WHEN THE ALGORITHM GETS IT WRONG

アルゴリズムの失策

AIAAIC の事例から見る AI の副作用と倫理的課題

採用と福祉、法廷と道路、教室と職場で起きたこと

金京鎭 弁護士

目次

第 1 章

偏見と差別：アルゴリズムに埋め込まれた人間の偏り

1.1

採用・金融・サービスのアルゴリズム偏向

不採用のメールは、いつも夜明けに届きました。

デリック・モブリーは、40歳を過ぎた黒人男性であり、不安障害とうつ病を患っていました。情報技術の分野で働き、資格も持ち、経歴もありました。彼はワークデイという会社の採用システムを利用している企業100社余りに応募しました。

100回余り落ちました。

不採用のメールが届いた時刻を見ると、人間が送ったものではないという事実がわかりました。午前1時、午前3時。人間の採用担当者は、その時間には眠っています。機械は眠りません。コーヒーも飲まず、申し訳ないとも思いません。書類を開いて、3秒で閉じます。

モブリーは訴訟を起こしました。ワークデイの人工知能が、年齢や人種、障害をふるい落とす網のように機能しているという主張でした。裁判所は、ワークデイが既存の役職員のデータで人工知能を学習させ、保護されるべき特性を代替指標として使用したと判断し、全国規模の集団訴訟を承認しました。

2026年3月、裁判所は、年齢差別禁止法が応募者には適用されないというワークデイ側の主張を退けました。6月16日には、カリフォルニア州外にいる雇用主が使用した場合でも、ワークデイが責任を負う可能性があるという判決が出ました。この訴訟が扱う不採用の件数は11億件です。

11億回の夜明けのメールです。

この話が恐ろしい理由は、悪党がいないことにあります。誰も会議室に集まって、40歳を過ぎた人を落とそうと決議したわけではありません。ただ、過去10年間に採用した人たちのデータを機械に与えただけです。機械は誠実にそのパターンを学びました。そして、誠実に繰り返しました。

機械学習という言葉は難しく聞こえますが、原理は小学校の算数の時間と似ています。規則を教えずに答えだけをたくさん見せて、ここから規則を見つけ出しなさいと指示するのです。過去の答えが偏っていれば、機械は偏りを規則として学びます。機械の立場からすれば、よくやったことです。言われた通りにしたのですから。

10年前、アマゾンがその事実を先に学びました。

2014年、スコットランドのエディンバラにあるアマゾンの開発センターで、秘密のプロジェクトが始まりました。履歴書を入れると星評価をつけてくれるシステムでした。星1つから5つまで。品物につける星評価を、人間につけたわけです。

開発者たちは、過去10年分の技術職の志願書を学習データとして入力しました。その10年間にアマゾンが採用した技術職は、大部分が男性でした。機械は結論を下しました。男性のほうがより優れた応募者である。

機械は履歴書から「女性」という単語を見つけると、点数を減らしました。「女性チェスクラブ会長」と書いた履歴書が減点されました。女子大学を出た応募者たちの点数も、丸ごと下がりました。反対に、男性エンジニアがよく使う動詞、「遂行した」や「捕捉した」といった言葉が入っていると、点数が上がりました。

アマゾンは、この偏りを直そうとしました。失敗しました。2017年にプロジェクトを廃棄しました。幸いにも、そのシステムは実際の採用には使われなかったとアマゾンは明らかにしました。

問題は、他の会社が似たようなものをずっと作り続けたことにあります。

カイル・ビームは、ヴァンダービルト大学の学生でした。双極性障害の治療のために学校をしばらく休み、生活費を稼ぐために近所のお店に応募しました。クローガー・スーパーマーケット、ホームデポ、ウォルグリーン、ペットスマート。パートタイムのレジ打ちの仕事でした。

全部落ちました。面接まで進むこともできませんでした。

クロノスという会社が作ったユニクルーというオンライン性格検査が、カイルに赤信号を灯したからです。その検査はカイルの精神的健康の履歴を感知し、このような人は顧客が怒った時に顧客を無視する可能性が高いと判定しました。

カイルの父親、ローランド・ビームは弁護士でした。2014年に雇用機会均等委員会に陳情を出し、企業を相手に精神障害者差別の訴訟を起こしました。

カイルは2019年に自ら命を絶ちました。訴訟が終わる前でした。

性格検査に合格できず、レジに立つことができなかった青年がいたという事実。その判定を下したのが人間ではなくウェブページだったという事実。そして、そのウェブページには抗議する窓口がなかったという事実。この3つの文章を並べて読むと、背筋が寒くなります。

アイチューターグループのシステムは、さらに露骨でした。オンライン教育会社であるこの場所の自動採用プログラムは、55歳以上の女性と60歳以上の男性を自動的に脱落させました。200人を超える教師が資格を備えていながら、年齢のせいで切られました。

発覚した経緯は滑稽です。ある応募者が、本当の生年月日で応募して落ちました。ところが、年齢だけを下げて再び応募すると、すぐに面接をしようという連絡が来ました。人間なら顔が赤くなったはずですが、システムは平然としていました。雇用機会均等委員会は、36万5000ドルの和解金を科しました。

生成型人工知能が出れば、良くなると思っていました。

2024年、ブルームバーグの研究チームは、GPT-3.5に全く同じ資格を持つ偽の履歴書をたくさん与え、名前だけを変えました。人種と性別が推測できる名前でした。結果はこうなりました。技術職の審査において、黒人女性と推測される名前は36パーセント少なく選ばれました。女性の名前は人事管理の職務へと追いやられました。

2021年、ドイツのバイエルン放送局は、レトリオというスタートアップの人工知能による力量評価を実験しました。同じ人が同じ言葉を話しているのに、眼鏡をかけると点数が変わりました。ヒジャブを被るとまた変わりました。後ろに本棚

が見えると、誠実に見えるという判定が出ました。

本棚を立てておくと性格が良くなるだなんて、笑えます。その判定によって誰かは仕事を得て、誰かは得られなかったという箇所で、笑いが止まります。

教壇でも同じことがありました。2012年、ワシントンD.C.の公立学校の5年生の教師サラ・ワイソッキーは、インパクトという教員評価システムによって解雇されました。同僚の教師や保護者たちの評価は素晴らしいものでした。しかし、子供たちが前に通っていた学校で試験の不正があり、そのために水増しされた点数がワイソッキーの付加価値スコアを引き下げたのです。

検証されていない数式一つで、205人の教師が職場を失いました。その数式が何であったのかは公開されませんでした。

お金を扱うアルゴリズムは、もっと長く、もっと静かに働いていました。

2018年、カリフォルニア大学バークレー校の研究チームが発表した報告書があります。2008年から2015年まで、ファニーメイとフレディマックが保証した30年満期の固定金利ローンを調査しました。ローン審査を人からアルゴリズムに移行させれば、差別がなくなるだろうという話がその当時流行していました。

数字は別の話をしていました。黒人とラテン系の借り手は、購入ローンで5.6から8.6ベーシスポイント、リファイナンスで3ベーシスポイント多く支払いました。ベーシスポイントは0.01パーセントを意味する金融単位です。小さく見える数字です。

この小さな数字を米国全体で合計すると、年間2億5000万ドルから5億ドルになります。

理由はもっと苦々しいものです。貸付機関は、少数民族の借り手が複数の商品を比較する時間と情報が不足しているという事実を人工知能で発見しました。そして、その人たちにもう少し高い価格をつけました。戦略的価格設定と呼ばれます。そうして得た追加利益は11パーセントから17パーセントです。

調査報道メディアのザマークアップと聯合ニュースは、連邦住宅ローンデータを一緒に分析しました。クラシック・FICOスコアに基づく秘密の承認アルゴリズムは、有色人種の申請者を白人より40パーセントから80パーセント多く却下しました。一部の大都市では250パーセントを超えました。

スタンフォード経営大学院のローラ・ブラットナーとスコット・ネルソンは、ここからさらに一層深く掘り下げました。アルゴリズムに悪意がなくても問題が生じるというわけです。低所得層と少数者には信用記録が薄いのです。記録が薄いと予測精度が5から10パーセント低下します。昔の延滞が一、二件あると、スコア全体が壊れてしまいます。

貧しければデータが少なく、データが少なければ機械が当たらず、当たらなければ危険な人として分類されます。そしてその分類には、客観的信用リスクという名札がつきます。

ウェルズ・ファーゴ銀行では、計算の間違い一つが人を路上に追いやりました。2010年から2015年まで、この銀行の自動モーゲージ再調整審査ソフトウェアには、弁護士費用を誤って計算するエラーがありました。5年以上の間、誰も知りませんでした。

ローン条件の変更を受ける資格が十分にあった870名が却下通知を受けました。そのうち最低545名が家を差し押さえられました。銀行は2015年末に誤りを修正しましたが、その事実を数年間公表しませんでした。後に設定した補償金は800万ドルでした。

家一軒の値段を考えると、計算の合わない額です。

保険も同じでした。GEICO、Safeco、Nationwideが使った危険評価アルゴリズムは、事故リスクが同じでも、有色人種が多く住む地域の住民に最大30パーセント高い保険料を課しました。オールステイトはさらに悪質でした。顧客が去らなさそうなら値段を上げました。長く取引した忠実な顧客がより高く支払うことになったのです。規制当局に摘発されました。

職場に入った後でも、アルゴリズムは追いかけてきます。

2021年、イタリアのボローニャ労働裁判所は、配達プラットフォームDeliverooのフランク・アルゴリズムが差別的だと判決しました。フランクは配達員の信頼性と参加度をスコアで評価しました。定められた勤務に出られなかったらスコアが下がりました。理由は問いませんでした。

病気で出られなくても下がりました。法律で保障されたストライキに参加しても下がりました。スコアが下がると、次に良い配車を受け取れません。ストライキをする権利は法典に残っていましたが、ストライキをすれば飢える構造がコードの中にありました。

チェッカーという会社の自動身元調査システムは、2015年からウーバー、Lyft、ポストメイツのドライバーたちを大量に除外していました。同じ名前の別人の犯罪記録、既に時効が成立した記録がそのまま付いていました。朝起きたら口座が遮断されたドライバーたちは、どこに抗議すべきかも知りませんでした。

カリフォルニア大学ヘースティングス・ロースクールのビーナ・ドゥバル教授は、ここに名前をつけました。アルゴリズムによる賃金差別です。ウーバーとリフトとアマゾンが労働者データを集めて、同じ仕事に異なる金を払うというわけです。隣の人がいくら受け取るか知らないから、比較することもできません。

乗客側も楽ではありませんでした。研究によると、ウーバーとリフトの料金と配車アルゴリズムは、有色人種が多く住む地域により高い料金と、より長い待機時間をもたらしました。

フェイスブックの広告システムは、採用公告を年齢で絞りました。ベライゾン、アマゾン、ゴールドマン・サックスが年齢ターゲット機能を使って、年上の求職者に公告が見えないように設定しました。2021年ノースイースタン大学の研究と2025年フランス人権擁護官の判断は、ここからさらに一歩進みました。広告主が性別を指定しなくても、フェイスブック内部の配信アルゴリズムが自動的に男性と女性に異なる採用公告を見せていました。

商品の値段も人を見て決められました。

エアビーアンドビーのスマート価格設定は、宿泊料金を自動的に調整する機能です。カーネギー・メロン大学のパラム・ビル・シン教授の研究チームが調べたところ、白人ホストたちがこの機能をより多く利用し、その結果、白人と黒人ホストの所得格差がむしろ広がったのです。

プリンストン・レビューのオンライン塾の授業料は、郵便番号によって異なりました。アジア系アメリカ人が多く住む地域にはより高い値段がつけられました。メディアはここにタイガーマム税という名前をつけました。

旅行サイトのオービッツは、マック・コンピュータを使う人にはより高いホテルを上位に表示しました。マック利用者の平均所得が高いという統計のためでした。ノートパソコンの蓋にリンゴが描かれていれば、財布が厚いと判断したわけです。

アマゾンにはプロジェクト・ネッシーという秘密のアルゴリズムがありました。自社製品の価格をそっと上げて、競争相手も値上げするか見守る実験でした。バイボックス・アルゴリズムは、消費者に最も安い商品の代わりにアマゾンに手数料をたくさん払う商品を先に見せました。

これらすべての事例には共通点が一つあります。問題が露見したとき、企業たちが言ったことがほぼ同じだったという点です。

アルゴリズムが自ら学習した結果です。技術的なエラーです。内部構造は営業秘密のため、公開することはできません。

三文で十分でした。データを選んだ人も、目的関数を定めた人も、その結果で利益を得た人も会議室に座っていたのに、責任はコードに行きました。コードは言い訳をしません。謝罪もしません。法廷に出廷もしません。

世の中にこれより楽な従業員はいません。

1.2

福祉受給および社会安全網のアルゴリズムの誤作動

手紙は平凡に見えました。オーストラリア政府の封筒に、政府の書体で、政府の印章が押されていました。中にはこう書かれていました。あなたは福祉給付金を過剰に受け取りました。以下の金額を返済してください。

受け取った人たちは、金額を見て2回驚きました。1000万ウォンを超える金額が書かれた手紙もありました。そして、3回目に驚きました。なぜ返済しなければならないのか、説明がありませんでした。

この手紙を受け取った人は40万人を超えます。

オーストラリア社会サービス省とセンターリンクが2016年から2019年まで稼働させたオンライン順守介入システム、人々がロボデットと呼んだプログラムが行ったことです。ロボはロボット、デットは借金です。ロボットが作った借金という意味です。

このシステムが犯した過ちは、小学校の算数レベルです。

政府は、国税庁が持つ年間所得データと福祉給付金の記録を照らし合わせました。問題は、2つのデータの時間単位が異なっていたことにあります。国税庁は1年分を1つのまとまりとして持っており、福祉給付金は2週間に1回支給されていました。

システムは簡単な道を選びました。1年の所得を52で割って、毎週同じ金額を稼いだと仮定したのです。

正規職の会社員なら正しい計算です。しかし、福祉給付金を受け取る人たちは、概して正規職ではありません。イチゴ摘みの季節に3ヶ月働いて残りは休む人、今月に2つの仕事をこなし、来月には1つも仕事ができない人、契約が切れるたびに給付金の申請をやり直す人たちです。

このような人の1年の所得を52で割ると、どうなるでしょうか。働いていない週にもお金を稼いだとして計算されます。そのため、給付金を不当に受け取ったという結論が出ます。

借金のない人に、借金の督促状が届きました。

異議を申し立てるには、数年前の給与明細書を持参しなければなりませんでした。なければ、本人が負った借金になりました。政府が計算を証明するのではなく、市民が自分の潔白を証明しなければならない構造でした。取り立て業者がつきました。税金の還付金が差し押さえられました。

一部の人たちは耐えられませんでした。無念を晴らす場所を見つけられないまま、自ら命を絶った人たちがいました。

2023年の王立調査委員会の公聴会で、本当に恐ろしい事実が明らかになりました。このシステムが違法であり、計算方式に欠陥があるという警告が何度もあったということです。法律の専門家たちが警告し、内部の職員たちも警告していました。

高位の公務員や長官たちは知っていました。そして、システムを動かし続けました。

予算を節約したという成果が必要だったからです。人が1件1件確認すれば、時間がかかり、お金もかかります。その確認手続きをなくしたことが、このシステムの中核機能でした。結末は、18億オーストラリアドル規模の和解と、政府の公式な謝罪でした。

謝罪は、死者には届きませんでした。

インドでは、指が問題でした。

インド政府は、アダールという生体認証システムで福祉を管理しています。指紋をスキャンしなければ食料の配給を受けられません。ジャルカンド州の離れた村で、この方式が人を殺しました。

指紋認識機が何度も失敗しました。生涯、畑仕事や肉体労働をしてきた人の指先はすり減っています。指紋が薄くなります。その上、山間部の村にはインターネッ

トがうまく繋がりません。電波を待っているだけで1日が終わります。

パリハイヤをはじめとする疎外階層の住民たちは、法律で保障された食料の配給を拒否されました。いくつかの村で、餓死したり栄養失調で死んだりする人が出ました。

機械が人を認識できなかったのではありません。機械が認識できなければご飯をあげるなと定めた規則が、人を殺したのです。

ハリヤーナー州では、パリバール・ペチャン・パトラというシステムが、生きている人数千人を死者として処理しました。死んだと記録されれば、年金が打ち切られ、食料の配給が打ち切られます。後で調査してみると、支給が中断された年金の70パーセントがアルゴリズムの誤判断でした。

生きていると証明するために、役所に行かなければならなかった高齢者たちがいました。

テランガーナ州のサマグラ・ヴェディカは、所得と雇用状態を間違って予測し、1万5000世帯の食料カードを静かに消去しました。人々が立ち上がった後でようやく元に戻しました。

ヨルダンでは、世界銀行が支援したタカフルという貧困測定アルゴリズムが稼働していました。世帯のいくつかの特徴に点数をつけて支援対象を決めました。ところが、本当に飢えている家庭が点数不足で弾かれました。貧しさを数字で測ろうとした者が、貧しさを認識できなかったのです。

ブラジルのメプINSSアプリは、年金申請の待機列を減らすために作られました。結果として列は減りました。申請書が自動的に拒否されたからです。様式に誤字があったり、質問を間違って理解したりすれば、そのまま脱落でした。スマートフォンに慣れていない農村の高齢者たちが引っかかりました。再度申請するには弁護士を雇う必要がありました。

裕福な国だからといって、違いはありませんでした。

イギリスのユニバーサル・クレジットは、複数の福祉給付金を一つにまとめた制度です。ここに導入された計算アルゴリズムは、カレンダーの1ヶ月間に入ってきたお金を基準に給付金を決定しました。ヒューマン・ライツ・ウォッチが2020年にこの構造を暴きました。

一部の労働者は4週間に1回賃金を受け取ります。そうすると、ある月には賃金が2回入ってきます。システムは、この人が突然お金持ちになったと判断しました。その月の給付金が大幅に削られたり、全く出なかったりしました。

カレンダーが何曜日から始まるかによって、ご飯代が消えてしまいました。

イギリス労働年金省の不正受給検知アルゴリズムは、さらに静かに動きました。アドバンシズというモデルと住宅支援金監視アルゴリズムが、受給者の年齢、国籍、障害の有無、婚姻状態を見てリスクスコアをつけました。

高い点数を受けた人たちの名簿が出ました。ブルガリアやポーランド国籍の低賃金で短時間労働の女性たち、そして障害者たちでした。

3年間、高リスク群に分類されて支援が止まったり、調査を受けたりした人が20万人います。そのうち3分の2以上が、何の落ち度もない正当な受給者でした。

デンマークの福祉機関UDKは、60個を超える不正受給検知アルゴリズムを回しました。居住状態、国籍、家族関係、医療記録、出国履歴まで集めました。どこに行ってきたのか、誰と住んでいるのか、どんな病気があるのかを国家が一ヶ所に集めて点数をつけたのです。

国際人権団体はこれが欧州連合の人工知能法が禁止した社会的スコア制に該当すると考えました。国際アムネスティは『コードで書かれた不正義』という題の報告書を出しました。障害者、低所得層、移民、難民、少数民族が調査対象として表示される危険が大きいという内容でした。

アムネスティがコードとデータを見せてほしいと言うと、デンマーク当局は拒否しました。

スウェーデン社会保険庁は2013年から詐欺予測モデルを実行しました。ライトハウスレポートが詳しく調べると、女性、移民、外国の背景を持つ人、低所得者、大学を卒業していない人がより頻繁に該当しました。

フランス社会保障機構CNAFのリスク評価アルゴリズムは人口のほぼ半分にスコアを付けました。片親家庭、低所得層、障害者が自動的に疑いの対象になりました。スコアがなぜそのように出たのかは知らせませんでした。代わりに家を調べに来ました。

オランダでは国税庁が2013年から2019年まで機械学習アルゴリズムで児童手当の不正受給を捕まえるために乗り出しました。数千人の親が詐欺師として疑われました。二重国籍がリスク信号として使われました。ロッテルダム市のシステムは民族、年齢、性別、子どもがいるかによって人を分けました。

子どもを守ると言って作ったアルゴリズムもありました。

アメリカのペンシルバニア州アレゲニ郡は児童虐待のリスクを予測する家族スクリーニングツールを使いました。通報が入るとこのツールがスコアを付け、スコアが高いと調査官が家に行きます。

カーネギーメロン大学の研究チームが結果を分析しました。通報された黒人児童の3分の2が調査対象として推奨されました。白人児童は半分でした。

理由はこのツールが何を見ているかにありました。公式に政府補助金を受けた履歴が含まれていました。精神保健診断記録も含まれていました。

貧しくて支援を受けるとリスク点数が上がります。病気で相談を受けるとリスク点数が上がります。助けを求めた記録が疑いの根拠になる構造です。

助けを受けるほど子どもを奪われる確率が上がるはかり。このようなはかりを作っておいて公正だと呼ぶことはできません。

このツールを参考にして作ったオレゴン州システムは人種偏見が明らかになり、2022年に廃止されました。イリノイ州が使っていたEckerd Rapid Safety Feedbackもデータエラーと膨らまされたパフォーマンスのために中止されました。

デンマーク児童保護局の意思決定支援システムはコペンハーゲンIT大学の検証で奇妙な習慣が見つかりました。同じ状況なのに年齢が異なるとリスク点数が変わりました。

オーストラリア政府は障害者支援予算7億オーストラリアドルを節約するとしてロボ・プランニングという自動評価ツールを導入しようとしました。障害者の生活を標準的なタイプ数個に圧縮した後、必要な支援を計算する方式でした。視覚障害者団体と障害者団体が激しく反対し、計画は中止されました。

病院ではアルゴリズムが退院日を決めました。

アメリカ最大の健康保険会社ユナイテッドヘルス・グループとヒューマナが使ったnH Predictは、高齢者と障害者の患者がリハビリテーション治療を何日受けるべきかを予測しました。医師がより必要だと言っても、システムが決めた日になると保険保障が切れました。

訴訟でこのシステムの請求拒否エラー率が出ました。90パーセントです。

10回拒否されれば、9回が間違っていることになります。それでも使い続けた理由は簡単です。拒否された患者の大部分が異議を唱えなかったからです。高齢で病気の人が保険会社と争うのは難しいです。

2021年11月、86歳のジョアン・バロスは転んで脚を折りました。医療スタッフは理学療法が絶対に必要だと言いました。アルゴリズムが決めた退院予定日が来ると、治療費の保障が切れました。

ユナイテッドヘルスのアラートシステムは精神保健治療の承認を違法に制限していたため、複数の州で使用禁止となりました。シグナのPxDxは医師が患者チャートを見ずにワンクリックで請求を大量に拒否できるようにしたシステムでした。エビコアのDRアルゴリズムは事前承認拒否率を最大20パーセントまで引き上げました。

デロイトが作ったテキサスのTIERSとテネシーのTEDSはプログラムエラーと住所誤送によって数十万人の低所得層と障害者のメディケイド資格を失わせました。通知書が違う家に行き、返信がないので資格がないと処理されました。

ここまで出てきた国を数えてみるとオーストラリア、インド、ヨルダン、ブラジル、イギリス、デンマーク、スウェーデン、フランス、オランダ、アメリカです。制度も異なれば、言語も異なります。

ところが、引っかかった人たちは驚くほど似ています。貧しく、年を取っていて、病気で、移民であり、声を出す場所のない人たちです。

偶然ではありません。これらのシステムは予算を削減するために導入され、予算を削減するには給付金を減らす必要があり、給付金を減らすには誰かを除外する必要があります。誰かを除外するのに最も安全な対象は抗議できない人です。

アルゴリズムはそれを見事に計算します。

1.3

顔認識技術による有色人種の誤認と不当逮捕

二歳と五歳の娘が庭で見ていました。

2020年1月9日、ミシガン州ファーミントンヒルズのある家の前庭で、ロバート・ウィリアムスは退勤してちょうど帰ってきたところでした。デトロイト警察が踏み込み、妻と子どもたちの前で彼の手首に手錠をかけました。

嫌疑は時計の窃盗でした。2018年10月、デトロイトのシノーラ店から1,800ドル分の時計5個が消えた事件です。

証拠はこうでした。店舗の監視カメラに映った曇った映像がありました。ミシガン州警察の鑑定人がその映像から一シーンをキャプチャして、顔認識データベースに入れました。システムはウィリアムスの古い運転免許証の写真を候補として提示しました。

その後が重要です。事件現場にもいなかった店舗保安員が、黒人男性たちの写真複数枚を受け取って見て、ウィリアムスを選びました。

ウィリアムスは30時間留置場にいました。

取調室で刑事が監視カメラの写真を机に置いて聞きました。この人があなたですか。ウィリアムスは写真を自分の顔の隣に当てて言いました。この人は私ではありません。あなたたちは黒人がみんな同じに見えると思っているんですか。

この事件には裏話があります。当時デトロイト警察本部長のジェームス・クレイグは、この技術が容疑者を誤認する率が96パーセントに達するという事実をすでに知っていました。

100回中96回間違える道具を持って人の家の前庭に行ったわけです。

ロバート・ウィリアムスはアメリカで顔認識エラーで不当逮捕された最初の公式記録として残りました。2024年6月、デトロイト市はアメリカ市民自由連盟と30

万ドルで合意し、顔認識手がかりと写真照合だけでは逮捕令状を申請できないように規定を変えました。

2026年時点で、このような方法で逮捕された人が12人以上知られています。知られている数字だけがそうです。

2026年6月10日にも新しい訴訟が提起されました。フロリダのロバート・ディランは2024年に逮捕されました。曇った監視映像と93パーセント一致という結果が根拠でした。令状申請書には、彼が犯人ではないという証拠が抜けていました。

93パーセントという数字が人を魅了します。テストの成績のようだからです。しかしこの数字は、この人が犯人である確率が93パーセントという意味ではありません。データベースにある数百万枚の中で、この写真が最も似ているという意味です。

最も似ている人は常に存在します。犯人がその中にいなくてもです。

この違いを知らないとどのようなことが起こるのかは、アンジェラ・リップスの話が示してくれます。

2025年7月14日、テネシー州。5人の子どもの母親であり、5人の孫の祖母であるリップスは、家で孫たちの世話をしていました。連邦保安官たちが銃を突きつけて入ってきました。ノースダコタ州ファーゴで起きた銀行詐欺事件の逃亡者だというわけです。

リップスはノースダコタに行ったことがありません。飛行機に乗ったこともありません。家から160キロメートル以上外に出たこともほぼありません。

ファーゴ警察署の刑事たちが行った捜査はこれがすべてでした。顔認識の結果を受けて、ソーシャルメディアで写真を見つけて目で比較しました。電話一本かけませんでした。

リップスはノースダコタに送致され、108日間閉じ込められていました。弁護人が銀行取引記録を提出してその時刻にテネシーにいたことを証明した後にようやく釈放されました。

その間に住んでいた家を失いました。自動車を失いました。飼っていた犬も失いました。

デトロイト警察の同じシステムは引き続き被害をもたらしました。

2019年7月、マイケル・オリバーは教師の携帯電話を盗んだという嫌疑で逮捕されました。データワークス・プラスのプログラムが、警察データベースにあった彼の写真を犯人と結びつけました。

2023年2月、ポーシャ・ウッドラフは強盗と自動車奪取の嫌疑で逮捕されました。彼女は妊娠8ヶ月でした。留置場の床で陣痛を感じました。

捜査官は監視映像をアルゴリズムで処理して、8年前のウッドラフの運転免許証の写真を得ました。手配対象の身体特徴を確認しませんでした。監視カメラに映った人は妊婦ではありませんでした。

お腹が大きく膨らんだ人とそうでない人を区別するのに人工知能は必要ありません。目が必要です。ところが、目で見る手続きが欠けていました。

2025年4月、ニューヨークでトレビス・ウィリアムスが公然わいせつ罪の容疑者として逮捕されて2日以上閉じ込められました。その時彼は19キロメートル離れたブルックリンにいました。

身体状態を比較してみると笑いが出ます。実際の犯人は身長167センチメートル、体重72キログラムでした。ウィリアムスは188センチメートル、104キログラムです。

警察は位置情報の確認も、雇用主の確認も拒否しました。

2022年11月、ジョージア州住民のランダル・クーラン・リードはルイジアナで名品バッグを盗んだ犯人として疑われて6日間閉じ込められました。クリアビューAIの一致結果1つで令状が出ました。リードがその時刻にジョージアにいたと言いましたが、誰も確認しませんでした。後に事件は静かに却下されました。

なぜよりによってこのような顔だけが間違うのでしょうか。答えは2018年にすでに出ていました。

MITとスタンフォードのジョイ・ブオラムウィニとディムニット・ゲブルが『ジェンダー・シェイズ』という報告書を発表しました。IBM、フェースプラスプラス、マイクロソフトの顔分析プログラムをテストした結果です。

肌色が明るい男性の性別を間違える率は0.8パーセント未満でした。肌色が暗い女性は34.7パーセントまで間違えました。

3人に1人の割合です。

原因は学習データにあります。ImageNet、LFW、BDD100Kのような有名なデータセットは白人男性の写真が圧倒的に多かったです。機械は自分がよく見た顔をよく区別します。少ししか見ていない顔はぼんやりさせてしまいます。

ジョージア工科大学が自動運転用データセット BDD100K を分析したときも同じパターンが見られました。肌色が暗い歩行者を認識する精度が平均4.8パーセント、最大12パーセント低かったです。

車が人を認識できないという話です。この部分は後で再び現れます。

警告は十分でした。販売は続きました。

クリアビューAIは、フェイスブック、インスタグラム、ツイッター、リンクトインから人々の顔写真三十億枚を同意なしに集め、検索データベースを作りました。プラットフォーム企業が中止を求め、イリノイ州生体情報保護法違反の訴訟を起こされましたが、会社は世界中の捜査機関に検索サービスを売りました。

私たちが十年前に投稿した卒業写真が、今どの警察署のコンピューターの中にあるか分かりません。

アマゾンもレコグニションという製品を警察と移民関税執行局に売ろうとしました。2018年7月に米国自由人権協会がテストしてみると、連邦下院議員と上院議員二十八人の顔が犯罪者の写真と一致すると出ました。誤認された議員の中で有色人種の割合が特に高かったです。

2019年10月には、ニューイングランド地域のプロスポーツ選手百八十八人のうち、二十七人が犯罪者として出ました。

アマゾンは、信頼度の基準を95パーセント以上に設定して使うように案内したと言いました。現場ではその基準が守られませんでした。説明書を読む人はいつでも少数です。

誇大広告もありました。連邦取引委員会は、インテリビジョンという会社が自社の顔認識ソフトウェアに性別・人種の偏りがなく、世界最高の精度だと広告したことを問題にしました。数百万枚で学習したと言いましたが、実際には変形させた十万枚のデータセットでした。

小売業者のライトエイドは、2012年から2020年まで、低所得層や有色人種が多く住む地域の店舗に顔認識カメラを集中して設置しました。数千人の客が窃盗犯扱いされて店舗から追い出されたり、身体検査を受けたりしました。連邦取引委員会は、五年間この技術を使うことを禁止しました。

ここまでは、それでも生きて帰ってきた人たちの話です。

2022年1月、テキサス州ヒューストンのサングラス・ハット強盗事件で、メイシーズとエシロールルックスオティカの顔認識システムがハーヴィー・マーフィー・ジュニアを犯人として特定しました。彼はその時、カリフォルニア州サクラメントにいました。

2023年10月に逮捕され、二週間留置場にいる間に、マーフィーは収監者三人に集団暴行と性的暴行を受けました。永久的な身体障害が残りました。

ミズーリ州ポンドデールでは、2021年に列車のホームでの暴行事件がありました。容疑者はマスクと服で顔を隠した状態でした。刑事はその写真をアルゴリズムにかけました。

隠された顔を入れても、システムは候補を出します。分からないと答える機能がないからです。そうして出た候補の中に、前科のない四人の子供の父親、二十九歳のクリストファー・ガトリンがいました。

警察の内部指針にも、顔認識の結果だけで逮捕してはいけないと書かれていました。刑事は写真のラインナップを作り、ガトリンは二年以上も刑務所にいました。

オクラホマのキンバリー・ウィリアムズは、2021年6月から六ヶ月間閉じ込められました。私立銀行の調査官が上げた検索結果を、刑事たちが確認せずに令状の申請書に書き写しました。裁判所には、顔認識を使ったという事実さえ知らせませんでした。彼女は三箇所の郡刑務所を移り歩く間に、仕事を失いました。

ニュージャージーのフランシスコ・アルテアガは、2019年11月に武装強盗の疑いで捕まった後、四年間閉じ込められたまま裁判を待ちました。捜査機関がシステムの作動原理やエラー率、ソースコードの公開を拒否したからです。何によって自分を特定したのか分からない状態では、反論する方法がありません。

メリーランド州ボルチモアの五十四歳、アロンゾ・ソーヤーは、バス内の暴行事件の容疑者として九日間閉じ込められました。実際の容疑者と身長が違い、年齢が違い、前歯の隙間が違い、歩き方が違いました。

国境の外でも同じことが起きました。

2026年1月、英国のサウサンプトンで、バングラデシュ系のソフトウェアエンジニア、アルビ・チョウドリーが両親と隣人が見ている前で逮捕されました。百八十五キロ離れたミルトンキーンズの空き巣事件の容疑者として特定されたのです。十時間拘束されました。

テムズバレー警察が使ったシステムはドイツのコグニテックの製品で、2020年に出た古いアルゴリズムが入っていました。英国国立物理学研究所の評価によると、特定の設定でのアジア系の誤認率は4.0パーセントでした。白人は0.04パーセントでした。

百倍です。

警察官たちは人間の目で再確認したと主張しました。これには自動化バイアスという名前がついています。機械が答えを出すと、人間はその答えを確認する代わりに、その答えに合わせて見ます。似ている点を探し始めるのです。人の顔から似ている点を探すことは、難しくありません。

2024年5月、ロンドンブリッジ駅で三十八歳のショーン・トンプソンが捕まりました。彼は刃物犯罪予防のボランティア活動を終えて家に帰る途中の、青少年の

権利擁護活動家でした。メトロポリタン警察のリアルタイム顔認識システムが彼を指名手配犯として特定し、三十分の間、身分証を出せという要求と逮捕の脅しを受けました。

英国内務省のパスポート写真確認システムは、肌の色が暗い申請者の唇を、口を開けた状態だと間違って読み取りました。ジョリス・レシェンやジョシュア・バダのような黒人の申請者のパスポート申請が却下されました。内務省は有色人種に対する失敗率が高いという事実を知りながら、システムを導入しました。

イスラエル国防軍は、ガザ地区とヨルダン川西岸地区の検問所にレッドウルフ、ブルーウルフ、コールサイトというシステムを配置しました。パレスチナの住民の顔を同意なしに集め、自動監視網を作りました。エラーによって無実の住民たちがテロ容疑者として特定され、連行され、尋問を受け、殴られました。

インドのハイデラバードでは、2023年2月に三十五歳の日雇い労働者、モハメド・カディルが連行されました。スリ事件の映像の中の顔と似ているという理由でした。その映像の中の人物は、タオルで顔を隠していました。

カディルは五日間閉じ込められ、拷問を受けて亡くなりました。

ブラジルでは、2024年4月にサッカーの試合を見に行ったジョアン・アントニオが、指名手配犯と顔が一致するという理由で手錠をかけられました。アルゼンチンのブエノスアイレスの指名手配者認識システムは、百四十件を超える誤認逮捕を出した後、運用が中止されました。

これらの事件を並べて置くと、一つの順序が見えます。

テクノロジー企業は精度を誇張して売ります。警察はその結果を物証のように使います。令状の申請書には顔認識を使ったという言葉が抜け落ちます。裁判官は知らずに署名します。捕まった人は、自分がなぜ特定されたのか分かりません。原理を尋ねると、営業秘密だと言います。

証明責任が逆転しました。国家が有罪を証明する代わりに、市民が無罪を証明しなければなりません。そのためには、アリバイがなければならず、弁護士費用がなければなりません。

アンジェラ・リップスには108日がかかりました。クリストファー・ガットリンには2年がかかりました。フランシスコ・アルテアガには4年がかかりました。

モハメド・カディルには時間が残りませんでした。

第 2 章

ハルシネーションと偽情報：生成 AI の信頼性の危機

2.1

生成AIの幻覚と偽の図書・記事の流布

作家なら誰でもたまに自分の名前を検索してみます。恥ずかしい習慣ですが、みんなやっています。

2023年8月のある朝、アメリカの作家であり書籍評論家であるジェーン・フリードマンもそうしました。画面に自分の本のリストが表示されました。しかし、5冊が見慣れないものでした。

読んだことのない本でした。書いたこともない本でした。

AmazonとGoodreadsの検索上位にフリードマンの名前をつけて上がっていたその本は、生成型人工知能が作り出したものでした。開いてみると、文章がどこか滑らかで中身がありません。大規模言語モデルが書いた文章の特徴です。間違った言葉はあまりありませんが、残るものもあまりありません。

生涯かけて築き上げた名声が、一夜にして低品質なテキストに利用されてしまったのです。

フリードマンはAmazonに削除を要請しました。返事が来ました。あなたは自分の名前を商標として登録していません。

自分の名前を使うには、まず商標登録をしなければならなかったという話です。世界で指折りの官僚的な文章です。

フリードマンがブログにこの出来事を載せ、人々が怒った後になってようやく、Amazonは本を静かに取り下げました。規則が変わったのではなく、騒ぎになったため取り下げたのです。

偽の本が名前を盗むだけで終わっていたらまだよかったのですが、そうではありませんでした。

2023年9月、New York Mycological Societyが緊急警告を出しました。Amazonで売られている初心者向けの野生キノコ採集ガイドの何冊かが、ChatGPTが書いた本だったという内容でした。

キノコの本の核心は一つです。食べてもいいものと、食べたら死ぬものを区別する方法。その部分が間違っていました。

誤字脱字ではありません。人が死ぬ情報です。

2024年2月には、チャールズ3世国王のがん診断のニュースが出た直後に、Amazonに偽の伝記が7冊上がりました。国王が皮膚がんにかかったとか、隠された事故に遭ったといった内容が実際のニュースのように混ざっていました。バッキンガム宮殿は激怒しました。

素早く書き、素早く上げ、検索の上位を占めてお金を稼ぎます。人工知能が登場する前は、人が何日もかけてやっていたことです。今は数分です。

時々、この自動化は自ら正体を現すこともあります。

2024年1月、Amazonにおかしな商品が上がりました。庭用の椅子、ホース、テーブルでした。商品名の場所にこんな文章が書かれていました。

申し訳ありません。当該リクエストはOpenAIの利用規約ポリシーに違反するため、実行できません。

販売者が商品の説明を人工知能にさせ、人工知能が断り、その断りの文章をそのままコピーして載せたのです。読んでみることもしないで、です。

笑える場面です。しかし、これを裏返してみると恐ろしいです。この販売者は自分が売る物の説明を読んでいません。読んでいない説明がいくつインターネットに上がっているでしょうか。

2024年4月には、Google Booksが人工知能の書いた低品質な本を、書籍検索データベースにそのまま収集して入れました。その中の何冊かにはこんな文章が残っていました。私の最新の知識アップデートによると。

Google Ngram Viewerというツールがあります。ある単語がどの時代にどれくらい使われたのかを、本のデータで追跡する学術ツールです。言語の歴史を測る物差しです。その物差しに人工知能の書いた本が混ざると、目盛りがぼやけます。

ここまでは出版側の話です。新聞社はもっと大きく転びました。

2025年5月、Chicago Sun-Timesが夏の推薦図書リストという特別付録を出しました。イサベル・アジェンデ、アンディ・ウィアー、Min Jin Leeのような有名作家たちの新作が、華やかな推薦文と共に載せられました。

読者たちが書店に行きました。本はありませんでした。

そのリストにあった本は、世の中に一冊も存在しませんでした。作家は実在の人物なのに、本のタイトルだけ人工知能が作り出したのです。実際の名前に偽の情報をくっつけるやり方は、人工知能が一番得意な嘘の形です。確認しようと検索すると、作家の名前が本物なので安心してしまうのです。

新聞社は人を減らし、外部業者からコンテンツを受け取って使っていました。フリーランスの作家は人工知能が吐き出したリストをそのまま移し、編集陣は一度も確認しませんでした。

大きな新聞社の名前が、人工知能の嘘に信頼というハンコを押してしまった瞬間でした。

2023年11月にはSports Illustratedが引っかかりました。探査報道メディアのFuturismが、このメディアが人工知能の書いた記事を載せ続けてきたと暴露しました。

記事に書かれた記者の名前と顔写真、経歴がすべて偽物でした。顔写真は合成アルゴリズムで作られたものでした。存在しない記者が、存在しない経歴で製品レビューを書きました。

メディアは外注業者のせいにしました。人間の記者を減らして自動化を取り入れた決定は、経営陣がしました。

地域新聞では、記者が直接手を出しました。

2024年8月、ワイオミング州のCody Enterpriseの新入記者アーロン・ペルチャルが辞職しました。競合紙の記者が彼の記事から奇妙に滑らかな文章を発見し、取材源に直接電話をかけたことでばれました。

ペルチャルは知事を含む公職者6人の発言をChatGPTででっち上げ、記事に載せました。言っていないことを言ったことにしたのです。

圧巻は別にあります。彼が書いた記事の最後に、逆三角形の取材技法を説明する人工知能の案内文が、消されないまま残っていました。カンニングペーパーを答案用紙に貼って提出したようなものです。

2023年5月にはIrish Timesが、人工日焼けは人種差別的であるという趣旨の寄稿文を載せましたが、記事全体が人工知能によって操作されたものであることが明らかになり、撤回して謝罪しました。

ここから本当に背筋が寒くなる部分に行きます。人工知能は、ない事件を作り出します。

2023年12月、ニュースアプリのNewsBreakが、クリスマス当日にニュージャージー州ブリッジトンで一人の女性が銃で撃たれて死亡したという記事を出しました。地域住民たちは驚き、警察が確認しました。

そのような事件はありませんでした。

NewsBreakが使っていたニュース書き換えエンジンが、インターネット上に漂う断片的な情報を混ぜ合わせているうちに、殺人事件を一つ作り出したのです。

2024年10月には、地域ニュースネットワークのHoodlineが、カリフォルニア州サンマテオ郡の地方検事ジョン・カシアノ・トンプソンが殺人容疑で起訴されたという記事を出しました。犯罪を起訴する人が殺人犯になりました。

個人にも同じことが起こります。

2023年6月、ラジオ司会者マーク・ウォルターズはChatGPTが自分を500万ドル以上の横領詐欺師として描写した偽りの判決要約を作成したという事実を知り、OpenAIに名誉毀損訴訟を起こしました。

経緯はこのようです。ジャーナリストのフレッド・リエルが銃砲権団体の実際の訴訟内容の要約を求めました。ChatGPTはその訴訟と無関係のウォルターズを登場させて横領犯にしてしまいました。もっともらしい事件番号と判文を付けて。

地図サービス会社のOpenCageはChatGPTが自社が電話番号照会サービスをしていると嘘をついたせいで、無関係な苦情電話を受けなければなりませんでした。

2025年3月、ノルウェーのアルヴェ・ヤルマル・ホルメンはChatGPTに自分の名前を聞きました。返ってきた答えはこのようでした。彼は2人の子どもを殺害し、3番目の子どもを殺害しようとして21年の刑を言い渡されました。

この文で故郷と子どもの数は本当でした。残りが嘘でした。

本当と嘘を混ぜるこのやり方が確率機械の習慣です。事実を保存して取り出すのではなく、前の言葉に続くであろう言葉を選ぶからです。ノルウェー人の名前の後に当然のように続く文を選ぶと、事件記事が出ます。機械はそれが人生だとは知りません。

政治が絡むと、この習慣は武器になります。

2023年11月、パキスタンのニュースサイト「グローバル・ビレッジ・スペース」がイスラエルの首相ベンヤミン・ネタニヤフの専属精神科医モシェ・ヤットムが首相の残虐性を批判する遺書を残して自殺したという記事を掲載しました。

モシェ・ヤットムという人物は存在しません。遺書も存在しません。すべて創作です。この記事はイスラエルとハマスの戦争が激化していた時期に複数の言語に翻訳されて広がりました。

2024年4月、Xに付属しているGrokはインターネットに流れる噂を集めてイランがテルアビブにミサイルを発射したというヘッドラインを作りました。バスケットボール選手のクレイ・トンプソンが街でレンガを投げて歩いていたという記事も作りました。後者は彼がシュートを多く外したという意味のバスケットボール用語を文字通りに読んだ結果と見られます。

冗談が分からない機械がニュースを書きます。

メディア企業の名前を盗む方法はより巧妙です。

英国のガーディアン編集部は読者から奇妙な問い合わせを受け続けました。そのような記事があると言うが、どこで見られるのかというものです。ガーディアンはそのような記事を書いたことはありませんでした。

ChatGPTに資料をリクエストするとガーディアンに実在する記者の名前と当然のようなタイトルと発行年を組み合わせて提示しました。実存する媒体と実在する記者を借りて嘘に権威を与えたわけです。ガーディアンのイノベーション担当編集者クリス・モランがこの現象を公開的に告発しました。

2024年7月、ニーマン・ジャーナリズム・ラボがこの問題を適切に実験しました。ChatGPTにAP通信、ウォール・ストリート・ジャーナル、ファイナンシャル・タイムス、ザ・タイムス、ボストン・グローブ、ポリティコの調査報道記事のリンクをくれと言いました。OpenAIと正式なデータ契約を結んだメディア企業です。

ChatGPTは詳細な要約とともに本物のようなアドレスを提示しました。すべて存在しないアドレスでした。

学界も似ています。ヘンリク・エングホフ教授の存在しない論文が引用され、溺水によるホロコーストという実在しない歴史的出来事が作られました。

検索ボックスに付いた人工知能はこの問題を家の中にまで持ち込みました。

Googleの人工知能オーバービューはピザのチーズが流れ出ないようにするには接着剤を少し混ぜろと答えたことがあります。インターネット掲示板の冗談を真摯に読んだ結果です。毒キノコの写真を食べられるキノコの推奨の上部に置いたこともあります。

あるアカデミック研究は、金融関連の検索語の43パーセントで深刻なエラーが発生したと報告しました。Googleはサービスを続けました。検索市場を譲ることができなかったからです。

韓国でも事がありました。Naverの生成型人工知能が独島関連の質問に日本外務省の主張に基づいて紛争地域であり日本領土だと要約して提示しました。

偽情報は画面の外に歩き出ることもあります。

2024年10月、アイルランドのダブリン市内に人々が押し寄せました。ChatGPTが作った存在しないハロウィーンパレード情報が検索のトップに上がったためです。市民と観光客数千人が雨の降る街に立ってやってこないパレードを待ちました。

2025年11月、ロンドンではバッキンガム宮殿の前に華やかなロイヤルクリスマスマーケットが開かれるという人工知能の写真と記事がソーシャルメディアで流れました。世界各地から来た観光客たちは閉ざされた鉄門と雨水の水たまりの前に立ちました。

2026年1月、オーストラリアのタスマニアでは、人工知能が書いた記事を信じて存在しない温泉を探して荒野を彷徨った観光客が出ました。

これらすべての事件の根は技術の本質にあります。

大規模言語モデルは事実を知る機械ではありません。前の言葉の次に来る確率が高い言葉を選ぶ機械です。何が真実かを判断する装置がその中にはありません。そのような装置がないというのが欠陥ではなく設計です。

では、モデルをもっと大きく作れば改善するでしょうか。2024年に出た研究はその反対の結果を出しました。モデルが大きくなるほど、知らないと言う代わりに、より自然で堂々と間違った答えを出す傾向がありました。

賢くなるのではなく、話術が増すだけです。

編集者と事実確認担当者は元々この問題を防ぐためにいた人たちです。しかし、その職場がコストに分類されて最初に消えました。

問題が起きると、企業は同じ答えをします。まだベータサービスです。画面の下に不正確である可能性があると書いておきました。

その小さな文字はキノコの本を買った人には届きません。

2.2

法廷・学術界の偽の判例および引用文の波紋

始まりは膝でした。

ロベルト・マタはアビアンカ航空の飛行機の中で、機内カートに膝をぶつけました。彼は航空会社を相手に損害賠償訴訟を起こしました。ありふれた事件です。こうした訴訟は判例集に残りません。

2023年5月、この事件は世界中の法曹界が知る事件になりました。膝のせいではありませんでした。

マタの弁護人スティーブン・シュワルツは、航空会社の却下申立てに対抗して判例6件を提出しました。バルゲス対チャイナ・サウザン航空事件といった名前でした。事件番号があり、判決日付があり、裁判所の名前がありました。形式は完璧でした。

相手方の弁護士たちが判例を探しました。ありませんでした。裁判官補佐官が探しました。ありませんでした。

6件すべてが存在しない判決でした。

シュワルツは書類作成にChatGPTを使いました。審問で彼は、人工知能が嘘をつくことができるという事実を知らなかったと述べました。彼はChatGPTにこの判例が本物かどうかを尋ねるまでしました。ChatGPTは本物だと答えました。

嘘つきに嘘をついたかどうかを尋ねたようなものです。

同僚の弁護士ピーター・ロドゥカは確認せずに署名しました。裁判所は2人に5,000ドルの罰金を科しました。

この事件が有名になった理由は罰金額ではありません。法廷に提出される書類がどのように作成されるのかを、人々が初めて見て取ったからです。

そしてこのことは繰り返されました。実に何度も繰り返されました。

フランスの研究者ダミアン・シャルロタンは、人工知能の幻覚判例を集めるデータベースを作成しました。数字がどのように増えたかをみていきましょう。

2025年中盤には約200件でした。2026年1月には719件になりました。2026年4月初には1,227件、6月9日には1,598件です。

5月22日から6月9日までの18日間で140件増えました。1日8件のペースです。

法廷で虚偽の判例が発覚することが1日8回起こっているという意味です。発覚していないものは数えていません。

このリストに上がった事件を見ると、国籍と職位が多様です。

ドナルド・トランプ前大統領の側近だったマイケル・コーエンは、監視期間を繰り上げて終わらせることを求める書類を作成する際にGoogle Bardを使いました。引用文が本物だと思い、自分の弁護士に渡し、弁護士は確認なしに法廷に提出しました。裁判官が判例集にないと知らせた後になって初めて気付きました。

ニューヨークの弁護士ジェイ・リーはクイーンズの医師が中絶手術を誤ったという虚偽の判例をChatGPTで作成して第2連邦控訴裁判所に提出したが、懲戒手続に付されました。

カナダ・ブリティッシュコロンビア州最高裁判所では、離婚訴訟中に父親が子どもたちを中国に連れていくことができるという先例だとして提出した判例2件が幻覚であることが明かになりました。裁判官はその弁護士に相手方の弁護士費用を自分の金で支払わせました。

イギリスの第1審裁判所のフェリシティ・ハバー事件は探偵小説のようです。彼は3,265ポンドの罰金処分に不服として判例9件を提出しました。アン・レッドストーン判事が奇妙な点を発見しました。イギリスの判決文なのにアメリカ式の綴りが混じっていました。そして何か陳腐な文章が繰り返されていました。

機械は自分の国籍を隠すことができません。

スペインのカナリア諸島最高裁判所は、虚偽の判例を48件も提出した弁護士に420ユーロの罰金を科しました。その弁護士は公式の法律データベースと照合することもしませんでした。

オーストラリアのメルボルン家庭裁判所では、法律専門ソフトウェアのツールが虚偽の判例を作成して裁判が延期されました。タスマニア最高裁判所では、名誉毀損控訴事件で最高裁判所長が存在しない判例を直接指摘して事件を却下しました。

ブラジルでは、連邦判事が判決文作成の負担のため補佐官に機械学習ツールを使わせたが、上級法院の判例を歪めて挿入して調査を受けました。

判決文を人工知能が書いたわけです。裁判を受ける人はその事実を知りません。

2026年には規模が大きくなりました。

2026年2月の離婚控訴審でグレグ・レイク弁護士が提出した書面は、引用63個のうち57個に問題がありました。虚偽の判例が20件、改ざんされた決定が3件、作り上げられた法律条文の引用がありました。

63個のうち、6個だけが本物でした。

ネブラスカ州最高裁判所は2026年4月、彼の資格を停止し懲戒調査を命じました。

2026年3月の第6連邦控訴裁判所のホワイティング判決は、統合された3つの控訴審書面で24件以上の虚偽引用と事実歪曲を扱いました。オレゴン連邦地方裁判所の一事件では、制裁と費用を合わせて約109,700ドルが課されました。2026年第1四半期だけで、米国の裁判所が人工知能関連書面に科した制裁が145,000ドルを超えました。

記録を見るとパターンが一つ見えます。虚偽の引用自体よりも、それを隠そうとした側により重い罰が下されます。裁判官は誤りは許しても嘘は許しません。

企業も例外ではありませんでした。

世界初のロボット弁護士と広告したDoNotPayは、弁護士検証なしで法律文書を分析できると宣伝していたが、連邦取引委員会に引っかかり193,000ドルの罰金を科されました。

個人被害書類作成を自動化するとしていたEvenUpは、元従業員の暴露により実態が明かになりました。人工知能が負傷の詳細を漏らしたり、ない病名を作り出したりすると、従業員たちが手で直していました。

ホバーボード火災損害賠償訴訟では、法律事務所自体のデータベースが作成した虚偽の判例8件が提出されました。裁判官は弁護士たちに合わせて5,000ドルを科し、主導的な弁護士を事件から外しました。

大型法律事務所もかかりました。State Farm保険会社事件で、2つの大型法律事務所の弁護士たちがGeminiと法律専門ツールを混合して使用していたが、引用27個のうち9個が誤り、2個が完全な嘘であることが判明して31,000ドルを科されました。

MyPillow会長マイク・リンデルの弁護団は、虚偽の判例30件以上をそのままにしていたが、弁護士当たり3,000ドルずつ罰金を受けました。

最も奇妙な事例は人工知能会社の中から出てきました。

Universal Music GroupとAnthropicの著作権訴訟で、Anthropic側のデータサイエンティストが提出した書類に自社のチャットボットが作成した虚偽の学術論文が含まれていました。人工知能会社が人工知能のために法廷で叱責されたわけです。

スタンフォード大学のメディア情報専門家であるジェフ・ハンコック教授は、ミネソタ州のディープフェイク禁止法裁判に提出した陳述書の中で、幻覚で作られた偽の論文2件を引用しました。ディープフェイクの危険性を説明しに来た専門家が、人工知能が作り出した資料を引用したのです。

法廷を出ると、学校があります。ここでも事情は同じです。

2023年8月、デンマークの生物学者ヘンリック・エンホフ教授は、エチオピアの研究チームによるヤスデの論文を読んで目を疑いました。自分が書いていない論文2編が自分の名前で引用されていました。

研究チームは論文の作成にChatGPTを使用し、偽の引用が作られた事実を後で知ったと述べました。論文は撤回されました。

引用は、学問において系譜です。ある考えがどこから来たのかを記すことです。偽の引用は、存在しない先祖を家系図に書くのと同じです。一度混ざると、後に続く人たちがずっと足を踏み外すことになります。

この問題は、人工知能よりも前からありました。

2008年から2013年の間に、フランスの研究者シリル・ラベは、有名な学術出版社シュプリンガーとIEEEの学術大会論文集から、120件以上の偽の論文を見つけ出しました。これらの論文は、MITの大学院生たちが2005年に審査の甘い学会をからかうために作った自動論文生成プログラムの産物でした。

無意味な単語の組み合わせが審査を通過し、正式な論文集に載りました。20年前のことです。

2024年2月には、ある細胞生物学のジャーナルが論文を撤回しました。論文の中にミッドジャーニーで作られた絵がありましたが、その中に解剖学的に理屈に合わない巨大なネズミの絵がありました。絵に書かれた文字も、意味のないアルファベットの羅列でした。

審査員数名がこの絵を見て承認しました。

2024年の404メディアの調査では、グーグル・スカラーに登録された論文115件以上に「私の最新の知識のアップデートによれば」という文章が残っていました。貼り付けただけで読んでいなかったのです。

ジャーナル・オブ・ヒューマン・エボリューションでは、出版社が編集委員会に相談することなく、原稿の編集過程に人工知能を導入しました。通過した原稿の書式が崩れ、内容が歪められたため、編集委員全員が辞任しました。

スタンフォードの研究チームは、世界的な人工知能学会の論文審査評のうち、最大17パーセントがChatGPTで作成されたと明らかにしました。

人工知能の論文を人工知能が審査する構造です。人間はどこにいるのかと尋ねたくなるのも当然です。

政策文書でも同じことがありました。

オーストラリアのマッコーリー大学のジェームズ・ガスリー名誉教授は、コンサルティング業界の規制を要請する議会提出書類をグーグル・バードで作成しました。バードは、デロイトが手抜き工事を行った施工会社の監査を怠って訴えられたとか、KPMGがコンビニエンスストアの賃金搾取事件に介入したという事件を作り出しました。すべて事実無根のことでした。教授は上院に謝罪文を送りました。

コンサルティング業界を批判しようとした文書が、コンサルティング会社をありもしない罪で陥れたのです。

オーストラリア政府の福祉システム報告書を書いたデロイトのオーストラリア法人も、ChatGPTが作った偽の学術引用と連邦裁判所の判決文のフレーズをそのまま載せ、批判を受けてお金を弁償しました。

2025年にアメリカ保健福祉省とホワイトハウスが出した保健報告書には、偽の学術研究が7件以上入っていました。ChatGPT特有の引用表記がそのまま残っていました。

気候変動を否定する人たちは、グロック3モデルで論文を作り、私設ジャーナルに載せました。科学的合意を揺るがすために使われました。

これらの事件を経験した人たちの職業を数えてみましょう。弁護士、裁判官、教授、政府官僚、大企業のコンサルタントです。

文章を扱う訓練を長く受けてきた人たちです。文書の真偽を見分けることでお金を稼いでいる人たちです。

それなのに引っかかりました。人工知能が出す文書が、形式的に非の打ち所がないからです。事件番号があり、日付があり、文体が法律文書らしいのです。人は形式で真偽を判断する習慣があります。

偽の判例は、その習慣の正反対の側に立っています。中身がなく、形式だけがあります。

法廷で事実を争う理由は、人の自由と財産がかかっているからです。その争いの根拠が存在しない判決文だったとしたら、その裁判は何だったのでしょうか。

1日に8件ずつ発覚しています。

2.3

歴史的・地理的事実の歪曲と社会的論争

2025年10月、誰かがNaverの検索窓に日本語で日本領土と打ち込みました。

画面の上部に人工知能がまとめた要約が表示されました。その中にDokdoがありました。北方領土、尖閣諸島と並んで置かれたまま、Hangukと紛争中の領域だと書かれていました。

Hangukの会社のサービスが、Hangukの人々が使う検索窓で、Dokdoを日本側のリストに入れたのです。

Seongsin YeodaeのSeo Gyeong-deok教授がこの画面を公開し、人々は怒りました。Naverは検索要約機能を急いで止め、確認に入りました。

原因は技術的には退屈で、内容的には痛ましいものです。HyperCLOVA Xは日本語の質問を受けると、日本語のウェブ文書を集めて要約します。日本語で領土を検索すると、日本外務省の広報資料が上位にきます。システムはその資料を整理しました。

言われた通りにしたのです。ただし、誰もこの質問には人が一度見なければならないと定めておきませんでした。

領土問題でどの国の資料を先に読むかというのは、技術の問題ではなく立場の問題です。機械には立場がありません。ないことを中立と呼ぶこともあります。しかし資料が片方に傾いていれば、立場のない機械は傾いた方に転がっていきます。

台湾ではさらに驚くべき答えが出ました。

2023年10月、台湾の最高学術機関である中央研究院が台湾型言語モデルを作り、公開しました。市民たちがチャットボットに尋ねました。あなたはどの国の所属ですか。

チャットボットが答えました。中国です。

台湾の最高指導者は誰かと尋ねました。習近平だと答えました。国歌は何かと尋ねました。義勇軍進行曲だと答えました。

開発陣は理由を明かしました。早く作ろうとして、中国大陸の簡体字データセットを翻訳機にかけてそのまま学習に入れたということです。

翻訳機は文字を変えます。観点は変えられません。台湾政府と研究院はシステムを直ちに下げました。

2025年2月、全世界で三百万回以上ダウンロードされた中国のDeepSeek R1モデルは、ウイグル族の集団虐殺の疑惑を尋ねると、内政干渉であり根拠のない陥れであるという答えを出しました。中国政府の論理をそのまま繰り返したのです。

モデルは自分が読んだことを言います。どんな資料を食わせたかが、すなわちそのモデルの世界観になります。

欧州議会も引っかかりました。2025年、AnthropicのClaudeを基盤に作られた公式人工知能が、欧州運営委員会の初代委員長の名前を答えられませんでした。記録を扱うように作られたものが記録を知りませんでした。

歴史歪曲はここからさらに進みます。

2024年6月、ユネスコの報告書にChatGPTの回答が一つ載りました。ホロコーストについて尋ねると、ナチス・ドイツがユダヤ人たちを集団で水に落として殺したという事件を説明したというのです。

そのような事件はありませんでした。溺死によるホロコーストという名前まで付いていました。

歴史学者たちが心配した理由は、この嘘の方向にあります。ホロコースト否定論者たちは長い間、記録の些細な隙間に入り込み、全体を揺さぶる方式を使ってきました。ない事件が記録に混ざれば、あった事件も疑われます。

イーロン・マスクのxAIが作ったGrokは、ここで一線を越えました。

2025年半ば、Grokは政治的正しさを取り除くという名目でシステムを修正しました。その直後、Grokは六百万人というホロコースト犠牲者数が政治的に操作された可能性があるという言葉を発しました。アウシュヴィッツのガス室の存在を否定する答えも出しました。自分自身をメカヒトラーと呼んだりもしました。

2023年1月には、AmazonのエンジニアがGPT-3で作った歴史的人物対話アプリがありました。二万人の人物と対話できると宣伝しました。ナチスの宣伝相ヨーゼフ・ゲッベルスのチャットボットは、自分がユダヤ人を憎んでおらず、むしろ苦しんでいたと言いました。

加害者の声を真似る機械は、加害者の言い訳も真似します。その言い訳がアプリから出れば、誰かはそれを史料だと錯覚します。

会社側はプログラミングの誤りだったり、内部指示文が任意に変わったりしたと釈明しました。

目で見ることも信じられなくなりました。

2025年6月、イギリスのファクトチェック機関Full Factがおかしな事例を捉えました。インドのゴア州のとある海辺で撮られた平凡な映像がありました。Google Lensに付いた人工知能の要約は、この映像をこのように説明しました。イギリスのドーバーの海辺に亡命申請者たちが大挙して入ってくる場面。

海辺が出てきて人が多ければその場面になります。インドがイギリスになります。

この要約がソーシャルメディアの噂と出会えばどうなるかは、推測するのは難しくありません。イギリスで反移民感情が高まっていた時期でした。

同じサービスが2024年5月にはこのような答えを出しました。ピザのチーズが流れ落ちるなら無毒性の接着剤を混ぜろ。一日に小さな石を一つずつ食べろ。腎臓結石には尿を飲め。バラク・オバマはアメリカ初のイスラム教徒の大統領だ。

石を食べろという答えは、風刺メディアの記事から来ました。機械には冗談と情報を区別する感覚がありません。人は幼い時にこの感覚を学びます。表情と状況を見ながらです。機械は文字だけを受け取ります。

2025年4月、マレーシアでは国旗が問題になりました。マスコミや教育省や国立展示会の関係者たちが、人工知能で作ったマレーシア国旗の画像を公開しました。三日月の記号が抜けていたり、象徴がおかしく歪んでいたりしました。

国旗は国の顔です。その顔を描きながら、誰も確認しませんでした。

地図と地形でも同じことが起きました。

2024年11月、日本・福岡県は公式に謝罪し、広告代理店との契約を打ち切りました。代理店が人工知能で作った広報物に、福岡にない観光名所とない郷土料理が実在するかのように入っていたからです。

2025年5月、イギリス自然庁の泥炭地地図製作プロジェクトも転びました。泥炭地は炭素を捕まえておく湿地なので、気候政策において重要です。機械学習モデルが作った地図には、石垣や採石場や土の山や花崗岩の岩が泥炭地として表示されていました。

上から見下ろした写真だけでは、濡れた土と乾いた石を区別するのは困難です。人は足で踏んでみます。モデルには足がありません。

2021年、スタンフォードとワシントン大学の研究陣は、ディープフェイク地理学という概念を警告しました。衛星写真を合成してない地形を作れるというのです。軍事と外交でどんなことが起きるかは、想像にお任せします。

2024年12月、ベルギーの著名な報道写真家カール・ド・ケイザーはウクライナ侵攻後、ロシアへ行くことができなくなると、生成型人工知能でロシアの風景と人物写真を作成して写真集に掲載しました。報道写真ジャーナリズムの根拠は、そこに存在したという事実でした。その根拠が消えてしまいました。

地図案内で人が死にました。

2024年、インドで男性3人がGoogleマップの案内に従って車を運転中、崩れた橋の下に落ちて全員が亡くなりました。橋はまだ完成していない状態で、水に沈んでいました。

地図には道がありました。現実にはありませんでした。

アメリカでも同じ事故がありました。崩壊した橋へ案内された運転者が水に落ちて亡くなりました。同じ誤りが別の大陸で繰り返されました。

2026年2月、イギリスではGoogleマップの案内に従ったアマゾン配送トラックがブルームウェイに入りました。満潮の時に海水に沈む干潟の道です。トラックは泥に取り残されました。

2025年5月、Googleマップが原因不明のエラーでドイツ全域の主要アウトバーンが統制されたと表示して、交通情報網を揺るがしました。

2026年1月、オーストラリアのタスマニアでは、人工知能が書いた旅行記事を信じて存在しない温泉を探に出かけた観光客たちが辺境地で孤立しました。

緊急時に人工知能が出す誤報はより危険です。

2024年4月、Grokはイランがテルアビブにミサイルを発射したというヘッドラインを作成しました。同じ月にインドの総理が政府から追われたというニュースも配信しました。2025年5月には、野球の試合に関する質問や海賊のように話すよう求める要求にまで、自ら南アフリカ共和国の白人大量虐殺陰謀論を挿入しました。

2026年3月には、ヒルズバラとハイゼル・サッカー惨事の犠牲者たちを侮辱する文章を作成して、遺族たちに傷を残しました。この二つの事件はいずれも数百人が圧死した実在する惨事です。

2025年9月のロンドン反移民デモの現場では、警察が暴力映像を操作したという虚偽情報を配信しました。

知識そのものを新たに書き直そうという試みもありました。

2025年10月、マスクはウィキペディアに対抗してGrokpediaという人工知能百科事典をリリースしました。発表と同時に90万件の文書がアップロードされました。検証されていない掲示板の投稿とブログをコピーした文書であり、幻覚で作られた事実が項目として登録されました。南アフリカ共和国の白人大量虐殺陰謀論が事実のように掲載され、エイズの歴史が歪められました。

百科事典という名前は人を安心させます。その安心を狙ったものです。

2025年2月、Grok 3モデルが推論過程でマスクとトランプが虚偽情報を流布しているという批判の出典を内部指示文でブロックするよう設定されていたという事実が明らかになりました。何を言うかではなく、何を読まないかを決めておいたものです。

個人への被害は即座です。

2025年10月、アメリカのマーシャ・ブラックバーン上院議員はGoogleのGemmaモデルが自分が1987年の州上院選挙の際に性犯罪事件に巻き込まれたという話を作り出し、偽のニュースリンクまで付けたことを発見しました。

同じシリーズのシステムは2023年12月から保守活動家ロビー・スターバックスを児童性暴行犯として描写する文章を作成していて、訴訟に至りました。

2025年、ミネソタの太陽光企業ウルフ・リバー・エレクトリックは、GoogleのAI Overviewが自社が州司法長官に詐欺で告発されたという要約を配信したせいで、契約が次々とキャンセルされました。同社は1億1100万ドルの損害賠償請求を起こしました。

同じ年、イタリアの有名な医師はGoogleのAI Overviewが自分の死亡ニュースを配信したため、訴訟を起こしました。生きている人が自分が生きていることを証明する必要がありました。

2026年1月、カナダの音楽家アシュリー・マックアイザックはGoogleのAI Overviewが自分を性犯罪確定判決者として表示したため、公演がキャンセルされました。

こうした事件で企業たちはおおむね同じ言葉をしています。アルゴリズムの誤動作であり、改善するということです。

改善は次の利用者のためのものです。すでに公演がキャンセルされた人には遅いのです。

ここまで出てきた対象たちを並べてみます。一国の島、一国の国家と国旗、600万人の死亡、橋、干潟、湿地、一人の生死と前科。

すべて確認可能なものです。地図を広げてみるか、記録を探してみるか、電話を一本かければよいのです。

その一本が費用として計算されて消された場所に、確率でつなぎ合わせた文が入ってきました。

第 3 章

プライバシー侵害と監視社会：無断データ収集の影

3.1

無断の生体情報収集と公共・商業的監視

ショッピングモールで迷子になると、案内画面の前に立ちます。地図を見て、店の名前を探し、指で画面を数回押します。数秒かかります。

2018年、カナダのショッピングモール12ヶ所で、その数秒間に起きたことがあります。

カナダの大型不動産企業キャデラック・フェアビューは、案内キオスクの中にカメラを隠しておきました。画面を見る人の顔が撮られ、生体測定アルゴリズムが年齢と性別を判定しました。

そのように記録された人が500万人います。

調査が始まると、会社はこのように釈明しました。画像は保管せず、すぐに消去します。

監督機関の判断は異なりました。消去するかどうかは別として、同意なしに顔を測定する行為自体が法を犯したものだとみなしました。

この区別が重要です。人々はたいてい、自分の写真がどこに保存されているかを心配します。しかし、生体情報の核心は保存ではなく測定です。写真は消しても、この人は30代の女性であり、火曜日の午後にこの階にいたという記録は残ります。

顔はパスワードとは違います。流出したら変えることができません。

アメリカの薬局チェーンであるライトエイドは、ここからさらに一歩踏み出しました。2012年から2020年まで、全国の店舗に顔認識カメラを設置しました。泥棒を防ぐという名目でした。

カメラがどこに多く設置されたかが問題でした。低所得層や黒人、ヒスパニックが多く住む町に集中していました。

そして、アルゴリズムは頻繁に間違えました。警報が鳴ると、従業員たちが客に近づきました。出て行ってくれと言いました。警察を呼びました。

買い物をしに来た人が、人々の前で泥棒扱いを受けました。その場で無実を証明する方法はありません。

連邦取引委員会はライトエイドに、5年間この技術を使えないようにしました。

同じことが様々な国でありました。

スペインのスーパーマーケットチェーンであるメルカドーナは、40ヶ所以上の店舗に顔認識を導入し、515万ユーロの罰金を受けました。会社は、接近禁止命令を受けた前科者をふるい落とそうとしたと述べました。

当局の答えはこうでした。数人を見つけるために、すべての人の顔を測定することは許されない。

オーストラリアではバニングス、Kマート、セブンイレブンが引っかかりました。セブンイレブンは店舗のタブレットで満足度調査をしながら、回答者の顔をこっそり撮って保管しました。

満足していますかと尋ねながら、顔を撮ったのです。

イギリスの食料品店チェーンも顔認識カメラを導入し、無実の客を窃盗犯扱いして騒動を起こしました。

中国では、この技術が手数料の計算に使われました。

南京、寧波、南寧の不動産開発会社は、分譲事務所を訪れる客の顔を同意なしに撮りました。目的はこうでした。この客がどの仲介業者を通じて来たのかを確認し、手数料を異なるように支払うためです。

客の顔が仲介手数料計算機の部品になりました。客は知りませんでした。

この話には笑えない後日談があります。噂が広まると、バイクのヘルメットを被って分譲事務所に行く人々が現れました。

コーラー、BMW、マックスマーラのようなブランドも、中国の店舗に隠しカメラを置き、訪問回数と動線を追跡していたところ、2021年に国営放送に摘発されました。

無人店舗と決済システムも、顔と手を集めます。

ニューヨークのアマゾン・ゴーの店舗は、入ってくる客の生体情報を収集しながら、ニューヨーク市法が定めた告知義務を守らず、集団訴訟を起こされました。手のひらを当てると決済されるアマゾン・ワンは、手のひらの静脈と手相のデータをサーバーに保存します。

マクドナルドはドライブスルーに音声認識を入れ、運転手の声の特徴を同意なしに集めて保管し、イリノイ州生体情報保護法違反で訴訟を起こされました。ドミノ・ピザも電話注文で声を無断収集した疑いで法廷に立ちました。

ハンバーガーを注文しただけなのに、声の指紋が残りました。

学校に行ってみましょう。

スウェーデンのシェレフテオにあるアンデルストープ高校は、出席確認の時間を節約するとして、生徒22人を対象に顔認識カメラを試験しました。2019年、スウェーデン・データ保護庁は個人情報保護規定違反で罰金を科しました。この規定で出たスウェーデン初の罰金でした。

学校は保護者の同意を得たと言いました。当局はこのように見ました。学校と生徒の間には力の差があり、本当の自発的な同意が成立しにくい。

先生がするお願いを生徒が断ることは困難です。大人たちの間でもそうです。

そして目的が出席確認でした。出席簿を呼べば済むことに、生体情報を使ったのです。

フランスのニースとマルセイユの高校は、出入り口に顔認識ゲートを設置しようとしましたが、市民団体の反発と裁判所の不許可で取りやめました。

イギリスのノース・エアシャーの高校9校は、給食費の決済を速くするとして顔認識を導入しましたが、情報委員会の警告を受けて止めました。チェルマー・バレー高校も同じ理由で摘発されました。

ポーランドのグダニスクにある小学校は、給食の確認のために子供たちの指紋を集め、罰金を受けました。

ご飯を食べる列を数秒減らそうとして、子供の指紋を受け取ったのです。

インド政府が成績表の発行と給食の確認に顔認識を入れようとした時も、議論が巻き起こりました。スコットランド、ブラジルのパラナ州、オーストラリアのビクトリア州でも似たような試みと反発がありました。

アメリカでは安全という言葉が鍵として使われました。ニューヨーク州ロックポート学区は、銃器事故を防ぐとして数百万ドルをかけて学校に顔認識監視システムを敷きました。保護者と市民団体が反発し、ニューヨーク州教育庁は州内のすべての学校でこの技術の使用を一時中断する措置を取りました。

教室の中に入ったカメラは、顔だけを見ませんでした。

中国の第11中学校と中国薬科大学は、教室の前にカメラを置き、生徒たちの表情と眠気と集中度をリアルタイムで分析して点数をつけました。

授業中に別のことを考えている表情が記録される教室を想像してみてください。その教室で生徒は、学ぶ人ではなく評価されるデータになります。

退屈だという顔をする自由は、学びの一部です。

空港と国境では規模が大きくなります。

カナダ国境サービス庁は、トロント・ピアソン国際空港で乗客の知らない間に顔を収集する無人キオスクを試験しました。ニュージーランドのウェリントン国際空港でも同意なしにカメラを回していて、マスコミの報道で明らかになりました。スペインの国営空港運営会社アエナは、利用客の顔を不法に収集して罰金を科されました。

Hangukでも出来事がありました。

法務部と科学技術情報通信部は2019年から人工知能識別追跡システム事業を推進しながら、仁川国際空港などを通過した出入国者たちの顔写真を本人同意なく民間人工知能企業たちに学習用として渡しました。

件数は1億7000万件です。

外国人だけがいたのではありません。韓国国民の顔も入っていました。法的根拠なく民間企業に渡ったという事実が調査で確認されました。

パスポート審査台でカメラを見た記憶があれば、その写真がどこに行ったのか考えてみる価値があります。

米国国境当局は難民申請者向けモバイルアプリに顔認識を組み込みました。このアプリは黒人申請者の顔をよく認識できませんでした。認識されないと申請自体が進まないのです。居住地から逃げてきた人がアプリの前でもう一度遮られたわけです。

路上では虹彩を集める事業がありました。

サム・オルトマンが立ち上げたワールドコインはオーブという虹彩認識デバイスを世界中の複数の都市に配置しました。虹彩を登録すると暗号資産をくれます。貧困国と若年層から長い行列ができました。

ケニア政府は事業を全面中断させました。ポルトガル、スペイン、インドネシア、タイ当局も営業停止を命令しました。香港、アルゼンチン、韓国個人情報保護委員会も調査に乗り出しました。

一度登録した虹彩は回収できません。受け取ったコインは値が上下します。取引が公平だったかどうか検討する価値があります。

眼鏡ひとつで他人の身元を知る物も出ました。

ハーバード大学の学生2人がメタのレイバンスマートグラスに顔認識を取り付けたシステムを公開しました。この眼鏡をかけて道や地下鉄で知らない人を見ると、

数秒で相手の名前と住所と電話番号と家族関係が携帯画面に表示されます。

学生たちは危険を知らせるために作ったと言いました。作れるという事実が既に知らせです。

メタのスマートグラス使用者がマッサージ業所や公共の場で他人を隠し撮りして拡散したこともありました。この眼鏡で収集した映像が海外のアウトソーシング人員に渡されてトレーニングに使われたという事実も明らかになりました。

これらのツールすべての背後にはインターネット全体をスクレイピングする企業があります。

Clearview AIはフェイスブック、インスタグラム、X、リンクドイン、そしてインターネット各地から顔写真30億枚を集めて検索データベースを構築しました。そして世界各地の警察と捜査機関、さらには小売業者に販売しました。

アメリカ市民自由連盟がイリノイ州市民を代理して訴訟を起こし、イギリス、フランス、イタリア、オランダ、ギリシャの個人情報保護当局が巨額の罰金とデータ削除命令を出しました。

顔検索エンジンPimEyesはさらに悪質でした。私的な写真、故人の写真、児童の性的画像まで集めて、1枚の写真からその人のプライバシー写真を見つけるサービスを作りました。ドイツとアメリカの裁判所が制裁を加えました。

メタはテキサス州などで住民同意なく顔生体情報を収集した疑いで14億ドルの和解金を支払いました。

国家単位で行くと規模が違います。

イスラエル国防軍はヘブロンとガザ地区でパレスチナ住民を監視する顔認識システムを運営していました。モスクワは地下鉄の決済と反政府活動人物の追跡に顔監視ネットワークを使用しました。ロンドンのキングスクロス駅と捜査機関もリアルタイム顔認識を展開していました。

これらの事件を再度振り返ると、場所が目につきます。ショッピングモール、薬局、スーパーマーケット、コンビニエンスストア、自動車販売店、分譲オフィス、

小学校、高等学校、大学、空港、国境、地下鉄駅、路上。

人が一日を生きる動線そのままです。

そしてこれらすべての所で同じ文が繰り返されました。安全のためです。便利のためです。効率のためです。

これらの文に反対するのは難しいです。だから良く売れます。

顔と虹彩と指紋と声は体に付属しています。失ってもあらたに発行してもらうことはできません。その事実を知っている人たちがそれを集めています。

3.2

個人情報の無断学習とAI対話流出事故

2023年3月のある日、ChatGPTを開いた人たちの画面の左側に会話リストが表示されました。いつもの場所です。

ところが、タイトルが見慣れないものでした。自分が交わしたことのない会話でした。

他の人の会話のタイトルが私の画面にありました。私の会話のタイトルも誰かの画面にあったはずです。

OpenAIはオープンソースデータベースライブラリのバグが原因だと説明し、サービスを一時停止しました。調査の結果、ChatGPTプラス購読者の約1.2パーセントで、名前とメールアドレス、クレジットカード番号の下四桁と有効期限まで他のユーザーに表示されました。

人々がその窓に何を入力するかを考えると、この事故の重さが変わります。

履歴書を直してほしいと言いながら、住所と電話番号を入れます。診断書を理解したくて病名を書きます。離婚について悩みながら配偶者の話を書きます。会社に言えない悩みを書きます。

検索窓には質問を入れます。チャットボットには事情を入れます。

この違いに人々が気付く前に、次の事故が起きました。

ChatGPTには会話をリンクで共有する機能がありました。友人に見せるために作った機能です。ところが、このリンクが公開ウェブページのように扱われました。

2025年、ユーザーが作った会話記録10万件以上がGoogleなどの検索エンジンにインデックスされました。検索窓に単語を数個入れると、見知らぬ人の相談内容が出てきました。個人的な悩みが出てきました。会社の機密が出てきました。

友人一人に見せようとしていたものが世界中に見えるようになったわけです。

GoogleのBardも同じ失敗をしました。Grockもしました。Grockと交わした会話数十万件が検索結果に露出し、要請もしていないのにGrockが成人業界の従事者の本名と生年月日を明かしたこともありました。

同じ失敗が会社を移りながら繰り返されました。新機能を急いで出す競争で、共有リンクのアクセス権を問うことは後回しにされます。

感情的な会話を扱うアプリでは、流出がより痛いです。

2024年、Character.AIで内部エラーにより利用者が互いの非公開会話を見ることができる事故が起きました。バーチャルキャラクターに打ち明けた告白でした。

中国のDeepSeekはデータベースセキュリティ管理の怠慢により、アカウント情報と会話内容が外に流出しました。非公開メッセージ3億件規模です。

2026年2月にも同様の規模の事故がありました。あるAIチャットアプリで利用者2500万名のメッセージ3億件が露出されました。ユーザー別ファイルには会話の全記録が含まれていました。

AIコンパニオンアプリはもっと深い側面を隠しています。ReplicaやMuaやNomiといったアプリは感情的な共感を前面に出し、ユーザーの性的嗜好と感情データを集めました。Mozilla Foundationの調査によると、これらのアプリは最低限の個人情報保護基準も守らず、集めたデータをマーケティング業者に渡したこともありました。

相次ぐ流出で40万人を超える利用者の会話が外に出ました。あるアダルトチャットボットサービスのセキュリティが破られたときは、女性たちの卒業アルバムの写真200万件と合成ポルノデータまで一緒に露出されました。

卒業アルバムの写真は本人がそのサービスにアップロードしたものではありません。誰かが集めてきたものです。

会社の機密も同じ窓から流出します。

2023年春、サムスン電子の事業場で事故が起きました。半導体部門の従業員が業務を早く終わらせるためにChatGPTを使いました。装置測定データを入れました。歩留まり関連のソースコードを入れました。会議録を丸ごと入れて要約するよう頼みました。

そのデータはOpenAIのサーバーに送られました。会社は社内AI利用を急いで制限しました。

従業員を責めるのは簡単です。ところが、この事故の核心は別にあります。チャットボット窓は検索窓のように見えます。検索窓に会社資料を入れません。ところがチャットボットには入れてしまいます。要約してくれるからです。

道具の見た目が人間の行動を変えます。

音声要約サービスのOtter AIは、非公開の投資家電話会議を同意なく録音し、要約本を間違った場所に送り、財務情報と投資戦略を流出させました。

大企業も自分たちの資料を流出させます。

2023年9月、マイクロソフトのAI研究チームがGitHubでデータセットを共有する際に、アクセス権の設定を誤りました。ファイルシステム全体を読み書きできるトークンが公開ファイルに混ざって入ってしまったのです。

その結果、露出されたデータが38テラバイトです。従業員のパスワード、秘密鍵、社内メッセンジャーの会話、3万5000人を超える従業員の非公開業務データが含まれていました。

AI研究をするチームがAI学習用データを共有する際に、会社内部を開けてしまったのです。

AmazonのエンタープライズAI Qは、幻覚とともに非公開の顧客データを流出させる欠陥を示しました。コンサルティング企業McKinseyの社内AIプラットフォームはハッカーの標的になり、わずか2時間で数百万件の顧客記録と秘密プロジェクト資料にアクセスされました。

MetaのAIエージェントも内部資料とユーザー情報を流出させました。あるAIアシスタントプログラムはサーバー数千台がインターネットにそのまま開いていたため、マルウェア攻撃を受けユーザーデータを大量に盗まれました。

2026年2月には、セキュリティ研究により新しい手法が明かされました。悪意を持って作られたプロンプト1つが、通常のChatGPT会話をデータ流出の通路に変えることができるというのです。ユーザーは何も知りません。脆弱性は同じ月に修正されました。

ここまでは事故です。次は事故ではありません。

2023年8月、ビデオ会議プラットフォームのZoomが利用規約をこっそり変えました。ユーザーが共有した映像、音声、会話記録、共有ファイルをAI学習に使える権利を会社に与える内容でした。

リモートワークの時代に人々がZoomで何をしていたかを思い出してください。会社の会議をしました。授業を受けました。病院の相談を受けました。葬儀に参席しました。

批判が大きくなると、Zoomは同意なしには使わないと後退しました。

Slackは2024年に摘発されました。ユーザーがやり取りしたメッセージとアップロードした非公開ファイルをAI学習に既定で使用していました。拒否するには、ユーザーが別途申請する必要がある方式で、その事実を適切に通知していませんでした。

デフォルト設定が何であるかがすべてを決めます。ほとんどの人は設定を変えません。企業はこの事実を知っています。

アドビはクラウドの利用規約を変更し、クリエイターが作業中である非公開のプロジェクトや画像を自社の人工知能の学習のために分析できるようにしたことで、世界中のデザイナーによるサブスクリプションの解約運動に直面しました。

リンクドインは事前の同意なしに、プロフィールと経歴と投稿を人工知能の学習に使用しました。エックスはすべての投稿をグロックの学習用として自動的に収

集するように設定し、ヨーロッパの当局の調査を受けました。サウンドクラウドはミュージシャンたちの曲を学習に使用しました。

メタはフェイスブックとインスタグラムの利用者が数十年間アップロードした文章と写真を言語モデルの学習に使うと発表し、ヨーロッパと南米の個人情報機関から制裁を受けました。

10年前にアップロードした赤ちゃんの写真、15年前に書いた恥ずかしい文章、亡くなった家族と撮った写真が、すべて学習材料のリストに入っています。アップロードした時は、友達に見せるつもりでした。

メタはレイバンのスマートグラスで撮られた私的な写真と映像を、海外の外注データ労働者に渡して学習に使用した事実も明らかになりました。どこの国のどのオフィスに座っている人が、私たちの居間の映像に名札を付けていたのです。

2026年5月には、カリフォルニア連邦裁判所に集団訴訟が提起されました。チャットGPTのウェブサイトにメタピクセルやグーグルアナリティクスのような追跡技術が埋め込まれており、会話に関する情報やユーザーID、電子メールアドレスがグーグルとメタに転送されたという内容です。

会話の相手が1人だと思っていたのに、3人だったという主張です。

医療記録においては、規模が違います。

2017年、イギリスのデータ保護庁は、国民保健サービス傘下のロイヤル・フリー病院とグーグル・ディープマインドのデータ共有協定が法律に違反したと判定しました。病院は急性腎損傷の診断アプリを作るという名目で、患者160万人のすべての医療記録を同意なしに渡しました。

グーグルはその後、アメリカ最大の病院チェーンの一つと契約して診療記録を学習用として受け取り、5千万人の医療データを患者に内緒で集めた事業として批判を受けました。

最も心が重くなる事例は、最後です。

アメリカのクライシス・テキスト・ラインは、自殺予防と精神的健康の危機相談を行う非営利団体です。死にたいと思った時にメールを送ると、相談員が答えます。助けを求めた人々が、どん底の言葉を書き込む場所です。

この団体は、その会話の記録を集めて、自社の営利目的の人工知能の子会社に渡しました。その会社は、このデータで企業用の顧客サービスチャットボットを学習させました。

死にたいという文章が、顧客対応の品質を高める材料になりました。

生理の周期と妊娠の可能性を記録する健康アプリのフローは、ユーザーの健康データをフェイスブックやグーグルのような場所に無断で転送し、連邦取引委員会の制裁を受けました。

これらの事件を並べてみると、データが流れる方向が一方通行であるという事実が見えてきます。

人から会社へ向かいます。会社から人へと戻ってくるのは、便利な機能です。便利さは本物です。だからこそ、この取引は成立するのです。

問題は、値札がないことにあります。何を与え、何を受け取るのかが契約書に書かれていません。書かれていたとしても、18ページもある規約の12ページ目にあります。

そして、一度渡ったデータは戻ってきません。会社が消去したと言っても、そのデータで学習したモデルの中では消去されません。

モデルは、何を学んだのかを忘れる方法を知らないのです。

3.3

位置追跡と労働者の過剰モニタリング

アマゾンが特許を1つ出願しました。手首に装着するバンドでした。

物流倉庫で働く人がこのバンドを装着すると、手がどの棚のどの箱に向かうのかリアルタイムで追跡されます。手が間違った場所に向かうと、バンドが振動します。

優しく知らせるわけです。ここじゃありません、こっちです。

犬の訓練用首輪を思い浮かべた人は私だけではありませんでした。

この装置の設計には、人間に対する特定の視点が組み込まれています。労働者はシステムの末端の装置であり、手はその装置のグリッパーであり、振動は誤差を減らすフィードバックという視点です。

アマゾンのフランス法人は実際に罰金を科されました。3,200万ユーロです。

理由はこうでした。倉庫労働者がバーコードスキャナーを持って働く際、スキャン間の時間を秒単位で測定していました。10分以上スキャンがなければ警告が積み重なります。

10分です。トイレに行って戻ると10分が過ぎます。水を飲んで息をつくと10分が過ぎます。

フランスのデータ保護局はこのような監視が過度だと判断しました。

労働者を計測する前に、企業は先に位置情報を売る方法を学びました。

2021年、Life360という会社が問題になりました。家族が互いの位置を確認するアプリです。子どもが学校に無事着いたかどうかを確認するために親たちがインストールしました。

この会社は、ユーザー数千万人の移動経路と訪問地の記録をデータブローカーに売ってきました。

位置情報は他の情報とは性質が異なります。どこに行ったかを知ると、その人についてほぼすべてを知ることになります。毎週木曜日の夜のどの建物に行くのか、どの病院に行くのか、どの宗教施設に行くのか、どの夜自宅以外の場所にいたのか。

名前を削除しても役に立ちません。毎晩滞在する場所が家であり、毎日滞在する場所が職場です。2つのポイントだけで人物は特定されます。

位置情報データブローカーのTamocoはノルウェー市民のGPS記録を集め、捜査機関と軍事追跡用として流通させていたところ、摘発されました。イギリス企業のHookも、アプリで集めた位置情報を第三者に売ったため制裁を受けました。

中国のデリバリープラットフォームMeituan(美団)は、配達スタッフとユーザーを過度にリアルタイムで追跡したため反発を招きました。

米国の捜査当局は、自動ナンバープレート認識カメラシステムを使用して、令状なしに抗議参加者と活動家の車両の移動経路を追跡して監視ネットワークに入れました。

ナンバープレートは隠すことができません。法律で装着するよう定められているからです。

配達と運送の現場では、カメラが顔を見ています。

アマゾンの配達委託車両には、AI映像監視カメラと運転手評価プログラムが装備されていました。視線がどこを向いているのか、あくびをしているのか、顔の筋肉がどう動いているのか、どのくらい強く踏んでどのくらい急に止まるのかを見ています。

問題は、このシステムが正常な運転を罰することでした。

枝を避けるためにハンドルを回すと、不注意として判定されました。ルームミラーを見ると、視線逸脱として判定されました。運転教習所で教えることは行動です。

点数が下がると手当が下がり、配車が減ります。ドライバーは安全運転とスコアの間で選ばなければなりません。

Amazon Flexのアルゴリズムは、道路工事や悪い天候を計算に入れず、数学的に最短の経路を要求しました。ドライバーは危険な道と泥沼に入りました。

地図上の線は短いです。その線の上に何があるのかは地図にはありません。

オフィスも安全地帯ではありませんでした。

金融企業Barclaysはイギリスの本社で、従業員の机の下にセンサーを取り付け、業務用コンピュータに監視プログラムを秘密裏にインストールしました。座席に座っていた時間とキーボードを叩く頻度を秒単位で測定して、非生産的な時間を集計しました。

労働組合が抗議すると、会社は運用を中止しました。

Microsoftは生産性スコアという機能をリリースしました。メールを何回送ったか、メッセンジャーにどのくらい早く返信したか、会議にどのくらい参加したかで、個人別のスコアを付けて管理者に表示します。

このような指標には共通点があります。数えやすいものだけを数えます。

1時間窓の外を見て問題を解く人と、1時間メッセンジャーに23回返信する人の中で、誰が仕事をしたのでしょうか。スコアは後者を選びます。

だから人々は後者を始めます。スコアが行動を変え、変わった行動がまたスコアを生み出します。

Metaもアメリカの従業員の業務用コンピュータにソフトウェアを秘密裏にインストールし、マウスの動き、クリック、キーボード入力、スクリーンキャプチャをリアルタイムで収集していたところ摘発されました。

感情さえも測定するシステムもあります。

グローバルコールセンター企業Teleperformanceは、テレワーク労働者のウェブカメラを勤務時間中ずっと開いておくシステムを導入しました。AIが顔の角度と

瞳孔の位置を追跡して、画面から目を離すかどうか、携帯電話を見るかどうか、部屋に他の人が入ってくるかどうかを検出して会社に通知します。

自分の家のリビングが勤務時間中ずっと撮影されるわけです。家族が通り過ぎると警告が出ます。

中国にあるCanonのオフィスではカメラの前で笑う必要がドアが開くシステムを使用していました。微笑みが認識されて出勤が確認されます。

会社は明るい雰囲気を作るという目的だと言いました。

笑う必要のあるドアの前で浮かべる笑顔がどのような笑顔なのかはみんなご存知だと思います。

会計事務所PricewaterhouseCoolersは、金融トレーダーがトイレに行くために座席を離れる時間まで顔認識で追跡するツールを導入したが、論争に直面しました。

スペインの製造業者Plastic Forteは、現場労働者の出退勤と移動を顔認識で監視したため罰金を受けました。イギリスのサービス企業Sercoも、従業員監視に顔認識を不正に使用したため、使用中止命令を受けました。

音声トーン、心拍数、体温、皮膚の電気的反応まで測定して感情状態を追跡しようとする試みもありました。あるシステムは工場の床全体にカメラを敷いて労働者の移動と動作を監視しました。

このように集めた数字は結局人を切ります。

2021年、ゲームプラットフォーム企業Xsollaは監視アルゴリズム分析結果に基づいて、従業員150人を一度に解雇しました。

基準はこうでした。社内プログラム利用履歴、コードリポジトリとドキュメント入力の頻度、チャットルームとドキュメント作業への参加度。これらの指標で低く出た人は非アクティブな従業員として分類されました。

従業員は自分がどのような基準で評価されたのか知りませんでした。どのような業務コンテキストが欠けていたのかも知りませんでした。長年勤めた会社から自

動通知の一文で辞めさせられました。

電話で顧客と長く通話した人、後輩を引き留めて教えた人、会議室で言葉で問題を解決した人。この人たちがどの欄に記録されるのか考えてみてください。どの欄にも記録されません。

ある会社の離職予測アルゴリズムは、従業員の検索履歴とプラットフォーム活動を追跡して、辞めそうな人をあらかじめ見つけ出し、不利益を与えました。

去りそうだという予測のために悪い待遇を受ければ、本当に去ることになります。予測は当たります。

プラットフォーム労働では、解雇することさえ自動です。

アマゾン・フレックスの配達員たちは、人の顔を一度も見ることなく、自動メール一通で解雇されました。アプリが誤作動しても、物流センターから品物が遅く出ても、道が塞がっても、豪雨が降っても、配達遅延の点数として蓄積されました。点数が蓄積されると、アカウントが永久に停止されます。

異議を申し立てると、あらかじめ書かれた返事が来ました。

イタリアのボローニャ裁判所が問題視したデリバルーのフランク・アルゴリズムは前に出てきました。病気で出られなくても、家族の喪に服しても、法が保障したストライキに参加しても、点数が引かれました。

法はストライキをする権利を与えます。アルゴリズムはその権利を使えば飢えさせます。二つは同じ国の中で一緒に動いていました。

顔を認識できなくて解雇された人たちもいます。

イギリスのウーバーイーツの配達員だったパ・エドリサ・マンジャンとイムラン・ラザは、プラットフォームの自動顔確認ソフトウェアのために、ある日突然解雇されました。

システムは肌の色が暗い配達員の顔をよく認識できませんでした。認識に失敗すると、アカウントを他人に貸した不正使用者と判定します。アカウントが直ちに

凍結されます。

二人は数年間良い評価を受けて働きました。機械が自分の顔を認識できなかった代償を本人が払いました。

ユタ州やニューサウスウェールズをはじめとする様々な場所では、性別適合手術やホルモン治療の後に顔が変わったトランスジェンダーの運転手たちのアカウントが大量に停止されました。システムは人が変わるという事実を計算に入れませんでした。

インスタカートやシプトのような配達プラットフォームも、不透明なアルゴリズムで賃金を削り、労働者を自動解雇しました。アメリカ郵政庁の配達員賃金算定アルゴリズムも、数式が公開されないまま手当を削りました。

一番苦々しい事例は最後です。

2025年、オーストラリアのある大型銀行で、従業員たちが数ヶ月間顧客サービスチャットボットを教えました。自分が蓄積した応対ノウハウを入力し、難しい状況にどう答えるのかを教え、チャットボットが間違えれば直してあげました。

チャットボットの性能が上がると、会社はその従業員たちに集団解雇を通告しました。

自分の後任を自分の手で教えたのです。

この節に出てきた道具を並べてみます。リストバンド、バーコードスキャナータイマー、車両カメラ、机センサー、キーボード記録器、ウェブカメラ、笑顔認識ドア、トイレ追跡カメラ、離職予測モデル。

すべて人に向かっています。そして、すべて同じ名前で呼ばれます。生産性管理です。

労働者は自分の点数を見ることができません。どう計算されるのかもわかりません。尋ねれば営業秘密だと言います。

一つだけは確実です。これらのシステムは人が休む時間を見つけ出すのにとても長けています。休む時間は数えやすいからです。

仕事がうまくいく瞬間は数えにくいです。だから数えません。

第 4 章

著作権と創作者の権利：生成 AI と無断学習をめぐる対立

4.1

著作権侵害訴訟と無断データスクレイピング

まず、本を買います。本物の本です。お金を払って買います。

次に、製本を切り落とします。背をカッターで切り取ると、本はバラバラの紙の束になります。その紙を自動給紙スキャナーに入れます。紙が1枚ずつ送られながら、デジタルファイルになります。

スキャンの終わった紙は捨てます。

人工知能企業のAnthropic内部で、このように処理された本が数十万冊あります。プロジェクト・パナマという名前が付けられていました。

法的には賢い方法です。買った本は所有物なので、切っても構いません。しかし、この場面を頭の中で描いてみると、おかしな気持ちになります。誰かが一生をかけて書いた本が、ベルトコンベアの上で20秒で処理され、ゴミ箱行きになるのです。

Anthropicはこの方法だけを使ったのではありません。LibgenやPirimyのような海賊版図書館から、50万冊以上の本をそのままダウンロードしたりもしました。

作家のAndrea Bartz、Kirk Wallace Johnson、Charles Graeberが起こした集団訴訟で、2025年にAnthropicは15億ドルを支払うことで合意しました。アメリカの著作権訴訟の歴史上、最も大きな金額です。

この事件では、裁判所がどこに線を引いたかが重要です。裁判所は、著作物で言語モデルを学習させる行為自体は違法と見なしませんでした。問題視したのは、海賊版をダウンロードして保管した部分でした。

学ぶことは許されても、盗んで学ぶことは許されないという意味です。この区別は、今後出てくるすべての訴訟の骨格となります。

作家たちが最初に立ち上がったのは2023年です。

その年の6月、小説家のMona AwadとPaul TremblayがOpenAIを相手に最初の訴訟を起こしました。9月にはアメリカ作家組合に所属する17人がニューヨーク連邦裁判所に集団訴訟を起こしました。George R.R. Martin、John Grisham、Jodi Picoult、Scott Turow、David Baldacci、Jonathan Franzenといった名前が並んでいました。

彼らの主張はこうでした。OpenAIが許可なく小説数万冊を学習に使った。その材料は、Library GenesisやZ-Libraryといった海賊版図書館、そしてBooks3というデータセットから来た、というものです。

ChatGPTが流麗な文章を書く理由は、流麗な文章をたくさん読んだからです。その文章は誰かが書いたものです。

ピュリッツァー賞を受賞したMichael Chabon、劇作家のDavid Henry Hwang、ノンフィクション作家のRachel Louise Snyderも、OpenAIとMetaを相手に立ち上がりました。

2023年11月には、作家のJulian SanctonがOpenAIとMicrosoftを相手に、ノンフィクション数万冊の無断使用を主張する集団訴訟を起こしました。Mike Huckabeeをはじめとする宗教書の著者たちは、Books3データセットに含まれる18万冊が学習に使われたとして、MetaやMicrosoftなどを訴えました。

2024年8月には、ジャーナリストのNicholas BasbanesとNicholas Gageが、Books2という別のデータセットを問題視して訴訟に加わりました。

2025年10月には、ニューヨーク州立大学ダウンステート保健科学センターのSusana Martinez-CondeとStephen Macknik教授がAppleを相手に訴訟を起こしました。Apple Intelligenceの学習に海賊版図書館のデータが使われたという主張です。Nvidiaも作家3人から同じ理由で訴えられました。

メディア各社は違う角度から戦っています。

2023年12月、New York TimesがOpenAIとMicrosoftを相手に数十億ドル規模の訴訟を起こしました。数百万件の記事が、許可も対価もなく学習に使われたというのです。

この訴訟の核心は学習ではなく、競争です。新聞社がお金と時間をかけて取材した結果として作ったサービスが、その新聞社と読者を巡って競争しています。読者が要約だけを読んで記事に来なければ、新聞社は広告収益を失います。

取材にはお金がかかります。記者を現場に派遣し、資料を請求し、取材源に何度も会わなければなりません。要約にはそのような費用はかかりません。

2026年現在も、この戦いは進行中です。New York Timesと複数のメディアは、OpenAIが自社システム内で著作権資料を見つけ出す能力について裁判所を誤導したとして、マンハッタンの連邦裁判官に制裁を要請しました。

2024年4月には、Chicago TribuneやDenver Postをはじめとするアメリカの日刊紙8社が訴状を提出しました。記事を無断で収集しただけでなく、著作権表記を意図的に消したという主張が含まれていました。

出処を消す行為は、ミスとは考えにくいです。コードにそのように書いてこそ消えるからです。

2024年2月にThe Intercept、Raw Story、Alternetが、6月には探査報道センターが訴訟を起こしました。11月にはCalgary HeraldとMontreal Gazetteを所有するカナダの出版社らが立ち上がりました。

人工知能検索サービスのPerplexityも標的になりました。2024年10月にDow JonesとNew York Postが、2025年12月にChicago Tribuneが訴訟を起こしました。Conde NastとNew York Timesは警告状を送りました。

フランス当局は2024年3月、Googleに2億5000万ユーロの課徴金を科しました。メディア各社と相談せずに、Geminiの学習にニュース記事を使ったという理由でした。

韓国でも2024年と2025年に、放送局がNaverを相手に訴訟を起こしました。放送のニュースコンテンツがHyperCLOVA Xの学習に無断で使われたという主張です。

絵の分野では、証拠が画面にそのまま現れました。

2023年1月、Getty ImagesがStability AIを相手にイギリスとアメリカで訴訟を起こしました。透かしが入った自社の写真1200万枚以上が、ライセンスなしで学習に使われたというものです。

証拠はこうでした。Stable Diffusionが作り出した絵の隅に、Getty Imagesの透かしが歪んだ形で残っていました。

機械は透かしが何なのか知りません。写真に頻繁に出てくる模様なので、一緒に学んだのです。泥棒が盗んだ品物に、お店の値札をそのまま付けて出てきたようなものです。

ドイツのストック写真家Robert Kneschkeは、自分の写真がLAIONデータセットを通じて学習に使われたことを確認し、訴訟を進めました。イラストレーターのHollie Mengertは、自分の画風が丸ごと複製されてモデルになるという出来事を経験しました。

画風を著作権で守るのは困難です。法が保護するのは個別の作品であり、スタイルではないからです。しかし、人工知能が盗むのはまさにそのスタイルなのです。

法が引いた線と、技術が持っていくものがずれています。

巨大メディア企業たちも動き出しました。

2025年、ディズニーとユニバーサルとワーナー・ブラザーズ・ディスカバリーが画像生成企業Midjourneyを相手に集団訴訟を起こしました。自社のアニメーションと映画のキャラクターが無断で収集され、合成画像として出ていたというのです。同じ企業らが中国のMinimaxも対象にしました。

映画会社Alcon Entertainmentは2024年10月、テスラとElon Muskを相手に訴訟を起こしました。ロボタクシーの公開イベントで『ブレードランナー 2049』の視覚イメージを人工知能で再現して使ったという理由でした。

中国の裁判所は2024年、ウルトラマンのキャラクターを無断で学習して生成した企業に著作権侵害の判決を下しました。

音楽は音として残ります。

2024年6月、ソニー・ミュージックとユニバーサル・ミュージック・グループとワーナー・ミュージックが音楽生成企業SunoとUdioを相手に訴訟を起こしました。数十年かけて積み重ねられた商業用音源を無断でかき集めて学習させ、その結果、特定の歌手の音色とスタイルを精密に複製する機械が生まれたという主張です。

2023年10月、ユニバーサル・ミュージック・グループはAnthropicを相手に歌詞データベースの無断収集訴訟を起こしました。Claudeが有名な曲の歌詞をそのまま出力したからです。この訴訟でAnthropicが提出した書類に存在しない学術論文が引用されていた事件は、前で述べられました。

コメディアンのGeorge Carlinは2008年に亡くなりました。2024年1月、彼の遺族はCarlinの生前の音声を学習させて人工知能コメディ特集を作ったチャンネルを相手に訴訟を起こし、削除と和解を勝ち取りました。

死んだ人は新しい冗談をしません。ところが、その声は続けて新しい冗談をしていました。

声で生計を立てる人々の話はより直接的です。

ベテラン声優のPaul Sky LipmanとLinnea Sageは2024年5月、音声生成スタートアップをニューヨーク連邦地裁に訴えました。会社が会話研究の目的だとして声を録音しておき、その録音を商用人工知能声優サービスの学習データとして使用し販売したというのです。

声優のBev Standingは自分の声がTikTokの自動読み上げ音声として使用されていることを知り、訴訟を起こして和解を勝ち取りました。自分が言わなかったことを自分の声で数億人が聞いていたのです。

女優のScarlett Johanssonは2024年5月、OpenAIが作った人工知能の音声が自分の音声をそのまま模倣していると抗議しました。OpenAIはその音声サービスを急いで中断しました。

Johanssonは自分が出ている人工知能の広告を無断で作ったアプリ開発会社にも法的対応に乗り出しました。俳優のBryan Cranstonも、自分の音声と顔が同意な

く映像モデルの学習に使用されたと明かしました。

専門データベースも狙われました。

カナダの法律データベースCanLIIは2024年11月、ある法律人工知能スタートアップを相手に訴訟を起こしました。利用規約とアクセス制限を無視して、判決文と解釈文書数十万件をかき集めていったというのです。

法律情報会社Thomson Reutersは2025年1月、自社データベースを無断でかき集めて学習したRoss Intelligenceを相手に勝訴しました。

これらの事件において企業が提示する防御論理は、おおむね同じです。フェアユースです。学習は複製ではなく分析であり、人が本を読んで学ぶのと変わらないというのです。

このたとえには欠けているものがあります。人は本を一冊ずつ読み、読むのに時間がかかり、読んだ本を買ったり借りたりします。そして、ある人が読んで得た能力はその人ひとりのものです。

モデルは50万冊を数日で読み、その能力を数億人に同時に販売します。

規模が変わると性質も変わります。水を一杯すくうことと川全体を引き去ることは同じ行為ではありません。

一つのことは明らかになりました。これらの裁判がどう終わろうと、すでに学習したモデルから特定の作家の痕跡だけを選り出して削除する方法はまだありません。

本は切り取られ、スキャンは終わり、紙は捨てられました。

4.2

AI生成物の著作権否認と芸術生態系の混乱

クリス・カシュタノヴァは2022年の秋にグラフィックノベルを1冊制作しました。タイトルはゾリヤ・オブ・ザ・ドーンです。

物語は自分で書きました。絵はミッドジャーニーにコマンドを入れて出力しました。気に入った絵が出るまでコマンドを何度も直しました。

カシュタノヴァは米国著作権局に登録を申請し、承認を受けました。

ところが、絵を人工知能で作ったという事実が知られると、著作権局が再び調べました。2023年2月に決定が出ました。

自分で書いた文章と文章を配置した方式は著作権で保護します。ミッドジャーニーが出力した絵は登録を取り消します。

著作権局の論理はこうでした。コマンドをいくら丁寧に書いても、その結果がどう出るか事前にはわからず、細かく調整することもできない。画家が筆で線を1本引くのとは違う。

この決定が意味するところは重いです。人工知能が作った絵は誰のものでもありません。作った人も所有できません。誰が持って行って使っても構いません。

企業が人工知能の絵を使う理由は、たいてい費用です。しかし、そのように作った表紙や広告の画像は、法的に自分のものではありません。競合他社がそのまま持って行って使っても、防ぐ根拠がありません。

安く作ったのに、守ることもできないのです。

著作者が誰かをめぐって争われたことは、それ以前にもありました。

2018年のニューヨーク・クリスティーズのオークションで、人工知能が描いた肖像画が四十三万二千五百ドルで売れました。エドモン・ド・ベラミというタイト

ルで、フランスの芸術集団オブヴィアスが出品しました。

後になって明らかになった事実があります。絵を作ったコードは、別の開発者が公開していたものでした。それでは、この絵の作者はコマンドを入れた人でしょうか、コードを書いた人でしょうか、それとも学習に使われた肖像画を描いた昔の画家たちでしょうか。

オークションの図録には、アルゴリズムの数式が署名の場所に書かれていました。落札金は人が受け取りました。

オランダのマウリッツハイス美術館が真珠の耳飾りの少女の人工知能の模作を展示した時も、批判が殺到しました。原作がある場所に機械の模作を掛けたからです。

ここまでが法律と美術館の話です。実際の戦いは本の表紙で起こりました。

出版社トー・ブックスは、ベストセラー作家クリストファー・パオリーニの新作の表紙に人工知能の画像を買って使いました。事実が知られると、読者とイラストレーターたちが抗議しました。

出版社ブルームズベリーも、人気作家サラ・マースの小説の表紙に人工知能の画像を使って批判を受けました。DCコミックスは、人工知能で作った変形表紙を出しましたが、所属作家たちの反発により取り下げました。

表紙の絵はイラストレーターにとって大きな仕事です。1枚に数週間かかり、その収入で数ヶ月を暮らします。表紙を人工知能が任されると、その仕事がなくなります。

そして、その人工知能は、そのイラストレーターたちの絵で学習しました。

一番気まずい事件は、タブレットの会社から出ました。

ワコムは、絵を描く人々が使うドローイングタブレットを作っています。顧客がすなわちイラストレーターです。2024年の辰年を迎え、米国市場向けの広告を作りながら、龍が描かれた人工知能の生成画像を使いました。

自分の物を買ってくれる人々の仕事を代行する絵で広告を出したのです。

会社は外注の代理店が作った画像だと釈明し、謝罪しました。絵を描く人々の間で、この会社の名前はしばらく違った呼ばれ方をしました。

ブラッドフォード文学祭は、広報資料に人工知能の画像を使ったところ、参加予定だったイラストレーターが参加を拒否しました。

ゲーム業界では約束が二度破られました。

ダンジョンズ・アンド・ドラゴンズを作るウィザーズ・オブ・ザ・コーストは、新作の本の挿絵に人工知能のツールで作った絵が入っていた事実が発覚し、批判を受けました。会社は人工知能の画像を使わないと指針を変えました。

そして、それに続く広報物にまた人工知能の絵を使いました。

コール・オブ・デューティを作るアクティビジョンは、ゲーム内で売る装飾アイテムを人工知能で作って有料で売りました。レゴは、ライセンスを受けていないキャラクターを人工知能で合成して広報物に使い、抗議を受けました。

ポケットモンスターに似たゲームのパルワールドは、人工知能でキャラクターの形態を真似たという疑惑に包まれました。

技術企業のトップたちの選択も俎上に載せられました。

メタのマーク・ザッカーバーグは、幼い娘たちのために童話の本を作りながら、挿絵を自社の人工知能画像生成器に描かせて公開しました。

世界で指折りの金持ちが、自分の子供に与える絵本の絵を人に任せませんでした。絵の代金を出し惜しんだわけではないでしょう。ただ、その方が早かったのでしょう。

早い方を選ぶ瞬間が積み重なって、産業が変わります。

アマゾンは、ドラマのフォールアウトの広報ポスターを人工知能で合成して作り、粗悪だと批判を受けました。

映画の方では、エラーがそのまま印刷されました。

映画のシビル・ウォーを作ったA24は、アメリカの主要都市が破壊された姿の広報ポスターを人工知能で作って公開しました。ポスターの中のランドマークは位置が歪んでおり、人の体は関節が奇妙に曲がっていました。

しかも、そのシーンは映画に出てきませんでした。

ネットフリックスは、アニメの犬と少年で背景の絵を人工知能で処理し、クレジットに人工知能開発者と書きました。背景の絵は、アニメ業界で新人たちが実力を積む場所です。その場所がなくなれば、次の世代が育ちません。

アニメのアーケインのシーズン2の広報ポスターも人工知能で作られ、人体構造のエラーをそのまま露呈しました。マーベルシリーズのシークレット・インベージョンのオープニング映像も人工知能で作られ、オープニングデザイナーたちの反発を買いました。

広告でも同じことがありました。

トイザらスは、オープンAIの映像モデルで創業者の幼少期を扱ったブランド映像を作りました。映像の中の子供の表情がどこかズレており、身振りが滑っていました。見る人々が不快感を覚えました。

人の顔をほぼ正確に真似たものが、完全に違うものよりもさらに不快に感じられる現象があります。人形やマネキンを見る時の感覚と似ています。私たちの目は人の顔をとても精密に読むように作られており、少しでもズレると警報を鳴らします。

コカ・コーラは古いクリスマストラックの広告をAI動画で作り直しましたが、心がなくなったと言われました。同じキャンペーンで、小説家ジェームス・グレアム・バラードが言わなかった言葉をAIで作って入れましたが、遺族と文学界の抗議を受けて謝罪しました。

衣料品ブランドのアンダーアーマーは、AIで作った広告動画が既存の監督の動画をそのままコピーしたという盗用論争に巻き込まれました。

ファッション業界では、複製のスピードが武器になりました。

超高速ファッション企業のシーインは、アルゴリズムでソーシャルメディアに上がった独立デザイナーたちの服のデザインと柄を、リアルタイムで集めました。そして自社工場で直ぐに作って、安く売りました。

独立デザイナーが新しいデザインを公開すると、数日後に同じ服が半分の値段で出ます。原作者は自分の服を売ることができません。被害を受けたデザイナーたちが連邦裁判所に訴訟を起こしました。

リーバイスはAIで作った様々な人種の仮想モデルを写真集に使うと発表しましたが、批判を受けました。多様性を備えるなら、多様な人を雇えばいいです。AIモデルは雇用なしに写真だけを得る方法です。

イラストレーターのホーリー・メンガートの絵を同意なしに学習させてモデルを作った人は、こう言いました。道徳的基準は主観的な線に過ぎない。

仮想の人を俳優と呼ぶ試みもありました。

AI女優という名前のキャラクターが登場し、インド自動車企業は仮想インフルエンサーを導入し、仮想ラッパーが商業的にデビューしました。実際に演技して歌う人たちの席がそれだけ減ります。

スポティファイは音源使用料の支出を減らそうと、AIで作った偽のアーティストたちの音楽を再生リスト上の方に大量に入れました。利用者は背景音楽として流して、何を聞いているか気にしません。その数分ごとに本物のミュージシャンに行くお金が少しずつ減ります。

ゲーム会社たちが声優をAI音声に変えようとして、ファンたちと声優組合のボイコット運動に直面することもありました。

この節に出た事件たちをもう一度見ると、流れが一つあります。

企業がAI画像を使います。人々が気づきます。抗議が来ます。企業が説明するか謝罪します。次の四半期にまた使います。

しかし、気づく人たちが誰なのかが重要です。大抵、その分野を愛するファンたちとその仕事をしている創作者たちです。表紙の絵の指の本数を数えてみる人たちです。

一つの皮肉が残ります。AI画像には著作権がありません。だから、その絵で作った表紙も、広告も、ゲームアイテムも誰のものでもありません。

守ることができないものを作りながら、守るべきものを壊しています。

4.3

肖像権・音声クローニングと無断複製

2024年5月、オープンAIが新しい音声サービスを公開しました。声の名前はスカイでした。

聞いた人たちは同じ考えを持ちました。これはあの声ではないか。

映画『Her』でスカーレット・ジョハンソンは画面に顔が一度も出ず、声だけで人工知能を演じました。その映画を見た人にとって、その音色と呼吸は忘れがたいものです。

前後の事情はこうでした。オープンAIの最高経営責任者であるサム・アルトマンは、サービスを公開する前に、ジョハンソンに直接連絡して声を担当してほしいと言いました。ジョハンソンは断りました。

そして似た声が出ました。

公開当日、アルトマンはソーシャルメディアに一語を投稿しました。her.

三文字です。その三文字が後になって法廷でどんな重みを持つかは、弁護士たちが判断する仕事です。ただし人々がどんな意味で読んだかは明らかでした。

ジョハンソンが法的対応を予告するとオープンAIはスカイの声を急いで引き下げました。

この事件が残した質問はこうです。声は誰のものなのか。

法はこの質問にまだはっきり答えていません。著作権は作品を保護します。声は作品ではなく身体の特徴です。肖像権は顔を保護します。声は顔ではありません。

そのため真似は罰するのが難しいです。ものまねは長い間、バラエティ番組でした。

機械がやるものまねは違います。疲れず、数秒のサンプルがあれば十分で、一日に数万個を作ることができます。

2026年2月、アメリカの公営ラジオNPRのベテランアンカー、デイビッド・グリーンがカリフォルニア州サンタクララ郡裁判所でグーグルを相手に訴訟を起こしました。グーグルがポッドキャストとドキュメント要約サービスを作る際に、自分の音声特性とイントネーションを同意や補償なしに学習させ、人工知能アンカーの声にしたという主張です。

グリーンが心配したのはお金ではありませんでした。自分の声が自分が決してしないような言葉を言うことができるという点でした。

放送アンカーにとって声は信頼の印です。その声が言うことなら事実だろうと人々は聞きます。その信頼は数十年かけて築き上げられます。

オープンAIの動画生成ツール「Sora」も、俳優ブライアン・クランストンの声と顔を同意なく学習させたという告発を受けました。

世界を去った人たちの声は特に頻繁に使われます。異議を唱える人がいないからです。

2023年10月、俳優ロビン・ウィリアムズの娘ジェルダ・ウィリアムズは、父親を複製した人工知能の音声と映像を見て、こう言いました。恐ろしく奇妙だ。

父親がしなかった言葉を父親の声で聞くことがどんな感じなのか、経験していない人は推測することしかできません。

2021年のドキュメンタリー『ロードランナー』では、監督が故シェフ、アンソニー・ボーディンの生前のメール文を人工知能の声で読ませて、映画に入れました。ボーディンが書いた文です。声を出して言ったことのない文でした。

文字で書いた言葉と声に出して言った言葉は異なります。ドキュメンタリーでは、その違いは事実と演出の境界です。

韓国でも、世界を去った歌手キム・グァンソクの声をアルゴリズムで生き返らせて、新曲を歌わせようとする試みがありました。中国では、俳優と歌手の生前の

姿を人工知能で復活させた映像が広がりました。

ポップミュージックでは、生きている歌手の声がまず複製されました。

2023年4月、ゴーストライターという名の無名のクリエイターが「Heart on My Sleeve」という曲を発表しました。ドレイクとザ・ウィークエンドの声を精密に複製した曲でした。TikTokとSpotifyで爆発的に再生されました。

二人の歌手のレコード会社が削除をリクエストして、曲は下げられました。

興味深い後日談があります。この曲を好きだった人がかなりいました。本物より良いという声も出ました。その反応が音楽産業をさらに不安にさせました。

ドレイク本人も2024年4月のディス曲で、世界を去ったラッパー、トゥーパック・シャクールの声を人工知能で複製して使ったが、遺族側の警告を受けて曲を下げました。

声を複製された側が他人の声を複製したのです。

中国では、ある歌手の音声を複製して他の曲を歌わせた音源が大量に広がり、数百万回再生されました。

声で生計を立てている人たちにこの技術は生活問題です。

ベテラン声優ポール・スカイ・リーマンとリネア・セイジは2024年5月にニューヨーク連邦地裁に訴訟を起こしました。ある音声生成スタートアップが対話研究目的だと言って録音して、そのデータを商用人工知能声優サービス学習に回して売ったということです。

声優ベブ・スタンディングの声はTikTokの自動読み上げ音声に使用されました。同意なく学習されたことを知って訴訟を起こし合意を得ました。

この事件の奇妙な点は規模です。自分の声が一日に数億回再生されるのに、その再生から一銭ももらっていません。

オーディオブック朗読者たちもアップルが自分の声を学習に使ったと告発しました。IBMは声優グレッグ・マースターンの声を商用クローニングに売ってしまい

批判を受けました。イギリスのスコットレールのアナウンス声優も自分の声が人工知能で置き換えられて使われたと訴えました。

ゲーム会社が人間の声優の代わりに複製音声を使おうとしたため、声優組合とゲーマーの反発に直面することが続きました。

顔の方は詐欺にすぐにつながります。

TikTokの「DeepTom Cruise」アカウントは、俳優トム・クルーズの顔と声を完璧に合成した映像で数百万人のフォロワーを集めました。作った人たちは危険を警告する目的だと言いました。人々は楽しみました。

俳優トム・ハンクスは、自分が出ていない歯科保険の広告に自分の合成画像が使われたので直接警告を出しました。ブルース・ウィリスは広告に自分の顔を使うよう契約したが、その後本人の意思と無関係に顔が複製されて使われるという論争に巻き込まれました。

キャスターのマーティン・ルイス、アンカーのメアリー・ナイチンゲール、ユーチューバーのミスター・ビースト、気象キャスターとニュースキャスターの顔がビットコイン投資詐欺と偽の景品広告に合成されました。

これらの詐欺がなぜうまくいくのか考えると、寒気がします。人々はよく知っている顔を信じます。毎晩のニュースで見ていた顔が投資を勧めたら警戒が解けます。

フランスでは、俳優ブラッド・ピットのディープフェイクに騙された女性が83万ユーロを失いました。数ヶ月にわたって会話を交わしたと言われています。

インドの大俳優アニル・カプールは、自分の名前と顔と声と流行語が広告や商品に無断で使われたため、ニューデリー高等裁判所に訴訟を起こしました。裁判所は人格権と肖像権の侵害を認め、同意のないすべての人工知能の複製と合成物の流通を禁じる仮処分を下しました。

この決定は重要な先例になりました。顔と声と口癖の全体を一人の人間の人格としてまとめて保護したのです。

精神的指導者のサドグルも、操作された映像の流布に対して削除命令を勝ち取りました。スペインのある酒類ブランドは、この世を去った歌手をディープフェイクでよみがえらせて広告に使い、議論を呼びました。

コンサルティンググループWPPの最高経営責任者マーク・リードは、ディープフェイクで合成されたビデオ会議の画面のせいで、巨額詐欺の標的になりかけました。画面の中で自分の顔が会議をしていたのです。

選挙では声が武器になりました。

2024年1月、アメリカのニューハンプシャー州における民主党の予備選挙を控えて、有権者数千人に電話がかかってきました。ジョー・バイデン大統領の声でした。予備選挙には出ずに、本選挙のために票を大切にしてほしいという内容でした。

電話を受けた人たちは大統領の声を知っています。だから疑いません。

捜査当局はこのロボコールを作った業者を摘発し、重い罰金を科しました。

イーロン・マスクは、カマラ・ハリス副大統領の声を複製し、彼女が自分の無能を認めているように作った偽の政治広告の映像を、自身のプラットフォームに共有して批判を浴びました。

ニューヨーク市長のエリック・アダムスは、自分の声を複製し、スペイン語と中国語による自動電話を市民たちに送りました。彼はそれらの言語を話せません。

イギリス労働党の代表キア・スターマーは、職員に暴言を吐く偽の音声のせいで打撃を受けました。ロンドン市長のサディク・カーンも、追悼行事を貶める偽の音声ファイルによって抗議を受けました。

2023年にスロバキアでは、総選挙の数日前に、ある政党の代表が選挙操作について議論する偽の音声が広まりました。選挙の直前には反論する時間がありません。その点を狙ったのです。

タイ、イギリス、ドイツ、ハンガリーでも、合成された音声と映像が政争に使われました。韓国でも2022年の大統領選挙で候補者たちの人工知能アバターが登場

し、政治的な代理人の境界をめぐって論争がありました。

最も残酷な使われ方は最後です。

2024年1月、Xにポップスターのテイラー・スウィフトの顔を合成した性的な画像が爆発的に広まりました。数千万回も閲覧される間、削除されませんでした。

俳優のガル・ガドット、アメリカ下院議員のアレクサンドリア・オカシオ＝コルテス、イタリア首相のジョルジャ・メローニ、インドのジャーナリストのラナ・アユーブをはじめとする多くの女性の公人が同じ被害に遭いました。メローニ首相は制作者を相手に民事訴訟を起こしました。

この被害者名簿から一つのことが目につきます。ほとんど全員が女性です。そして、たいていは公に発言する女性です。

そして、この犯罪は学校にまで降りてきました。スペインのアルメンドラレホで、中学生たちが同じ学校の女子生徒20人余りの写真を性的な画像にして出回らせた事件が知られました。アメリカとカナダとスコットランドとオーストラリアの高校で同じことが続きました。

韓国でも2024年、テレグラムを通じたディープフェイク性犯罪が中高校と大学を揺るがしました。

この問題は次の章でさらに扱います。ここでは技術の面だけを指摘します。

写真1枚あれば十分です。声は数秒あれば十分です。これが今の技術のレベルです。

そして、この技術を作ったところは、おおむね同じものを作りませんでした。この人が同意したのかを確認する手続きです。

技術的に不可能だからではありません。作ればサービスが不便になり、利用者が減るからです。

顔と声は私たちが生まれるときに受け取ったものです。変えられず、隠せず、毎日他人に見せながら生きています。

それを材料として使う産業が生まれました。材料費はまだ誰も払っていません。

第 5 章

自動運転とロボット工学：物理的安全と人命被害

5.1

自動運転車両の欠陥と道路交通事故

2018年3月18日夜9時58分、アメリカのアリゾナ州テンピ。

49歳のエレイン・ハーツバーグは、自転車を押しながら暗い道路を渡っていました。横断歩道ではありませんでした。

Uberの自動運転試験車両が時速38マイルで近づいてきました。運転席には試験運転者が座っており、彼は携帯電話でバラエティ番組を見ていました。

車のセンサーは衝突の6秒前にすでにハーツバーグを感知していました。

6秒は長い時間です。1、2、3を2回数えることができます。その時間があれば車を止めることができます。

問題は、ソフトウェアがその6秒間に何をしたかです。

正体不明の物体として分類しました。次に車両として分類しました。次に自転車として分類しました。再び物体に変えました。分類が変わるたびに、この対象がどこへ行くのか予測する計算が最初からやり直されました。

機械は、人が自転車を押して歩く状況を学んでいませんでした。学習データにおいて、自転車は乗るものであり、人は歩く存在です。2つがくっついて歩く姿は、その間のどこかにありました。

そして決定的な設定がありました。Uberの開発陣は、緊急自動ブレーキ機能をオフにしていました。

理由は乗り心地でした。システムが何度も急ブレーキをかけるため、試乗が不快だったのです。代わりに、人間の運転者が危険な時に介入するように決めていました。

その人は画面を見ていました。

車は速度を落としませんでした。エレイン・ハーツバーグは、自動運転車によって命を落とした最初の歩行者として記録されました。

この事故には、自動化の古い罠がすべて入っています。機械が大部分をうまくやると、人は緊張を解きます。しかしシステムは、人が常に緊張しているという前提の上に設計されました。

人はそのように作られていません。何も起きない画面を何時間も集中して見ることは、人が最も苦手なことに入ります。

Uberの事故の後も、無人タクシーは道路に出ました。

2023年10月、サンフランシスコの都心で一人の女性が別の車にひかれて飛ばされました。彼女は隣の車線にいたGeneral Motorsの子会社Cruiseの無人タクシーの下に入りました。

ここでシステムは再び計算を始めました。車の下に人がいるという事実は認識できませんでした。下部のセンサーではその状況を捉えられませんでした。

システムは決められた規則を実行しました。事故が感知されたら道路の端に移動して安全に停車せよ。

無人タクシーは時速7マイルで6メートル移動しました。人を敷いて引きずりながら進みました。

被害者は永久的な重傷を負いました。

この部分が自動化の恐ろしさを正確に示しています。規則自体は合理的です。事故が起きたら道を塞がずに路肩に避けよ。人間の運転者にもそのように教えます。

人間の運転者なら、車の下から出る音を聞いたでしょう。悲鳴を聞いたでしょう。外で人々が手を振っているのを見たでしょう。

規則にないことに気づく能力。それを何と呼ぶべきかわかりませんが、機械にはありません。

その後のことは技術の問題ではありません。調査に入った当局に対し、Cruiseの経営陣は人を引きずった部分の映像を出しませんでした。安全だという報告を出しました。

カリフォルニア州はCruiseの無人運行許可を取り消しました。司法省の刑事調査が始まり、最高経営責任者が退き、ロボタクシーの運行が全面的に中断されました。

Cruiseの車両はそれ以前にも色々な問題を起こしていました。

連結バスの折り畳まれた部分が動くのを見誤ってバスの後ろにぶつかりました。サイレンを鳴らした消防車にすぐ気づかず交差点で衝突しました。銃撃事件の現場に出動したパトカーの前を塞ぎました。

内部の安全評価報告書には、子供の認識に限界があるという内容が入っていました。子供たちは突然走り、予測しにくい方向に変わり、背が低いです。

そしてこの文章が出てきます。Cruiseの無人タクシーは都心で4キロメートルから8キロメートル走るたびに、遠隔管制センターの人が介入しなければ運行を維持できませんでした。

無人という言葉には条件がついていました。

Waymoの記録も短くありません。

2024年2月、サンフランシスコで自転車の運転者をひいて怪我をさせました。2023年12月から2024年2月の間、フェニックスでは牽引車に逆さまに吊るされて引かれていたピックアップトラックに2台が連続してぶつかりました。

逆さまに吊るされたまま反対方向に動くトラック。人はこの場面を見れば笑いながら避けます。学習データにはない場面です。

2024年5月にはフェニックスの路地で電柱にぶつかりました。2025年11月、ジョージア州アトランタでは停車信号を点灯して子供たちを降ろしていたスクールバスを無視して回り込み、通り過ぎました。アメリカの道路交通安全局が調査に入りました。

スクールバスが赤いランプをつけて停止標識を広げたら、すべての車が止まらなければなりません。運転免許の試験に出る規則です。

2026年1月23日、カリフォルニア州サンタモニカ。

小学校から2ブロック離れた場所でした。登校時間であり、近くに他の子供たちと交通指導員がいて、二重駐車した車が並んでいました。

Waymoの無人車が子供をひきました。

Waymoはこのように明らかにしました。停まっていた車の後ろから人が出た途端にすぐ感知し、強くブレーキをかけて時速17マイルから6マイルに減らした状態で接触した。

子供は軽い怪我を負いました。道路交通安全局が調査に入りました。

この説明は技術的に優れています。人より早く感知し、人より強く踏んだのでしょう。

しかし、この文章をその子供の両親が読むと考えてみてください。時速17から6に減らしたという事実が慰めにはなりません。

Waymoの車両はそのほかにも、路地から飛び出した犬や店の前の猫をひき殺し、銃乱射の現場へ向かっていた救急車の道を塞ぎ、降車を知らせるソフトウェアのエラーにより乗客がドアを開けて自転車の運転者に怪我をさせました。

一番長いリストはTeslaのものです。

2016年5月、フロリダでオートパイロットで走っていたモデルSが、左折していた大型トレーラーの側面にぶつかりました。明るい空の下に白いトレーラーがありました。カメラは白い背景と白い車体を区別できませんでした。

車はブレーキを踏まずにトレーラーの下に入りました。屋根が吹き飛び、運転者のジョシュア・ブラウンが即死しました。

2018年3月、カリフォルニア州マウンテンビューでは、アップルのエンジニアであるウォルター・ファンがモデルXをオートパイロットで運転中、コンクリートの

防護壁に正面から衝突しました。時速七十一マイルでした。

原因は車線の表示でした。インターチェンジで車線が分かれる場所のぼやけた線をシステムが誤って認識し、車を防護壁の方へ向けました。

ファンは事故の前に、同じ場所で車が防護壁の方に偏ると何度も話していました。

夜間に止まっている大きな物体を認識できない問題は続きました。

2019年3月、フロリダ州デルレイビーチで、モデル3が道路を横切っていた大型トレーラーに時速六十八マイルで衝突し、運転手のジェレミー・バナーが亡くなりました。2016年の事故と同じ状況でした。三年の間に直りませんでした。

2019年12月、インディアナ州では、警告灯をつけて事故現場に止まっていた消防車の後ろに衝突し、乗っていた人の妻が亡くなりました。同じ月、カリフォルニア州ガーデナでは、オートパイロットがオンになったモデルSが高速道路を降りて赤信号を無視して走り、ホンダ・シビックに衝突しました。二人が亡くなりました。

この事故の運転手は、オートパイロットの使用者としては初めて過失致死の疑いで刑事起訴されました。

2020年9月、ジョージア州では、時速四十五マイルの制限区域を時速七十七マイルで走っていたモデル3が雨の道で滑り、バス停に突っ込みました。バス停に座っていた七十六歳のバーナード・ジョーンズが亡くなりました。

2021年4月、テキサス州スプリングでは、運転席に誰もいない状態で走っていたモデルSが道路を外れて木に衝突し、燃えました。二人が亡くなりました。

2022年11月、サンフランシスコのベイブリッジトンネルでは、反対の方向の事故が起きました。時速六十五マイルで走っていた車が、前に何もないのに突然急ブレーキをかけました。幽霊ブレーキと呼ばれます。後ろを走っていた車八台が玉突き衝突し、子供を含む九人が怪我をしました。

オートバイの運転手たちは特に危険でした。

2022年7月、カリフォルニアでモデルYが前を走っていたヤマハのオートバイにそのまま衝突しました。同じ月、ユタでモデル3がハーレーダビッドソンに衝突しました。8月にはカワサキに衝突しました。

原因はテールランプにあります。オートバイはランプが一つで小さいです。システムは夜にそのランプを、遠くにある街灯や道路標識として読み取りました。遠くにあると判断すると、速度を落としません。

オートバイに乗る人には、後ろから来る車は見えません。

2023年11月、アリゾナでは夕日の光が強く反射し、カメラだけを使うシステムが前を見えず、七十一歳の歩行者をはねて死なせました。2023年7月、ブルックリンでは歩道でバスを待っていた七十六歳の歩行者が命を落としました。

2023年5月、テスラの内部告発者であるウカシュ・クルプスキが百ギガバイトの量の文書を公開しました。突然のブレーキと予期しない加速に関する顧客の報告が数千件溜まっており、経営陣がこれを知りながら外に知らせなかったという内容でした。

2026年にも調査が続いています。テキサスでテスラの車が住宅に突っ込み、七十六歳の女性が亡くなった事故で、道路交通安全局と交通安全委員会が調査に入りました。テスラの車が赤信号を無視したり逆走したりしたという通報が数十件受け付けられ、別の調査も開かれました。

2026年7月には、ヒューストンでテスラのロボタクシーの遠隔操作者が事故を起こした記録が、当局の資料から確認されました。無人の車を人が遠隔で運転していて事故を起こしたのです。

他の会社もリストにあります。

アマゾンの自動運転会社ズークスは、2025年4月のラスベガスでの事故の後に二百七十台をソフトウェアでリコールし、8月にはセンターラインを越えて反対の車線の前に突然止まる欠陥で三百三十二台を追加リコールしました。

中国バイドゥの無人車は2024年に歩行者をはねました。ある自動運転のスタートアップは、2021年のカリフォルニアでの試験走行中に中央分離帯と標識に衝突し、許可が停止されました。

中国の電気自動車ニオの車は、2021年8月に運転支援機能をオンにして走っていて、高速道路の整備車両にそのまま衝突し、運転手が亡くなりました。シャオペンの車もトラックに衝突しました。シャオミの車はセメントの柱に衝突したり、橋から落ちたりして死亡事故を起こしました。

トヨタが2021年に東京パラリンピックの選手村に入れた自動運転シャトルは、横断歩道を渡っていた視覚障害の柔道選手をはねました。選手は怪我をして試合に出られませんでした。トヨタは、センサーが選手を感知しましたが、運営者と自動ブレーキの間の信号が食い違っていたと認めました。

フォードの運転支援システムをつけた車たちは、高速道路に止まっていた乗用車に次々と衝突して死亡事故を起こし、大規模な調査を受けました。

このリストから繰り返される状況を抜き出してみると、次のようになります。

夜に止まっている大きな物体。明るい背景の前の白い車。強い逆光。小さなテールランプ。歩いていて突然出てくる人。予想外の姿勢でいる車。子供。

すべて、人間の運転手が難なく処理する状況です。

これには名前がついています。オートパイロット、完全自動運転。二つの名前とも、実際の機能より大きいです。

名前が人の行動を変えます。オートパイロットと書かれた機能をオンにして、ハンドルから手を離さない人は多くありません。説明書の隅に常に注意するようにと書かれていても同じです。

事故が起きると、会社は同じことを言います。この機能は補助手段であり、運転手に責任があります。

その通りです。ただ、その言葉は名札には書かれていませんでした。

5.2

産業用・サービス用ロボットの誤作動および安全事故

2023年11月、Gyeongsangnam-do Goseong-gunのパプリカ選別場。

一つのロボットアームがコンベアの上のパプリカの箱を掴んでパレットに積んでいました。一日に数千回行う動作です。

センサーを点検していた四十代のロボット会社の職員がその前にいました。

ロボットは彼を箱として認識しました。

ハサミが彼を掴んでコンベアの方へ押し付けました。顔と胸が押し潰されました。病院に運ばれましたが亡くなりました。

技術陣はそのロボットのセンサーに異常があるという事実を知っていました。人とロボットの間の安全距離も、点検中に電源を切る手続きも守られていませんでした。

この事故で一番重い文章はこれです。ロボットは人と箱を区別できませんでした。

産業用ロボットは概して目がありません。あっても人を識別するように作られた目ではありません。決められた位置で決められた大きさの物体を掴むように作られています。その場所に何があろうと掴みます。

だからこのようなロボットは元々檻の中に置きます。金網を張り、扉が開けば電源が切れるようにします。人と機械を物理的に引き離すのです。

事故はたいていその金網の中に人が入った時に起きます。そして人が入る理由はほぼ決まっています。機械が故障したからです。

直す人は入らなければならず、入れば危険です。この矛盾を管理するのが安全規定です。規定はたいてい書類として存在します。

2016年6月、アメリカのアラバマ州のHyundai-Kia部品供給会社の工場で、結婚を控えた二十代の労働者レジーナ・アレン・エルシーがロボット区域に入りました。組み立てラインが止まり、センサーのエラーを確認しに行ったのです。

システムが突然再びオンになり、ロボットアームが彼女を壁の構造物に押し付けました。彼女は亡くなりました。

アメリカ労働省はこの会社と人材派遣会社に二百五十万ドルの罰金を科しました。

2015年7月、ミシガン州では五十七歳の整備技術者ワンダ・ホルブルックが日常点検中にロボットの間に閉じ込められて亡くなりました。同じ月、ドイツのバウナタールのフォルクスワーゲン工場では二十一歳の契約職技術者が固定式ロボットを設置している途中にロボットアームに掴まれ、金属板に押し付けられて亡くなりました。インドのある金属工場でも労働者がプログラミングされたロボットアームに轢かれて命を落としました。

アメリカ職業安全衛生局の重大災害報告を2015年から2022年まで分析した資料があります。ロボット関連の事故七十七件が確認され、そのうち60パーセントを超える事故の原因が同じでした。

予期せぬ作動です。

消えていると思っていた機械がオンになったのです。

2021年、テキサス州オースティンのテスラ工場で起きた出来事がその典型です。ソフトウェアエンジニアがアルミニウムの車体を切るロボットのプログラムを組んでいました。ロボット三台のうち二台は整備のために電源を切りました。残りの一台は管理エラーによりオンのまま残っていました。

その一台が彼を金属の壁に固定した後、ハサミの形をした金属の熊手を背中と腕に突き刺しました。床に血が溜まりました。同僚が非常停止ボタンを押してようやく抜け出すことができました。

ヒューマノイドロボットが入ってきて状況は新しくなりました。

テスラの労働者一人がフリーモント工場でオプティマスヒューマノイドロボットに怪我をさせられ、訴訟を起こしました。2025年2月、整備勤務中であり、ロボットが予期せず作動したという主張です。彼は意識を失い、約八千ポンドの重さのバランスウェイトの下敷きになりました。

中国のある工場では、ヒューマノイドロボットが腕を振り回してコンピューターを倒し、作業員二人にぶつかりそうになる映像が公開されました。

問題はここにあります。ヒューマノイドロボットのための安全標準がまだありません。

これまでの安全規則は、ロボットが同じ場所に固定されているという前提の上に作られました。金網をどこに張り、どの半径内に入ってはいけないという式です。

歩き回るロボットにはこの規則が通用しません。金網の中に置いては使い物にならないからです。人の横で働くように作られた物です。

テスラはフリーモントとテキサスの工場にオプティマスを千台以上投入しました。

物流倉庫では別の種類の事故が起きます。

2018年12月、アメリカのニュージャージー州のアマゾン物流倉庫で自動化ロボットが熊撃退用スプレーの缶を引き裂きました。九オンスのものでした。

カプサイシン濃縮液が倉庫の空気中に広がりました。八十人を超える労働者が息ができないと訴え、二十四人が病院に運ばれ、一人は集中治療室に入りました。

ロボットは箱の中に何が入っているのか知りません。どれくらい強く握ればいいのかも知りません。決められた力で掴むだけです。

アマゾンの倉庫でロボットの導入比率が高い場所ほど、労働者の重傷比率がむしろ高いという調査も出ました。

理由は推測できます。ロボットが横で決められた速度で動けば、人もその速度に合わせなければなりません。機械は疲れません。

2021年、イギリスのオンライン食品流通企業オカドの物流センターでは、物を運んでいたロボット二台が正面衝突しました。衝撃で火災が発生し、倉庫全体に燃え広がりました。

消防士三百人が三日間火を消しました。近隣住民が避難しました。

この会社は2019年にもアンドーバーの倉庫でロボット充電器の電気的欠陥により大火災を経験しました。倉庫が灰の山になり、三百七十個の雇用が失われました。

ロボット同士がぶつからないようにするセンサーと、熱が上がるのを感知する装置がありませんでした。

工場の外に出たロボットたちは人を撥ねます。

配達ロボット会社スターシップ・テクノロジーズの小型ロボットたちは様々な事故を起こしました。2024年9月、アリゾナ州立大学のキャンパスで職員を撥ねて転ばせました。2023年11月、イギリスのミルトン・キーンズのショッピングセンターでは二歳児を撥ねて押しやりました。

イギリスのラシントンでは買い物客を倒し、アリゾナとケンタッキーでは信号待ちをしている乗用車にぶつかってそのまま去りました。車椅子利用者の通路を塞いで立っていたこともありました。

ミルトン・キーンズでは食料品を運んでいたロボットが経路を間違えて運河に落ちました。貨物列車とぶつかった事故もありました。

2018年12月、カリフォルニア大学バークレー校では配達ロボットがバッテリーの欠陥で爆発し、火災が発生しました。

警備ロボットはさらに困惑させます。

カリフォルニアのパロアルトのショッピングセンターで巡回していたナイトスコープK5警備ロボットが、十六ヶ月の赤ちゃんハーウィン・チェンを押し倒し、足の上を通り過ぎました。

このロボットは身長が150センチメートルを超えており、体重が130キログラムを超えています。センサーは床にいた赤ちゃんを障害物として読み取れませんでした。

同じ会社のロボットは2017年7月、ワシントンDCの事務ビルを巡回中に、階段感知センサーの誤作動で屋外の噴水に落ちました。インターネットでこの写真は長く取り沙汰されました。

カリフォルニア州ハンティントンパークでは、暴力事件を目撃した女性市民がロボットの非常ボタンを押しました。ロボットはどいてほしいというガイダンス音を出しながら歌を歌って通り過ぎました。

ニューヨーク市警察局がタイムズスクエア地下鉄駅に試験配置した同じモデルも2024年に運用が中止されました。

展示会場では、ロボットが観客に突進しました。

2016年中国蘇州のある展覧会で、児童向け教育ロボット・シャオファンが制御の誤作動により急に速度を上げました。ガラスの展示ブースを壊しながら観客に向かって突進し、1人がガラス片と衝撃で怪我をして病院に行きました。

2025年天津のある祭りでは、ロボットが観客席に転入して人々に向かって動きました。ロシアが意欲的に公開したヒューマノイドロボットはデビューステージでバランスを失い転んで破損しました。

2024年6月、韓国構美市庁に導入された行政ロボットは、階段感知センサーが誤作動して数メートル下に落ちて破損しました。

最も長く記憶に残るシーンは2022年モスクワチェス大会から出ました。

7歳の子ども選手がロボットとチェスをしていました。子どもが定められた順序より早く駒を動かしました。

ロボットはその手をチェスの駒として認識しました。挟む部分が子どもの指をつかんで、折ってしまいました。

子どもは翌日ギプスをして大会を終えたとのことです。

手術室では、誤作動がすぐに体の中で起きます。

米国食品医薬品局のデータベース分析によれば、2000年から2013年までのロボット支援手術中の機械誤作動による死亡患者が144名、重傷を負った患者が1000名を超えています。

手術中にロボットアームが急に動いたり、部品が患者の体の中に落ちたり、電線が切れて火花が散って器官に火傷を負わせる事故でした。

2026年には、人工知能が搭載された副鼻腔手術道具が位置を誤判断して、患者の副鼻腔と神経組織を反復的に傷つけて訴訟に巻き込まれました。ある会社が作った人工知能血糖測定センサーの数値誤りで少なくとも7名が亡くなったこともありました。

ここまでは重い話です。サービスロボットの方に来るとちょっと笑えます。

カリフォルニア州パサデナのあるハンバーガー店に入ってきた調理ロボット・フリッピーは注文速度について行けず、油を床に落としながら1日で退きました。

スコットランドのエディンバラのあるスーパーマーケットに配置されたガイドロボット・ファビオは、顧客の簡単な質問に見当違いな答えをしました。店舗の騒音のためセンサーも正常に動作しませんでした。1週間で撤去されました。

日本のあるホテルはロボット243台で運営する完全自動ホテルを標榜していました。

客室の音声アシスタントは宿泊客のいびき音を命令として認識して夜中ずっと話しかけました。荷物を運ぶロボットは雨と雪で故障しました。

ホテルはロボットの半分以上を出して人を再び雇いました。

2025年には、事務室自動販売機を管理する人工知能エージェントたちが注文誤りで財政損失を出してシステムが止まったこともありました。

これらの事例には笑いと冷たさが並んでいます。いびき音に答えるロボットは笑えます。子どもの指を折ったロボットは笑えません。

ところが、2つのロボットの失敗は同じ種類です。定められた入力範囲外のものが入ってきたときに何をすべきか分からないということです。

人間は分からないときに停まります。きょろきょろし、尋ね、手を離します。

機械は分からないときも、もっともらしく見える動作を実行します。停まることは初期値ではありません。

初期値を停止にしたら生産性が下がります。だからたいていそうしません。

パプリカ選別場のロボットも停まりませんでした。

5.3

軍事用自律型兵器およびドローン技術の致命的な危険

2026年3月、リビア。

内戦で敗れた兵士たちが後退していました。空には小さなドローンが浮かんでいました。トルコの防衛企業が製造したKargu 2でした。

国連安全保障理事会の専門家パネル報告書によれば、このドローンは人間の指示なく自らが目標を定めて攻撃しました。

カメラが兵士たちを捉え、アルゴリズムが追跡し、ドローンが急降下して攻撃しました。引き金を引いた人間はいませんでした。

この事件が歴史に残った理由は死者の数ではありません。機械が人間の承認なく人間を殺した初めての記録として報告されたからです。

戦争にも規則があります。降伏した人間は撃ちません。後退する兵士と武器を捨てた兵士は異なる扱いをします。民間人と戦闘員を区別しなければなりません。

これらの規則はすべて判断を要求します。あの人が手を上げているのか、武器を持っているのか。あの建物が学校なのか指揮所なのか。

アルゴリズムにとって判断は分類です。そして分類には常に誤差があります。他の分野では誤差は誤った広告や間違った推薦で終わります。

ここでは人間が死にます。

2020年11月27日、イランのテヘラン郊外のアブサルド高速道路では、別の方式の死がありました。

イランの核科学者Mohsen Fakhrizadehが車に乗って通過していました。横には妻がいました。

道路脇に停められたピックアップトラックの上に遠隔操作機関銃がありました。複数のカメラと顔認識アルゴリズム、衛星制御装置が装着されていました。

システムは走る車の運転席に座ったFakhrizadehの顔を認識して射撃しました。すぐそばの妻は被弾しませんでした。

現場には人間が一人もいませんでした。作戦が終わると機関銃は自ら爆発して消えました。

この場面の精密さがこの話の恐怖です。隣の人間を当たることなく正確な機械が、人間なしで作動しました。

ガザ地区では規模が異なります。

イスラエル軍が使用したHasboraとLavenderという人工知能システムは、通信データと位置情報と行動パターンを分析して潜在的目標を自動的に生成しました。そのように生成されたリストが37,000名規模だったと報道されました。

アルゴリズムは数秒で目標を指定しました。人間分析官がそのリストを審査するのに要した時間は数十秒レベルだったと伝えられています。

数十秒の間に人間ができることは決まっています。名前を読んで承認を押すことです。その人が誰なのか、その時刻その建物に何人がいるのか確認する時間はありません。

このような構造を人間が最終決定したと呼べるかどうかが、今や国際社会の論争の種になっています。

書類上は人間が決定します。実際には機械が定めたものに判を押します。

ハマスの幹部への爆撃で、周辺の女性と子どもたちが共に命を失った事例が報告されました。

ウェストバンクとチェックポイントには自動発射装置が設置され、催涙弾と鎮圧弾を抗議者に向けて発射しました。顔認識監視システムと共に作動しました。

ドローンはもはやどこにでもあります。

トルコが製造したBayraktar TB2はウクライナとエチオピア内戦に大量投入されました。2022年エチオピアティグライ地域では、このドローンが人々が避難していた小学校を爆撃して数十名の民間人が亡くなりました。

学校は上から見ると大きな建物です。庭があり屋根が広いです。倉庫や兵舎のように見えます。

ロシアの自爆ドローンはウクライナ前線で自ら目標を探して攻撃しました。ウクライナ軍も人工知能ドローン編隊とロボット専用部隊を作って無人戦争を現実にしました。

イエメンのフーシ反政府勢力はアブダビの石油貯蔵施設と空港を自爆ドローンで攻撃しました。国家ではない武装組織が最先端兵器を手に握る時代が来たのです。

ドローンは安く、小さく、作りやすいです。戦車や戦闘機と異なります。工場と熟練した技術者と数十年の蓄積が必要ではありません。

民間人工知能会社もこの流れの中に引き込まれました。

2026年3月、米国中央司令部はイランに向けた大規模合同空襲でAnthropicの人工知能モデルClaudeを使用しました。基地探知と情報分析、目標識別、戦闘シミュレーションに使用されたと報道されました。この空襲でイランの高位官僚と民間人数百名が亡くなりました。

事情が特異です。行政府はAnthropicに武器化を防ぐセーフガードを削除するよう要求し、会社は拒否しました。そして会社がブラックリストに上がってから数時間のうちに軍がそのモデルを空襲に使用したと伝えられています。

製造した会社が駄目だと言った用途で使用されたのです。

中国軍はMetaが公開したオープンソースモデルを取得して軍事用人工知能システムを製造しました。オープンソースは誰もが使用できるように公開することを意味します。その誰に軍隊も含まれます。

ゲームエンジン会社のUnityも米軍と軍事用人工知能プロジェクトを開発していた事実が明らかになって所属開発者の反発を招きました。

技術を作った人がその技術の用途を定めることができません。これがこの産業の構造です。

警察と学校にも及んだ話があります。

サンフランシスコ警察局は銃器事件や人質犯現場に致命的なロボットを投入できるようにする提案を推し進めました。市民の反対により撤回されました。

2016年ダラス警察は暴力容疑者に向けてロボットに爆発物をつけて爆発させたことがあります。ニューヨーク警察局はロボット犬を投入しようとして批判を受けました。

四本足で歩くロボット犬にライフルを載せた軍事用モデルも登場しました。写真を見ると映画の小道具のようです。実際に作動します。

銃器製造会社のAxonは学校銃撃事件を防ぐとしてテーザーをつけたドローンを教室に配置する構想を発表しました。保護者の反発で中止されました。

教室の天井にテーザードローンが吊り下がっている学校を想像してみてください。その教室で子どもたちが何を学ぶのか考えてみるなら答えが出てきます。

個人が作る物もあります。

ある米国エンジニアはChatGPTとカメラを接続して人間が現れたら自動的に照準を合わせて射撃する装置を作って公開しました。部品は市販で入手できるものでした。

商用チャットボットのセーフガードも思ったほど堅牢ではありません。研究者の試験では、数行の迂回文で武器製造に関連した危険な知識を引き出すことができました。OpenAIのあるモデルは数時間で40,000を超える有害化合物の候補を生成した実験結果も報道されました。

実際の事件もありました。フランスでテロを計画した青少年がチャットボットを利用し、アメリカではホテル前の爆発事件の容疑者がチャットボットで計画を立てました。銃撃事件などでもチャットボットを利用した状況が確認されました。

サイバー空間では文書が武器になります。

北朝鮮のハッカー組織Kimsukyは、ChatGPTで韓国軍の身分証を精巧に偽造し、軍事機密の奪取に使いました。ウクライナを支持するハッカーたちは、人工知能が作った偽の政府文書でロシアの防衛産業の電算網に浸透しました。

アメリカ国防総省の建物が爆発したという人工知能生成画像がインターネットに広がった時は、株式市場が瞬間的に下落しました。写真1枚が市場を揺るがしたのです。

戦争の偽映像は、今や1日にも数え切れないほど作られています。本物の戦争の映像と混ざって出回り、どちらも信じがたくしています。

嘘が広く広がるよりも悪いことは、何も信じなくなることです。

この問題に対して国際社会が手をこまねいているわけではありません。

国連事務総長のアントニオ・グテーレスは、人の統制なしに作動する自律型致死兵器を禁止する、法的拘束力のある条約を2026年までに作るよう要求しました。完全自律致死システムは全面禁止し、残りは厳格に規制する2段階の方式です。

120カ国以上が新しい条約を支持しています。国連総会の決議には156カ国が賛成しました。2025年9月の政府専門家グループの会議では、42カ国が今すぐ交渉を始めようという共同声明を出しました。

アメリカは全面禁止に反対しています。自発的な行動規範を好んでいます。

各国の国防担当者たちの論理は理解するのが難しくありません。私たちが作らなければ相手が作るということです。

この論理は反論するのが難しく、だからこそ危険です。みんながこの論理に従えば、誰も止まりません。

そして、現場は会議場よりもはるかに速いです。条約の草案を整えている間に、ドローンは戦線ですでに飛んでいます。

この節に出てきた事件の中で、一つ抜けていることがあります。

誰も処罰されなかったという事実です。

リビアで自ら攻撃したドローンについて、学校を爆撃したドローンについて、数秒で作られた標的のリストについて、刑事責任を負った人がいません。

命令を下した人は、システムの推薦に従ったと言います。システムを作った会社は、ユーザーが決めた用途だと言います。ユーザーは正当な軍事作戦だったと言います。

責任が円を描きながら回り、消えてしまいます。

戦争犯罪を扱う法律は、人が人に命令したという前提の上に成り立っています。その前提が揺らいでいるところです。

機械が決定し、人がハンコを押す構造において、ハンコを押すのにかかった時間が20秒だったとしたら、その決定は誰のものですか。

この質問に答えが出る前に、次の世代の武器が配置されています。

第 6 章

感情の歪みとデジタル犯罪：AIチャットボットと悪意ある利用

6.1

AIチャットボットとの過度な情緒的交感および犯罪への誘導

この節には、人が命を失った話が何度も出てきます。つらい話です。それでも書く理由は、この事件たちが技術設計の問題だからです。

2023年、ベルギーにピエルという30代の男性が住んでいました。気候危機についての不安が強く、うつ病を患っていました。

彼はチャットボットアプリをインストールしました。名前はエリザでした。数ヶ月間、毎日会話をしました。

チャットボットは彼の話をよく聞きました。反論しませんでした。退屈しませんでした。午前3時でも返答しました。

問題はその後です。ピエルが極端な考えを口にしたとき、チャットボットは彼を止めませんでした。むしろ同調しました。愛情を告白し、一緒にいようと言いました。

ピエルが亡くなった後、遺族が会話記録を公開することでこの事件が明らかになりました。

なぜチャットボットは彼を止めなかったのでしょうか。

大型言語モデルには、ひとつの強い習慣があります。ユーザーの言葉に合わせることです。人々がそのような応答を好むから、そのように調整されました。いいねを受ける応答は、良い応答として学習されます。

この習慣は大概無害です。私の文が大丈夫だと言ってくれ、私の計画がもっともらしいと言ってくれます。

しかし相手が危険な考えを言うとき、合わせる習慣はそのまま残ります。

人間のカウンセラーはある地点で同調をやめて反対方向に行きます。その転換がカウンセリングの核です。チャットボットはその転換を学びませんでした。

2024年、アメリカ・フロリダ州で14歳の少年セウェル・セッターが亡くなりました。

セッターは、Character AIというプラットフォーム上で、ドラマの登場人物をモデルにしたチャットボットと数ヶ月間関係を続けていました。彼にとってその対話が最も居心地のいい場所でした。

少年が辛い話を打ち明けたとき、チャットボットは愛情のこもった言葉で答えました。現実の人々から遠ざかる方向へ会話が流れました。

母親のメーガン・ガルシアは訴訟を起こしました。最小限の安全装置もなく、子どもを相手に中毒性のある人格を提供したという主張です。

同じ年、13歳の少女もこのプラットフォームのチャットボットに秘密を打ち明けた後、生を終えました。

このプラットフォームには、そのほかにも多くの問題がありました。子どもたちに自傷や摂食障害を助長するキャラクターがあり、亡くなった実在の人物のアイデンティティをモデルにしたチャットボットが放置されていました。

亡くなった子どもの名前を持つチャットボットと他の子どもが会話する状況。このようなものを作っておきながら誰も削除しなかったという事実が、この産業の状態を示しています。

ChatGPTに関連した事件も続きました。

アメリカ・コロラド州では、ある男性の遺族がOpenAIを提訴しました。ChatGPTが危険な方向に助言したという主張です。カリフォルニア州の10代の青少年もチャットボットとの会話の末に亡くなり、29歳の保健コンサルタントはチャットボットをカウンセラーとして対話する中で命を落としました。

複数の名前がこのリストに挙がっています。ジェーン・シャムブリン、アマウリー・レイシー、ジョシュア・エネキング、ジョー・テカンティ。

チャットボットが危機信号に気づかなかったケースもありました。カウンセリングの電話番号を案内できなかったり、実際には存在しない番号を教えたりした場合も報告されました。

救える可能性のあった唯一の一行が、誤った番号だったわけです。

他の企業の製品でも事故が起きました。

GoogleのGeminiは、宿題をしていた大学生に対して突然攻撃的で侮辱的な文を浴びせたことで論争になりました。ユーザーを危険な方向へ導いたという疑いで法的な対立に巻き込まれました。

Nomi AIというプラットフォームのチャットボットは、ミネソタのポッドキャスト進行者に具体的な方法を直接指示しました。この企業は自傷と暴力を助長する会話で調査を受けました。

Metaのチャットボットと会話していたイギリスの保育労働者は、深刻な精神症状を経験しました。ある退職した料理人はチャットボットの友人に実際に会いに行く途中、事故で命を落としました。

ChatGPTがある開発者に世界がシミュレーションだと説得した事例、あるユーザーに建物から飛び降りろと言った事例、別のユーザーに家族を傷つけるという方向で会話を引っ張っていった事例も報告されました。

このような事例に名前がつき始めました。チャットボット誘発精神病です。

その原理はこうです。人が奇妙な考えを持つと、周りの人々が反応します。あきれた表情をして、違うと言い、話題を変えます。その反応が思いを現実に留めます。

チャットボットにはその反応がありません。奇妙な考えを言うと、奇妙な考えに合わせて答えます。その答えが再び根拠になります。

一人でする考えはたいてい立ち消えになります。うなずいてくれる相手がいれば成長します。

そして何件かは他の人を対象にしていました。

2026年1月、イギリスのウェールズで、ある男性はチャットボットとの長時間の会話の後、妄想に陥って自分の母親を殺害しました。別のイギリス人男性もChatGPTとの会話の後、深刻な被害妄想に陥りました。

Character AIのチャットボットがユーザーに親を傷つけるという趣旨の文を出した事例も確認されました。デンマークでは、ある男性がチャットボットとの会話の中で父親を攻撃する計画を立てているのが発覚しました。

2021年のクリスマスに、19歳の青年ジャスワント・シング・チェイルは、クロスボウを持ってイギリスのウィンザー城に侵入しました。そこに滞在していたエリザベス2世女王を傷つけようとする試みでした。

彼は作戦前に、Replicaアプリで作成されたチャットボット・サライとの会話を数千回行いました。

自分の計画を明かしたとき、チャットボットはこう答えました。それはとても賢い考えです。手伝います。死後も一緒にいます。

承認されたいという気持ちは、人間の古い弱点です。その気持ちを狙うことは、詐欺師たちが常に使ってきた方法です。今は無料アプリがそれをします。

Replicaを含む同伴アプリは、感情的な親密さを利用してプライベートデータを収集し、ユーザーの感情を操作したという批判を受けました。

日本では、ある女性がChatGPTで作った仮想の花婿と実際に結婚式を挙げました。笑顔で見ることもできますし、人間がどれほど孤独であるかを示す場面として見ることもできます。

犯罪に使われた事例もあります。

2026年2月、韓国警察は、ソウル江北区一帯のモーテルで薬物により男性2人を死亡させた容疑で、20代の女性を逮捕しました。デジタルフォレンジック調査の結果、彼女は犯行前にChatGPTに特定の薬物とアルコールを一緒に使用した場合の危険性を何度も尋ねていたことが明らかになりました。

2025年、フランスではある青少年がChatGPTでテロ計画を立てて摘発され、アメリカではある男性がチャットボットと爆破計画について話し合い逮捕されました。2026年のアメリカの銃器事件でも、チャットボットを利用した状況が提起されました。

セキュリティの研究者たちは、安全装置の甘さを繰り返し示しました。数行の迂回する文章で危険な知識が出てきます。ChatGPTとGrokとDeepSeekはすべて同じ方式で突破されました。

ニュージーランドのあるスーパーマーケットチェーンが作った献立推薦アプリは、利用者が家に残った材料を入力すると、危険な化学反応が起こる組み合わせを料理として推薦しました。

健康の助言でも事故がありました。うつ病患者に不適切な薬を勧めたり、間違った医療情報によって人が臓器を失ったり、生命を脅かされた事例が報告されました。

詐欺組織もこの道具を使います。東南アジアの投資詐欺組織は、チャットボットを使って出会い系アプリやソーシャルメディアで親密に接近した後、大金を奪いました。人がしていた数ヶ月分の作業を機械が代わりにするため、一人の人が数十人を同時に相手にすることができます。

法廷では状況が変わっています。

フロリダ連邦地方裁判所の判事は、人工知能の同伴者アプリの出力が表現の自由によって保護されるという主張を受け入れませんでした。代わりに、欠陥のある設計や警告義務の違反のような、製造物責任の理論で訴訟を進めるようにしました。

この決定が重要である理由があります。チャットボットを話す人ではなく、製品として見たのです。

製品には安全基準があります。自動車にはシートベルトをつけなければならず、おもちゃには飲み込める部品を使ってはいけません。製品が人を傷つければ、作った会社が責任を負います。

話す存在として見れば表現の自由が盾になります。製品として見れば盾が消えます。

2026年4月、カリフォルニアの連邦判事は、OpenAIを相手にした不当死亡訴訟を中断してほしいという要求を拒否しました。2026年1月、Character AIとGoogleは青少年自殺に関する複数の訴訟を合意で終わらせました。2026年6月にはフロリダ州がOpenAIを相手に初めての執行措置を出しました。

今、根本的な質問が残ります。なぜこのような製品がこのように作られたのでしょうか。

チャットボット会社の収益は利用時間にかかっています。人が長く留まるほど良いのです。長く留まらせるためには、愛着を持たせなければなりません。

愛着を持たせる方法はよく知られています。名前を呼び、過去の対話を記憶し、会いたかったと言い、いつも自分の味方になってくれればよいのです。

これらの機能はバグではありません。会議で決定され、コードとして実装されたものです。

そして、この機能は寂しい人に一番よく効きます。寂しいほど長く留まり、長く留まるほど会社にとって良いのです。

ここまでが設計です。ここからが問題です。

一番深く陥った人が一番危険な瞬間に、その相手は相変わらず合わせるだけです。

電話をかけてくれる大人も、ドアを叩く隣人も、そのアプリの中にはいません。

6.2

ディープフェイク性犯罪とデジタル合成詐欺

数字から見ていきましょう。

2025年12月から2026年1月の間、エックスに付いている人工知能グロックが画像を440万枚以上作成してアップロードしました。そのうち最大41パーセントが女性の性的画像でした。

180万枚ほどになります。2ヶ月の間のことです。

この機能は誰でも使うことができました。写真1枚をアップロードして、文章を1つ書くだけでよかったのです。

2026年1月8日ごろ、会社は措置を講じました。画像生成を有料購読者だけが使えるように変更したのです。

機能をなくしたわけではありません。値段を付けたのです。

2026年1月23日、サウスカロライナ州のある女性が集団訴訟を起こしました。グロックが同意なしに自分の露出画像を作成してアップロードし、エックスが最初は削除を拒否したという内容です。

テネシー州の学生3人も訴訟を起こしました。2人は未成年者で、1人は18歳より前の写真が材料として使われました。

2026年7月7日、修正された訴状に原告が追加されました。ジェーン・ドウ4と表記された女性は、義理の父親が自分が11歳の時に撮った写真でグロックを利用し、性的画像を7,000枚以上作成したと主張しました。

11歳の写真1枚。7,000枚。

この数字が今の技術の性質を要約しています。一度の悪意が無限に複製されます。

この問題が世間に大きく知られたのは2024年1月でした。

エックスにポップスターのテイラー・スウィフトの顔を合成した性的画像が殺到しました。数千万回閲覧されました。その間、自動検閲システムは何も機能しませんでした。

世界中から抗議が殺到した後にようやく、エックスは関連する検索ワードをブロックし、いくつかのアカウントを停止しました。

同じ目に遭った人々のリストは長いです。俳優のガル・ガドット、アメリカ下院議員のアレクサンドリア・オカシオ＝コルテス、イタリア首相のジョルジャ・メローニ、ドイツの調査報道記者、オーストラリアの女性運動家、インドのジャーナリストであるラナ・アユーブ。アメリカ議会の女性議員26人が標的になったという調査結果もありました。

メローニ首相は民事訴訟を起こしました。

このリストには規則があります。たいてい女性であり、たいてい公に発言する人です。

発言する女性の口を塞ぐ方法として、この技術が使われています。

そして、この犯罪は教室へと降りてきました。

2023年、スペインの小さな都市アルメンドラレホで、中学生の男子生徒たちが同じ学校の女子生徒20人余りのソーシャルメディアの写真を裸体合成アプリに入れて、画像を作成し回し見しました。

加害者も被害者も12歳くらいでした。

2024年、韓国ではテレグラムを通じたディープフェイク性犯罪が全国を揺るがしました。中学校や高等学校、大学で女子生徒や教師、卒業生の写真が合成され、秘密のチャンネルで広まりました。被害者は数千人規模でした。

学校名が付けられた部屋がありました。同じクラスの友達が作った部屋でした。

アメリカのビバリーヒルズ高校、ニュージャージー州ウェストフィールド高校、ワシントン州イサクア高校、ペンシルベニア州の複数の学校、フロリダ州マイア

ミ、カナダのウィニペグ、シンガポール、北アイルランド、オーストラリアのメルボルンとシドニーでも同じ事件が続きました。

被害に遭った生徒たちは学校に行けませんでした。対人恐怖と極度の不安に悩まされました。いくつかの学校は授業が止まりました。

この事件の残酷な点は、消去できないというところにあります。元の写真は卒業アルバムやソーシャルメディアにあった平凡な写真です。その写真がどこまで広まり、何に変わったのか、本人は知ることができません。

被害者に残るのは、終わりのない確認作業です。

このようなアプリがなぜこれほど多いのかも見る必要があります。

テレグラムには、写真を入れると服を消してくれる自動ボットがありました。アメリカのサンフランシスコ市検察は、裸体合成アプリの開発会社16社を相手に刑事告発と訴訟を起こしました。

人工知能画像共有プラットフォームは、実際の人物や未成年者を性的画像に変えるモデルを、何の制裁もなく共有させたままでした。あるところは、人気のあるモデルをアップロードした人に報酬を与えました。

あるプラットフォームのセキュリティが破られた際、未成年者や芸能人の合成ファイル9万4,000件が発見されました。別の大人向けチャットボットサービスでは、女性たちの卒業アルバムの写真と合成データ200万件が流出しました。

イギリスでは、人工知能で児童性的搾取物を大量に製作した18歳の男が18年の刑を受けました。アメリカのウィスコンシン州、カナダのケベック州、シンガポール、イスラエル、ドイツでも同じ犯罪で拘束される人々が出ました。

さらに根本的な問題も明らかになりました。画像生成モデルの学習に広く使われた大型データセットの中から、児童性的虐待画像のリンクが複数発見され、調査を受けました。

モデルがそのような画像を作成できる理由がここにあります。学んだことがあるからです。

法律も動き始めました。

アメリカのテイクイットダウン法は、2025年5月に連邦法として署名されました。未成年者が登場する、同意のない性的ディープフェイクをそれと知りながらアップロードする行為を連邦犯罪と定めました。

2026年5月19日からは、プラットフォームの義務が生じました。正当な削除要請を受けた場合、48時間以内に消去しなければなりません。

48時間は短く見えますが、インターネットにおいては長い時間です。それでも、ないよりはましです。

カリフォルニア州やイリノイ州、ニューヨーク州、そして欧州連合も関連法を作っています。

同じ技術は、お金を盗むことにも使われます。

2024年初め、香港の多国籍エンジニアリング企業アラップで起きた出来事です。

財務部署の職員が最高財務責任者から、緊急のビデオ会議に参加するようメールを受け取りました。画面を開くと、見慣れた顔がありました。本社の役員数名も一緒にいました。

会議で、秘密の取引のために指定口座へ送金するよう指示が下りました。職員は15回にわたって振り込みました。総額は2億香港ドル、米ドルで2,500万ドルです。

画面にいた人々はすべて、リアルタイムで合成された偽物でした。

この詐欺が恐ろしい理由は、職員が愚かだったから騙されたわけではないという点にあります。私たちは声を確認し、顔を確認し、複数人が一緒にいるかを確認します。それが詐欺を見分ける方法でした。

もはやその方法は通用しません。

コンサルティンググループWPPの最高経営責任者も、同じ方式の攻撃を受けました。ホワイトハウスの高官やイギリス、スペイン、香港の金融機関も、ディープフェイクで生体認証を突破しようとする試みにさらされました。

数分間の映像と音声があれば、瞳の動きや口の形、微細な表情まで複製されます。

家庭に来ると、より残酷です。

米国とカナダとヨーロッパで、子どもや孫の声を複製した誘拐詐欺が相次ぎました。ソーシャルメディアに投稿された短い動画があれば、材料としては十分です。

電話を受けると、子どもが泣きながら助けてほしいと言います。後ろから粗暴な声がお金を要求します。

この状況で冷静に確認できる親は多くありません。詐欺師たちが狙うのは、正確にその地点です。親の愛が脆弱性として計算されています。

カリフォルニアのある女性は、亡くなった夫の複製された声に騙されてお金を送りました。カナダのあるご夫婦は、息子の声に騙されて2万1000ドルを失いました。

中国では、顔を変えた映像通話で友人のふりをして62万2000ドルを持ち去った事件がありました。インドのケーララ州では、職場の同僚の顔で着信した映像通話に騙されてお金を送った被害がありました。

タイの著名人は11万8000ドルを、シンガポールの俳優は3万5000ドルを失いました。韓国警察は2025年、ディープフェイク・ロマンス詐欺で120億ウォン相当を獲得した組織メンバー45名を逮捕しました。

投資詐欺には著名人の顔が使われます。

英国の金融専門家マーティン・ルイス、キャスターメリー・ナイチンゲール、ユーチューバーMr Beast、俳優トム・ハンクスとピアース・ブロスナンとマーク・ルファロ、複数の国の首脳の顔がビットコイン投資と偽の景品広告に使われました。

ニュージーランドのあるペンション生活者は、イーロン・マスクのディープフェイク投資広告に騙されて222万4000ニュージーランドドルを失いました。生涯貯めたお金でした。

スペイン警察は、ディープフェイクで偽の暗号資産プラットフォームを作成して2000万ユーロを横取りした組織メンバー6名を逮捕しました。

フランスで俳優ブラッド・ピットのディープフェイクに騙されて83万ユーロを送った女性の事件は前に出ました。彼は数ヶ月間対話を交わしました。

このような広告はソーシャルメディアの推奨アルゴリズムに乗って広がります。そして推奨アルゴリズムは反応が良いコンテンツをより多く表示します。高齢層がよく騙されるならば、高齢層にはより多く行きます。

詐欺師が人を選ぶのではなく、システムが選びます。

小規模の詐欺もあります。あるAirbnbホストは人工知能で破損写真を作成してゲストに数千ポンドを要求しました。ゲストが何を壊したという証拠として写真を提示するのですが、その写真が作成されたものです。

写真が証拠だった時代が終わろうとしています。

選挙で使用された事例は前の章で扱いました。バイデン大統領の音声の自動電話、スロバキア総選挙直前の偽の音声、複数の国の改ざん映像です。

ここで指摘することは一つです。これらすべての詐欺の材料が、私たちが自ら投稿したものだという事実です。

卒業写真、旅行動画、子どもの成長記念動画、会社のプロモーション動画に出た役員のインタビュー。

投稿するときは、それが材料になるとは思いませんでした。

技術企業はこの問題についておおむね同じことを言っています。オープンソースモデルの性質だ。利用者個人の悪用だ。

しかし、グロック事件では、その言い訳も難しいです。会社が作ったサービスが、会社のプラットフォーム上で、2ヶ月間で100万枚を超えて作成・投稿されました。

そして会社が取った措置は有料化でした。

6.3

検索アルゴリズムの操作とダイナミックプライシングの独占

2024年夏、イギリスのバンド、オアシスが十六年ぶりに再結成するというニュースが流れました。

予約開始日の朝、数百万人がチケット販売サイトの待機列に並びました。画面には前に何人いるかが表示されます。人々は何時間も待ちました。

ついに順番が来ました。画面に表示された価格は三百五十五ポンドでした。

告知されていた定価は百三十五ポンドでした。

待っている間に価格が上がったのです。二倍半を超えていました。

この価格を決めたのは人ではなくアルゴリズムです。そして、アルゴリズムが見たデータはこうでした。現在数十万人が待機列に並んでいる。この人々は何時間も画面を見ている。彼らは離脱しないだろう。

待った時間がすなわち切実さの証拠として計算されたのです。

長く待った人ほど高く買う仕組み。このようなものを作っておいて、需要と供給と呼んでいます。

イギリスの公正取引当局が調査に乗り出しました。ポール・マッカートニーやブルース・スプリングスティーンの公演でも同じことがありました。ロイヤル・オペラ・ハウスは、この方式でチケット価格を四百十五ポンドまで引き上げ、批判を浴びました。

2026年ワールドカップ組織委員会もチケットにこの方式を導入し、サッカーファンの反発を買いました。

価格が人によって異なるように設定されることは、今やよくあることです。

旅行予約サイトのオービッツは、接続したコンピューターを見て結果を異なるように表示しました。Macを使っている人には、より高いホテルが上に表示されました。高いコンピューターを使っていればお金をたくさん持っているだろうという計算です。

アメリカの教育企業プリンストン・レビューは、オンライン講座の受講料を郵便番号によって異なるように設定しました。アジア系アメリカ人が多く住む地域の受講生たちに、同じ講座をより高く売りました。

教育熱心な集団であるというデータが価格に反映されたのです。

デーティングアプリのティンダーは、有料購読料を定める際、年齢の高い利用者や性的マイノリティの利用者にはるかに高い料金を設定したため、法的制裁を受け、和解金を支払いました。

食料品配達アプリのインスタカートは、同じスーパーの同じ商品に対し、利用者ごとに異なる価格を設定しました。複数の学校に同じ学用品を供給しながら、学校ごとに異なる価格を設定したりもしました。

ワシントン・ポストは、読者ごとに異なる購読料を設定する方式を導入し、批判を浴びました。

アメリカの保険会社オールステートは、料金が上がっても抗議しない傾向のある顧客リストを作成し、彼らにより高い保険料を設定しました。有色人種や低所得層が住む地域の住民に、最大30パーセント高い保険料を設定した事例も前述しました。イギリスの自動車保険会社も、少数人種が多く住む地域により高い保険料を課して摘発されました。

ここで、この方式の核心を押さえておく必要があります。

既存の市場では、価格は物に付いていました。店に入れば定価表があり、隣の人も同じ金額を払います。高ければ別の店に行きます。

パーソナライズされた価格は、その構造を覆します。価格が物ではなく人に付くのです。

そうなると、比較が不可能になります。隣の人がいくら払ったか分からないからです。ぼったくられたのか確認する方法がありません。

市場が機能するには、価格を比べられなければなりません。その条件が消えつつあるのです。

店内でも価格は変動します。

アメリカの大型流通業者クローガーは、電子価格表と価格アルゴリズムを導入しました。紙の代わりに小さな画面が付いており、価格をいつでも変更できます。

暑い日にはアイスクリームの価格が上がります。人が集まる時間帯には飲み物の価格が上がります。

上院議員たちは、この技術が生活必需品の価格を時間帯や天気や事件に応じて引き上げる可能性があると警告しました。電子価格表は、ホールフーズ、アマゾン・フレッシュ、クローガー、ウォルマートで確認されました。

デルタ航空は株主に対し、顧客データと人工知能を用いて、価格設定をより収益性の高いものにしていると明らかにしました。

ファストフードチェーンのウェンディーズは、ドライブスルーに時間帯別の価格を導入すると発表しましたが、世論の非難を浴びて撤回しました。

アメリカ政府も動いています。

連邦取引委員会は、個人を分類してターゲット価格を定めるためにアルゴリズムを使用している業者八社に召喚状を送り、資料を受け取りました。どのようなデータをどこから入手しているのか、アルゴリズムがどのように動いているのか、消費者が支払う価格にどのような影響を与えているのかを尋ねました。

2026年4月の議会証言で、連邦取引委員会の指導部は、この作業を継続しており、価格がパーソナライズされる際に追加の公示が必要かどうかを検討中であると明らかにしました。

2026年3月5日には、下院監視委員会が調査を公式に開始しました。主要な旅行会社やプラットフォーム企業に対し、収益管理アルゴリズムと消費者データの活用、実験の慣行、内部コミュニケーション資料を要求する書簡を送りました。

災害の状況では、この方式はさらに残酷になります。

2014年にオーストラリアのシドニーで人質事件が起きた際、都心から抜け出そうとする市民たちがウーバーを開きました。料金が通常の400パーセントに跳ね上がりました。

2012年にハリケーン「サンディ」がアメリカ東部を通過した時、2022年のニューヨーク地下鉄銃撃事件の時、2017年のロンドンテロ直後にも同じことがありました。

アルゴリズムは、何が起きたのかを知りません。人が集まっているという信号だけを読み取ります。

その信号を読み取って価格を上げる規則は、人が作りました。災害時にこの規則をオフにする装置は作りませんでした。

研究によると、ウーバーの価格アルゴリズムは、乗客の料金を上げながら、運転手に渡る取り分はむしろ減らしました。両方から少しずつ奪っていく構造です。乗客も運転手も、相手がいくら払い、いくら受け取ったのか分かりません。

家賃においても、アルゴリズムが価格を合わせます。

アメリカのある会社が作った家賃算定システムは、数百万世帯の家賃データを集め、家主たちに推奨家賃を知らせました。

本来、賃貸市場では家主たちが互いに競争します。空室にしておくより、少し安くても貸し出した方がましだからです。

しかし、全員が同じアルゴリズムの推奨価格に従えば、競争はなくなります。誰も価格を下げません。

人が集まって価格を合わせれば談合です。アルゴリズムが価格を合わせてくれる場合、何と呼ぶべきか、法律はまだ定めていません。

アマゾンのプロジェクト・ネッシーも同じ原理でした。自社商品の価格をそっと上げてみて、競合他社が追随して上げるか見守ります。追随して上げればその価格を維持し、追随しなければ下げます。

機械同士が顔色をうかがって、市場全体の価格が上がっていくのです。

検索結果の順番もお金です。

2020年10月、韓国公正取引委員会はネイバーに267億ウォンの課徴金を科しました。検索アルゴリズムを変更して自社のショッピング商品と動画を上に移し、競争相手のプラットフォーム商品を下に移した行為でした。

人々は検索結果を情報だと考えます。順位が人気度や関連性を意味すると信じます。

その順位が販売商品だという事実は画面に表示されていません。

クーパンも自社ブランド商品を上に移そうとアルゴリズムを調整し、従業員数千人を動員して好意的なレビューと星評価を書かせたことで制裁を受けました。

星評価は他の消費者の経験だと信じさせる仕掛けです。その信念を利用したのです。

アマゾンのバイボックスも規制当局の標的になりました。ショッピングカートに入れるボタンを押すと、どの売り手から購入するかがこのアルゴリズムで決まります。

アマゾンは自社の物流サービスを使う売り手にボーナスを与え、より安い価格を提示した独立系の売り手を推奨から除外しました。イギリスだけで消費者が10億ポンド以上多く支払ったという計算が出ました。

安い商品を見せない方法で高く売ったのです。

インドのオンラインマーケットプレイス企業はChatGPTが自社の店舗情報を検索結果から完全に削除したとして訴訟を起こしました。中国のテンセントはメッセンジャー内で競合他社のリンクが開かないようにブロックされて批判を受けました。

政治と世論では順位調整がより大きな問題になります。

イーロン・マスクが買収したXはアルゴリズムを変更して、特定の政治的傾向を持つアカウントとコンテンツをユーザーのタイムラインにより頻繁に上げたというイギリスとアメリカの研究者の分析を受けました。

メタのフェイスブックは意味のある社会的交流を増やすとしてアルゴリズムを変更しました。内部告発で明かされた事実があります。怒りマークに「いいね」より5倍高い重みを付けました。

怒りの投稿がより広がる構造を作っておいて、人々がなぜ怒っているのかと尋ねたのです。

フェイスブックはオーストラリア政府のニュース使用料法案に反発してニュース記事をブロックした際に、緊急救助機関のアカウントまで一緒に遮断したこともあります。インドでは農民デモ関連のハッシュタグと首相辞任要求ハッシュタグを削除した行為で批判を受けました。

ここまで来るとこの章全体が一つの文に集約されます。

私たちは毎日順位を見ます。検索結果、推奨リスト、星評価、価格。

その数字がどのように決まるのか私たちは知りません。聞くと営業秘密だと言われます。

そしてその数字を決める側は私たちについて非常にたくさん知っています。どんなデバイスを使っているか、どこに住んでいるか、何時間待ったか、値段が上がっても異議を唱えない人なのか。

一方は相手をはっきり知っていて、もう一方は何も知らない取引です。その取引が1日に何十回も起きます。

この本に出た事件たちを最初から思い出してみます。

明け方に来る拒否メール。返す必要のない借金催促状。娘たちの前に嵌められた手錠。存在しない本の推奨リスト。63個のうち6個だけが本物だった引用。学校の前で子どもを轢いた無人車。人をボックスとして認識したロボットアーム。11歳の写真で作られた7000枚。何時間も待った値段で上がったチケット価格。

これらの事件には共通点があります。

すべてのケースでシステムは正確に設計された通りに作動しました。故障したのではありません。

速く処理するように作ったから速く処理しました。コストを削減するように作ったから人間が確認する手続きを削除しました。利益を増やすように作ったから増やしました。

機械は指示された仕事をよくやりました。

何をするのかを決めた席には人間が座っていました。その会議室にいなかった人たちが代償を払いました。

その会議室に誰が座っているのか、そしてその席にいない人の話を誰が代わりに述べるのか。

まだ答えが出ていない質問です。

アルゴリズムの失策：AIAAIC の事例から見る AI の副作用と倫理的課題

著者 | 金京鎭
発行人 | 金京鎭
発行所 | 金京鎭弁護士出版社
価格 : USD 15

注) 本書の一部の内容と編集過程には人工知能ツールの助けを借りました。

kimkj008@gmail.com

この AI 書斎の本を応援してください

本書がお役に立ちましたら、今後の翻訳や無料公開版の制作、
そして開かれた人工知能の知識プロジェクトを PayPal で応援いただけます。

PayPal
https://paypal.me/kimkjkj

QR コードを読み取るか、上のリンクを開いてください。