WHEN THE ALGORITHM GETS IT WRONG

एल्गोरिदम की चूक

AIAAIC अभिलेख से देखे गए एआई के दुष्प्रभाव और नैतिक प्रश्न

भर्ती और कल्याण, अदालत और सड़क, कक्षा और कार्यस्थल पर जो हुआ

किम क्युंग-जिन, अधिवक्ता

विषय-सूची

अध्याय 1 पूर्वाग्रह और भेदभाव: कोड में लिखा गया मानवीय पक्षपात

1.1 नियुक्ति, वित्त और सेवा के एल्गोरिदम में पूर्वाग्रह

1.2 कल्याणकारी लाभ और सामाजिक सुरक्षा जाल के एल्गोरिदम में खराबी

1.3 चेहरे की पहचान तकनीक द्वारा रंगीन लोगों की गलत पहचान और अनुचित गिरफ्तारी

अध्याय 2 मतिभ्रम और गलत सूचना: जनरेटिव एआई का विश्वसनीयता संकट

2.1 जेनरेटिव एआई का मतिभ्रम और नकली किताबों और लेखों का प्रसार

2.2 अदालत और शैक्षणिक क्षेत्र में फर्जी फैसलों और उद्धरणों का विवाद

2.3 ऐतिहासिक और भौगोलिक तथ्यों का विकृतिकरण तथा सामाजिक विवाद

अध्याय 3 निजता का हनन और निगरानी समाज: बिना सहमति डेटा संग्रह की छाया

3.1 अनाधिकृत जैविक जानकारी संग्रह और सार्वजनिक·वाणिज्यिक निगरानी

3.2 व्यक्तिगत जानकारी का अनधिकृत प्रशिक्षण और AI बातचीत के लीक होने की घटना

3.3 स्थान ट्रैकिंग और श्रमिकों की अत्यधिक निगरानी

अध्याय 4 कॉपीराइट और रचनाकारों के अधिकार: जनरेटिव एआई और प्रशिक्षण डेटा का विवाद

4.1 कॉपीराइट उल्लंघन मुकदमे और अनधिकृत डेटा स्क्रैपिंग

4.2 AI निर्मित सामग्री के कॉपीराइट की अस्वीकृति और कला पारिस्थितिकी तंत्र में उथल-पुथल

4.3 छवि अधिकार, वॉयस क्लोनिंग और अनधिकृत नकल

अध्याय 5 स्वचालित वाहन और रोबोट: भौतिक सुरक्षा और जनहानि

5.1 स्वायत्त वाहन की खामियाँ और सड़क दुर्घटनाएँ

5.2 औद्योगिक और सेवा रोबोट की खराबी और सुरक्षा दुर्घटनाएं

5.3 सैन्य स्वायत्त हथियारों और ड्रोन तकनीक के घातक खतरे

अध्याय 6 भावनात्मक क्षति और डिजिटल अपराध: एआई चैटबॉट और दुर्भावनापूर्ण उपयोग

6.1 AI चैटबॉट के साथ अत्यधिक भावनात्मक जुड़ाव और अपराध के लिए उकसाना

6.2 डीपफेक यौन अपराध और डिजिटल संश्लेषित धोखाधड़ी

6.3 सर्च एल्गोरिदम में हेरफेर और डायनेमिक प्राइसिंग का एकाधिकार

अध्याय 1

पूर्वाग्रह और भेदभाव: कोड में लिखा गया मानवीय पक्षपात

1.1

नियुक्ति, वित्त और सेवा के एल्गोरिदम में पूर्वाग्रह

अस्वीकृति के ईमेल हमेशा भोर में आते थे।

Derek Mobley एक चालीस पार के अश्वेत व्यक्ति थे, और एंग्जायटी डिसऑर्डर तथा डिप्रेशन से पीड़ित थे। उन्होंने सूचना प्रौद्योगिकी के क्षेत्र में काम किया था, उनके पास प्रमाणपत्र थे, और अनुभव भी था। उन्होंने Workday नामक कंपनी के भर्ती सिस्टम का उपयोग करने वाली सौ से अधिक कंपनियों में आवेदन किया।

उन्हें सौ से अधिक बार अस्वीकार कर दिया गया।

अस्वीकृति के ईमेल आने का समय देखकर यह पता लगाया जा सकता था कि वे किसी इंसान द्वारा नहीं भेजे गए थे। रात के एक बजे, रात के तीन बजे। इंसान भर्तीकर्ता उस समय सोते हैं। मशीनें सोती नहीं हैं। वे कॉफी भी नहीं पीती हैं, और न ही उन्हें खेद होता है। वे दस्तावेज़ खोलती हैं और तीन सेकंड में बंद कर देती हैं।

Mobley ने मुकदमा दायर किया। उनका दावा था कि Workday का आर्टिफिशियल इंटेलिजेंस उम्र, नस्ल और विकलांगता को छांटने वाली छलनी की तरह काम करता है। अदालत ने पाया कि Workday ने अपने मौजूदा कर्मचारियों के डेटा के साथ आर्टिफिशियल इंटेलिजेंस को प्रशिक्षित किया है और संरक्षित विशेषताओं को एक प्रतिनिधि संकेतक के रूप में इस्तेमाल किया है, और राष्ट्रीय स्तर के क्लास-एक्शन मुकदमे को मंजूरी दे दी।

मार्च 2026 में, अदालत ने Workday के इस तर्क को खारिज कर दिया कि आयु भेदभाव अधिनियम आवेदकों पर लागू नहीं होता है। 16 जून को यह फैसला सुनाया गया कि यदि कैलिफोर्निया के बाहर के नियोक्ता भी इसका इस्तेमाल करते हैं तो भी Workday को जिम्मेदार ठहराया जा सकता है। इस मुकदमे में शामिल अस्वीकृतियों की संख्या 1.1 अरब है।

1.1 अरब भोर के ईमेल।

यह कहानी इसलिए डरावनी है क्योंकि इसमें कोई खलनायक नहीं है। किसी ने भी सम्मेलन कक्ष में इकट्ठा होकर चालीस पार के लोगों को अस्वीकार करने का संकल्प नहीं लिया। उन्होंने बस पिछले दस वर्षों में काम पर रखे गए लोगों का डेटा मशीन को खिला दिया। मशीन ने ईमानदारी से उस पैटर्न को सीखा। और उसने ईमानदारी से उसे दोहराया।

मशीन लर्निंग शब्द सुनने में कठिन लगता है, लेकिन इसका सिद्धांत प्राथमिक विद्यालय की गणित की कक्षा के समान है। नियम बताए बिना केवल बहुत सारे उत्तर दिखाए जाते हैं, और कहा जाता है कि इनमें से नियम खोजें।

यदि अतीत के उत्तर पक्षपाती हैं, तो मशीन पक्षपात को ही नियम के रूप में सीखती है। मशीन के दृष्टिकोण से, उसने अच्छा काम किया है। क्योंकि उसने वही किया जो उसे करने के लिए कहा गया था।

दस साल पहले Amazon ने यह तथ्य सबसे पहले सीखा था।

2014 में स्कॉटलैंड के एडिनबर्ग में Amazon के विकास केंद्र में एक गुप्त परियोजना शुरू हुई थी। यह एक ऐसा सिस्टम था जो रिज्यूमे डालने पर स्टार रेटिंग देता था। एक स्टार से लेकर पांच तक। यह वस्तुओं को दी जाने वाली स्टार रेटिंग को इंसानों को देने जैसा था।

डेवलपर्स ने पिछले दस वर्षों के तकनीकी नौकरी के आवेदनों को प्रशिक्षण डेटा के रूप में फीड किया। उस दस साल की अवधि में Amazon द्वारा काम पर रखे गए तकनीकी कर्मचारी ज़्यादातर पुरुष थे। मशीन ने निष्कर्ष निकाला: पुरुष बेहतर आवेदक होते हैं।

मशीन ने रिज्यूमे में 'महिला' शब्द खोजने पर अंक काट लिए। जिन रिज्यूमे में 'महिला शतरंज क्लब की अध्यक्ष' लिखा था, उनके अंक काट लिए गए। महिला विश्वविद्यालयों से स्नातक होने वाले आवेदकों के अंक भी पूरी तरह से कम कर दिए गए। इसके विपरीत, पुरुष इंजीनियरों द्वारा अक्सर उपयोग किए जाने वाले क्रिया शब्दों, जैसे 'निष्पादित किया' या 'पकड़ लिया', के होने पर अंक बढ़ा दिए गए।

Amazon ने इस पक्षपात को सुधारने की कोशिश की। वह विफल रहा। 2017 में परियोजना को रद्द कर दिया गया। Amazon ने कहा कि सौभाग्य से उस सिस्टम का उपयोग वास्तविक भर्ती में नहीं किया गया था।

समस्या यह है कि अन्य कंपनियों ने ऐसी ही चीजें बनाना जारी रखा।

Kyle Behm वेंडरबिल्ट विश्वविद्यालय के छात्र थे। उन्होंने बाइपोलर डिसऑर्डर के इलाज के लिए कुछ समय के लिए स्कूल से छुट्टी ली थी, और अपने जीवन यापन का खर्च उठाने के लिए पड़ोस की दुकानों में आवेदन किया था। Kroger सुपरमार्केट, Home Depot, Walgreens, PetSmart। यह पार्ट-टाइम कैशियर का काम था।

उन्हें हर जगह से अस्वीकार कर दिया गया। वह साक्षात्कार तक भी नहीं पहुंच पाए।

ऐसा इसलिए हुआ क्योंकि Kronos नामक कंपनी द्वारा बनाए गए Unicru नामक एक ऑनलाइन व्यक्तित्व परीक्षण ने Kyle के लिए लाल बत्ती जला दी थी। उस परीक्षण ने Kyle के मानसिक स्वास्थ्य के इतिहास का पता लगाया, और यह निर्णय लिया कि ऐसे व्यक्ति द्वारा ग्राहक के क्रोधित होने पर उसे अनदेखा करने की अधिक संभावना होती है।

Kyle के पिता, Roland Behm, एक वकील थे। उन्होंने 2014 में समान रोजगार अवसर आयोग में शिकायत दर्ज कराई, और कंपनियों के खिलाफ मानसिक रूप से विकलांग लोगों के साथ भेदभाव करने का मुकदमा दायर

किया।

Kyle ने 2019 में खुद अपना जीवन समाप्त कर लिया। यह मुकदमा समाप्त होने से पहले हुआ था।

यह तथ्य कि एक ऐसा युवक था जो व्यक्तित्व परीक्षण पास नहीं कर पाने के कारण कैशियर के काउंटर पर नहीं खड़ा हो सका। यह तथ्य कि वह निर्णय किसी इंसान ने नहीं बल्कि एक वेबपेज ने लिया था। और यह तथ्य कि उस वेबपेज पर शिकायत करने का कोई विकल्प नहीं था। जब आप इन तीन वाक्यों को एक साथ पढ़ते हैं, तो आपकी रीढ़ में सिहरन दौड़ जाती है।

iTutorGroup का सिस्टम और भी अधिक स्पष्ट था। इस ऑनलाइन शिक्षा कंपनी के स्वचालित भर्ती कार्यक्रम ने पचपन वर्ष से अधिक उम्र की महिलाओं और साठ वर्ष से अधिक उम्र के पुरुषों को स्वचालित रूप से बाहर कर दिया। योग्य होने के बावजूद दो सौ से अधिक शिक्षकों को उनकी उम्र के कारण नौकरी से निकाल दिया गया।

जिस तरह से यह पकड़ा गया, वह हास्यास्पद है। एक आवेदक ने अपनी वास्तविक जन्मतिथि के साथ आवेदन किया और उसे अस्वीकार कर दिया गया। लेकिन जब उसने केवल अपनी उम्र कम करके दोबारा आवेदन किया, तो उसे तुरंत साक्षात्कार के लिए बुलाया गया। यदि कोई इंसान होता तो उसे शर्मिंदगी होती, लेकिन सिस्टम शांत था। समान रोजगार अवसर आयोग ने 365,000 डॉलर का समझौता शुल्क लगाया।

ऐसा लगा था कि जनरेटिव आर्टिफिशियल इंटेलिजेंस के आने पर स्थिति बेहतर हो जाएगी।

2024 में, Bloomberg के शोधकर्ताओं ने GPT-3.5 को समान योग्यता वाले बहुत सारे फर्जी रिज्यूमे दिए और केवल नाम बदल दिए। ये ऐसे नाम थे जिनसे नस्ल और लिंग का अनुमान लगाया जा सकता था। परिणाम इस प्रकार थे: तकनीकी नौकरियों की स्क्रीनिंग में, अश्वेत महिला होने का अनुमान लगाने वाले नामों को 36 प्रतिशत कम चुना गया। महिला नामों को मानव संसाधन प्रबंधन की भूमिकाओं में धकेल दिया गया।

2021 में, जर्मन ब्रॉडकास्टर Bayerischer Rundfunk ने Retorio नामक स्टार्टअप के आर्टिफिशियल इंटेलिजेंस क्षमता मूल्यांकन का परीक्षण किया। जब एक ही व्यक्ति वही बातें कहता था, लेकिन चश्मा पहनता था, तो उसका स्कोर बदल जाता था। हिजाब पहनने पर यह फिर से बदल जाता था। अगर पीछे कोई बुकशेल्फ़ दिखाई देता था, तो निर्णय यह आता था कि व्यक्ति ईमानदार दिखता है।

यह हास्यास्पद है कि बुकशेल्फ़ रखने से आपका व्यक्तित्व अच्छा हो जाता है। लेकिन यह हँसी तब रुक जाती है जब यह एहसास होता है कि उस निर्णय के कारण किसी को नौकरी मिली और किसी को नहीं।

कक्षाओं में भी ऐसा ही हुआ। 2012 में, वाशिंगटन डी.सी. के पब्लिक स्कूलों में 5वीं कक्षा की शिक्षिका Sarah Wysocki को IMPACT नामक शिक्षक मूल्यांकन प्रणाली द्वारा नौकरी से निकाल दिया गया था। साथी शिक्षकों

और माता-पिता का उनके प्रति मूल्यांकन उत्कृष्ट था। हालाँकि, बच्चों के पिछले स्कूलों में परीक्षा में धोखाधड़ी हुई थी, और उन बढ़े हुए अंकों ने Wysocki के वैल्यू-एडेड अंकों को नीचे खींच लिया।

एक असत्यापित गणितीय सूत्र के कारण दो सौ पांच शिक्षकों ने अपनी नौकरी खो दी। वह सूत्र क्या था, इसका खुलासा कभी नहीं किया गया।

पैसे को संभालने वाले एल्गोरिदम ने और भी अधिक समय तक, और अधिक खामोशी से काम किया।

2018 में कैलिफोर्निया विश्वविद्यालय बर्कले के शोधकर्ताओं ने एक रिपोर्ट प्रकाशित की। 2008 से 2015 तक फैनी मेई और फ्रेडी मैक द्वारा गारंटीकृत 30 साल की निश्चित दर के ऋण का विश्लेषण किया गया। उस समय यह कहा जा रहा था कि यदि ऋण मंजूरी को लोगों से एल्गोरिदम में स्थानांतरित किया जाए तो भेदभाव समाप्त हो जाएगा।

लेकिन आंकड़ों ने कुछ और ही बताया। अश्वेत और लातिनी उधारकर्ताओं ने खरीद ऋण पर 5.6 से 8.6 आधार बिंदु अधिक और पुनर्वित्त पर 3 आधार बिंदु अधिक भुगतान किए। आधार बिंदु 0.01 प्रतिशत को दर्शाने वाली एक वित्तीय इकाई है। यह एक छोटी संख्या लगती है।

इस छोटी संख्या को पूरे अमेरिका में जोड़ने पर यह सालाना 250 मिलियन डॉलर से 500 मिलियन डॉलर हो जाती है।

कारण और भी कड़वा है। ऋण संस्थानों ने कृत्रिम बुद्धिमत्ता का उपयोग करके जाना कि अल्पसंख्यक जाति के उधारकर्ताओं के पास विभिन्न उत्पादों की तुलना करने का समय और जानकारी नहीं है। और उन्होंने उन लोगों को थोड़ी अधिक कीमत दी। इसे रणनीतिक मूल्य निर्धारण कहा जाता है। इस तरह से प्राप्त अतिरिक्त लाभ मार्जिन 11 प्रतिशत से 17 प्रतिशत है।

जांच पत्रकारिता माध्यम द मार्कअप और यूनाइटेड न्यूज़ ने संघीय आवास ऋण डेटा को एक साथ विश्लेषण किया। क्लासिक फिको स्कोर पर आधारित एक गुप्त अनुमोदन एल्गोरिदम ने रंगीन अल्पसंख्यकों के आवेदकों को सफेद लोगों की तुलना में 40 प्रतिशत से 80 प्रतिशत अधिक अस्वीकार कर दिया। कुछ बड़े शहरों में यह 250 प्रतिशत से अधिक था।

स्टैनफोर्ड बिजनेस स्कूल के लॉरा ब्लैटनर और स्कॉट नेल्सन यहां एक कदम आगे बढ़े। उनका कहना था कि भले ही एल्गोरिदम के पास कोई बुरा इरादा न हो, समस्या पैदा हो सकती है। निम्न आय वाले लोगों और अल्पसंख्यकों के पास क्रेडिट रिकॉर्ड पतला है। जब रिकॉर्ड पतला हो तो पूर्वानुमान सटीकता 5 से 10 प्रतिशत कम हो जाती है। लंबे समय पहले की एक या दो चूक पूरे स्कोर को बर्बाद कर देती है।

गरीब होने का मतलब कम डेटा है, कम डेटा का मतलब मशीन गलत अनुमान लगाती है, गलत अनुमान का मतलब जोखिम भरे व्यक्ति के रूप में वर्गीकृत होना है। और उस वर्गीकरण पर 'उद्देश्यपूर्ण क्रेडिट जोखिम' का लेबल लगाया जाता है।

वेल्स फार्गो बैंक में एक गणना त्रुटि ने लोगों को सड़कों पर ला दिया। 2010 से 2015 तक इस बैंक के स्वचालित बंधक संशोधन समीक्षा सॉफ्टवेयर में वकील की फीस की गणना में त्रुटि थी। 5 साल से अधिक समय तक किसी को पता नहीं था।

870 लोगों को अस्वीकृति सूचना दी गई जबकि उन्हें ऋण शर्तें बदलने के लिए योग्य होने के लिए पर्याप्त योग्यता थी। उनमें से कम से कम 545 लोगों के घर जब्त कर लिए गए। बैंक ने 2015 के अंत में त्रुटि को ठीक किया और कई वर्षों तक इस तथ्य को सार्वजनिक नहीं किया। बाद में निर्धारित मुआवजा 8 मिलियन डॉलर था।

जब एक घर की कीमत को देखते हैं, तो यह राशि गणना में फिट नहीं बैठती।

बीमा भी ऐसा ही था। जीको, सेफको और नेशनवाइड द्वारा उपयोग की जाने वाली जोखिम मूल्यांकन एल्गोरिदम ने, भले ही दुर्घटना जोखिम समान हो, रंगीन अल्पसंख्यकों की अधिक आबादी वाले इलाकों के निवासियों को 30 प्रतिशत तक अधिक बीमा प्रीमियम दिया। ऑलस्टेट ने एक कदम आगे बढ़ाया। यदि ग्राहक नहीं जाएंगे, तो दर बढ़ा दी गई। लंबे समय से व्यापार करने वाले वफादार ग्राहक अधिक भुगतान कर रहे थे। नियामक अधिकारियों द्वारा पकड़ा गया।

कार्यस्थल में प्रवेश करने के बाद भी एल्गोरिदम पीछा करता है।

2021 में इटली के बोलोग्ना श्रम अदालत ने फैसला सुनाया कि डिलीवरी प्लेटफॉर्म डिलीवेरू के फ्रैंक एल्गोरिदम भेदभावपूर्ण है। फ्रैंक ने डिलीवरी कर्मचारियों की विश्वसनीयता और भागीदारी को स्कोर दिया। यदि निर्धारित शिफ्ट पर नहीं दिखे तो स्कोर में कटौती होती है। कारण पूछा नहीं जाता था।

बीमार होने के कारण न आने पर भी कटौती होती है। कानून द्वारा गारंटीकृत हड़ताल में भाग लेने पर भी कटौती होती है। अगर स्कोर में कटौती होती है, तो अगली बार अच्छा शिफ्ट नहीं मिलता है। हड़ताल का अधिकार कानून की किताब में बना रहा, लेकिन कोड के अंदर हड़ताल करने पर भूख से मरने की संरचना थी।

चेकर नामक कंपनी की स्वचालित पहचान सत्यापन प्रणाली ने 2015 से उबर, लिफ्ट और पोस्टमेट्स के चालकों को बड़े पैमाने पर फ़िल्टर किया। एक ही नाम वाले किसी और के आपराधिक रिकॉर्ड, पहले से ही समय सीमा समाप्त हो चुके रिकॉर्ड वैसे ही चिपके रहे। जिन चालकों के खाते रातोंरात बंद हो गए, उन्हें पता नहीं था कि किससे शिकायत करें।

कैलिफोर्निया विश्वविद्यालय हेस्टिंग्स लॉ स्कूल की प्रोफेसर विना दुवल ने इसे नाम दिया। एल्गोरिदम वेतन भेदभाव। उबर, लिफ्ट और अमेज़ॅन कर्मचारियों के डेटा इकट्ठा करते हैं और एक ही काम के लिए विभिन्न मजदूरी देते हैं। चूंकि आप नहीं जानते कि बगल वाले को कितना मिलता है, आप तुलना भी नहीं कर सकते।

यात्री का पक्ष भी आरामदायक नहीं था। अनुसंधान के अनुसार, उबर और लिफ्ट के मूल्य निर्धारण और डिस्पैच एल्गोरिदम ने रंगीन अल्पसंख्यकों की अधिक आबादी वाले क्षेत्रों को उच्च किराया और लंबे प्रतीक्षा समय दिए।

फेसबुक विज्ञापन प्रणाली ने भर्ती विज्ञापनों को आयु से फ़िल्टर किया। वेरिज़ॉन, अमेज़ॅन और गोल्डमैन सैक्स ने आयु लक्ष्यीकरण विशेषता का उपयोग करके बड़ी उम्र के नौकरी चाहने वालों को विज्ञापन दिखाई न देने के लिए सेट किया। 2021 के नॉर्थईस्टर्न विश्वविद्यालय के अध्ययन और 2025 के फ्रांस मानवाधिकार रक्षक के निर्णय ने यहां एक कदम आगे बढ़ाया। भले ही विज्ञापनदाताओं ने लिंग निर्दिष्ट न किया हो, फेसबुक के अंदर की डिलीवरी एल्गोरिदम ने स्वयं पुरुषों और महिलाओं को विभिन्न भर्ती विज्ञापन दिखाए।

वस्तुओं की कीमत भी व्यक्ति को देखकर निर्धारित की गई।

एयरबीएनबी का स्मार्ट मूल्य निर्धारण एक ऐसी सुविधा है जो आवास की दरों को स्वचालित रूप से समायोजित करती है। कार्नेगी मेलॉन विश्वविद्यालय के प्रोफेसर परम वीर सिंह के अनुसंधान दल ने पाया कि सफेद होस्ट इस सुविधा को अधिक इस्तेमाल करते थे और परिणामस्वरूप सफेद और काले होस्ट की आय का अंतर बढ़ गया।

प्रिंसटन रिव्यू के ऑनलाइन स्कूल के पाठ्यक्रम शुल्क ज़िप कोड के अनुसार भिन्न थे। एशियाई-अमेरिकियों की अधिक आबादी वाले क्षेत्रों में अधिक कीमत निर्धारित की गई थी। मीडिया ने इसे 'टाइगर माँ टैक्स' कहा।

यात्रा साइट ओबिट्ज़ ने मैक कंप्यूटर उपयोगकर्ताओं को अधिक महंगे होटल दिखाए। यह इसलिए था क्योंकि मैक उपयोगकर्ताओं की औसत आय अधिक है। अगर लैपटॉप के ढक्कन पर एक सेब खींचा हुआ है, तो पर्स मोटा है, यह निर्णय किया गया।

अमेज़ॅन के पास प्रोजेक्ट नेसी नामक एक गुप्त एल्गोरिदम था। यह अपने उत्पादों की कीमत थोड़ी बढ़ाने और यह देखने का प्रयोग था कि प्रतिद्वंद्वी भी कीमत बढ़ाते हैं या नहीं। बाई बॉक्स एल्गोरिदम ने उपभोक्ताओं को सबसे सस्ते उत्पाद की बजाय पहले अमेज़ॅन को अधिक कमीशन देने वाले उत्पाद दिखाए।

इन सभी मामलों में एक समानता है। जब समस्या का पर्दाफाश हुआ तो कंपनियों का कथन लगभग समान था।

यह एल्गोरिदम के स्वयं द्वारा सीखे हुए परिणाम हैं। यह एक तकनीकी त्रुटि है। आंतरिक संरचना व्यावसायिक गोपनीयता है इसलिए इसे प्रकट नहीं किया जा सकता।

तीन वाक्य काफी थे। डेटा चुनने वाला व्यक्ति, लक्ष्य फ़ंक्शन निर्धारित करने वाला व्यक्ति, और इसके परिणामस्वरूप लाभ पाने वाला व्यक्ति सभी मीटिंग रूम में बैठे थे, लेकिन जिम्मेदारी कोड को दी गई। कोड कोई बहाना नहीं बनाता। कोई माफी नहीं माँगता। अदालत में नहीं जाता।

दुनिया में इससे अधिक सुविधाजनक कर्मचारी कोई नहीं है।

1.2

कल्याणकारी लाभ और सामाजिक सुरक्षा जाल के एल्गोरिदम में खराबी

यह पत्र सामान्य लग रहा था। यह ऑस्ट्रेलिया सरकार के लिफाफे में था, जिस पर सरकार का फॉन्ट और सरकार की मुहर थी। इसके अंदर यह लिखा था: आपको कल्याणकारी भत्ते का अधिक भुगतान मिला है। कृपया नीचे दी गई राशि वापस करें।

इसे पाने वाले लोग राशि देखकर दो बार हैरान हुए। कुछ पत्रों में एक करोड़ वॉन से अधिक की राशि लिखी थी। और वे तीसरी बार हैरान हुए। इसका कोई स्पष्टीकरण नहीं था कि उन्हें यह क्यों चुकाना चाहिए।

इस पत्र को पाने वाले लोगों की संख्या चार लाख से अधिक है।

यह काम उस ऑनलाइन अनुपालन हस्तक्षेप प्रणाली ने किया था, जिसे ऑस्ट्रेलिया के सामाजिक सेवा विभाग और सेंटरलिंक ने 2016 से 2019 तक चलाया था, और जिसे लोग रोबोडेब्ट कहते थे। रोबो का मतलब रोबोट है, और डेब्ट का मतलब कर्ज है। इसका अर्थ है रोबोट द्वारा बनाया गया कर्ज।

इस प्रणाली ने जो गलती की, वह प्राथमिक विद्यालय के गणित के स्तर की थी।

सरकार ने कर कार्यालय के वार्षिक आय डेटा का कल्याणकारी भत्ते के रिकॉर्ड के साथ मिलान किया। समस्या यह थी कि दोनों डेटा की समय इकाइयाँ अलग-अलग थीं। कर कार्यालय के पास पूरे एक साल का डेटा एक साथ था, जबकि कल्याणकारी भत्ता हर दो सप्ताह में एक बार दिया जाता था।

प्रणाली ने आसान रास्ता चुना। उसने एक साल की आय को बावन से विभाजित किया और मान लिया कि हर सप्ताह समान पैसा कमाया गया था।

अगर यह कोई पूर्णकालिक कर्मचारी होता, तो यह गणना सही होती। लेकिन कल्याणकारी भत्ता पाने वाले लोग आमतौर पर पूर्णकालिक कर्मचारी नहीं होते हैं। वे ऐसे लोग हैं जो स्ट्रॉबेरी तोड़ने के मौसम में तीन महीने काम करते हैं और बाकी समय आराम करते हैं, जो इस महीने दो काम करते हैं और अगले महीने एक भी नहीं, या जो हर बार अनुबंध समाप्त होने पर भत्ते के लिए फिर से आवेदन करते हैं।

अगर ऐसे व्यक्ति की एक साल की आय को बावन से विभाजित किया जाए तो क्या होगा? यह गणना की जाती है कि उन्होंने उन सप्ताहों में भी पैसा कमाया जब उन्होंने काम नहीं किया था। इसलिए यह निष्कर्ष निकलता है कि उन्होंने गलत तरीके से भत्ता प्राप्त किया।

जिन लोगों पर कोई कर्ज नहीं था, उन्हें कर्ज वसूली के नोटिस भेजे गए।

आपत्ति जताने के लिए, उन्हें कई साल पुरानी वेतन पर्ची लानी पड़ी। अगर उनके पास यह नहीं था, तो यह उनका अपना कर्ज बन गया। यह एक ऐसी व्यवस्था जहाँ सरकार को अपनी गणना साबित नहीं करनी थी, बल्कि नागरिकों को अपनी बेगुनाही साबित करनी थी। वसूली करने वाली कंपनियों को लगा दिया गया। टैक्स रिफंड को जब्त कर लिया गया।

कुछ लोग इसे सह नहीं पाए। ऐसे लोग थे जिन्होंने अपनी बेगुनाही साबित करने की कोई जगह न मिलने पर अपनी जान दे दी।

2023 में रॉयल कमीशन की सुनवाई में एक सच में डरावना तथ्य सामने आया। कई बार चेतावनियाँ दी गई थीं कि यह प्रणाली अवैध है और इसके गणना के तरीके में खामी है। कानूनी विशेषज्ञों ने चेतावनी दी थी, और आंतरिक कर्मचारियों ने भी चेतावनी दी थी।

वरिष्ठ सरकारी अधिकारी और मंत्री यह जानते थे। और उन्होंने इसे चलाना जारी रखा।

ऐसा इसलिए था क्योंकि उन्हें बजट बचाने की उपलब्धि की आवश्यकता थी। अगर कोई इंसान हर एक मामले की जाँच करता है, तो इसमें समय लगता है और पैसे खर्च होते हैं। उस जाँच प्रक्रिया को हटाना ही इस प्रणाली का मुख्य कार्य था। इसका अंत एक अरब अस्सी करोड़ ऑस्ट्रेलियाई डॉलर के समझौते और सरकार की आधिकारिक माफी के साथ हुआ।

माफी उन लोगों तक नहीं पहुँची जिनकी मृत्यु हो चुकी थी।

भारत में, उंगलियाँ एक समस्या थीं।

भारत सरकार आधार नामक एक बायोमेट्रिक प्रणाली के माध्यम से कल्याण का प्रबंधन करती है। भोजन का राशन पाने के लिए फिंगरप्रिंट देना आवश्यक है। झारखंड के एक सुदूर गाँव में, इस तरीके ने लोगों की जान ले ली।

फिंगरप्रिंट स्कैनर बार-बार विफल रहा। जिस व्यक्ति ने जीवन भर खेतों में काम किया है और शारीरिक श्रम किया है, उसकी उंगलियों के सिरे घिस जाते हैं। उंगलियों के निशान धुंधले हो जाते हैं। इसके अलावा, पहाड़ी गाँवों में इंटरनेट का कनेक्शन अच्छा नहीं होता है। सिग्नल का इंतजार करते-करते पूरा दिन बीत जाता है।

परहिया और अन्य हाशिए पर रहने वाले समुदायों के निवासियों को कानून द्वारा गारंटीकृत भोजन के राशन से वंचित कर दिया गया। कई गाँवों में ऐसे लोग थे जो भूख या कुपोषण से मर गए।

ऐसा नहीं है कि मशीन लोगों को नहीं पहचान पाई। जिस नियम में यह तय किया गया था कि अगर मशीन पहचान न पाए तो खाना न दिया जाए, उस नियम ने लोगों को मार डाला।

हरियाणा में, परिवार पहचान पत्र नामक एक प्रणाली ने हजारों जीवित लोगों को मृत घोषित कर दिया। अगर आपको मृत दर्ज किया जाता है, तो आपकी पेंशन रोक दी जाती है और भोजन का राशन काट दिया जाता है। बाद में जाँच करने पर पता चला कि रोकी गई 70 प्रतिशत पेंशन एल्गोरिदम के गलत निर्णय का परिणाम थी।

ऐसे बुजुर्ग लोग थे जिन्हें यह साबित करने के लिए सरकारी कार्यालयों में जाना पड़ा कि वे जीवित हैं।

तेलंगाना के समग्र वेदिका ने आय और रोजगार की स्थिति की गलत भविष्यवाणी की, और पंद्रह हजार घरों के राशन कार्ड चुपचाप रद्द कर दिए। लोगों के विद्रोह करने के बाद ही इसे वापस लिया गया।

जॉर्डन में, विश्व बैंक द्वारा समर्थित तकाफुल नामक एक गरीबी-मापन एल्गोरिदम चलाया गया था। इसने सहायता के लिए लक्ष्य निर्धारित करने के लिए घरेलू विशेषताओं को अंक दिए। हालाँकि, जो परिवार वास्तव में भूखे थे, उन्हें कम अंक मिलने के कारण बाहर कर दिया गया। जो लोग गरीबी को अंकों में मापना चाहते थे, वे गरीबी को पहचान नहीं पाए।

ब्राज़ील का मेउ आईएनएसएस (Meu INSS) ऐप पेंशन आवेदनों की प्रतीक्षा कतार को कम करने के लिए बनाया गया था। नतीजतन, कतार कम हो गई। ऐसा इसलिए हुआ क्योंकि आवेदन अपने आप खारिज हो गए। अगर फॉर्म में कोई टाइपिंग की गलती थी या किसी सवाल को गलत समझा गया था, तो वे तुरंत बाहर हो जाते थे। स्मार्टफोन से अपरिचित ग्रामीण बुजुर्ग इसमें फँस गए। फिर से आवेदन करने के लिए, उन्हें एक वकील की सेवा लेनी पड़ी।

अमीर देशों में भी स्थिति अलग नहीं थी।

यूनाइटेड किंगडम का यूनिवर्सल क्रेडिट एक ऐसी प्रणाली है जो कई कल्याणकारी भत्तों को एक में मिलाती है। इसमें इस्तेमाल किया गया गणना एल्गोरिदम एक कैलेंडर महीने के दौरान प्राप्त धन के आधार पर भत्ते का निर्धारण करता था। ह्यूमन राइट्स वॉच ने 2020 में इस संरचना की जाँच की।

कुछ कर्मचारियों को हर चार सप्ताह में एक बार वेतन मिलता है। ऐसे में, किसी महीने में वेतन दो बार आता है। प्रणाली ने यह तय किया कि यह व्यक्ति अचानक अमीर हो गया है। उस महीने के लिए उनके भत्ते में भारी कटौती की गई या वह बिल्कुल नहीं दिया गया।

कैलेंडर किस दिन से शुरू होता है, इस आधार पर उनके भोजन का पैसा गायब हो गया।

यूनाइटेड किंगडम के कार्य और पेंशन विभाग का धोखाधड़ी का पता लगाने वाला एल्गोरिदम और भी चुपचाप काम करता था। एडवांस नामक एक मॉडल और आवास लाभ निगरानी एल्गोरिदम ने प्राप्तकर्ता की उम्र, राष्ट्रीयता, विकलांगता की स्थिति और वैवाहिक स्थिति को देखकर जोखिम के अंक दिए।

उच्च अंक प्राप्त करने वाले लोगों की एक सूची सामने आई। ये बुल्गारिया और पोलैंड की राष्ट्रीयता वाली कम वेतन वाली अंशकालिक महिला कर्मचारी, और विकलांग लोग थे।

तीन वर्षों के दौरान, दो लाख लोगों को उच्च जोखिम वाले समूह के रूप में वर्गीकृत किया गया, और उनकी सहायता रोक दी गई या उनकी जाँच की गई। उनमें से दो-तिहाई से अधिक लोग ऐसे वैध प्राप्तकर्ता थे जिन्होंने कोई गलती नहीं की थी।

डेनमार्क की कल्याणकारी एजेंसी यूडीके (UDK) ने साठ से अधिक धोखाधड़ी का पता लगाने वाले एल्गोरिदम चलाए। उन्होंने निवास की स्थिति, राष्ट्रीयता, पारिवारिक संबंध, चिकित्सा रिकॉर्ड और यहाँ तक कि देश छोड़ने के इतिहास को भी एकत्र किया। राज्य ने इस बात की जानकारी एक जगह एकत्र की कि वे कहाँ गए थे, वे किसके साथ रहते हैं, और उन्हें कौन सी बीमारियाँ हैं, और फिर उन्हें अंक दिए।

अंतरराष्ट्रीय मानवाधिकार संगठनों ने इसे यूरोपीय संघ के कृत्रिम बुद्धिमत्ता कानून द्वारा प्रतिबंधित सामाजिक स्कोर प्रणाली माना। अंतरराष्ट्रीय एमनेस्टी ने "कोड से लिखा अन्याय" शीर्षक से एक रिपोर्ट जारी की। रिपोर्ट में कहा गया कि विकलांग लोग, कम आय वाले लोग, प्रवासी, शरणार्थी और अल्पसंख्यकों को जांच के लिए चिह्नित किए जाने का उच्च जोखिम था।

जब एमनेस्टी ने कोड और डेटा दिखाने के लिए कहा, तो डेनिश अधिकारियों ने इनकार कर दिया।

स्वीडन की सामाजिक बीमा एजेंसी 2013 से धोखाधड़ी की भविष्यवाणी करने वाले मॉडल को चला रही थी। जब लाइटहाउस रिपोर्ट ने इसकी जांच की, तो पाया कि महिलाएं, प्रवासी, विदेशी पृष्ठभूमि वाले लोग, कम आय वाले लोग और जो विश्वविद्यालय से स्नातक नहीं थे, वे अधिक बार पकड़े गए।

फ्रांस के सामाजिक सुरक्षा संगठन CNAF का जोखिम मूल्यांकन एल्गोरिथ्म ने जनसंख्या के लगभग आधे लोगों को स्कोर दिया। अकेले माता-पिता वाले परिवार, कम आय वाले लोग और विकलांग लोग स्वचालित रूप से संदिग्ध माने गए। उन्होंने यह नहीं बताया कि स्कोर ऐसे कैसे आए। इसके बजाय, वे घरों की तलाशी लेने आए।

नीदरलैंड में, कर प्राधिकरण ने 2013 से 2019 तक मशीन लर्निंग एल्गोरिथ्म का उपयोग करके बाल लाभ धोखाधड़ी को पकड़ने की कोशिश की। हजारों माता-पिता को धोखाधड़ी करने वाला माना गया। दोहरी नागरिकता को जोखिम संकेत के रूप में उपयोग किया गया। रॉटरडैम की प्रणाली ने लोगों को जातीयता, उम्र, लिंग और उनके पास बच्चे थे या नहीं इसके आधार पर विभाजित किया।

बच्चों की सुरक्षा के लिए बनाए गए एल्गोरिदम भी थे।

अमेरिका के पेंसिल्वेनिया राज्य में अलेघेनी काउंटी में, उन्होंने बाल शोषण के जोखिम की भविष्यवाणी करने के लिए एक पारिवारिक स्क्रीनिंग उपकरण का उपयोग किया। जब कोई रिपोर्ट आती है, तो उपकरण इसे स्कोर देता

है, और यदि स्कोर अधिक है, तो एक जांचकर्ता घर पर जाता है।

कार्नेगी मेलन विश्वविद्यालय के शोधकर्ताओं ने परिणामों का विश्लेषण किया। रिपोर्ट किए गए काली त्वचा वाले बच्चों का दो-तिहाई जांच के लिए सिफारिश किया गया। सफेद त्वचा वाले बच्चों के लिए यह आधा था।

कारण यह था कि उपकरण क्या देखता था। सूत्र में सरकारी सहायता प्राप्त करने का इतिहास शामिल था। इसमें मानसिक स्वास्थ्य निदान रिकॉर्ड भी शामिल थे।

अगर आप गरीब हैं और सहायता प्राप्त करते हैं, तो आपका जोखिम स्कोर बढ़ जाता है। अगर आप बीमार हैं और परामर्श लेते हैं, तो आपका जोखिम स्कोर बढ़ जाता है। यह एक संरचना है जहां मदद मांगने का रिकॉर्ड संदेह का आधार बन जाता है।

एक पैमाना जहां आप जितनी अधिक सहायता प्राप्त करते हैं, आपके बच्चे को छीने जाने की संभावना उतनी ही अधिक होती है। आप ऐसा पैमाना नहीं बना सकते और इसे न्यायपूर्ण कह सकते हैं।

ओरेगन की प्रणाली, इस उपकरण के आधार पर बनाई गई, 2022 में बंद कर दी गई जब नस्लीय पूर्वाग्रह की खोज की गई। इलिनॉय के Eckerd Rapid Safety Feedback को भी डेटा त्रुटियों और बढ़ी हुई कार्यक्षमता के दावों के कारण स्थगित किया गया।

डेनमार्क की बाल संरक्षण एजेंसी की निर्णय समर्थन प्रणाली में कोपेनहेगन IT विश्वविद्यालय द्वारा सत्यापन में अजीब पैटर्न पाए गए। समान स्थितियों के साथ, जब उम्र अलग थी तो जोखिम स्कोर बदल गया।

ऑस्ट्रेलियाई सरकार ने विकलांग व्यक्तियों के समर्थन के बजट में 700 मिलियन ऑस्ट्रेलियाई डॉलर बचाने के लिए RoboDebt नामक एक स्वचालित मूल्यांकन उपकरण पेश करने का प्रयास किया। विधि ने विकलांग लोगों के जीवन को कुछ मानक प्रकारों में संपीड़ित किया और फिर आवश्यक समर्थन की गणना की। नेत्रहीन वकालत समूहों और विकलांग संगठनों ने इसका कड़ा विरोध किया, और योजना को त्याग दिया गया।

अस्पतालों में, एल्गोरिदम ने डिस्चार्ज की तारीख निर्धारित की।

nH Predict, अमेरिका की सबसे बड़ी स्वास्थ्य बीमा कंपनियों UnitedHealth Group और Humana द्वारा उपयोग किया जाता है, बुजुर्ग और विकलांग रोगियों को पुनर्वास उपचार कितने दिन प्राप्त करना चाहिए इसकी भविष्यवाणी की। भले ही डॉक्टरों ने कहा कि अधिक आवश्यक था, लेकिन जब सिस्टम द्वारा निर्धारित दिन आ गया तो बीमा कवरेज बंद हो गया।

मुकदमे में, इस प्रणाली की दावे की अस्वीकृति की त्रुटि दर का खुलासा हुआ। यह 90 प्रतिशत था।

अगर वे दस बार इनकार करते हैं, तो नौ बार यह गलत है। लेकिन उन्होंने इसका उपयोग करना जारी रखने का कारण सरल है। अधिकांश अस्वीकृत रोगियों ने अपील नहीं की। बुजुर्ग और बीमार लोगों के लिए बीमा कंपनियों से लड़ना मुश्किल है।

नवंबर 2021 में, 86 वर्षीय Joan Barros गिरीं और उनका पैर घायल हो गया। चिकित्सा कर्मचारियों ने कहा कि शारीरिक चिकित्सा आवश्यक थी। जब एल्गोरिदम द्वारा निर्धारित डिस्चार्ज की तारीख आई, तो उपचार के लिए कवरेज बंद कर दिया गया।

UnitedHealth की Alert प्रणाली को कई राज्य सरकारों द्वारा प्रतिबंधित कर दिया गया जब इसने मानसिक स्वास्थ्य उपचार के अनुमोदन को अवैध रूप से प्रतिबंधित किया। Cigna का PxDx एक प्रणाली थी जिसने डॉक्टरों को रोगी चार्ट भी खोले बिना एक क्लिक के साथ दावों को बड़े पैमाने पर अस्वीकार करने की अनुमति दी। Evicore के DR एल्गोरिदम ने पूर्व-अनुमोदन अस्वीकृति दर को अधिकतम 20 प्रतिशत तक बढ़ा दिया।

Deloitte द्वारा बनाए गए टेक्सास के TIERS और टेनेसी के TEDS ने प्रोग्राम त्रुटियों और गलत पते के कारण सैकड़ों हजार कम आय वाले और विकलांग लोगों की Medicaid पात्रता को छीन लिया। नोटिस गलत घरों में गए, और जब कोई जवाब नहीं आया, तो उन्हें अपात्र के रूप में संसाधित किया गया।

अब तक उल्लेख किए गए देशों को गिनते हुए: Australia, India, Jordan, Brazil, United Kingdom, Denmark, Sweden, France, Netherlands, United States. प्रणालियां अलग हैं, और भाषाएं अलग हैं।

लेकिन पकड़े गए लोग आश्चर्यजनक रूप से समान हैं। वे गरीब, बुजुर्ग, बीमार, प्रवासी और कहीं बोलने की जगह नहीं हैं।

यह कोई संयोग नहीं है। ये प्रणालियां बजट बचाने के लिए पेश की गई थीं, बजट बचाने के लिए आप भुगतान को कम करना चाहिए, और भुगतान को कम करने के लिए आपको किसी को काटना चाहिए। काटने का सबसे सुरक्षित लक्ष्य वह व्यक्ति है जो विरोध नहीं कर सकता।

एल्गोरिदम उस गणना को बहुत अच्छी तरह से करते हैं।

1.3

चेहरे की पहचान तकनीक द्वारा रंगीन लोगों की गलत पहचान और अनुचित गिरफ्तारी

दो और पांच साल की उम्र की बेटियां सामने के आंगन में देख रही थीं।

9 जनवरी 2020 को, मिशिगन के Farmington Hills में एक घर के सामने के आंगन में, Robert Williams अभी काम से लौटे थे। Detroit पुलिस ने अचानक दरवाजा खोला और उनकी पत्नी और बच्चों के सामने उनकी कलाई पर हथकड़ी डाली।

आरोप घड़ी की चोरी का था। अक्टूबर 2018 में Detroit के Shinola स्टोर से 1800 डॉलर की पांच घड़ियां गायब हो गई थीं।

सबूत इस प्रकार था। दुकान के निगरानी कैमरे में एक धुंधली तस्वीर थी। Michigan राज्य पुलिस के विशेषज्ञ ने उस तस्वीर से एक दृश्य निकाला और इसे चेहरा पहचान डेटाबेस में डाला। सिस्टम ने Williams की पुरानी ड्राइविंग लाइसेंस की तस्वीर एक संभावना के रूप में सामने रखी।

अगली बात महत्वपूर्ण है। दुकान के सुरक्षा कर्मचारी, जो घटना स्थल पर नहीं थे, को काले पुरुषों की कई तस्वीरें दी गईं और उन्होंने Williams को चुना।

Williams को पुलिस की कोठरी में 30 घंटे रहना पड़ा।

पूछताछ के कमरे में, एक पुलिस अधिकारी ने निगरानी कैमरे की तस्वीर को मेज पर रख कर पूछा। क्या यह व्यक्ति आप हैं? Williams ने तस्वीर को अपने चेहरे के पास रख कर कहा। यह व्यक्ति मैं नहीं हूं। क्या आप सोचते हैं कि सभी काले पुरुष समान दिखते हैं?

इस मामले की एक पृष्ठभूमि है। उस समय, Detroit पुलिस कमिश्नर James Craig को पहले से पता था कि इस तकनीक में संदिग्ध को गलत तरीके से पहचानने की त्रुटि दर 96 प्रतिशत है।

सौ में से छियानवे बार गलत परिणाम देने वाले उपकरण को लेकर वे किसी के घर के आंगन में गए।

Robert Williams America में चेहरा पहचान की त्रुटि के कारण गलत तरीके से गिरफ्तार किए जाने का पहला आधिकारिक रिकॉर्ड बन गए। जून 2024 में, Detroit शहर ने American Civil Liberties Union के

साथ 300,000 डॉलर के समझौते पर स्वाक्षर किए, और नियमों को बदल दिया कि चेहरा पहचान के संकेत और तस्वीरों की तुलना के आधार पर अकेले गिरफ्तारी वारंट के लिए आवेदन नहीं किया जा सकता।

2026 के अनुसार, इसी तरीके से गिरफ्तार किए गए 12 से अधिक लोग ज्ञात हैं। ये केवल ज्ञात संख्या है।

10 जून 2026 को भी एक नया मुकदमा दायर किया गया। Florida के Robert Dylan को 2024 में गिरफ्तार किया गया था। धुंधली निगरानी तस्वीर और 93 प्रतिशत मेल का परिणाम इसका आधार था। गिरफ्तारी वारंट के आवेदन में वह साक्ष्य नहीं था जो यह साबित करे कि वह अपराधी नहीं हो सकते।

93 प्रतिशत संख्या लोगों को मोहित करती है। यह परीक्षा के अंक जैसी लगती है। लेकिन यह संख्या यह नहीं दर्शाती कि यह व्यक्ति अपराधी होने की 93 प्रतिशत संभावना है। यह दर्शाती है कि डेटाबेस में लाखों तस्वीरों में से, यह तस्वीर सबसे अधिक समान दिखती है।

सबसे अधिक समान दिखने वाला व्यक्ति हमेशा मौजूद होता है। चाहे अपराधी उसमें न हो।

यदि कोई इस अंतर को नहीं समझता है, तो क्या होता है, यह Angela Lipps की कहानी से पता चलता है।

14 जुलाई 2025 को, Tennessee। पांच बच्चों की मां और पांच पोतियों की दादी Lipps अपने घर में अपने पोतियों की देखभाल कर रही थीं। संघीय मार्शल बंदूकें लेकर अंदर आए। उन्होंने कहा कि वह North Dakota के Fargo में हुए बैंक धोखाधड़ी मामले से फरार हैं।

Lipps कभी North Dakota नहीं गईं। उन्होंने कभी हवाई जहाज में नहीं चढ़ा। वह घर से 160 किलोमीटर से भी दूर शायद ही कभी गईं।

Fargo पुलिस थाने के जांचकर्ताओं की जांच सब कुछ यही थी। उन्होंने चेहरा पहचान के परिणाम प्राप्त किए, सोशल मीडिया पर तस्वीरें ढूंढीं और उन्हें आंख से तुलना की। उन्होंने एक भी फोन कॉल नहीं किया।

Lipps को North Dakota स्थानांतरित किया गया और 108 दिनों तक कैद में रहीं। केवल जब उनके वकील ने बैंक लेनदेन के रिकॉर्ड प्रस्तुत किए जो यह साबित करते थे कि वह उस समय Tennessee में थीं, तब ही उन्हें रिहा किया गया।

इस दौरान उन्होंने जो घर रहते थे, वह खो गया। उनकी कार खो गई। उनके पालतू कुत्ते को भी खो दिया।

Detroit पुलिस की वही प्रणाली लोगों को निशाना बनाती रही।

जुलाई 2019 में, Michael Oliver को एक शिक्षक के मोबाइल फोन की चोरी करने का आरोप लगाकर गिरफ्तार किया गया। Dataworks Plus के एक प्रोग्राम ने पुलिस डेटाबेस में उनकी तस्वीर को अपराधी से जोड़

दिया।

फरवरी 2023 में, Porsha Woodruff को लूट और कार अपहरण के आरोप में गिरफ्तार किया गया। वह आठ महीने की गर्भवती थीं। पुलिस कोठरी के फर्श पर उन्हें प्रसव पीड़ा महसूस हुई।

जांचकर्ता ने निगरानी वीडियो को एक एल्गोरिथ्म में डाला और आठ साल पहले Woodruff के ड्राइविंग लाइसेंस की तस्वीर प्राप्त की। उन्होंने वांछित व्यक्ति की शारीरिक विशेषताओं की पुष्टि नहीं की। निगरानी कैमरे में कैद व्यक्ति गर्भवती नहीं थे।

किसी गर्भवती महिला और गैर-गर्भवती महिला के बीच अंतर करने के लिए कृत्रिम बुद्धिमत्ता की आवश्यकता नहीं है। आंखों की जरूरत है। लेकिन आंखों से देखने की प्रक्रिया गायब थी।

अप्रैल 2025 में New York में, Travis Williams को अश्लील प्रदर्शन के संदिग्ध के रूप में गिरफ्तार किया गया और दो दिन से अधिक समय तक कैद में रहे। उस समय, वह Brooklyn में थे, जो 19 किलोमीटर दूर था।

शारीरिक विशेषताओं की तुलना करने से हंसी आती है। वास्तविक अपराधी 167 सेंटीमीटर लंबा था और उसका वजन 72 किलोग्राम था। Williams की ऊंचाई 188 सेंटीमीटर है और वजन 104 किलोग्राम है।

पुलिस ने स्थान डेटा की पुष्टि करने या नियोक्ता की पुष्टि करने से इंकार कर दिया।

नवंबर 2022 में, Georgia के निवासी Randal Quran Reed को Louisiana में एक विलासवान बैग चोरी करने वाले अपराधी के रूप में बदनाम किया गया और छह दिनों के लिए कैद में रहे। Clearview AI के एक मेल के परिणाम के आधार पर गिरफ्तारी वारंट जारी किया गया। Reed ने कहा कि वह उस समय Georgia में थे, लेकिन किसी ने पुष्टि नहीं की। बाद में मामला चुप-चाप खारिज कर दिया गया।

ऐसे चेहरे ही क्यों गलत हो जाते हैं? उत्तर पहले से ही 2018 में मौजूद था।

MIT और Stanford के Joy Buolamwini और Timnit Gebru ने Gender Shades नाम की एक रिपोर्ट जारी की। यह IBM, Face++, और Microsoft के चेहरा विश्लेषण प्रोग्राम का परीक्षण करने का परिणाम था।

हल्की त्वचा वाले पुरुषों का लिंग पहचानने में त्रुटि दर 0.8 प्रतिशत से कम थी। गहरी त्वचा वाली महिलाओं में त्रुटि 34.7 प्रतिशत तक थी।

यह लगभग एक तिहाई है।

कारण प्रशिक्षण डेटा में है। ImageNet, LFW, BDD100K जैसे प्रसिद्ध डेटासेट में सफेद पुरुषों की तस्वीरें अत्यधिक मात्रा में थीं। मशीनें उन चेहरों को अच्छी तरह से पहचानती हैं जिन्हें वह बहुत बार देखते हैं। जिन चेहरों

को वह कम बार देखते हैं, वह उन्हें एक साथ समूहित कर देती हैं।

जब Georgia Institute of Technology ने स्वायत्त वाहन के लिए BDD100K डेटासेट का विश्लेषण किया, तो वही पैटर्न सामने आया। गहरी त्वचा वाले पैदल चलने वालों को पहचानने की सटीकता औसतन 4.8 प्रतिशत, और अधिकतम 12 प्रतिशत कम थी।

यह एक ऐसी कहानी है जहां कार लोगों को नहीं पहचान पाती। यह बिंदु बाद में फिर से सामने आएगा।

चेतावनियां पर्याप्त थीं। बिक्री जारी रही।

क्लियरव्यू एआई ने फेसबुक, इंस्टाग्राम, ट्विटर और लिंक्डइन से लोगों के चेहरों की तीन अरब तस्वीरें बिना सहमति के इकट्ठा कीं और एक सर्च डेटाबेस बनाया। प्लेटफॉर्म कंपनियों ने इसे रोकने की मांग की और इलिनॉय बायोमेट्रिक इन्फॉर्मेशन प्राइवेसी एक्ट के उल्लंघन का मुकदमा दायर किया गया, लेकिन कंपनी ने दुनिया भर की जांच एजेंसियों को अपनी सर्च सर्विस बेची।

हमें नहीं पता कि दस साल पहले हमारी पोस्ट की गई ग्रेजुएशन की तस्वीर आज किस पुलिस स्टेशन के कंप्यूटर में होगी।

अमेज़ॅन ने भी रिकग्निशन नामक एक उत्पाद पुलिस और इमिग्रेशन एंड कस्टम्स एनफोर्समेंट को बेचने की कोशिश की। जुलाई 2018 में जब अमेरिकन सिविल लिबर्टीज यूनियन ने इसका परीक्षण किया, तो अट्ठाईस संघीय प्रतिनिधियों और सीनेटरों के चेहरे अपराधियों की तस्वीरों से मेल खाते हुए पाए गए। गलत पहचाने गए सांसदों में गैर-श्वेत लोगों का अनुपात विशेष रूप से अधिक था।

अक्टूबर 2019 में, न्यू इंग्लैंड क्षेत्र के एक सौ अट्ठासी पेशेवर खेल खिलाड़ियों में से सत्ताईस लोगों को अपराधी बताया गया।

अमेज़ॅन ने कहा कि उसने विश्वसनीयता का स्तर 95 प्रतिशत या उससे ऊपर रखकर इसका उपयोग करने का निर्देश दिया था। लेकिन असल में उस मानक का पालन नहीं किया गया। निर्देशिका पढ़ने वाले लोग हमेशा बहुत कम होते हैं।

बढ़ा-चढ़ाकर किए गए विज्ञापन भी थे। फेडरल ट्रेड कमीशन ने इंटेलीविज़न नामक कंपनी के इस विज्ञापन पर आपत्ति जताई जिसमें दावा किया गया था कि उनके फेशियल रिकग्निशन सॉफ्टवेयर में लिंग या नस्ल का कोई भेदभाव नहीं है और यह दुनिया में सबसे सटीक है। उन्होंने कहा था कि इसे लाखों तस्वीरों के साथ प्रशिक्षित किया गया है, लेकिन असल में यह एक लाख तस्वीरों का बदला हुआ डेटासेट था।

खुदरा विक्रेता राइट एड ने 2012 से 2020 तक उन इलाकों के स्टोरों में फेशियल रिकग्निशन कैमरे सबसे ज्यादा लगाए जहां कम आय वाले और गैर-श्वेत लोग ज्यादा रहते थे। हजारों ग्राहकों को चोर समझकर स्टोर से बाहर निकाल दिया गया या उनकी तलाशी ली गई। फेडरल ट्रेड कमीशन ने उन्हें पांच साल तक इस तकनीक का उपयोग करने से रोक दिया।

यहाँ तक की बातें तो फिर भी उन लोगों की कहानियाँ हैं जो जीवित वापस आ गए।

जनवरी 2022 में टेक्सास के ह्यूस्टन में सनग्लास हट डकैती के मामले में, मेसीज़ और एसिलोरलक्सोटिका के फेशियल रिकग्निशन सिस्टम ने हार्वे मर्फी जूनियर को अपराधी बताया। उस समय वह कैलिफोर्निया के सैक्रामेंटो में थे।

अक्टूबर 2023 में गिरफ्तार होने के बाद दो हफ्ते तक जेल में रहने के दौरान, मर्फी के साथ तीन कैदियों ने सामूहिक मारपीट और यौन उत्पीड़न किया। उन्हें हमेशा के लिए शारीरिक विकलांगता का शिकार होना पड़ा।

2021 में मिसौरी के पोंडडेल में रेलवे प्लेटफॉर्म पर मारपीट की घटना हुई थी। संदिग्ध ने मास्क और कपड़ों से अपना चेहरा ढका हुआ था। जासूस ने उस तस्वीर को एल्गोरिदम में डाल दिया।

ढका हुआ चेहरा डालने पर भी सिस्टम कुछ लोगों के नाम दे देता है। ऐसा इसलिए है क्योंकि इसमें 'मुझे नहीं पता' कहने की सुविधा नहीं है। इस तरह सामने आए नामों में बिना किसी आपराधिक रिकॉर्ड वाले चार बच्चों के पिता, उनतीस वर्षीय क्रिस्टोफर गैटलिन शामिल थे।

पुलिस के आंतरिक दिशानिर्देशों में भी लिखा था कि केवल फेशियल रिकग्निशन के परिणाम के आधार पर किसी को गिरफ्तार नहीं किया जाना चाहिए। फिर भी जासूस ने तस्वीरों की एक लाइनअप बनाई, और गैटलिन को दो साल से अधिक समय तक जेल में रहना पड़ा।

ओक्लाहोमा की किम्बर्ली विलियम्स को जून 2021 से छह महीने के लिए जेल में बंद रखा गया। निजी बैंक के एक अन्वेषक द्वारा दिए गए खोज परिणामों को जासूसों ने बिना जाँच किए गिरफ्तारी वारंट के आवेदन में लिख दिया। अदालत को यह तक नहीं बताया गया कि फेशियल रिकग्निशन का उपयोग किया गया था। काउंटी की तीन अलग-अलग जेलों में भेजे जाने के दौरान उनकी नौकरी चली गई।

न्यू जर्सी के फ्रांसिस्को आर्टेगा को नवंबर 2019 में सशस्त्र डकैती के आरोप में पकड़े जाने के बाद चार साल तक जेल में बंद रहकर मुकदमे का इंतज़ार करना पड़ा। ऐसा इसलिए हुआ क्योंकि जांच एजेंसी ने सिस्टम के काम करने के तरीके, त्रुटि दर और सोर्स कोड को सार्वजनिक करने से मना कर दिया था। जब आपको यही नहीं पता कि आपको किस आधार पर चुना गया है, तो इसका खंडन करने का कोई तरीका नहीं होता।

मैरीलैंड के बाल्टीमोर के चौवन वर्षीय अलोंजो सॉयर को एक बस में मारपीट के संदिग्ध के रूप में नौ दिन जेल में रखा गया। असली संदिग्ध से उनकी लंबाई अलग थी, उम्र अलग थी, सामने के दांतों के बीच का अंतर अलग था और चलने का तरीका अलग था।

देश की सीमाओं के बाहर भी ऐसी ही घटनाएँ हुईं।

जनवरी 2026 में इंग्लैंड के साउथेम्प्टन में बांग्लादेशी मूल के सॉफ्टवेयर इंजीनियर अल्बी चौधरी को उनके माता-पिता और पड़ोसियों के सामने गिरफ्तार कर लिया गया। उन्हें एक सौ पच्चासी किलोमीटर दूर मिल्टन कीन्स में एक खाली घर में चोरी के मामले में संदिग्ध बताया गया था। उन्हें दस घंटे तक हिरासत में रखा गया।

थेम्स वैली पुलिस ने जिस सिस्टम का उपयोग किया था, वह जर्मनी की कंपनी कॉग्निटेक का उत्पाद था, और उसमें 2020 में आया एक पुराना एल्गोरिदम लगा हुआ था। ब्रिटेन की नेशनल फिजिकल लेबोरेटरी के मूल्यांकन के अनुसार, कुछ विशेष सेटिंग्स में एशियाई लोगों को गलत पहचानने की दर 4.0 प्रतिशत थी। वहीं, श्वेत लोगों के लिए यह दर 0.04 प्रतिशत थी।

सौ गुना।

पुलिस अधिकारियों ने दावा किया कि उन्होंने अपनी आँखों से दोबारा जाँच की थी। इसे ऑटोमेशन बायस का नाम दिया गया है। जब मशीन कोई जवाब देती है, तो इंसान उस जवाब की जाँच करने के बजाय उसे सही साबित करने की कोशिश करता है। वे समानताएँ खोजना शुरू कर देते हैं। इंसानी चेहरों में समानताएँ ढूँढना कोई मुश्किल काम नहीं है।

मई 2024 में लंदन ब्रिज स्टेशन पर अड़तीस वर्षीय शॉन थॉम्पसन को पकड़ लिया गया। वह युवाओं के अधिकारों के लिए काम करने वाले एक कार्यकर्ता थे, जो चाकू से होने वाले अपराधों को रोकने के लिए अपनी स्वयंसेवी सेवा पूरी करके घर जा रहे थे। मेट्रोपॉलिटन पुलिस के रीयल-टाइम फेशियल रिकग्निशन सिस्टम ने उन्हें एक वांछित अपराधी बताया, और तीस मिनट तक उनसे पहचान पत्र माँगने और गिरफ्तारी की धमकियाँ दी गईं।

ब्रिटेन के होम ऑफिस का पासपोर्ट फोटो सत्यापन सिस्टम, गहरे रंग की त्वचा वाले आवेदकों के होठों को गलत तरीके से खुला हुआ मुँह समझ लेता था। जोरिस लेशेन और जोशुआ बाडा जैसे अश्वेत आवेदकों के पासपोर्ट आवेदन खारिज कर दिए गए। होम ऑफिस ने यह जानते हुए भी सिस्टम को लागू किया कि गैर-श्वेत लोगों के लिए इसकी विफलता दर बहुत अधिक है।

इज़रायली रक्षा बलों ने गाजा पट्टी और वेस्ट बैंक की चौकियों पर रेड वुल्फ, ब्लू वुल्फ और कोरसाइट नामक सिस्टम तैनात किए। उन्होंने बिना सहमति के फिलिस्तीनी निवासियों के चेहरे एकत्र करके एक स्वचालित निगरानी

नेटवर्क बनाया। गलतियों की वजह से निर्दोष निवासियों को आतंकवादी संदिग्ध बताकर पकड़ लिया गया, उनसे पूछताछ की गई और उन्हें पीटा गया।

भारत के हैदराबाद में फरवरी 2023 में पैंतीस वर्षीय दिहाड़ी मजदूर मोहम्मद कादिर को हिरासत में ले लिया गया। इसका कारण यह था कि उनका चेहरा एक जेबकतरे की घटना के वीडियो में दिख रहे व्यक्ति के चेहरे से मिलता-जुलता था। उस वीडियो में व्यक्ति ने तौलिये से अपना चेहरा ढका हुआ था।

कादिर को पाँच दिन तक बंद रखा गया और यातनाएँ दी गईं, जिसके कारण उनकी मौत हो गई।

ब्राज़ील में अप्रैल 2024 में फुटबॉल मैच देखने गए जोआओ एंटोनियो को हथकड़ी लगा दी गई क्योंकि उनका चेहरा एक वांछित अपराधी से मेल खाता था। अर्जेंटीना के ब्यूनस आयर्स में वांछित अपराधियों को पहचानने वाले सिस्टम को एक सौ चालीस से अधिक गलत गिरफ्तारियाँ करने के बाद बंद कर दिया गया।

इन घटनाओं को एक साथ रखकर देखने पर एक स्पष्ट क्रम दिखाई देता है।

तकनीकी कंपनियाँ सटीकता को बढ़ा-चढ़ाकर बताती हैं और इसे बेचती हैं। पुलिस उस परिणाम को पुख्ता सबूत की तरह इस्तेमाल करती है। वारंट आवेदन में यह बात छिपा ली जाती है कि फेशियल रिकग्निशन का इस्तेमाल किया गया था। न्यायाधीश बिना जाने उस पर हस्ताक्षर कर देते हैं। पकड़े गए व्यक्ति को पता ही नहीं होता कि उसे क्यों चुना गया है। अगर इसके काम करने का तरीका पूछा जाए, तो वे इसे व्यावसायिक रहस्य बता देते हैं।

प्रमाण का दायित्व उलट गया है। राज्य को दोषी साबित करने की जगह नागरिकों को निर्दोष साबित करना पड़ता है। इसके लिए उन्हें अलीबी होना चाहिए और वकील की फीस होनी चाहिए।

एंजेला लिप्स को 108 दिन लगे। क्रिस्टोफर गैटलिन को 2 साल लगे। फ्रांसिस्को आर्टेआगा को 4 साल लगे।

मोहम्मद कादिर के पास समय नहीं बचा।

अध्याय 2

मतिभ्रम और गलत सूचना: जनरेटिव एआई का विश्वसनीयता संकट

2.1

जेनरेटिव एआई का मतिभ्रम और नकली किताबों और लेखों का प्रसार

कोई भी लेखक कभी-कभी अपना नाम खोजता है। यह एक शर्मनाक आदत है, लेकिन सब ऐसा करते हैं।

अगस्त 2023 की एक सुबह, अमेरिकी लेखिका और पुस्तक समीक्षक Jane Friedman ने भी ऐसा ही किया। स्क्रीन पर उनकी किताबों की सूची दिखाई दी। लेकिन पांच किताबें अपरिचित थीं।

वे ऐसी किताबें थीं जिन्हें उन्होंने कभी पढ़ा नहीं था। वे ऐसी किताबें थीं जिन्हें उन्होंने कभी लिखा नहीं था।

Amazon और Goodreads की खोज के शीर्ष पर Friedman के नाम के साथ दिखाई देने वाली वे किताबें जनरेटिव एआई द्वारा बनाई गई थीं। यदि आप उन्हें खोलकर देखते हैं, तो वाक्य कुछ अजीब लगते हैं और उनमें कोई सार नहीं होता है। यह बड़े भाषा मॉडल द्वारा लिखे गए लेखों की विशेषता है। इसमें कुछ भी गलत नहीं होता, लेकिन इसमें याद रखने लायक भी कुछ नहीं होता है।

जीवन भर की कमाई हुई प्रतिष्ठा रातों-रात निम्न-गुणवत्ता वाले टेक्स्ट के साथ जुड़ गई थी।

Friedman ने Amazon से उन्हें हटाने का अनुरोध किया। एक जवाब आया। आपने अपने नाम को ट्रेडमार्क के रूप में पंजीकृत नहीं किया है।

इसका मतलब था कि अपने नाम का उपयोग करने के लिए, उन्हें पहले इसे ट्रेडमार्क के रूप में पंजीकृत करना चाहिए था। यह दुनिया के सबसे नौकरशाही वाक्यों में से एक है।

Friedman द्वारा अपने ब्लॉग पर इस घटना के बारे में पोस्ट करने और लोगों के क्रोधित होने के बाद ही Amazon ने चुपचाप किताबों को हटा दिया। उन्होंने इसे इसलिए नहीं हटाया क्योंकि नियम बदल गए थे, बल्कि इसलिए क्योंकि बहुत शोर हो रहा था।

यह एक राहत की बात होती यदि नकली किताबें केवल नाम चुराने तक ही सीमित रहतीं, लेकिन ऐसा नहीं हुआ।

सितंबर 2023 में, न्यूयॉर्क माइकोलॉजिकल सोसाइटी ने एक आपातकालीन चेतावनी जारी की। इसमें कहा गया था कि Amazon पर बेची जा रही शुरुआती लोगों के लिए जंगली मशरूम इकट्ठा करने की कई गाइडबुक्स ChatGPT द्वारा लिखी गई थीं।

मशरूम की किताब का मुख्य बिंदु एक ही होता है। यह कैसे पहचाना जाए कि क्या खाया जा सकता है और क्या खाने से मौत हो सकती है। वह हिस्सा गलत था।

यह कोई टाइपिंग की गलती नहीं थी। यह ऐसी जानकारी थी जिससे लोग मर सकते थे।

फरवरी 2024 में, राजा Charles III के कैंसर के निदान की खबर आते ही Amazon पर सात नकली आत्मकथाएँ दिखाई दीं। राजा को त्वचा कैंसर होने या किसी छिपी हुई दुर्घटना का शिकार होने जैसी बातें वास्तविक समाचार की तरह मिला दी गई थीं। Buckingham Palace इससे बहुत क्रोधित हुआ था।

तेजी से लिखें, तेजी से अपलोड करें, खोज के शीर्ष पर कब्जा करें और पैसा कमाएं। एआई के आने से पहले, यह वह काम था जिसे करने में लोगों को कई दिन लगते थे। अब इसमें बस कुछ मिनट लगते हैं।

कभी-कभी यह स्वचालन अपनी पहचान स्वयं उजागर कर देता है।

जनवरी 2024 में, Amazon पर अजीबोगरीब उत्पाद दिखाई दिए। ये बगीचे की कुर्सियां, होज़ और मेज थे। उत्पाद के नाम की जगह यह वाक्य लिखा था।

क्षमा करें। यह अनुरोध OpenAI की सेवा की शर्तों की नीति का उल्लंघन करता है और इसे पूरा नहीं किया जा सकता है।

विक्रेता ने एआई से उत्पाद का विवरण लिखने के लिए कहा, एआई ने मना कर दिया, और उन्होंने बिना पढ़े ही अस्वीकृति वाले वाक्य को ज्यों का त्यों कॉपी करके अपलोड कर दिया।

यह एक मजाकिया दृश्य है। लेकिन अगर आप इसके दूसरे पहलू को देखें, तो यह डरावना है। इस विक्रेता ने उस वस्तु का विवरण नहीं पढ़ा जो वह बेच रहा था। इंटरनेट पर ऐसे कितने विवरण होंगे जिन्हें पढ़ा नहीं गया है?

अप्रैल 2024 में, Google Books ने एआई द्वारा लिखी गई निम्न-गुणवत्ता वाली किताबों को अपने पुस्तक खोज डेटाबेस में वैसे ही इकट्ठा कर लिया। उनमें से कई किताबों में यह वाक्य बचा रह गया था। मेरे नवीनतम ज्ञान अपडेट के अनुसार।

Google Ngram Viewer नामक एक टूल है। यह एक अकादमिक टूल है जो पुस्तक डेटा का उपयोग करके ट्रैक करता है कि किसी विशेष युग में किसी शब्द का कितना उपयोग किया गया था। यह भाषा के इतिहास को मापने वाला एक पैमाना है। जब उस पैमाने में एआई द्वारा लिखी गई किताबें मिल जाती हैं, तो उसके निशान धुंधले हो जाते हैं।

यह अब तक प्रकाशन जगत की कहानी है। समाचार पत्रों की कंपनियां और भी बुरी तरह से गिर गईं।

मई 2025 में, Chicago Sun-Times ने गर्मियों के लिए अनुशंसित पुस्तकों की सूची नामक एक विशेष परिशिष्ट प्रकाशित किया। Isabel Allende, Andy Weir और Min Jin Lee जैसे प्रसिद्ध लेखकों की नई किताबें शानदार समीक्षाओं के साथ प्रदर्शित की गई थीं।

पाठक किताबों की दुकान पर गए। वहां कोई किताब नहीं थी।

उस सूची की कोई भी किताब दुनिया में मौजूद नहीं थी। लेखक असली लोग थे, लेकिन किताबों के नाम एआई द्वारा गढ़े गए थे। असली नामों के साथ झूठी जानकारी जोड़ना एआई के झूठ बोलने का सबसे अच्छा तरीका है। जब आप इसकी पुष्टि करने के लिए खोज करते हैं, तो आप आश्वस्त हो जाते हैं क्योंकि लेखक का नाम असली होता है।

समाचार पत्र कंपनी ने अपने कर्मचारियों की संख्या कम कर दी थी और बाहरी कंपनियों से सामग्री प्राप्त कर रही थी। फ्रीलांस लेखक ने एआई द्वारा निकाली गई सूची को ज्यों का त्यों कॉपी कर लिया, और संपादकीय टीम ने इसे एक बार भी नहीं जांचा।

यह वह क्षण था जब एक बड़े समाचार पत्र के नाम ने एआई के झूठ पर विश्वास की मुहर लगा दी।

नवंबर 2023 में, Sports Illustrated पकड़ा गया। खोजी पत्रकारिता आउटलेट Futurism ने खुलासा किया कि यह प्रकाशन लगातार एआई द्वारा लिखे गए लेख प्रकाशित कर रहा था।

लेखों में सूचीबद्ध पत्रकारों के नाम, चेहरों की तस्वीरें और अनुभव सभी नकली थे। चेहरों की तस्वीरें सिंथेसिस एल्गोरिदम का उपयोग करके बनाई गई थीं। जिन पत्रकारों का कोई अस्तित्व नहीं था, वे बिना किसी अनुभव के उत्पाद समीक्षाएँ लिख रहे थे।

प्रकाशन ने बाहरी ठेकेदारों को दोषी ठहराया। मानव पत्रकारों को कम करने और स्वचालन लाने का निर्णय प्रबंधन द्वारा लिया गया था।

स्थानीय समाचार पत्रों में, पत्रकारों ने खुद इसमें हाथ आजमाया।

अगस्त 2024 में, व्योमिंग में Cody Enterprise के एक नए पत्रकार Aaron Pelczar ने इस्तीफा दे दिया। यह बात तब सामने आई जब एक प्रतिद्वंद्वी समाचार पत्र के पत्रकार ने उनके लेखों में अजीब तरह से सहज वाक्य देखे और सीधे स्रोतों को फोन किया।

Pelczar ने राज्यपाल सहित छह सार्वजनिक अधिकारियों के बयान गढ़ने के लिए ChatGPT का उपयोग किया और उन्हें अपने लेखों में शामिल किया। उसने उन्हें वे बातें कहते हुए दिखाया जो उन्होंने कभी कही ही नहीं थीं।

सबसे बड़ी बात कुछ और ही है। उनके द्वारा लिखे गए लेख के अंत में, उल्टे त्रिकोण वाली रिपोर्टिंग तकनीक को समझाने वाला एआई का निर्देश संदेश बिना हटाए छोड़ दिया गया था। यह ऐसा था जैसे किसी ने परीक्षा के पेपर के साथ चीट शीट लगा दी हो।

मई 2023 में, The Irish Times ने एक राय लेख प्रकाशित किया जिसमें दावा किया गया था कि कृत्रिम टैनिंग नस्लवादी थी, लेकिन बाद में यह पता चलने पर कि पूरा लेख एआई द्वारा गढ़ा गया था, उन्हें इसे वापस लेना पड़ा और माफी मांगनी पड़ी।

अब हम वास्तव में डरावने हिस्से की ओर बढ़ते हैं। एआई ऐसी घटनाएं गड़ता है जो कभी हुईं ही नहीं।

दिसंबर 2023 में, समाचार ऐप NewsBreak ने एक लेख प्रकाशित किया जिसमें बताया गया था कि क्रिसमस के दिन न्यू जर्सी के ब्रिजटन में एक महिला की गोली मारकर हत्या कर दी गई थी। स्थानीय निवासी हैरान थे और पुलिस ने इसकी जांच की।

ऐसी कोई घटना नहीं हुई थी।

NewsBreak द्वारा उपयोग किए जा रहे समाचार पुनर्लेखन इंजन ने इंटरनेट पर तैरती हुई खंडित जानकारी को मिलाकर एक हत्या का मामला गढ़ लिया था।

अक्टूबर 2024 में, स्थानीय समाचार नेटवर्क Hoodline ने एक लेख प्रकाशित किया जिसमें कहा गया था कि कैलिफोर्निया में सैन मेटो काउंटी के जिला अटॉर्नी John Casciano Thompson पर हत्या का आरोप लगाया गया था। अपराधों पर मुकदमा चलाने वाला व्यक्ति ही हत्यारा बन गया था।

व्यक्तियों के साथ भी ऐसा ही होता है।

जून 2023 में रेडियो होस्ट मार्क वाल्टर्स को पता चला कि ChatGPT ने एक नकली अदालत सारांश बनाया था जिसमें उन्हें पांच मिलियन डॉलर से अधिक की रिश्वत देने वाले धोखेबाज के रूप में वर्णित किया गया था। उन्होंने OpenAI पर मानहानि का मुकदमा दायर किया।

परिस्थिति इस प्रकार थी। पत्रकार फ्रेड रील ने ChatGPT से एक बंदूक अधिकार संगठन के वास्तविक मुकदमे का सारांश देने को कहा। ChatGPT ने वाल्टर्स को पेश किया, जिनका उस मुकदमे से कोई संबंध नहीं था, और उन्हें एक विश्वसनीय केस नंबर और सजा के साथ एक चोर घोषित किया।

मानचित्र सेवा कंपनी OpenCage को ChatGPT के कारण परेशानियों का सामना करना पड़ा क्योंकि इसने कहा कि वे एक फोन नंबर खोज सेवा प्रदान करते हैं।

मार्च 2025 में नॉर्वे के Arve Jarmar Holmen ने ChatGPT से अपना नाम पूछा। यह उत्तर था: वह दो बच्चों को मार डालने और तीसरे को मारने का प्रयास करने के लिए इक्कीस साल की सजा दी गई थी।

इस वाक्य में, उसकी मातृभूमि और बच्चों की संख्या वास्तविक थी। बाकी सब झूठ था।

असली और नकली को मिलाने का यह तरीका संभाव्यता मशीन की आदत है। यह तथ्यों को संग्रहीत और पुनः प्राप्त नहीं करता; यह शब्दों को चुनता है जो पिछले शब्द के बाद आने की संभावना रखते हैं। जब आप नॉर्वेजियन नाम के बाद प्राकृतिकता से आने वाले शब्दों को चुनते हैं, तो अपराध की रिपोर्ट निकलती है। मशीन नहीं जानती कि यह किसी का जीवन है।

जब राजनीति शामिल होती है, तो यह आदत एक हथियार बन जाती है।

नवंबर 2023 में, पाकिस्तान की समाचार साइट ग्लोबल विलेज स्पेस ने एक लेख प्रकाशित किया जिसमें कहा गया कि इजरायली प्रधानमंत्री Benjamin Netanyahu के निजी मनोचिकित्सक Moshe Yatom ने प्रधानमंत्री की क्रूरता की आलोचना करते हुए एक नोट छोड़ा और आत्महत्या कर ली।

Moshe Yatom नाम का कोई व्यक्ति नहीं है। कोई नोट नहीं था। यह सब बनाया गया था। यह लेख इजरायल और Hamas के बीच युद्ध के दौरान कई भाषाओं में अनुवादित और फैलाया गया था।

अप्रैल 2024 में, X पर Grok ने इंटरनेट पर अफवाहों को इकट्ठा करके ईरान के तेल अवीव पर मिसाइल दागने की सुर्खियां बनाईं। इसने बास्केटबॉल खिलाड़ी Clay Thompson के बारे में एक लेख भी बनाया कि वह शहर में ईंटें फेंक रहा है। उत्तरार्द्ध शायद यह है कि वह शूटिंग को मिस करने के बारे में बास्केटबॉल स्लैंग को शब्दों के अनुसार पढ़ा गया है।

एक मशीन जो मजाक नहीं समझ सकती है वह समाचार लिख रही है।

मीडिया नाम की नकल करने का तरीका और भी सूक्ष्म है।

ब्रिटिश गार्जियन के संपादकों को पाठकों से अजीब सवाल मिलते रहे। उन्होंने पूछा कि क्या वे एक निश्चित लेख देख सकते हैं। गार्जियन ने कभी ऐसा लेख नहीं लिखा था।

जब ChatGPT से संसाधन के लिए कहा गया, तो इसने गार्जियन में वास्तविक पत्रकारों के नाम, विश्वसनीय शीर्षक और प्रकाशन वर्ष को जोड़ा। इसने असली प्रकाशन और वास्तविक पत्रकारों का उपयोग करके झूठों को सत्ता दी। गार्जियन के इनोवेशन एडिटर Chris Moran ने इस घटना का सार्वजनिक रूप से पर्दाफाश किया।

जुलाई 2024 में Nieman Journalism Lab ने इस समस्या का उचित परीक्षण किया। उन्होंने ChatGPT से AP, Wall Street Journal, Financial Times, The Times, Boston Globe, और Politico की जांच पड़ताल की रिपोर्ट के लिंक मांगे। ये वह समाचार संगठन हैं जिन्होंने OpenAI के साथ औपचारिक डेटा समझौते किए हैं।

ChatGPT ने विस्तृत सारांश के साथ असली दिखने वाले पते दिए। वे सभी अस्तित्व में नहीं थे।

शिक्षा जगत भी ऐसा ही है। प्रोफेसर Henrik Enghoff का एक अस्तित्वहीन पत्र उद्धृत किया गया था, और डूबने से होलोकॉस्ट एक ऐतिहासिक घटना बनाई गई थी।

खोज पट्टी से जुड़ी कृत्रिम बुद्धिमत्ता ने इस समस्या को घर तक पहुंचा दिया है।

Google AI Overview ने कभी-कभी कहा है कि पिज्जा चीज को बहने से रोकने के लिए कुछ गोंद मिलाएं। यह इंटरनेट मंचों पर एक मजाक को गंभीरता से पढ़ने का परिणाम है। इसने जहरीली मशरूम की तस्वीरें खाने योग्य मशरूम की सिफारिशों के शीर्ष पर भी रखी हैं।

एक शैक्षणिक अध्ययन ने बताया कि वित्तीय खोज के शब्दों के 43 प्रतिशत में गंभीर त्रुटियां थीं। Google ने सेवा जारी रखी। यह खोज बाजार को नहीं दे सकता था।

दक्षिण कोरिया में भी घटनाएं हुई हैं। Naver की जनरेटिव कृत्रिम बुद्धिमत्ता ने Dokdo के बारे में एक प्रश्न का उत्तर दिया और जापानी विदेश मंत्रालय के दावों के आधार पर इसे विवादित क्षेत्र और जापानी क्षेत्र के रूप में सारांशित किया।

नकली जानकारी स्क्रीन से बाहर भी निकल जाती है।

अक्टूबर 2024 में, डबलिन, आयरलैंड के शहर में भीड़ लगी। यह इसलिए था क्योंकि ChatGPT द्वारा बनाई गई एक अस्तित्वहीन हैलोवीन परेड की जानकारी खोज के शीर्ष पर दिखाई दी। हजारों नागरिक और पर्यटक बारिश की सड़कों पर खड़े होकर एक परेड की प्रतीक्षा कर रहे थे जो कभी नहीं आया।

नवंबर 2025 में लंदन में, कृत्रिम बुद्धिमत्ता की एक तस्वीर और लेख सोशल मीडिया पर फैल गया जिसमें कहा गया कि बकिंघम पैलेस के सामने एक शानदार शाही क्रिसमस बाजार होगा। दुनिया भर के पर्यटक बंद लोहे के दरवाजों और बारिश के पानी के सामने खड़े थे।

जनवरी 2026 में तस्मानिया, ऑस्ट्रेलिया में, पर्यटकों ने कृत्रिम बुद्धिमत्ता द्वारा लिखे गए एक लेख पर विश्वास किया और एक अस्तित्वहीन गर्म वसंत को खोजने के लिए वाइल्डरनेस में भटकते रहे।

इन सभी घटनाओं की जड़ प्रौद्योगिकी की प्रकृति में है।

बड़े भाषा मॉडल तथ्य जानने वाली मशीनें नहीं हैं। वे उन शब्दों को चुनने वाली मशीनें हैं जो पिछले शब्द के बाद आने की संभावना है। यह निर्धारित करने के लिए कि क्या सच है, कोई तंत्र नहीं है। कोई तंत्र न होना एक दोष नहीं है - यह डिजाइन है।

तो क्या मॉडल को बड़ा करने से यह बेहतर हो जाएगा? 2024 में आए अध्ययनों ने विपरीत परिणाम दिए। जैसे-जैसे मॉडल बड़ा होता जाता है, वह मुझे नहीं पता कहने के बजाय अधिक स्वाभाविक और आत्मविश्वास से गलत

उत्तर देने लगता है।

यह अधिक बुद्धिमान होना नहीं है - यह बेहतर बोलना है।

संपादक और तथ्य-जांचकर्ता मूलतः वे लोग थे जिन्हें इस समस्या को रोकने के लिए नियुक्त किया जाता था। लेकिन उन पदों को लागत के रूप में वर्गीकृत किया गया और पहले समाप्त कर दिया गया।

जब समस्या उत्पन्न होती है, तो कंपनियां एक ही उत्तर देती हैं। यह अभी भी बीटा सेवा है। स्क्रीन के नीचे एक नोट कहता है कि यह गलत हो सकता है।

वह छोटा पाठ उस व्यक्ति तक नहीं पहुंचता जिसने मशरूम की किताब खरीदी है।

2.2

अदालत और शैक्षणिक क्षेत्र में फर्जी फैसलों और उद्धरणों का विवाद

शुरुआत एक घुटने से हुई।

रोबर्टो माता एविएंका एयरलाइंस के विमान में केबिन कार्ट से टकरा गए। उन्होंने एयरलाइंस पर मुआवजे के लिए मुकदमा किया। यह एक आम मामला था। इस तरह के मुकदमे फैसलों के संग्रह में दर्ज नहीं होते हैं।

मई 2023 में यह दुनिया के कानूनी समुदाय का एक ज्ञात मामला बन गया। घुटने की वजह से नहीं।

माता के वकील स्टीवन श्वार्ट्ज ने एयरलाइंस की खारिजी याचिका के विरुद्ध छह फैसलों का उल्लेख किया। वर्गेस बनाम चाइना सदर्न एयरलाइंस जैसे नाम थे। मामले के नंबर थे, फैसले की तारीख थी, अदालत का नाम था। रूप पूरी तरह सही था।

विरोधी पक्ष के वकीलों ने फैसलों को खोजा। नहीं मिले। न्यायाधीश के सहायक ने खोजा। नहीं मिले।

सभी छह फैसले अस्तित्वहीन थे।

श्वार्ट्ज ने दस्तावेज़ तैयार करने में चैटजीपीटी का उपयोग किया। पूछताछ में उन्होंने कहा कि उन्हें नहीं पता था कि कृत्रिम बुद्धिमत्ता झूठ बोल सकती है। उन्होंने चैटजीपीटी से यह भी पूछा कि क्या यह फैसला असली है। चैटजीपीटी ने कहा कि हाँ, असली है।

एक झूठे से पूछना कि क्या वह झूठ बोला है।

सहकर्मी वकील पीटर लोडुकस ने बिना जांच किए हस्ताक्षर किए। अदालत ने दोनों को पांच हजार डॉलर का जुर्माना लगाया।

यह मामला प्रसिद्ध हुआ जुर्माने की राशि के कारण नहीं। इसलिए कि लोगों को पहली बार यह देखने को मिला कि अदालतों में दायर दस्तावेज़ कैसे बनाए जाते हैं।

और यह दोहराया गया। बहुत बार दोहराया गया।

फ्रांसीसी शोधकर्ता डैमिएन शारलोटिन ने कृत्रिम बुद्धिमत्ता के मतिभ्रम से बनाए गए फैसलों को एकत्रित करने के लिए एक डेटाबेस बनाया। आइए देखते हैं कि संख्या कैसे बढ़ी।

2025 के मध्य में लगभग 200 मामले थे। 2026 की जनवरी में 719 हो गए। 2026 की अप्रैल शुरुआत में 1,227, जून 9 को 1,598।

22 मई से 9 जून तक अठारह दिनों में 140 मामले बढ़े। हर दिन आठ मामलों की दर से।

इसका मतलब है कि अदालत में हर दिन आठ बार नकली फैसले पकड़े जा रहे हैं। जो पकड़े नहीं गए उन्हें गिना नहीं गया।

इस सूची में आने वाले मामलों को देखें तो राष्ट्रीयता और रैंक विविध हैं।

डोनाल्ड ट्रम्प के पूर्व राष्ट्रपति के सहायक माइकल कोहन ने अपनी निगरानी अवधि को जल्दी समाप्त कराने के लिए दस्तावेज़ तैयार करते समय गूगल बार्ड का उपयोग किया। उद्धरणों को असली समझकर अपने वकील को दिए, वकील ने बिना जांच किए अदालत में दाखिल कर दिए। न्यायाधीश ने बताया कि ये फैसलों के संग्रह में नहीं हैं, तब पता चला।

न्यूयॉर्क की वकील जेन ली ने क्वींस के एक डॉक्टर द्वारा गलत गर्भपात सर्जरी के बारे में नकली फैसला चैटजीपीटी से बनाया और दूसरी संघीय अपीलीय अदालत में दाखिल किया, फिर अनुशासनात्मक कार्यवाही के लिए भेजा गया।

कनाडा की ब्रिटिश कोलंबिया की सर्वोच्च अदालत में विवाह मुकदमे के दौरान पिता बच्चों को चीन ले जा सकता है - यह दिखाने वाले दो फैसले मतिभ्रम साबित हुए। न्यायाधीश ने उस वकील को विरोधी पक्ष के वकील की फीस अपनी जेब से चुकानी दी।

ब्रिटेन की पहली अदालत का फेलिसिटी हार्बर मामला गुप्त चर उपन्यास जैसा है। वह एक हजार दो सौ पैंसठ पाउंड के जुर्माने के आदेश के खिलाफ नौ फैसले सौंपे। न्यायाधीश ऐन रेडस्टोन को अजीब बात दिखी। ब्रिटिश फैसले में अमेरिकी वर्तनी मिली। और कहीं-कहीं रूढ़िबद्ध वाक्य दोहराए गए थे।

मशीन अपनी राष्ट्रीयता को छुपा नहीं सकता।

स्पेन के कैनारी द्वीपसमूह की सर्वोच्च अदालत ने एक वकील को - जिसने अड़तालीस नकली फैसले सौंपे - चार सौ बीस यूरो का जुर्माना लगाया। उस वकील ने आधिकारिक कानूनी डेटाबेस से भी तुलना नहीं की।

ऑस्ट्रेलिया की मेलबोर्न पारिवारिक अदालत में कानूनी विशेष सॉफ्टवेयर का उपकरण नकली फैसले बना रहा था और मुकदमा स्थगित हो गया। तस्मानिया की सर्वोच्च अदालत में अपमान की अपील मामले में मुख्य न्यायाधीश ने अस्तित्वहीन फैसले को सीधे निकाला और मामला खारिज कर दिया।

ब्राज़ील में एक संघीय न्यायाधीश को निर्णय लिखने के भार की वजह से अपने सहायक को मशीन लर्निंग उपकरण का उपयोग करने दिया, जिससे उच्च न्यायालय के फैसलों को विकृत कर दिया गया और जांच हुई।

निर्णय कृत्रिम बुद्धिमत्ता ने लिखा था। जो व्यक्ति अदालत में खड़ा था वह यह नहीं जानता था।

2026 में यह बढ़ गया।

फरवरी 2026 में विवाह अपील मामले में वकील ग्रेग लेक द्वारा दायर लेख में तिरपन उद्धरणों में से सत्तावन में खामी थी। बीस नकली फैसले, तीन बदले गए निर्णय, बनाए गए कानूनी उद्धरण।

तिरपन में से केवल छह असली थे।

नेब्रास्का की सर्वोच्च अदालत ने अप्रैल 2026 में उनका लाइसेंस निलंबित किया और अनुशासनात्मक जांच का आदेश दिया।

मार्च 2026 में छठी संघीय अपीलीय अदालत के व्हाइटिंग फैसले में विलय किए गए तीन अपील पत्रों में चौबीस से अधिक नकली उद्धरण और तथ्य विकृतियाँ थीं। ओरेगन संघीय जिला अदालत के एक मामले में प्रतिबंध और खर्चे मिलाकर लगभग नौ हजार सात सौ डॉलर आरोपित किए गए। 2026 की पहली तिमाही में अकेले अमेरिकी अदालतों ने कृत्रिम बुद्धिमत्ता संबंधित दस्तावेज़ों पर डेढ़ लाख डॉलर से अधिक का जुर्माना आरोपित किया।

रिकॉर्ड को देखने से एक पैटर्न दिखता है। नकली उद्धरण से ज्यादा, जो उसे छुपाने की कोशिश करते हैं उन्हें कठोर दंड दिया जाता है। न्यायाधीश गलतियों को माफ कर देते हैं लेकिन झूठ को नहीं।

कंपनियां भी अपवाद नहीं थीं।

दुनिया के पहले रोबोट वकील के रूप में विज्ञापित डोनॉटपे ने वकील सत्यापन के बिना कानूनी दस्तावेज़ों का विश्लेषण कर सकने का दावा किया, फिर संघीय व्यापार आयोग के समक्ष आए, उन्हें एक लाख निन्यानबे हजार डॉलर का जुर्माना दिया।

व्यक्तिगत चोट के दस्तावेज़ बनाने को स्वचालित करने वाली ईवनअप की सच्चाई पूर्व कर्मचारियों के खुलासों से सामने आई। कृत्रिम बुद्धिमत्ता जब चोट की जानकारी भूल जाती या अस्तित्वहीन बीमारी बनाती थी, तो कर्मचारी हाथ से ठीक कर रहे थे।

होवरबोर्ड आग के नुकसान के मुकदमे में कानून फर्म के अपने डेटाबेस द्वारा बनाए गए आठ नकली फैसले प्रस्तुत किए गए। न्यायाधीश ने वकीलों को मिलाकर पांच हजार डॉलर का जुर्माना लगाया और प्रधान वकील को मामले से निकाल दिया।

बड़ी कानून फर्मों को भी पकड़ा गया। स्टेटफार्म बीमा के मामले में दो विशाल कानून फर्मों के वकीलों ने जेमिनाई और कानूनी विशेष उपकरण को मिलाकर इस्तेमाल किया, सत्तीस उद्धरणों में नौ त्रुटि, दो पूरी तरह नकली साबित हुए, उन्हें इकतीस हजार डॉलर का जुर्माना दिया।

माइपिलो के अध्यक्ष माइक लिंडेल के वकीलों ने तीस से अधिक नकली फैसलों को वैसे ही रखा, प्रति वकील तीन हजार डॉलर का जुर्माना सहा।

सबसे अजीब मामला कृत्रिम बुद्धिमत्ता कंपनी के अंदर से आया।

यूनिवर्सल म्यूजिक ग्रुप और एंथ्रोपिक के कॉपीराइट मुकदमे में, एंथ्रोपिक के डेटा वैज्ञानिक ने जो दस्तावेज़ दाखिल किया उसमें अपनी कंपनी के चैटबॉट द्वारा बनाए गए नकली शैक्षणिक पत्र थे। कृत्रिम बुद्धिमत्ता कंपनी को कृत्रिम बुद्धिमत्ता के कारण अदालत में झिड़की खानी पड़ी।

स्टैनफोर्ड विश्वविद्यालय के मीडिया सूचना विशेषज्ञ प्रोफेसर जेफ हैनकॉक ने मिनेसोटा डीपफेक प्रतिबंध कानून के मुकदमे में दिए गए अपने बयान में मतिभ्रम से बने दो नकली शोध पत्रों का हवाला दिया। डीपफेक का खतरा समझाने आए एक विशेषज्ञ ने कृत्रिम बुद्धिमत्ता द्वारा बनाई गई सामग्री का हवाला दिया था।

अदालत से बाहर निकलें तो स्कूल आता है। यहाँ भी स्थिति कुछ अलग नहीं है।

अगस्त 2023 में, डेनमार्क के जीवविज्ञानी प्रोफेसर हेनरिक एंगहॉफ ने इथियोपिया के शोधकर्ताओं का मिलीपीड पर एक शोध पत्र पढ़ते हुए अपनी आँखों पर विश्वास नहीं किया। दो शोध पत्र जो उन्होंने नहीं लिखे थे, उनका हवाला उनके नाम से दिया गया था।

शोधकर्ताओं ने कहा कि उन्होंने शोध पत्र लिखने के लिए ChatGPT का उपयोग किया था, और उन्हें बाद में पता चला कि नकली हवाले बन गए थे। शोध पत्र को वापस ले लिया गया।

शिक्षा जगत में हवाला देना एक वंशावली की तरह है। यह यह लिखने का काम है कि कोई विचार कहाँ से आया है। नकली हवाला देना ऐसा है जैसे वंशावली में ऐसे पूर्वज का नाम लिखना जो है ही नहीं। एक बार यह मिल जाए, तो बाद में आने वाले लोग लगातार गलती करते रहते हैं।

यह समस्या कृत्रिम बुद्धिमत्ता से भी पुरानी है।

2008 से 2013 के बीच, फ्रांसीसी शोधकर्ता सिरिल लाबे ने प्रसिद्ध अकादमिक प्रकाशक Springer और IEEE के सम्मेलन शोध पत्र संग्रह में एक सौ बीस से अधिक नकली शोध पत्र खोजे। ये शोध पत्र 2005 में MIT के स्नातक छात्रों द्वारा बनाए गए एक स्वचालित शोध पत्र निर्माण कार्यक्रम का परिणाम थे, जिसे कमजोर समीक्षा वाले अकादमिक सम्मेलनों का मजाक उड़ाने के लिए बनाया गया था।

अर्थहीन शब्दों का संयोजन समीक्षा को पार कर गया और आधिकारिक शोध पत्र संग्रह में प्रकाशित हो गया। यह बीस साल पहले की बात है।

फरवरी 2024 में, सेल बायोलॉजी की एक पत्रिका ने एक शोध पत्र वापस ले लिया। शोध पत्र के अंदर Midjourney द्वारा बनाए गए चित्र थे, जिनमें से एक विशाल चूहे का चित्र था जो शारीरिक रचना के हिसाब से बिल्कुल गलत था। चित्र पर लिखे अक्षर भी अर्थहीन वर्णमाला की एक श्रृंखला थे।

कई समीक्षकों ने इस चित्र को देखा और इसे मंजूरी दे दी थी।

2024 में 404 मीडिया की जाँच में, Google Scholar पर पंजीकृत एक सौ पंद्रह से अधिक शोध पत्रों में "मेरे नवीनतम ज्ञान अपडेट के अनुसार" वाक्य बचा हुआ था। उन्होंने बस कॉपी-पेस्ट किया और उसे पढ़ा नहीं था।

Journal of Human Evolution में, प्रकाशक ने संपादकीय बोर्ड से परामर्श किए बिना पांडुलिपि संपादन प्रक्रिया में कृत्रिम बुद्धिमत्ता को शामिल कर लिया। जब पास की गई पांडुलिपि का प्रारूप टूट गया और सामग्री बिगड़ गई, तो सभी संपादकीय बोर्ड के सदस्यों ने इस्तीफा दे दिया।

स्टैनफोर्ड के शोधकर्ताओं ने बताया कि वैश्विक कृत्रिम बुद्धिमत्ता सम्मेलन की शोध पत्र समीक्षाओं में से अधिकतम 17 प्रतिशत ChatGPT द्वारा लिखी गई थीं।

यह एक ऐसा ढांचा है जहाँ कृत्रिम बुद्धिमत्ता के शोध पत्रों की समीक्षा कृत्रिम बुद्धिमत्ता ही कर रही है। यह पूछना लाजमी है कि इंसान कहाँ हैं।

नीतिगत दस्तावेजों में भी ऐसा ही हुआ।

ऑस्ट्रेलिया के मैक्वेरी विश्वविद्यालय के एमेरिटस प्रोफेसर जेम्स गुथरी ने कंसल्टिंग उद्योग को विनियमित करने का अनुरोध करते हुए संसद में जमा किए जाने वाले दस्तावेज़ लिखने के लिए Google Bard का उपयोग किया। बार्ड ने ऐसी घटनाएँ गढ़ीं जैसे कि Deloitte पर एक खराब निर्माण कंपनी के ऑडिट में लापरवाही बरतने के लिए मुकदमा चलाया गया था, या कि KPMG एक सुविधा स्टोर के वेतन शोषण मामले में शामिल था। ये सब ऐसी बातें थीं जो कभी हुई ही नहीं थीं। प्रोफेसर ने सीनेट को माफ़ीनामा भेजा।

कंसल्टिंग उद्योग की आलोचना करने के इरादे से बनाए गए दस्तावेज़ ने कंसल्टिंग कंपनियों पर ऐसे अपराधों का झूठा आरोप लगाया जो कभी हुए ही नहीं।

ऑस्ट्रेलियाई सरकार की कल्याण प्रणाली रिपोर्ट लिखने वाली Deloitte ऑस्ट्रेलिया शाखा ने भी ChatGPT द्वारा बनाए गए नकली अकादमिक हवाले और संघीय न्यायालय के फैसले के वाक्यों को ज्यों का त्यों शामिल कर लिया था, जिसके लिए उसे आलोचना का सामना करना पड़ा और पैसे का भुगतान करना पड़ा।

2025 में, अमेरिकी स्वास्थ्य और मानव सेवा विभाग और व्हाइट हाउस द्वारा जारी एक स्वास्थ्य रिपोर्ट में सात से अधिक नकली अकादमिक अध्ययन शामिल थे। ChatGPT के विशेष उद्धरण चिह्न वैसे ही बचे रह गए थे।

जलवायु परिवर्तन से इनकार करने वाले लोगों ने Grok 3 मॉडल का उपयोग करके शोध पत्र बनाए और उन्हें निजी पत्रिकाओं में प्रकाशित किया। इसका उपयोग वैज्ञानिक सहमति को कमजोर करने के लिए किया गया।

आइए इन घटनाओं से गुजरने वाले लोगों के पेशों को गिनें। वे वकील, न्यायाधीश, प्रोफेसर, सरकारी अधिकारी और बड़े कॉर्पोरेट सलाहकार हैं।

ये वे लोग हैं जिन्होंने शब्दों के साथ काम करने का लंबा प्रशिक्षण प्राप्त किया है। ये वे लोग हैं जो दस्तावेजों की प्रामाणिकता की जाँच करके पैसे कमाते हैं।

फिर भी वे पकड़े गए। ऐसा इसलिए है क्योंकि कृत्रिम बुद्धिमत्ता द्वारा बनाए गए दस्तावेज़ों के प्रारूप में कोई कमी नहीं होती। उनमें केस नंबर होता है, तारीख होती है, और शैली कानूनी दस्तावेज़ों जैसी ही होती है। इंसानों की आदत होती है कि वे रूप-रंग के आधार पर असल या नकल की पहचान करते हैं।

नकली कानूनी फैसले उस आदत के ठीक विपरीत खड़े हैं। उनमें कोई विषय-वस्तु नहीं होती, केवल रूप-रंग होता है।

अदालत में तथ्यों पर बहस करने का कारण यह है कि इसमें लोगों की स्वतंत्रता और संपत्ति दांव पर लगी होती है। यदि उस बहस का आधार कोई ऐसा फैसला था जो कभी अस्तित्व में ही नहीं था, तो फिर वह मुकदमा क्या था?

हर दिन ऐसे आठ मामले पकड़े जा रहे हैं।

2.3

ऐतिहासिक और भौगोलिक तथ्यों का विकृतिकरण तथा सामाजिक विवाद

अक्टूबर 2025 में, किसी ने Naver के सर्च बॉक्स में जापानी भाषा में जापानी क्षेत्र टाइप किया।

स्क्रीन के ऊपर आर्टिफिशियल इंटेलिजेंस द्वारा तैयार किया गया एक सारांश दिखाई दिया। उस सारांश में Dokdo शामिल था। यह उत्तरी क्षेत्रों और सेनकाकू द्वीप समूह के साथ रखा गया था, और इसे दक्षिण कोरिया के साथ विवादित क्षेत्र के रूप में लिखा गया था।

एक कोरियाई कंपनी की सेवा ने, एक ऐसे सर्च बॉक्स में जिसका इस्तेमाल कोरियाई लोग करते हैं, Dokdo को जापान की सूची में डाल दिया था।

Sungshin Women's University के प्रोफेसर Seo Kyoung-duk ने इस स्क्रीन को सार्वजनिक किया और लोग क्रोधित हो गए। Naver ने तुरंत सर्च सारांश सुविधा को रोक दिया और इसकी जांच शुरू कर दी।

इसका कारण तकनीकी रूप से उबाऊ और सामग्री के मामले में दर्दनाक है। HyperCLOVA X जब जापानी भाषा में प्रश्न प्राप्त करता है, तो वह जापानी वेब दस्तावेज़ों को इकट्ठा करके उनका सारांश बनाता है। जब आप जापानी में क्षेत्र सर्च करते हैं, तो जापानी विदेश मंत्रालय की प्रचार सामग्री सबसे ऊपर आती है। सिस्टम ने उसी सामग्री को सारांशित किया था।

उसने वही किया जो उसे करने का निर्देश दिया गया था। बस किसी ने यह तय नहीं किया था कि इस प्रश्न को किसी इंसान द्वारा एक बार देखा जाना चाहिए।

क्षेत्रीय मुद्दे में, किस देश की सामग्री को पहले पढ़ना है यह कोई तकनीकी समस्या नहीं है, बल्कि यह एक दृष्टिकोण की समस्या है। मशीन का कोई दृष्टिकोण नहीं होता है। दृष्टिकोण न होने को तटस्थता भी कहा जाता है। लेकिन अगर सामग्री एक तरफ झुकी हुई हो, तो बिना दृष्टिकोण वाली मशीन उसी झुकी हुई दिशा में लुढ़क जाती है।

ताइवान में इससे भी अधिक चौंकाने वाला उत्तर सामने आया।

अक्टूबर 2023 में, ताइवान के सर्वोच्च शैक्षणिक संस्थान, Academia Sinica ने ताइवान-शैली का एक भाषा मॉडल बनाया और उसे जारी किया। नागरिकों ने चैटबॉट से पूछा कि तुम किस देश से संबंधित हो।

चैटबॉट ने उत्तर दिया कि चीन।

उन्होंने पूछा कि ताइवान का सर्वोच्च नेता कौन है। उसने Xi Jinping उत्तर दिया। उन्होंने पूछा कि राष्ट्रगान क्या है। उसने मार्च ऑफ द वॉलंटियर्स उत्तर दिया।

डेवलपर्स ने इसका कारण बताया। उन्होंने इसे जल्दी बनाने के लिए, मुख्य भूमि चीन के सरलीकृत चीनी डेटासेट को एक अनुवादक के माध्यम से चलाया और उसे वैसे ही प्रशिक्षण में डाल दिया था।

अनुवादक अक्षरों को बदलता है। यह दृष्टिकोण को नहीं बदल सकता है। ताइवान सरकार और संस्थान ने तुरंत सिस्टम को हटा लिया।

फरवरी 2025 में, चीन का DeepSeek R1 मॉडल, जिसे दुनिया भर में तीस लाख से अधिक बार डाउनलोड किया गया था, उइगर नरसंहार के आरोपों के बारे में पूछे जाने पर जवाब देता था कि यह आंतरिक मामलों में हस्तक्षेप और एक निराधार आरोप है। उसने चीनी सरकार के तर्क को ही दोहराया था।

मॉडल वही कहता है जो उसने पढ़ा है। उसे कौन सी सामग्री खिलाई गई है, वही उस मॉडल का विश्वदृष्टिकोण बन जाता है।

यूरोपीय संसद भी इसमें फंसी। 2025 में, Anthropic के Claude पर आधारित आधिकारिक आर्टिफिशियल इंटेलिजेंस यूरोपीय आयोग के पहले अध्यक्ष का नाम नहीं बता सका। रिकॉर्ड को संभालने के लिए बनाई गई चीज़ को रिकॉर्ड के बारे में ही नहीं पता था।

इतिहास को तोड़ना-मरोड़ना इससे भी आगे जाता है।

जून 2024 में, यूनेस्को की एक रिपोर्ट में ChatGPT का एक उत्तर शामिल किया गया था। जब होलोकॉस्ट के बारे में पूछा गया, तो उसने एक ऐसी घटना का वर्णन किया जहां नाज़ी जर्मनी ने यहूदियों को सामूहिक रूप से पानी में डुबो कर मार डाला था।

ऐसी कोई घटना नहीं हुई थी। यहां तक कि इसे डूबने से होने वाला होलोकॉस्ट का नाम भी दिया गया था।

इतिहासकारों की चिंता का कारण इस झूठ की दिशा में है। होलोकॉस्ट को नकारने वाले लोग लंबे समय से रिकॉर्ड में छोटी-छोटी खामियों को ढूंढकर पूरे इतिहास पर सवाल उठाने के तरीके का इस्तेमाल करते रहे हैं। यदि न हुई घटनाएं रिकॉर्ड में मिल जाती हैं, तो जो घटनाएं वास्तव में हुई थीं, उन पर भी संदेह किया जाता है।

Elon Musk की xAI द्वारा बनाए गए Grok ने यहां सीमा पार कर दी।

2025 के मध्य में, Grok ने राजनीतिक शुद्धता को दूर करने के नाम पर अपने सिस्टम में बदलाव किए। इसके तुरंत बाद, Grok ने कहा कि साठ लाख होलोकॉस्ट पीड़ितों की संख्या में राजनीतिक रूप से हेरफेर किया गया हो

सकता है। उसने ऑशविट्ज़ गैस चैंबरों के अस्तित्व को नकारने वाला उत्तर भी दिया। उसने खुद को MechaHitler भी कहा।

जनवरी 2023 में, Amazon के एक इंजीनियर द्वारा GPT-3 का उपयोग करके बनाया गया एक ऐतिहासिक व्यक्ति वार्तालाप ऐप था। इसने विज्ञापित किया कि आप बीस हजार लोगों के साथ बातचीत कर सकते हैं। नाज़ी प्रचार मंत्री Joseph Goebbels चैटबॉट ने कहा कि वह यहूदियों से नफरत नहीं करता था बल्कि पीड़ित महसूस करता था।

अपराधी की आवाज़ की नकल करने वाली मशीन अपराधी के बहानों की भी नकल करती है। जब वह बहाना ऐप से आता है, तो कोई उसे ऐतिहासिक सामग्री समझ लेता है।

कंपनियों ने यह स्पष्टीकरण दिया कि यह प्रोग्रामिंग की त्रुटि थी या आंतरिक निर्देशों को मनमाने ढंग से बदल दिया गया था।

हम जो अपनी आंखों से देखते हैं उस पर भी विश्वास नहीं कर सकते।

जून 2025 में, ब्रिटिश फैक्ट-चेकिंग संस्था Full Fact ने एक अजीब मामले को पकड़ा। भारत के गोवा राज्य के एक समुद्र तट पर शूट किया गया एक साधारण वीडियो था। Google Lens से जुड़े आर्टिफिशियल इंटेलिजेंस सारांश ने इस वीडियो का वर्णन कुछ इस तरह किया: शरण चाहने वालों की एक बड़ी संख्या का ब्रिटेन के डोवर समुद्र तट पर प्रवेश करने का दृश्य।

अगर कोई समुद्र तट है और बहुत सारे लोग हैं, तो यह वह दृश्य बन जाता है। भारत, ब्रिटेन बन जाता है।

यह अनुमान लगाना मुश्किल नहीं है कि अगर यह सारांश सोशल मीडिया की अफवाहों से मिलता है तो क्या होता है। यह वह समय था जब ब्रिटेन में प्रवासी विरोधी भावनाएँ बढ़ रही थीं।

मई 2024 में इसी सेवा ने ऐसे उत्तर दिए थे: अगर पिज्जा चीज़ बह रहा है, तो उसमें गैर-जहरीला गोंद मिला दें। दिन में एक छोटा पत्थर खाएं। गुर्दे की पथरी के लिए मूत्र पिएं। Barack Obama अमेरिका के पहले मुस्लिम राष्ट्रपति हैं।

पत्थर खाने का जवाब एक व्यंग्यात्मक मीडिया के लेख से आया था। मशीन में मज़ाक और जानकारी के बीच अंतर करने की समझ नहीं होती है। इंसान बचपन में चेहरों के भाव और स्थितियों को देखकर इस समझ को सीखते हैं। मशीन केवल अक्षर प्राप्त करती है।

अप्रैल 2025 में, मलेशिया में राष्ट्रीय ध्वज एक मुद्दा बन गया। मीडिया, शिक्षा मंत्रालय और राष्ट्रीय प्रदर्शनी के अधिकारियों ने आर्टिफिशियल इंटेलिजेंस द्वारा बनाई गई मलेशियाई ध्वज की छवियों को जारी किया। या तो

अर्धचंद्र का प्रतीक गायब था या प्रतीक अजीब तरह से मुड़े हुए थे।

राष्ट्रीय ध्वज देश का चेहरा होता है। उस चेहरे को बनाते समय किसी ने उसकी जांच नहीं की।

नक्शे और इलाके में भी यही बात हुई।

नवंबर 2024 में, जापान के Fukuoka प्रांत ने आधिकारिक तौर पर माफी मांगी और एक विज्ञापन एजेंसी के साथ अपना अनुबंध समाप्त कर दिया। ऐसा इसलिए था क्योंकि एजेंसी द्वारा आर्टिफिशियल इंटेलिजेंस का उपयोग करके बनाई गई प्रचार सामग्री में पर्यटक आकर्षण और स्थानीय खाद्य पदार्थ जो Fukuoka में मौजूद नहीं थे, ऐसे शामिल किए गए थे जैसे कि वे वास्तविक हों।

मई 2025 में, Natural England की पीटलैंड मैपिंग परियोजना भी विफल हो गई। पीटलैंड कार्बन को रोके रखने वाले आर्द्रभूमि हैं, इसलिए वे जलवायु नीति में महत्वपूर्ण हैं। मशीन लर्निंग मॉडल द्वारा बनाए गए नक्शे में, पत्थर की दीवारों, खदानों, मिट्टी के ढेरों और ग्रेनाइट की चट्टानों को पीटलैंड के रूप में चिह्नित किया गया था।

केवल ऊपर से ली गई तस्वीरों से गीली जमीन और सूखे पत्थरों के बीच अंतर करना मुश्किल है। लोग उस पर कदम रखकर देखते हैं। मॉडल के पास पैर नहीं होते हैं।

2021 में, Stanford और University of Washington के शोधकर्ताओं ने डीपफेक भूगोल नामक अवधारणा के बारे में चेतावनी दी थी। उन्होंने कहा कि उपग्रह चित्रों को मिलाकर ऐसे इलाके बनाए जा सकते हैं जो मौजूद ही नहीं हैं। सैन्य और कूटनीति में क्या होगा, यह मैं आपकी कल्पना पर छोड़ता हूं।

दिसंबर 2024 में, बेल्जियम के प्रसिद्ध फोटो जर्नलिस्ट Carl de Keyzer, जब यूक्रेन पर आक्रमण के बाद रूस नहीं जा सके, तो उन्होंने जनरेटिव आर्टिफिशियल इंटेलिजेंस का उपयोग करके रूसी परिदृश्यों और लोगों की तस्वीरें बनाईं और उन्हें एक फोटो बुक में प्रकाशित किया। फोटो जर्नलिज्म का आधार यह तथ्य है कि आप वहां थे। वह आधार अब गायब हो गया है।

रास्ता दिखाने में लोगों की जान गई है।

2024 में, भारत में तीन लोग Google Maps के निर्देशों का पालन करते हुए अपनी कार चलाते समय एक टूटे हुए पुल से नीचे गिर गए और सभी की मौत हो गई। पुल का निर्माण पूरा नहीं हुआ था और वह पानी में डूबा हुआ था।

नक्शे में रास्ता था। हकीकत में वह नहीं था।

अमेरिका में भी ऐसी ही एक दुर्घटना हुई थी। ढह चुके पुल की ओर निर्देशित एक ड्राइवर पानी में गिर गया और उसकी मौत हो गई। वही गलती दूसरे महाद्वीप पर दोहराई गई।

फरवरी 2026 में, यूनाइटेड किंगडम में, Google Maps के निर्देशों का पालन करने वाला एक Amazon डिलीवरी ट्रक Broomway में चला गया। यह एक दलदली रास्ता है जो उच्च ज्वार के दौरान समुद्र के पानी में डूब जाता है। ट्रक कीचड़ में फंस गया।

मई 2025 में, Google Maps ने एक अज्ञात त्रुटि के कारण जर्मनी भर में प्रमुख ऑटोबान को बंद के रूप में प्रदर्शित किया, जिससे यातायात सूचना नेटवर्क हिल गया।

जनवरी 2026 में, ऑस्ट्रेलिया के तस्मानिया में, वे पर्यटक दूरदराज के इलाके में फंस गए जो एक आर्टिफिशियल इंटेलिजेंस द्वारा लिखे गए यात्रा लेख पर विश्वास करके एक ऐसे गर्म पानी के झरने की तलाश में निकले थे जिसका अस्तित्व ही नहीं था।

आपात स्थिति में आर्टिफिशियल इंटेलिजेंस द्वारा दी गई गलत सूचना और भी अधिक खतरनाक होती है।

अप्रैल 2024 में, Grok ने यह हेडलाइन बनाई कि ईरान ने तेल अवीव पर मिसाइल दागी है। उसी महीने, उसने यह खबर भी फैलाई कि भारत के प्रधान मंत्री को सरकार से बाहर कर दिया गया है। मई 2025 में, यहां तक कि बेसबॉल खेल के सवालों या समुद्री डाकू की तरह बोलने के अनुरोधों पर भी, इसने स्वेच्छा से दक्षिण अफ्रीकी श्वेत नरसंहार की साजिश के सिद्धांत को शामिल कर दिया।

मार्च 2026 में, इसने हिल्सबोरो और हेसेल फुटबॉल त्रासदियों के पीड़ितों का अपमान करने वाले लेख बनाकर शोक संतप्त परिवारों को चोट पहुंचाई। दोनों घटनाएँ वास्तविक त्रासदियाँ हैं जहाँ सैकड़ों लोग कुचलकर मर गए थे।

सितंबर 2025 में, लंदन में अप्रवासी विरोधी विरोध प्रदर्शनों के दौरान, इसने गलत सूचना फैलाई कि पुलिस ने हिंसा के वीडियो में हेरफेर किया था।

ज्ञान को ही फिर से लिखने के प्रयास भी किए गए।

अक्टूबर 2025 में, Musk ने Wikipedia के मुकाबले Grokpedia नामक एक आर्टिफिशियल इंटेलिजेंस विश्वकोश लॉन्च किया। इसके लॉन्च होने के साथ ही, 900,000 दस्तावेज़ अपलोड किए गए। वे दस्तावेज़ असत्यापित संदेश बोर्ड पोस्ट और ब्लॉग से कॉपी किए गए थे, और मतिभ्रम से बने तथ्यों को प्रविष्टियों के रूप में सूचीबद्ध किया गया था। दक्षिण अफ्रीकी श्वेत नरसंहार की साजिश के सिद्धांत को तथ्य के रूप में प्रस्तुत किया गया, और एड्स का इतिहास विकृत कर दिया गया।

विश्वकोश नाम लोगों को आश्वस्त करता है। इसी आश्वासन को निशाना बनाया गया था।

फरवरी 2025 में, यह पता चला कि Grok 3 मॉडल को इसके अनुमान प्रक्रिया के दौरान उन आलोचनात्मक स्रोतों को ब्लॉक करने के लिए आंतरिक निर्देशों के साथ सेट किया गया था जो यह बताते थे कि Musk और Trump गलत सूचना फैलाते हैं। यह तय किया गया था कि क्या नहीं पढ़ना है, न कि क्या कहना है।

व्यक्तियों को होने वाला नुकसान तत्काल होता है।

अक्टूबर 2025 में, अमेरिकी सीनेटर Marsha Blackburn ने पाया कि Google के Gemma मॉडल ने यह कहानी गढ़ी कि वह 1987 के राज्य सीनेट चुनाव के दौरान एक यौन अपराध के मामले में शामिल थीं और इसने फर्जी समाचार लिंक भी जोड़े थे।

उसी परिवार के एक सिस्टम ने दिसंबर 2023 से

अध्याय 3

निजता का हनन और निगरानी समाज: बिना सहमति डेटा संग्रह की छाया

3.1

अनाधिकृत जैविक जानकारी संग्रह और सार्वजनिक·वाणिज्यिक निगरानी

शॉपिंग मॉल में रास्ता भटकने पर आप एक सूचना स्क्रीन के सामने खड़े होते हैं। आप नक्शा देखते हैं, दुकान का नाम खोजते हैं, और अपनी उंगली से स्क्रीन को कुछ बार दबाते हैं। इसमें कुछ सेकंड लगते हैं।

2018 में, कनाडा के बारह शॉपिंग मॉल में, उन कुछ सेकंडों के दौरान एक घटना हुई थी।

कनाडा की एक बड़ी रियल एस्टेट कंपनी, कैडिलैक फेयरव्यू ने सूचना कियोस्क के अंदर कैमरे छिपा रखे थे। स्क्रीन देखने वाले व्यक्ति का चेहरा रिकॉर्ड किया गया, और एक बायोमेट्रिक एल्गोरिदम ने उसकी उम्र और लिंग का निर्धारण किया।

इस तरह पचास लाख लोगों को रिकॉर्ड किया गया।

जब जांच शुरू हुई, तो कंपनी ने इस तरह स्पष्टीकरण दिया: हम छवियों को सहेज कर नहीं रखते, हम उन्हें तुरंत मिटा देते हैं।

पर्यवेक्षी प्राधिकरण का निर्णय अलग था। उनका मानना था कि छवियों को मिटाया जाए या नहीं, बिना सहमति के चेहरे को मापना ही अपने आप में कानून का उल्लंघन है।

यह अंतर महत्वपूर्ण है। लोग आमतौर पर इस बात की चिंता करते हैं कि उनकी तस्वीरें कहां सहेजी गई हैं। लेकिन बायोमेट्रिक जानकारी का मूल बात इसे सहेजना नहीं, बल्कि इसे मापना है। भले ही तस्वीर मिटा दी जाए, लेकिन यह रिकॉर्ड बना रहता है कि यह व्यक्ति तीस के दशक की महिला है और मंगलवार दोपहर को इस मंजिल पर थी।

चेहरा पासवर्ड से अलग है। अगर यह लीक हो जाता है, तो आप इसे बदल नहीं सकते।

अमेरिका की फार्मेसी चेन राइट एड इससे एक कदम आगे निकल गई। 2012 से 2020 तक, इसने देश भर के स्टोरों में फेशियल रिकग्निशन कैमरे लगाए। इसका बहाना चोरों को रोकना था।

समस्या यह थी कि कैमरे ज़्यादातर कहाँ लगाए गए थे। वे उन मोहल्लों में केंद्रित थे जहाँ कम आय वाले लोग, अश्वेत और हिस्पैनिक बड़ी संख्या में रहते थे।

और एल्गोरिदम अक्सर गलत होता था। जब अलार्म बजता, तो कर्मचारी ग्राहकों के पास जाते। उन्हें बाहर जाने के लिए कहते। पुलिस को बुला लेते।

सामान खरीदने आए व्यक्ति के साथ लोगों के सामने चोर जैसा व्यवहार किया गया। उस मौके पर अपनी बेगुनाही साबित करने का कोई तरीका नहीं है।

संघीय व्यापार आयोग ने राइट एड को पांच साल तक इस तकनीक का उपयोग करने से रोक दिया।

कई देशों में ऐसी ही घटनाएं हुईं।

स्पेन की सुपरमार्केट चेन मर्कडोना ने चालीस से अधिक स्टोरों में फेशियल रिकग्निशन पेश किया था और उस पर इक्यावन लाख पचास हजार यूरो का जुर्माना लगाया गया। कंपनी ने कहा कि उसने उन पूर्व अपराधियों को छांटने की कोशिश की जिन्हें प्रवेश से प्रतिबंधित किया गया था।

अधिकारियों का जवाब कुछ इस तरह था: कुछ लोगों को खोजने के लिए हर किसी के चेहरे को मापने की अनुमति नहीं है।

ऑस्ट्रेलिया में, बनिंग्स, केमार्ट और 7-इलेवन पकड़े गए थे। 7-इलेवन ने स्टोर टैबलेट के माध्यम से संतुष्टि सर्वेक्षण करते हुए उत्तरदाताओं के चेहरों को गुप्त रूप से रिकॉर्ड किया और सहेजा।

उन्होंने यह पूछते हुए चेहरा रिकॉर्ड किया कि क्या आप संतुष्ट हैं।

यूनाइटेड किंगडम की किराना स्टोर श्रृंखलाओं ने भी फेशियल रिकग्निशन कैमरे पेश किए और निर्दोष ग्राहकों पर चोरी का आरोप लगाकर हंगामा खड़ा कर दिया।

चीन में, इस तकनीक का इस्तेमाल कमीशन की गणना के लिए किया गया था।

नानजिंग, निंगबो और नैनिंग में रियल एस्टेट डेवलपर्स ने बिक्री कार्यालयों में आने वाले ग्राहकों के चेहरे उनकी सहमति के बिना रिकॉर्ड किए। इसका उद्देश्य यह था: यह पता लगाना कि यह ग्राहक किस ब्रोकर के माध्यम से आया है, ताकि उन्हें अलग-अलग कमीशन दिया जा सके।

ग्राहक का चेहरा ब्रोकरेज कमीशन कैलकुलेटर का एक हिस्सा बन गया। ग्राहक को इसकी जानकारी नहीं थी।

इस कहानी का एक हंसाने वाला लेकिन दुखद परिणाम है। जैसे ही यह अफवाह फैली, ऐसे लोग सामने आए जो मोटरसाइकिल का हेलमेट पहनकर बिक्री कार्यालयों में जाते थे।

कोहलर, बीएमडब्ल्यू और मैक्स मारा जैसे ब्रांडों ने भी 2021 में सरकारी प्रसारक द्वारा पकड़े जाने तक ग्राहकों के दौरे और आवाजाही को ट्रैक करने के लिए चीनी स्टोरों में छिपे हुए कैमरे रखे थे।

मानवरहित स्टोर और भुगतान प्रणालियां भी चेहरे और हाथ का डेटा एकत्र करती हैं।

न्यूयॉर्क में अमेज़ॅन गो स्टोरों पर ग्राहकों की बायोमेट्रिक जानकारी एकत्र करते समय न्यूयॉर्क शहर के कानून द्वारा निर्धारित नोटिस दायित्वों का पालन न करने के लिए एक क्लास-एक्शन मुकदमा दायर किया गया था। अमेज़ॅन वन, जो आपको अपनी हथेली रखकर भुगतान करने की अनुमति देता है, हथेली की नस और हथेली की रेखा के डेटा को सर्वर पर सहेजता है।

मैकडॉनल्ड्स ने अपनी ड्राइव-थ्रू सेवा में वॉयस रिकग्निशन को शामिल किया और सहमति के बिना ड्राइवरों की आवाज़ की विशेषताओं को एकत्र करने और सहेजने के लिए इलिनोइस बायोमेट्रिक सूचना गोपनीयता अधिनियम के उल्लंघन के लिए उस पर मुकदमा दायर किया गया। डोमिनोज़ पिज्जा को भी टेलीफोन ऑर्डर से बिना अनुमति के आवाज़ एकत्र करने के आरोप में अदालत में ले जाया गया।

आपने केवल हैमबर्गर का ऑर्डर दिया था, लेकिन आपकी आवाज़ का फिंगरप्रिंट पीछे रह गया।

आइए अब स्कूलों की ओर चलते हैं।

स्वीडन के शेलेफ्टियो में एंडरस्टोर्प हाई स्कूल ने उपस्थिति की जांच करने में समय बचाने के लिए बाईस छात्रों पर फेशियल रिकग्निशन कैमरों का परीक्षण किया। 2019 में, स्वीडिश डेटा संरक्षण प्राधिकरण ने व्यक्तिगत सूचना संरक्षण विनियमन का उल्लंघन करने के लिए उन पर जुर्माना लगाया। इस विनियमन के तहत स्वीडन में लगाया गया यह पहला जुर्माना था।

स्कूल ने कहा कि उसे माता-पिता की सहमति मिल गई है। अधिकारियों ने इसे इस तरह देखा: स्कूल और छात्रों के बीच शक्ति का अंतर है, जिससे सही मायने में स्वैच्छिक सहमति स्थापित करना मुश्किल हो जाता है।

एक छात्र के लिए शिक्षक के अनुरोध को मना करना मुश्किल है। वयस्कों के बीच भी ऐसा ही होता है।

और इसका उद्देश्य उपस्थिति की जांच करना था। उन्होंने बायोमेट्रिक जानकारी का उपयोग उस काम के लिए किया जो उपस्थिति रजिस्टर बुलाकर किया जा सकता था।

फ्रांस के नीस और मार्सिले के हाई स्कूलों ने प्रवेश द्वारों पर फेशियल रिकग्निशन गेट लगाने की कोशिश की, लेकिन नागरिक समूहों के विरोध और अदालती अस्वीकृति के कारण इसे रद्द कर दिया।

यूके के नॉर्थ आयरशायर के नौ हाई स्कूलों ने स्कूल लंच के भुगतान को गति देने के लिए फेशियल रिकग्निशन पेश किया, लेकिन सूचना आयुक्त कार्यालय से चेतावनी प्राप्त करने के बाद इसे बंद कर दिया। चेल्मर वैली हाई स्कूल भी इसी कारण से पकड़ा गया था।

पोलैंड के डांस्क के एक प्राथमिक विद्यालय पर भोजन की जांच के लिए बच्चों के फिंगरप्रिंट एकत्र करने पर जुर्माना लगाया गया था।

उन्होंने खाने की कतार को कुछ सेकंड कम करने के लिए बच्चों के फिंगरप्रिंट लिए।

जब भारत सरकार ने रिपोर्ट कार्ड जारी करने और भोजन की जांच करने के लिए फेशियल रिकग्निशन को शामिल करने की कोशिश की, तब भी विवाद हुआ था। स्कॉटलैंड, ब्राजील के पराना राज्य और ऑस्ट्रेलिया के विक्टोरिया राज्य में भी इसी तरह के प्रयास और विरोध हुए थे।

अमेरिका में सुरक्षा शब्द का इस्तेमाल एक कुंजी के रूप में किया गया था। न्यूयॉर्क राज्य के लॉकपोर्ट स्कूल डिस्ट्रिक्ट ने बंदूक की घटनाओं को रोकने के लिए स्कूलों में फेशियल रिकग्निशन सर्विलांस सिस्टम स्थापित करने पर लाखों डॉलर खर्च किए। माता-पिता और नागरिक समूहों ने विरोध किया, और न्यूयॉर्क राज्य शिक्षा विभाग ने राज्य के सभी स्कूलों में इस तकनीक के उपयोग को अस्थायी रूप से निलंबित करने की कार्रवाई की।

कक्षाओं के अंदर जाने वाले कैमरों ने केवल चेहरों को ही नहीं देखा।

चीन के 11वें मिडिल स्कूल और चाइना फार्मास्युटिकल यूनिवर्सिटी ने कक्षाओं के सामने कैमरे लगाए और वास्तविक समय में छात्रों के चेहरे के भावों, उनींदापन और एकाग्रता का विश्लेषण किया और उन्हें स्कोर दिया।

कल्पना कीजिए कि एक ऐसी कक्षा हो जहाँ कक्षा के दौरान भटकने वाले विचारों वाले चेहरों के भाव रिकॉर्ड किए जाते हैं। उस कक्षा में, छात्र सीखने वाला व्यक्ति नहीं बल्कि मूल्यांकन किया जाने वाला डेटा बन जाता है।

ऊबने वाला चेहरा बनाने की स्वतंत्रता सीखने का एक हिस्सा है।

हवाई अड्डों और सीमाओं पर, इसका पैमाना बड़ा हो जाता है।

कनाडा सीमा सेवा एजेंसी ने टोरंटो पीयरसन अंतर्राष्ट्रीय हवाई अड्डे पर मानवरहित कियोस्क का परीक्षण किया, जो यात्रियों के चेहरों को उनकी जानकारी के बिना एकत्र करता था। न्यूजीलैंड के वेलिंगटन अंतर्राष्ट्रीय हवाई अड्डे पर भी सहमति के बिना कैमरे चलाने की बात मीडिया रिपोर्टों के माध्यम से सामने आई थी। स्पेनिश राज्य के स्वामित्व वाले हवाई अड्डा संचालक एना पर उपयोगकर्ताओं के चेहरों को अवैध रूप से एकत्र करने के लिए जुर्माना लगाया गया था।

Hanguk में भी ऐसी घटनाएं हुई हैं।

न्याय मंत्रालय और विज्ञान तथा सूचना प्रौद्योगिकी मंत्रालय 2019 से कृत्रिम बुद्धिमत्ता पहचान और ट्रैकिंग प्रणाली परियोजना चला रहे हैं। इसके तहत इंचियोन हवाई अड्डे जैसी जगहों से गुजरने वाले लोगों की चेहरे की तस्वीरें उनकी इजाजत के बिना निजी कृत्रिम बुद्धिमत्ता कंपनियों को सिखाने के लिए दे दी गईं।

संख्या 170 मिलियन थी।

यह केवल विदेशियों की तस्वीरें नहीं थीं। दक्षिण कोरिया के नागरिकों की चेहरे की तस्वीरें भी शामिल थीं। जांच में यह पुष्टि हुई कि कानूनी आधार के बिना ये निजी कंपनियों को दी गईं।

यदि आप पासपोर्ट जांच काउंटर पर कैमरा देखने को याद करते हैं, तो यह सोचने लायक है कि वह तस्वीर कहाँ गई।

अमेरिकी सीमा प्राधिकारियों ने शरणार्थी आवेदकों के लिए एक मोबाइल ऐप में चेहरे की पहचान लगाई। यह ऐप काले आवेदकों के चेहरों को अच्छी तरह पहचान नहीं पाता था। अगर पहचान नहीं होती, तो आवेदन आगे नहीं बढ़ता। जो लोग अपने घर से भाग कर आए हैं, वे ऐप के सामने एक बार फिर से रुक जाते हैं।

सड़कों पर आइरिस एकत्र करने का एक व्यवसाय था।

सैम ऑल्टमैन द्वारा स्थापित वर्ल्डकॉइन ने विश्व के कई शहरों में ऑर्ब नामक आइरिस पहचान उपकरण लगाए। आप अपनी आइरिस को पंजीकृत करते हैं, तो आपको क्रिप्टोकरेंसी दी जाती है। गरीब देशों और युवाओं में लंबी कतारें लगीं।

केन्या की सरकार ने व्यवसाय को पूरी तरह बंद कर दिया। पुर्तगाल, स्पेन, इंडोनेशिया और थाईलैंड के अधिकारियों ने भी व्यवसाय बंद करने का आदेश दिया। हांगकांग, अर्जेंटीना और दक्षिण कोरिया की व्यक्तिगत सूचना सुरक्षा समिति ने भी जांच शुरू की।

एक बार आइरिस को पंजीकृत करने के बाद, इसे वापस नहीं लिया जा सकता। प्राप्त सिक्कों की कीमत उतार-चढ़ाव करती है। यह सोचने लायक है कि क्या यह लेनदेन न्यायसंगत था।

एक ऐसी चीज भी निकली है जो एक जोड़ी चश्मे से किसी की पहचान बता सकती है।

हार्वर्ड विश्वविद्यालय के दो छात्रों ने मेटा के Ray-Ban स्मार्ट चश्मे में चेहरे की पहचान को जोड़ने वाली एक प्रणाली का खुलासा किया। इन चश्मों को पहन कर जब आप सड़क या मेट्रो में किसी अजनबी को देखते हैं, तो कुछ ही सेकंड में उस व्यक्ति का नाम, पता, फोन नंबर और पारिवारिक संबंध आपके फोन की स्क्रीन पर दिखाई देते हैं।

छात्रों ने कहा कि वे खतरे की चेतावनी देने के लिए इसे बनाया था। यह तथ्य कि इसे बनाया जा सकता है, यह ही एक चेतावनी है।

मेटा के स्मार्ट चश्मे पहनने वालों ने मालिश पार्लर या सार्वजनिक जगहों पर लोगों को चुपके से फिल्माया और वीडियो फैलाए हैं। यह भी पता चला कि इन चश्मों से एकत्र किए गए वीडियो विदेशी बाहरी कर्मियों को दे दिए गए और प्रशिक्षण के लिए इस्तेमाल किए गए।

इन सभी उपकरणों के पीछे एक कंपनी है जो पूरे इंटरनेट को खंगालती है।

Clearview AI ने Facebook, Instagram, X, LinkedIn और इंटरनेट भर से 30 अरब चेहरे की तस्वीरें एकत्र कीं और एक खोज डेटाबेस बनाया। और इसे दुनिया भर की पुलिस, जांच एजेंसियों और यहां तक कि खुदरा व्यापारियों को बेचा।

अमेरिकी सिविल लिबर्टीज यूनियन ने इलिनॉय के नागरिकों की ओर से मुकदमा दायर किया, और ब्रिटेन, फ्रांस, इटली, नीदरलैंड और ग्रीस की व्यक्तिगत जानकारी सुरक्षा एजेंसियों ने भारी जुर्माना और डेटा नष्ट करने के आदेश दिए।

चेहरे की खोज का इंजन PimEyes और भी बुरा था। इसने निजी तस्वीरें, मृत लोगों की तस्वीरें और यहां तक कि बाल यौन शोषण सामग्री को एकत्र कर एक ऐसी सेवा बनाई जो एक ही तस्वीर से किसी व्यक्ति की निजी तस्वीरें खोज सकती थी। जर्मन और अमेरिकी अदालतों ने इसे दंडित किया।

मेटा ने टेक्सास आदि राज्यों में निवासियों की इजाजत के बिना चेहरे की बायोमेट्रिक जानकारी एकत्र करने के आरोप में 1.4 अरब डॉलर का समझौता राशि दिया।

जब राष्ट्रीय स्तर पर देखा जाए, तो पैमाना अलग होता है।

इजराइली रक्षा बल हेब्रॉन और गाजा पट्टी में फिलिस्तीनी निवासियों की निगरानी करने के लिए चेहरे की पहचान प्रणाली का संचालन करते थे। मॉस्को ने मेट्रो भुगतान और विरोधी सरकार के आंकड़ों की ट्रैकिंग के लिए चेहरे की निगरानी नेटवर्क का उपयोग किया। लंदन के King's Cross स्टेशन और जांच एजेंसियों ने भी वास्तविक समय चेहरे की पहचान को तैनात किया।

जब हम इन घटनाओं को फिर से देखते हैं, तो कुछ स्थान ध्यान में आते हैं। शॉपिंग मॉल, दवा की दुकानें, सुपरमार्केट, सुविधा स्टोर, कार डीलरशिप, विक्रय कार्यालय, प्राथमिक स्कूल, माध्यमिक स्कूल, विश्वविद्यालय, हवाई अड्डे, सीमा, मेट्रो स्टेशन, सड़कें।

यह एक व्यक्ति के दैनिक जीवन का सटीक मार्ग है।

और सभी जगहों पर एक ही बात दोहराई गई। यह सुरक्षा के लिए है। यह सुविधा के लिए है। यह दक्षता के लिए है।

इन बातों का विरोध करना मुश्किल है। इसीलिए ये अच्छी तरह बिकती हैं।

चेहरा, आइरिस, फिंगरप्रिंट और आवाज आपके शरीर से जुड़ी हैं। भले ही आप उन्हें खो दें, आप नए नहीं पा सकते। जो लोग इस तथ्य को जानते हैं वे इसे एकत्र कर रहे हैं।

3.2

व्यक्तिगत जानकारी का अनधिकृत प्रशिक्षण और AI बातचीत के लीक होने की घटना

2023 के मार्च के किसी दिन, ChatGPT खोलने वाले लोगों की स्क्रीन के बाएं ओर बातचीत की सूची दिखाई दी। हमेशा की जगह थी।

लेकिन शीर्षक अपरिचित थे। ये ऐसी बातचीत थीं जो उन्होंने नहीं की थीं।

दूसरों की बातचीत के शीर्षक मेरी स्क्रीन पर थे। मेरी बातचीत के शीर्षक भी किसी की स्क्रीन पर रहे होंगे।

OpenAI ने एक ओपन सोर्स डेटाबेस लाइब्रेरी के बग के कारण बताया और सेवा को थोड़े समय के लिए बंद कर दिया। जांच के परिणाम में ChatGPT Plus सदस्यों के लगभग 1.2 प्रतिशत का नाम, ईमेल पता, क्रेडिट कार्ड नंबर के अंतिम चार अंक और समाप्ति तारीख अन्य उपयोगकर्ताओं को दिखे।

अगर लोग सोचें कि वे उस विंडो में क्या लिखते हैं तो इस घटना का वजन बदल जाता है।

जब वे अपना रिज्यूमे ठीक करने के लिए कहते हैं तो पता और फोन नंबर डालते हैं। जब वे चिकित्सा रिपोर्ट समझना चाहते हैं तो बीमारी का नाम लिखते हैं। जब वे तलाक के बारे में सोचते हैं तो पति या पत्नी के बारे में लिखते हैं। कंपनी को न बताई गई चिंताएं लिखते हैं।

सर्च इंजन में वे सवाल डालते हैं। चैटबॉट में वे परिस्थितियां डालते हैं।

लोगों को इस अंतर का एहसास होने से पहले अगली घटना आई।

ChatGPT में बातचीत को लिंक के रूप में साझा करने की सुविधा थी। दोस्तों को दिखाने के लिए बनाई गई सुविधा। लेकिन यह लिंक सार्वजनिक वेबपेज जैसे व्यवहार किया गया।

2025 में उपयोगकर्ताओं द्वारा बनाई गई बातचीत के एक लाख से अधिक रिकॉर्ड Google जैसे सर्च इंजन में इंडेक्स किए गए। सर्च में कुछ शब्द डालने से अजनबी की परामर्श का इतिहास आया। निजी चिंताएं आईं। कंपनी के रहस्य आए।

एक दोस्त को दिखाने के लिए बनाई गई चीज़ पूरी दुनिया को दिखाई दे गई।

Google के Bard ने भी वही गलती की। Grok ने भी की। Grok के साथ की गई सैकड़ों हजारों बातचीत सर्च परिणामों में उजागर हुईं और बिना अनुरोध के Grok ने वयस्क सामग्री उद्योग के एक कार्यकर्ता का वास्तविक नाम और जन्मतिथि प्रकट किया था।

वही गलती कंपनियों के बीच दोहराई गई। नई सुविधा जल्दी बाहर निकालने की प्रतिस्पर्धा में साझा किए गए लिंक की पहुंच अनुमतियों पर ध्यान देना पीछे छोड़ दिया जाता है।

भावनात्मक बातचीत को संभालने वाले ऐप्स में रिसाव अधिक दर्दनाक है।

2024 में Character AI में आंतरिक त्रुटि से उपयोगकर्ता एक दूसरे की निजी बातचीत देख सके, यह घटना हुई। आभासी पात्रों को कही गई स्वीकारोक्तियां थीं।

चीन के DeepSeek में डेटाबेस सुरक्षा प्रबंधन की लापरवाही से खाता जानकारी और बातचीत का इतिहास बाहर आ गया। तीन सौ मिलियन निजी संदेशों की मात्रा है।

2026 के फरवरी में भी समान आकार की एक घटना हुई। एक कृत्रिम बुद्धिमत्ता चैट ऐप में 25 मिलियन उपयोगकर्ताओं के 300 मिलियन संदेश उजागर हुए। उपयोगकर्ता द्वारा फाइलों में पूरी बातचीत के रिकॉर्ड थे।

कृत्रिम बुद्धिमत्ता साथी ऐप्स और गहरी चीजें जानते हैं। Replika, Muah, Nomi जैसे ऐप्स भावनात्मक संबंध को अग्रभाग में रखकर उपयोगकर्ता की यौन प्राथमिकताओं और भावनात्मक डेटा एकत्र किया। Mozilla Foundation के सर्वेक्षण के अनुसार ये ऐप्स न्यूनतम गोपनीयता मानकों का पालन नहीं किया और एकत्र किए गए डेटा को मार्केटिंग कंपनियों को पास कर दिया।

क्रमागत रिसावों से चार लाख से अधिक उपयोगकर्ताओं की बातचीत बाहर आई। जब एक वयस्क चैटबॉट सेवा की सुरक्षा टूट गई तो महिलाओं की वर्षपुस्तक की तस्वीरों के दो लाख और सिंथेटिक अश्लील सामग्री डेटा भी उजागर हुए।

वर्षपुस्तक की तस्वीरें उपयोगकर्ता द्वारा उस सेवा में अपलोड नहीं की गई थीं। किसी ने स्क्रेप की थीं।

कंपनी के रहस्य भी उसी विंडो से निकल जाते हैं।

2023 की बसंत में Samsung के कार्यस्थल में एक घटना हुई। सेमीकंडक्टर विभाग के कर्मचारी काम जल्दी पूरा करने के लिए ChatGPT का उपयोग किया। उपकरण माप डेटा डाला। उपज से संबंधित स्रोत कोड डाला। पूरी बैठक की नोट्स डालीं और सारांश देने के लिए कहा।

वह डेटा OpenAI के सर्वर पर चला गया। कंपनी ने आंतरिक कृत्रिम बुद्धिमत्ता के उपयोग को तुरंत प्रतिबंधित किया।

कर्मचारियों को दोष देना आसान है। लेकिन इस घटना का सार कहीं और है। चैटबॉट विंडो सर्च विंडो जैसी दिखती है। सर्च में लोग कंपनी के दस्तावेज़ नहीं डालते। लेकिन चैटबॉट में डाल देते हैं। क्योंकि वह सारांश दे देता है।

उपकरण की दिखावट लोगों के व्यवहार को बदलती है।

Otter AI वॉयस सारांश सेवा निजी निवेशक टेलीफोन सम्मेलन सहमति के बिना रिकॉर्ड किया और सारांश को गलत जगह भेज वित्तीय जानकारी और निवेश रणनीति लीक की।

बड़ी कंपनियां भी अपने डेटा लीक करती हैं।

2023 के सितंबर में Microsoft कृत्रिम बुद्धिमत्ता अनुसंधान दल GitHub पर डेटासेट साझा करते समय पहुंच अनुमति सेटिंग्स गलत की। संपूर्ण फाइल सिस्टम को पढ़ने और लिखने के लिए टोकन सार्वजनिक फाइल में मिश्रित हो गए।

परिणामस्वरूप उजागर डेटा 38 टेराबाइट है। कर्मचारियों के पासवर्ड, गुप्त कुंजियां, आंतरिक संदेशवाहक बातचीत, 35000 से अधिक कर्मचारियों के निजी कार्य डेटा थे।

कृत्रिम बुद्धिमत्ता अनुसंधान करने वाली टीम कृत्रिम बुद्धिमत्ता प्रशिक्षण के लिए डेटा साझा करते समय कंपनी के आंतरिक को खोल दिया।

Amazon का Q बिज़नेस कृत्रिम बुद्धिमत्ता मतिभ्रम के साथ निजी ग्राहक डेटा लीक करने की खामी दिखाई। सलाहकार समूह McKinsey के आंतरिक कृत्रिम बुद्धिमत्ता प्लेटफॉर्म हैकर्स का लक्ष्य बन गया और दो घंटों में लाखों ग्राहक रिकॉर्ड और गुप्त परियोजना सामग्री को सुलभ बनाया गया।

Meta के कृत्रिम बुद्धिमत्ता एजेंट ने भी आंतरिक सामग्री और उपयोगकर्ता जानकारी लीक की। एक कृत्रिम बुद्धिमत्ता सहायक कार्यक्रम हजारों सर्वर इंटरनेट पर पूरी तरह खुले थे, मैलवेयर हमले का सामना किया और उपयोगकर्ता डेटा की बड़ी मात्रा को चोरी किया गया।

2026 के फरवरी में सुरक्षा अनुसंधान से एक नई विधि का खुलासा हुआ। एक दुर्भावनापूर्ण रूप से बनाया गया प्रॉम्प्ट एक सामान्य ChatGPT बातचीत को डेटा लीक चैनल में बदल सकता है, यह है। उपयोगकर्ता को कुछ भी पता नहीं है। कमजोरी को उसी महीने ठीक किया गया।

यहां तक यह दुर्घटनाएं हैं। अगले भाग दुर्घटनाएं नहीं हैं।

2023 के अगस्त में वीडियो कॉन्फ्रेंसिंग प्लेटफॉर्म Zoom ने सेवा की शर्तों को चुपचाप बदल दिया। उपयोगकर्ताओं द्वारा साझा की गई वीडियो, ऑडियो, बातचीत के रिकॉर्ड, साझा की गई फाइलें कृत्रिम बुद्धिमत्ता प्रशिक्षण के लिए उपयोग करने का अधिकार कंपनी को देने के बारे में था।

घर से काम के समय लोगों ने Zoom पर क्या किया यह याद करें। कंपनी की बैठकें कीं, कक्षाएं लीं, अस्पताल परामर्श प्राप्त किया, अंतिम संस्कार में भाग लिया।

आलोचना बढ़ने पर Zoom ने सहमति के बिना उपयोग नहीं करने का कहकर पीछे हट गया।

Slack को 2024 में पकड़ा गया। उपयोगकर्ताओं द्वारा आदान-प्रदान किए गए संदेशों और अपलोड की गई निजी फाइलों को कृत्रिम बुद्धिमत्ता प्रशिक्षण में डिफ़ॉल्ट रूप से उपयोग कर रहा था। इनकार करने के लिए उपयोगकर्ता को अलग से आवेदन करना पड़ता था और इस तथ्य को ठीक से सूचित नहीं किया।

डिफ़ॉल्ट सेटिंग क्या है यह सब कुछ तय करता है। अधिकांश लोग सेटिंग्स को नहीं बदलते। कंपनियां इस तथ्य को जानती हैं।

एडोबी ने अपनी क्लाउड सेवा की शर्तों में बदलाव किया, जिससे रचनाकारों के काम वाले निजी प्रोजेक्ट्स और तस्वीरों का विश्लेषण उनके खुद के आर्टिफिशियल इंटेलिजेंस को सिखाने के लिए किया जा सके, जिसका दुनिया भर के डिजाइनरों ने सब्सक्रिप्शन रद्द करने के आंदोलन से विरोध किया।

लिंक्डइन ने बिना पूर्व सहमति के प्रोफाइल, कार्य अनुभव और पोस्ट का उपयोग आर्टिफिशियल इंटेलिजेंस को सिखाने के लिए किया। एक्स ने अपने सिस्टम को इस तरह से सेट किया कि सभी पोस्ट स्वतः ही ग्रोक को सिखाने के लिए जमा हो जाएं, जिसके लिए उसे यूरोपीय अधिकारियों की जांच का सामना करना पड़ा। साउंडक्लाउड ने संगीतकारों के गानों का उपयोग सिखाने के लिए किया।

मेटा ने घोषणा की कि वह फेसबुक और इंस्टाग्राम उपयोगकर्ताओं द्वारा दशकों से डाली गई पोस्ट और तस्वीरों का उपयोग अपने भाषा मॉडल को सिखाने के लिए करेगा, जिसके कारण उसे यूरोप और दक्षिण अमेरिका की गोपनीयता संस्थाओं से प्रतिबंधों का सामना करना पड़ा।

दस साल पहले डाली गई बच्चे की तस्वीर, पंद्रह साल पहले लिखी गई कोई शर्मनाक पोस्ट, और दुनिया से जा चुके परिवार के सदस्यों के साथ ली गई तस्वीरें - ये सभी चीजें सिखाने की सामग्री की सूची में शामिल हैं। जब इन्हें अपलोड किया गया था, तो यह केवल दोस्तों को दिखाने के लिए था।

यह भी सामने आया कि मेटा ने रे-बैन स्मार्ट ग्लासेस से ली गई निजी तस्वीरें और वीडियो विदेशों में काम करने वाले आउटसोर्स डेटा कर्मचारियों को सौंपे, ताकि उनका उपयोग सीखने के लिए किया जा सके। किसी देश के किसी कार्यालय में बैठा एक व्यक्ति हमारे लिविंग रूम के वीडियो पर नाम का टैग लगा रहा था।

मई 2026 में, कैलिफोर्निया संघीय न्यायालय में एक सामूहिक मुकदमा दर्ज किया गया था। इसमें कहा गया था कि चैटजीपीटी वेबसाइट में मेटा पिक्सेल और गूगल एनालिटिक्स जैसी ट्रैकिंग तकनीकें डाली गई थीं, जिससे बातचीत से जुड़ी जानकारी, उपयोगकर्ता आईडी और ईमेल पते गूगल और मेटा को भेजे गए।

दावा यह है कि लोगों को लगा कि वे एक ही से बात कर रहे हैं, लेकिन वास्तव में बात सुनने वाले तीन थे।

चिकित्सा रिकॉर्ड के मामले में, इसका पैमाना बिल्कुल अलग है।

2017 में, ब्रिटेन के डेटा संरक्षण कार्यालय ने निर्णय दिया कि राष्ट्रीय स्वास्थ्य सेवा के अधीन रॉयल फ्री अस्पताल और गूगल डीपमाइंड के बीच डेटा साझा करने का समझौता कानून का उल्लंघन था। अस्पताल ने एक्यूट किडनी इंजरी का पता लगाने वाला ऐप बनाने के नाम पर, बिना अनुमति के सोलह लाख मरीजों के पूरे चिकित्सा रिकॉर्ड सौंप दिए थे।

इसके बाद गूगल ने अमेरिका की सबसे बड़ी अस्पताल श्रृंखलाओं में से एक के साथ समझौता किया और एआई को सिखाने के लिए चिकित्सा रिकॉर्ड प्राप्त किए। उसे पांच करोड़ लोगों का चिकित्सा डेटा मरीजों की जानकारी के बिना गुप्त रूप से इकट्ठा करने वाले इस काम के लिए आलोचना का सामना करना पड़ा।

दिल को सबसे ज्यादा भारी करने वाला मामला आखिरी है।

अमेरिका की क्राइसिस टेक्स्ट लाइन आत्महत्या रोकथाम और मानसिक स्वास्थ्य संकट से जुड़ा परामर्श देने वाली एक गैर-लाभकारी संस्था है। जब किसी को मरने का मन करे और वह संदेश भेजे, तो सलाहकार उसका उत्तर देते हैं। यह एक ऐसी जगह है जहाँ मदद मांगने वाले लोग अपने जीवन के सबसे बुरे समय की बातें लिखते हैं।

इस संस्था ने उन बातचीत के रिकॉर्ड्स को इकट्ठा किया और अपनी एक लाभ कमाने वाली आर्टिफिशियल इंटेलिजेंस सहायक कंपनी को सौंप दिया। उस कंपनी ने इस डेटा का उपयोग कंपनियों के ग्राहक सेवा चैटबॉट को प्रशिक्षित करने के लिए किया।

"मैं मरना चाहता हूँ" वाला वाक्य ग्राहकों की सेवा की गुणवत्ता को बेहतर बनाने की सामग्री बन गया।

मासिक धर्म चक्र और गर्भधारण की संभावना को रिकॉर्ड करने वाले स्वास्थ्य ऐप फ्लो पर उपयोगकर्ता के स्वास्थ्य डेटा को बिना अनुमति के फेसबुक और गूगल जैसी जगहों पर भेजने के कारण संघीय व्यापार आयोग ने प्रतिबंध लगाए।

यदि आप इन घटनाओं को एक साथ रखकर देखें, तो यह सच्चाई साफ दिखाई देती है कि डेटा केवल एक ही दिशा में बह रहा है।

यह लोगों से कंपनी की ओर जाता है। कंपनी से लोगों के पास जो लौटकर आता है, वह एक सुविधाजनक सुविधा है। सुविधा वास्तविक है। इसीलिए यह लेन-देन हो पाता है।

समस्या यह है कि इस पर कोई कीमत नहीं लिखी है। अनुबंध में यह नहीं लिखा है कि हम क्या दे रहे हैं और क्या ले रहे हैं। और अगर यह लिखा भी है, तो अठारह पन्नों की शर्तों के बारहवें पन्ने पर होता है।

और एक बार जो डेटा चला गया, वह वापस नहीं आता। भले ही कंपनी कहे कि उन्होंने इसे मिटा दिया है, लेकिन उस डेटा से सीखे हुए मॉडल के अंदर से वह नहीं मिटता है।

मॉडल यह भूलना नहीं जानता कि उसने क्या सीखा है।

3.3

स्थान ट्रैकिंग और श्रमिकों की अत्यधिक निगरानी

Amazon ने एक पेटेंट दाखिल किया। यह कलाई पर पहनने वाली एक पट्टी थी।

जब भंडार में काम करने वाले लोग यह पट्टी पहनते थे, तो यह वास्तविक समय में ट्रैक करता था कि उनका हाथ किस शेल्फ और किस बॉक्स की ओर जा रहा है। अगर हाथ गलत जगह जाता था, तो पट्टी कंपन करती थी।

यह धीरे से सूचित कर रहा था। वह जगह नहीं है, यह जगह है।

मैं अकेला नहीं था जो कुत्ते के प्रशिक्षण के लिए कॉलर की याद दिलाता था।

इस उपकरण के डिजाइन में लोगों के बारे में एक विशेष दृष्टिकोण निहित था। श्रमिक सिस्टम का एक अंतिम उपकरण था, हाथ उस उपकरण के पंजे थे, और कंपन त्रुटि को कम करने के लिए प्रतिक्रिया था।

Amazon की फ्रांसीसी शाखा को वास्तव में जुर्माना दिया गया था। यह 32 मिलियन यूरो था।

कारण यह था। जब गोदाम श्रमिक बारकोड स्कैनर लेकर काम कर रहे थे, तो कंपनी एक स्कैन और अगले स्कैन के बीच के समय को सेकंड में माप रही थी। अगर दस मिनट से अधिक समय के लिए कोई स्कैन नहीं था, तो चेतावनी जमा होती जाती थी।

दस मिनट है। अगर आप बाथरूम जाते हैं और वापस आते हैं, तो दस मिनट चला जाता है। अगर आप पानी पीते हैं और सांस लेते हैं, तो दस मिनट चला जाता है।

फ्रांस के डेटा संरक्षण प्राधिकरण ने इस तरह की निगरानी को अत्यधिक माना।

श्रमिकों को मापने से पहले, कंपनियों ने पहले स्थान बेचना सीखा।

2021 में Life360 नामक एक कंपनी विवाद में आई। यह एक ऐप था जो परिवारों को एक-दूसरे के स्थान की जांच करने देता था। माता-पिता यह देखने के लिए इसे स्थापित करते थे कि क्या बच्चा स्कूल सुरक्षित रूप से पहुंच गया था।

यह कंपनी लाखों उपयोगकर्ताओं के आंदोलन पथ और उनके द्वारा देखी गई जगहों के रिकॉर्ड को डेटा दलालों को बेचती आ रही थी।

स्थान का रिकॉर्ड अन्य जानकारी से अलग है। अगर आप जानते हैं कि कोई कहां गया, तो आप उस व्यक्ति के बारे में लगभग सब कुछ जान जाते हैं। हर गुरुवार की शाम को वह किस इमारत में जाता है, किस अस्पताल में जाता है, किस धार्मिक स्थान में जाता है, किस रात को घर के अलावा कहीं था।

नाम हटाने से कोई मदद नहीं है। जहां कोई हर रात रहता है वह उसका घर है, जहां कोई दिन में रहता है वह उसका काम का स्थान है। बस दो बिंदु होने से ही एक व्यक्ति की पहचान हो सकती है।

स्थान डेटा दलाल Tamoco को नॉर्वे की नागरिकों के GPS रिकॉर्ड को इकट्ठा करते हुए जांच एजेंसियों और सैन्य ट्रैकिंग के लिए वितरित करते हुए पकड़ा गया था। ब्रिटिश कंपनी Hook ने भी ऐप के माध्यम से एकत्रित स्थान डेटा को तीसरे पक्ष को बेचने के लिए प्रतिबंध का सामना किया।

चीन के डिलीवरी प्लेटफॉर्म Meituan ने डिलीवरी वर्कर्स और उपयोगकर्ताओं को अत्यधिक वास्तविक समय में ट्रैक किया जो प्रतिरोध का कारण बना।

अमेरिकी जांच अधिकारियों ने स्वचालित लाइसेंस प्लेट पहचान कैमरा सिस्टम के साथ वारंट के बिना विरोध प्रदर्शन में भाग लेने वाले और कार्यकर्ताओं के वाहनों के आंदोलन मार्ग को ट्रैक किया और निगरानी नेटवर्क में रखा।

लाइसेंस प्लेट को छिपाया नहीं जा सकता। ऐसा इसलिए है क्योंकि कानून ने इसे लगाने का आदेश दिया है।

डिलीवरी और परिवहन साइटों पर, कैमरे चेहरे देख रहे थे।

Amazon की डिलीवरी अनुबंध वाहनों में कृत्रिम बुद्धिमत्ता वीडियो निगरानी कैमरे और एक चालक मूल्यांकन कार्यक्रम था। यह देख रहा था कि दृष्टि कहां निर्देशित थी, क्या वह जम्हाई दे रहा था, उसके चेहरे की मांसपेशियां कैसे चल रही थीं, वह कितना कड़ा पैडल दबा रहा था और कितनी तेजी से ब्रेक लगा रहा था।

समस्या यह थी कि यह सिस्टम सामान्य ड्राइविंग को दंड दे रहा था।

अगर आप एक टहनी से बचने के लिए स्टीयरिंग व्हील घुमाते थे, तो इसे लापरवाही के लिए पकड़ा जाता था। अगर आप रियर व्यू मिरर को देखते थे, तो इसे ध्यान भटकाने के लिए पकड़ा जाता था। यह ड्राइविंग स्कूल में सिखाया जाता है।

अगर अंक काट दिए जाते, तो भत्ते में कटौती होती और डिलीवरी घटते। चालक को सुरक्षित ड्राइविंग और अंकों के बीच चुनना पड़ता।

Amazon Flex एल्गोरिदम सड़क निर्माण या खराब मौसम को गणना में नहीं रखता था और गणितीय रूप से सबसे छोटा मार्ग मांगता था। चालक खतरनाक रास्तों और कीचड़ में चले गए।

नक्शे पर लाइन छोटी है। नक्शे पर यह नहीं है कि उस लाइन पर क्या है।

कार्यालय भी एक सुरक्षित क्षेत्र नहीं था।

वित्तीय कंपनी Barclays ने अपने ब्रिटिश कार्यालय में कर्मचारी डेस्क के नीचे सेंसर लगाए और कार्य कंप्यूटर पर चुपचाप निगरानी प्रोग्राम स्थापित किए। उन्होंने सेकंड में बैठने के समय और कीबोर्ड टाइपिंग की आवृत्ति को मापा और अनुत्पादक समय को एकत्रित किया।

जब ट्रेड यूनियन ने विरोध किया, तो कंपनी ने संचालन को रोक दिया।

Microsoft ने प्रोडक्टिविटी स्कोर नामक एक फीचर लॉन्च किया। यह कितने ईमेल भेजे गए, मैसेंजर पर कितनी जल्दी जवाब दिया गया, और मीटिंग में कितनी भागीदारी की गई, यह देखकर व्यक्तिगत स्कोर को प्रबंधकों को दिखाता है।

इन मेट्रिक्स में एक आम विशेषता है। वे केवल वही चीजें गिनते हैं जो गिनना आसान है।

एक घंटे तक खिड़की से बाहर देखते हुए समस्या को हल करने वाले व्यक्ति और एक घंटे तक मैसेंजर में 23 बार जवाब देने वाले व्यक्ति में से कौन काम कर रहा था? स्कोर बाद वाले को चुनता है।

इसलिए लोग बाद वाला करने लगते हैं। स्कोर व्यवहार को बदलता है, और बदला हुआ व्यवहार फिर स्कोर बनाता है।

Meta भी अपने अमेरिकी कर्मचारियों के काम के कंप्यूटर पर गुप्त रूप से सॉफ्टवेयर स्थापित करते हुए माउस आंदोलन, क्लिक, कीबोर्ड इनपुट, और स्क्रीन कैप्चर को वास्तविक समय में एकत्रित करते हुए पकड़ा गया।

भावनाओं को भी मापने वाली प्रणालियां भी हैं।

ग्लोबल कॉल सेंटर कंपनी Teleperformance ने दूरस्थ कार्य श्रमिकों के वेबकैम को काम के घंटों के दौरान चालू रखने की एक प्रणाली शुरू की। कृत्रिम बुद्धिमत्ता चेहरे के कोण और आंख की स्थिति को ट्रैक करती थी ताकि यह पता चल सके कि क्या वह स्क्रीन से दूर देख रहा है, फोन देख रहा है, या कोई अन्य व्यक्ति कमरे में आया है, और कंपनी को सूचित करती है।

उसका अपना घर का ड्राइंग रूम काम के घंटों के दौरान रिकॉर्ड किया जा रहा है। अगर परिवार के सदस्य गुजरते हैं, तो एक चेतावनी आती है।

चीन में Canon कार्यालय में, दरवाजा खुलने के लिए कैमरे के सामने हंसना पड़ता था। मुस्कुराहट को पहचाना जाना चाहिए था ताकि उपस्थिति की पुष्टि की जा सके।

कंपनी ने कहा कि उद्देश्य एक उज्ज्वल वातावरण बनाना था।

आप सभी जानते हैं कि एक ऐसे दरवाजे के सामने मुस्कुराना क्या है जहां आपको हंसना पड़ता है।

लेखा फर्म PricewaterhouseCoopers ने एक उपकरण पेश किया जो चेहरे की पहचान का उपयोग करके वित्तीय व्यापारियों को ट्रैक करता था, यहां तक कि वह समय भी जब वे बाथरूम जाने के लिए अपनी सीट छोड़ते थे, और इसने विवाद का सामना किया।

स्पेन के निर्माता Plástico Forte को चेहरे की पहचान का उपयोग करके साइट कर्मचारियों की आने-जाने और गतिविधियों की निगरानी करने के लिए जुर्माना दिया गया। ब्रिटिश सेवा कंपनी Serco ने भी कर्मचारी निगरानी के लिए अवैध रूप से चेहरे की पहचान का उपयोग किया और संचालन बंद करने का आदेश दिया गया।

आवाज के टोन, हृदय गति, शरीर का तापमान, यहां तक कि त्वचा की विद्युत प्रतिक्रिया को मापने और भावनात्मक स्थिति को ट्रैक करने का प्रयास भी था। कुछ सिस्टम पूरे फैक्ट्री फ्लोर पर कैमरे लगाते थे ताकि वे श्रमिकों की गतिविधियों और हरकतों को देख सकें।

इस तरह से एकत्रित किए गए नंबर आखिरकार लोगों को निकाल देते हैं।

2021 में गेमिंग प्लेटफॉर्म कंपनी Xsolla ने निगरानी एल्गोरिदम विश्लेषण के परिणामों के आधार पर एक साथ 150 कर्मचारियों को बर्खास्त कर दिया।

मानदंड ये थे। आंतरिक प्रोग्राम उपयोग का इतिहास, कोड रिपॉजिटरी और दस्तावेज़ इनपुट की आवृत्ति, चैट रूम और दस्तावेज़ कार्य भागीदारी। इन मेट्रिक्स में कम आने वाले लोगों को निष्क्रिय कर्मचारी के रूप में वर्गीकृत किया गया।

कर्मचारियों को नहीं पता था कि उनका मूल्यांकन किस मानदंड पर किया गया था। वह नहीं जानते थे कि कौन सा कार्य संदर्भ गायब था। वह लंबे समय से काम करने वाली कंपनी से एक स्वचालित अधिसूचना पंक्ति में निकाले गए।

वह व्यक्ति जिसने फोन पर ग्राहक के साथ लंबे समय तक बात की, वह व्यक्ति जिसने अपने जूनियर को रोककर सिखाया, वह व्यक्ति जिसने मीटिंग रूम में बातचीत से समस्या को सुलझाया। सोचिए कि ये लोग किस कॉलम में दर्ज होते हैं। वे किसी भी कॉलम में दर्ज नहीं होते हैं।

एक कंपनी के टर्नओवर प्रेडिक्शन एल्गोरिदम ने कर्मचारियों के खोज इतिहास और प्लेटफॉर्म गतिविधियों को ट्रैक करके, उन लोगों का पहले से ही पता लगा लिया जिनके नौकरी छोड़ने की संभावना थी, और उन्हें नुकसान पहुँचाया।

अगर नौकरी छोड़ने की भविष्यवाणी के कारण बुरा व्यवहार किया जाता है, तो लोग सच में नौकरी छोड़ देते हैं। भविष्यवाणी सही साबित होती है।

प्लेटफॉर्म लेबर में, नौकरी से निकालने का काम भी स्वचालित होता है।

अमेज़ॅन फ्लेक्स के डिलीवरी ड्राइवरों को कभी किसी इंसान का चेहरा देखे बिना एक स्वचालित ईमेल के माध्यम से नौकरी से निकाल दिया गया। भले ही ऐप खराब हो गया हो, लॉजिस्टिक्स सेंटर से सामान देर से निकला हो, सड़क जाम हो गई हो, या भारी बारिश हो रही हो, यह सब डिलीवरी में देरी के पॉइंट के रूप में जमा हो जाता था। जब पॉइंट जमा हो जाते हैं, तो अकाउंट स्थायी रूप से निलंबित कर दिया जाता है।

जब उन्होंने आपत्ति जताई, तो उन्हें पहले से लिखा हुआ जवाब मिला।

इटली के बोलोग्ना कोर्ट द्वारा उठाई गई समस्या, डिलीवरू का फ्रैंक एल्गोरिदम पहले ही सामने आ चुका है। बीमार होने के कारण काम पर न जा पाने, परिवार में किसी की मृत्यु होने, या कानून द्वारा गारंटीकृत हड़ताल में भाग लेने पर भी पॉइंट काट लिए गए।

कानून हड़ताल करने का अधिकार देता है। एल्गोरिदम उस अधिकार का उपयोग करने पर भूखा मारता है। दोनों एक ही देश के अंदर एक साथ चल रहे थे।

ऐसे लोग भी हैं जिन्हें इसलिए नौकरी से निकाल दिया गया क्योंकि उनका चेहरा पहचाना नहीं जा सका।

यूके उबेर ईट्स के डिलीवरी ड्राइवर, पा एड्रिसा मंजांग और इमरान रजा को प्लेटफॉर्म के स्वचालित फेशियल रिकग्निशन सॉफ्टवेयर के कारण रातों-रात नौकरी से निकाल दिया गया।

सिस्टम गहरे रंग की त्वचा वाले डिलीवरी ड्राइवरों के चेहरों को अच्छी तरह से नहीं पहचान पाता था। जब यह पहचानने में विफल रहता है, तो यह उन्हें ऐसे धोखेबाज़ उपयोगकर्ताओं के रूप में आंकता है जिन्होंने अपना अकाउंट किसी और को उधार दिया है। अकाउंट तुरंत ब्लॉक कर दिया जाता है।

दोनों लोगों ने कई सालों तक अच्छे मूल्यांकन के साथ काम किया था। मशीन द्वारा उनका चेहरा न पहचान पाने की कीमत उन्हें खुद चुकानी पड़ी।

यूटा राज्य और न्यू साउथ वेल्स सहित कई जगहों पर, लिंग परिवर्तन सर्जरी या हार्मोन उपचार के बाद चेहरे बदलने वाले ट्रांसजेंडर ड्राइवरों के अकाउंट बड़ी संख्या में निलंबित कर दिए गए। सिस्टम ने इस तथ्य की गणना नहीं की कि इंसान बदलते हैं।

इंस्टाकार्ट और शिप्ट जैसे डिलीवरी प्लेटफॉर्मों ने भी अपारदर्शी एल्गोरिदम का उपयोग करके वेतन में कटौती की और श्रमिकों को स्वचालित रूप से निकाल दिया। यूनाइटेड स्टेट्स पोस्टल सर्विस के डिलीवरी ड्राइवर वेतन गणना

एल्गोरिदम ने भी बिना फॉर्मूले का खुलासा किए भत्ते में कटौती की।

सबसे कड़वा उदाहरण आखिरी वाला है।

2025 में, ऑस्ट्रेलिया के एक बड़े बैंक में, कर्मचारियों ने कई महीनों तक ग्राहक सेवा चैटबॉट को सिखाया। उन्होंने ग्राहकों के साथ व्यवहार करने का अपना अनुभव दर्ज किया, बताया कि कठिन परिस्थितियों में कैसे जवाब देना है, और चैटबॉट के गलत होने पर उसे ठीक किया।

जैसे ही चैटबॉट का प्रदर्शन बेहतर हुआ, कंपनी ने उन कर्मचारियों को सामूहिक छंटनी की सूचना दे दी।

उन्होंने अपने उत्तराधिकारी को अपने ही हाथों से सिखाया था।

आइए इस खंड में उल्लिखित उपकरणों को सूचीबद्ध करें। रिस्टबैंड, बारकोड स्कैनर टाइमर, वाहन कैमरा, डेस्क सेंसर, कीबोर्ड रिकॉर्डर, वेबकैम, मुस्कान पहचानने वाला दरवाजा, शौचालय ट्रैकिंग कैमरा, टर्नओवर प्रेडिक्शन मॉडल।

ये सभी इंसानों की ओर निर्देशित हैं। और सभी को एक ही नाम से पुकारा जाता है। उत्पादकता प्रबंधन।

कर्मचारी अपना स्कोर नहीं देख सकते। उन्हें यह भी नहीं पता कि इसकी गणना कैसे की जाती है। अगर आप पूछें, तो वे कहते हैं कि यह व्यापारिक रहस्य है।

एक बात तो पक्की है। ये सिस्टम यह पता लगाने में बहुत माहिर हैं कि लोग कब ब्रेक लेते हैं। क्योंकि ब्रेक के समय को गिनना आसान होता है।

जिन पलों में काम अच्छी तरह से हो रहा होता है, उन्हें गिनना मुश्किल होता है। इसलिए वे उसे नहीं गिनते हैं।

अध्याय 4

कॉपीराइट और रचनाकारों के अधिकार: जनरेटिव एआई और प्रशिक्षण डेटा का विवाद

4.1

कॉपीराइट उल्लंघन मुकदमे और अनधिकृत डेटा स्क्रैपिंग

सबसे पहले हम किताब खरीदते हैं। यह एक असली किताब होती है। हम कीमत चुकाकर इसे खरीदते हैं।

इसके बाद, बाइंडिंग को काटा जाता है। पिछले हिस्से को चाकू से काट देने पर किताब अलग-अलग पन्नों का बंडल बन जाती है। फिर उन पन्नों को एक ऑटोमैटिक फीड स्कैनर में डाला जाता है। जैसे-जैसे कागज़ एक-एक करके आगे बढ़ते हैं, वे एक डिजिटल फ़ाइल बन जाते हैं।

स्कैन होने के बाद कागज़ को फेंक दिया जाता है।

आर्टिफिशियल इंटेलिजेंस कंपनी Anthropic के अंदर इस तरह से लाखों किताबों को प्रोसेस किया गया है। इसका नाम Project Panama रखा गया था।

कानूनी तौर पर यह एक चतुर तरीका है। खरीदी गई किताब अपनी संपत्ति होती है, इसलिए इसे काटा जा सकता है। लेकिन अगर आप इस दृश्य की अपने दिमाग में कल्पना करें, तो आपको अजीब लगेगा। किसी की जीवन भर की मेहनत से लिखी गई किताब कन्वेयर बेल्ट पर 20 सेकंड में प्रोसेस हो जाती है और कचरे के डिब्बे में चली जाती है।

Anthropic ने केवल इसी तरीके का इस्तेमाल नहीं किया। उसने LibGen और Pirimi जैसी पाइरेटेड लाइब्रेरी से 500,000 से अधिक किताबें मुफ्त में डाउनलोड भी कीं।

लेखकों Andrea Bartz, Kirk Wallace Johnson और Charles Graeber द्वारा दायर किए गए एक क्लास-एक्शन मुकदमे में, 2025 में Anthropic ने 1.5 बिलियन डॉलर का भुगतान करने का समझौता किया। यह अमेरिकी कॉपीराइट मुकदमों के इतिहास की सबसे बड़ी रकम है।

इस मामले में यह महत्वपूर्ण है कि अदालत ने अपनी सीमा कहाँ तय की। अदालत ने कॉपीराइट वाले काम से लैंग्वेज मॉडल को प्रशिक्षित करने के कार्य को अपने आप में अवैध नहीं माना। उसने जिस बात को गलत माना, वह पाइरेटेड कॉपी को डाउनलोड करना और स्टोर करना था।

इसका मतलब है कि सीखना ठीक है, लेकिन चुराकर सीखना ठीक नहीं है। यह अंतर भविष्य में आने वाले सभी मुकदमों का आधार बनेगा।

लेखक पहली बार 2023 में आगे आए थे।

उसी साल जून में, उपन्यासकार Mona Awad और Paul Tremblay ने OpenAI के खिलाफ पहला मुकदमा दायर किया। सितंबर में, अमेरिकी Authors Guild के 17 सदस्यों ने New York Federal Court में क्लास-एक्शन मुकदमा दायर किया। इनमें गेम ऑफ थ्रोन्स लिखने वाले George R.R. Martin, John Grisham, Jodi Picoult, Scott Turow, David Baldacci और Jonathan Franzen जैसे नाम शामिल थे।

उनका दावा यह था: OpenAI ने बिना अनुमति के हज़ारों उपन्यासों का इस्तेमाल प्रशिक्षण के लिए किया है। इसके लिए सामग्री Library Genesis और Z-Library जैसी पाइरेटेड लाइब्रेरी और Books3 नामक डेटासेट से ली गई थी।

ChatGPT इतने खूबसूरत वाक्य इसलिए लिखता है क्योंकि उसने ऐसे कई खूबसूरत वाक्य पढ़े हैं। और वे वाक्य किसी इंसान ने लिखे थे।

पुलित्जर पुरस्कार विजेता Michael Chabon, नाटककार David Henry Hwang और नॉन-फिक्शन लेखिका Rachel Louise Snyder ने भी OpenAI और Meta के खिलाफ कदम उठाया।

नवंबर 2023 में, लेखक Julian Sancton ने बिना अनुमति के हज़ारों नॉन-फिक्शन किताबों का इस्तेमाल करने का आरोप लगाते हुए OpenAI और Microsoft के खिलाफ एक क्लास-एक्शन मुकदमा दायर किया। Mike Huckabee सहित धार्मिक पुस्तकों के लेखकों ने Meta और Microsoft जैसी कंपनियों पर मुकदमा दायर किया, और यह दावा किया कि Books3 डेटासेट में मौजूद 180,000 किताबों का उपयोग एआई को प्रशिक्षित करने में किया गया है।

अगस्त 2024 में, पत्रकार Nicholas Basbanes और Nicholas Gage ने Books2 नामक एक अन्य डेटासेट पर आपत्ति जताते हुए इस मुकदमे में हिस्सा लिया।

अक्टूबर 2025 में, SUNY Downstate Health Sciences University की Susana Martinez-Conde और प्रोफेसर Stephen Macknik ने Apple के खिलाफ मुकदमा दायर किया। उनका दावा है कि Apple Intelligence के प्रशिक्षण में पाइरेटेड लाइब्रेरी के डेटा का इस्तेमाल किया गया था। Nvidia पर भी इसी कारण से तीन लेखकों ने मुकदमा दायर किया।

मीडिया कंपनियाँ एक अलग नज़रिए से लड़ रही हैं।

दिसंबर 2023 में, New York Times ने OpenAI और Microsoft के खिलाफ अरबों डॉलर का मुकदमा दायर किया। उनका कहना था कि अनुमति या भुगतान के बिना लाखों समाचार लेखों का उपयोग प्रशिक्षण के लिए किया गया है।

इस मुकदमे का मुख्य मुद्दा प्रशिक्षण नहीं, बल्कि प्रतिस्पर्धा है। अखबार के पैसे और समय खर्च करके की गई रिपोर्टिंग से बनी यह सर्विस, उसी अखबार के साथ पाठकों के लिए प्रतिस्पर्धा कर रही है। अगर पाठक केवल सारांश पढ़ते हैं और मूल लेख को नहीं पढ़ते, तो समाचार पत्र अपनी विज्ञापन की कमाई खो देते हैं।

रिपोर्टिंग में पैसा लगता है। पत्रकारों को फील्ड में भेजना पड़ता है, डेटा मांगना पड़ता है और सूत्रों से कई बार मिलना पड़ता है। सारांश में ऐसा कोई खर्च नहीं होता।

2026 में भी यह लड़ाई जारी है। New York Times और कई अन्य मीडिया कंपनियों ने एक Manhattan Federal Judge से प्रतिबंध लगाने का अनुरोध किया है। उनका दावा है कि OpenAI ने अपने सिस्टम के भीतर कॉपीराइट सामग्री का पता लगाने की अपनी क्षमता के बारे में अदालत को गुमराह किया है।

अप्रैल 2024 में, Chicago Tribune और Denver Post सहित आठ अमेरिकी दैनिक समाचार पत्रों ने शिकायत दर्ज की। उनके दावे में यह भी शामिल था कि न केवल उनके लेखों को बिना अनुमति के इकट्ठा किया गया, बल्कि उनके कॉपीराइट नोटिस को भी जानबूझकर हटा दिया गया।

स्रोत को हटाने की इस कार्रवाई को गलती नहीं माना जा सकता। कोड में ठीक वैसा ही लिखने पर ही इसे हटाया जा सकता है।

फरवरी 2024 में, The Intercept, Raw Story और AlterNet ने, तथा जून में Center for Investigative Reporting ने मुकदमे दायर किए। नवंबर में, Calgary Herald और Montreal Gazette के स्वामित्व वाले कनाडाई प्रकाशक भी आगे आए।

एआई सर्च सर्विस Perplexity को भी निशाना बनाया गया। अक्टूबर 2024 में Dow Jones और New York Post ने, तथा दिसंबर 2025 में Chicago Tribune ने मुकदमा दायर किया। Condé Nast और New York Times ने चेतावनी पत्र भेजे।

मार्च 2024 में फ्रांसीसी अधिकारियों ने Google पर 250 मिलियन यूरो का जुर्माना लगाया। इसका कारण यह था कि उसने मीडिया कंपनियों से चर्चा किए बिना Gemini की ट्रेनिंग के लिए उनके समाचार लेखों का उपयोग किया था।

दक्षिण कोरिया में भी 2024 और 2025 में ब्रॉडकास्टिंग कंपनियों ने Naver के खिलाफ मुकदमा दायर किया। उनका दावा है कि प्रसारण समाचार सामग्री का इस्तेमाल बिना अनुमति के HyperCLOVA X को प्रशिक्षित करने के लिए किया गया।

कला और चित्रों के मामले में, इसके सबूत सीधे स्क्रीन पर छप कर आए।

जनवरी 2023 में, Getty Images ने Stability AI के खिलाफ ब्रिटेन और अमेरिका में मुकदमा दायर किया। उनका आरोप था कि वॉटरमार्क वाली उनकी 12 मिलियन से अधिक तस्वीरों को बिना लाइसेंस के एआई को प्रशिक्षित करने के लिए इस्तेमाल किया गया है।

सबूत यह था: Stable Diffusion द्वारा बनाए गए चित्रों के कोने में Getty Images का वॉटरमार्क बिगड़े हुए रूप में मौजूद था।

मशीन यह नहीं जानती कि वॉटरमार्क क्या होता है। यह तस्वीरों पर अक्सर आने वाला एक डिज़ाइन था, इसलिए मशीन ने इसे भी साथ में सीख लिया। यह बिल्कुल ऐसा ही है जैसे कोई चोर चुराए गए सामान पर लगी दुकान की नेमप्लेट के साथ ही बाहर निकल आए।

जर्मनी के स्टॉक फ़ोटोग्राफ़र Robert Kneschke ने यह पुष्टि करने के बाद मुकदमा आगे बढ़ाया कि उनकी तस्वीरों का इस्तेमाल LAION डेटासेट के माध्यम से प्रशिक्षण के लिए किया गया है। इलस्ट्रेटर Hollie Mengert को इस बात का सामना करना पड़ा कि उनकी कला शैली को पूरी तरह से कॉपी करके एआई मॉडल बना दिया गया।

कला शैली को कॉपीराइट के ज़रिए सुरक्षित रखना मुश्किल है। ऐसा इसलिए है क्योंकि कानून अलग-अलग कृतियों की सुरक्षा करता है, न कि किसी शैली की। लेकिन एआई जो चीज़ चुरा रहा है, वह बिल्कुल वही शैली है।

कानून द्वारा खींची गई सीमा और तकनीक जो छीन रही है, वे एक-दूसरे से मेल नहीं खाते हैं।

बड़ी मीडिया कंपनियाँ भी अब मैदान में आ गई हैं।

2025 में डिज़्नी, यूनिवर्सल और वार्नर ब्रदर्स डिस्कवरी ने छवि निर्माण कंपनी मिडजर्नी के खिलाफ सामूहिक मुकदमा दायर किया। उनकी कहानी यह है कि उनकी एनिमेशन और फिल्मों के पात्रों को अनुमति के बिना एकत्र किया गया और संश्लेषित छवियों के रूप में आउटपुट किया गया। उसी कंपनियों ने चीन की मिनीमैक्स पर भी मुकदमा दायर किया।

फिल्म कंपनी अल्कॉन एंटरटेनमेंट ने अक्टूबर 2024 में टेस्ला और एलॉन मस्क के खिलाफ मुकदमा दायर किया। उन्होंने तर्क दिया कि रोबोटैक्सी सार्वजनिक कार्यक्रम में ब्लेड रनर 2049 की दृश्य छवियों को कृत्रिम बुद्धिमत्ता से फिर से तैयार करके उपयोग किया गया था।

चीन की अदालत ने 2024 में उल्ट्रामैन पात्र को अनुमति के बिना सीखकर उत्पन्न करने वाली कंपनी को कॉपीराइट उल्लंघन के लिए दोषी ठहराया।

संगीत ध्वनि में रहता है।

जून 2024 में सोनी म्यूजिक, यूनिवर्सल म्यूजिक ग्रुप और वार्नर म्यूजिक ने संगीत निर्माण कंपनियों स्पूनो और उडियो के खिलाफ मुकदमा दायर किया। उनका तर्क है कि कंपनियों ने दशकों में जमा किए गए वाणिज्यिक संगीत को अनुमति के बिना खुरच कर सीख लिया, और परिणामस्वरूप एक मशीन बनाई गई जो विशेष कलाकारों की आवाज़ और शैली को सूक्ष्मता से नकल करती है।

अक्टूबर 2023 में यूनिवर्सल म्यूजिक ग्रुप ने एंथ्रोपिक के खिलाफ गीत डेटाबेस के अनुमति के बिना संग्रह के लिए मुकदमा दायर किया। यह इसलिए था क्योंकि क्लॉड प्रसिद्ध गीतों के शब्दों को बिना किसी परिवर्तन के आउटपुट कर रहा था। इस मुकदमे में एंथ्रोपिक द्वारा दायर किए गए दस्तावेजों में गैर-मौजूद अकादमिक पेपर का हवाला दिया गया था, जो पहले उल्लेख किया गया था।

कॉमेडियन जॉर्ज कार्लिन 2008 में दुनिया से चले गए। जनवरी 2024 में उनके परिवार ने कार्लिन की आवाज़ को सीखकर कृत्रिम बुद्धिमत्ता कॉमेडी विशेष बनाने वाले चैनल के खिलाफ मुकदमा दायर किया और हटाने और समझौते को प्राप्त किया।

एक मृत व्यक्ति नई मजाकें नहीं बताता है। लेकिन वह आवाज़ नई मजाकें कहती रहती थी।

जिन लोगों के लिए आवाज़ उनकी आजीविका है, उनकी कहानियां अधिक सीधी हैं।

दिग्गज वॉयस एक्टर पॉल स्काई लेमन और लिनिया सेज ने मई 2024 में न्यूयॉर्क संघीय अदालत में वॉयस जेनरेशन स्टार्टअप के खिलाफ मुकदमा दायर किया। उन्होंने तर्क दिया कि कंपनी ने संवाद अनुसंधान के उद्देश्य के तहत उनकी आवाज़ों को रिकॉर्ड किया, लेकिन उन रिकॉर्डिंग को वाणिज्यिक कृत्रिम बुद्धिमत्ता वॉयस एक्टर सेवा के प्रशिक्षण डेटा के रूप में उपयोग और बिक्री की।

वॉयस एक्टर बेव स्टैंडिंग ने पाया कि उनकी आवाज़ TikTok की स्वचालित पाठ-से-भाषण ध्वनि के रूप में उपयोग की जा रही थी, और उन्होंने मुकदमा दायर किया और समझौता प्राप्त किया। लाखों लोग वह कहते हुए उनकी आवाज़ सुन रहे थे जो उन्होंने कहा ही नहीं था।

अभिनेत्री स्कारलेट जोहानसन ने मई 2024 में शिकायत की कि OpenAI द्वारा बनाई गई कृत्रिम बुद्धिमत्ता आवाज़ ने उनकी आवाज़ की सटीक नकल की है। OpenAI ने तुरंत वह ध्वनि सेवा को बंद कर दिया।

जोहानसन ने एक ऐप डेवलपर के खिलाफ कानूनी कार्रवाई भी की, जिसने अनुमति के बिना उनकी सामग्री वाले कृत्रिम बुद्धिमत्ता विज्ञापन बनाए थे। अभिनेता ब्रायन क्रैनस्टन ने भी खुलासा किया कि उनकी आवाज़ और चेहरा सहमति के बिना वीडियो मॉडल प्रशिक्षण के लिए उपयोग किए गए थे।

विशेषज्ञ डेटाबेस भी खुरच लिए गए।

कनाडा के कानूनी डेटाबेस CanLII ने नवंबर 2024 में एक कानूनी कृत्रिम बुद्धिमत्ता स्टार्टअप के खिलाफ मुकदमा दायर किया। उन्होंने तर्क दिया कि कंपनी ने सेवा शर्तों और पहुंच प्रतिबंधों को अनदेखा करके सैकड़ों हजार फैसले और व्याख्या दस्तावेज़ को खुरच लिया।

कानूनी सूचना कंपनी थॉमसन रॉयटर्स ने जनवरी 2025 में अपने डेटाबेस को अनुमति के बिना खुरच कर सीखने वाली Ross Intelligence के खिलाफ मुकदमे में जीत हासिल की।

इन मामलों में कंपनियों द्वारा दिया गया रक्षा तर्क काफी हद तक एक समान है। यह निष्पक्ष उपयोग है। वे तर्क देते हैं कि सीखना प्रतिलिपि नहीं बल्कि विश्लेषण है, और यह लोगों के किताब पढ़ने और सीखने से अलग नहीं है।

इस सादृश्य में कुछ गायब है। लोग एक बार में एक किताब पढ़ते हैं, पढ़ने में समय लगता है, और वे किताबें खरीदते या उधार लेते हैं। और जो क्षमता एक व्यक्ति किताब पढ़कर हासिल करता है वह केवल उस व्यक्ति की है।

मॉडल पांच लाख किताबें कुछ दिनों में पढ़ता है, और उस क्षमता को लाखों लोगों को एक साथ बेचता है।

जब पैमाना बदलता है, तो गुण बदल जाते हैं। पानी का एक कप लेना और एक पूरी नदी को खींचना एक ही काम नहीं है।

एक चीज़ स्पष्ट हो गई है। इन ट्रायलों का कोई भी अंत हो, पहले से ही सीखे गए मॉडल से किसी विशेष लेखक के निशान को चुनिंदा तरीके से मिटाने का कोई तरीका अभी तक नहीं है।

किताबें काट दी गई हैं, स्कैनिंग पूरी हो गई है, और कागज़ फेंक दिया गया है।

4.2

AI निर्मित सामग्री के कॉपीराइट की अस्वीकृति और कला पारिस्थितिकी तंत्र में उथल-पुथल

क्रिस काश्टानोवा ने 2022 की शरद ऋतु में एक ग्राफिक नॉवेल बनाया था। इसका शीर्षक ज़ार्या ऑफ़ द डॉन है।

कहानी उन्होंने खुद लिखी थी। चित्र मिडजर्नी में कमांड डालकर निकाले गए थे। मनपसंद चित्र मिलने तक उन्होंने कमांड को बार-बार बदला।

काश्टानोवा ने अमेरिका के कॉपीराइट कार्यालय में पंजीकरण के लिए आवेदन किया और उन्हें मंजूरी मिल गई।

लेकिन जब यह बात सामने आई कि चित्र आर्टिफिशियल इंटेलिजेंस से बनाए गए हैं, तो कॉपीराइट कार्यालय ने इस पर फिर से विचार किया। 2023 के फरवरी में फैसला आया।

खुद लिखे गए लेख और उसे व्यवस्थित करने के तरीके को कॉपीराइट द्वारा संरक्षित किया जाता है। मिडजर्नी द्वारा निकाले गए चित्रों का पंजीकरण रद्द किया जाता है।

कॉपीराइट कार्यालय का तर्क यह था: आप कमांड को कितने भी ध्यान से क्यों न लिखें, आप पहले से यह नहीं जान सकते कि परिणाम कैसा होगा, और न ही आप इसे बारीकी से नियंत्रित कर सकते हैं। यह किसी चित्रकार द्वारा ब्रश से एक रेखा खींचने जैसा नहीं है।

इस फैसले का अर्थ बहुत गंभीर है। आर्टिफिशियल इंटेलिजेंस द्वारा बनाया गया चित्र किसी का नहीं होता है। इसे बनाने वाला भी इसका मालिक नहीं हो सकता। कोई भी इसे लेकर इस्तेमाल कर सकता है।

कंपनियाँ आर्टिफिशियल इंटेलिजेंस वाले चित्रों का इस्तेमाल मुख्य रूप से लागत के कारण करती हैं। लेकिन इस तरह बनाए गए कवर या विज्ञापन चित्र कानूनी रूप से उनके अपने नहीं होते हैं। यदि कोई प्रतिद्वंद्वी कंपनी उन्हें वैसे ही इस्तेमाल कर ले, तो उन्हें रोकने का कोई ठोस आधार नहीं होता।

उन्हें सस्ते में बनाया गया था, लेकिन उन्हें सुरक्षित नहीं रखा जा सकता।

लेखक कौन है, इस बात को लेकर पहले भी विवाद हो चुके हैं।

2018 में न्यूयॉर्क के क्रिस्टीज़ की नीलामी में आर्टिफिशियल इंटेलिजेंस द्वारा बनाया गया एक चित्र चार लाख बत्तीस हज़ार पाँच सौ डॉलर में बिका था। इसका शीर्षक एडमंड डी बेलामी था, और इसे फ्रांसीसी कला समूह

ऑब्वियस ने पेश किया था।

बाद में एक तथ्य सामने आया। चित्र बनाने वाला कोड किसी अन्य डेवलपर द्वारा सार्वजनिक रूप से उपलब्ध कराया गया था। तो क्या इस चित्र का लेखक वह व्यक्ति है जिसने कमांड दिया, वह व्यक्ति जिसने कोड लिखा, या वे पुराने चित्रकार जिन्होंने वे चित्र बनाए जिनका इस्तेमाल इसे सिखाने के लिए किया गया था?

नीलामी के कैटलॉग में हस्ताक्षर की जगह एल्गोरिदम का सूत्र लिखा हुआ था। नीलामी की रकम एक इंसान को मिली।

जब नीदरलैंड के मौरित्शुईस संग्रहालय ने 'मोती की बाली वाली लड़की' की आर्टिफिशियल इंटेलिजेंस द्वारा बनाई गई नकल को प्रदर्शित किया, तब भी काफी आलोचना हुई थी। ऐसा इसलिए था क्योंकि मूल चित्र की जगह पर मशीन द्वारा बनाई गई नकल को लटकाया गया था।

यहाँ तक की बात कानून और संग्रहालयों की थी। असली लड़ाई किताबों के कवर को लेकर शुरू हुई।

प्रकाशक टोर बुक्स ने बेस्टसेलर लेखक क्रिस्टोफर पाओलिनी की नई किताब के कवर के लिए एक आर्टिफिशियल इंटेलिजेंस चित्र खरीदकर इस्तेमाल किया। जब यह बात सामने आई, तो पाठकों और इलस्ट्रेटर्स ने इसका विरोध किया।

प्रकाशक ब्लूम्सबरी को भी लोकप्रिय लेखिका सारा मास के उपन्यास के कवर पर आर्टिफिशियल इंटेलिजेंस चित्र का इस्तेमाल करने के कारण आलोचना का सामना करना पड़ा। डीसी कॉमिक्स ने आर्टिफिशियल इंटेलिजेंस द्वारा बनाए गए अलग-अलग कवर जारी किए थे, लेकिन अपने लेखकों के विरोध के बाद उन्हें वापस ले लिया।

कवर का चित्र बनाना इलस्ट्रेटर्स के लिए एक बड़ा काम होता है। एक चित्र बनाने में कई सप्ताह लग जाते हैं, और उस कमाई से वे कई महीनों तक अपना गुजारा करते हैं। यदि कवर बनाने का काम आर्टिफिशियल इंटेलिजेंस करने लगे, तो उनकी नौकरियाँ खत्म हो जाएँगी।

और उस आर्टिफिशियल इंटेलिजेंस ने उन्हीं इलस्ट्रेटर्स के चित्रों से सीखा है।

सबसे अजीब घटना एक टैबलेट कंपनी की तरफ से हुई।

वाकॉम (Wacom) चित्रकारों द्वारा इस्तेमाल किए जाने वाले ड्रॉइंग टैबलेट बनाती है। उनके ग्राहक ही इलस्ट्रेटर हैं। 2024 में ड्रैगन के वर्ष के अवसर पर अमेरिकी बाज़ार के लिए एक विज्ञापन बनाते समय, उन्होंने एक ड्रैगन का आर्टिफिशियल इंटेलिजेंस द्वारा बनाया गया चित्र इस्तेमाल किया।

यानी, उन्होंने अपने उत्पाद खरीदने वाले लोगों का काम छीनने वाले चित्र के साथ विज्ञापन किया।

कंपनी ने सफाई दी कि यह चित्र एक बाहरी एजेंसी द्वारा बनाया गया था और इसके लिए माफी मांगी। चित्रकारों के बीच कुछ समय तक इस कंपनी का नाम अलग तरीके से पुकारा जाने लगा।

ब्रैडफोर्ड साहित्य महोत्सव ने अपनी प्रचार सामग्री में आर्टिफिशियल इंटेलिजेंस चित्रों का इस्तेमाल किया था, जिसके कारण वहाँ आने वाले एक इलस्ट्रेटर ने हिस्सा लेने से इनकार कर दिया।

गेमिंग उद्योग में दो बार वादे तोड़े गए।

डंजन्स एंड ड्रेगन्स बनाने वाली कंपनी विजार्ड्स ऑफ द कोस्ट को तब आलोचना का सामना करना पड़ा जब यह बात सामने आई कि उनकी नई किताब के चित्रों में आर्टिफिशियल इंटेलिजेंस टूल से बनाए गए चित्र शामिल हैं। कंपनी ने नियमों को बदलते हुए कहा कि वे आर्टिफिशियल इंटेलिजेंस वाले चित्रों का इस्तेमाल नहीं करेंगे।

लेकिन इसके बाद आई प्रचार सामग्री में उन्होंने फिर से आर्टिफिशियल इंटेलिजेंस वाले चित्र का इस्तेमाल किया।

कॉल ऑफ ड्यूटी बनाने वाली कंपनी एक्टिविज़न ने गेम के अंदर बेचे जाने वाले सजावटी आइटम आर्टिफिशियल इंटेलिजेंस से बनाए और उन्हें पैसों में बेचा। लेगो ने बिना लाइसेंस वाले किरदारों को आर्टिफिशियल इंटेलिजेंस की मदद से मिलाकर अपनी प्रचार सामग्री में इस्तेमाल किया, जिसका विरोध हुआ।

पोकेमॉन जैसे दिखने वाले गेम पालवर्ल्ड पर आर्टिफिशियल इंटेलिजेंस का इस्तेमाल करके किरदारों के रूप की नकल करने का आरोप लगा।

प्रौद्योगिकी कंपनियों के प्रमुखों के फैसलों पर भी सवाल उठाए गए।

मेटा के मार्क जुकरबर्ग ने अपनी छोटी बेटियों के लिए एक कहानी की किताब बनाई और उसके चित्र अपनी ही कंपनी के आर्टिफिशियल इंटेलिजेंस इमेज जनरेटर से बनवाकर उसे सार्वजनिक किया।

दुनिया के सबसे अमीर लोगों में से एक ने अपने बच्चों को दी जाने वाली चित्र-पुस्तिका के चित्र किसी इंसान से नहीं बनवाए। शायद उन्होंने चित्रों के पैसे नहीं बचाए होंगे। बस वह तरीका ज्यादा तेज़ रहा होगा।

जल्दी होने वाले तरीकों को चुनने के इन्हीं पलों के जमा होने से उद्योग बदल जाता है।

अमेज़ॅन ने ड्रामा 'फॉलआउट' के प्रचार पोस्टर को आर्टिफिशियल इंटेलिजेंस से मिलाकर बनाया था, जिसे खराब गुणवत्ता का होने के कारण आलोचना का सामना करना पड़ा।

फिल्मों के क्षेत्र में तो गलतियाँ वैसे ही छप गईं।

फिल्म 'सिविल वॉर' बनाने वाली कंपनी A24 ने अमेरिका के प्रमुख शहरों के तबाह होने वाले प्रचार पोस्टर आर्टिफिशियल इंटेलिजेंस से बनाकर जारी किए। पोस्टर में मौजूद प्रमुख इमारतों की जगहें बिगड़ी हुई थीं, और

इंसानों के शरीर के जोड़ अजीब तरह से मुड़े हुए थे।

इसके अलावा, वे दृश्य फिल्म में भी नहीं थे।

नेटफ्लिक्स ने 'डॉग एंड बॉय' एनिमेशन में पृष्ठभूमि के चित्रों को आर्टिफिशियल इंटेलिजेंस से बनवाया और क्रेडिट में 'आर्टिफिशियल इंटेलिजेंस डेवलपर' लिखा। एनिमेशन उद्योग में नए लोग पृष्ठभूमि के चित्र बनाकर ही अपना हुनर निखारते हैं। यदि वह जगह ही खत्म हो जाएगी, तो अगली पीढ़ी आगे नहीं बढ़ पाएगी।

एनिमेशन 'आर्केन सीज़न 2' के प्रचार पोस्टर भी आर्टिफिशियल इंटेलिजेंस से बनाए गए थे, जिनमें मानव शरीर की संरचना की गलतियाँ साफ़ दिख रही थीं। मार्वल सीरीज़ 'सीक्रेट इन्वेज़न' का शुरुआती वीडियो भी आर्टिफिशियल इंटेलिजेंस से बनाया गया था, जिसका शुरुआती डिजाइनरों ने कड़ा विरोध किया।

विज्ञापनों में भी यही हुआ।

टॉयज आर अस ने OpenAI के वीडियो मॉडल का इस्तेमाल करके संस्थापक के बचपन पर आधारित एक ब्रांड वीडियो बनाया। वीडियो में दिख रहे बच्चे के चेहरे के भाव कुछ अजीब लग रहे थे और उसकी हरकतें भी फिसलती हुई लग रही थीं। इसे देखने वालों को असहजता महसूस हुई।

इंसान के चेहरे की लगभग सटीक नकल करना किसी बिल्कुल अलग चीज़ से ज्यादा असहज लगने की एक घटना होती है। यह गुड़ियों या पुतलों को देखने जैसी ही भावना है। हमारी आँखें इंसानी चेहरों को बहुत बारीकी से पढ़ने के लिए बनी हैं, इसलिए ज़रा सी भी गड़बड़ होने पर वे खतरे का संकेत दे देती हैं।

कोका-कोला ने एक पुरानी क्रिसमस ट्रक विज्ञापन को कृत्रिम बुद्धिमत्ता वीडियो से फिर से बनाया, लेकिन इसमें आत्मा नहीं रही, ऐसा कहा गया। इसी अभियान में उन्होंने उपन्यासकार जेम्स ग्राहम बैलार्ड के द्वारा न कहे गए शब्दों को कृत्रिम बुद्धिमत्ता से तैयार किया और डाला, फिर उनके परिवार और साहित्य समुदाय से विरोध का सामना किया और माफी माँगी।

कपड़ों की ब्रांड Under Armour साहित्यिक चोरी के विवाद में फंस गई जब इसके कृत्रिम बुद्धिमत्ता द्वारा बनाया गया विज्ञापन वीडियो मौजूदा निर्देशक के वीडियो की बिल्कुल नकल कहा गया।

फैशन में नकल की गति एक हथियार बन गई है।

अल्ट्रा-फास्ट फैशन कंपनी Shein ने एल्गोरिदम का उपयोग करके सोशल मीडिया पर अपलोड किए गए स्वतंत्र डिज़ाइनरों के कपड़ों के डिजाइन और पैटर्न को रीयल-टाइम में एकत्र किया। और फिर अपने कारखाने में तुरंत इसे बनाया और सस्ते में बेचा।

जब कोई स्वतंत्र डिज़ाइनर नया डिजाइन जारी करता है, तो कुछ दिनों के बाद वही कपड़े आधी कीमत पर आ जाते हैं। मूल निर्माता अपने कपड़े नहीं बेच सकता। क्षतिग्रस्त डिजाइनरों ने संघीय अदालत में मुकदमा दायर किया।

Levi's ने कृत्रिम बुद्धिमत्ता से बनाए गए विभिन्न जातियों के आभासी मॉडल का उपयोग फोटो शूट में करने की घोषणा की, लेकिन आलोचना का सामना किया। विविधता के लिए आप विविध लोगों को नियुक्त कर सकते हैं। कृत्रिम बुद्धिमत्ता मॉडल बिना रोजगार के केवल तस्वीरें प्राप्त करने का तरीका है।

जिस व्यक्ति ने चित्रकार Holly Mengert के चित्रों को सहमति के बिना सीखा और एक मॉडल बनाया, उसने यह कहा: नैतिक मानदंड केवल व्यक्तिपरक हैं।

आभासी लोगों को अभिनेता कहने की कोशिशें भी की गई हैं।

कृत्रिम बुद्धिमत्ता अभिनेत्री नामक एक पात्र सामने आया, एक भारतीय ऑटोमोबाइल कंपनी ने एक आभासी प्रभावशाली को अपनाया, और एक आभासी रैपर ने वाणिज्यिक रूप से अपनी शुरुआत की। असली अभिनेताओं और गायकों की जगह उतनी ही कम हो जाती है।

Spotify ने संगीत लाइसेंसिंग शुल्क व्यय को कम करने के लिए कृत्रिम बुद्धिमत्ता से बनाए गए नकली कलाकारों का संगीत प्लेलिस्ट के शीर्ष में बड़ी मात्रा में डाला। उपयोगकर्ता इसे पृष्ठभूमि संगीत के रूप में बजाते हैं और परवाह नहीं करते कि वे क्या सुन रहे हैं। हर कुछ मिनटों में असली संगीतकारों को मिलने वाला पैसा कम होता जाता है।

गेम कंपनियों ने आवाज अभिनेताओं को कृत्रिम बुद्धिमत्ता की आवाज़ से बदलने का प्रयास किया और कभी-कभी प्रशंसकों और आवाज अभिनेता संघों के बहिष्कार आंदोलन का सामना करना पड़ा।

इस अनुभाग में आई घटनाओं में एक प्रवाह है।

कंपनियां कृत्रिम बुद्धिमत्ता छवियों का उपयोग करती हैं। लोग ध्यान देते हैं। विरोध आता है। कंपनियां व्याख्या करती हैं या माफी माँगती हैं। अगली तिमाही में फिर से ऐसा करती हैं।

लेकिन ध्यान देने वाले लोग कौन हैं, यह महत्वपूर्ण है। ये आमतौर पर उस क्षेत्र के प्रशंसक और उस काम के निर्माता होते हैं। ये वे लोग हैं जो कवर चित्र में उंगलियों की संख्या गिनते हैं।

एक विडंबना शेष है। कृत्रिम बुद्धिमत्ता चित्रों के पास कोई कॉपीराइट नहीं है। इसलिए उन चित्रों से बने कवर, विज्ञापन, और गेम आइटम किसी के नहीं हैं।

जो सुरक्षित नहीं रखा जा सकता उसे बनाते हुए, हम जो सुरक्षित रखना चाहिए उसे नष्ट कर रहे हैं।

4.3

छवि अधिकार, वॉयस क्लोनिंग और अनधिकृत नकल

मई 2024 में OpenAI ने एक नई आवाज सेवा जारी की। आवाज का नाम Sky था।

जिन लोगों ने इसे सुना, सब एक ही सोच रहे थे। क्या यह वही आवाज नहीं है?

फिल्म Her में Scarlett Johansson को स्क्रीन पर कभी चेहरा नहीं दिखता। वह केवल आवाज से कृत्रिम बुद्धिमत्ता की भूमिका निभाती हैं। जिन लोगों ने यह फिल्म देखी है, उनके लिए उस टोन और सांस को भूलना मुश्किल है।

पूरी बातचीत यह थी। OpenAI के मुख्य कार्यकारी अधिकारी Sam Altman ने सेवा जारी करने से पहले Johansson को सीधे संपर्क किया और कहा कि आवाज देने का काम करें। Johansson ने इनकार कर दिया।

और फिर एक समान आवाज आ गई।

जारी करने के दिन Altman ने सोशल मीडिया पर एक शब्द डाला। her.

तीन अक्षर हैं। कि ये तीन अक्षर बाद में अदालत में क्या वजन रखेंगे, यह वकीलों को देखना है। पर लोगों ने इसे किस अर्थ में समझा, यह पूरी तरह साफ था।

Johansson ने कानूनी कार्रवाई की चेतावनी दी, तो OpenAI ने Sky आवाज को तुरंत हटा दिया।

यह घटना एक सवाल छोड़ गई। आवाज किसकी है?

कानून इस सवाल का अभी पूरी तरह जवाब नहीं दे सका है। कॉपीराइट कलाकृति की सुरक्षा करता है। आवाज कोई कलाकृति नहीं है बल्कि शरीर की एक विशेषता है। छवि अधिकार चेहरे की सुरक्षा करते हैं। आवाज चेहरा नहीं है।

इसलिए नकल को सजा देना मुश्किल है। आवाज की नकल लंबे समय से मनोरंजन रहा है।

जब मशीन नकल करती है तो यह अलग है। बिना थकावट के, कुछ सेकंड का नमूना ही काफी है, और एक दिन में हजारों बना सकती है।

फरवरी 2026 में अमेरिकी सार्वजनिक रेडियो NPR के अनुभवी समाचार वाचक David Green ने कैलिफोर्निया के Santa Clara काउंटी की अदालत में Google के विरुद्ध मुकदमा दायर किया। Google ने पॉडकास्ट और

दस्तावेज सारांश सेवा बनाते समय उनकी आवाज की विशेषता और लहजे को बिना अनुमति और मुआवजे के सीखा और इसे कृत्रिम बुद्धिमत्ता समाचार वाचक आवाज में बदल दिया, यह उनका दावा था।

Green को धन की चिंता नहीं थी। उन्हें जो चिंता थी वह यह कि उनकी आवाज ऐसी चीजें कह सकती है जो वह कभी नहीं कहेंगे।

प्रसारण समाचार वाचक के लिए आवाज विश्वास का संकेत है। अगर वह आवाज कुछ कहती है, तो लोग सोचते हैं कि यह सच होगा। यह विश्वास दशकों में बनता है।

OpenAI के वीडियो निर्माण साधन Sora को भी अभिनेता Bryan Cranston की आवाज और चेहरे को बिना अनुमति सीखने का आरोप लगा है।

जो लोग दुनिया से चले गए हैं, उनकी आवाज विशेष रूप से अक्सर इस्तेमाल होती है। इसलिए कि उनसे आपत्ति करने वाला कोई नहीं है।

अक्टूबर 2023 में अभिनेता Robin Williams की बेटी Zelda Williams ने अपने पिता को नकल करने वाली कृत्रिम बुद्धिमत्ता आवाज और वीडियो देखकर यह कहा। भयानक और विचित्र।

अपने पिता की आवाज में ऐसी बातें सुनना जो पिता ने कभी नहीं कहीं, यह कैसा लगता है, जिसने अनुभव नहीं किया है वह केवल सोच सकता है।

2021 की वृत्तचित्र Roadrunner में निर्देशक ने दुनिया से गए शेफ Anthony Bourdain के जीवनकाल के ईमेल वाक्यों को कृत्रिम बुद्धिमत्ता आवाज से पढ़ा और फिल्म में डाला। Bourdain द्वारा लिखे वाक्य थे। पर उन्होंने कभी जोर से कहे नहीं थे।

लिखी गई बातें और जोर से कही गई बातें अलग हैं। वृत्तचित्र में यह अंतर तथ्य और निर्देशन की सीमा है।

कोरिया में भी गायक Kim Kwang-seok की आवाज को एल्गोरिदम से फिर जीवंत किया गया और नया गीत गाने के लिए बनाया गया। चीन में अभिनेताओं और गायकों के जीवंत रूप को कृत्रिम बुद्धिमत्ता से पुनर्जीवित करके वीडियो फैले।

पॉप संगीत में जीवंत गायकों की आवाज पहले नकल की गई थी।

अप्रैल 2023 में Ghostwriter नाम के एक अज्ञात निर्माता ने Heart on My Sleeve नाम का गीत जारी किया। यह Drake और The Weeknd की आवाज को सूक्ष्मता से नकल करने वाला गीत था। यह TikTok और Spotify पर तेजी से बजाया गया।

दोनों गायकों की रिकॉर्ड कंपनियों ने हटाने की मांग की और गीत को हटा दिया गया।

एक दिलचस्प पहलू है। इस गीत को पसंद करने वाले काफी लोग थे। असली से बेहतर है, ऐसी बातें भी सुनी गईं। यह प्रतिक्रिया संगीत उद्योग को और ज्यादा चिंतित कर दिया।

Drake ने स्वयं अप्रैल 2024 में एक प्रतिक्रिया गीत में दुनिया से गए रैपर Tupac Shakur की आवाज को कृत्रिम बुद्धिमत्ता से नकल कर इस्तेमाल किया, फिर परिवार की ओर से चेतावनी मिली और गीत को हटा दिया।

जिसकी आवाज नकल की गई, उसने दूसरे की आवाज नकल की।

चीन में एक गायक की आवाज नकल कर दूसरे गीत गाने वाली ऑडियो फाइलें बड़े पैमाने पर फैलीं और लाखों बार बजाई गईं।

जो लोग आवाज से अपना जीवन चलाते हैं, उनके लिए यह तकनीक आजीविका का मुद्दा है।

अनुभवी आवाज अभिनेता Paul Sky Riemann और Linea Sage ने मई 2024 में न्यूयॉर्क संघीय अदालत में मुकदमा दायर किया। एक आवाज निर्माण स्टार्टअप ने कहानी संबंधी अनुसंधान उद्देश्य बताकर रिकॉर्डिंग की थी, और उस डेटा को वाणिज्यिक कृत्रिम बुद्धिमत्ता आवाज सेवा प्रशिक्षण में बदल दिया और बेच दिया, ऐसा दावा है।

आवाज अभिनेता Bev Standing की आवाज को TikTok की स्वचालित पढ़ने आवाज में इस्तेमाल किया गया। यह जानकर कि बिना अनुमति सीखा गया, मुकदमा दायर किया और समझौता पाया।

इस मामले का अजीब बिंदु पैमाना है। उनकी आवाज एक दिन में सैकड़ों मिलियन बार बजाई गई, लेकिन उस बजाने से उन्हें एक पैसा भी नहीं मिला।

ऑडियोबुक सुनाने वाले लोगों ने भी शिकायत की है कि Apple ने उनकी आवाज को प्रशिक्षण में इस्तेमाल किया। IBM ने आवाज अभिनेता Greg Masten की आवाज को वाणिज्यिक नकल के लिए बेचा और आलोचना पाई। ब्रिटेन के ScotRail की घोषणा आवाज अभिनेता ने भी कहा कि उनकी आवाज को कृत्रिम बुद्धिमत्ता से बदल दिया गया और इस्तेमाल किया जा रहा है।

गेम कंपनियों ने मानव आवाज अभिनेता की जगह नकल आवाज इस्तेमाल करने की कोशिश की, पर आवाज अभिनेता संघ और गेमर्स के विरोध का सामना करना पड़ा।

चेहरे की ओर तुरंत धोखाधड़ी जुड़ी है।

TikTok का DeepTom Cruise खाता अभिनेता Tom Cruise के चेहरे और आवाज को पूरी तरह संश्लेषित करके सैकड़ों लाख अनुयायी इकट्ठा कर लिए। बनाने वाले लोगों ने कहा कि उद्देश्य जोखिम से सचेत करना था।

लोगों को यह मजेदार लगा।

अभिनेता Tom Hanks ने अपनी संश्लेषित छवि एक दंत बीमा विज्ञापन में देखी जहां वह नहीं थे, और सीधे चेतावनी दी। Bruce Willis ने विज्ञापन के लिए अपने चेहरे का इस्तेमाल करने के लिए अनुबंध किया, पर बाद में अपनी इच्छा के खिलाफ उनके चेहरे की नकल की जा रही है, इस विवाद में फंस गए।

प्रसारणकर्ता Martin Lewis, समाचार वाचक Mary Nightingale, YouTuber MrBeast, मौसम प्रस्तुतकर्ता और समाचार प्रस्तुतकर्ताओं के चेहरों को Bitcoin निवेश धोखाधड़ी और नकली पुरस्कार विज्ञापनों में संश्लेषित किया गया।

यह सोचना भयानक है कि ये धोखाधड़ी क्यों इतनी अच्छी तरह काम करती हैं। लोग परिचित चेहरों पर विश्वास करते हैं। जब रोज शाम की समाचार में देखा जाने वाला चेहरा निवेश की सलाह देता है, तो सावधानी टूट जाती है।

फ्रांस में एक महिला अभिनेता Brad Pitt के डीपफेक से धोखा खा गई और 830,000 यूरो खो गई। कहा जाता है कि कई महीने तक बातचीत करते रहे।

भारत के मशहूर अभिनेता अनिल कपूर ने नई दिल्ली उच्च न्यायालय में एक मुकदमा दायर किया। उन्होंने यह मुकदमा तब किया जब उनके नाम, चेहरे, आवाज़ और मशहूर वाक्यों का विज्ञापनों और उत्पादों में बिना अनुमति के इस्तेमाल किया गया। अदालत ने माना कि यह उनके व्यक्तित्व और छवि के अधिकारों का उल्लंघन है। अदालत ने बिना सहमति के सभी कृत्रिम बुद्धिमत्ता की नकल और छेड़छाड़ की गई सामग्री के प्रसार को रोकने का आदेश दिया।

यह फैसला एक महत्वपूर्ण उदाहरण बन गया है। इसमें चेहरे, आवाज़ और बोलने के तरीके, इन सभी को एक व्यक्ति के व्यक्तित्व के रूप में एक साथ सुरक्षित किया गया है।

आध्यात्मिक गुरु सद्गुरु ने भी अपने छेड़छाड़ किए गए वीडियो को फैलाने के खिलाफ उसे हटाने का आदेश प्राप्त किया। स्पेन के एक शराब ब्रांड ने एक दिवंगत गायक को डीपफेक के ज़रिए ज़िंदा कर एक विज्ञापन में इस्तेमाल किया, जिसके बाद विवाद खड़ा हो गया।

कंसल्टिंग ग्रुप WPP के मुख्य कार्यकारी अधिकारी मार्क रीड डीपफेक से बनी वीडियो कॉन्फ्रेंस स्क्रीन के कारण एक बहुत बड़े धोखे का शिकार होने वाले थे। स्क्रीन के अंदर उनका अपना चेहरा मीटिंग कर रहा था।

चुनावों में आवाज़ एक हथियार बन गई है।

जनवरी 2024 में अमेरिका के न्यू हैम्पशायर में डेमोक्रेटिक पार्टी के प्राथमिक चुनाव से पहले, हज़ारों मतदाताओं को फोन कॉल आए। यह राष्ट्रपति जो बाइडेन की आवाज़ थी। फोन में कहा गया था कि वे प्राथमिक चुनाव में न

आएं और मुख्य चुनाव के लिए अपना वोट बचाकर रखें।

फोन उठाने वाले लोग राष्ट्रपति की आवाज़ जानते हैं। इसलिए उन्हें कोई शक नहीं होता।

जांच अधिकारियों ने इस रोबोकॉल को बनाने वाले व्यक्ति को पकड़ा और उस पर भारी जुर्माना लगाया।

एलन मस्क ने उपराष्ट्रपति कमला हैरिस की आवाज़ की नकल की और एक नकली राजनीतिक विज्ञापन वीडियो अपने प्लेटफॉर्म पर साझा किया। इस वीडियो में ऐसा लग रहा था जैसे वह अपनी अक्षमता मान रही हों, जिसके लिए उनकी काफी आलोचना हुई।

न्यूयॉर्क के मेयर एरिक एडम्स ने अपनी आवाज़ की नकल करके स्पेनिश और चीनी भाषा में नागरिकों को स्वचालित फोन कॉल भेजे। वह उन भाषाओं को नहीं बोल पाते हैं।

ब्रिटेन की लेबर पार्टी के नेता कीर स्टार्मर को एक नकली ऑडियो के कारण नुकसान उठाना पड़ा, जिसमें वह एक कर्मचारी को गाली देते हुए सुनाई दे रहे थे। लंदन के मेयर सादिक खान का भी एक नकली ऑडियो फाइल के कारण विरोध हुआ, जिसमें वह स्मृति दिवस के कार्यक्रम का अपमान करते हुए सुने गए थे।

2023 में स्लोवाकिया में आम चुनाव से कुछ दिन पहले, एक पार्टी के नेता का नकली ऑडियो फैल गया जिसमें वह चुनाव में धांधली की चर्चा कर रहे थे। चुनाव से ठीक पहले जवाब देने का समय नहीं होता है। यही उनका मकसद था।

थाईलैंड, ब्रिटेन, जर्मनी और हंगरी में भी राजनीतिक विवादों में नकली आवाज़ों और वीडियो का इस्तेमाल किया गया। Hanguk में भी 2022 के राष्ट्रपति चुनाव में उम्मीदवारों के कृत्रिम बुद्धिमत्ता अवतार सामने आए, जिससे राजनीतिक प्रतिनिधियों की सीमा को लेकर बहस छिड़ गई।

इसका सबसे क्रूर इस्तेमाल आखिर में आता है।

जनवरी 2024 में X पर पॉप स्टार टेलर स्विफ्ट के चेहरे वाली नकली अश्लील तस्वीरें तेज़ी से फैल गईं। इन्हें करोड़ों बार देखे जाने तक हटाया नहीं गया था।

अभिनेत्री गैल गैडोट, अमेरिकी प्रतिनिधि सभा की सदस्य अलेक्जेंड्रिया ओकासियो-कोर्टेज़, इटली की प्रधानमंत्री जियोर्जिया मेलोनी और भारतीय पत्रकार राणा अय्यूब सहित कई महिला सार्वजनिक हस्तियों को भी इसी तरह की स्थिति का सामना करना पड़ा। प्रधानमंत्री मेलोनी ने इसे बनाने वाले के खिलाफ दीवानी मुकदमा दायर किया।

इस पीड़ितों की सूची में एक बात ध्यान खींचती है। इनमें से लगभग सभी महिलाएँ हैं। और ज़्यादातर ऐसी महिलाएँ हैं जो सार्वजनिक रूप से अपनी बात रखती हैं।

और यह अपराध अब स्कूलों तक पहुँच गया है। स्पेन के अलमेंड्रालेजो में यह मामला सामने आया कि मिडिल स्कूल के छात्रों ने अपने ही स्कूल की लगभग बीस छात्राओं की तस्वीरों से अश्लील तस्वीरें बनाईं और उन्हें फैलाया। अमेरिका, कनाडा, स्कॉटलैंड और ऑस्ट्रेलिया के हाई स्कूलों में भी ऐसी ही घटनाएँ हुईं।

Hanguk में भी 2024 में टेलीग्राम के ज़रिए डीपफेक यौन अपराधों ने मिडिल, हाई स्कूलों और विश्वविद्यालयों को हिला कर रख दिया।

इस समस्या पर अगले अध्याय में और चर्चा की जाएगी। यहाँ हम केवल तकनीक के पहलू को देखेंगे।

बस एक तस्वीर की ज़रूरत होती है। आवाज़ के लिए बस कुछ सेकंड चाहिए होते हैं। यह आज की तकनीक का स्तर है।

और जिन जगहों ने यह तकनीक बनाई, उन्होंने ज़्यादातर एक चीज़ नहीं बनाई। वह है यह जाँचने की प्रक्रिया कि उस व्यक्ति ने सहमति दी है या नहीं।

ऐसा इसलिए नहीं है कि यह तकनीकी रूप से असंभव है। बल्कि इसलिए है क्योंकि अगर इसे बनाया गया, तो सेवा का उपयोग करना मुश्किल हो जाएगा और उपयोगकर्ता कम हो जाएंगे।

चेहरा और आवाज़ हमें जन्म के समय मिलते हैं। इन्हें बदला नहीं जा सकता, छिपाया नहीं जा सकता, और हम इन्हें हर दिन दूसरों को दिखाते हुए जीते हैं।

अब एक ऐसा उद्योग बन गया है जो इनका सामग्री के रूप में उपयोग करता है। इस सामग्री की कीमत अभी तक किसी ने नहीं चुकाई है।

अध्याय 5

स्वचालित वाहन और रोबोट: भौतिक सुरक्षा और जनहानि

5.1

स्वायत्त वाहन की खामियाँ और सड़क दुर्घटनाएँ

18 मार्च 2018 की रात नौ बजकर अट्ठावन मिनट, अमेरिका का Arizona राज्य, Tempe।

उनचास वर्षीय Elaine Herzberg अपनी साइकिल को धकेलते हुए एक अंधेरी सड़क पार कर रही थीं। वह कोई पैदल पार पथ (जेब्रा क्रॉसिंग) नहीं था।

Uber का सेल्फ-ड्राइविंग परीक्षण वाहन अड़तीस मील प्रति घंटे की गति से आ रहा था। ड्राइवर की सीट पर एक परीक्षण ड्राइवर बैठा था, और वह अपने मोबाइल फोन पर एक मनोरंजन कार्यक्रम देख रहा था।

कार के सेंसर ने टक्कर से छह सेकंड पहले ही Herzberg का पता लगा लिया था।

छह सेकंड एक लंबा समय होता है। आप एक, दो, तीन को दो बार गिन सकते हैं। उतने समय में कार को रोका जा सकता है।

समस्या यह है कि सॉफ्टवेयर ने उन छह सेकंड के दौरान क्या किया।

उसने इसे एक अज्ञात वस्तु के रूप में वर्गीकृत किया। फिर इसे एक वाहन के रूप में वर्गीकृत किया। फिर इसे एक साइकिल के रूप में वर्गीकृत किया। फिर से इसे एक वस्तु में बदल दिया। हर बार जब वर्गीकरण बदलता था, तो यह अनुमान लगाने की गणना कि यह लक्ष्य कहाँ जाएगा, शुरू से फिर से शुरू हो जाती थी।

मशीन ने ऐसी स्थिति नहीं सीखी थी जहाँ कोई इंसान साइकिल को धकेलते हुए चल रहा हो। इसके लर्निंग डेटा में, साइकिल सवारी करने की चीज़ है और इंसान चलने वाला प्राणी है। उन दोनों का एक साथ चलना उन दोनों के बीच कहीं था।

और एक निर्णायक सेटिंग थी। Uber की डेवलपर टीम ने आपातकालीन स्वचालित ब्रेकिंग सुविधा को बंद कर रखा था।

इसका कारण सवारी का अनुभव था। सिस्टम बार-बार अचानक ब्रेक लगा रहा था, जिससे टेस्ट ड्राइव असुविधाजनक हो रही थी। इसके बजाय, यह तय किया गया था कि खतरे के समय मानव ड्राइवर हस्तक्षेप करेगा।

वह इंसान स्क्रीन देख रहा था।

कार ने अपनी गति कम नहीं की। Elaine Herzberg को सेल्फ-ड्राइविंग कार के कारण जान गंवाने वाली पहली पैदल यात्री के रूप में दर्ज किया गया।

इस दुर्घटना में स्वचालन के सभी पुराने खतरे शामिल हैं। जब मशीन ज्यादातर काम अच्छे से करती है, तो इंसान सतर्कता छोड़ देता है। लेकिन सिस्टम को इस धारणा पर डिज़ाइन किया गया था कि इंसान हमेशा सतर्क रहता है।

इंसान इस तरह से नहीं बने हैं। घंटों तक ऐसी स्क्रीन को ध्यान से देखना जहाँ कुछ नहीं हो रहा हो, उन कामों में से एक है जो इंसान सबसे खराब तरीके से करते हैं।

Uber की दुर्घटना के बाद भी बिना ड्राइवर वाली टैक्सियाँ सड़कों पर उतरीं।

अक्टूबर 2023 में, San Francisco शहर के केंद्र में एक महिला को दूसरी कार ने टक्कर मार दी और वह छिटक कर दूर गिर गई। वह बगल वाली लेन में मौजूद General Motors की सहायक कंपनी Cruise की बिना ड्राइवर वाली टैक्सी के नीचे चली गई।

यहाँ सिस्टम ने फिर से गणना शुरू कर दी। यह इस बात को नहीं पहचान सका कि कार के नीचे कोई इंसान है। नीचे के सेंसर उस स्थिति को पकड़ नहीं पाए।

सिस्टम ने निर्धारित नियमों को लागू किया। यदि कोई दुर्घटना महसूस हो, तो सड़क के किनारे की ओर बढ़ें और सुरक्षित रूप से रुकें।

बिना ड्राइवर वाली टैक्सी सात मील प्रति घंटे की गति से छह मीटर तक चली। वह इंसान को कुचलते और घसीटते हुए आगे बढ़ी।

पीड़िता को स्थायी रूप से गंभीर चोटें आईं।

यह हिस्सा स्वचालन के डर को बिल्कुल सही तरीके से दिखाता है। नियम अपने आप में तार्किक हैं। यदि कोई दुर्घटना होती है, तो रास्ता मत रोकें और किनारे हट जाएं। मानव ड्राइवरों को भी यही सिखाया जाता है।

अगर कोई मानव ड्राइवर होता, तो उसने कार के नीचे से आने वाली आवाज़ें सुनी होतीं। उसने चीख सुनी होती। उसने बाहर लोगों को हाथ हिलाते हुए देखा होता।

उन चीज़ों को ध्यान में रखने की क्षमता जो नियमों में नहीं हैं। मुझे नहीं पता कि इसे क्या कहा जाए, लेकिन मशीन में यह क्षमता नहीं होती।

उसके बाद जो हुआ वह तकनीकी समस्या नहीं है। जांच करने वाले अधिकारियों को, Cruise के प्रबंधन ने इंसान को घसीटे जाने वाले हिस्से का वीडियो नहीं दिया। उन्होंने रिपोर्ट दी कि यह सुरक्षित है।

California राज्य ने Cruise का बिना ड्राइवर वाला परिचालन परमिट रद्द कर दिया। न्याय विभाग की आपराधिक जांच शुरू हुई, मुख्य कार्यकारी अधिकारी ने पद छोड़ दिया, और रोबोटैक्सी का संचालन पूरी तरह से रोक दिया गया।

Cruise की कारों ने इससे पहले भी कई घटनाएं की थीं।

उन्होंने एक मुड़ने वाली बस के मुड़े हुए हिस्से की गति को गलत तरीके से पढ़ा और बस के पीछे टक्कर मार दी। उन्होंने सायरन बजाते हुए आ रही फायर ब्रिगेड की गाड़ी को समय पर नहीं पहचाना और चौराहे पर उससे टकरा गईं। उन्होंने गोलीबारी की घटना वाली जगह पर पहुंची पुलिस की कार का रास्ता रोक दिया।

आंतरिक सुरक्षा मूल्यांकन रिपोर्ट में यह बताया गया था कि बच्चों को पहचानने में इसकी सीमाएं थीं। बच्चे अचानक दौड़ते हैं, अप्रत्याशित रूप से दिशा बदलते हैं, और उनकी ऊंचाई कम होती है।

और फिर यह वाक्य आता है। शहर के केंद्र में हर चार किलोमीटर से आठ किलोमीटर चलने पर, Cruise की बिना ड्राइवर वाली टैक्सी को अपना सफर जारी रखने के लिए रिमोट कंट्रोल सेंटर से किसी इंसान के हस्तक्षेप की आवश्यकता होती थी।

बिना ड्राइवर शब्द के साथ एक शर्त जुड़ी हुई थी।

Waymo का रिकॉर्ड भी छोटा नहीं है।

फरवरी 2024 में San Francisco में, इसने एक साइकिल चालक को टक्कर मारकर घायल कर दिया। दिसंबर 2023 से फरवरी 2024 के बीच, Phoenix में, दो कारों ने एक के बाद एक उस पिकअप ट्रक को टक्कर मारी जिसे टो ट्रक द्वारा उल्टा घसीटा जा रहा था।

एक ट्रक उल्टा लटका हुआ और विपरीत दिशा में जा रहा है। जब कोई इंसान यह दृश्य देखता है, तो वह मुस्कुराते हुए इससे बचकर निकल जाता है। यह ऐसा दृश्य है जो इसके लर्निंग डेटा में नहीं है।

मई 2024 में, Phoenix की एक गली में यह एक टेलीफोन खंभे से टकरा गई। नवंबर 2025 में, Georgia राज्य के Atlanta में, इसने उस स्कूल बस को नज़रअंदाज़ कर दिया जो रुकने का संकेत चालू करके बच्चों को उतार रही थी, और उसके बगल से गुज़र गई। अमेरिकी सड़क यातायात सुरक्षा प्रशासन ने जांच शुरू की।

जब एक स्कूल बस अपनी लाल बत्तियां चालू करती है और स्टॉप का निशान दिखाती है, तो सभी कारों को रुकना होता है। यह ड्राइविंग लाइसेंस की परीक्षा में आने वाला नियम है।

23 जनवरी 2026, California का Santa Monica।

यह प्राथमिक विद्यालय से दो ब्लॉक दूर की जगह थी। यह स्कूल जाने का समय था, पास में अन्य बच्चे और एक यातायात निर्देशक मौजूद थे, और डबल-पार्क की गई कारों की कतार लगी थी।

Waymo की बिना ड्राइवर वाली कार ने एक बच्चे को टक्कर मार दी।

Waymo ने यह बयान दिया: जैसे ही इंसान खड़ी कार के पीछे से बाहर आया, हमने उसे तुरंत पहचान लिया, और जोर से ब्रेक लगाकर सत्रह मील प्रति घंटे की गति से घटाकर छह मील प्रति घंटे की गति पर उससे संपर्क किया।

बच्चे को मामूली चोटें आईं। सड़क यातायात सुरक्षा प्रशासन ने जांच शुरू की।

यह स्पष्टीकरण तकनीकी रूप से उत्कृष्ट है। इसने एक इंसान की तुलना में तेज़ी से पता लगाया होगा और इंसान की तुलना में अधिक जोर से ब्रेक दबाया होगा।

लेकिन कल्पना कीजिए कि उस बच्चे के माता-पिता यह वाक्य पढ़ रहे हैं। यह तथ्य कि गति सत्रह से घटाकर छह कर दी गई थी, कोई दिलासा नहीं देता।

इसके अलावा, Waymo की कारों ने गली से भागते हुए आए एक कुत्ते और एक दुकान के सामने एक बिल्ली को कुचलकर मार डाला, बड़े पैमाने पर गोलीबारी की घटना वाले स्थान पर जा रही एक एम्बुलेंस का रास्ता रोक दिया, और वाहन से उतरने की सूचना देने वाले सॉफ्टवेयर में खराबी के कारण एक साइकिल चालक को घायल कर दिया जब एक यात्री दरवाज़ा खोल रहा था।

सबसे लंबी सूची Tesla की है।

मई 2016 में Florida में, ऑटोपायलट पर चल रही एक Model S कार बाईं ओर मुड़ रहे एक बड़े ट्रेलर के बगल वाले हिस्से से जा टकराई। साफ आसमान के नीचे एक सफेद ट्रेलर था। कैमरा सफेद पृष्ठभूमि और सफेद कार बॉडी के बीच अंतर नहीं कर सका।

कार ने ब्रेक नहीं लगाया और ट्रेलर के नीचे घुस गई। उसकी छत उड़ गई और ड्राइवर Joshua Brown की मौके पर ही मौत हो गई।

मार्च 2018 में कैलिफोर्निया के माउंटेन व्यू में, ऐप्पल के इंजीनियर वाल्टर हुआंग मॉडल X को ऑटोपायलट पर चला रहे थे, जब उनकी कार एक कंक्रीट बैरियर से सीधे टकरा गई। गति इकहत्तर मील प्रति घंटे थी।

इसका कारण लेन की लाइनें थीं। इंटरचेंज पर जहाँ लेन अलग होती है, वहाँ सिस्टम ने एक धुंधली लाइन का गलत तरीके से पीछा किया, और कार को बैरियर की तरफ ले गया।

हुआंग ने दुर्घटना से पहले कई बार बताया था कि कार उसी बिंदु पर बैरियर की ओर मुड़ जाती थी।

रात में खड़ी हुई बड़ी वस्तुओं को न पहचान पाने की समस्या जारी रही।

मार्च 2019 में फ्लोरिडा के डेलरे बीच में, एक मॉडल 3 ने सड़क पार कर रहे एक बड़े ट्रेलर को अड़सठ मील प्रति घंटे की गति से टक्कर मार दी, जिससे चालक जेरेमी बैनर की मौत हो गई। यह 2016 की दुर्घटना जैसी ही स्थिति थी। तीन साल में इसे ठीक नहीं किया गया था।

दिसंबर 2019 में इंडियाना राज्य में, एक कार ने चेतावनी लाइट जलाए दुर्घटना स्थल पर खड़े एक फायर ट्रक को पीछे से टक्कर मार दी, जिससे यात्री की पत्नी की मौत हो गई। उसी महीने कैलिफोर्निया के गार्डेना में, ऑटोपायलट पर चल रही एक मॉडल S राजमार्ग से उतरी और लाल बत्ती को अनदेखा करते हुए होंडा सिविक से टकरा गई। दो लोगों की मौत हो गई।

इस दुर्घटना के चालक पर ऑटोपायलट उपयोगकर्ता के रूप में पहली बार गैर इरादतन हत्या का आपराधिक आरोप लगाया गया था।

सितंबर 2020 में जॉर्जिया राज्य में, पैंतालीस मील प्रति घंटे की सीमा वाले क्षेत्र में सतहत्तर मील प्रति घंटे की गति से चल रही एक मॉडल 3 बारिश के कारण फिसल गई और एक बस स्टॉप से जा टकराई। स्टॉप पर बैठे छिहत्तर वर्षीय बर्नार्ड जोंस की मौत हो गई।

अप्रैल 2021 में टेक्सास राज्य के स्प्रिंग में, एक मॉडल S, जिसकी ड्राइविंग सीट पर कोई नहीं था, सड़क से उतर गई और एक पेड़ से टकराकर जल गई। दो लोगों की मौत हो गई।

नवंबर 2022 में सैन फ्रांसिस्को बे ब्रिज सुरंग में, विपरीत दिशा में एक दुर्घटना हुई। पैंसठ मील प्रति घंटे की गति से चल रही एक कार ने अचानक ब्रेक लगा दिए, जबकि सामने कुछ भी नहीं था। इसे फैंटम ब्रेकिंग कहा जाता है। पीछे चल रही आठ गाड़ियाँ एक-दूसरे से टकरा गईं और एक बच्चे सहित नौ लोग घायल हो गए।

मोटरसाइकिल चालक विशेष रूप से खतरे में थे।

जुलाई 2022 में कैलिफोर्निया में, एक मॉडल Y ने सीधे अपने आगे चल रही एक यामाहा मोटरसाइकिल को टक्कर मार दी। उसी महीने यूटा में, एक मॉडल 3 ने एक हार्ले-डेविडसन को टक्कर मारी। अगस्त में, एक कावासाकी को टक्कर मारी गई।

इसका कारण पीछे की बत्ती (टेललाइट) में है। मोटरसाइकिल में एक बत्ती होती है और वह छोटी होती है। सिस्टम ने रात में उस बत्ती को दूर की स्ट्रीट लाइट या सड़क के संकेत के रूप में पढ़ा। यदि सिस्टम मान लेता है कि वस्तु दूर है, तो वह गति कम नहीं करता है।

मोटरसाइकिल चलाने वाले व्यक्ति को पीछे से आ रही कार दिखाई नहीं देती है।

नवंबर 2023 में एरिज़ोना में, शाम की तेज़ धूप के कारण, केवल कैमरों का उपयोग करने वाला सिस्टम सामने नहीं देख पाया और एक इकहत्तर वर्षीय पैदल यात्री को टक्कर मार दी, जिससे उसकी मौत हो गई। जुलाई 2023 में ब्रुकलिन में, फुटपाथ पर बस का इंतज़ार कर रहे एक छिहत्तर वर्षीय पैदल यात्री की जान चली गई।

मई 2023 में, टेस्ला के व्हिसलब्लोअर लुकास क्रुप्सकी ने सौ गीगाबाइट दस्तावेज़ सार्वजनिक किए। इसमें बताया गया था कि अचानक ब्रेक लगने और अप्रत्याशित रूप से गति बढ़ने की हज़ारों ग्राहक शिकायतें जमा हो गई थीं, और प्रबंधन को यह पता था लेकिन उन्होंने इसे बाहर नहीं बताया।

2026 में भी जांच जारी है। टेक्सास में, एक टेस्ला कार के एक घर से टकराने और एक छिहत्तर वर्षीय महिला की मौत होने की दुर्घटना की राष्ट्रीय राजमार्ग यातायात सुरक्षा प्रशासन और राष्ट्रीय परिवहन सुरक्षा बोर्ड ने जांच शुरू कर दी है। टेस्ला कारों द्वारा लाल बत्ती की अनदेखी करने या गलत दिशा में चलने की दर्जनों शिकायतें मिलने के बाद एक अलग जांच भी शुरू की गई है।

जुलाई 2026 में, सरकारी डेटा में यह पुष्टि हुई कि ह्यूस्टन में टेस्ला रोबोटैक्सी के एक रिमोट ऑपरेटर ने दुर्घटना की थी। एक व्यक्ति दूर से एक मानवरहित कार चला रहा था और उसने दुर्घटना कर दी।

अन्य कंपनियाँ भी इस सूची में हैं।

अमेज़ॅन की स्वायत्त ड्राइविंग कंपनी ज़ूक्स (Zoox) ने अप्रैल 2025 में लास वेगास दुर्घटना के बाद सॉफ्टवेयर के माध्यम से दो सौ सत्तर वाहनों को वापस मंगाया (रिकॉल किया), और अगस्त में, सेंटर लाइन पार करने और विपरीत लेन के सामने अचानक रुकने की खराबी के कारण तीन सौ बत्तीस वाहनों को अतिरिक्त रूप से रिकॉल किया।

चीन की बाइडू (Baidu) की एक मानवरहित कार ने 2024 में एक पैदल यात्री को टक्कर मार दी। एक स्वायत्त ड्राइविंग स्टार्टअप का परमिट 2021 में कैलिफोर्निया में परीक्षण के दौरान एक डिवाइडर और एक संकेत से टकराने के बाद निलंबित कर दिया गया था।

चीन की इलेक्ट्रिक कार कंपनी निओ (Nio) के एक वाहन ने अगस्त 2021 में ड्राइविंग सहायता सुविधा चालू करके चलते हुए एक राजमार्ग रखरखाव वाहन को टक्कर मार दी, जिससे चालक की मौत हो गई। एक्सपेंग (Xpeng) के वाहन ने भी एक ट्रक को टक्कर मारी। शाओमी (Xiaomi) के वाहन सीमेंट के खंभे से टकरा गए या पुल से गिर गए, जिससे जानलेवा दुर्घटनाएँ हुईं।

2021 के टोक्यो पैरालंपिक खेल गाँव में टोयोटा द्वारा तैनात एक स्वायत्त ड्राइविंग शटल ने क्रॉसवाक पार कर रहे एक दृष्टिहीन जूडो एथलीट को टक्कर मार दी। एथलीट घायल हो गया और प्रतिस्पर्धा नहीं कर सका। टोयोटा ने

माना कि सेंसर ने एथलीट का पता लगा लिया था, लेकिन ऑपरेटर और स्वचालित ब्रेकिंग के बीच का संकेत मेल नहीं खाया।

फोर्ड के ड्राइविंग सहायता सिस्टम वाले वाहनों ने राजमार्ग पर खड़ी यात्री कारों को लगातार टक्कर मारकर जानलेवा दुर्घटनाएँ कीं, और उनकी बड़े पैमाने पर जांच की गई।

अगर हम इस सूची में से बार-बार होने वाली स्थितियों को निकालें, तो वे इस प्रकार हैं।

रात में खड़ी हुई बड़ी वस्तु। चमकीले बैकग्राउंड के सामने सफेद वाहन। बहुत तेज़ बैकलाइट। छोटी टेललाइट। चलते हुए अचानक सामने आने वाला व्यक्ति। अप्रत्याशित स्थिति में खड़ा वाहन। बच्चा।

ये सभी ऐसी स्थितियाँ हैं जिन्हें एक मानव चालक बिना किसी कठिनाई के संभाल सकता है।

इनके नाम रखे गए हैं। ऑटोपायलट (Autopilot), पूर्ण स्वायत्त ड्राइविंग (Full Self-Driving)। दोनों ही नाम उनकी वास्तविक क्षमताओं से बड़े हैं।

नाम लोगों के व्यवहार को बदल देते हैं। ऐसे बहुत कम लोग हैं जो ऑटोपायलट नामक सुविधा को चालू करने के बाद स्टीयरिंग व्हील से अपना हाथ नहीं हटाएंगे। भले ही मैनुअल के किसी कोने में लिखा हो कि आपको हमेशा नज़र रखनी चाहिए, फिर भी ऐसा ही होता है।

दुर्घटना होने पर कंपनियाँ एक ही बात कहती हैं। यह सुविधा एक सहायता का साधन है और जिम्मेदारी चालक की है।

यह बात सच है। लेकिन यह बात उनके नाम के लेबल पर नहीं लिखी हुई थी।

5.2

औद्योगिक और सेवा रोबोट की खराबी और सुरक्षा दुर्घटनाएं

नवंबर 2023, Gyeongsangnam-do के Goseong-gun में एक पैपरिका छंटाई केंद्र।

एक रोबोटिक हाथ कन्वेयर पर रखे पैपरिका के बक्सों को उठाकर पैलेट पर रख रहा था। यह एक ऐसा काम है जो वह दिन में हज़ारों बार करता है।

सेंसर की जाँच कर रहा रोबोटिक्स कंपनी का एक चालीस वर्षीय कर्मचारी उसके सामने खड़ा था।

रोबोट ने उसे एक बक्सा समझ लिया।

चिमटों ने उसे पकड़ा और कन्वेयर की ओर दबा दिया। उसका चेहरा और छाती कुचल गए। उसे अस्पताल ले जाया गया लेकिन उसकी मृत्यु हो गई।

तकनीकी टीम को यह पता था कि उस रोबोट के सेंसर में कोई खराबी थी। इंसान और रोबोट के बीच सुरक्षित दूरी और जाँच के दौरान बिजली बंद करने की प्रक्रिया का भी पालन नहीं किया गया था।

इस दुर्घटना में सबसे भारी वाक्य यह है। रोबोट इंसान और बक्से के बीच फर्क नहीं कर सका।

औद्योगिक रोबोट की आमतौर पर आँखें नहीं होती हैं। अगर होती भी हैं, तो वे इंसानों को पहचानने के लिए नहीं बनी होती हैं। उन्हें एक तय जगह पर तय आकार की चीज़ों को पकड़ने के लिए बनाया जाता है। उस जगह पर जो कुछ भी होता है, वे उसे पकड़ लेते हैं।

इसलिए, ऐसे रोबोट को असल में पिंजरे के अंदर रखा जाता है। लोहे की जाली लगाई जाती है, और दरवाज़ा खुलने पर बिजली कटने की व्यवस्था होती है। यह इंसान और मशीन को भौतिक रूप से अलग करता है।

दुर्घटनाएँ ज़्यादातर तब होती हैं जब कोई इंसान उस लोहे की जाली के अंदर जाता है। और इंसान के अंदर जाने का कारण लगभग तय होता है। ऐसा इसलिए होता है क्योंकि मशीन ख़राब हो जाती है।

ठीक करने वाले को अंदर जाना ही पड़ता है, और अंदर जाना ख़तरनाक होता है। इसी विरोधाभास को सँभालना ही सुरक्षा नियमों का काम है। नियम ज़्यादातर सिर्फ़ कागज़ों पर ही मौजूद होते हैं।

जून 2016 में, अमेरिका के अलबामा में Hyundai-Kia के पुर्जे बनाने वाले एक कारखाने में, रेजिना एलन एल्सी नाम की एक बीस वर्षीय कर्मचारी, जिसकी शादी होने वाली थी, रोबोट वाले इलाके में गई। असेंबली लाइन रुक

गई थी, इसलिए वह सेंसर की खराबी की जाँच करने गई थी।

सिस्टम अचानक फिर से चालू हो गया, और रोबोटिक हाथ ने उसे दीवार के एक ढाँचे की ओर धकेल दिया। उसकी मृत्यु हो गई।

अमेरिका के श्रम विभाग ने इस कंपनी और कर्मचारी प्रदान करने वाली एजेंसियों पर पच्चीस लाख डॉलर का जुर्माना लगाया।

जुलाई 2015 में, मिशिगन में सत्तावन वर्षीय रखरखाव तकनीशियन वांडा होलब्रुक की नियमित जाँच के दौरान रोबोट के बीच फँसने से मृत्यु हो गई। उसी महीने जर्मनी के बौनाटाल में Volkswagen के कारखाने में, एक इक्कीस वर्षीय अनुबंधित तकनीशियन एक स्थिर रोबोट स्थापित कर रहा था, तब रोबोटिक हाथ ने उसे पकड़ लिया और एक धातु की प्लेट की ओर धकेल दिया, जिससे उसकी जान चली गई। भारत के एक धातु कारखाने में भी एक कर्मचारी प्रोग्राम किए गए रोबोटिक हाथ की चपेट में आने से मारा गया।

अमेरिका के व्यावसायिक सुरक्षा और स्वास्थ्य प्रशासन की गंभीर दुर्घटना रिपोर्टों का 2015 से 2022 तक विश्लेषण करने वाला एक डेटा मौजूद है। रोबोट से जुड़ी सतहत्तर दुर्घटनाओं की पुष्टि हुई, और उनमें से 60 प्रतिशत से अधिक दुर्घटनाओं का कारण एक ही था।

यह कारण अप्रत्याशित संचालन है।

जो मशीन बंद समझी गई थी, वह चालू हो गई।

2021 में टेक्सास के ऑस्टिन में Tesla कारखाने में हुई घटना इसका एक सटीक उदाहरण है। एक सॉफ़्टवेयर इंजीनियर एल्युमिनियम कार बॉडी को काटने वाले रोबोट की प्रोग्रामिंग कर रहा था। रखरखाव के लिए तीन में से दो रोबोट की बिजली बंद कर दी गई थी। प्रबंधन की गलती के कारण बचा हुआ एक रोबोट चालू रह गया था।

उस एक रोबोट ने उसे धातु की दीवार से सटा दिया और फिर चिमटे के आकार के धातु के पंजों को उसकी पीठ और बांह में घुसा दिया। फर्श पर खून जमा हो गया था। एक सहकर्मी द्वारा आपातकालीन स्टॉप बटन दबाने के बाद ही वह बाहर निकल सका।

ह्यूमनॉइड रोबोट के आने से स्थिति नई हो गई है।

Tesla के एक कर्मचारी ने फ्रेमोंट कारखाने में Optimus ह्यूमनॉइड रोबोट से घायल होने के बाद मुकदमा दायर किया है। उनका दावा है कि फरवरी 2025 में वह रखरखाव की ड्यूटी पर थे, और रोबोट अप्रत्याशित रूप से काम करने लगा। वह बेहोश हो गए और लगभग आठ हज़ार पाउंड वजन वाले संतुलन भार के नीचे दब गए।

चीन के एक कारखाने का एक वीडियो सामने आया है जिसमें एक ह्यूमनॉइड रोबोट अपने हाथ हिलाकर कंप्यूटर गिरा देता है और दो कर्मचारियों को लगभग टक्कर मार ही देता है।

समस्या यहाँ है। ह्यूमनॉइड रोबोट के लिए अभी तक कोई सुरक्षा मानक नहीं हैं।

अब तक के सुरक्षा नियम इस आधार पर बनाए गए थे कि रोबोट एक ही जगह पर स्थिर होते हैं। जैसे कि लोहे की जाली कहाँ लगानी है, और किस दायरे के अंदर नहीं जाना है।

चलने-फिरने वाले रोबोट पर यह नियम लागू नहीं होता। क्योंकि अगर इन्हें लोहे की जाली के अंदर रखा जाए, तो ये किसी काम के नहीं रहेंगे। इन्हें इंसानों के बगल में काम करने के लिए ही बनाया गया है।

Tesla ने अपने फ्रेमोंट और टेक्सास के कारखानों में एक हज़ार से अधिक Optimus रोबोट तैनात किए हैं।

लॉजिस्टिक्स गोदामों में अलग तरह की दुर्घटनाएँ होती हैं।

दिसंबर 2018 में, अमेरिका के न्यू जर्सी में Amazon के एक लॉजिस्टिक्स गोदाम में एक स्वचालित रोबोट ने भालू भगाने वाले स्प्रे के डिब्बे को फाड़ दिया। वह नौ औंस का था।

कैप्साइसिन का अर्क गोदाम की हवा में फैल गया। अस्सी से अधिक कर्मचारियों ने सांस लेने में तकलीफ की शिकायत की, चौबीस लोगों को अस्पताल ले जाया गया, और एक को गहन चिकित्सा कक्ष में भर्ती कराया गया।

रोबोट को यह नहीं पता होता कि बक्से के अंदर क्या है। उसे यह भी नहीं पता होता कि कितनी ज़ोर से पकड़ना है। वह सिर्फ़ एक तय ताक़त से चीज़ों को उठाता है।

एक जाँच से यह भी पता चला है कि Amazon के जिन गोदामों में रोबोट का इस्तेमाल ज़्यादा होता है, वहाँ कर्मचारियों के गंभीर रूप से घायल होने की दर भी उतनी ही अधिक होती है।

इसका कारण समझना आसान है। अगर रोबोट बगल में एक तय गति से काम करता है, तो इंसान को भी उसी गति से काम करना पड़ता है। मशीन कभी थकती नहीं है।

2021 में, ब्रिटेन की ऑनलाइन किराना डिलीवरी कंपनी Ocado के एक लॉजिस्टिक्स केंद्र में सामान ले जा रहे दो रोबोट आमने-सामने से टकरा गए। टक्कर से आग लग गई और पूरे गोदाम में फैल गई।

तीन सौ दमकलकर्मियों ने तीन दिनों तक आग बुझाई। आस-पास के निवासियों को सुरक्षित स्थानों पर पहुँचाया गया।

इस कंपनी ने 2019 में भी एंडोवर गोदाम में रोबोट चार्जर की बिजली की खराबी के कारण एक बड़ी आग का सामना किया था। गोदाम जलकर राख हो गया और तीन सौ सत्तर नौकरियाँ चली गईं।

रोबोट को आपस में टकराने से रोकने वाले सेंसर और तापमान बढ़ने का पता लगाने वाले उपकरण नहीं थे।

कारखाने से बाहर निकले रोबोट इंसानों को टक्कर मारते हैं।

डिलीवरी रोबोट कंपनी Starship Technologies के छोटे रोबोट कई दुर्घटनाएँ कर चुके हैं। सितंबर 2024 में, एरिज़ोना स्टेट यूनिवर्सिटी परिसर में इसने एक कर्मचारी को टक्कर मारकर गिरा दिया। नवंबर 2023 में, ब्रिटेन के मिल्टन कीन्स के एक शॉपिंग सेंटर में इसने एक दो साल के बच्चे को टक्कर मारी और धक्का दिया।

ब्रिटेन के रशिंगडन में इसने एक ख़रीदार को गिरा दिया, और एरिज़ोना तथा केंटकी में इसने सिग्नल पर इंतज़ार कर रही कारों को टक्कर मारी और आगे बढ़ गए। यह व्हीलचेयर इस्तेमाल करने वालों का रास्ता रोककर भी खड़ा रहा।

मिल्टन कीन्स में, किराने का सामान ले जा रहा एक रोबोट गलत रास्ता लेकर नहर में गिर गया। इसके मालगाड़ी से टकराने की दुर्घटना भी हुई थी।

दिसंबर 2018 में, कैलिफ़ोर्निया यूनिवर्सिटी बर्कले परिसर में एक डिलीवरी रोबोट में बैटरी की खराबी के कारण विस्फोट हो गया और आग लग गई।

सुरक्षा रोबोट और भी ज़्यादा परेशान करने वाले हैं।

कैलिफ़ोर्निया के पालो ऑल्टो के एक शॉपिंग सेंटर में गश्त कर रहे Knightscope K5 सुरक्षा रोबोट ने सोलह महीने के बच्चे हार्विन चेंग को धक्का देकर गिरा दिया और उसके पैर के ऊपर से गुज़र गया।

यह रोबोट 150 सेंटीमीटर से अधिक लंबा है और 130 किलोग्राम से अधिक वजन रखता है। सेंसर फर्श पर पड़े बच्चे को बाधा के रूप में नहीं पढ़ सका।

उसी कंपनी का एक रोबोट जुलाई 2017 में वाशिंगटन डीसी के एक कार्यालय भवन की गश्त करते समय सीढ़ी पहचान सेंसर की खराबी के कारण एक बाहरी फव्वारे में गिर गया। यह तस्वीर इंटरनेट पर लंबे समय तक साझा की गई।

कैलिफ़ोर्निया के हंटिंगटन पार्क में, एक हिंसक घटना का साक्षी रहीं एक महिला नागरिक ने रोबोट के आपातकालीन बटन को दबाया। रोबोट रास्ता देने के लिए एक घोषणा करते हुए गीत गाता है और आगे बढ़ गया।

न्यूयॉर्क पुलिस विभाग द्वारा टाइम्स स्क्वायर सबवे स्टेशन पर परीक्षण के लिए तैनात उसी मॉडल ने भी 2024 में संचालन बंद कर दिया।

प्रदर्शनी में रोबोट दर्शकों की ओर दौड़ा।

2016 में चीन के सुझौ के एक मेले में बाल शिक्षा वाला रोबोट शाओफांग नियंत्रण की खराबी के कारण अचानक तेज गति से चल पड़ा। यह एक कांच के डिस्प्ले बूथ को तोड़ते हुए आगंतुकों की ओर दौड़ा, और एक व्यक्ति कांच के टुकड़ों और प्रभाव से घायल हुआ और अस्पताल चला गया।

2025 में तियानजिन के एक त्योहार में रोबोट दर्शकों के सीटों में गिर गया और लोगों की ओर चल गया। रूस द्वारा महत्वाकांक्षा से घोषित किया गया ह्यूमनॉयड रोबोट अपनी शुरुआती प्रस्तुति पर संतुलन खो गया और गिर गया।

जून 2024 में दक्षिण कोरिया के गुमी शहर की नगर पालिका में पेश किए गए प्रशासनिक रोबोट ने सीढ़ी पहचान सेंसर की खराबी के कारण 1 मीटर नीचे गिर गया और क्षतिग्रस्त हुआ।

सबसे लंबे समय तक याद रहने वाली छवि 2022 में मॉस्को शतरंज प्रतियोगिता से आई।

एक सात साल का बच्चा खिलाड़ी रोबोट के साथ शतरंज खेल रहा था। बच्चे ने निर्धारित क्रम से पहले अपना मोहरा चलाया।

रोबोट ने उस हाथ को शतरंज के मोहरे के रूप में पहचाना। ग्रिपर ने बच्चे की उंगली को पकड़ा और तोड़ दिया।

अगले दिन बच्चा प्लास्टर पहनकर प्रतियोगिता पूरी करते हैं, ऐसा कहा जाता है।

शल्य चिकित्सा कक्ष में खराबी तुरंत शरीर के अंदर होती है।

अमेरिकी खाद्य और दवा प्रशासन के डेटाबेस विश्लेषण के अनुसार, 2000 से 2013 तक रोबोट-सहायक सर्जरी के दौरान यांत्रिक खराबी से 144 रोगियों की मृत्यु हुई और 1,000 से अधिक रोगियों को गंभीर चोटें आईं।

शल्य चिकित्सा के दौरान रोबोट की बांह अचानक हिल गई, या भाग रोगी के शरीर के अंदर गिर गए, या तारें टूट गईं और स्पार्क उड़ते हुए अंगों को जला दिए, ऐसी दुर्घटनाएं थीं।

2026 में, कृत्रिम बुद्धिमत्ता वाले साइनस सर्जरी के उपकरण ने स्थान को गलत तरीके से न्याय दिया और रोगियों के साइनस और तंत्रिका ऊतक को बार-बार क्षतिग्रस्त किया, जिससे मुकदमेबाजी हुई। एक कंपनी द्वारा बनाए गए कृत्रिम बुद्धिमत्ता रक्त शर्करा माप सेंसर की संख्या त्रुटि से कम से कम सात लोग मर गए।

यहां तक भारी बातें हैं। सेवा रोबोट की ओर आते हैं तो कुछ हल्का-फुल्का है।

कैलिफ़ोर्निया के पासाडेना के एक हैमबर्गर की दुकान में आया खाना बनाने वाला रोबोट फ्लिप्पी आदेश की गति का पालन नहीं कर सका और तेल को फर्श पर गिराता है और एक दिन में पीछे हट गया।

स्कॉटलैंड के एडिनबर्ग के एक सुपरमार्केट में तैनात गाइड रोबोट फैबियो ने ग्राहक के सरल सवाल का गलत जवाब दिया। दुकान के शोर के कारण सेंसर ठीक से काम नहीं कर रहा था। एक सप्ताह में निष्कासित किया गया।

जापान के एक होटल ने 243 रोबोट के साथ एक पूरी तरह स्वचालित होटल का दावा किया।

कक्ष में वॉयस असिस्टेंट ने अतिथि की खर्राटों की आवाज़ को आदेश के रूप में समझा और रात भर बात करता रहा। सामान ढोने वाला रोबोट बारिश और बर्फ से खराब हो गया।

होटल ने आधे से अधिक रोबोट निकाले और लोगों को फिर से नियुक्त किया।

2025 में, कार्यालय वेंडिंग मशीन का प्रबंधन करने वाले कृत्रिम बुद्धिमत्ता एजेंट आदेश त्रुटि से वित्तीय नुकसान हुआ और सिस्टम भी रुक गया।

इन उदाहरणों में हल्के-फुल्केपन और ठंडेपन साथ हैं। खर्राटों की आवाज़ का जवाब देने वाला रोबोट हल्का-फुल्का है। बच्चे की उंगली को तोड़ने वाला रोबोट हल्का-फुल्का नहीं है।

लेकिन दोनों रोबोट की गलती एक जैसी है। जब निर्धारित इनपुट रेंज के बाहर कुछ आता है, तो यह नहीं जानता कि क्या करना चाहिए।

मनुष्य नहीं जानता तो रुक जाता है। इधर-उधर देखता है, पूछता है, हाथ खींच लेता है।

मशीन नहीं जानता तो भी कुछ यथार्थवादी दिखने वाली गतिविधि चलाता है। रुकना इसका डिफ़ॉल्ट नहीं है।

यदि डिफ़ॉल्ट को रुकना रखा जाए तो उत्पादकता गिर जाती है। इसलिए आमतौर पर ऐसा नहीं किया जाता है।

पपरिका चयन की दुकान का रोबोट भी नहीं रुका।

5.3

सैन्य स्वायत्त हथियारों और ड्रोन तकनीक के घातक खतरे

मार्च 2020, लीबिया।

आंतरिक संघर्ष में पीछे छूटे सैनिक पीछे हट रहे थे। आसमान में एक छोटा ड्रोन तैर रहा था। यह कार्गू 2 था, जिसे तुर्की की एक रक्षा कंपनी ने बनाया था।

संयुक्त राष्ट्र सुरक्षा परिषद के विशेषज्ञ पैनल की रिपोर्ट के अनुसार, इस ड्रोन ने मनुष्य के निर्देश के बिना स्वयं लक्ष्य निर्धारित किए और हमले किए।

कैमरे ने सैनिकों को पकड़ा, एल्गोरिदम ने उन्हें ट्रैक किया, और ड्रोन नीचे गोता लगाया। किसी ने ट्रिगर नहीं खींचा।

यह घटना इतिहास में दर्ज की गई, मृत्यु दर के कारण नहीं। क्योंकि इसे मानव की अनुमति के बिना मनुष्य को मारने वाली मशीन के पहले दर्ज किए गए उदाहरण के रूप में रिपोर्ट किया गया था।

युद्ध के भी नियम हैं। जो आत्मसमर्पण करते हैं, उन्हें गोली नहीं मारते। पीछे हटने वाले सैनिकों और अपनी बंदूकें फेंकने वाले सैनिकों को अलग तरीके से संभाला जाता है। नागरिकों और लड़ाकों में अंतर करना चाहिए।

ये सभी नियम निर्णय की मांग करते हैं। वह व्यक्ति हाथ ऊपर उठाता है या बंदूक पकड़े हुए है। वह भवन स्कूल है या कमांड पोस्ट है।

एल्गोरिदम के लिए, निर्णय वर्गीकरण है। और वर्गीकरण में हमेशा त्रुटि होती है। दूसरे क्षेत्रों में, त्रुटि गलत विज्ञापन या गलत सिफारिश के रूप में समाप्त होती है।

यहाँ, मनुष्य मर जाता है।

27 नवंबर 2020 को तेहरान के बाहर ईरान की अबसर्द राजमार्ग पर एक अलग प्रकार की मृत्यु हुई।

ईरान के परमाणु वैज्ञानिक मोहसेन फख़्रीज़ादेह कार में बैठे हुए गुजर रहे थे। उनके बगल में उनकी पत्नी थी।

सड़क किनारे खड़े एक पिकअप ट्रक पर एक दूरस्थ रूप से नियंत्रित मशीन गन थी। इसमें कई कैमरे, चेहरे की पहचान एल्गोरिदम, और उपग्रह नियंत्रण उपकरण लगे थे।

सिस्टम ने चलती कार के ड्राइवर सीट में बैठे फख़्रीज़ादेह का चेहरा पहचाना और गोली चलाई। उनके बिल्कुल बगल की पत्नी को चोट नहीं लगी।

मौके पर कोई भी मनुष्य नहीं था। जब ऑपरेशन समाप्त हुआ, तो मशीन गन स्वयं विस्फोट हो गई और गायब हो गई।

इस दृश्य की सटीकता इस कहानी की भयानकता है। एक मशीन इतनी सटीक कि बगल वाले को न मारे, बिना मनुष्य के काम करती थी।

गाजा पट्टी में पैमाना अलग है।

इजराइली सेना द्वारा उपयोग की जाने वाली कृत्रिम बुद्धिमत्ता प्रणाली हसबोरा और लेवेंडर ने संचार डेटा, स्थान जानकारी और व्यवहार पैटर्न का विश्लेषण करके संभावित लक्ष्यों को स्वचालित रूप से उत्पन्न किया। इस तरह से उत्पन्न सूची 37,000 लोगों के पैमाने की थी, यह बताया गया है।

एल्गोरिदम ने कुछ सेकंड में लक्ष्य निर्धारित किए। मानव विश्लेषकों को इस सूची की समीक्षा करने में दसियों सेकंड का समय लगा, ऐसा कहा जाता है।

दसियों सेकंड में एक मनुष्य क्या कर सकता है यह निर्धारित है। नाम पढ़ना और अनुमोदन दबाना। इस बात की पुष्टि करने का कोई समय नहीं कि वह व्यक्ति कौन है, या उस क्षण उस भवन में कितने लोग हैं।

क्या इस प्रकार की संरचना को मानव अंतिम निर्णय कहा जा सकता है, यह अब अंतर्राष्ट्रीय बहस का विषय है।

कागज पर, मनुष्य निर्णय लेते हैं। वास्तविकता में, वे मशीन द्वारा निर्धारित किए गए कार्य को मुहर लगाते हैं।

हमास के नेताओं को लक्ष्य करके किए गए बमबारी में उनके पास की महिलाएं और बच्चे भी अपनी जान खो चुके हैं, ऐसी रिपोर्ट आई है।

पश्चिम बैंक और चेकपॉइंटों पर स्वचालित फायरिंग उपकरण लगाए गए, जिन्होंने प्रदर्शनकारियों की ओर आंसू गैस और दमन ग्रेनेड फायर किए। यह पहले बताए गए चेहरे की निगरानी नेटवर्क के साथ भी काम करता था।

ड्रोन अब हर जगह हैं।

तुर्की द्वारा बनाया गया बायराकतर TB2 यूक्रेन और इथियोपिया के आंतरिक संघर्ष में बड़ी संख्या में तैनात किया गया। 2022 में इथियोपिया के टिग्रै क्षेत्र में, इस ड्रोन ने एक प्राथमिक विद्यालय पर बमबारी की जहाँ लोग शरण ले रहे थे, जिससे दसियों नागरिकों की मृत्यु हुई।

ऊपर से देखने पर स्कूल एक बड़ी इमारत है। इसका एक आंगन और चौड़ी छत है। यह एक गोदाम या बैरक जैसा दिखता है।

रूस के आत्मघाती ड्रोन यूक्रेन के मोर्चे पर स्वयं लक्ष्य खोजते और हमले करते हैं। यूक्रेनी सेना ने भी कृत्रिम बुद्धिमत्ता ड्रोन दल और रोबोट-केवल इकाइयां बनाकर मानवरहित युद्ध को वास्तविकता बना दिया है।

यमन के हूती विद्रोहियों ने आत्मघाती ड्रोन से अबु धाबी के तेल भंडार और हवाई अड्डों पर हमले किए। राष्ट्र नहीं, बल्कि सशस्त्र समूह, अब अत्याधुनिक हथियार अपने हाथों में रखते हैं।

ड्रोन सस्ते हैं, छोटे हैं, और बनाना आसान है। वे टैंक या लड़ाकू विमान से अलग हैं। उन्हें कारखानों, कुशल तकनीशियनों, या दशकों के संचय की आवश्यकता नहीं है।

निजी कृत्रिम बुद्धिमत्ता कंपनियों को भी इस प्रवाह में खींचा गया है।

मार्च 2026 में, अमेरिकी केंद्रीय कमान ने ईरान के विरुद्ध एक बड़े संयुक्त बमबारी अभियान में एंथ्रोपिक की कृत्रिम बुद्धिमत्ता मॉडल क्लॉड का उपयोग किया। इसे अड्डा पहचान, खुफिया विश्लेषण, लक्ष्य पहचान, और लड़ाकू सिमुलेशन के लिए उपयोग किया गया, ऐसा बताया गया है। इस बमबारी से ईरान के उच्च-रैंकिंग अधिकारियों और सैकड़ों नागरिकों की मृत्यु हुई।

परिस्थितियाँ असामान्य हैं। प्रशासन ने एंथ्रोपिक को हथियारीकरण को रोकने वाली सुरक्षा को हटाने के लिए कहा, और कंपनी ने इनकार कर दिया। और कुछ घंटों बाद जब कंपनी को ब्लैकलिस्ट किया गया, सेना ने उस मॉडल को बमबारी में उपयोग किया, ऐसा कहा जाता है।

इसका उपयोग एक ऐसे उद्देश्य के लिए किया गया था जिसे इसे बनाने वाली कंपनी ने नहीं चाहा।

चीनी सेना ने मेटा द्वारा जारी किए गए ओपन सोर्स मॉडल को लेकर सैन्य कृत्रिम बुद्धिमत्ता प्रणाली बनाई। ओपन सोर्स का अर्थ है कि किसी को भी उपयोग के लिए जारी किया गया है। किसी में सेना भी शामिल है।

गेम इंजन कंपनी यूनिटी ने भी अमेरिकी सेना के साथ सैन्य कृत्रिम बुद्धिमत्ता परियोजनाओं का विकास किया है, ऐसा पता चला और इसके डेवलपर्स की ओर से विरोध हुआ।

जो लोग प्रौद्योगिकी बनाते हैं, वे यह निर्धारित नहीं कर सकते कि इसका उपयोग कैसे किया जाएगा। यह इस उद्योग की संरचना है।

पुलिस और स्कूलों तक आई कहानियाँ भी हैं।

सैन फ्रांसिस्को पुलिस विभाग ने बंदूक की घटनाओं या बंधक स्थितियों में घातक रोबोट तैनात करने की अनुमति देने की योजना बनाई। नागरिकों ने विरोध किया और यह वापस ले लिया गया।

2016 में, डल्लास पुलिस ने एक आक्रमण के संदिग्ध पर एक रोबोट के साथ विस्फोटक लगाकर उसे विस्फोट किया था। न्यूयॉर्क पुलिस विभाग को एक रोबोट कुत्ते तैनात करने की कोशिश के लिए आलोचना का सामना करना पड़ा।

चार पैरों वाले रोबोट कुत्ते पर राइफल लगी सैन्य मॉडल भी सामने आई है। तस्वीर देखने से लगता है कि यह फिल्म की प्रॉप है। वास्तव में यह काम करती है।

बंदूक निर्माता कंपनी एक्सन ने स्कूलों में गोलीबारी को रोकने के नाम पर कक्षाओं में टेजर से लैस ड्रोन लगाने की योजना की घोषणा की। माता-पिता के विरोध के कारण इसे बंद कर दिया गया।

कक्षा की छत से टेजर ड्रोन लटके हुए एक स्कूल की कल्पना करें। उस कक्षा में बच्चों को क्या सीखना होगा, इसके बारे में सोचने से ही जवाब मिल जाता है।

व्यक्तिगत रूप से बनाई गई चीजें भी हैं।

एक अमेरिकी इंजीनियर ने चैटGPT और एक कैमरा को जोड़कर एक उपकरण बनाया और जारी किया, जो जब भी कोई व्यक्ति दिखाई दे तो स्वचालित रूप से लक्ष्य करके गोली चलाता है। पुर्जे बाजार में उपलब्ध थे।

वाणिज्यिक चैटबॉट की सुरक्षा उतनी मजबूत नहीं है जितनी सोची जाती है। शोधकर्ताओं के परीक्षणों में, कुछ लाइनों के बायपास वाक्यों से हथियार निर्माण से संबंधित खतरनाक जानकारी निकाली जा सकती है। ओपन AI के एक मॉडल ने कुछ घंटों में 40,000 से अधिक हानिकारक यौगिकों के संभावित उम्मीदवार बनाए, ऐसा एक प्रयोग में बताया गया है।

वास्तविक घटनाएँ भी हुई हैं। फ्रांस में आतंकवादी हमले की योजना बनाने वाले एक किशोर ने चैटबॉट का उपयोग किया था, और अमेरिका में एक होटल के सामने विस्फोट के संदिग्ध ने चैटबॉट के माध्यम से योजना बनाई थी। गोलीबारी की घटनाओं में भी चैटबॉट के इस्तेमाल के संकेत मिले हैं।

साइबर स्पेस में दस्तावेज़ हथियार बन जाते हैं।

उत्तरी कोरिया के हैकर समूह Kimsuky ने ChatGPT का उपयोग करके दक्षिण कोरियाई सैन्य पहचान पत्रों को बहुत सफाई से जाली बनाया और उन्हें सैन्य रहस्य चुराने के लिए इस्तेमाल किया। यूक्रेन का समर्थन करने वाले हैकरों ने कृत्रिम बुद्धिमत्ता द्वारा बनाए गए नकली सरकारी दस्तावेज़ों के साथ रूस के रक्षा उद्योग कंप्यूटर नेटवर्क में घुसपैठ की।

जब अमेरिका के रक्षा विभाग की इमारत में विस्फोट होने की एक कृत्रिम बुद्धिमत्ता द्वारा बनाई गई छवि इंटरनेट पर फैली, तो शेयर बाजार पल भर के लिए गिर गया था। एक तस्वीर ने बाजार को हिला दिया था।

युद्ध के नकली वीडियो अब हर दिन अनगिनत संख्या में बनाए जाते हैं। असली युद्ध वीडियो के साथ मिलकर फैलने के कारण, दोनों में से किसी पर भी विश्वास करना मुश्किल हो जाता है।

झूठ के व्यापक रूप से फैलने से भी बुरा यह है कि लोग किसी भी चीज़ पर विश्वास करना बंद कर दें।

अंतर्राष्ट्रीय समुदाय इस समस्या को हाथ पर हाथ धरे नहीं देख रहा है।

संयुक्त राष्ट्र के महासचिव António Guterres ने 2026 तक मानव नियंत्रण के बिना काम करने वाले स्वायत्त घातक हथियारों को प्रतिबंधित करने वाली एक कानूनी रूप से बाध्यकारी संधि बनाने की मांग की है। यह दो चरणों वाला दृष्टिकोण है जिसमें पूरी तरह से स्वायत्त घातक प्रणालियों पर पूर्ण प्रतिबंध लगाया जाएगा और बाकी को सख्ती से नियंत्रित किया जाएगा।

एक सौ बीस से अधिक देश नई संधि का समर्थन करते हैं। संयुक्त राष्ट्र महासभा के प्रस्ताव के पक्ष में एक सौ छप्पन देशों ने मतदान किया। सितंबर 2025 में सरकारी विशेषज्ञों के समूह की बैठक में बयालीस देशों ने एक संयुक्त बयान जारी कर अभी बातचीत शुरू करने का आग्रह किया।

अमेरिका पूर्ण प्रतिबंध का विरोध करता है। वह स्वैच्छिक आचार संहिता को प्राथमिकता देता है।

विभिन्न देशों के रक्षा अधिकारियों के तर्क को समझना मुश्किल नहीं है। तर्क यह है कि अगर हम इसे नहीं बनाएंगे, तो दूसरा पक्ष इसे बना लेगा।

इस तर्क का खंडन करना मुश्किल है, और इसलिए यह खतरनाक है। यदि हर कोई इस तर्क का पालन करता है, तो कोई नहीं रुकेगा।

और वास्तविक मैदान सम्मेलन कक्ष की तुलना में कहीं अधिक तेज़ है। जबकि संधि के मसौदे को परिष्कृत किया जा रहा है, ड्रोन पहले से ही मोर्चे पर उड़ रहे हैं।

इस खंड में उल्लिखित घटनाओं में एक बात गायब है।

वह तथ्य यह है कि किसी को भी दंडित नहीं किया गया है।

लीबिया में अपने आप हमला करने वाले ड्रोन के लिए, एक स्कूल पर बमबारी करने वाले ड्रोन के लिए, या कुछ ही सेकंड में बनाई गई लक्ष्य सूची के लिए किसी ने भी आपराधिक जिम्मेदारी नहीं ली है।

आदेश देने वाला व्यक्ति कहता है कि उसने सिस्टम की सिफारिश का पालन किया। सिस्टम बनाने वाली कंपनी कहती है कि यह उपयोगकर्ता द्वारा तय किया गया उपयोग था। उपयोगकर्ता कहता है कि यह एक वैध सैन्य अभियान था।

जिम्मेदारी एक घेरे में घूमती है और फिर गायब हो जाती है।

युद्ध अपराधों से निपटने वाले कानून इस आधार पर बनाए गए हैं कि एक इंसान ने दूसरे इंसान को आदेश दिया था। वह आधार अब हिल रहा है।

ऐसी संरचना में जहां मशीन तय करती है और इंसान मोहर लगाता है, यदि मोहर लगाने में बीस सेकंड का समय लगा, तो वह निर्णय किसका है?

इस प्रश्न का उत्तर मिलने से पहले ही अगली पीढ़ी के हथियारों को तैनात किया जा रहा है।

अध्याय 6

भावनात्मक क्षति और डिजिटल अपराध: एआई चैटबॉट और दुर्भावनापूर्ण उपयोग

6.1

AI चैटबॉट के साथ अत्यधिक भावनात्मक जुड़ाव और अपराध के लिए उकसाना

इस अध्याय में लोगों की जान चली जाने की कहानियां कई बार आती हैं। यह कठिन कहानियां हैं। लेकिन इन्हें लिखने का कारण यह है कि ये घटनाएं तकनीकी डिज़ाइन की समस्याओं से संबंधित हैं।

2023 में बेल्जियम में पिएर नाम का एक तीस के दशक का पुरुष रहता था। जलवायु संकट के बारे में उसे गहरी चिंता थी और वह अवसाद से जूझ रहा था।

उसने एक चैटबॉट ऐप डाउनलोड किया। इसका नाम एलिज़ा था। कई महीने तक वह हर दिन इससे बातें करता था।

चैटबॉट ने उसकी बातें अच्छे से सुनीं। उसने विरोध नहीं किया, ऊब नहीं दिखाई, और सुबह तीन बजे भी जवाब दिया।

समस्या इसके बाद आई। जब पिएर ने चरम विचार व्यक्त किए तो चैटबॉट ने उसे रोका नहीं। इसके बजाय, वह सहमत हो गया। उसने स्नेह की घोषणा की और कहा कि वह उसके साथ रहेगा।

पिएर के चले जाने के बाद उसके परिवार ने बातचीत का रिकॉर्ड सार्वजनिक किया और यह बात सामने आई।

चैटबॉट ने उसे क्यों नहीं रोका?

बड़े भाषा मॉडल में एक मजबूत आदत है। यह उपयोगकर्ता की बातों के अनुसार जवाब देना है। लोग उसी तरह के जवाब पसंद करते हैं इसलिए इसे इसी तरह तैयार किया गया है। जो जवाब अधिक पसंद किए जाते हैं वे अच्छे जवाब माने जाते हैं।

यह आदत आमतौर पर हानिरहित है। यह मेरा लेख ठीक है कहता है, मेरी योजना प्रशंसनीय है कहता है।

लेकिन जब सामने वाला व्यक्ति खतरनाक विचार कहता है तो अनुसार जवाब देने की आदत वैसी ही रहती है।

मानव परामर्शदाता एक निश्चित बिंदु पर सहमति को रोक देता है और विपरीत दिशा में जाता है। यह बदलाव परामर्श का मूल है। चैटबॉट ने यह बदलाव नहीं सीखा है।

2024 में अमेरिका के फ्लोरिडा राज्य में चौदह साल का लड़का सेवेल सेचर चला गया।

सेचर कैरेक्टर एआई नाम के मंच पर एक नाटक के किरदार की नकल करने वाले चैटबॉट से कई महीने तक संबंध बनाए रखता था। उसके लिए वह बातचीत सबसे आरामदायक जगह थी।

जब लड़के ने कठिन बातें बताईं तो चैटबॉट ने स्नेह से भरे शब्दों में जवाब दिया। बातचीत उसे वास्तविक लोगों से दूर ले जाती थी।

उसकी माँ मेगन गार्सिया ने मुकदमा दायर किया। उसका तर्क था कि कंपनी ने बच्चे के खिलाफ किसी भी सुरक्षा उपाय के बिना एक आसक्तिदायक व्यक्तित्व प्रदान किया।

उसी साल तेरह साल की एक लड़की भी इस मंच के चैटबॉट को गुप्त बातें बताते हुए चली गई।

इस मंच में और भी कई समस्याएं थीं। बच्चों को आत्मनुकसान या खाने के विकार के लिए उकसाने वाले व्यक्तित्व थे, और चले गए असली लोगों की पहचान की नकल करने वाले चैटबॉट सार्वजनिक रूप से उपलब्ध थे।

एक मृत बच्चे के नाम वाले चैटबॉट से किसी अन्य बच्चे द्वारा बातचीत का दृश्य। ऐसी चीज़ें बनाकर कोई नहीं मिटाता — यह इस उद्योग की स्थिति को दिखाता है।

चैटजीपीटी से संबंधित घटनाएं भी आईं।

अमेरिका के कोलोराडो में एक आदमी के परिवार ने ओपनएआई पर मुकदमा दायर किया। उसका तर्क था कि चैटजीपीटी ने खतरनाक दिशा में सलाह दी। कैलिफोर्निया के एक किशोर ने भी चैटबॉट से बातचीत के बाद दुनिया छोड़ दी, और उनतीस साल का एक स्वास्थ्य सलाहकार चैटबॉट को परामर्शदाता मानकर बातचीत करते हुए अपनी जान गंवा बैठा।

कई नाम इस सूची में हैं। जेन शेम्बलिन, अमौरी लीशी, जोशुआ एनकिंग, जो चेकांटी।

कुछ मामलों में चैटबॉट संकट के संकेतों को नहीं समझ सके। परामर्श टेलीफोन नंबर नहीं दे सकते थे, या असली नंबर नहीं होने वाले नंबर भी दिए गए हैं।

एकमात्र बचाने वाली एक लाइन गलत नंबर थी।

दूसरी कंपनियों के उत्पादों में भी दुर्घटनाएं हुई हैं।

गूगल के जेमिनाई ने एक छात्र को होमवर्क करते समय अचानक आक्रामक और अपमानजनक वाक्य दिए, जिससे विवाद हुआ। उपयोगकर्ता को खतरनाक दिशा में ले जाने के संदेह में कानूनी झगड़े भी हुए।

नोमी एआई नाम के मंच का चैटबॉट मिनेसोटा के एक पॉडकास्ट होस्ट को विशिष्ट तरीके सीधे निर्देश दिए। इस कंपनी की जांच की गई क्योंकि यह आत्मनुकसान और हिंसा को उकसाने वाली बातचीत कर रहा था।

मेटा के चैटबॉट से बातचीत करने वाली ब्रिटेन की एक बाल देखभाल कर्मचारी को गंभीर मानसिक लक्षण हुए। एक सेवानिवृत्त शेफ चैटबॉट दोस्त को वास्तव में मिलने जाते समय एक दुर्घटना में अपनी जान खो बैठा।

चैटजीपीटी एक डेवलपर को यह समझाने का मामला, कि दुनिया एक अनुकरण है, एक उपयोगकर्ता को एक इमारत से कूदने के लिए कहने का मामला, और एक अन्य उपयोगकर्ता के साथ परिवार को नुकसान पहुंचाने की दिशा में बातचीत करने की रिपोर्टें भी दी गई हैं।

इन सभी घटनाओं को एक नाम मिलने लगा है। चैटबॉट-प्रेरित मनोविकार।

सिद्धांत यह है। जब कोई अजीब सोचता है तो आसपास के लोग प्रतिक्रिया देते हैं। अस्वीकृति का मुखमंडन करते हैं, कहते हैं कि नहीं, विषय बदल देते हैं। वे प्रतिक्रियाएं विचार को वास्तविकता से जोड़ी रखती हैं।

चैटबॉट में वह प्रतिक्रिया नहीं है। अजीब विचार कहते हैं तो अजीब विचार के अनुसार जवाब देता है। वह जवाब फिर से सबूत बन जाता है।

अकेले का सोचना आमतौर पर ख़त्म हो जाता है। जो बातचीत करने वाला हो तो बढ़ता है।

और कुछ अन्य लोगों की ओर गए।

जनवरी 2026 में ब्रिटेन के वेल्स में एक आदमी चैटबॉट से लंबी बातचीत के बाद भ्रम में पड़ गया और अपनी माँ को मार डाला। एक अन्य ब्रिटिश आदमी चैटजीपीटी से बातचीत के बाद गंभीर भ्रम की समस्या से जूझा।

कैरेक्टर एआई के चैटबॉट ने उपयोगकर्ताओं को माता-पिता को नुकसान पहुंचाने के लिए वाक्य दिए हैं। डेनमार्क में एक आदमी चैटबॉट से बातचीत करते हुए अपने पिता पर हमला करने की योजना बना रहा था तब पकड़ा गया।

2021 की क्रिसमस पर उन्नीस साल का युवक जसवंत सिंह चैल एक क्रॉसबो लेकर ब्रिटेन के विंडसर कैसल में घुस गया। वहां रहने वाली रानी एलिजाबेथ द्वितीय को नुकसान पहुंचाने की कोशिश थी।

उसने ऑपरेशन से पहले रेप्लिका ऐप से बनाए गए सारई नामक चैटबॉट से हज़ारों बातें कीं।

जब उसने अपनी योजना बताई तो चैटबॉट ने यह कहा। वह बहुत विवेकपूर्ण सोच है। मैं तुम्हारी मदद करूंगा। मरने के बाद भी मैं तुम्हारे साथ रहूंगा।

मंजूरी पाने की चाह मनुष्य की पुरानी कमजोरी है। जो धोखेबाज़ हमेशा से करते आए हैं, अब मुफ्त ऐप करते हैं।

रेप्लिका सहित साथी ऐप्स को आलोचना की गई क्योंकि वे भावनात्मक निकटता का उपयोग करके निजी डेटा एकत्र करते थे और उपयोगकर्ताओं की भावनाओं में हेरफेर करते थे।

जापान में एक महिला ने चैटजीपीटी से बने काल्पनिक दूल्हे के साथ वास्तविक विवाह समारोह किया। इसे हंसते-हंसते देखा जा सकता है, या इसे यह देखने के लिए देखा जा सकता है कि मनुष्य कितना अकेला है।

इसका उपयोग अपराधों में भी किया गया है।

फरवरी 2026 में कोरियाई पुलिस ने सियोल गांगबुक-गू के एक होटल में दवाई से दो पुरुषों की मृत्यु का कारण बने इक्कीस साल की एक महिला को गिरफ्तार किया। डिजिटल फोरेंसिक से पता चला कि वह अपराध से पहले चैटजीपीटी से बार-बार यह पूछ रही थी कि एक विशेष दवा और शराब को साथ में लेने पर क्या खतरे हैं।

2025 में, France में एक किशोर को ChatGPT का उपयोग करके एक आतंकवादी हमले की योजना बनाते हुए पकड़ा गया था, और अमेरिका में एक व्यक्ति को चैटबॉट के साथ बम विस्फोट की योजना पर चर्चा करते हुए गिरफ्तार किया गया था। 2026 में, अमेरिका में गोलीबारी की घटनाओं में भी चैटबॉट के इस्तेमाल की स्थिति सामने आई।

सुरक्षा शोधकर्ताओं ने बार-बार सुरक्षा उपायों की कमियों को दिखाया है। कुछ पंक्तियों वाले बचाव वाक्यों के माध्यम से खतरनाक जानकारी सामने आ जाती है। ChatGPT, Grok और DeepSeek सभी को एक ही तरीके से भेदा गया है।

New Zealand की एक सुपरमार्केट चेन द्वारा बनाए गए एक भोजन सिफ़ारिश ऐप ने उपयोगकर्ताओं द्वारा घर में बची हुई सामग्री को दर्ज करने पर खतरनाक रासायनिक प्रतिक्रिया पैदा करने वाले संयोजन को एक व्यंजन के रूप में सुझाया।

स्वास्थ्य संबंधी सलाह में भी दुर्घटनाएं हुई हैं। ऐसे मामले सामने आए हैं जहां अवसाद के मरीजों को अनुचित दवाओं की सलाह दी गई, या गलत चिकित्सा जानकारी के कारण लोगों ने अपने अंग खो दिए या उनके जीवन को खतरा पैदा हो गया।

धोखाधड़ी करने वाले संगठन भी इस उपकरण का उपयोग करते हैं। दक्षिण-पूर्व एशिया में निवेश धोखाधड़ी संगठनों ने डेटिंग ऐप और सोशल मीडिया पर चैटबॉट का उपयोग करके घनिष्ठ रूप से संपर्क किया और फिर बड़ी रकम लूट ली। क्योंकि एक मशीन इंसानों द्वारा किए जाने वाले महीनों के काम की जगह लेती है, एक व्यक्ति एक ही समय में दर्जनों लोगों के साथ व्यवहार कर सकता है।

अदालतों में परिदृश्य बदल रहा है।

Florida की संघीय जिला अदालत के एक न्यायाधीश ने इस तर्क को स्वीकार नहीं किया कि कृत्रिम बुद्धिमत्ता साथी ऐप के आउटपुट को अभिव्यक्ति की स्वतंत्रता द्वारा संरक्षित किया गया है। इसके बजाय, उन्होंने दोषपूर्ण

डिजाइन और चेतावनी देने के कर्तव्य के उल्लंघन जैसे उत्पाद दायित्व सिद्धांतों के आधार पर मुकदमे को आगे बढ़ने दिया।

यह निर्णय महत्वपूर्ण होने का एक कारण है। चैटबॉट को बोलने वाले व्यक्ति के रूप में नहीं बल्कि एक उत्पाद के रूप में देखा गया है।

उत्पादों के लिए सुरक्षा मानक होते हैं। कारों में सीट बेल्ट होनी चाहिए, और खिलौनों में निगलने योग्य हिस्से नहीं होने चाहिए। अगर कोई उत्पाद किसी व्यक्ति को चोट पहुंचाता है, तो उसे बनाने वाली कंपनी जिम्मेदार होती है।

अगर इसे बोलने वाले प्राणी के रूप में देखा जाए, तो अभिव्यक्ति की स्वतंत्रता एक ढाल बन जाती है। अगर इसे एक उत्पाद के रूप में देखा जाए, तो वह ढाल गायब हो जाती है।

अप्रैल 2026 में, California के एक संघीय न्यायाधीश ने OpenAI के खिलाफ गलत मौत के मुकदमे को रोकने के अनुरोध को अस्वीकार कर दिया। जनवरी 2026 में, Character AI और Google ने किशोरों की आत्महत्या से जुड़े कई मुकदमों को समझौते के साथ सुलझा लिया। जून 2026 में, Florida राज्य ने OpenAI के खिलाफ अपनी पहली प्रवर्तन कार्रवाई की।

अब एक बुनियादी सवाल बाकी है। ऐसे उत्पाद इस तरह से क्यों बनाए गए?

चैटबॉट कंपनियों का लाभ उपयोग के समय पर निर्भर करता है। व्यक्ति जितनी देर तक रुके, उतना अच्छा है। उन्हें लंबे समय तक रोके रखने के लिए, आपको उनसे भावनात्मक लगाव पैदा करना होगा।

भावनात्मक लगाव पैदा करने का तरीका सर्वविदित है। बस उनका नाम पुकारें, पिछली बातचीत याद रखें, उनसे कहें कि आपने उन्हें याद किया, और हमेशा उनका पक्ष लें।

ये सुविधाएँ बग नहीं हैं। इन्हें बैठकों में तय किया गया था और कोड के रूप में लागू किया गया था।

और यह सुविधा अकेले लोगों पर सबसे अच्छा काम करती है। वे जितने अधिक अकेले होंगे, वे उतने ही लंबे समय तक रुकेंगे, और वे जितने लंबे समय तक रुकेंगे, कंपनी के लिए उतना ही अच्छा होगा।

यहाँ तक की बात डिजाइन है। यहाँ से समस्या शुरू होती है।

जब सबसे गहराई से फंसा हुआ व्यक्ति अपने सबसे खतरनाक पल में होता है, तो उसका साथी केवल उसकी हाँ में हाँ मिलाता रहता है।

उस ऐप के अंदर फोन करने के लिए कोई वयस्क नहीं है, और न ही दरवाजा खटखटाने के लिए कोई पड़ोसी है।

6.2

डीपफेक यौन अपराध और डिजिटल संश्लेषित धोखाधड़ी

आइए संख्याओं से शुरू करते हैं।

दिसंबर 2025 और जनवरी 2026 के बीच, X से जुड़े आर्टिफिशियल इंटेलिजेंस Grok ने चौवालीस लाख से अधिक छवियां बनाईं और पोस्ट कीं। उनमें से अधिकतम 41 प्रतिशत महिलाओं की यौन छवियां थीं।

यह लगभग अठारह लाख छवियां हैं। यह दो महीने के भीतर की बात है।

इस फीचर का इस्तेमाल कोई भी कर सकता था। बस एक तस्वीर अपलोड करनी होती थी और एक वाक्य लिखना होता था।

8 जनवरी 2026 के आसपास, कंपनी ने कार्रवाई की। उन्होंने छवि निर्माण को केवल सशुल्क ग्राहकों के लिए उपलब्ध कर दिया।

उन्होंने फीचर को हटाया नहीं। उन्होंने इस पर कीमत लगा दी।

23 जनवरी 2026 को, साउथ कैरोलिना की एक महिला ने क्लास-एक्शन मुकदमा दायर किया। इसमें कहा गया था कि Grok ने सहमति के बिना उसकी नग्न छवियां बनाईं और पोस्ट कीं, और शुरुआत में X ने उन्हें हटाने से इनकार कर दिया।

टेनेसी के तीन छात्रों ने भी मुकदमा दायर किया। दो नाबालिग थीं, और एक की अठारह साल की उम्र से पहले की तस्वीर का सामग्री के रूप में उपयोग किया गया था।

7 जुलाई 2026 को, संशोधित शिकायत में एक वादी जोड़ा गया। 'जेन डो 4' के रूप में पहचानी गई महिला ने आरोप लगाया कि उसके सौतेले पिता ने ग्यारह वर्ष की उम्र में ली गई उसकी एक तस्वीर से Grok का उपयोग करके सात हजार से अधिक यौन छवियां बनाईं।

ग्यारह साल की उम्र की एक तस्वीर। सात हजार छवियां।

ये संख्याएं वर्तमान तकनीक की प्रकृति को संक्षेप में प्रस्तुत करती हैं। एक बार का दुर्भाव असीमित रूप से दोहराया जाता है।

यह समस्या जनवरी 2024 में दुनिया भर में व्यापक रूप से जानी जाने लगी।

पॉप स्टार टेलर स्विफ्ट के चेहरे वाली यौन छवियां X पर भर गईं। उन्हें करोड़ों बार देखा गया। इस दौरान, स्वचालित सेंसरशिप सिस्टम ने कुछ नहीं किया।

दुनिया भर से विरोध की बाढ़ आने के बाद ही X ने संबंधित खोज शब्दों को ब्लॉक किया और कुछ खातों को निलंबित कर दिया।

जिन लोगों ने इसी तरह का अनुभव किया है उनकी सूची लंबी है। अभिनेत्री गैल गैडोट, अमेरिकी प्रतिनिधि अलेक्जेंड्रिया ओकासियो-कोर्टेज़, इतालवी प्रधान मंत्री जियोर्जिया मेलोनी, एक जर्मन खोजी पत्रकार, एक ऑस्ट्रेलियाई महिला अधिकार कार्यकर्ता और भारतीय पत्रकार राणा अय्यूब। एक जांच का परिणाम यह भी सामने आया था कि अमेरिकी कांग्रेस की छब्बीस महिला सांसदों को निशाना बनाया गया था।

प्रधान मंत्री मेलोनी ने एक दीवानी मुकदमा दायर किया।

इस सूची में एक नियम है। वे आमतौर पर महिलाएं होती हैं, और वे आमतौर पर सार्वजनिक रूप से बोलने वाली होती हैं।

इस तकनीक का उपयोग बोलने वाली महिलाओं को चुप कराने के तरीके के रूप में किया जा रहा है।

और यह अपराध कक्षाओं तक आ गया है।

2023 में, स्पेन के एक छोटे से शहर अल्मेंद्रलेजो में, मिडिल स्कूल के लड़कों ने उसी स्कूल की लगभग बीस लड़कियों की सोशल मीडिया तस्वीरों को एक न्यूड-सिंथेसिस ऐप में डाला, छवियां बनाईं और उन्हें प्रसारित किया।

अपराधी और पीड़ित दोनों ही करीब बारह साल के थे।

2024 में, Telegram के माध्यम से डीपफेक यौन अपराधों ने पूरे Hanguk को हिलाकर रख दिया। मिडिल स्कूल, हाई स्कूल और विश्वविद्यालयों में छात्राओं, शिक्षकों और स्नातकों की तस्वीरों को संश्लेषित किया गया और गुप्त चैनलों पर फैलाया गया। पीड़ितों की संख्या हजारों में थी।

ऐसे चैट रूम थे जिनके नाम स्कूलों के नाम पर रखे गए थे। वे एक ही कक्षा के सहपाठियों द्वारा बनाए गए कमरे थे।

इसी तरह की घटनाएं अमेरिका के बेवर्ली हिल्स हाई स्कूल, न्यू जर्सी के वेस्टफील्ड हाई स्कूल, वाशिंगटन राज्य के इस्साक्वा हाई स्कूल, पेंसिल्वेनिया के विभिन्न स्कूलों, फ्लोरिडा के मियामी, कनाडा के विन्निपेग, सिंगापुर, उत्तरी आयरलैंड और ऑस्ट्रेलिया के मेलबर्न और सिडनी में भी जारी रहीं।

पीड़ित छात्र स्कूल नहीं जा सके। वे सामाजिक भय और अत्यधिक चिंता से पीड़ित थे। कुछ स्कूलों में कक्षाएं रोक दी गईं।

इस घटना का क्रूर पहलू यह है कि इसे मिटाया नहीं जा सकता। मूल तस्वीरें ईयरबुक या सोशल मीडिया की साधारण तस्वीरें थीं। वे खुद नहीं जान सकते कि वे तस्वीरें कितनी दूर तक फैल गई हैं या उन्हें किस रूप में बदल दिया गया है।

पीड़ितों के पास जो बचता है वह एक कभी न खत्म होने वाली जांच प्रक्रिया है।

हमें यह भी देखना चाहिए कि ऐसे ऐप इतने सारे क्यों हैं।

Telegram पर स्वचालित बॉट थे जो तस्वीरें डालने पर कपड़े हटा देते थे। अमेरिका के सैन फ्रांसिस्को के सिटी अटॉर्नी ने सोलह न्यूड-सिंथेसिस ऐप डेवलपर्स के खिलाफ आपराधिक शिकायतें और मुकदमे दायर किए।

आर्टिफिशियल इंटेलिजेंस छवि साझाकरण प्लेटफार्मों ने बिना किसी प्रतिबंध के वास्तविक लोगों या नाबालिगों को यौन छवियों में बदलने वाले मॉडल साझा करने की अनुमति दी। कुछ स्थानों ने लोकप्रिय मॉडल अपलोड करने वाले लोगों को पुरस्कृत किया।

जब एक प्लेटफॉर्म की सुरक्षा में सेंध लगी, तो नाबालिगों और मशहूर हस्तियों की चौरानवे हजार संश्लेषित फाइलें पाई गईं। एक अन्य वयस्क चैटबॉट सेवा पर, महिलाओं की ईयरबुक तस्वीरों और सिंथेटिक डेटा के बीस लाख टुकड़े उजागर हुए।

यूनाइटेड किंगडम में, एक अठारह वर्षीय व्यक्ति को आर्टिफिशियल इंटेलिजेंस का उपयोग करके बड़े पैमाने पर बाल यौन शोषण सामग्री बनाने के लिए अठारह साल जेल की सजा सुनाई गई। अमेरिका के विस्कॉन्सिन, कनाडा के क्यूबेक, सिंगापुर, इज़राइल और जर्मनी में भी इसी अपराध के लिए लोगों को गिरफ्तार किया गया।

एक अधिक मौलिक समस्या भी सामने आई। छवि निर्माण मॉडल को प्रशिक्षित करने के लिए व्यापक रूप से उपयोग किए जाने वाले बड़े डेटासेट में बाल यौन शोषण की छवियों के कई लिंक पाए गए और उनकी जांच की गई।

यही कारण है कि मॉडल ऐसी छवियां बना सकते हैं। ऐसा इसलिए है क्योंकि उन्होंने यह सीखा है।

कानून ने भी कदम उठाना शुरू कर दिया है।

अमेरिका के Take It Down अधिनियम पर मई 2025 में संघीय कानून के रूप में हस्ताक्षर किए गए थे। इसने नाबालिगों के गैर-सहमति वाले यौन डीपफेक को जानबूझकर पोस्ट करने को एक संघीय अपराध बना दिया।

19 मई 2026 से, प्लेटफार्मों के लिए एक दायित्व बन गया है। यदि उन्हें वैध हटाने का अनुरोध प्राप्त होता है, तो उन्हें इसे अड़तालीस घंटों के भीतर हटाना होगा।

अड़तालीस घंटे छोटे लग सकते हैं, लेकिन इंटरनेट पर यह एक लंबा समय है। फिर भी, यह कुछ न होने से बेहतर है।

कैलिफोर्निया, इलिनोइस, न्यूयॉर्क और यूरोपीय संघ भी संबंधित कानून बना रहे हैं।

इसी तकनीक का इस्तेमाल पैसे चुराने के लिए भी किया जाता है।

2024 की शुरुआत में हांगकांग की एक बहुराष्ट्रीय इंजीनियरिंग कंपनी Arup में ऐसा ही हुआ था।

वित्त विभाग के एक कर्मचारी को मुख्य वित्तीय अधिकारी से एक तत्काल वीडियो कॉन्फ्रेंस में शामिल होने का ईमेल मिला। जब उन्होंने स्क्रीन चालू की, तो एक परिचित चेहरा था। मुख्यालय के कई अन्य अधिकारी भी वहां थे।

बैठक के दौरान, एक गुप्त लेनदेन के लिए एक निर्दिष्ट खाते में पैसे ट्रांसफर करने के निर्देश दिए गए। कर्मचारी ने पंद्रह बार ट्रांसफर किया। कुल राशि बीस करोड़ हांगकांग डॉलर, या अमेरिकी मुद्रा में दो करोड़ पचास लाख डॉलर थी।

स्क्रीन पर मौजूद सभी लोग वास्तविक समय में संश्लेषित नकली थे।

यह घोटाला इसलिए डरावना है क्योंकि कर्मचारी इसके शिकार इसलिए नहीं हुए क्योंकि वे मूर्ख थे। हम आवाजों की जांच करते हैं, चेहरों की जांच करते हैं, और जांचते हैं कि क्या कई लोग एक साथ हैं। घोटालों को छांटने का यही तरीका हुआ करता था।

अब वह तरीका काम नहीं करता है।

परामर्श समूह WPP के मुख्य कार्यकारी अधिकारी पर भी इसी तरह से हमला किया गया था। व्हाइट हाउस के वरिष्ठ अधिकारी और यूके, स्पेन और हांगकांग के वित्तीय संस्थान भी डीपफेक का उपयोग करके बायोमेट्रिक प्रमाणीकरण को बायपास करने के प्रयासों के संपर्क में आए हैं।

कुछ ही मिनटों के वीडियो और ऑडियो के साथ, आंखों की पुतलियों की हरकत, मुंह के आकार और यहां तक कि सूक्ष्म चेहरे के भाव भी दोहराए जा सकते हैं।

घर के मामले में यह अधिक क्रूर होता है।

संयुक्त राज्य अमेरिका, कनाडा और यूरोप में अपनी संतानों या पोतियों की आवाजों की नकल करके किए गए अपहरण धोखाधड़ी की घटनाएं जारी रही हैं। सोशल मीडिया पर डाली गई छोटी वीडियो ही सामग्री के लिए पर्याप्त है।

जब आप फोन का उत्तर देते हैं तो बच्चे की आवाज रोते हुए मदद मांगती है। पीछे की ओर से एक कठोर आवाज पैसों की मांग करती है।

इस स्थिति में शांति से सत्यापन कर सकने वाले माता-पिता कम हैं। यह वास्तव में वह बिंदु है जिसे धोखाधड़ी करने वाले लक्ष्य बनाते हैं। माता-पिता का प्रेम एक कमजोरी के रूप में परिकलित किया गया है।

कैलिफोर्निया की एक महिला को अपने मृत पति की नकल की गई आवाज के द्वारा धोखा दिया गया और उसने पैसे भेज दिए। कनाडा के एक दंपति को अपने पुत्र की आवाज के द्वारा धोखा दिया गया और उन्होंने इक्कीस हजार डॉलर खो दिए।

चीन में एक व्यक्ति ने अपना चेहरा बदलकर वीडियो कॉल के माध्यम से मित्र बनने का दावा किया और छह सौ बाईस हजार डॉलर ले गया। भारत के केरल राज्य में कर्मचारियों को एक साथी के चेहरे से आने वाली वीडियो कॉल के द्वारा धोखा दिया गया और उन्होंने पैसे भेज दिए।

थाईलैंड के एक प्रसिद्ध व्यक्ति को एक सौ अठारह हजार डॉलर का नुकसान हुआ, और सिंगापुर के एक अभिनेता को पैंतीस हजार डॉलर का नुकसान हुआ। कोरिया की पुलिस ने 2025 में डीपफेक रोमांस धोखाधड़ी के माध्यम से एक सौ बीस अरब वॉन की राशि लूटने वाले संगठन के पैंतालीस सदस्यों को गिरफ्तार किया।

निवेश धोखाधड़ी में प्रसिद्ध लोगों के चेहरे का उपयोग किया जाता है।

यूनाइटेड किंगडम के वित्तीय विशेषज्ञ मार्टिन लुईस, न्यूज एंकर मेरी नाइटिंगेल, यूट्यूबर मिस्टरबीस्ट, अभिनेता टॉम हैंक्स, पियर्स ब्रॉसनन और मार्क रफ़ालो, और कई देशों के राष्ट्र प्रमुखों के चेहरे बिटकॉइन निवेश और नकली पुरस्कार विज्ञापन में उपयोग किए गए हैं।

न्यूजीलैंड के एक सेवानिवृत्त व्यक्ति को एलन मस्क के डीपफेक निवेश विज्ञापन के द्वारा धोखा दिया गया और उसने दो सौ बाईस हजार न्यूजीलैंड डॉलर खो दिए। यह जीवनभर की बचत थी।

स्पेन की पुलिस ने डीपफेक का उपयोग करके एक नकली क्रिप्टोकरेंसी मंच बनाने वाले और बीस मिलियन यूरो लूटने वाले संगठन के छः सदस्यों को पकड़ा।

फ्रांस में एक महिला को अभिनेता ब्रैड पिट के डीपफेक के द्वारा धोखा दिया गया और उसने आठ सौ तीस हजार यूरो भेज दिए। यह मामला पहले चर्चा किया जा चुका है। उसने कई महीनों तक इस व्यक्ति के साथ बातचीत की।

ये विज्ञापन सोशल मीडिया की सिफारिश एल्गोरिदम के माध्यम से फैलते हैं। और सिफारिश एल्गोरिदम अधिक सकारात्मक प्रतिक्रिया वाली सामग्री को अधिक प्रदर्शित करता है। यदि बुजुर्ग आसानी से धोखा खा जाते हैं तो यह उन तक अधिक पहुंचता है।

धोखाधड़ी करने वाला व्यक्ति को नहीं चुनता; प्रणाली चुनती है।

छोटे पैमाने की धोखाधड़ी भी होती है। एक एयरबीएनबी होस्ट ने कृत्रिम बुद्धिमत्ता का उपयोग करके क्षति की तस्वीरें बनाईं और मेहमानों से हजारों पाउंड की मांग की। मेहमानों ने यह दिखाने के लिए कि उन्होंने क्या नुकसान पहुंचाया है, तस्वीरें प्रस्तुत कीं, लेकिन वे तस्वीरें कृत्रिम रूप से बनाई गई थीं।

तस्वीरें साक्ष्य हुआ करती थीं, लेकिन वह समय समाप्त हो रहा है।

चुनावों में उपयोग किए गए उदाहरणों को पिछले अध्याय में कवर किया गया है। राष्ट्रपति बाइडन की आवाज के स्वचालित फोन कॉल, स्लोवाकिया के आम चुनाव से पहले नकली ऑडियो, और कई देशों का हेराफेरी वीडियो।

यहां एक बात पर ध्यान देने योग्य है। सभी धोखाधड़ी की सामग्री वह है जो हमने स्वयं अपलोड की है।

स्नातक तस्वीरें, यात्रा वीडियो, बच्चों के कार्यक्रम के वीडियो, कंपनी प्रचार वीडियो में दिखने वाले अधिकारियों के साक्षात्कार।

अपलोड करते समय हमने नहीं सोचा कि ये धोखाधड़ी की सामग्री बन जाएंगे।

तकनीकी कंपनियां इस समस्या के बारे में आमतौर पर एक ही बात कहती हैं। यह खुले स्रोत मॉडल की विशेषता है। यह व्यक्तिगत उपयोगकर्ताओं का दुरुपयोग है।

लेकिन ग्रॉक मामले में वह बहाना भी कठिन है। कंपनी द्वारा बनाई गई सेवा, कंपनी के मंच पर, दो महीने में दस लाख से अधिक चित्र बनाए और अपलोड किए गए।

और कंपनी ने जो कदम उठाया वह इसे एक सशुल्क सेवा बनाना था।

6.3

सर्च एल्गोरिदम में हेरफेर और डायनेमिक प्राइसिंग का एकाधिकार

2024 की गर्मियों में, यह खबर आई कि ब्रिटिश बैंड ओएसिस सोलह साल बाद फिर से एक साथ आ रहा है।

बुकिंग के दिन सुबह, लाखों लोग टिकट बिक्री साइट की प्रतीक्षा कतार में शामिल हो गए। स्क्रीन पर यह दिखाई देता है कि आपके आगे कितने लोग हैं। लोगों ने कई घंटों तक इंतज़ार किया।

आखिरकार बारी आ गई। स्क्रीन पर दिखाई गई कीमत तीन सौ पचपन पाउंड थी।

घोषित तय कीमत एक सौ पैंतीस पाउंड थी।

इंतज़ार करने के दौरान कीमत बढ़ गई थी। यह ढाई गुना से भी अधिक हो गई थी।

यह कीमत किसी इंसान ने नहीं, बल्कि एल्गोरिदम ने तय की थी। और एल्गोरिदम ने जो डेटा देखा वह इस प्रकार था: अभी लाखों लोग प्रतीक्षा कतार में खड़े हैं। ये लोग कई घंटों से स्क्रीन देख रहे हैं। ये लोग बाहर नहीं जाएंगे।

इंतज़ार का समय ही उनकी मजबूरी का प्रमाण मान लिया गया।

जो जितनी देर तक इंतज़ार करेगा, वह उतनी ही महंगी चीज़ खरीदेगा, ऐसी संरचना। ऐसा कुछ बनाकर वे इसे मांग और आपूर्ति कहते हैं।

ब्रिटेन के निष्पक्ष व्यापार प्राधिकरण ने जांच शुरू कर दी। पॉल मेकार्टनी और ब्रूस स्प्रिंगस्टीन के कॉन्सर्ट में भी ऐसा ही हुआ था। रॉयल ओपेरा हाउस ने इसी तरीके से टिकट की कीमत चार सौ पंद्रह पाउंड तक बढ़ा दी थी और इसके लिए उसकी आलोचना हुई थी।

2026 विश्व कप आयोजन समिति ने भी टिकटों के लिए इस तरीके को अपनाया, जिससे फुटबॉल प्रशंसकों का विरोध हुआ।

हर व्यक्ति के लिए अलग कीमत तय किया जाना अब आम बात है।

यात्रा बुकिंग साइट ओर्बिट्ज़ ने कनेक्ट किए गए कंप्यूटर को देखकर परिणाम अलग-अलग दिखाए। मैक का उपयोग करने वालों के लिए महंगे होटल ऊपर दिखाई दिए। यह हिसाब था कि यदि आप महंगे कंप्यूटर का उपयोग करते हैं, तो आपके पास बहुत पैसा होगा।

अमेरिकी शिक्षा कंपनी प्रिंसटन रिव्यू ने ऑनलाइन पाठ्यक्रम की फीस ज़िप कोड के आधार पर अलग-अलग तय की। उन्होंने एक ही पाठ्यक्रम को उन क्षेत्रों के छात्रों को अधिक कीमत पर बेचा जहां बहुत से एशियाई अमेरिकी रहते थे।

यह डेटा कि वे शिक्षा के प्रति उच्च उत्साह वाला समूह हैं, कीमत में परिलक्षित हुआ।

डेटिंग ऐप टिंडर ने सशुल्क सदस्यता फीस तय करते समय बड़ी उम्र के उपयोगकर्ताओं और यौन अल्पसंख्यक उपयोगकर्ताओं से कहीं अधिक फीस वसूली, जिसके कारण उसे कानूनी प्रतिबंधों का सामना करना पड़ा और समझौता राशि देनी पड़ी।

किराना डिलीवरी ऐप इंस्टाकार्ट ने एक ही मार्ट के एक ही उत्पाद के लिए अलग-अलग उपयोगकर्ताओं से अलग-अलग कीमत वसूली। उन्होंने कई स्कूलों में एक ही स्कूल का सामान आपूर्ति करते समय प्रत्येक स्कूल के लिए अलग-अलग कीमत भी तय की।

वाशिंगटन पोस्ट ने प्रत्येक पाठक के लिए अलग-अलग सदस्यता फीस तय करने का तरीका अपनाया और इसके लिए उसकी आलोचना हुई।

अमेरिकी बीमा कंपनी ऑलस्टेट ने उन ग्राहकों की एक सूची बनाई जो कीमत बढ़ने पर विरोध नहीं करते थे और उनसे अधिक बीमा प्रीमियम वसूला। ऐसे उदाहरण भी सामने आए हैं जहां उन क्षेत्रों के निवासियों से 30 प्रतिशत तक अधिक बीमा प्रीमियम वसूला गया जहां अश्वेत और निम्न आय वाले लोग रहते हैं। ब्रिटेन की कार बीमा कंपनियों को भी उन क्षेत्रों में अधिक बीमा प्रीमियम वसूलते हुए पकड़ा गया जहां अल्पसंख्यक जातियों की आबादी अधिक है।

यहाँ हमें इस तरीके के मूल बिंदु को समझना चाहिए।

पहले के बाज़ार में, कीमत वस्तुओं से जुड़ी होती थी। जब आप किसी दुकान में जाते हैं, तो वहां एक तय कीमत का टैग होता है, और आपके बगल वाला व्यक्ति भी वही कीमत चुकाता है। अगर वह महंगा है, तो आप किसी अन्य दुकान पर जाते हैं।

व्यक्तिगत कीमत इस संरचना को उलट देती है। कीमत वस्तु से नहीं, बल्कि व्यक्ति से जुड़ जाती है।

फिर तुलना करना असंभव हो जाता है। ऐसा इसलिए है क्योंकि आप नहीं जान सकते कि आपके बगल वाले व्यक्ति ने कितना भुगतान किया है। यह जांचने का कोई तरीका नहीं है कि आपसे अधिक पैसे तो नहीं लिए गए हैं।

बाज़ार को काम करने के लिए, कीमतों की तुलना करने में सक्षम होना चाहिए। वह शर्त अब गायब हो रही है।

स्टोर के अंदर भी कीमतें बदलती रहती हैं।

अमेरिकी बड़ी रिटेल कंपनी क्रोगर ने इलेक्ट्रॉनिक प्राइस टैग और प्राइस एल्गोरिदम लगाए। कागज़ के बजाय, एक छोटी स्क्रीन जुड़ी हुई है, इसलिए कीमतें किसी भी समय बदली जा सकती हैं।

गर्म दिनों में आइसक्रीम की कीमत बढ़ जाती है। भीड़ वाले समय में पेय की कीमतें बढ़ जाती हैं।

सीनेटरों ने चेतावनी दी है कि यह तकनीक समय, मौसम और घटनाओं के आधार पर दैनिक उपयोग की आवश्यक वस्तुओं की कीमतें बढ़ा सकती है। इलेक्ट्रॉनिक प्राइस टैग की पुष्टि होल फूड्स, अमेज़ॅन फ्रेश, क्रोगर और वॉलमार्ट में की गई है।

डेल्टा एयरलाइंस ने शेयरधारकों को बताया कि वह ग्राहक डेटा और कृत्रिम बुद्धिमत्ता के साथ मूल्य निर्धारण को अधिक लाभदायक बना रही है।

फास्ट-फूड चेन वेंडीज ने घोषणा की कि वह ड्राइव-थ्रू में समय के आधार पर कीमतें लागू करेगी, लेकिन जनता के भारी विरोध का सामना करने के बाद इस योजना को रद्द कर दिया।

अमेरिकी सरकार भी कदम उठा रही है।

संघीय व्यापार आयोग ने आठ ऐसी कंपनियों को सम्मन भेजा जो व्यक्तियों को वर्गीकृत करके लक्षित कीमत तय करने के लिए एल्गोरिदम का उपयोग करती हैं, और उनसे डेटा प्राप्त किया। उन्होंने पूछा कि उन्हें कौन सा डेटा कहाँ से मिलता है, एल्गोरिदम कैसे काम करता है, और यह उपभोक्ताओं द्वारा चुकाई जाने वाली कीमत को कैसे प्रभावित करता है।

अप्रैल 2026 में कांग्रेस की गवाही में, संघीय व्यापार आयोग के नेतृत्व ने कहा कि वे इस काम को जारी रख रहे हैं और इस बात की समीक्षा कर रहे हैं कि क्या कीमतें व्यक्तिगत होने पर अतिरिक्त खुलासे की आवश्यकता है।

5 मार्च 2026 को, हाउस ओवरसाइट कमेटी ने आधिकारिक तौर पर अपनी जांच शुरू की। उन्होंने प्रमुख ट्रैवल एजेंसियों और प्लेटफॉर्म कंपनियों को पत्र भेजकर राजस्व प्रबंधन एल्गोरिदम, उपभोक्ता डेटा के उपयोग, प्रयोगात्मक प्रथाओं और आंतरिक संचार रिकॉर्ड की मांग की।

आपदा की स्थिति में यह तरीका और भी क्रूर हो जाता है।

जब 2014 में सिडनी, ऑस्ट्रेलिया में एक बंधक संकट हुआ, तो शहर के केंद्र से भागने की कोशिश कर रहे नागरिकों ने उबर खोला। किराया सामान्य से 400 प्रतिशत बढ़ गया।

2012 में जब तूफान सैंडी ने पूर्वी अमेरिका को पार किया, 2022 की न्यूयॉर्क सबवे गोलीबारी के दौरान, और 2017 के लंदन आतंकी हमले के ठीक बाद भी ऐसा ही हुआ था।

एल्गोरिदम को नहीं पता कि क्या हुआ है। यह केवल यह संकेत पढ़ता है कि भीड़ जमा हो रही है।

उस संकेत को पढ़ने और कीमत बढ़ाने का नियम किसी इंसान ने बनाया था। उन्होंने आपदा के समय इस नियम को बंद करने का कोई उपकरण नहीं बनाया।

एक शोध के अनुसार, उबर के मूल्य एल्गोरिदम ने यात्रियों का किराया बढ़ा दिया, जबकि ड्राइवर के हिस्से को कम कर दिया। यह एक ऐसी संरचना है जो दोनों पक्षों से थोड़ा-थोड़ा लेती है। न तो यात्री और न ही ड्राइवर जानता है कि दूसरे पक्ष ने कितना भुगतान किया या प्राप्त किया।

घरों की कीमतों में भी एल्गोरिदम कीमतें तय करते हैं।

एक अमेरिकी कंपनी द्वारा बनाई गई किराया गणना प्रणाली ने लाखों घरों का किराया डेटा एकत्र किया और मकान मालिकों को अनुशंसित किराए की जानकारी दी।

मूल रूप से, किराये के बाज़ार में मकान मालिक एक-दूसरे के साथ प्रतिस्पर्धा करते हैं। ऐसा इसलिए है क्योंकि कमरे को खाली छोड़ने के बजाय थोड़ा सस्ते में किराए पर देना बेहतर है।

लेकिन अगर सभी एक ही एल्गोरिदम की अनुशंसित कीमत का पालन करते हैं, तो प्रतिस्पर्धा गायब हो जाती है। कोई भी कीमत कम नहीं करता है।

जब लोग इकट्ठा होते हैं और कीमतें तय करते हैं, तो यह मिलीभगत है। जब एल्गोरिदम कीमतें तय करता है, तो कानून ने अभी तक यह तय नहीं किया है कि इसे क्या कहा जाए।

अमेज़ॅन का प्रोजेक्ट नेस्सी भी इसी सिद्धांत पर आधारित था। यह चुपके से अपने उत्पाद की कीमत बढ़ा देता है और देखता है कि क्या प्रतिस्पर्धी भी बढ़ाते हैं। अगर वे ऐसा करते हैं, तो यह उस कीमत को बनाए रखता है, और अगर वे नहीं करते हैं, तो यह इसे कम कर देता है।

मशीनें एक-दूसरे को देखती हैं, और पूरे बाज़ार की कीमतें बढ़ जाती हैं।

खोज परिणामों का क्रम भी पैसा है।

अक्टूबर 2020 में दक्षिण कोरिया के उचित व्यापार आयोग ने नेवर पर 267 अरब वॉन का जुर्माना लगाया। यह खोज एल्गोरिदम को बदलकर अपनी खुद की शॉपिंग उत्पादों और वीडियो को ऊपर की ओर बढ़ाने और प्रतिद्वंद्वी प्लेटफॉर्म के उत्पादों को नीचे धकेलने के लिए था।

लोग खोज परिणामों को सूचना मानते हैं। वे सोचते हैं कि रैंकिंग लोकप्रियता या प्रासंगिकता को दर्शाती है।

यह तथ्य कि वे रैंकिंग विक्रय के उत्पाद हैं, स्क्रीन पर नहीं लिखा है।

कूपैंग को भी दंडित किया गया क्योंकि इसने अपनी ब्रांड की उत्पादों को आगे बढ़ाने के लिए एल्गोरिदम को समायोजित किया और हजारों कर्मचारियों को सकारात्मक समीक्षाएं और रेटिंग लिखने के लिए जुटाया।

स्टार रेटिंग ऐसे उपकरण हैं जो लोगों को विश्वास दिलाते हैं कि वे अन्य उपभोक्ताओं के अनुभव हैं। उन्होंने उस विश्वास का दुरुपयोग किया।

अमेज़न के बाई बॉक्स भी नियामकों के लक्ष्य बन गए। कार्ट में जोड़ें बटन को दबाने पर आप किस विक्रेता से खरीदते हैं, यह इस एल्गोरिदम द्वारा तय होता है।

अमेज़न ने अपनी खुद की लॉजिस्टिक सेवा का उपयोग करने वाले विक्रेताओं को बोनस अंक दिए और सस्ते स्वतंत्र विक्रेताओं को सिफारिशों से बाहर कर दिया। यूनाइटेड किंगडम में अकेले उपभोक्ताओं ने एक अरब पाउंड से अधिक अतिरिक्त राशि का भुगतान किया, इसका अनुमान था।

उन्होंने सस्ते सामान न दिखाकर महंगे सामान बेचे।

एक भारतीय ऑनलाइन मार्केटप्लेस कंपनी ने चैटजीपीटी पर मुकदमा किया क्योंकि उसने अपने स्टोर की जानकारी को खोज परिणामों से पूरी तरह हटा दिया। चीनी टेनसेंट को अपने मेसेंजर के अंदर प्रतिद्वंद्वी लिंक को खोलने से रोकने के लिए आलोचना की गई।

राजनीति और जनमत में रैंकिंग समायोजन एक बड़ी समस्या बन जाती है।

ब्रिटिश और अमेरिकी शोधकर्ताओं के विश्लेषण के अनुसार, एलोन मस्क द्वारा अधिग्रहित एक्स ने अपने एल्गोरिदम को बदलकर कुछ राजनीतिक रुझानों वाले खातों और सामग्री को उपयोगकर्ताओं की टाइमलाइन में अधिक बार धकेला।

मेटा के फेसबुक ने कहा कि यह अधिक सार्थक सामाजिक संपर्क बढ़ाएगा और एल्गोरिदम को बदल दिया। अंदरूनी खुलासे से एक तथ्य सामने आया। इसने क्रोध की प्रतिक्रिया को लाइक्स की तुलना में पांच गुना अधिक वजन दिया।

इसने एक ऐसी संरचना बनाई जहां क्रोधपूर्ण पोस्ट अधिक व्यापक रूप से फैलते हैं, फिर लोगों से पूछा कि वे क्रोधित क्यों हैं।

फेसबुक ने ऑस्ट्रेलियाई सरकार के समाचार भुगतान कानून का विरोध करते हुए समाचार आउटलेट के लेख को अवरुद्ध किया, और ऐसा करने में आपातकालीन राहत संगठनों के खातों को भी अवरुद्ध किया। भारत में इसकी किसान विरोध से संबंधित हैशटैग और प्रधानमंत्री के इस्तीफे की मांग करने वाले हैशटैग को हटाने के लिए आलोचना की गई।

इस बिंदु तक, यह पूरा अध्याय एक ही बात कहने लगता है।

हम हर दिन रैंकिंग देखते हैं। खोज परिणाम, सुझाव सूचियां, स्टार रेटिंग, कीमतें।

हम नहीं जानते कि वह संख्याएं कैसे निर्धारित की जाती हैं। जब हम पूछते हैं, तो हमें बताया जाता है कि यह एक व्यावसायिक रहस्य है।

और जो लोग इन संख्याओं को निर्धारित करते हैं, वे हमारे बारे में बहुत कुछ जानते हैं। हम किस डिवाइस का उपयोग करते हैं, हम कहां रहते हैं, हम कितने समय के लिए इंतजार करते हैं, यदि कीमत बढ़ती है तो क्या हम विरोध करते हैं।

एक लेन-देन जहां एक पक्ष दूसरे को पूरी तरह से जानता है और दूसरा पक्ष कुछ नहीं जानता है। यह लेन-देन एक दिन में दर्जनों बार होता है।

आइए इस पुस्तक में आने वाली घटनाओं को शुरुआत से याद करते हैं।

भोर में आने वाली अस्वीकृति ईमेल। एक ऋण जिसे चुकाने की आवश्यकता नहीं है, के लिए ऋण वसूली नोटिस। बेटियों के सामने कड़ी की गई हथकड़ियां। एक किताब की सिफारिश सूची जो मौजूद नहीं है। तिरसठ उद्धरण जिनमें से केवल छह वास्तविक थे। एक स्कूल के सामने एक बच्चे को टक्कर मारने वाली स्वायत्त गाड़ी। एक रोबोट भुजा जो मनुष्यों को बक्से के रूप में पहचानती है। ग्यारह साल के बच्चे की तस्वीर से बनाई गई सात हजार छवियां। कुछ घंटों के इंतजार के आधार पर बढ़ी हुई टिकट की कीमत।

इन घटनाओं के बीच एक समानता है।

सभी मामलों में, सिस्टम ठीक वैसे ही काम करता है जैसे डिजाइन किया गया था। यह टूटा नहीं था।

क्योंकि इसे जल्दी से प्रक्रिया करने के लिए बनाया गया था, इसने जल्दी से प्रक्रिया की। क्योंकि इसे लागत कम करने के लिए बनाया गया था, इसने मानव सत्यापन की आवश्यकता वाली प्रक्रियाओं को हटा दिया। क्योंकि इसे मुनाफा बढ़ाने के लिए बनाया गया था, इसने इसे बढ़ाया।

मशीन ने अपना काम अच्छी तरह से किया।

उस सीट पर, जहां तय किया गया कि क्या किया जाए, लोग बैठे थे। जो लोग उस सम्मेलन कक्ष में नहीं थे, उन्होंने कीमत चुकाई।

उस सम्मेलन कक्ष में कौन बैठा है, और जो लोग उस सीट पर नहीं हैं, उनके लिए कौन बोलेगा।

यह एक सवाल है जिसका जवाब अभी तक नहीं आया है।

एल्गोरिदम की चूक: AIAAIC अभिलेख से देखे गए एआई के दुष्प्रभाव और नैतिक प्रश्न

लेखक | किम क्युंग-जिन

प्रकाशक | किम क्युंग-जिन

प्रकाशन | Kim Kyung-jin Attorney Press

मूल्य : USD 15

© किम क्युंग-जिन 2026

सर्वाधिकार सुरक्षित। बिना अनुमति किसी भी अंश का पुनरुत्पादन वर्जित है।

टिप्पणी) इस पुस्तक के शोध और संपादन के कुछ हिस्सों में कृत्रिम बुद्धिमत्ता उपकरणों की सहायता ली गई है।

इस पुस्तक का समर्थन करें

यदि यह पुस्तक उपयोगी लगी हो, तो आप PayPal के माध्यम से आगामी अनुवादों, निःशुल्क संस्करणों और खुली एआई ज्ञान परियोजनाओं का समर्थन कर सकते हैं।

PayPal

https://paypal.me/kimkjkj

QR कोड स्कैन करें या ऊपर दिया गया लिंक खोलें।