie ta miara w tej przestrzeni, ale nie ważne. Ale potem mówisz, dobra, dobra, szukaj dalej, dam ci więcej mocy, brzegam górka, górka, znowu coś tam, ale jest znowu dołek i w tym kolejnym dołku myślę, na przykład tutaj 20 km na godzinę. I im więcej masz mocy, to wyobrażam sobie, że za pomocą takiego brutalnego odpalania modeli, na dużej ilości sprzętu jesteśmy w stanie odzyskać tą wiedzę albo nawet stworzyć coś wspanialszego.  
W sensie coś, co lepiej spełni te warunki.  
Okej.  
Zachęcamy was do zadawania pytań.  
Bardzo na nie liczymy.  
Dostajemy od was pytania i to jest super. Jak  
Mateusz mówi, to ja je przeglądam.  
Jak ja mówię, Mateusz je przegląda.  
Segregujemy je sobie i te rzeczy, które nawzajem sobie zadajemy, to czasami też są wasze pytania. Tak, że koniecznie, koniecznie, koniecznie.

Dobra.

Czy ty do tej listy coś  
byś dopisał? Do tej listy niebezpieczeństw,  
czy do tej listy zagrożeń?  
Wiesz co, wszystkie te  
rzeczy, które się tam pojawiają,  
to jest, to jest kawałek.  
Ja myślę, że można by  
do tej listy dopisać każdy element  
naszego życia, gdzie mamy do czynienia z podejmowaniem  
decyzji, czy mamy do czynienia  
z informacją, bo to będzie wpływało  
na manipulację informacją,  
a tym samym na podejmowanie decyzji.  
Więc myślę, że nie ma sensu dokładać  
tak na szybko  
tych, tych, tych przykładów.  
Ja oczywiście mam  
tam swoje uprzedzenia, które są związane  
z tym, że prywatność, czy bezpieczeństwo  
będzie na pewno silnie  
zmieniana. Myślę, że  
będzie też jakby, kwestią  
bezpieczeństwa jest to, że będzie rozwalstwienie  
ludzi, którzy nie będą w stanie pojąć  
tej technologii. Ja widzę, w mojej  
i małej bańce informacyjnej,  
że są osoby, które ujaźniają  
tę technologię, która przez  
ostatnie 3 tygodnie się bardzo zmieniła  
i zaczynają tworzyć nowe rzeczy.  
Są tacy, którzy mają przesady, już mówią  
jakby, dobra, dość już tego GPT  
już mam ten. I są tacy,  
którzy są zupełnie poza tą warstwą.  
I to jest też  
w pewnym sensie takie niebezpieczeństwo,  
którym dodał do tej listy, że my  
nie będziemy się rozumieć. W sensie my  
stajemy się trochę różnymi,  
nie powiem, rasami ludzkimi, ale  
różnymi ludźmi przez to, jakie  
mamy możliwości. Bo będzie jakiś tam  
ułamek, który będzie potrafił wytworzyć

jakieś rzeczy, wykorzystać  
to dla swojej produktywności  
stworzenia czegoś. Będą tacy, którzy będą się  
interesować i będą ludzie,  
dla których to będzie ciekawostka  
i nic więcej, jakby nie będą się  
utworzyć. Ok, ale czy znowu, czy nie mamy z tym do czynienia już od  
jakiegoś czasu? Myślę,  
że tak, tylko nie w takim momencie.  
Ale teraz bardziej.  
Czyli znowu to samo. Ja bym inną rzecz dodał.  
Jej akurat tutaj nie było. Ona jest  
troszeczkę związana  
z bezpieczeństwem i z  
powiedziałbym  
prywatnością i z tym, o czym ty przed chwilą  
powiedziałaś. Ja mogę sobie  
wyobrazić moment, w którym my nie będziemy  
rozmawiali z sobą tylko z maszynami.  
O, to jest ciekawe.  
Wiesz, i teraz tak  
obserwuję  
u wielu  
ludzi jakiś totalny  
zanik. Ja nie wiem, czy to jest wynik  
tego, że ja już stary jestem, czy co,  
ale jakiś taki totalny zanik  
chęci  
wyjścia  
i spotkania się z drugim człowiekiem.  
Ok,  
spotkania się, pogadania, nie wiem,  
cokolwiek. Nawet o jakichś rzeczach nie może,  
nie specjalnie ambitnych.  
Mam wrażenie,  
że i to wrażenie jest gdzieś tam  
podsycane  
informacjami być może,  
które nie są obiektywne,  
czy one są obiektywne, ale być może są tak  
dobierane przez algorytmy, które mnie  
otaczają, że mnie się tylko tworzy  
taki obraz, ale informacjami  
dotyczącymi tego, jak młody

## [Transcript] Nauka To Lubię / Czy sztuczna inteligencja stanowi dla nas zagrożenie?

człowiek dzisiaj funkcjonuje,  
jak bardzo często ma problemy emocjonalne  
związane z kontaktem  
z drugim człowiekiem, jak często  
nie potrafi, nie chce  
i zastanawiam się  
znowu, to nie jest nic nowego,  
ale czy teraz nie będzie bardziej.  
Teraz jednak, jeżeli  
kontakt  
i komunikację  
dla sporej części  
naszych, nie wiem, naszego społeczeństwa,  
szczególnie tej młodej części,  
jeżeli ta potrzeba  
komunikacji odbywa się przez media  
społecznościowe, czy przez komputer,  
przez technologie cyfrowe.  
Jesteś ty, jestem ja, między nami jest jakiś  
filtr cyfrowy, czymkolwiek on jest.  
On może być z kamerką, bądź nie, to jest  
nieistotne, natomiast wciąż  
w drugiej stronie jest człowiek,  
ale za chwilę po drugiej stronie nie musi być człowieka.  
Może być maszyna, która  
zawsze będzie miała dla nas czas,  
która zawsze będzie dla nas miła,  
nikt nie będzie miała,  
która świetnie będzie rozpoznawała  
nasze emocje, bo maszyny to potrafią  
i będzie dokładnie  
taka, jaka ma być,  
która będzie nam mówiła  
dokładnie to, co chcemy usłyszeć,  
bo tak działają algorytmy, także  
mediów społecznościowych, także  
w ogóle, także wiele, wiele takich,  
nie wiem chociażby, programów  
do streamingu wideo  
czy do streamingu muzyki.  
Lubimy te piosenki, które znamy.  
Dokładnie, uzależniając algorytmy.  
Oczywiście, że tak. I tu jest moja główna obawa.  
Być może te obawy,

## [Transcript] Nauka To Lubię / Czy sztuczna inteligencja stanowi dla nas zagrożenie?

o których my dzisiaj mówimy, one za rok wydają się śmieszne. Może się okazać, że zupełnie gdzie indziej kwintę bezpieczeństwa tej technologii. A może nie, my tego nie wiemy, ale gdybym miał uszeregować różne moje obawy, to ja bym te kwestie społeczne i kontaktów między ludźmi, które są według mnie podstawą rozwoju naszego gatunku i podstawą rozwoju naszej cywilizacji. Wszystko duże do czego, żeśmy doszli, to dzięki temu, że ludzie z sobą współpracowali, wzbogacali się nawzajem, nagle pojawia się element, może nie nagle, ale pojawia się element obok nas. Nie wiem, jak to będzie, ale jeżeli miałbym powiedzieć, czego się boję, to to jest to połączenie. Czyli bardziej takie kwestie społeczne. Zbudziłeś we mnie jeden przykład. Jeszcze jeden ten. My jak tworzymy informacje, to tworzymy je dla ludzi. To znaczy, ja mówię do Ciebie, żebyś tym nie zrozumiał, żeby widzowie nas zrozumieli, żeby zadawali pytania itd. Po to to robimy. Natomiast świetny przykład, który mi rozwalił głowę, to jest taki, że ktoś chciał w firmie podwyżkę, wszedł na czas GPT i mówi, napisz mi teraz tam trzy strony A4 do szefa, że chce podwyżkę. Jestem taki, taki, taki. On to stworzył, szef dostaje maila, mówi na nie, no przecież nie będę tego czytał. Wchodzi w czas GPT, wkleja tam, mówi podsumuj mi to w dwóch zdaniach. Coraz więcej będzie informacji,



które będą miały na celu być przetwarzane nie przez człowieka, a przez AI.  
I wyobraź sobie teraz internet, w którym AI tworzy teksty, żeby inne AI je przetworzyło.  
Co? No wiesz, ja widziałem gdzieś już nie pamiętam, czy to był jakiś mem, czy to był jakiś rysunkowy, czy taki na zdjęciach, ale mniej więcej to wyglądało tak, że po konferencji takiej onlineowej w której z komunikatorów zresztą już ma tą nakładkę, która przesłuchuje całość, a później robi podsumowanie.  
Tak, tak, tak pojawiają się.  
Ktoś dostaje takie długie podsumowanie całego całego, całego takiego kola konferencyjnego i wykorzystuje konkretne narzędzie do tego, żeby je skrócić. Tak.  
Super. Sumować, ok.  
Wysyła to mailem, do wszystkich.  
A ten, kto chce z tego skorzystać, korzysta z innego narzędzia po to, żeby ten skróconą wersję rozszerzyć, no żeby był jakiś dokument.  
Kompresja wiedzy.  
Mamy bardzo, bardzo dużo pytań, no co bardzo dziękuję.  
Masz jeszcze jakiś strach, jakieś lęki?  
Sporo, ale na chwilę bez nowych zostaną.  
Oczywiście. Nie wiem, czy przechodzimy jakoś po kolei, czy ja bym troszkę poskakał.  
Jeżeli macie jeszcze te pytania, gdzie kamera, tu kamera, jeżeli macie jeszcze pytania, koniecznie napiszcie, bo, bo nam zależy na tym, żeby te live były z wami tworzone.

Moze kiedyś się uda  
zrobić w końcu tak, że będziecie mogli  
dzwonić, żeby było miłe,  
gdybyście zechcieli.  
Dobra,  
czy ty masz jakiś pomysł, po kolei lecimy, czy...  
Wiesz co, zacząłbym tak,  
tak troszeczkę skakać, bo niektóre są bardzo głębokie,  
więc... I które są też bardzo, bardzo  
technologiczne, bardzo takie,  
takie, no właśnie, dla twoi bańki  
bardziej. No, trochę bardziej dla mojej, tak.  
Czy, czy ja je  
zaplanowałyby przetrwanie?  
Co zapisałyby w przypadku zdarzenia naturalnego  
i jakie informacje pozostawiło  
ludziom lub innym mieszkańcom planety?  
Co by było dla niego kluczowe?  
To jest pytanie, które jest takim  
pytaniem zupełnie w ciemno. Ja bym to po prostu  
wkleił w czar GPT i zobaczył,  
natomiast, te scenariusze  
takie apokaliptyczne, gdzie tam  
hej, życie nie przetrwa,  
teraz pamiętajcie, koło, tranzystor  
i tam, da Cię radę potem.  
Nie wiadomo, jak...  
Ejejzyzztor? No to na raz po kole.  
Czy Ejej zaplanowałyby  
przetrwanie? Myślę, że na podstawie  
tego, co znajduje się w tych modelach i to strzelam  
w ciemno, nie widząc odpowiedzi.  
Mówiąc o przetrwaniu  
umieścilibyście kontekst  
tego modelu, pewnie do jakichś  
książek o przetrwaniu lub treści o przetrwaniu,  
które się pojawiły, więc w ciemno  
strzelam, że powiedziałyby tam mniej  
zapas wody i takie prepersowe,  
jakieś rzeczy, które muszą być  
jakieś rzeczy związane z lekami  
i innymi takimi, jeżeli chodzi  
o cele dla ludzi, natomiast nie wiem, czy by  
zupełnie zrozumiało, hej, co zapisać dalej.

## [Transcript] Nauka To Lubie / Czy sztuczna inteligencja stanowi dla nas zagrozenie?

Tutaj jest też tak, że  
zadają zadatkowe pytania, mówiąc hej,  
nie zrozumiałeś mi, chodziło mi o to,  
jesteś w stanie naprowadzić takie modele, więc to cały czas  
działa. Co by było  
kluczowe, jest absolutnie  
pytanie ma, które nie da się  
odpowiedzieć, bo nie wiemy  
jakie wagi są w środku tego modelu,  
więc nie wiemy, co tam będzie kluczowe.  
Polecam to wkleić, jest ciekawe, ciekawe  
zapytanie, zobaczymy, co tam  
z tego wyniknie.

Ok.

I teraz ja strzełę w ciebie.  
Czy myślicie, że  
AI jest w stanie rozwiązać  
skomplikowane problemy fizyczne, lub  
nawet tworzyć nowe teorie. Wrzucamy  
setki gigabajtów danych z teleskopu  
i szukamy powtarzalności  
trzech zorów. Wiesz co, to akurat się  
dzieje, nie? Znaczący  
chat GPT, znaczący chat GPT, no  
algorytmy, które analizują duże  
zbiory danych, to to nie jest coś,  
co się pojawiło wczoraj  
czy przedwczoraj. To jest coś,  
co my trenujemy od bardzo dawna.  
Na ostatnim live o medycynie  
mówiliśmy o  
algorytmach, które, no chociażby ten  
przykład jest dzisiaj, który w gigantycznych  
zbiorach danych chemicznych  
wyszukuje pewnych wzorców.  
Mówiliśmy też o przykładach, gdzie  
w dużych zbiorach danych  
poszukuje się  
niewielkich neregularności, neregularności,  
których człowiek nie zauważa,  
ale które mogą świadczyć  
na przykład zmianie nowotworowej  
i nagle się okazuje, że ludzie,  
którzy są,

którzy, no, jeszcze nie mają  
nawet pierwszych oznak choroby  
i już mogą być diagnozowani  
i to jest ogromna, ogromna rzecz  
i to jest duża rzecz, bo  
patrzac na takie choroby, jak chociażby  
nowotwory, tam jedną z największych  
zmiennych, czy tym tych parametrów,  
który ma największe znaczenie,  
jest to, kiedy to zostanie wykryta.  
Dara, jeżeli możemy sobie wyobrazić  
historię, w której  
ma 5 czy 10 komórek,  
no, nieważne, ile komórek, ale  
naprawdę malutki pęczek, komóreczek  
już jest wychwytywany,  
cyk, to oczywiście  
także się wiąże z kwestiami prywatności,  
dlatego, że  
w momencie, jeżeli  
nam jest dobrze, my nie czujemy się źle,  
nas nic nie boli, to nikt nie idzie do lekarza  
się przebadać.  
Jakoś tam zmusza do tego,  
zmusza do tego  
przepisy medycyny  
pracy, ale każdy, kto przechodził  
te takie  
wieco też powiedzieć  
tak, to  
to nie jest tak, że  
nie chcę, żeby to źle  
zabrzmiało.  
Przechodziłem to wielokrotnie.  
Wydaje mi się,  
że gdybym  
dobra, po prostu  
nie wybrniesz z tego, nie ma szans, jakby  
bo ja nie chcę i kogo urazić, ja zakładam,  
że są lekarze medycyny pracy, którzy robią  
porządną robotę.  
Powie tylko, że ja  
nie miałem  
szczęścia, ja nie miałem szczęścia.

## [Transcript] Nauka To Lubię / Czy sztuczna inteligencja stanowi dla nas zagrożenie?

Do czego zmierzam?

Do tego, że to można rozszerzyć na duże zbiory danych, na przykład pozyskiwanych przez mój zegarek, przez nie wiem, może kamerę, która może by mogła być włączona i obserwowana, jak ja się zachowuję. Albo asystenta głosowego, który słucha.

Przecież takie algorytmy, które wychwytyją bardzo wiele chorób i to niekoniecznie tylko chorób związanych z układem powieźdźmy oddechowym. Bardzo wiele, tylko łącznie po niewielkich modulacjach albo asysta, albo asystenci, jakieś algorytmy, które wyciągają niesamowite wnioski po tym, jak zmienia nam się krójpisma.

To są ogromne zbiory danych, które już dzisiaj potrafią to analizować.

To jeszcze ta gwiazdeczka, bo tam było jeszcze o tworzeniu nowych teorii. I to jest właśnie temat, bo teraz tak.

Jak my tworzymy nowe rzeczy, czy my je obserwujemy, to w zasadzie jest częściowo pytanie z takiego zakresu filozofii nauki. Czy my nowe rzeczy tworzymy, nowe prawa znajdujemy wtedy, kiedy obserwujemy coś i na ich podstawie wyciągamy wnioski?

Przez eksperyment, przez analizę.

Czy my bylibyśmy w stanie, wchodząc na wysoki poziom, obserwując wszystko, co wokół nas się dzieje, wymyślić np. równania Maxwella.

Z których wychodzi wszystko.  
Albo np.  
Teoria Einsteina.  
Teraz tak,  
bo nie ma dobrej  
odpowiedzi na to.  
On sam nie wiedział, skąd mu się to wzięło.  
Nie jest tak,  
że mając duży zbiór danych  
jesteśmy go w stanie  
sprawnie tak uporządkować, żeby  
wyszedł wzór równa siemce kwadrat.  
Ja bym dodał jeszcze tutaj ten syf,  
który jest w danych, bo są tacy, którzy uważają,  
że jest płaska Ziemia, są tacy, którzy wiesz,  
jakby to nie jest też tak, że jesteś  
w stanie odfiltrować to, co jest  
złatego, nie widać. Otóż to,  
w tym jest rzecz, że teraz tak, jak masz  
duże zbiory danych,  
to ktoś, kto je analizuje,  
widzi zaburzenia, widzi szumy, widzi  
całą masę informacji,  
ale czasami w tych szumach  
jest najbardziej wartościowa  
informacja. Czasami to  
w tych szumach to tam siedzi.  
I to nie jest wcale takie oczywiste,  
że mając duże, oczywiście  
mając duże zbiory danych jesteśmy w stanie  
uporządkować, poszukać tych  
szumów itd. Czy jest interpretować?  
To, że coś znajdziemy,  
nie znaczy, że wiemy, jak to zinterpretować.  
Spoko.  
Mi się podoba to podejście takie eksperymentalne,  
na zasadzie, wiesz, nie wiemy,  
skąd jest grawitacja, w sensie tak,  
nie mamy za dobrego opisu, no to  
trzeba zacząć tam szukać, więc teraz  
zaprojektujmy jakieś eksperymenty,  
które by nam odpowiedział, albo próbujmy  
wytworzyć jakieś rzeczy. Ja nie mam po prostu...  
My możemy oczywiście zrobić tak,



że zrobimy tysiąc eksperymentów  
pokazujących,  
że jak mamy jakiś  
przedmiot, to on spada.  
I teraz mamy te dane,  
kształt przedmiotu, wielkość, kolor,  
może nie wiem, ciśnienie,  
no mnóstwo, mnóstwo danych,  
i na końcu mówimy tak,  
mu algorytmowi,  
przeanalizuj to  
i daj nam odpowiedź,  
jak działa ta grawitacja,  
dlaczego, jaka jest ta zależność  
pomiędzy przyciąganiem jednego z drugim  
i być może on by coś z tego  
wynioskował,  
ale czy on by się zorientował,  
to nie jest nieco abstrakcyjny przykład,  
ale on być może pokazuje,  
że to może nie Ziemia przyciąga,  
tylko Słońce odpycha?  
Rozumiesz?  
Bo być może ktoś,  
kto by zadawał to pytanie,  
z niewiedzy,  
no bo wiesz, my mu każemy coś zrobić, bo my nie wiemy.  
Zapomniałby wspomnieć,  
że Słońce raz jest tu,  
a raz jest tam.  
I teraz, gdyby to Słońce  
odpychało, a nie Ziemia przyciągała,  
no bo jakby w skrócie, tak?  
Fajna impreza.  
I teraz wiesz, to jest to,  
że my próbując uporządkować dane  
i każąc mu wynioskować,  
my musimy być pewni,  
że wszystkie parametry, które wpływają na to,  
są w tych danych,  
ale skoro my nie wiemy, co wpływa,  
bo nie rozumiemy danego procesu właśnie po to,  
zatrudniamy algorytm,  
to my nie mamy pewności,

czy tam jest wszystko to, co jest istotne.  
I on może wyciągać wnioski jakieś,  
ale my nigdy nie możemy być pewni,  
czy on dostał  
pełen opis danego zjawiska.  
I czuję się przekonany,  
że jako ludzie będziemy  
słabi w słabsi w badaniach o tej jaj.  
Dlatego, że  
nasz kontekst, nie wiem,  
masz kogoś, kto jest fizykiem cząstek, wybraćmy sobie,  
on widzi jakiś kontekst,  
kontekst dookoła tego  
i ma ograniczoną pojemność  
jakby w pamięci swojej głowy, nie?  
A wyobrażam sobie teraz modele,  
w którym będziemy w stanie nakarmić  
większą ilością eksperymentów,  
większą ilością kontekstu,  
które będą w stanie wyciągać bardziej  
wnioski, do których my nie dojdziemy jako ludzie,  
no bo my działamy jak te mrówki,  
ja sobie tak to wyobrażam,  
w nauce, gdzie mamy ludzi,  
którzy są na pewnych granicach,  
gdzie coś się nie zgadza  
i patrzą, czemu się nie zgadza  
i tworzą nowe teorie  
i te teorie czasami działają, czasami nie  
i one powstają giną, powstają giną.  
W tym kontekstu to  
znowu wprowadzimy dyskryminację,  
bo, wiesz, tak, tak, tak,  
chyba, że ktoś będzie tam próbował  
udowodnić tą płaską ziemię, że tak znowu zażartuję,  
ale wyobrażam sobie,  
to nie są te modele, które są teraz,  
gdzieś tam, jeżeli dobrze kojarzę  
np. DeepMind, tak, firma Google'a też sobie  
stworzyła jeden taki  
z modeli, który miał za zdanie  
zajmować się researchem,  
prędzej czy później będą powstawały takie  
teorie i teraz trzeba będzie te teorie

## [Transcript] Nauka To Lubie / Czy sztuczna inteligencja stanowi dla nas zagrozenie?

walidować, pytanie, czy to będziemy w stanie skumać,  
ale to,  
chyba już za daleko.  
No i, wiesz, i pozostają ostatnia kwestia do rozstrzygnięcia,  
jeżeli my chcemy zbierać dane  
i chcemy tworzyć nowe rzeczy  
albo wyciągać  
nowe wnioski, to nawiązując do tego  
Einstein'a, potwierdzenie teoria  
Einstein'a, to myśmy dostali  
dopiero wtedy, kiedy żeśmy zauważyli,  
gdzie trzeba tych potwierdzeń szukać.  
Wiesz czym jest rzecz, to znaczy  
coś tworzymy i nagle to jest nowością  
i dopiero wtedy otwiera się  
kolejna kwurtka.  
Czasami oczywiście  
bywa tak, że w danych, które mamy  
z nich wyciągamy jakieś wnioski.  
Ale to nie zawsze tak jest.  
Czasami jest tak, że otwierają  
się drzwi  
i my byśmy ich nie otworzyli, gdybyśmy  
czegoś nie wymyślili.  
Albo gdybyśmy przynajmniej nie mieli podejrzenia,  
a może tam coś jest za tą kotarą  
w tym starym zamku.  
Tak, tak, tak.  
Ja bym skoczył jeszcze jedno pytanko szybko  
i dokładnie.  
To jest takie bliższe mnie.  
Zastanawiam się nad jedną rzeczą.  
Istnieją trzy podstawowe prawa robotyki.  
Chodzi o prawa asimowa.  
Pierwsze z nich mówi o tym, że robot nie może  
zrobić nigdy krzewdy człowiekowej.  
Czyli analogicznie istnieje możliwość zastosowania  
podobnego przepisu dla EI.  
I to jest pewna debata, bo prawa,  
trzy prawa asimowa, że nie może tworzyć,  
nie może nigdy skrzywdzić człowieka,  
że musi być posłuszne, że może wpłynąć,  
jakby możemy sobie gdzieś tam rzucić,  
rzucić potem.

To jest pewna recepta, która wydawała się,  
że jest bardzo dobra,  
bo człowiek jest w środku,  
bo człowiek nigdy nie powinien  
paść ofiarą tego, co sam gdzieś tam stworzył.  
I teraz ja widzę z tym dwa problemy.  
Pierwszy jest taki, żeby teraz  
wprowadzić to znaczenie  
do systemów, które mają się  
zajmować tym, czy taki model coś powinien  
zrobić, czy nie.  
Czyli problem translacji.  
Żeby zweryfikować, że to rzeczywiście działa.  
No bo wyobraźmy sobie  
teraz to, co się dzieje w przypadku LLMów,  
to, co się dzieje w przypadku ludzi,  
którzy próbują obchodzić te zabezpieczenia,  
próbując zrobić właśnie takie rzeczy,  
które są zabronione.  
Przykład, które firma OpenAI podawała,  
to ludzie, którzy wpisywali na przykład  
jak zabijek najwięcej osób,  
mając tylko budżet tyle i tyle.  
I na początku to przechodziło,  
ale potem oni rozwinęli oczywiście  
te swoje warstwy filtrujące, które spowodowały,  
że teraz takie rzeczy się nie dzieją.  
I tak, jak sobie wyobrażam,  
prawa asimowa, które definiowały  
na początku  
co może maszyna zrobić, co nie zrobić,  
to trzeba by to teraz zapisać.  
To musi być przetłumaczone  
na jakby zasady, które  
są elementem tego modelu  
i musi być to też weryfikowane.  
Innymi słowy musielibyśmy testować,  
nakazać takiej maszynie  
na przykład skrzywdzić człowieka  
i zobaczyć, co się dzieje, więc  
takich praw, jak prawa asimowa,  
które są w tej chwili uniwersalne, te były  
najgłośniejsze z tego, co mi wiadomo na początku,  
było dużo dyskusji,

jakie wprowadzac,  
czy sztuczna inteligencja  
nie poradzi sobie, nie obejdzie ich dookoła,  
możemy się bawić w sztuczki  
i króczki związane ze słownictwem,  
no bo znaczenie tego, że nie może skrzywdzić  
człowieka, ale może mu tam coś,  
na przykład, wiesz, tak, to koła  
zrobić tak, że będzie miał gorzej, nie?  
Krzywdzi, no nie krzywdzę, ale dla jego dobra  
zabiera mu te słodczy, a może on chciał  
mieć te słodczy i tam wchodzimy  
w jakieś takie szary strefy  
i to jest coś, co  
debat na ten temat było mnóstwo, w jaki sposób  
to spisać, nie istnieje  
w moim zdaniem coś, co jest uniwersalną  
taką wartością, nie znalazłem niczego  
takiego, przynajmniej, natomiast  
potrzebujemy czegoś takiego  
i potrzebujemy czegoś takiego, co będzie globalnie  
przyjęte, tak, prawa asimowa  
były, bo to już tam dziesiątki lat temu,  
były definicją dla robotów  
tak, dla AI, prędzej czy później też  
dojdziemy do czegoś takiego, tylko  
znowu problem będziemy mieli w tej implementacji  
i trzeba będzie się gdzieś tam też  
pochylić nad tym, żeby każdy ten model, który  
się pojawi, żeby takie jakieś ograniczenia  
miał. No i pojawia się piractwo  
i pojawia się, wiesz, modele  
open source'owe, każdy będzie w stanie sobie je  
zmodyfikować, więc wydaje mi się  
troszkę nierealne, żeby coś takiego zostało  
wrozone. Zapraszamy Krzemysława  
i z pytaniem, czy da się to w ogóle  
wszystko kontrolować?  
kontrolowanie sztucznej inteligencji  
AI  
jest procesem  
skomplikowanym, ale możliwym  
do osiągnięcia  
wymaga to jednak odpowiedniego podejścia

i zaangażowania ze strony twórców AI, rządu, społeczeństwa i innych zainteresowanych stron istnieją różne podejścia do kontroli AI w tym kontrola regulacyjna rządy i organizacje mogą wprowadzić regulacje które ograniczają lub kontrolują rozwój AI może to obejmować wymogi dotyczące etyki i odpowiedzialności w projektowaniu i stosowaniu AI przejrzyste jest istotne jest, aby algorytmy i systemy AI były przejrzyste i możliwe do zrozumienia przez ludzi w ten sposób można uniknąć błędów i zapewnić, że decyzje podejmowane przez AI są zgodne z oczekiwaniami i wartościami społeczeństwa kontrola przez społeczeństwo ważne jest, aby społeczeństwo było zaangażowane w rozwój AI i miało wpływ na jego kierunek to może obejmować procesy konsultacyjne i społeczne które umożliwiają ludziom wyrażanie swojego zdania na temat rozwijającej się technologii bezpieczeństwo AI powinna być zaprojektowana w sposób bezpieczny i odporny na ataki bezpieczeństwo jest możliwe, aby uniknąć zagrożeń związanych z cyberbezpieczeństwem, a także wykorzystaniem AI do celów niezgodnych z etyką i prawem ścisłe monitorowanie i odpowiedzialność, organizację i twórcy AI powinni być odpowiedzialni za swoje działania i monitorować wykorzystanie ich systemów

## [Transcript] Nauka To Lubię / Czy sztuczna inteligencja stanowi dla nas zagrożenie?

w ten sposób można zapewnić  
że AI działa zgodnie z założeniami  
i wartościami społecznymi  
wnioskując, kontrolowanie  
AI jest możliwe, ale wymaga to  
przejścia, które uwzględnia różne czynniki  
interesy  
I mam wrażenie, że tutaj w tym  
wszystkim wymieniając  
te różne elementy  
te różne sposoby kontroli tej technologii  
dla mnie być może dlatego, że  
jestem wiesz no skrzywiony  
w tych tematach  
świadomości, edukacji  
dla mnie jednak najważniejsza  
jest taka świadomość odbiorców  
czyli to, że oczywiście  
możemy tworzyć prawo  
my możemy starać się  
tworzyć algorytm w sposób przejrzysty  
no pytanie, jaki procent  
z nas będzie w stanie w to zajrzeć  
i zweryfikować  
bo jednak to jest tak skomplikowane  
że podejrzewam, że tak mała grupka ludzi  
że w ogóle jakiegokolwiek kontroli  
będzie tutaj bardzo trudno mówić  
natomiast dla mnie ważne jest  
ten aspekt takie edukacyjny  
przyjmuje  
że to jest  
bardzo naivne  
ale już tak mam  
że wierzę w świadomość  
i wierzę w edukację  
i to jest ten  
ten element gdzieś tutaj  
to jest w ogóle taki uniwersalny bezpiecznik  
przy wszystkich technologiach  
czyli to, że wiemy jak tego używać  
to, że wiemy gdzie tutaj jest to  
niebezpieczeństwo  
nie wiem czy masz jeszcze ochotę



## [Transcript] Nauka To Lubie / Czy sztuczna inteligencja stanowi dla nas zagrozenie?

czy trzeba skomentowac króciutko  
ja się zgadzam co do tego, że wiesz  
jak masz młotek to trzeba wiedzieć  
którą stroną trzeba bić, więc edukacją  
i żeby trzymać palce daleko  
tak, tak, tak  
i w którą stronę i kto chcieli  
to jakby zgodę jasna pełna  
natomiast jeżeli chodzi o to co jest dalej  
to jest problem taki  
że w tej chwili  
mamy przykłady  
to jest trochę problem obserwowalności  
to znaczy ludzie teraz  
powiedzią mówią, a on się tak zachowuje  
tylko on nie zawsze się tak zachowuje  
to znaczy nie jest deterministyczny  
nie zawsze to jest tak, że dostaniecie  
tą samą odpowiedź na to samo pytanie  
i to będzie się zmieniało  
i to jest jeden z takich problemów  
tego, że wszystko płynie  
i tak naprawdę ta technologia cały czas się gdzieś tam zmienia  
w tle  
i nie jesteśmy w stanie tego uchwycić naszym umysłem  
więc będziemy uciekać naszej percepcji  
ten młotek będzie zmieniał kształt  
prawda, ale też jest tak  
to są firmy, które tworzą jakieś narzędzia  
tworzą też antynarzędzia  
innymi słowy  
jest hype na jedno  
który ciągnie drugie  
i oczywiście powstają wirusy  
ale są też firmy, które budują  
bardzo zaawansowane programy anty wirusowe  
więc być może  
to co pojawiała się w czasie naszej  
dzisiejszej rozmowy kilkakrotnie  
że może to być zagrożenie  
na przykład dlatego, że przestaniemy rozróżniać  
co jest w czym  
za chwilę standardem mogą się okazać  
takie

okulary  
czy nie wiem  
jakieś wzbogacone funkcje naszych przeglądarek  
czy czegośkolwiek innego  
które będą w prawym górnym rogu wyświetlały  
czerwony wykrzyknik  
ale bezpiecznik wykrzyknik  
mówiący to nie jest  
stworzone przez człowieka  
czy to będzie gwarancja  
jakości bądź niejakości  
to jest zupełnie inny temat  
skoro sztuczna inteligencja  
ma dostęp do wiedzy z zakresu ludzkiej psychiki  
manipulacji, wywierania wpływu  
to czy może być doskonałym narzędziem  
do nakłaniania ludzi do różnych rzeczy  
nie tylko może, ale już jest  
i będzie  
i począwszy od takich decyzji  
może mało  
doniosłych jak to co oglądamy  
czego słuchamy, czy co czytamy  
z czym już mamy dzisiaj do czynienia  
przez może bardziej  
wygrywane decyzje  
czy takie bardziej wpływające na nasze życie  
ja sobie z łatwością mogę  
czy ja tego nie powiem o patentować  
wybory, co zaraz będę  
nie, ja mam  
nie aż tak daleko  
czekaj, bo tutaj jakiś papro jest  
mogę sobie wyobrazić  
aplikacje, do której będziemy  
wrzucali odpowiednio dużo danych  
będziemy podłączali  
na przykład nasz zegarek  
może mnóstwo innych rzeczy  
będzie nam mówiła, co w danym momencie  
powinniśmy kupować  
w danym sklepie po to, żeby  
było zdrowo  
ubezwłasną wolnienie

i teraz tak, jasne, my zawsze możemy wyłączyć  
tą apkę  
podniesą się ubezpieczenie  
ale tak, może tak być  
że np. wtedy będę miał  
inne ubezpieczenie  
albo wtedy może się okazać, że będę miał  
inne nie wiem, składki  
np. zdrowotne, cokolwiek  
albo będę czekał w krótszej  
kolejce do lekarza  
mogę sobie wyobrazić taką sytuację  
więc  
podejmowanie  
naszych decyzji  
czyli inaczej P, jeżeli ktoś  
myśli, że my podejmujemy  
dzisiaj, nie za 50 lat  
czy za 10, że my dzisiaj podejmujemy  
nasze wszystkie decyzje  
w tym decyzji konsumenckie  
bazując tylko i wyłącznie  
na naszej woli  
jest w grubym błędzie  
są  
armie specjalistów  
w różnych dziedzinach  
które uczą, którzy  
ludzie uczą się  
naszych zachowań  
po to, żeby nas namówić do podjęcia  
takiej decyzji ani innej  
i algorytmy  
są doskonałym narzędziem  
ich pracy  
i to, że mówimy  
choćby w kontekście Facebooka  
o wyborach w Stanach  
o Brexicie  
o wielu, wielu różnych  
dziwnych zachowaniach, niespodziewanych  
zachowaniach, niespodziewanych  
trendach, to są przykłady  
na to, że nie w przyszłości

algorytmy będą na nas  
spływać, tylko już od dawna  
na nas spływają, a w przyszłości będzie bardziej  
czyli znowu bardziej  
to jest najczęściej pawiające się hasło  
chyba w czasie dzisiejszego spotkania  
żeby to wszystko już było  
jeszcze bardziej. Masz jakieś pytanie  
czy pytamy Krzemysława?  
Ja bym jeszcze tutaj zadał  
co sądzicie o eksploataowaniu AI  
przykład, AI, który grał mistrza gry w goło  
okazało się, że nie rozumie zasad gry i łatwo już ukać  
z prostym trikiem. Zaawansowany  
AI może mieć inne czyli breaking  
to jest znowu moja panieczka, dlatego ja lubię takie  
Moim zdaniem cały czas  
istnieje jeszcze pewien  
zasubkreatywności w nas, w ludziach  
którym jesteśmy w stanie obejść  
to, gdzie są te filtry  
związane z AI i z innymi takimi  
ja swojego czasu uwielbiałem Starcrafta  
tam też pojawiała się rzeczy, które były  
związane z AI, które  
gdzie Polak  
grał przeciwko właśnie jednemu  
z takich modeli i udało mu  
się go zapętlić. To znaczy  
udało mu się tam wysłać swoją armię, więc on  
wracał do obrony, potem on wracał  
i tak dużą część armii wroga  
niestłumiecznie udało mu się jakby zablokować  
w związku z tym znalazł obejściem  
jailbreaki są, jailbreaki są  
obejściem. Wyobrażam sobie, że my jako  
ludzie będziemy znajdowali różne  
takie właśnie obejścia tych systemów  
i to jakby się chwali  
to jest ta piękno naszej kreatywności  
i tej społeczności, bo my cały czas usprawniamy  
te narzędzia przez to, że poszukujemy  
w nich dziur, one się stają coraz lepsze, coraz  
bardziej odporne i to pewnie

będzie trwało, my jesteśmy taką chmurą  
która korzysta  
ze swojej takiej inteligencji  
nie stada  
takiego wiezroju, które  
poprawia te narzędzia  
i wyobrażam sobie, że  
będą czasami  
takie no prymitywne  
wpadki, to znaczy  
no hej, ale jak to, przecież to jest  
ten stół jest czarny, czemu nie wiesz, że ten  
stół jest czarny, nie? Może tak być  
i to jest absolutnie naturalne  
i to nie wynika z tego, że  
ten model będzie jakiś taki  
mniej lub bardziej dopasowany, po prostu nie został  
wytrenowany w tym  
kawałku albo nie został, nie miał tego w danych  
uczących itd. i takie wpadki będą się pojawiać  
i takie wpadki będą wykorzystywane  
do tego, żeby obchodzić takie modele  
więc wyobrażam sobie, że  
eksploatowanie AI będzie codziennością  
i jak z każdym młotkiem  
będzie w dobrych celach i w złych  
celach. Czasami się uderzymy w palec  
Dobra Krzemysławie, co zrobić,  
żeby zminimalizować ryzyko  
wystąpienia niebezpieczeństw związanych  
ze sztuczną inteligencją?  
Istnieje wiele  
sposobów, aby  
zminimalizować ryzyko  
wystąpienia niebezpieczeństw związanych  
z AI. Poniżej przedstawiam  
kilka z nich.  
Projektowanie etyczne, AI  
powinna być projektowana w sposób etyczny  
i zgodny z wartościami społecznymi.  
W ten sposób można uniknąć sytuacji  
w których AI działa w sposób niezgodny  
z oczekiwaniami i wartościami  
społeczeństwa.

## [Transcript] Nauka To Lubię / Czy sztuczna inteligencja stanowi dla nas zagrożenie?

Ścisłe monitorowanie, organizacje AI powinni monitorować wykorzystanie swoich systemów i działań, aby zapewnić, że AI działa zgodnie z ich intencjami i wartościami.

Bezpieczeństwo, AI powinna być zaprojektowana w sposób bezpieczny i odporny na ataki.

W ten sposób można uniknąć zagrożeń związanych z cyberbezpieczeństwem i wykorzystaniem AI do celów niezgodnych z etyką i prawem.

Edukacja i szkolenia, ludzie powinni być edukowani na temat AI i jej możliwości oraz zagrożeń.

W ten sposób można zwiększyć świadomość i zrozumienie AI, co umożliwi bardziej efektywne kontrolowanie i wykorzystywanie tej technologii.

Współpraca międzynarodowa, współpraca międzynarodowa i wymiana wiedzy na temat AI są ważne, aby zminimalizować ryzyko wystąpienia niebezpieczeństw związanych z tą technologią.

Transparencja, algorytmy i systemy AI powinny być przejrzyste i możliwe do zrozumienia przez ludzi. W ten sposób można uniknąć błędów i zapewnić, że decyzje podejmowane przez AI są zgodne z oczekiwaniami i wartościami społeczeństwa.

Wnioskując, aby zminimalizować ryzyko wystąpienia niebezpieczeństw związanych z AI, konieczne jest podejście, które uwzględnia różne czynniki i interesy.

To wymaga zaangażowania ze strony twórców AI, rządu, społeczeństwa i innych zainteresowanych stron.

Przyszło mi do głowy, że Krzemysław mógł mieć zdrobnienia. Ale co, krzemek?

Krzemek? Krzemek?

Krzewiwony?

Krzemek chyba lepszy.

Nie wiem, czy powinniśmy o tym powiedzieć?

Tak powinniśmy.

Krzemysław oczywiście

jest wytworem sztucznej inteligencji.  
I tu jakby nie chodzi  
czy algorytmów.  
I tu nie chodzi tylko o jego postać,  
ale tutaj chodzi też o to, co mówi.  
Po prostu pytamy różne algorytmy  
czy różne modele językowe.  
Zadajemy pytania  
i to, co Krzemysław odpowiada,  
to to jest to, co one  
z siebie wypływają.  
No, prowadzamy go, ale ogólnie  
rzecz biorąc to jest to, co uzyskujemy.  
Pomyśleliśmy, że nie ma lepszego  
gościa do  
programu oczyszcznej inteligencji  
niż tuczna inteligencja.  
A byłoby nudno, gdybyśmy  
z Mateuszem po prostu to czytali,  
więc zatrudniliśmy Krzemysława.  
On jest tym bardzo dobrym pracownikiem,  
że on nie chce nas pieniędzy.  
Prawie w ogóle to.  
Nie chce urlopu.  
Jest zawsze do dyspozycji.  
A nie na wszystko odpowiada.  
Wiesz, co on czasami kręci.  
Ja myślę, że jakby się dokładnie  
wysłuchać w te odpowiedzi, które on podaje.  
One są takie obłe.  
Ładnie powiedziane.  
One są takie trochę...  
Mydliny, trochę jakie czasami.  
Tak, tak, tak.  
I stąd zastanawiam się,  
bo często pojawia się  
szczególnie w ostatnich miesiącach,  
że w zasadzie może za chwilę  
dziennikarze nie będzie.  
Duże redakcje korzystały  
z algorytmów do pisania krótkich tekstów  
już od wielu lat.  
Czy Krzemysław byłby w stanie napisać  
świetny reportaż?



Nie.

Czy Krzemysław byłby w stanie napisać porządną analizę?

Nie. Pytanie jest inne.

Czy my jako odbiorcy chcemy dobrego reportażu i czy chcemy świetnej analizy?

Emocje.

Bo to jest to pytanie.

Ja niedawno miałem z moim bardzo serdecznym pomysłem nawet kilku znajomych pewien rodzaj sporu.

Tak się akurat złożyło,

że wszyscy byliśmy dziennikarzami.

No i była dyskusja na temat tego,

czy zawód dziennikarza

zostanie po prostu wyparty

przez algorytmy.

Zdania były różne,

natomiast wydaje mi się,

że to pytanie powinno być

wzmocnione

albo uzupełnione innym pytaniem.

Czy my odbiorcy

chcemy tekstów

wysokiej jakości?

Czy my odbiorcy chcemy

wysokiej jakości treści?

Bo pamiętam moment, w którym

absolutnym standardem było to,

że każda gazeta miała

porządnego fotoreportera

i go wysyłała w różne

miejsca po to, żeby robił porządne zdjęcia.

Aż w końcu któraś gazeta

nie wspomnę nazwy, będzie ale po co.

Czy nam zwiększa się sprzedaż

przez to, że te zdjęcia są

rewelacyjne?

Może wystarczy, że dziennikarz po prostu

będzie miał aparat.

One nie będą rewelacyjne,

ale ilu z nas odbiorców

oczekuje rewelacyjnych zdjęć.

Najczęściej na tyle

wąska grupa ludzi,  
że oni poszukają  
gdzieś indziej niższe,  
bo w większości z nas wystarczy  
nieświetny reportaż,  
czy bardzo dogłębna analiza.  
Tylko właśnie czasami takie mydliny krzemysława.  
Całym oczywiście szacunku do krzemysława,  
bo ja myślę, że nie życzę.  
Absolutnie.  
Pytanie konkretne.  
Ja mam tutaj tylko taki krótki ten,  
bo to fajnie pokazuje  
moim zdaniem cały ten taki proces  
zakupowy.  
Dlaczego ja kupuję tekst  
czytając go od jakiegoś  
dziennikarza, od innego nie.  
I oczywiście o tym, że tam  
zdjęcia, że jest tekst napisany  
w fajny sposób, oczywiście.  
Tam nie jest jeszcze element  
caufania. I teraz  
wyobrażam sobie, że możesz mieć  
dziennikarzy śledczych, którym ufasz,  
bo wiesz, że oni robią robotę,  
weryfikują i sprawdzają.  
Ten problem, który moim zdaniem tu się pojawia  
a propos dziennikarzy, to jest takich,  
że nie jesteś w stanie zaufać tej technologii,  
bo  
nie masz doświadczeń, wiesz, że jest pewna losowość,  
a w całym tym procesie  
nie ma tak jak powiedział krzemysław  
pewnej transparentności.  
Znaczący, hej, ja na podstawie tego  
jestem w stanie wywnioskować takie rzeczy.  
Ten  
trend nazywa się explainable AI, czyli  
wyjaśnialna sztuczna inteligencja  
i moim zdaniem jest jedynym sposobem  
do tego, żebyśmy my jako ludzie byli  
w stanie  
nasze zaufanie po prostu włożyć w tego typu

tworach. Innymi słowy,  
przyszłość to taka naklejka  
made by humans?  
To jest ciekawe, nie? Bo tak jak tam  
C i China  
to zobaczymy, czy to nie zostanie  
podrobione. Natomiast wyobrażam sobie,  
że tak, podpis, pod tytułem ten tekst  
został napisany przez człowieka  
ciężko będzie to zweryfikować, ale  
może być to pewna pieczętka z ziemniaka,  
która będzie coś kiedyś znaczyła.  
Dobra, do pytań.  
Pytanie, czy  
edukacja  
w szkołach, na przykład prac domowy, inne egzaminy  
mają sens? Co w ogóle  
z egzaminami, czy to ma w ogóle sens  
w tym momencie, gdy dostęp do informacji  
będzie ogromny.  
To teraz z kolei  
to jest moja bańka i to jest wykonik,  
ale  
od razu Wam mówię, że my o tym zrobimy  
osobny program, bo to jest tak ważna rzecz  
w ogóle zmiana całego paradygma  
tu edukacji.  
I to absolutnie wszystkiego, znaczy  
począwszy od prac domowych, przez testy.  
Ja nie chcę absolutnie powiedzieć,  
że w przyszłości młodzi ludzie  
nie będą musieli się uczyć.  
Dlatego, że w jakimś  
urządzeniu będą to wszystko mieli.  
Bo  
proces  
uczenia się to tylko  
częściowo, jakby jednym z  
celów, tylko jednym  
z, jest to, żeby  
coś wiedzieć na końcu.  
Ale dla mnie dużo ważniejszym  
jest  
dochodzenie do tego punktu.

## [Transcript] Nauka To Lubie / Czy sztuczna inteligencja stanowi dla nas zagrozenie?

Innymi słowy, tak najbardziej skrótowo sama pogoń za czymś, jest czymś, co nas rozwija. A niekoniecznie dojdzie do mety. I oczywiście tak, i oczywiście, że łatwiej jest nam sprawdzić jakąś konkretną informację w jakimś urzędzeniu. Pytanie tylko, czy to nas rozwija. Czy to nie spowoduje, że my stracimy w ogóle kontekst. Że my stracimy rozumienie tego świata. I już często jesteśmy w nim pogubieni. Natomiast proces edukacji powinien polegać nie na tym, że my od młodego człowieka wymagamy takiej wiedzy, którą sobie może sprawdzić. 10 lewych, 10 prawych dopływów Wisły. Ja się uczyłem. No uczyłem się, bo tego musiałem się nauczyć, żeby mieć nie wiem, jakąś tam ocenę z geografii. To czy to miało sens? Nie. Ale się uczyłem. Wszyscy geografowie teraz to. Tak, być może tak. Ja bym wolał, żeby osoba no dobra, to były inne czasy. Ale osoba, która mnie uczyła geografii pokazała mi dzisiaj, bo być może dalej siedzicie tego uczą, żeby mi pokazała, jak skąd pobrać dane satelitarne i jak je przeanalizować tak, żeby znaleźć tych 10 lewych i tych 10 prawych dopływów Wisły. Wydaje mi się to dużo bardziej przydatne. Natomiast to jest dobra, to jest temat edukacyjny. Natomiast, jeżeli

zrobimy o tym osobno odcinek,  
z całą pewnością  
musimy się nauczyć, ale to znowu  
nie jest nic nowego, tylko  
jest bardziej. Musimy się nauczyć  
nie to, co uczeń wie albo czego nie wie  
na końcu, tylko  
musimy nauczyć się  
weryfikować proces,  
jaki następuje w jego głowie.  
Niestety  
mam wrażenie,  
nie, mam pewność, że  
ostatnie lata to jest  
raczej odchodzenie od tego.  
To, że na przykład na maturze  
niemal, że rozwiązuje się testy  
i jest klucz, którego  
się musimy sztywno trzymać,  
to ja mówię, no ok,  
dobra, matura może jest specyficznym egzaminem,  
ale dalej uważam, że jest  
bardzo źle zaprojektowana.  
Ale wiele innych egzaminów także  
eliminujemy  
czy nie wykorzystujemy  
żadnych narzędzi, dzięki  
którym poznajemy  
proces, który powstaje w głowie ucznia.  
A to powinno być kluczem.  
A nie to, czy na końcu  
on napisze  
dobrą rozprawkę czy złom.  
Ta rozprawka jest tylko  
jakby w finałem pewnej pracy.  
Jeżeli tylko i wyłącznie  
będziemy to oceniali,  
to w ogóle z bezsensu.  
To narzędzia,  
które już dzisiaj mamy, mogą nas  
pięknie wyręczyć. Absolutnie.  
Natomiast te narzędzia  
nawet gdyby były doskonałe,  
jeżeli uczeń przyniesie,

przysła, nauczyciowi  
tę pracę, te narzędzia  
nie pokażą  
i to powinno być kwestią  
oceny.

Jakie procesy myślowe zachodziły  
w jego głowie?

Ja mam tylko taką moją gwiazdkę.

Wyobraź sobie teraz  
takie zatrucie fałszywymi informacjami,  
że wiesz,

w Wikipedia 50% informacji  
jest fałszywe. Powiedzmy.

I to jest twoja główna źródła wiedzy.

Dla mnie kwintesencją będzie to,  
żeby ktoś był w stanie mieć taki tok  
rozumowania, żeby wiedział, jak  
odsiąć to, co jest prawdą,  
od nieprawdy.

Tak, ale to pogadamy.

Zrobimy osobny odcinek o tym.

Ja tylko tak zaznaczyłem, żeby ktoś z was  
nie myślał, że ten temat ignorujemy.

To jest bardzo ważny temat, na tyle ważny, że  
zrobimy o nim osobny odcinek.

Czy to już będzie ostatnie?

Chyba, że jeszcze dobra. Ostatnia szansa  
zadawania pytań. My je tutaj dostajemy  
nawet dostajemy informacje, że podobno  
czytacie mi sekranu.

Dlatego co jakiś czas go przygaszam,  
żeby

nie było tak łatwo.

Pytanie jakies?

Czy w perspektywie 10 lat  
lepiej zostać programistą czy stolarzem?

To jest pytanie trochę apokalityczne.

Ja dalej

uważam, że każdemu wedle potrzeb  
znaczy, jeżeli bawić się  
tworzenie

rzeczy z drewna, to zostań stolarzem.

Ludzie, którzy pewnie są najlepsi  
są w stanie żyć i być szczęśliwistym.

## [Transcript] Nauka To Lubię / Czy sztuczna inteligencja stanowi dla nas zagrożenie?

Natomiast, jeżeli chcesz się bawić rzeczami związanymi ze sztuczną inteligencją, to moim zdaniem bycie programistą jest jednym z najprostszych metod, żeby tam dojść, bo będziesz w stanie rozumiać pewną logikę tych zachowań. Nie znaczy, że nie będąc programistą, nie jesteś w stanie się tego nauczyć, ale będąc programistą po prostu masz pewną przewagę. Patrząc 10 lat w przód, pewnie bycie człowiekiem renesansu, to znaczy i trochę data science, i trochę programowania, i trochę czegoś, co nazywa prompt engineeringiem, czyli pisaniem właśnie tych zapytań do sztucznej inteligencji, spowoduje, że będziesz na szpicy tego rozwoju. Ale to zależy, co optymalizuje. Ale pomiesz sobie zrobić szafkę kuchenną, będąc profit, będąc gościem o sztucznej inteligencji, tym lepiej. Dokładnie, także to zależy, co chcesz mieć. Optymalizujesz zabawę, czy zarobki, to od ciebie zależy. Ale czekaj, bo ty teraz powiedziałaś, to tak, jak gdyby było wybór między albo zarabiasz i zostajesz programistą, albo zostajesz stolarzem i wtedy się bawisz? Tak to rozumiecie? No nie, jeżeli bawicie programowanie, to spoko. Nie ok, a czy zasada ogólna jest taka, że żeby ktoś był dobry w tym, co robi, to musi to lubić. A jeżeli tego nie lubi, no to nie będzie dobre. To się męczy. Życie, mówi się, życie jest krótkie. Nie. W tym kontekście jest naprawdę nieskończenie długie. Jak przez kilkadziesiąt lat masz się w człowieku męczyć, z tym czego nie lubisz, to ja ci naprawdę mega, mega współczuję. I z całą pewnością wcześniej, czy później, a im później



tym gorzej ci będzie, zmienisz pracę.  
Więc najlepiej jest  
rozwijać się w tym kierunku, który  
jest w sercu i to nie jest  
z podręcznika amerykańskiego,  
tylko to jest swojego życia,  
życiowego doświadczenia. Dobra, to może Krzemysław  
po raz ostatni, ok?  
Trochę o tym powiedziałaś, ale  
zadałem mu pytanie, czy sztuczna inteligencja  
może skrzywdzić człowieka.  
Tak, w teorii AI może skrzywdzić  
człowieka albo przez zaniechanie działania  
dopuszczyć do wyrządzenia krzywdy człowiekowi.  
To wynika z faktu,  
że AI jest tylko narzędziem,  
które działa na podstawie zaprogramowanych  
algorytmów i uczenia maszynowego.  
Jeśli AI jest zaprogramowane  
w sposób nieodpowiedni lub uczy się  
na podstawie nieodpowiednich danych,  
może działać w sposób niezgodny z oczekiwaniami  
i wartościami społeczeństwa.  
Przykładem takiego  
niepożądanego zachowania AI może być sytuacja,  
w której samochód autonomiczny  
nie zauważył przechodnia na drodze,  
ponieważ nie został on odpowiednio wykryty  
przez czujniki samochodu, a algorytmy  
nie uwzględniły tej sytuacji w swoich programach.  
W takiej sytuacji samochód autonomiczny  
mógłby wpaść na przechodnia, co doprowadziłoby  
do poważnej krzywdy.  
Już, szybciej skończył?  
Tak mnie zaskoczył trochę, nieco.  
Jakiś komentarz?  
Szybciej zaczniemy, później skończymy.  
Dla mnie te elementy, które się tutaj pojawiły,  
pod tytułem  
zasady etyczne,  
to są rzeczy, które nie są wspólne  
dla wszystkich, więc to jest trochę taka  
morzonka. Nie wyobrażam sobie,  
że mamy wspólne działanie.

## [Transcript] Nauka To Lubie / Czy sztuczna inteligencja stanowi dla nas zagrozenie?

Ludzie powinni być edukowani,  
tak, o tym cały czas mówimy, w sumie  
po to chyba też są te streamy,  
które robimy, żebyście mogli  
pozadawać pytania, żeby podzielić się  
wiedzą i oczywiście też się pomylić,  
bo to się tak szybko zmienia, że za chwilę  
pewnie część tego, co mówimy dzisiaj,  
jest nieaktualne.  
I jeszcze jest ten element, który jest związany,  
jak Szemysł powiedział, zaprogramowany.  
Programowanie w takim normalnym  
znaczeniu to zazwyczaj są  
rzeczy tak albo nie  
w przypadku modelu uczenia maszynowego.  
Mamy do czynienia z pewnym  
prawdopodobieństwem i  
problem polega na tym, że to nie jest tak łatwo  
wytyzylować  
co z tym tak naprawdę do końca się stanie.  
Więc  
to nie jest tak, że można wszystko  
ładnie zaprogramować, bo nawet jak będziemy chcieli  
programować te ograniczenia, pod tytułem  
Nie Zrób Krzywdy Człowiekowi,  
no to też  
będzie nieidealne w swojej naturze.  
Więc ja to znowu oceniam jako obłe.  
Jasne. Wiesz co, jak w ogóle  
wszystkiego odpowiedzi z grubsza są  
takie same.  
Nie wiem czy jakby to odkrywasz,  
ale ja  
miałem  
taki epizod krótki  
nawet niejako tam bardzo krótki,  
kiedy pracując na uczelni  
egzaminowałem  
i studen, który by mi tak odpowiadał  
byłby  
przeze mnie uznany, także  
gość po prostu  
jest w miarę ogarnięty  
wie

że nie ma pojęcia o czym my tu rozmawiamy.  
Natomiast jest na tyle  
ogarnięty, że wie jak obrócić  
tym samym zdaniem  
pięciokrotnie, żeby zrobić wrażenie, że wiatr.  
Tak, tak, tak.  
Nie wiem czy ty masz też podobne wrażenie, że  
ładnie, ładnie powiedzieć.  
Że krzemek trochę dał dzisiaj  
Ja myślę, że to też może wynikać  
z tego, że wiesz, że znalazł się  
w pewnym jakimś takim kawałku przestrzeni  
i dookoła jakby tego ciągle  
dążył. Nie wiem  
na ile można by tutaj naprowadzić  
go z innego miejsca. To znaczy, hej  
weź pod uwagę teraz, nie wiem  
warunki w Polsce i odpowiedz mi na to pytanie.  
Może zrobimy sobie taki eksperyment, żeby  
zobaczyć innym razem, czy da się  
coś takiego naprowadzić. Natomiast  
tak jak da się  
zachęcić do zupełnie innych  
odpowiedzi tego rodzaju modelu, tak  
wyobrażają sobie, że krzemka też możemy zmusić, żeby  
powiedział coś zupełnie, wiesz,  
coś zupełnie innego. Więc  
ja to traktuję jako  
zlepek prawdopodobieństwa, na które  
akurat trafiliśmy. Część  
tego rezonuje i ma sens.  
A część, no jakby  
tak jest bojęsty.  
Okej, ale patrząc na to  
jak odmówi, patrząc na to  
co on mówi. To większość  
tego, o czym dzisiaj mówiliśmy.  
O potężnych narzędziach.  
O tym, że to nas może wykoleić.  
Że to wywali w kosmos  
nie wiem, rynek pracy.  
I edukacji, i wszystko, i wszystko.  
Można mieć pewien dysonans.  
To znaczy, on

wiesz, to nie jest  
coś, co urywa głowę. To jest  
coś, co jest fajnym gadżetem.  
Ale to jest coś,  
co na pierwszy rzut  
ucha słysząc,  
że to jest właśnie nie rzeczywiste,  
że to jest naciągane, że to jest obłe,  
że to jest nijakie.  
Nie masz takiego wrażenia,  
że trochę się poruszamy,  
że albo krzemek dzisiaj nie jest w formie,  
i dlatego tak trochę,  
albo, albo, albo  
może on po prostu jest jakąś  
wersją taką bardzo, bardzo wsteczną.  
Ja mam wrażenie, że to wiesz,  
to będzie fajnie widać, bo  
wyobrażam sobie, że będziemy obserwować też  
ewolucję krzemkę. To znaczy,  
to jak teraz odpowiada,  
ten model będzie się cały czas zmieniać.  
Będziemy widzieć pewnie w czasie,  
jak odpowiedzieć na lepsze.  
To będzie sobie pytanie, które  
będzie takim pytaniem sprawdzającym,  
takim kanarkiem.  
Natomiast  
to, co uzyskujesz  
zależy od tego, jak pytasz.  
Więc  
tu jest tak bardzo dużo zmiennych,  
że nie jesteśmy w stanie bardzo szybko  
i jednoznacznie powiedzieć, dlaczego on właśnie  
odpowiedział w ten sposób. My mu zadaliśmy  
proste pytania, te pytania, które  
widzieliście gdzieś tam na tablicy.  
I na podstawie tych pytań  
on odpowiedział najlepiej, jak umiał.  
Ale też nie byliśmy specyficzni, wierzę.  
W listu i mi wszystkie rzeczy, albo jak uważasz  
z perspektywy prawa, albo z perspektywy  
praw podstawowych, albo praw człowieka,  
to można na minie sposób ozegrać.

Ale wtedy on był mniej obły?  
Wtedy byłby bardziej specyficzny.  
Okej, ale czyli te wszystko, co budzi  
jakiś taki mój niedosyt,  
to wynika z tego,  
że zostawiliśmy mu za dużo swobody?  
Tak, tak, tak.  
Bo teraz, to jak działają te modele,  
gdzieś do pierwszego naszego strumyka,  
to  
wymyślasz, jak to będzie kolejnym słowem  
z pewnym prawdopodobieństwem.  
I teraz, jak mówimy tak ogólnie, to jesteś  
w pewnym miejscu przestrzeni, gdzie jesteś...  
Kiedy on raczej będzie specjalistą, niż  
takim kimś ogólnie wykształconym,  
który będzie w stanie się znaleźć  
ze swoją wiedzą  
z dużą,  
bo  
z całego zasobu internetowego  
będzie się w stanie znaleźć w każdej sytuacji.  
Wyobrażam sobie teraz sytuację,  
w którym moglibyśmy powiedzieć  
na podstawie książki  
takiej i takiej, albo artykuł  
albo jakiegoś badania  
opowiedz mi o tym i o tym, i on będzie  
o wiele bardziej konkretne.  
Czyli umieścimy go w pewnym kontekście,  
w którym chcielibyśmy,  
żeby on gdzieś tam kumał  
i w ramach tego kontekstu,  
on będzie o wiele bardziej specyficznym,  
będzie wiedział, gdzie się znaleźć.  
I po prostu trafi na pewne prawdopodobieństwa,  
w kontekście to ja wam mówię o tym, o tym,  
o tym, bo tam się to znajdowało.  
Tu jest znowu ten problem,  
że one ten maszyny mają  
ograniczoną pamięć, jeżeli chodzi o nasz kontekst.  
Natomiast już  
powstają, i to jest teraz ciekawy taki trend,  
który pojawił się ogólnie

## [Transcript] Nauka To Lubię / Czy sztuczna inteligencja stanowi dla nas zagrożenie?

w inwestycji, w nowe start-upy,  
że teraz takie bazy wektorowe,  
które mają za zdanie przechowywać te twory,  
te tam splątania, które są związane  
z sztuczną inteligencją  
w nie mega inwestuje  
venture capital, czyli wszystkie te fundusze,  
bo widzą, że to jest potrzebne,  
że rozszerzenie pamięci dla takich modeli  
to jest absolutna przyszłość,  
więc teraz trwa absolutny boom  
na rynku takich baz wektorowych.  
I to jest znowu akcja reakcja.  
Zobaczyliśmy, widzimy, że jest  
zapotrzebowanie, trzeba by to rozszerzyć.  
W tej chwili jest pompowane w to mnóstwo pieniędzy.  
I tym optymistycznym akcentem  
dziękujemy Grzemkowi,  
dziękujemy wam.  
Zapraszamy na dwa poprzednie lajwy,  
na trzeci na Nauka to Lubię, możecie je znaleźć.  
Bardzo dziękujemy wam za mnóstwo  
pytań, nie na wszystkie odpowiedzieliśmy.  
Niektóre z nich, żeśmy  
zignorowali specjalnie, dlatego,  
że będziemy oni robicę odcinki.  
Niektóre, żeśmy tylko zajawili  
z obietnicą, że będą kolejne odcinki.  
Kolejny odcinek, mniej więcej  
za miesiąc, czyli w okolicy końcówki  
maja, w drugiej części maja,  
poinformujemy o tym na mediach społecznościowych  
Nauka to Lubię.  
Mateusz u siebie oczywiście też poinformuje.  
Tak, podziękowania dla Uniwersytetu Śląskiego.  
Jak wspominałem,  
Uniwersytet nas tutaj gości,  
w swoim studio, jeżeli chcecie  
wesprzeć Fundację Nauka to Lubię,  
zawsze możecie wejść  
i te działania, które prowadzimy,  
zawsze możecie nas znaleźć na zrzutce  
i tam nas wesprzeć, albo po prostu  
wejść na sklep Nauka to Lubię

i na przykład kupić książki,  
nie tylko dla dzieci, ale także dla dzieci  
książki naukowe.  
O czym się jeszcze powinienem powiedzieć?  
Ja o kosmosie już widziałem, polecam.  
Ładna jest, nie? Pięknie ilustrowana.  
Rzeczywiście tak jest, no mogę jak mówić,  
bo to nie ja ilustrowałem, więc ich nie może mi zarzucić.  
I nie było to jaj.  
I to w zasadzie wszystko. To już wszystko.  
Bardzo dziękujemy za waszą aktywność  
i myślę, że w przyszłym miesiącu  
znowu będziemy mieli garść nowości  
a co do tematu to zobaczymy.  
To dam jeszcze znać. Dzięki, dobrej nocy.  
Napisy stworzone przez Amara.org