

Max Paradox

PROBLEM? WYLUZUJ. OGARNIESZ TO!

Seria poradników z przymrużeniem oka ;)

Nie wiem już, czy to zdjęcie jest prawdziwe

Jak przeżyć epokę deepfake'ów, AI i internetu,
który potrafi podrobić wszystko



Mamy problem. Ale spokojnie. Da się go rozwiązać
i jeszcze będzie z czego się pośmiać.

SPIS TREŚCI

INTRO	1
Rozdział 1 - Oczy już nie wystarczą	10
Rozdział 2 - Najgroźniejszy deepfake może wyglądać jak twoja mama	19
Rozdział 3 - Najłatwiej oszukać cię tym, w co już chcesz uwierzyć	29
Rozdział 4 - Zrzut ekranu nie jest dokumentem z archiwum państwowego	39
Rozdział 5 - Detektor AI też może się pomylić	51
Rozdział 6 - Tysiąc komentarzy nadal może się mylić	62
Rozdział 7 - „Pilne” to często tylko inny sposób na wyłączenie ci myślenia	73
Rozdział 8 - Nie wszystko, co fałszywe, zostało stworzone przez AI	85
Rozdział 9 - Zanim uwierzysz, znajdź oryginał	97
Rozdział 10 - Metadane, znaki wodne i inne rzeczy, które brzmiały mądrzej, niż są trudne	109
Rozdział 11 - Procedura 90 sekund przed udostępnieniem	122
Rozdział 12 - Jak sprawdzić zdjęcie bez zostawiania cyfrowym Sherlockiem	134
Rozdział 13 - Film też trzeba sprawdzać poza filmem	148
Rozdział 14 - Nie sprawdzaj tylko treści. Sprawdzaj, kto do ciebie mówi	161
Rozdział 15 - A jeśli nie da się ustalić, czy to prawdziwe?	176

Rozdział 16 - Co zrobić, kiedy twoje własne oczy zostały oszukane	189
Rozdział 17 - Kiedy wszystko zaczyna wyglądać podejrzanie	202
Rozdział 18 - Zbuduj sobie cyfrowy pas bezpieczeństwa..	216
Rozdział 19 - Nie wygrasz z internetem. I bardzo dobrze .	230
Rozdział 20 - Nie musisz wiedzieć, co jest prawdziwe od razu.....	243

INTRO

Dostajesz wiadomość od znajomego. „Widziałeś to?”. Otwierasz film. Znany polityk mówi coś kompletnie absurdałnego. Albo aktor reklamuje inwestycję, dzięki której podobno każdy może zarabiać kilka tysięcy tygodniowo, wystarczy wpłacić skromne 999 złotych i najwyraźniej nie zadawać żadnych pytań. Głos brzmi prawidłowo. Twarz wygląda prawidłowo. Usta poruszają się we właściwym rytmie. Nawet charakterystyczne zmarszczki robią mniej więcej to, co powinny.

Patrzysz jeszcze raz.

No przecież mówi.

Pięć lat temu prawdopodobnie uznałbyś, że skoro istnieje nagranie, to sprawa jest jasna. Obraz był dowodem. Film był jeszcze lepszym dowodem. Nagranie głosu? Proszę bardzo, zamykamy śledztwo, dziękujemy wszystkim za udział. Dzisiaj sytuacja wygląda trochę inaczej. Masz film, dźwięk, zdjęcie i nadal możesz znajdować się dokładnie w tym samym miejscu co człowiek, któremu ktoś powiedział: „Mój kuzyn pracuje w urzędzie i mówił, że od przyszłego tygodnia zamykają internet”.

Czyli właściwie nigdzie.

Problem nie polega na tym, że internet nagle zaczął kłamać. Internet kłamał, odkąd pierwszy człowiek odkrył, że może napisać „lekarze go nienawidzą” nad zdjęciem dziwnego owocu i nikt nie przyjdzie do niego do domu z komisją kontrolną. Zmieniła się jakość podróbki. Kiedyś fałszywe zdjęcie mogło wyglądać jak efekt piętnastu minut pracy człowieka, który właśnie odkrył Photoshopa i zamierza wykorzystać wszystkie dostępne filtry jednocześnie. Dzisiaj

wygenerowany obraz potrafi być piękny, realistyczny, szczegółowy i przekonujący.

Czyli dokładnie taki, jaki nie powinien być, jeśli liczyłeś na łatwe życie.

Sztuczna inteligencja potrafi tworzyć twarze osób, które nigdy nie istniały. Może imitować głos, zmieniać wypowiedzi na filmie, generować sceny wyglądające jak fotografia reporterska, produkować fałszywe dokumenty i tworzyć materiały, które przy pierwszym spojrzeniu nie mają żadnego obowiązku wyglądać podejrzanie. Jednocześnie zwykłe narzędzia do edycji obrazu są prostsze niż kiedykolwiek. Nie trzeba już być grafikiem, montażystą i czarodziejem trzeciego poziomu Photoshopa. Wystarczy urządzenie, kilka minut i odpowiednia aplikacja.

To oczywiście nie oznacza, że od tej chwili wszystko jest fałszywe.

I właśnie tutaj zaczyna się drugi problem.

Kiedy ludzie odkrywają, że obrazu można już łatwo podrobić, pojawia się pokusa, by przyjąć bardzo wygodną zasadę: niczemu nie wierzę. Film pokazuje coś, czego nie lubię? AI. Zdjęcie jest niewygodne? Photoshop. Nagranie kompromituje kogoś, komu kibicuję? Deepfake. W ten sposób technologia, która pozwala tworzyć fałszywe dowody, daje nam również fantastyczne narzędzie do ignorowania dowodów prawdziwych.

Idealnie.

Najpierw nie wiadomo, czy można wierzyć temu, co widzimy.

Potem każdy może powiedzieć, że nie wierzy temu, czego nie chce widzieć.

Witamy w epoce informacyjnej, w której nawet rzeczywistość musi mieć dział PR.

Dlatego ta książka nie będzie próbowała nauczyć cię rozpoznawania każdej fałszywej fotografii po liczbie palców. Ta metoda przez chwilę była bardzo popularna. „Sprawdź dłonie!” — radził internet, ponieważ generatory rzeczywiście miały z nimi spore problemy. Człowiek na obrazku mógł wyglądać jak model, fotografię oświetlało zachodzące słońce, a ręka przypominała projekt studenta biologii, który nigdy wcześniej nie widział ręki. Było łatwo.

Technologia dostała jednak tę notatkę.

Podobnie jest z wieloma innymi „niezawodnymi oznakami AI”. Dziwne zęby, rozmazane napisy, asymetryczne okulary, podejrzanе uszy, światło zachowujące się jak po dwóch drinkach. Wszystkie mogą być sygnałem. Żaden nie jest magicznym wykrywaczem prawdy. Im lepsze stają się narzędzia, tym mniej sensowne jest budowanie całego bezpieczeństwa na zasadzie: „Jeżeli palec wygląda dziwnie, to robot”.

To trochę jak zabezpieczanie mieszkania przed włamywaczem poprzez sprawdzanie, czy ma pasiastą koszulkę i worek z napisem „ŁUP”.

Potrzebujemy czegoś lepszego.

Najważniejsza zmiana nie dotyczy oczu. Dotyczy sposobu myślenia. W świecie łatwych manipulacji nie wystarczy już pytać: „Czy to wygląda prawdziwie?”. Trzeba częściej pytać: „Skąd to pochodzi?”, „Kto opublikował materiał jako pierwszy?”, „Czy istnieją inne niezależne źródła?”, „Czy pokazano mi pełny kontekst?”, „Czy ktoś próbuje wywołać we mnie tak silną reakcję, żebym udostępnił to szybciej, niż pomyślę?”.

Ten ostatni punkt jest szczególnie interesujący, ponieważ fałszywe treści nie muszą pokonać twojej inteligencji.

Wystarczy, że pokonają twój refleks.

Widzisz coś oburzającego. Klik. Wysyłasz. Widzisz coś przerażającego. Klik. Rodzinna grupa na WhatsAppie właśnie dostała materiał dowodowy. Widzisz nagranie potwierdzające dokładnie to, co od dawna podejrzewałeś. Klik. Witaj, dopamino. Dziękujemy za współpracę.

Najbardziej skuteczna dezinformacja rzadko podchodzi do człowieka i mówi: „Dzień dobry, mam kompletnie absurdalną historię, w którą proszę uwierzyć”. Znacznie częściej mówi: „Wiedziałeś, że coś tu nie gra”. Potwierdza wcześniejsze przekonanie, wzmacnia strach, podkreśla złość albo oferuje sensację tak atrakcyjną, że rozsądek przez chwilę siedzi w poczekalni i przegląda stare gazety.

Dlatego będziemy zajmować się nie tylko samymi deepfake'ami. Przyjrzymy się również wygenerowanym zdjęciom, fałszywym nagraniom głosowym, przerobionym filmom, wyrwanym z kontekstu prawdziwym materiałom, fikcyjnym zrzutom ekranu, podszywaniu się pod ludzi i coraz bardziej przekonującym oszustwom. Bo czasami największym problemem nie będzie obraz stworzony od zera. Będzie nim prawdziwy obraz podpisany fałszywym opisem.

Zdjęcie może być autentyczne.

Tylko że z innego kraju.

Z innego roku.

Z innego wydarzenia.

A internet właśnie wsadził mu nowe życie jak kurtkę po starszym bracie.

Nie chodzi jednak o to, żebyś po przeczytaniu tej książki zamienił się w cyfrowego detektywa, który przed obejrzeniem zdjęcia śniadania znajomego sprawdza metadane, pogodę satelitarną i kierunek padania cienia na maselniczkę. Celem nie jest paranoja. Paranoja jest równie mało użyteczna jak naiwność, tylko zużywa znacznie więcej energii.

Chodzi o wypracowanie prostego systemu.

Nie każda treść wymaga śledztwa. Zdjęcie kota w kartonie może prawdopodobnie przejść bez międzynarodowej weryfikacji. Jeśli jednak materiał ma wpływać na twoje pieniądze, bezpieczeństwo, decyzje, opinię o człowieku albo zachęcić cię do natychmiastowego działania, warto podnieść poziom ostrożności.

Im większe konsekwencje, tym wyższy standard dowodu.

To jedna z najważniejszych zasad tej książki.

Jeżeli ktoś przesyła ci zabawne zdjęcie papieża na deskorolce, możesz wzruszyć ramionami. Jeżeli ktoś przesyła nagranie twojego szefa proszącego o pilny przelew, sytuacja wymaga czegoś więcej niż wzruszenia ramionami i odpowiedzi „OK”. Zwłaszcza jeżeli szef zaczyna rozmowę od „nie mogę teraz rozmawiać telefonicznie”.

W takim momencie właśnie rozmowa telefoniczna staje się wyjątkowo interesującym pomysłem.

Nauczysz się więc przede wszystkim zwalniać dokładnie tam, gdzie oszust, manipulator albo zwykły internetowy chaos chce, żebyś przyspieszył. Zobaczysz, które sygnały powinny podnosić czujność, jak sprawdzać pochodzenie materiałów, kiedy warto szukać drugiego źródła, dlaczego zrzut ekranu jest wyjątkowo słabym świadkiem oraz jak reagować, gdy fałszywa treść dotyczy ciebie lub kogoś bliskiego.

Będziemy też rozmawiać o tym, czego nie robić.

Nie staniesz się bezpieczniejszy tylko dlatego, że zainstalujesz siedem „detektorów AI” i będziesz wrzucał do nich każde zdjęcie. Narzędzia mogą pomagać, ale same również potrafią się mylić. Nie wystarczy też wejść w komentarze, ponieważ tłum internetowych specjalistów potrafi z pełnym przekonaniem ustalić, że materiał jest jednocześnie prawdziwy, fałszywy, stary, nowy i prawdopodobnie wykonany na zlecenie CIA.

Komentarze są niezwykłym miejscem.

Można wejść po odpowiedź i wyjść z sześcioma nowymi pytaniami oraz człowiekiem o pseudonimie „Prawda_77”, który właśnie wyjaśnił całą geopolitykę przy pomocy trzech emotikonów.

Nie będziemy także udawać, że istnieje metoda zapewniająca stuprocentową pewność. Czasami po prostu nie będziesz wiedzieć. I to też jest odpowiedź. „Nie wiem, czy to prawdziwe” jest znacznie dojrzałym stanowiskiem niż „na pewno prawdziwe” albo „na pewno AI”, jeśli dowody są słabe.

Internet bardzo lubi pewność.

Rzeczywistość znacznie częściej korzysta z wersji roboczej.

Najważniejsza umiejętność najbliższych lat może więc okazać się zaskakująco mało technologiczna. Nie będzie nią rozpoznawanie każdej manipulacji w sekundę. Będzie nią zdolność powiedzenia sobie: „Nie wiem jeszcze. Sprawdzę, zanim uwierzę albo podam dalej”.

Kilka sekund przerwy.

Jedno dodatkowe źródło.

Czasem jeden telefon.

Nie brzmi spektakularnie. Nie wymaga okularów rozszerzonej rzeczywistości ani czerwonego pokoju z sześcioma monitorami.

Ale działa znacznie lepiej niż wpatrywanie się przez dziesięć minut w ucho człowieka na zdjęciu i zastanawianie się, czy wygląda wystarczająco biologicznie.

Bo w epoce, w której można podrobić niemal wszystko, nie musisz być człowiekiem, którego nie da się oszukać.

Wystarczy, że coraz trudniej będzie cię oszukać szybko.

I od tego zaczniemy.

Prospekt i zasady serii zastosowane zgodnie z przesłanym materiałem.

Rozdział 1 - Oczy już nie wystarczą

Widzisz zdjęcie. Na zdjęciu płonie budynek. Ludzie stoją na ulicy, ktoś biegnie, ktoś nagrywa telefonem, w tle jest samochód strażacki. Obraz wygląda tak przekonująco, że twój mózg wykonuje swoją standardową procedurę: przyjmuje materiał do wiadomości, dopisuje emocję i przechodzi dalej.

Problem w tym, że budynek nie istnieje.

Strażacy nie istnieją.

Ulica też nie.

A człowiek, który właśnie biegnie w stronę płomieni, ma prawdopodobnie równie bogatą historię życia jak tost.

Jeszcze niedawno fotografia miała szczególną pozycję. Mogła być zmanipulowana, oczywiście. Retusz istnieje od bardzo dawna, propaganda również, a Photoshop nie został wynaleziony wczoraj po południu. Mimo to stworzenie naprawdę przekonującego fałszywego obrazu zwykle wymagało czasu, umiejętności i materiałów źródłowych. Dzisiaj bariera wejścia spadła dramatycznie. Można opisać scenę, nacisnąć przycisk i po chwili dostać kilka wersji wydarzenia, które nigdy się nie wydarzyło.

To zmienia bardzo ważną rzecz.

Realizm przestał być dowodem.

Jeżeli zdjęcie wygląda źle, może być fałszywe. Jeżeli wygląda świetnie, również może być fałszywe. Jeżeli wygląda przeciętnie, ziarnisto, jakby ktoś zrobił je starszym telefonem podczas deszczu i jeszcze przypadkiem zasłonił pół obiektywu palcem — nadal może być wygenerowane. Generatory nauczyły się bowiem nie tylko tworzyć idealne