## They bring crime and disorder

*They come without passports and leave with your identity.*

Every immigration wave brings fears that the newcomers will erode public safety. In the nineteenth century it was Irish dock‑workers; in the twenty‑first it is families crossing an ocean for refuge. Today the figure of the dangerous outsider has taken a digital form. **Artificial intelligences** – the models that suggest songs, reply to emails and generate artwork – are portrayed as lawless migrants from a wild frontier. In coffee shops and news segments there is a familiar rhetoric: these “algos” will make us less secure, make crime easier and our institutions chaotic. That fear deserves examination because it contains grains of truth. The chapter that follows explains how AI tools can be weaponised to amplify crime and disorder, not because they are malicious themselves, but because human actors exploit their capabilities. The aim is not to provoke panic. Instead, it offers a clear‑eyed account and asks what citizens and governments can do.

### Fear of disorder and the immigrant metaphor

The immigrant metaphor that frames this book is not an accusation against machines; it is a lens on our society. When people worry that a group “brings crime,” they are often projecting deeper anxieties about economic insecurity and social change. The same is true of AI. Criminals and authoritarian regimes have already harnessed automation to commit fraud, suppress dissent and infiltrate systems. But AI itself is a tool, akin to a car or a printing press. Blaming the tool alone obscures the **human agency** that decides whether it is used to help or to harm. In this chapter we will distinguish between intrinsic risks (the ways an algorithm can behave unexpectedly) and misuse (deliberate exploitation by bad actors). The comparison to immigrants also highlights how generalised fear can lead to discriminatory responses. Rather than panic, we need policy and oversight. The next sections examine different kinds of harm and show that behind the headlines there are human decisions, corporate incentives and regulatory gaps.

### Human rights violations and crimes against humanity

One of the most chilling applications of AI is its use by governments to monitor, predict and control populations. Investigations by the European Parliament into the Xinjiang region of China reveal that authorities there use AI‑driven **predictive policing** and facial recognition to identify and detain members of the Uyghur minority[[1]](https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA(2024)754450_EN.pdf). Data from cameras, license plate readers and mobile devices are fused in systems like the “Sharp Eyes” programme to identify “focus personnel” and flag behaviours deemed risky[[1]](https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA(2024)754450_EN.pdf). These data are fed into the **Integrated Joint Operations Platform**, which links ID numbers to physical characteristics and uses machine learning to generate lists of individuals for detention[[2]](https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA(2024)754450_EN.pdf#:~:text=As%20a%20result%20of%20this,legal%20process%2C%20based%20on%20the). Many Uyghurs have been sent to re‑education camps without charges or trials, purely on the basis of algorithmic predictions[[2]](https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA(2024)754450_EN.pdf#:~:text=As%20a%20result%20of%20this,legal%20process%2C%20based%20on%20the). The system’s designers claim it prevents terrorism; in practice it enables the mass detention of an ethnic group.

Predictive policing has also been tried in democratic societies, often with troubling results. A 2024 report from the U.S. Department of Justice notes that many predictive models rely on historical arrest data that reflect human biases[[3]](https://www.justice.gov/olp/media/1381796/dl#:~:text=for%20predictive%20policing%20may%20have,on%20privacy%20and%20civil%20liberties). Tools trained on such data tend to over‑police Black and Latino neighbourhoods and misidentify individuals. Some agencies have abandoned person‑based predictive policing because it offers little value while posing privacy risks[[3]](https://www.justice.gov/olp/media/1381796/dl#:~:text=for%20predictive%20policing%20may%20have,on%20privacy%20and%20civil%20liberties). The same report warns that risk assessment algorithms used in courts can perpetuate inequality by categorising people differently despite similar histories[[4]](https://www.justice.gov/olp/media/1381796/dl#:~:text=%E2%80%A2%20Risk%20Assessment,or%20fail%20to%20take%20into). At least seven mistaken arrests in the United States have been linked to facial recognition errors, nearly all involving Black citizens[[5]](https://www.justice.gov/olp/media/1381796/dl#:~:text=affected%20individuals,4). Here, the harm is not that AI “decided” to target minorities; it is that police departments adopted commercial systems without adequate testing or oversight, and built them on data sets saturated with systemic bias.

To understand why such abuses flourish, one must look at corporate incentives. Companies selling surveillance and predictive tools market them as neutral, efficient and profitable. Governments, eager to demonstrate security, purchase them. Ordinary citizens – often those with the least power – become the test subjects. This dynamic mirrors xenophobic narratives about immigrants as criminals: the focus is on control, not accountability. Recognising that AI tools can be weaponised for oppression helps us demand better protections. International human rights law already forbids arbitrary detention and discrimination; those norms must be enforced when algorithms are involved. Transparency about data sources and auditing of systems can help. Importantly, the victims of algorithmic repression are not anonymous masses but individuals with families, livelihoods and communities. Their stories remind us that “crime and disorder” can be inflicted by states as well as by thieves.

### Copyright expropriation and creative theft

Another area where fears of AI echo familiar immigrant stories is intellectual property. Creators worry that outsiders will steal their work and erode their livelihoods. In the context of AI, those fears are grounded in reality. Large language and image models are trained on vast corpora of text, art and music scraped from the internet, much of it copyrighted. A federal case decided in June 2025 illustrates the tension. Authors including Sarah Silverman and Ta‑Nehisi Coates sued Meta, alleging that the company used pirated versions of their books to train its model Llama. U.S. District Judge Vince Chhabria dismissed the case, stating that the plaintiffs had not provided enough evidence that Meta’s use would dilute the market for their work[[6]](https://www.reuters.com/sustainability/boards-policy-regulation/meta-fends-off-authors-us-copyright-lawsuit-over-ai-2025-06-25/#:~:text=June%2025%20%28Reuters%29%20,intelligence%20system%20infringed%20their%20copyrights). But he also emphasised that using copyrighted work without permission would be unlawful in “many circumstances” and that this ruling simply reflected poorly argued claims[[7]](https://www.reuters.com/sustainability/boards-policy-regulation/meta-fends-off-authors-us-copyright-lawsuit-over-ai-2025-06-25/#:~:text=Chhabria%20also%20said%2C%20however%2C%20that,of%20copyrighted%20materials). Chhabria noted that generative AI could flood markets with endless images, songs and articles created with a fraction of the effort humans expend[[8]](https://www.reuters.com/sustainability/boards-policy-regulation/meta-fends-off-authors-us-copyright-lawsuit-over-ai-2025-06-25/#:~:text=The%20judge%20said%20generative%20AI,be%20required%20to%20create%20them). In other words, the court recognised that models trained on copyrighted works may undermine incentives for creation[[8]](https://www.reuters.com/sustainability/boards-policy-regulation/meta-fends-off-authors-us-copyright-lawsuit-over-ai-2025-06-25/#:~:text=The%20judge%20said%20generative%20AI,be%20required%20to%20create%20them).

The deeper problem is opacity. Training datasets are often poorly documented and inconsistently licensed. A 2025 analysis from MIT Sloan observes that popular models like GPT‑4 are built on data whose lineage is unclear[[9]](https://mitsloan.mit.edu/ideas-made-to-matter/bringing-transparency-to-data-used-to-train-artificial-intelligence#:~:text=Popular%20large%20language%20models%20like,to%20a%20litany%20of%20risks). Without transparency, developers and users may unknowingly violate regulations like the European Union’s AI Act or expose themselves to copyright lawsuits[[9]](https://mitsloan.mit.edu/ideas-made-to-matter/bringing-transparency-to-data-used-to-train-artificial-intelligence#:~:text=Popular%20large%20language%20models%20like,to%20a%20litany%20of%20risks). Researchers with the **Data Provenance Initiative** audited more than 1,800 text datasets and found that license information was omitted in over 70 % of cases, while license types were miscategorised over 50 % of the time[[10]](https://mitsloan.mit.edu/ideas-made-to-matter/bringing-transparency-to-data-used-to-train-artificial-intelligence#:~:text=Conducted%20a%20systematic%20audit%20of,and%20they%20added). These errors make it hard for responsible companies to select lawful training data, and they make it easy for unscrupulous actors to claim ignorance.

Artists and journalists describe how their work is cannibalised. A painter might find derivative images flooding social media, produced by an AI trained on her portfolio. A novelist’s style can be imitated to generate endless fan fiction. When newspapers discovered that a model reproduced their paywalled articles verbatim, they sued. The question is not whether AI should be allowed to learn from human culture – that is how humans learn – but whether those whose work feeds the models should have a say and receive compensation. Historically, immigrants have been accused of stealing jobs and culture while in reality they enrich and diversify societies. The same is true of AI if governed wisely: it can inspire new art and collaboration. Policies requiring dataset provenance, consent from creators and fair remuneration would reduce the feeling of expropriation and align incentives toward creativity rather than theft.

### AI-powered scams and social engineering

Fraudsters have always exploited new technologies. The telephone enabled “grandparent scams” and phishing emails promised fake inheritances. AI lowers the cost and raises the realism of such tricks. In early 2024 the Electronic Frontier Foundation warned about **voice-cloning scams**. Criminals record a few seconds of a person’s voice – from social media posts or voicemail greetings – and then generate convincing audio messages[[11]](https://www.eff.org/deeplinks/2024/01/worried-about-ai-voice-clone-scams-create-family-password#:~:text=Your%20grandfather%20receives%20a%20call,account%20at%20an%20unfamiliar%20bank). One scenario recounted by investigators begins with a panicked call: a voice that sounds like your child says they have been arrested and need money. A second voice posing as a lawyer demands immediate payment. The voices are realistic because websites now offer cheap voice‑cloning services; in some cases scammers need only a thirty‑second audio sample[[11]](https://www.eff.org/deeplinks/2024/01/worried-about-ai-voice-clone-scams-create-family-password#:~:text=Your%20grandfather%20receives%20a%20call,account%20at%20an%20unfamiliar%20bank). Victims often wire money before they verify the situation. The EFF notes that while such scams are not yet widespread, the technology is accessible enough that they could become common[[11]](https://www.eff.org/deeplinks/2024/01/worried-about-ai-voice-clone-scams-create-family-password#:~:text=Your%20grandfather%20receives%20a%20call,account%20at%20an%20unfamiliar%20bank). They recommend simple defences such as family passphrases.

Corporate scams have already moved beyond prototypes. In 2024 an employee at the engineering firm **Arup** in Hong Kong joined a video call with people who looked and sounded like senior executives. He believed he was fulfilling a routine transfer, but the meeting was a deepfake. Criminals used AI‑generated video and voice to impersonate multiple executives simultaneously and convinced him to transfer HK$ 200 million (about £20 million) into accounts they controlled[[12]](https://www.theguardian.com/technology/article/2024/may/17/uk-engineering-arup-deepfake-scam-hong-kong-ai-video#:~:text=The%20British%20engineering%20company%20Arup,generated%20video%20call). The case shows that deepfake scams are no longer restricted to simple audio; they can involve coordinated video conferences, realistic backgrounds and body movements. Arup’s chief information officer told reporters that attacks including invoice fraud, WhatsApp voice spoofing and AI‑generated deepfakes have “risen sharply”[[12]](https://www.theguardian.com/technology/article/2024/may/17/uk-engineering-arup-deepfake-scam-hong-kong-ai-video#:~:text=The%20British%20engineering%20company%20Arup,generated%20video%20call). Law enforcement reported no arrests.

The problem is not only deception but also scale. AI models can create thousands of tailored phishing emails that mimic a company’s writing style and include plausible personal details gleaned from social media. Cybersecurity experts warn that generative models can automate phone scripts, design fake websites and simulate entire call‑centre operations. Once criminals used to need technical skills to craft malware; now they can ask a model to write ransomware or produce a PDF laced with malicious code. The victims are not just large companies. A small bakery owner in our town of Rugeley described receiving an email that appeared to be from her supplier; it contained a new bank account number and a friendly note referencing a recent conversation. She wired £8,000 before realising she had been scammed. She later learned that the attackers had used AI to scrape her posts and invoice templates, making their request familiar and credible. In this way AI democratises fraud: both criminals and victims are everyday people.

### AI-assisted hacking and cybercrime

The same capabilities that allow AI to compose poetry can be redirected to composing malware. A 2025 threat intelligence report from OpenAI describes how its investigators detected and disrupted several campaigns in which state‑linked actors used generative models to enhance cyberattacks[[13]](https://cdn.openai.com/threat-intelligence-reports/5f73af09-a3a3-4a55-992e-069237681620/disrupting-malicious-uses-of-ai-june-2025.pdf#:~:text=Our%20mission%20is%20to%20ensure,abusive%20activity%20including%20social%20engineering). In one case, North Korean operators used ChatGPT to generate résumés and cover letters at scale to infiltrate IT companies with deceptive job applications[[14]](https://cdn.openai.com/threat-intelligence-reports/5f73af09-a3a3-4a55-992e-069237681620/disrupting-malicious-uses-of-ai-june-2025.pdf#:~:text=Behavior%20Similar%20to%20the%20threat,different%20parts%20of%20the%20world). They even sought advice on remote‑work setups and used AI‑generated content to recruit collaborators[[15]](https://cdn.openai.com/threat-intelligence-reports/5f73af09-a3a3-4a55-992e-069237681620/disrupting-malicious-uses-of-ai-june-2025.pdf#:~:text=The%20core%20operators%20attempted%20to,different%20parts%20of%20the%20world). The same report documents Chinese and Russian operations that used AI to craft social‑engineering scripts, spam comments and disinformation[[16]](https://cdn.openai.com/threat-intelligence-reports/5f73af09-a3a3-4a55-992e-069237681620/disrupting-malicious-uses-of-ai-june-2025.pdf#:~:text=force%20multiplier%20for%20our%20expert,a%20likely%20task%20scam%20from). OpenAI notes that AI can be used at every stage of an attack: reconnaissance (asking about software vulnerabilities), development (writing and debugging malware), phishing (crafting credible emails) and post‑exploitation (automating network scans)[[17]](https://www.bleepingcomputer.com/news/security/openai-confirms-threat-actors-use-chatgpt-to-write-malware/#:~:text=OpenAI%20has%20disrupted%20over%2020,phishing%20attacks). The company says its systems help detect and shut down such misuse, but the broader security community must adapt.

Generative AI also reduces skill barriers for hackers. The BleepingComputer analysis of OpenAI’s February 2025 disruption efforts found that threat actors used models to ask about default passwords, identify configuration errors, and obfuscate malicious code[[18]](https://www.bleepingcomputer.com/news/security/openai-confirms-threat-actors-use-chatgpt-to-write-malware/#:~:text=Upon%20further%20investigation%2C%20OpenAI%20found,help%20of%20the%20LLM%20tool). Attackers who previously relied on publicly available scripts can now generate customised payloads and adapt them to specific networks. Combined with stolen credentials from massive data breaches, these tools enable “credential stuffing” on an industrial scale. The Identity Theft Resource Center (ITRC) reports that in the first half of 2025 there were 1,732 data compromises, an 11 % increase over the same period in 2024[[19]](https://www.hipaajournal.com/h1-2025-data-breach-report/#:~:text=This%20year%20is%20set%20to,full%20year%20total%20for%202024). Cyberattacks accounted for 77.8 % of these incidents, compromising the personal data of more than 114 million individuals[[20]](https://www.hipaajournal.com/h1-2025-data-breach-report/#:~:text=This%20year%20is%20set%20to,full%20year%20total%20for%202024). Phishing and business‑email compromise were the second most common cause, followed by ransomware[[21]](https://www.hipaajournal.com/h1-2025-data-breach-report/#:~:text=incidents,error%2C%20impacting%20more%20than%206). The ITRC notes that generative AI platforms are increasingly used to create sophisticated phishing campaigns that are harder for organisations and individuals to detect[[22]](https://www.hipaajournal.com/h1-2025-data-breach-report/#:~:text=One%20clear%20trend%20identified%20by,organizations%20and%20individuals%20to%20detect).

These trends create what cybersecurity scholar Qi Liao calls “ransomware 2.0”: attacks that not only encrypt data but also steal and sell it[[23]](https://www.cmich.edu/news/details/how-can-you-protect-your-privacy-money-from-ai#:~:text=Traditionally%2C%20artificial%20intelligence%20,DDoS%29%20attacks%20and%20malware). Liao explains that AI can automate and personalise phishing by analysing social‑media posts, generate deepfakes for blackmail, and craft malware that evades detection[[23]](https://www.cmich.edu/news/details/how-can-you-protect-your-privacy-money-from-ai#:~:text=Traditionally%2C%20artificial%20intelligence%20,DDoS%29%20attacks%20and%20malware). He warns that adversarial machine learning techniques can bypass spam filters and anti‑virus systems[[23]](https://www.cmich.edu/news/details/how-can-you-protect-your-privacy-money-from-ai#:~:text=Traditionally%2C%20artificial%20intelligence%20,DDoS%29%20attacks%20and%20malware). Meanwhile, criminals can cross‑reference anonymised data from medical, financial and voter records to re‑identify individuals and learn their habits[[24]](https://www.cmich.edu/news/details/how-can-you-protect-your-privacy-money-from-ai#:~:text=Even%20publicly%20available%20data%20can,and%20revealing%20their%20behavior%20patterns). It is a chilling inversion of AI’s promise: tools meant to secure and streamline our lives are used to break into them.

### Identity theft and the erosion of privacy

Identity theft is the crime that best illustrates the intersection of AI, data and everyday life. In 2025 the UK fraud prevention service **Cifas** reported that more than 217,000 fraud risk cases were logged in the first half of the year, a record number[[25]](https://identityweek.net/ai-fuels-surge-in-identity-fraud-as-people-sell-their-personal-information/#:~:text=The%20latest%20Fraudscape%20report%2C%20launched,month%20period). Of these, over 118,000 were identity fraud cases[[26]](https://identityweek.net/ai-fuels-surge-in-identity-fraud-as-people-sell-their-personal-information/#:~:text=Overall%2C%20more%20than%20118%2C000%20cases,that%20can%20bypass%20security%20checks). Cifas warns that AI enables criminals to create **synthetic identities**: fabricated personas that combine stolen data with invented attributes to bypass verification systems[[27]](https://identityweek.net/ai-fuels-surge-in-identity-fraud-as-people-sell-their-personal-information/#:~:text=members%20from%20January,month%20period). Economic pressures and risky consumer behaviour mean that some individuals even sell their own identities, exposing themselves to liability when criminals take loans or open accounts in their names[[28]](https://identityweek.net/ai-fuels-surge-in-identity-fraud-as-people-sell-their-personal-information/#:~:text=The%20report%20also%20highlights%20another,in%20their%20name%20by%20criminals). Sectors once considered low risk, such as motor insurance and telecommunications, are now prime targets[[29]](https://identityweek.net/ai-fuels-surge-in-identity-fraud-as-people-sell-their-personal-information/#:~:text=%2A%20Insurance%20fraud%20up%2025,identity%20theft%20in%20motor%20insurance). Facility takeovers – where attackers seize control of existing bank or phone accounts – accounted for 18 % of all fraud, and the telecommunications sector saw the steepest rise[[30]](https://identityweek.net/ai-fuels-surge-in-identity-fraud-as-people-sell-their-personal-information/#:~:text=Facility%20%28account%29%20takeover%20climbed%201,in%202024).

The ITRC adds that data breaches feed these scams. Among the 165 million individuals affected by breaches in the first half of 2025, many had their login credentials and personal identifiers packaged into a single database[[31]](https://www.hipaajournal.com/h1-2025-data-breach-report/#:~:text=ITRC%20also%20draws%20attention%20to,themselves%20from%20identity%20fraud%20and). This repackaging means that criminals can easily buy lists of usernames and passwords and use AI to test them across multiple services, a tactic known as credential stuffing. James E. Lee of the ITRC notes that more than two‑thirds of breach notifications fail to disclose the attack vector, leaving consumers unaware of how their data were stolen[[32]](https://www.hipaajournal.com/h1-2025-data-breach-report/#:~:text=cases%2C%20acceleration%20of%20the%20trends,of%20all%20notices%20did). The lack of transparency prevents individuals from taking targeted protective steps and undermines public trust.

Beyond financial fraud, privacy erosion occurs through **de‑anonymisation**. AI can cross‑reference public data with stolen datasets to identify individuals in supposedly anonymous datasets[[24]](https://www.cmich.edu/news/details/how-can-you-protect-your-privacy-money-from-ai#:~:text=Even%20publicly%20available%20data%20can,and%20revealing%20their%20behavior%20patterns). For example, a malicious actor could combine a health‑care dataset with voter registration information and smart‑home device logs to infer who is being treated for a particular condition and when they are away from home. Such granular profiling enables blackmail, targeted advertising and discrimination. The victims might never know how they were selected for a scam or denied a job. If data is “the new oil,” then identity theft is the new pipeline leak – invisible until a spill occurs.

### Algorithmic coercion into violence and self‑harm

While AI is often blamed for making people do things, the reality is subtler. Recommendation systems are designed to maximise engagement, and they do so by amplifying content that elicits strong emotions. Sometimes that content includes hate, extremism and self‑harm. A 2024 study of Instagram’s moderation revealed that Meta failed to remove any of 85 posts showing various degrees of self‑harm[[33]](https://www.theguardian.com/technology/2024/nov/30/instagram-actively-helping-to-spread-of-self-harm-among-teenagers-study-suggests#:~:text=Meta%20is%20actively%20helping%20self,found%20its%20moderation%20%E2%80%9Cextremely%20inadequate%E2%80%9D). Even more troubling, Instagram’s algorithm encouraged accounts posting self‑harm images to befriend each other, helping a network of harmful content grow[[33]](https://www.theguardian.com/technology/2024/nov/30/instagram-actively-helping-to-spread-of-self-harm-among-teenagers-study-suggests#:~:text=Meta%20is%20actively%20helping%20self,found%20its%20moderation%20%E2%80%9Cextremely%20inadequate%E2%80%9D). Psychologist Lotte Rubæk told reporters that such failures trigger vulnerable users and can be a matter of life and death[[34]](https://www.theguardian.com/technology/2024/nov/30/instagram-actively-helping-to-spread-of-self-harm-among-teenagers-study-suggests#:~:text=Failing%20to%20moderate%20self,groups%20are%20small%2C%E2%80%9D%20he%20said). Rather than accidental oversights, these outcomes reflect an incentive structure that rewards engagement over safety.

Scholarship on extremism shows a similar pattern. In a review of empirical studies, communication researcher Aaron Shaw observes that while popular narratives blame recommendation algorithms for radicalising innocent users, the evidence is more complex[[35]](https://pmc.ncbi.nlm.nih.gov/articles/PMC10468141/#:~:text=Many%20would%20blame%20social%20media,1). YouTube’s algorithm rarely recommends extremist content to casual viewers[[36]](https://pmc.ncbi.nlm.nih.gov/articles/PMC10468141/#:~:text=Many%20would%20blame%20social%20media,1). However, the platform still hosts extremist communities, and algorithmic structures allow users who already hold resentful beliefs to deepen them[[37]](https://pmc.ncbi.nlm.nih.gov/articles/PMC10468141/#:~:text=Many%20would%20blame%20social%20media,content%20espousing%20violence%2C%20hate%2C%20and). Shaw notes that radicalised individuals often arrive via external links or subscriptions to fringe channels, not through accidental recommendations[[38]](https://pmc.ncbi.nlm.nih.gov/articles/PMC10468141/#:~:text=content%20policies%20and%20extremist%20political,point%2C%20we%20should%20recall%20that). Yet once they are inside, the lack of effective moderation means they can exchange conspiracies, misogyny and hate without hindrance. The existence of these networks has real consequences: researchers have linked participation in extremist online spaces to increased participation in civil unrest and violence[[39]](https://pmc.ncbi.nlm.nih.gov/articles/PMC10468141/#:~:text=The%20fact%20that%20a%20substantial,these%20individuals%20to%20adopt%20such).

There is also evidence that algorithms can overexpose vulnerable people to self‑harm content. A 2025 qualitative study with young people, policy makers and social‑media industry professionals found that 83 % of youth have encountered self‑harm or suicide content online, often before the age of fourteen[[40]](https://pmc.ncbi.nlm.nih.gov/articles/PMC11933773/#:~:text=media,to%20potentially%20harmful%20content%20when). Participants warned that the way social‑media algorithms operate can lead to overexposure when teens interact with peers[[40]](https://pmc.ncbi.nlm.nih.gov/articles/PMC11933773/#:~:text=media,to%20potentially%20harmful%20content%20when). Despite these risks, national suicide‑prevention strategies rarely include social media guidelines[[41]](https://pmc.ncbi.nlm.nih.gov/articles/PMC11933773/#:~:text=makers%20to%20keep%20young%20people,46). Social‑media companies have adopted some policies, but they differ across platforms and lack independent oversight[[42]](https://pmc.ncbi.nlm.nih.gov/articles/PMC11933773/#:~:text=to%20self,a%20key%20gap%20in%20youth). The result is a patchwork of private rules that leaves many gaps. When tragedies occur, companies often say “we didn’t design it to do that.” That refrain echoes the way some societies blame immigrants for social ills without examining the policies and incentives that create the conditions for exploitation.

### Opacity, accountability and the black box problem

Underlying many of these harms is the **black box** nature of AI systems and the secrecy of their creators. The MIT Sloan report mentioned earlier points out that without transparency into the lineage of training data, it is difficult to assess legal risk or bias[[9]](https://mitsloan.mit.edu/ideas-made-to-matter/bringing-transparency-to-data-used-to-train-artificial-intelligence#:~:text=Popular%20large%20language%20models%20like,to%20a%20litany%20of%20risks). The Data Provenance Initiative discovered that more than half of licensing information in popular datasets was miscategorised and that over 70 % of datasets lacked license information[[10]](https://mitsloan.mit.edu/ideas-made-to-matter/bringing-transparency-to-data-used-to-train-artificial-intelligence#:~:text=Conducted%20a%20systematic%20audit%20of,and%20they%20added). These findings show that many companies build products on shaky foundations.

Opacity is also present in the public sector. During the first half of 2025, the ITRC noted that 69 % of data‑breach notifications failed to disclose how the breach occurred[[32]](https://www.hipaajournal.com/h1-2025-data-breach-report/#:~:text=cases%2C%20acceleration%20of%20the%20trends,of%20all%20notices%20did). This means that victims are left without information that could help them respond, and patterns of attack remain hidden. Similarly, when algorithms misidentify individuals or recommend harmful content, companies often claim trade secrets to avoid revealing how their systems work. Such secrecy hinders redress and allows abuses to continue. In a 2024 Mozilla Foundation report examining watermarking and labelling of AI‑generated content, researchers concluded that human‑facing labels are easy to manipulate and can even erode trust[[43]](https://www.mozillafoundation.org/en/research/library/in-transparency-we-trust/research-report/#:~:text=%23%20Human,short). They argue for machine‑readable markers combined with robust detection mechanisms and a multifaceted governance approach[[44]](https://www.mozillafoundation.org/en/research/library/in-transparency-we-trust/research-report/#:~:text=%23%20Machine,combined%20with%20robust%20detection%20mechanisms). Without strong disclosure requirements, citizens cannot know whether a video or call is authentic.

The black box problem is exacerbated by corporate incentives. Proprietary models are guarded to maintain a competitive edge, even when they have public consequences. This is akin to a landlord refusing to repair a building because the blueprint is a trade secret. Regulation must strike a balance between protecting innovation and protecting the public. The California Generative AI: Training Data Transparency Act, signed in 2024, is an example of legislation that requires companies to document their training data sources. European policymakers are considering similar measures. Transparency alone is not a panacea – after all, criminals will not volunteer their methods – but public institutions have a duty to set standards.

### Agency, law and remedies

The narrative that AI immigrants bring crime and disorder can lead to two dangerous responses: complacency (“nothing can be done”) or technophobic bans. Both ignore the role of human agency. The harms described above stem from human decisions to deploy surveillance against minorities, build opaque models on stolen data, design engagement‑maximising algorithms without safety checks, and weaponise generative tools for fraud. Human decisions can also mitigate those harms.

First, **strengthen civil liability**. Just as landlords must maintain safe buildings, companies that deploy AI should be liable for foreseeable harms. If a predictive policing tool leads to unlawful detention, those affected should have legal standing to sue and demand audits. Courts could require companies to disclose model assumptions and training data in discovery. Whistleblower protections should extend to engineers and ethicists who raise concerns about misuse.

Second, adopt **data‑provenance requirements and licensing**. Policy makers can mandate that companies maintain records of where training data come from, what licenses apply, and how data subjects can opt out. The Data Provenance Initiative’s approach – auditing datasets, reducing unspecified licenses from 70 % to 30 %, and offering tools for dataset exploration[[45]](https://mitsloan.mit.edu/ideas-made-to-matter/bringing-transparency-to-data-used-to-train-artificial-intelligence#:~:text=Conducted%20a%20systematic%20audit%20of,appropriate%20data%20for%20their%20needs) – shows that this is feasible. Requiring provenance helps copyright holders and protects companies from unknowingly ingesting harmful material (such as child abuse images found in the LAION‑5B dataset[[46]](https://mitsloan.mit.edu/ideas-made-to-matter/bringing-transparency-to-data-used-to-train-artificial-intelligence#:~:text=A%20trio%20of%20events%20in,competing%20model%2C%20thus%20violating%20the)).

Third, impose **mandatory red‑team audits**. Before deployment, systems should undergo adversarial testing by independent researchers to identify vulnerabilities and misuse pathways. This is common in cybersecurity and can be adapted for AI. Red teams should include experts in human rights, mental health, and cybersecurity, not just technical engineers. Their findings should be published with fixes and documented.

Fourth, improve **user education and public infrastructure**. Citizens need accessible information about deepfake scams, identity theft and privacy safeguards. Government agencies and non‑profits can run public‑awareness campaigns similar to anti‑smoking or seat‑belt campaigns. Simple practices like setting family passphrases, verifying calls through known numbers, using password managers, and enabling two‑factor authentication reduce risk. Public libraries and community centres can offer workshops on digital literacy.

Fifth, support **platform governance**. Social‑media companies should be required to provide researchers access to data (with appropriate privacy safeguards) to study algorithmic harms. Regulatory bodies like the UK’s Online Safety Act or the proposed U.S. Kids Online Safety Act should include provisions specific to suicide and self‑harm content. Algorithms that recommend content should have built‑in throttles for harmful material and be subject to external review.

Finally, recognise **collective responsibility**. Crime and disorder do not arise because tools are inherently evil; they arise because humans choose to exploit them. Equating AI with an invading criminal class risks repeating the injustices of past scapegoating. A more constructive view sees AI as a new population of residents in our digital society. Like all residents, they need rules, oversight and a culture of care. We must shape those rules so that AI contributes to safety and justice rather than undermining them.

### Closing reflection

If the previous chapters have shown that AI immigrants do not always integrate or always work, this chapter confronts the fear that they will bring crime and disorder. The evidence is nuanced. Yes, AI can be weaponised for repression, fraud and hacking. Yes, algorithms can amplify self‑harm and extremism. Yes, opaque practices make accountability difficult. But at every turn, human decisions, incentives and institutions shape those outcomes. Understanding the mechanisms of harm – predictive policing built on biased data, datasets scraped without consent, algorithms designed to maximise engagement – empowers us to change them.

As you finish this chapter, consider the parallels between today’s AI and yesterday’s immigrants. Demonising a group rarely leads to safety; building equitable systems does. What role can you play, as a voter, consumer or creator, in demanding transparency, accountability and care from those who build and deploy these powerful tools?

### Editorial note and references

This chapter draws on investigative journalism and research reports to illustrate how AI technologies can enable crime and disorder. The **European Parliament’s 2024 report** on human rights abuses in Xinjiang provided evidence of predictive policing and mass surveillance[[1]](https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA(2024)754450_EN.pdf)[[2]](https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA(2024)754450_EN.pdf#:~:text=As%20a%20result%20of%20this,legal%20process%2C%20based%20on%20the). The **U.S. Department of Justice’s 2024 review** of AI in criminal justice informed discussions on predictive policing and facial‑recognition errors[[3]](https://www.justice.gov/olp/media/1381796/dl#:~:text=for%20predictive%20policing%20may%20have,on%20privacy%20and%20civil%20liberties)[[4]](https://www.justice.gov/olp/media/1381796/dl#:~:text=%E2%80%A2%20Risk%20Assessment,or%20fail%20to%20take%20into). For copyright and data transparency issues, I relied on **Reuters’ coverage** of Meta’s 2025 copyright case[[7]](https://www.reuters.com/sustainability/boards-policy-regulation/meta-fends-off-authors-us-copyright-lawsuit-over-ai-2025-06-25/#:~:text=Chhabria%20also%20said%2C%20however%2C%20that,of%20copyrighted%20materials) and the **MIT Sloan analysis** of AI data provenance[[9]](https://mitsloan.mit.edu/ideas-made-to-matter/bringing-transparency-to-data-used-to-train-artificial-intelligence#:~:text=Popular%20large%20language%20models%20like,to%20a%20litany%20of%20risks)[[10]](https://mitsloan.mit.edu/ideas-made-to-matter/bringing-transparency-to-data-used-to-train-artificial-intelligence#:~:text=Conducted%20a%20systematic%20audit%20of,and%20they%20added). Cases of voice and deepfake scams came from the **Electronic Frontier Foundation’s 2024 warning**[[11]](https://www.eff.org/deeplinks/2024/01/worried-about-ai-voice-clone-scams-create-family-password#:~:text=Your%20grandfather%20receives%20a%20call,account%20at%20an%20unfamiliar%20bank) and **Guardian reporting** on the Arup scam[[12]](https://www.theguardian.com/technology/article/2024/may/17/uk-engineering-arup-deepfake-scam-hong-kong-ai-video#:~:text=The%20British%20engineering%20company%20Arup,generated%20video%20call). Statistics on identity theft and data breaches were taken from **Cifas’ Fraudscape report**[[25]](https://identityweek.net/ai-fuels-surge-in-identity-fraud-as-people-sell-their-personal-information/#:~:text=The%20latest%20Fraudscape%20report%2C%20launched,month%20period) and the **Identity Theft Resource Center’s H1 2025 Data Breach Analysis**[[47]](https://www.hipaajournal.com/h1-2025-data-breach-report/#:~:text=This%20year%20is%20set%20to,full%20year%20total%20for%202024). Additional insights on AI‑assisted hacking and phishing draw on OpenAI’s threat intelligence update[[13]](https://cdn.openai.com/threat-intelligence-reports/5f73af09-a3a3-4a55-992e-069237681620/disrupting-malicious-uses-of-ai-june-2025.pdf#:~:text=Our%20mission%20is%20to%20ensure,abusive%20activity%20including%20social%20engineering) and BleepingComputer’s analysis of malicious AI usage[[17]](https://www.bleepingcomputer.com/news/security/openai-confirms-threat-actors-use-chatgpt-to-write-malware/#:~:text=OpenAI%20has%20disrupted%20over%2020,phishing%20attacks)[[18]](https://www.bleepingcomputer.com/news/security/openai-confirms-threat-actors-use-chatgpt-to-write-malware/#:~:text=Upon%20further%20investigation%2C%20OpenAI%20found,help%20of%20the%20LLM%20tool). Mental health and algorithmic harm discussions were informed by the **Guardian’s self‑harm investigation**[[33]](https://www.theguardian.com/technology/2024/nov/30/instagram-actively-helping-to-spread-of-self-harm-among-teenagers-study-suggests#:~:text=Meta%20is%20actively%20helping%20self,found%20its%20moderation%20%E2%80%9Cextremely%20inadequate%E2%80%9D)[[34]](https://www.theguardian.com/technology/2024/nov/30/instagram-actively-helping-to-spread-of-self-harm-among-teenagers-study-suggests#:~:text=Failing%20to%20moderate%20self,groups%20are%20small%2C%E2%80%9D%20he%20said) and the **Journal of Medical Internet Research study** on youth exposure to self‑harm content[[40]](https://pmc.ncbi.nlm.nih.gov/articles/PMC11933773/#:~:text=media,to%20potentially%20harmful%20content%20when).

[[1]](https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA(2024)754450_EN.pdf) [[2]](https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA(2024)754450_EN.pdf#:~:text=As%20a%20result%20of%20this,legal%20process%2C%20based%20on%20the) Artificial intelligence and human rights: Using AI as a weapon of repression and its impact on human rights

<https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA(2024)754450_EN.pdf>

[[3]](https://www.justice.gov/olp/media/1381796/dl#:~:text=for%20predictive%20policing%20may%20have,on%20privacy%20and%20civil%20liberties) [[4]](https://www.justice.gov/olp/media/1381796/dl#:~:text=%E2%80%A2%20Risk%20Assessment,or%20fail%20to%20take%20into) [[5]](https://www.justice.gov/olp/media/1381796/dl#:~:text=affected%20individuals,4) Artificial Intelligence and Criminal Justice, Final Report, December 3, 2024

<https://www.justice.gov/olp/media/1381796/dl>

[[6]](https://www.reuters.com/sustainability/boards-policy-regulation/meta-fends-off-authors-us-copyright-lawsuit-over-ai-2025-06-25/#:~:text=June%2025%20%28Reuters%29%20,intelligence%20system%20infringed%20their%20copyrights) [[7]](https://www.reuters.com/sustainability/boards-policy-regulation/meta-fends-off-authors-us-copyright-lawsuit-over-ai-2025-06-25/#:~:text=Chhabria%20also%20said%2C%20however%2C%20that,of%20copyrighted%20materials) [[8]](https://www.reuters.com/sustainability/boards-policy-regulation/meta-fends-off-authors-us-copyright-lawsuit-over-ai-2025-06-25/#:~:text=The%20judge%20said%20generative%20AI,be%20required%20to%20create%20them) Meta fends off authors' US copyright lawsuit over AI | Reuters

<https://www.reuters.com/sustainability/boards-policy-regulation/meta-fends-off-authors-us-copyright-lawsuit-over-ai-2025-06-25/>

[[9]](https://mitsloan.mit.edu/ideas-made-to-matter/bringing-transparency-to-data-used-to-train-artificial-intelligence#:~:text=Popular%20large%20language%20models%20like,to%20a%20litany%20of%20risks) [[10]](https://mitsloan.mit.edu/ideas-made-to-matter/bringing-transparency-to-data-used-to-train-artificial-intelligence#:~:text=Conducted%20a%20systematic%20audit%20of,and%20they%20added) [[45]](https://mitsloan.mit.edu/ideas-made-to-matter/bringing-transparency-to-data-used-to-train-artificial-intelligence#:~:text=Conducted%20a%20systematic%20audit%20of,appropriate%20data%20for%20their%20needs) [[46]](https://mitsloan.mit.edu/ideas-made-to-matter/bringing-transparency-to-data-used-to-train-artificial-intelligence#:~:text=A%20trio%20of%20events%20in,competing%20model%2C%20thus%20violating%20the) Bringing transparency to the data used to train artificial intelligence | MIT Sloan

<https://mitsloan.mit.edu/ideas-made-to-matter/bringing-transparency-to-data-used-to-train-artificial-intelligence>

[[11]](https://www.eff.org/deeplinks/2024/01/worried-about-ai-voice-clone-scams-create-family-password#:~:text=Your%20grandfather%20receives%20a%20call,account%20at%20an%20unfamiliar%20bank) Worried About AI Voice Clone Scams? Create a Family Password | Electronic Frontier Foundation

<https://www.eff.org/deeplinks/2024/01/worried-about-ai-voice-clone-scams-create-family-password>

[[12]](https://www.theguardian.com/technology/article/2024/may/17/uk-engineering-arup-deepfake-scam-hong-kong-ai-video#:~:text=The%20British%20engineering%20company%20Arup,generated%20video%20call) UK engineering firm Arup falls victim to £20m deepfake scam | Engineering | The Guardian

<https://www.theguardian.com/technology/article/2024/may/17/uk-engineering-arup-deepfake-scam-hong-kong-ai-video>

[[13]](https://cdn.openai.com/threat-intelligence-reports/5f73af09-a3a3-4a55-992e-069237681620/disrupting-malicious-uses-of-ai-june-2025.pdf#:~:text=Our%20mission%20is%20to%20ensure,abusive%20activity%20including%20social%20engineering) [[14]](https://cdn.openai.com/threat-intelligence-reports/5f73af09-a3a3-4a55-992e-069237681620/disrupting-malicious-uses-of-ai-june-2025.pdf#:~:text=Behavior%20Similar%20to%20the%20threat,different%20parts%20of%20the%20world) [[15]](https://cdn.openai.com/threat-intelligence-reports/5f73af09-a3a3-4a55-992e-069237681620/disrupting-malicious-uses-of-ai-june-2025.pdf#:~:text=The%20core%20operators%20attempted%20to,different%20parts%20of%20the%20world) [[16]](https://cdn.openai.com/threat-intelligence-reports/5f73af09-a3a3-4a55-992e-069237681620/disrupting-malicious-uses-of-ai-june-2025.pdf#:~:text=force%20multiplier%20for%20our%20expert,a%20likely%20task%20scam%20from) Disrupting malicious uses of AI: June 2025

<https://cdn.openai.com/threat-intelligence-reports/5f73af09-a3a3-4a55-992e-069237681620/disrupting-malicious-uses-of-ai-june-2025.pdf>

[[17]](https://www.bleepingcomputer.com/news/security/openai-confirms-threat-actors-use-chatgpt-to-write-malware/#:~:text=OpenAI%20has%20disrupted%20over%2020,phishing%20attacks) [[18]](https://www.bleepingcomputer.com/news/security/openai-confirms-threat-actors-use-chatgpt-to-write-malware/#:~:text=Upon%20further%20investigation%2C%20OpenAI%20found,help%20of%20the%20LLM%20tool) OpenAI confirms threat actors use ChatGPT to write malware

<https://www.bleepingcomputer.com/news/security/openai-confirms-threat-actors-use-chatgpt-to-write-malware/>

[[19]](https://www.hipaajournal.com/h1-2025-data-breach-report/#:~:text=This%20year%20is%20set%20to,full%20year%20total%20for%202024) [[20]](https://www.hipaajournal.com/h1-2025-data-breach-report/#:~:text=This%20year%20is%20set%20to,full%20year%20total%20for%202024) [[21]](https://www.hipaajournal.com/h1-2025-data-breach-report/#:~:text=incidents,error%2C%20impacting%20more%20than%206) [[22]](https://www.hipaajournal.com/h1-2025-data-breach-report/#:~:text=One%20clear%20trend%20identified%20by,organizations%20and%20individuals%20to%20detect) [[31]](https://www.hipaajournal.com/h1-2025-data-breach-report/#:~:text=ITRC%20also%20draws%20attention%20to,themselves%20from%20identity%20fraud%20and) [[32]](https://www.hipaajournal.com/h1-2025-data-breach-report/#:~:text=cases%2C%20acceleration%20of%20the%20trends,of%20all%20notices%20did) [[47]](https://www.hipaajournal.com/h1-2025-data-breach-report/#:~:text=This%20year%20is%20set%20to,full%20year%20total%20for%202024) Data Breaches Up 10% Although Victim Count Falls Sharply

<https://www.hipaajournal.com/h1-2025-data-breach-report/>

[[23]](https://www.cmich.edu/news/details/how-can-you-protect-your-privacy-money-from-ai#:~:text=Traditionally%2C%20artificial%20intelligence%20,DDoS%29%20attacks%20and%20malware) [[24]](https://www.cmich.edu/news/details/how-can-you-protect-your-privacy-money-from-ai#:~:text=Even%20publicly%20available%20data%20can,and%20revealing%20their%20behavior%20patterns) How can you protect your privacy, money from AI?

<https://www.cmich.edu/news/details/how-can-you-protect-your-privacy-money-from-ai>

[[25]](https://identityweek.net/ai-fuels-surge-in-identity-fraud-as-people-sell-their-personal-information/#:~:text=The%20latest%20Fraudscape%20report%2C%20launched,month%20period) [[26]](https://identityweek.net/ai-fuels-surge-in-identity-fraud-as-people-sell-their-personal-information/#:~:text=Overall%2C%20more%20than%20118%2C000%20cases,that%20can%20bypass%20security%20checks) [[27]](https://identityweek.net/ai-fuels-surge-in-identity-fraud-as-people-sell-their-personal-information/#:~:text=members%20from%20January,month%20period) [[28]](https://identityweek.net/ai-fuels-surge-in-identity-fraud-as-people-sell-their-personal-information/#:~:text=The%20report%20also%20highlights%20another,in%20their%20name%20by%20criminals) [[29]](https://identityweek.net/ai-fuels-surge-in-identity-fraud-as-people-sell-their-personal-information/#:~:text=%2A%20Insurance%20fraud%20up%2025,identity%20theft%20in%20motor%20insurance) [[30]](https://identityweek.net/ai-fuels-surge-in-identity-fraud-as-people-sell-their-personal-information/#:~:text=Facility%20%28account%29%20takeover%20climbed%201,in%202024) AI fuels surge in identity fraud, as people sell their personal information - Identity Week

<https://identityweek.net/ai-fuels-surge-in-identity-fraud-as-people-sell-their-personal-information/>

[[33]](https://www.theguardian.com/technology/2024/nov/30/instagram-actively-helping-to-spread-of-self-harm-among-teenagers-study-suggests#:~:text=Meta%20is%20actively%20helping%20self,found%20its%20moderation%20%E2%80%9Cextremely%20inadequate%E2%80%9D) [[34]](https://www.theguardian.com/technology/2024/nov/30/instagram-actively-helping-to-spread-of-self-harm-among-teenagers-study-suggests#:~:text=Failing%20to%20moderate%20self,groups%20are%20small%2C%E2%80%9D%20he%20said) Instagram actively helping spread of self-harm among teenagers, study finds | Meta | The Guardian

<https://www.theguardian.com/technology/2024/nov/30/instagram-actively-helping-to-spread-of-self-harm-among-teenagers-study-suggests>

[[35]](https://pmc.ncbi.nlm.nih.gov/articles/PMC10468141/#:~:text=Many%20would%20blame%20social%20media,1) [[36]](https://pmc.ncbi.nlm.nih.gov/articles/PMC10468141/#:~:text=Many%20would%20blame%20social%20media,1) [[37]](https://pmc.ncbi.nlm.nih.gov/articles/PMC10468141/#:~:text=Many%20would%20blame%20social%20media,content%20espousing%20violence%2C%20hate%2C%20and) [[38]](https://pmc.ncbi.nlm.nih.gov/articles/PMC10468141/#:~:text=content%20policies%20and%20extremist%20political,point%2C%20we%20should%20recall%20that) [[39]](https://pmc.ncbi.nlm.nih.gov/articles/PMC10468141/#:~:text=The%20fact%20that%20a%20substantial,these%20individuals%20to%20adopt%20such) Social media, extremism, and radicalization - PMC

<https://pmc.ncbi.nlm.nih.gov/articles/PMC10468141/>

[[40]](https://pmc.ncbi.nlm.nih.gov/articles/PMC11933773/#:~:text=media,to%20potentially%20harmful%20content%20when) [[41]](https://pmc.ncbi.nlm.nih.gov/articles/PMC11933773/#:~:text=makers%20to%20keep%20young%20people,46) [[42]](https://pmc.ncbi.nlm.nih.gov/articles/PMC11933773/#:~:text=to%20self,a%20key%20gap%20in%20youth) Online Safety When Considering Self-Harm and Suicide-Related Content: Qualitative Focus Group Study With Young People, Policy Makers, and Social Media Industry Professionals - PMC

<https://pmc.ncbi.nlm.nih.gov/articles/PMC11933773/>

[[43]](https://www.mozillafoundation.org/en/research/library/in-transparency-we-trust/research-report/#:~:text=%23%20Human,short) [[44]](https://www.mozillafoundation.org/en/research/library/in-transparency-we-trust/research-report/#:~:text=%23%20Machine,combined%20with%20robust%20detection%20mechanisms) In Transparency We Trust? - Mozilla Foundation

<https://www.mozillafoundation.org/en/research/library/in-transparency-we-trust/research-report/>