



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/213396/>

Version: Published Version

---

**Article:**

Bellaby, R. (2024) The ethical problems of 'intelligence–AI'. *International Affairs*, 100 (6). pp. 2525-2542. ISSN: 0020-5850

<https://doi.org/10.1093/ia/iae227>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

# The ethical problems of 'intelligence–AI'

ROSS BELLABY

Interest in artificial intelligence (AI) has grown rapidly in recent years, along with the ability to process and analyse information in far vaster quantities from a more diverse range of sources, and to propose new forms of evaluation, far beyond what humans can achieve. Given this processing power, the potential that it represents to state security and intelligence actors should come as no surprise. In order to protect the political community across its social, economic, political and military interests, the intelligence community relies on being able to collect as much information from as many sources as possible—both open and secret—so as to detect threats and prevent them from materializing. However, this drive for ever greater quantities of information has generated a significant bottleneck at the processing and analysis stage, where there is too much information to review. AI therefore offers the potential to process this information in quantities and at speeds previously not seen, while also offering new analytical insights. Indeed, the United States' National Security Commission on Artificial Intelligence stated in 2023 that 'AI will help intelligence professionals find needles in haystacks, connect the dots, and disrupt dangerous plots by discerning trends and discovering previously hidden or masked indications and warnings'.<sup>1</sup>

This growing interest in AI is part of a much broader drive to advance AI capabilities as quickly and widely as possible, reflecting Russian president Vladimir Putin's statement that whoever becomes the leader in this field will rule the world.<sup>2</sup> Indeed, research into the field of AI has grown rapidly, outlining the good AI can play in citizens' lives, especially in areas of government and bureaucratic efficiency, the provision of health care, well-being initiatives and education, tackling climate challenges or providing greater national security.<sup>3</sup> The US' 2018 *National Defense*

<sup>1</sup> National Security Commission on Artificial Intelligence, *Final report*, 2023, <https://reports.nscai.gov/final-report>, p. 109. (Unless otherwise noted at point of citation, all URLs cited in this article were accessible on 21 Aug. 2024.)

<sup>2</sup> Tom Simonite, 'For superpowers, artificial intelligence fuels new global arms race', *Wired*, 8 Sept. 2017, <https://www.wired.com/story/for-superpowers-artificial-intelligence-fuels-new-global-arms-race>.

<sup>3</sup> John Danaher, *Automation and Utopia: human flourishing in a world without work* (Cambridge, MA: Harvard University Press, 2019); Antti Kauppinen, 'Flourishing and finitude', *Journal of Ethics and Social Philosophy* 8: 2, 2014, pp. 1–6, <https://doi.org/10.26556/jesp.v8i2.163>; Samuel Scheffler, *Why worry about future generations?* (Oxford: Oxford University Press: 2018); Sven Nyholm, *This is technology ethics: an introduction* (Hoboken, NJ: Wiley-Blackwell, 2023); John Tasioulas, 'Artificial intelligence, humanistic ethics', *Dædalus* 151: 2, 2022, pp. 232–43, [https://doi.org/10.1162/daed\\_a\\_01912](https://doi.org/10.1162/daed_a_01912).

Strategy identified AI as one of the key technologies that will ‘ensure [the US] will be able to fight and win the wars of the future’.<sup>4</sup> In a dedicated AI strategy document, the Department of Defense directed itself ‘to accelerate the adoption of AI and the creation of a force fit for our time’,<sup>5</sup> while the administration of President Joe Biden has warned against the growing assertiveness of China and Russia in the AI space.<sup>6</sup> In response, the US intelligence sector, spearheaded by the Office of the Director of National Intelligence through its Intelligence Advanced Research Projects Activity (IARPA), has prioritized AI by harnessing its ability to process massive and noisy data sources, where there is a large amount of meaningless information, analyse language and forecast major geopolitical trends and societal crises.<sup>7</sup> This commitment is evident in the CIA’s ‘over 100 AI initiatives’ and the Department of Defense’s substantial investments. These reached US\$2.5 billion in 2021, with the military already integrating AI into live combat via drone programmes.<sup>8</sup> Similarly, China and Russia are making significant investments in AI, with China setting itself the goal of becoming a world leader in AI by 2023, and Russia placing significant focus on military AI, with the emphasis on robotics.<sup>9</sup> In the United Kingdom, the Government Communications Headquarters has set out its aims for using AI to improve its intelligence provision.<sup>10</sup>

However, this boom in AI technology has also sparked some concern over its tendency to inherit biases from its training as well as over its inaccurate technology, which leads to misrepresentation and unfair results, with a lack of sufficient transparency, oversight and reliability.<sup>11</sup> Work on the governance of AI has resulted in the development of a series of ad hoc and patchwork international agreements,

<sup>4</sup> US Department of Defense, *Summary of the 2018 National Defense Strategy of the United States of America: sharpening the American military’s competitive edge*, 2018, <https://dod.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf>, p. 3.

<sup>5</sup> US Department of Defense, *Summary of the 2018 Department of Defense Artificial Intelligence Strategy: harnessing AI to advance our security and prosperity*, 2018, <https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/summary-of-dod-ai-strategy.pdf>, p. 4.

<sup>6</sup> Joseph R. Biden, Jr, *Interim national security strategic guidance* (Washington DC: The White House, 2021), <https://www.whitehouse.gov/wp-content/uploads/2021/03/NSC-1v2.pdf>, p. 8.

<sup>7</sup> Office of the Director of National Intelligence, ‘IARPA’, <https://www.iarpa.gov/>.

<sup>8</sup> Brandi Vincent, ‘How the CIA is working to ethically deploy artificial intelligence’, *NextGov*, 31 May 2019, <https://www.nextgov.com/artificial-intelligence/2019/05/how-cia-working-ethically-deploy-artificial-intelligence/157395>;

Kelley M. Sayler, *Artificial intelligence and national security* (Washington DC: Congressional Research Service, 2020), p. 2; Marcus Weisgerber, ‘The Pentagon’s new algorithmic warfare cell gets its first mission: hunt ISIS’, *Defense One*, 14 May 2017, <https://www.defenseone.com/technology/2017/05/pentagons-new-algorithmic-warfare-cell-gets-its-first-mission-hunt-isis/137833>.

<sup>9</sup> Sayler, *Artificial intelligence and national security*, p. 25.

<sup>10</sup> Government Communications Headquarters, *Pioneering a new national security: the ethics of artificial intelligence*, 2021, <https://www.gchq.gov.uk/files/GCHQAIpaper.pdf>; UK Government, *National AI strategy*, 2021, <https://www.gov.uk/government/publications/national-ai-strategy>.

<sup>11</sup> See Karolina La Fors, Bart Custers and Esther Keymolen, ‘Reassessing values for emerging big data technologies: integrating design-based and application-based approaches’, *Ethics and Information Technology*, vol. 21, 2019, pp. 209–26, <https://doi.org/10.1007/s10676-019-09503-4>; Yuval Noah Harari, ‘Reboot for the AI revolution’, *Nature*, vol. 550, 2017, pp. 324–7, <https://doi.org/10.1038/550324a>; U. Pagallo, ‘Cracking down on autonomy: three challenges to design in IT law’, *Ethics and Information Technology*, vol. 14, 2012, pp. 319–28, <https://doi.org/10.1007/s10676-012-9295-9>; Marijn Sax, ‘Big data: finders keepers, losers weepers?’, *Ethics and Information Technology*, vol. 18, 2016, pp. 25–31, <https://doi.org/10.1007/s10676-016-9394-0>; Bart W. Schermer, Bart Custers and Simone van der Hof, ‘The crisis of consent: how stronger legal protection may lead to weaker consent in data protection’, *Ethics and Information Technology*, vol. 16, 2014, pp. 171–82, <https://doi.org/10.1007/s10676-014-9343-8>; James Zou and Londa Schiebinger, ‘AI can be sexist and racist—it’s time to make it fair’, *Nature* 559: 7714, 2018, pp. 324–6, <https://doi.org/10.1038/d41586-018-05707-8>.

domestic policies and guidelines. Internationally, this has included the United Nations' regulation of autonomous weapons in 2014, as well as the adoption by the OECD in 2019 of AI ethical principles, UNESCO's 2021 Recommendation on the Ethics of Artificial Intelligence, the G7's Hiroshima AI Process and the Council of Europe's work towards a legally binding international convention on AI and human rights.<sup>12</sup> At the domestic level, a variety of guidelines have been developed that argue for the need to embed key ethical principles such as transparency, accountability, respect for human dignity, freedom of the individual, equality and unbiased analysis and application, and privacy and data governance into AI design and implementation.<sup>13</sup>

The difference with what I shall term 'intelligence–AI', however, is that in combination, intelligence and AI exacerbate each of the initial problems while simultaneously undermining the currently proposed solutions, creating new ethical dilemmas and potential harms to citizens. For instance, state intelligence actors have the legal authority, physical infrastructures and institutional drive to collect data, to a greater extent than any private AI company, and when these collection mechanisms are combined with AI it enables new datasets that would otherwise be unmanageable or too disconnected to be of use. This widens the reach of data collection beyond that of any private actor and creates an exponential threat to privacy. The inherently secretive nature of intelligence further exacerbates this impact by preventing many of the proposed oversight and transparency mechanisms, meaning that populations are unaware—and unable to challenge—how their information is collected or used. It also prevents both critical internal reflection of the AI's calculations and external review. Unlike private actors, intelligence organizations using AI possess the coercive power of the state, representing an important symbolic and actual threat to people's human rights. In combination, therefore, intelligence–AI raises new and unique harms to both individuals and society that are not yet fully detailed in other works.

In order to fully explore the ethical implications of intelligence–AI, this article will examine how emerging technology has been used across the collection, processing and analysis phases of intelligence.<sup>14</sup> Despite the inherently secretive nature of the world's leading intelligence organizations, they have publicly

<sup>12</sup> See Huw Roberts et al., 'Global AI governance: barriers and pathways forward', *International Affairs* 100: 3, 2024, pp. 1275–86, <https://doi.org/10.1093/ia/iaae073>; European Commission, *Ethics guidelines for trustworthy AI*, 2019, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

<sup>13</sup> See Government Communications Headquarters, *Pioneering a new national security*; Anna Jobin, Marcello Lenca and Effy Vayena, 'The global landscape of AI ethics guidelines', *Nature Machine Intelligence* 1: 2, 2019, pp. 389–99, <https://doi.org/10.1038/s42256-019-0088-2>; Jeffrey S. Saltz and Neil Dewar, 'Data science ethical considerations: a systematic literature review and proposed project framework', *Ethics and Information Technology* 21: 5, 2019, pp. 197–208, <https://doi.org/10.1007/s10676-019-09502-5>; Luc Steels and Ramon Lopez de Mantaras, 'The Barcelona declaration for the proper development and usage of artificial intelligence in Europe', *AI Communications*, vol. 31, 2018, pp. 485–94, <https://doi.org/10.3233/AIC-180607>; The Japanese Society for Artificial Intelligence, *The Japanese Society for Artificial Intelligence Ethical Guidelines*, 2017, <https://www.ai-gakkai.or.jp/ai-elsi/wp-content/uploads/sites/19/2017/05/JSAI-Ethical-Guidelines-1.pdf>; US Department of Defense, 'DoD adopts ethical principles for AI', 24 Feb. 2020, <https://www.defense.gov/News/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence>.

<sup>14</sup> Damien Van Puyvelde, Stephen Coulthart and M. Shahriar Hossain, 'Beyond the buzzword: big data and national security decision-making', *International Affairs* 93: 6, 2017, pp. 1397–416, <https://doi.org/10.1093/ia/iix184>.

reported data on intelligence–AI, most notably in the US and the UK, as a result of their reliance on academic and industrial leaders for co-developing AI technology. Mapping the use of intelligence–AI is possible through the examination of public documentation by private companies, research institutions and state security actors, with examples often shared via media outlets as well as across academic writings.<sup>15</sup> In this article, I will first argue that open-source does not necessarily mean the same as ethical, as the AI collection *en masse* of people’s social media information violates their privacy and undermines their consent and their autonomy. Second, in terms of the processing phase, I will argue that AI-aided categorization is overly reductive and perpetuates harmful social binaries, while also revealing new private information beyond what people have intentionally shared. Finally, I will argue that the inherently secretive environment of intelligence prevents critical insight or reflection, further embedding the social divides AI necessarily creates, while also promoting intelligence practices that, through the coercive power of the state, cause unequal harms across society.

## Intelligence–AI

AI is an umbrella term that refers to a variety of computer-aided processes and data analytics that mimic human thought processes or are tasked with activities that usually require human intelligence and insight. This includes machine learning, a subset of AI that involves training algorithms to learn patterns to make predictions from data without explicit programming, ‘to accomplish a task without being explicitly programmed to do so (i.e. learn on its own)’.<sup>16</sup> ‘Deep learning’ is a specialized subset of this process, ‘inspired by ... the human brain’, that involves layers of analysis conceived as neural networks that are capable of automatically learning representations from data via trial and error at multiple levels of abstraction.<sup>17</sup> This can involve ‘supervised learning’ using known and labelled data to train the algorithms; ‘unsupervised learning’ which identifies patterns in unlabelled data to uncover insights or group similar data points; or a hybrid approach to combine different AI techniques, including supervised and unsupervised learning, along with rule-based systems, to solve complex problems by integrating diverse learning and decision-making strategies.<sup>18</sup>

For intelligence actors, this promises that they will be able to take truly massive quantities and varieties of data and through descriptive, diagnostic, predictive and prescriptive analysis determine what has happened, why it happened, what will happen next and what the response should be, detecting and forecasting new

<sup>15</sup> Ronja Kniep, ‘Another layer of opacity: how spies use AI and why we should talk about it’, *About: Intel*, 20 Dec. 2019, <https://aboutintel.eu/how-spies-use-ai>; Max Tegmark, *Life 3.0: being human in the age of artificial intelligence* (New York: Alfred A. Knopf, 2017); Mark Coeckelbergh, *AI ethics* (Cambridge, MA: MIT Press, 2020).

<sup>16</sup> Matthew Ivey, ‘The ethical midfield in artificial intelligence: practical reflections for national security lawyers’, *Georgetown Journal of Legal Ethics* 33: 109, 2020, pp. 109–38 at p. 114, <https://www.law.georgetown.edu/legal-ethics-journal/wp-content/uploads/sites/24/2020/01/GT-GJLE190067.pdf>.

<sup>17</sup> Ivey, ‘The ethical midfield in artificial intelligence’, p. 115.

<sup>18</sup> Government Communications Headquarters, *Pioneering a new national security*, p. 12.

threats of societal crises, disease outbreaks, cyber attacks or major geopolitical trends.<sup>19</sup> For example, in the US the National Security Agency uses machine learning on open and closed signals intelligence, analysing language and speech to provide insights into the decision-making of individuals, groups and organizations.<sup>20</sup> Meanwhile, the National Geospatial-Intelligence Agency and the US military use AI on satellite images to aid in object classification and to identify potential threats and appropriate targets,<sup>21</sup> and in the social sphere, 'imagery analysis, biometric technologies (such as face, voice, and gait recognition), natural language processing [hereafter NLP], and algorithmic search and query functions' mean that intelligence actors can analyse the interconnections of social life of both the individual and society in ways that were not previously possible.<sup>22</sup> This includes gathering and analysing social media data to identify, locate and track individuals or groups of concern and predict their behaviour. For example, X (formerly Twitter), Instagram, YouTube, Facebook and various other online platforms are now realistic options for data collection, as AI can analyse the billions of data points generated daily around the world. This can include using NLP, a developing area of AI that enables machines to understand, interpret and generate meaning from human language sources.<sup>23</sup> What is key is that the AI develops its own route to determining the correct answer. It can be given correct inputs and marked on successful outputs, but the route through the layers and algorithms used is decided by the AI, the complexity of which often means that this process is beyond the developers' or users' ability to track or necessarily understand.

### **Collection: open-source does not mean ethical**

One practised assumption of intelligence–AI is that the acquisition of open-source data is ethically unproblematic. Many intelligence–AI programmes use the *en masse* collection of open-source data across a range of different sources without any indication of a limit, whether at the collection, processing or analysis stages. This involves the continuous, automated analysis of publicly available data in order to anticipate and/or detect significant societal events such as political unrest, humanitarian disasters and disease outbreaks, and includes gathering information from social media platforms, official government data and online data from websites, blogs and vlogs.<sup>24</sup> It does not involve the collection of information from targeted individuals where there is some indication that they represent a threat. Rather, this is the practice of routinely, pervasively and expansively collecting informa-

<sup>19</sup> Van Puyvelde et al., 'Beyond the buzzword'.

<sup>20</sup> Alexa O'Brien, 'The power and pitfalls of AI for US intelligence', *Wired*, 21 June 2022, <https://www.wired.com/story/ai-machine-learning-us-intelligence-community>.

<sup>21</sup> Brian Seamus Haney, 'Applied artificial intelligence in modern warfare and national security policy', *Hastings Science and Technology Law Journal* 11: 1, 2020, pp. 61–97 at p. 65.

<sup>22</sup> National Security Commission on Artificial Intelligence, *Final report*, p. 109.

<sup>23</sup> See Office of the Director of National Intelligence (IARPA), 'Better', 2018, <https://www.iarpa.gov/research-programs/better>. Also see the Bias Effects and Notable Generative AI Limitations (Bengal) and Rapid Explanation, Analysis and Sourcing Online (Reason) programmes: IARPA, 'Bengal', 2023, <https://www.iarpa.gov/research-programs/bengal>; and IARPA 'Reason', 2023, <https://www.iarpa.gov/research-programs/reason>.

<sup>24</sup> IARPA, 'Mercury', 2015, <https://www.iarpa.gov/research-programs/mercury>.

tion from sources often considered as ‘open’, as there is no direct restriction on who can access the information.

One example that uses this type of data is the Mercury programme, which collects both open and secret information ‘to anticipate and/or detect significant events, including military and terrorist activities, political crises, and disease outbreaks’.<sup>25</sup> Another is the Early Model Based Event Recognition using Surrogates (EMBERS) programme, which gathers information from a variety of sources, including ‘Twitter’s public API . . . , RSS news and blog feeds, Talkwalker alerts, NASA satellite meteorological data, Google Flu trends, Bloomberg financial news, TOR usage data, OpenTable’s restaurants’ cancellation data, the PAHO health survey, and webpages [referencing] tweets’ in order to ‘forecast population-level changes’.<sup>26</sup> EMBERS can mine ‘up to 2,000 messages a second and purchases open-source data such as Twitter’s “firehose”, which streams hundreds of millions of real-time tweets a day’.<sup>27</sup>

Other programmes use ‘image scraping’, scanning social media platforms such as Instagram and Facebook to collect people’s physical images, alongside other identifying data and posted activity, to aid in facial-recognition operations and surveillance. This can include ‘face-matching’, which matches two or more faceprints to determine if they are the same person, using photographs of unknown people and linking them to their real identities gathered from a database. ‘Face-tracking’ involves following and identifying individuals via CCTV as they move through various spaces, for example, security services ‘might use [the] technology to follow an unidentified protester from a rally to their home or car, and then identify them with an address or license plate database’.<sup>28</sup> ‘Face analysis’ uses facial recognition ‘to try to guess a person’s demographic traits, emotional state, and more, based on their facial features’, with some providers claiming to be able to ‘assign demographic attributes to their targets, including gender, race, ethnicity, sexual orientation, and age’. ‘Emotion analysis’ is similar and is claimed by some providers to be able to ‘determine how a person is feeling based on their face’.<sup>29</sup> One prominent provider in this space is Clearview AI, which offers a system that allows law-enforcement agencies to upload a photograph of a face and match it in a database of billions of images it has collected. It then provides links to where matching images appear online. Clearview AI has allegedly ‘harvested more than 20 billion images from the internet and social media to create its global facial recognition database’,<sup>30</sup> including from platforms such as Facebook, taken without

<sup>25</sup> IARPA, ‘Mercury’.

<sup>26</sup> Naren Ramakrishnan et al., “Beating the news” with EMBERS: forecasting civil unrest using open source indicators’, in *Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining*, 2014, pp. 1799–808, <https://doi.org/10.1145/2623330.2623373>.

<sup>27</sup> Leah McGrath Goodman, ‘The EMBERS project can predict the future with Twitter’, *Newsweek*, 7 May 2015, <https://www.newsweek.com/2015/03/20/embers-project-can-predict-future-twitter-312063.html>.

<sup>28</sup> Bennett Cyphers, Adam Schwartz and Nathan Sheard, ‘Face recognition isn’t just face identification and verification: it’s also photo clustering, race analysis, real-time tracking, and more’, Electronic Frontier Foundation, 7 Oct. 2021, <https://www.eff.org/deeplinks/2021/10/face-recognition-isnt-just-face-identification-and-verification>.

<sup>29</sup> Cyphers et al., ‘Face recognition isn’t just face identification and verification’.

<sup>30</sup> Robert Hart, ‘Clearview AI fined \$9.4 million in the UK for illegal facial recognition database’, *Forbes*, 23 May

user permission. Its customers include, in the UK, the Metropolitan Police, the Ministry of Defence and the National Crime Agency, and in the US, the Immigration and Customs Enforcement, the Department of Justice, the FBI and the Miami Police Department, which has confirmed that it uses Clearview AI across all crime types for face-matching purposes.<sup>31</sup> Clearview AI is considered one of the most powerful and accurate facial-recognition tools in the world.<sup>32</sup>

However, the fact that a platform is open-source does not automatically mean it is ethical to collect, process and analyse the information, as doing so can still violate people's privacy by using their information beyond the limits of how they have intended it to be used and without obtaining informed consent. According to one view, privacy can be conceptualized as existing as a collection of different property rights that people have in relation to themselves, information about themselves and information that is generated by their authorship or activity.<sup>33</sup> This cluster of rights includes positive rights, such as the right to sell or give the information away, as well as negative rights that prevent others from using, viewing, selling or damaging the information.<sup>34</sup> Even if looking at or using a person's information does not materially damage it or the individual, it still violates their right of control and privacy.<sup>35</sup> Moreover, these are strong rights of control since while 'we have fairly stringent rights over our property, we have very much more stringent rights over our own persons', which places a greater set of protections over information either about an individual or created by that individual.<sup>36</sup>

For individuals who put their information online, this is still their information to control, including who has access to it and how it is used. It is an individual's right to allow some people to see or even use it. They do not necessarily waive all their control over that information by putting it on social media, especially in terms of intelligence collection and use. For an individual to waive such protective rights, they must consent to that exchange: they must willingly, knowingly and, without coercion, agree. Consent is 'morally important because it expresses an agent's autonomous will'.<sup>37</sup> For an individual to truly consent, they must therefore have their autonomy intact, meaning that consent is both a manifestation and a reflection of that autonomy. The individual should therefore be able to direct their own will by basing their decisions on all the relevant information

2022, <https://www.forbes.com/sites/roberthart/2022/05/23/clearview-ai-fined-94-million-in-uk-for-illegal-facial-recognition-database>.

<sup>31</sup> Kim Lyons, 'Clearview AI's client list includes 2,200 organizations spanning law enforcement to universities', *The Verge*, 27 Feb. 2020, <https://www.theverge.com/2020/2/27/21156678/clearview-ai-client-macy-fbi-doj-twitter-facebook-youtube>.

<sup>32</sup> James Clayton and Ben Derico, 'Clearview AI used nearly 1m times by US police, it tells the BBC', BBC News, 27 March 2023, <https://www.bbc.co.uk/news/technology-65057011>.

<sup>33</sup> Judith Jarvis Thomson, 'The right to privacy', *Philosophy and Public Affairs* 4: 4, 1975, pp. 295–314 at pp. 298–303.

<sup>34</sup> Thomson, 'The right to privacy', p. 299.

<sup>35</sup> Thomson, 'The right to privacy', p. 301.

<sup>36</sup> Thomson, 'The right to privacy', p. 303. For authorship, see James Boyle, *Shamans, software and spleens: law and the construction of the information society* (Cambridge, MA: Harvard University Press, 1997), p. 54; Jerry Kang, 'Information privacy in cyberspace transactions', *Stanford Law Review* 50: 4, 1998, pp. 1193–294 at p. 1207, <https://doi.org/10.2307/1229286>.

<sup>37</sup> Roseanna Sommers, 'Commonsense consent', *Yale Law Journal* 129: 8, 2020, pp. 2232–605 at pp. 2232–5. See also Michael B. Gill, 'Presumed consent, autonomy, and organ donation', *Journal of Medicine and Philosophy* 29: 1, 2004, pp. 37–59, <https://doi.org/10.1076/jmep.29.1.37.30412>.

and by possessing the capacity and opportunity for self-reflection, with an understanding of the implications of their decisions.<sup>38</sup> The individual should have all the relevant knowledge available to them and should reflect on that knowledge, with the necessary capacity and the opportunity to (withhold) consent and be free from coercion. This means that people must have a realistic understanding of how their information is going to be used and of the implications (and harm) that such use can represent for themselves and others. If all these criteria are not satisfied, the individual has not waived their normal protective rights and collecting, processing or analysing the information is a violation of both their privacy and autonomy.<sup>39</sup>

Importantly, it can be argued that the consent required for intelligence agencies to use such information should be *informed* rather than *implied* consent, given the potential for harm that intelligence can cause, whether to the targeted individual, to other individuals or to their social groups. It can be argued that, given the coercive power of the state and the threat to people's privacy, autonomy, liberty, and even physical and mental integrity, security and intelligence is an area of activity that carries with it a higher degree of threat and risk to people's well-being, not only to those targeted directly but to wider social groups as well.<sup>40</sup> Moreover, intelligence–AI represents an even greater threat to social groups, as it has been demonstrated across several studies and investigations that in practice these AI algorithms have resulted in more people of colour being arrested, jailed or physically harmed and that the same algorithms encourage intelligence–AI to become biased, discriminatory and promote unequal treatment.<sup>41</sup>

Harm is therefore caused not only by violating the privacy and autonomy of those whose data is collected, but also by the repercussions of the problematic analysis and intelligence practices promoted in terms of how they are used against others. The individual should be explicitly aware of how their information is used rather than assuming they are aware, making their consent akin to the types of consent seen in the fields of medicine and research.<sup>42</sup> The expectation of informed consent means that individuals are not only notified, but that there is an indication that they have understood and reflected on the request made, and on the implications or consequences of their consent.<sup>43</sup> Moreover, there should be an alignment between the underlying intent of the consent and the resulting action: 'if A

<sup>38</sup> See Harry G. Frankfurt, 'Freedom of the will and the concept of the person', *Journal of Philosophy* 68: 1, 1971, pp. 5–20 at p. 7; Barbara Herman, *The practice of moral judgment* (Cambridge, MA: Harvard University Press, 1996), p. 228; Andrew E. Monroe and Bertram F. Malle, 'Free will without metaphysics', in Alfred Mele, ed., *Surrounding free will* (New York: Oxford University Press, 2014), pp. 25–48; Martha C. Nussbaum, *Women and human development: the capabilities approach* (Cambridge, UK: Cambridge University Press, 2000), p. 79.

<sup>39</sup> Sommers, 'Commonsense consent', p. 2236.

<sup>40</sup> Ross W. Bellaby, *The ethics of intelligence: a new framework* (Abingdon and New York: Routledge, 2014).

<sup>41</sup> Andrew D. Selbst, 'Disparate impact in big data policing', *Georgia Law Review* 52: 1, 2017, pp. 109–96, <https://georgialawreview.org/article/3373-disparate-impact-in-big-data-policing>.

<sup>42</sup> Peter H. Schuck, 'Rethinking informed consent', *The Yale Law Journal* 103: 4, 1994, pp. 899–959.

<sup>43</sup> Henriette Rau et al., 'The generic informed consent service gICS: implementation and benefits of a modular consent software tool to master the challenge of electronic consent management in research', *Journal of Translational Medicine*, vol. 18, 2020, pp. 287–99, <https://doi.org/10.1186/s12967-020-02457-y>; Kusa Kumar Shaha, Ambika Prasad Patra and Siddhartha Das, 'The importance of informed consent in medicine', *Scholars Journal of Applied Medical Sciences* 1: 5, 2013, pp. 455–63, <https://doi.org/10.36347/sjams.2013.vor105.0025>; O. O'Neill, 'Some limits of informed consent', *Journal of Medical Ethics* 29: 1, 2003, pp. 4–7, <https://doi.org/10.1136/jme.29.1.4>.

consents to B's using his car and (without A's knowing it) B uses the car to carry out a bank robbery, it would ordinarily be misleading to say that A consented to B's use of the getaway vehicle'.<sup>44</sup> While A could have asked whether B was intending to use the vehicle for an illegal activity, the question is: how reasonable is it for A to expect to enquire as to the use of their item, especially as B is the one seeking the consent?

Across public social media pages, people see access to their data as being closer to a walled garden rather than an open field, visible to one's peers rather than society as a whole.<sup>45</sup> There is no expectation that the wider world should be able to access their online data, with research showing a particularly strong aversion to authority figures having access. Indeed, a 2018 Pew Research Center review of people's attitudes towards online privacy in the US revealed that '81% say they feel very or somewhat concerned with how companies use the data they collect about them. Fully 71% say the same regarding the government's use of data'. With 79 per cent of people not feeling in control of their data collected by the government. 77 per cent report that they do not understand what data is collected by the government. And '72% of Americans say they have little to no understanding about the laws and regulations that are currently in place to protect their data privacy'. In terms of social media specifically, 76 per cent have little or no trust in social media companies to use their data ethically and not sell it; and 70 per cent say they have little to no trust in companies to make responsible decisions about how they use AI in their products.<sup>46</sup> Indeed, in terms of online social media, even though there should be a greater awareness that others could access such information given its outward-looking nature, there is a discrepancy between the level of privacy people expect and the number of people who have access to their information.<sup>47</sup>

<sup>44</sup> John Kleinig, 'The nature of consent', in Franklin Miller and Alan Wertheimer, eds, *The ethics of consent: theory and practice* (Oxford: Oxford University Press, 2009), p. 18.

<sup>45</sup> See Ross W. Bellaby 'Going dark: anonymising technology in cyberspace', *Ethics and Information Technology*, vol. 20, 2018, pp. 189–204 at p. 197, <https://doi.org/10.1007/s10676-018-9458-4>; Cliff Lampe, Nicole B. Ellison and Charles W. Steinfield, 'Changes in use and perception of Facebook', *Proceedings of the ACM 2008 conference on computer supported cooperative work*, 2008, p. 729, <https://doi.org/10.1145/1460563.1460675>; Sonia Livingstone, 'Taking risky opportunities in youthful content creation: teenagers' use of social networking sites for intimacy, privacy and self-expression', *New Media & Society* 10: 3, 2008, pp. 393–411 at p. 396, <https://doi.org/10.1177/1461444808089415>.

<sup>46</sup> Colleen McClain et al., 'Views of data privacy risks, personal data and digital privacy laws', Pew Research Center, 18 Oct. 2023, <https://www.pewresearch.org/internet/2023/10/18/views-of-data-privacy-risks-personal-data-and-digital-privacy-laws>. Also see Kuo-Cheng Chung et al., 'Social media privacy management strategies: a SEM analysis of user privacy behaviors', *Computer Communications*, vol. 174, 2021, pp. 122–30, <https://doi.org/10.1016/j.comcom.2021.04.012>; Hyunjin Kang, Wonsun Shin and Junru Huang, 'Teens' privacy management on video-sharing social media: the roles of perceived privacy risk and parental mediation', *Internet Research* 32: 1, 2022, pp. 312–34, <https://www.doi.org/10.1108/INTR-01-2021-0005>; Alex Koohang et al., 'Social media privacy concerns, security concerns, trust, and awareness: empirical validation of an instrument', *Issues in Information Systems* 22: 2, 2021, pp. 133–45, [https://doi.org/10.48009/2\\_iis\\_2021\\_136-149](https://doi.org/10.48009/2_iis_2021_136-149).

<sup>47</sup> Nadine Barrett-Maitland and Jenice Lynch, 'Social media, ethics and the privacy paradox', in Christos Kalloniatis and Carlos M. Travieso-Gonzalez, eds, *Security and privacy from a legal, ethical, and technical perspective* (London: IntechOpen, 2020), pp. 49–62. Also see Zeynep Tufekci, 'Can you see me now? Audience and disclosure regulation in online social network sites', *Bulletin of Science, Technology & Society* 28: 1, 2008, pp. 20–36, <https://doi.org/10.1177/0270467607311484>.

## Processing and analysis

In addition to ethical concerns raised by the collection of the data, intelligence–AI creates further issues through its processing and analysis of that data. As two distinct but intertwined parts of the intelligence cycle, the processing is the means through which the information is categorized, labelled, aggregated and presented, while the analysis phase is how that information is understood and given meaning. Machine learning enables ‘the interpretation and analysis of otherwise inaccessible patterns in large amounts of data’ and ‘may involve filtering, analysis of relationships between entities, or . . . sophisticated image- or voice-recognition’.<sup>48</sup> Machine learning can also build its own models based on the data provided, and then use these models to make inferences, allowing machines to determine for themselves what to look for in the data and ‘to learn and improve from their experience automatically without further programming’.<sup>49</sup> AI can carry out various types of analysis, including descriptive analytics—determining what happened in the past; diagnostic analytics—why it happened; predictive analytics—what will happen in the future; and prescriptive analytics to outline the optimal course of action.<sup>50</sup> Social network analysis examines how actors, entities or phenomena interrelate with each other and draws out meaning from those linkages; it can be used to draw maps of urban gangs, city-wide alert systems and crime-spot predictions.<sup>51</sup> Text analysis is used to identify patterns in texts and recompose them into a useful summary, while NLP can draw out the underlying sentiment of the text and its author.<sup>52</sup> For intelligence actors, given the exponential rate at which data is generated across the globe and cyberspace, AI offers a means of processing and analysing those large databases, with intelligence organizations citing the ability to leverage ‘artificial intelligence, automation and augmentation technologies to amplify the effectiveness of our workforce’ and as a result ‘advance mission capability’.<sup>53</sup>

## Cross-referencing and secret analysis

The access that intelligence actors have to citizens’ data is particularly unique and powerful. State security actors are able to gain access to information stores across an individual’s most personal fields, including their home life, physical location,

<sup>48</sup> Kathleen McKendrick, *Artificial intelligence prediction and counterterrorism*, research paper (London: Royal Institute of International Affairs, 2019), <https://www.chathamhouse.org/sites/default/files/2019-08-07-AICounterterrorism.pdf>, p. 8.

<sup>49</sup> David Quinn and Janice Goldstraw-White, *Artificial intelligence and its applications in security* (Sutton, UK: G4S Academy, 2022), [https://www.g4s.com/en-gb/-/media/g4s/unitedkingdom/indexed-files/files/ai\\_paper\\_july\\_22\\_v1.ashx](https://www.g4s.com/en-gb/-/media/g4s/unitedkingdom/indexed-files/files/ai_paper_july_22_v1.ashx); Brian No, ‘Artificial intelligence and deep learning’ (Reston, VA: CACI, 2022), <https://www.caci.com/artificial-intelligence-deep-learning>.

<sup>50</sup> Iqbal H. Sarker, ‘Multi-aspects AI-based modelling and adversarial learning for cybersecurity intelligence and robustness: a comprehensive overview’, *Security and Privacy* 6: 5, 2023, p. 11, <https://doi.org/10.1002/spy2.295>.

<sup>51</sup> McKendrick, *Artificial intelligence prediction and counterterrorism*.

<sup>52</sup> Adam C and Richard Carter, *Large language models and intelligence analysis* (London: Centre for Emerging Technology and Security, Alan Turing Institute, 2023), [https://cetas.turing.ac.uk/sites/default/files/2023-07/cetas\\_expert\\_analysis\\_-\\_large\\_language\\_models\\_and\\_intelligence\\_analysis.pdf](https://cetas.turing.ac.uk/sites/default/files/2023-07/cetas_expert_analysis_-_large_language_models_and_intelligence_analysis.pdf).

<sup>53</sup> *The AIM initiative: a strategy for augmenting intelligence using machines* (Office of the Director of National Intelligence, 2019), <https://www.dni.gov/files/ODNI/documents/AIM-Strategy.pdf>, p.iii.

finances, medical and health, online activity, purchase history and services such as water, internet and electricity providers. AI has the potential to condense diverse and large datasets automatically into digestible intelligence reports, or provide the intelligence analysts themselves with new searching tools that 'extract knowledge from massive corpora of information [that] relates not just to words but acts and entities, the state of the world, and how they relate to each other'.<sup>54</sup> This processing and analytical power of AI, in conjunction with the reach of the intelligence community, means that it can analyse multiple databases in a quantity and at a processing speed previously unseen because it would have been impossible for either human operators or software.

This ability to meaningfully cross-reference different databases, however, can create additional privacy violations, as the AI can make statements about an individual that would not be knowable from any of the singular datasets. For example, studies have demonstrated that people can be re-identified from anonymous data using postcodes, date of birth and gender with 87 per cent accuracy.<sup>55</sup> This type of cross-referencing processing and analysis can then be used to determine individuals' personal and political features—which would normally be considered to be of a very intimate nature and which were not directly revealed in the singular databases. Such features might include sexual orientation, age, gender and religious or political views.<sup>56</sup> This represents a greater ultimate violation of people's privacy, and so results in harms that are greater than the individual parts. Moreover, this type of analysis has been shown to undermine efforts to protect or anonymize the data collected. As researchers at the Alan Turing Institute have written:

Simply removing primary keys (e.g. name, birthday, postcode, etc) from a database and replacing them with some pseudo-random numbers ... doesn't work in general—due to the many diverse holders of records it's often possible to link data from different sources and infer who the subject is. ... For example, [when] Netflix released a large collection of [viewer information], removing people's names and other identifying details ... researchers were able [to] cross-reference the Netflix data with public review data on IMDB ... and add names back into Netflix's supposedly anonymous database.<sup>57</sup>

Furthermore, not only does this re-identification represent an increased privacy violation, but due to the inherently secretive nature of intelligence citizens are not aware that their information has been accessed—nor are they aware of the type of assumptions that are being made about them. People are not aware of how that information is being used (potentially against them), or if it is being used to cause harm to other people: importantly, they are not able to challenge such usage. This is a particular issue for intelligence–AI, as the inherent 'black box' problem of AI

<sup>54</sup> C and Carter, 'Large language models', p. 7.

<sup>55</sup> Andra Gumbus and Frances Gradzinsky, 'Era of big data: danger of discrimination', *ACM SIGCAS Computers and Society* 45: 3, 2015, pp. 118–25, <https://doi.org/10.1145/2874239.2874256>.

<sup>56</sup> European Commission, *Ethics guidelines for trustworthy AI*.

<sup>57</sup> Jon Crowcroft and Adria Gascón, 'A personal issue: how can we enable personal data to be shared without compromising our privacy?', Alan Turing Institute, 8 May 2018, <https://www.turing.ac.uk/news/personal-issue-how-can-we-enable-personal-data-be-shared-without-compromising-our-privacy>.

is exacerbated by the secret environment of intelligence, which in combination prevent internal and external interrogation, as well as preventing any real opportunity to convey to those affected how their information was used. The black box problem is a term used to describe the inability of humans to understand, in any relatable fashion, how an AI system came to its conclusions. This is because, while humans might set the parameters or input training materials, and can initially verify results, the AI is left to determine its own internal routes for reaching that result, which quickly becomes too complex for humans to understand.

In one study, ‘researchers analyzed a program designed to identify curtains in images and discovered that the AI algorithm first looked for a bed rather than a window, at which point it stopped searching the image’, and where it was ‘later learned that this was because most of the images in the training data set that featured curtains were bedrooms’.<sup>58</sup> The greater the volume of data and the more complex the processing or analysis required, the less knowable the means become. For example, Google’s BERT NLP model consists of 110 million parameters, and even if these were to be examined, they would ‘not yield an understanding of the logic of the model’.<sup>59</sup> So, while analysts can ‘mathematically explain how algorithms optimize their objective functions, the complexity of the algorithms make it nearly impossible to describe the optimization in understandable and intuitive terms’.<sup>60</sup> The use of machine learning and its specialized subset of deep learning, with the latter’s multiple levels of abstraction, increasingly obfuscate the means by which the decision is made, making the understanding impenetrable and necessarily reducing the accountability, understandability and traceability of the process as well as—in the long term—the predictability of the result.

More problematically, the black box problem will further exaggerate and be exaggerated by the distortive effect of the secretive nature of the intelligence field, as the use of AI will embed already existing biases through feedback loops, tech-washing (‘the process by which proponents of the outcomes can defend those outcomes as unbiased because they were derived from “math”’) <sup>61</sup> will ratify the results as scientifically derived by a machine, and groupthink will limit interrogation. This is because secretive environments within security are problematic, creating an internal culture that isolates the intelligence community on the inside from the rest of society on the outside, reinforcing the need for greater isolation and hyper-secrecy. When overly secretive environments or cultures are placed above or outside the normal political sphere, they isolate their members and their structures, separating off those on the outside who are unaware and unable to engage, and preventing them from acting as a counterbalance and reference point to the internal cultures and escalation. Those on the inside are subjected to a

<sup>58</sup> Saylor, *Artificial intelligence and national security*, p. 32.

<sup>59</sup> Anna Knack, Richard J. Carter and Alexander Babuta, *Human-machine teaming in intelligence analysis requirements for developing trust in machine learning systems* (London: Centre for Emerging Technology and Security, Alan Turing Institute, 2022), [https://cetas.turing.ac.uk/sites/default/files/2022-12/cetas\\_research\\_report\\_-\\_hmt\\_and\\_intelligence\\_analysis\\_vfinal.pdf](https://cetas.turing.ac.uk/sites/default/files/2022-12/cetas_research_report_-_hmt_and_intelligence_analysis_vfinal.pdf), p. 8.

<sup>60</sup> Ivey, ‘The ethical midfield in artificial intelligence’, p. 119.

<sup>61</sup> Matthew Guariglia, ‘Police use of artificial intelligence: 2021 in review’, Electronic Frontier Foundation, 1 Jan. 2022, <https://www.eff.org/deeplinks/2021/12/police-use-artificial-intelligence-2021-review>.

process of in-group/out-group differentiation that dehumanizes 'others'. When this is coupled with a lack of outside input, there is no means of measuring one's moral compass.<sup>62</sup> As a result, officers learn to exclude those considered as outsiders from their universe of obligation.<sup>63</sup> Cognitive restructuring means violence or harm is redefined as honourable, for a greater abstract good, and becomes increasingly socially and morally acceptable to those inside.<sup>64</sup> Secretive environments normalize this process while also reinforcing both the need for greater secrecy and a lack of regard for the negative consequences for those on the outside. In such an environment internal criticism is limited, as it is seen as a betrayal to the group, and so restricts alternative analysis as the group mentality smothers dissenting points of view.<sup>65</sup>

Secrecy does play an ethically important role in enabling intelligence organizations to perform their role, and this is not a call for complete transparency. Nor is the resulting behaviour the product of some nefarious intent, but a natural result of the divisions and isolation that the secrecy itself creates. The secretive and isolated culture does not empower the required critical inquiry of a system such as intelligence–AI, which is itself inherently closed off and difficult to unravel. The secretive nature of intelligence means that individuals are unaware of what types of conclusion are being made, and there is no opportunity for individuals to appeal or challenge these assumptions, or the methodology on which the assumptions are based. Those who have suffered harm might not even be aware of how they are being categorized or the conclusions being drawn about either them or their social group. This unequal treatment is, for intelligence–AI, beyond using poor sets of training data, but becomes more deeply rooted as well as more difficult to avoid, detect and remedy because of how AI and secretive intelligence interact.

### *Social sorting*

A key underlying problem with AI is that it relies on digitizing, quantifying and aggregating data in order to turn it into an analysable and usable product. One of the selling points of intelligence–AI is that it can take data from unmanageably large and diverse datasets and create user-friendly means of searching and presenting the information. This process relies, however, on the AI's ability to format the data—whether it consists in biometric details, online activity, video or photographs of people's faces, gait, behaviour or transcripts of text—into a numerical value and to attach this digital output to some socially derived label. Complex social and physical activities are therefore necessarily reduced in order that they can be quantified. This sorting and processing rely on the allocation of labels and then the aggregating, grouping, separating and further labelling of

<sup>62</sup> Albert Bandura, 'Moral disengagement in the perpetration of inhumanities', *Personality and Social Psychology Review* 3: 3, 1999, p. 194, [https://doi.org/10.1207/s15327957pspr0303\\_3](https://doi.org/10.1207/s15327957pspr0303_3).

<sup>63</sup> Helen Fein, *Human rights and wrongs: slavery, terror and genocide* (Boulder, CO: Paradigm Publishers, 2007), p. 11.

<sup>64</sup> Albert Bandura, *Social foundations of thought and action: a social cognitive theory* (Englewood Cliffs, NJ: Prentice Hall, 1986), p. 376.

<sup>65</sup> US Senate Select Committee on Intelligence, *Committee study of the Central Intelligence Agency's detention and interrogation program*, 2012, [https://irp.fas.org/congress/2014\\_rpt/ssci-rdi.pdf](https://irp.fas.org/congress/2014_rpt/ssci-rdi.pdf), p. 2.

social attributes through repeated runs to take the vast quantity of data collected and turn it into something usable.

This is problematic, in the first instance because such digital systems rely on being able to allocate a fixed numerical data point to create something which is analysable. While this process might not necessarily be a problem with discrete or factual data such as addresses, age or some types of biometric information, it becomes more problematic for socially constructed or socially dependent information, most notably in cases involving race, gender, identity, behaviour, culture, attitudes and beliefs. As a process, this necessarily creates categories and divisions within social phenomena that are complex, fluid, context-relevant and contested, representing a modern panoptic sorting process that describes a system of categorizing the population to a hitherto unimagined degree.<sup>66</sup> Simone Browne has argued that through this type of processing the visual human body is classified, and by doing so ‘race, as a constructed category, is defined and made visible’,<sup>67</sup> reducing and classifying people ‘into sets of logical manipulatable signs’, which has been ‘a hallmark of racializing scientific and administrative techniques going back several hundred years’.<sup>68</sup> Marta Maria Maldonado and Adela Licona describe this process of racialization (by categorization/facial recognition) as ‘the production, reproduction of and contest over racial meanings and the social structures in which such meanings become embedded’,<sup>69</sup> articulating a connection between systems of racial oppression and quantification and driven by a logic of conceptualization that is concerned with arbitrarily dividing human populations.<sup>70</sup> ‘The mechanistic gaze of facial recognition’, for example, ‘consists solely of the extraction and abstraction of ... personal features from what are essentially statistical images’.<sup>71</sup>

Second, this reduced data is then allocated to a label, which is an inherently political, predefined understanding of the complex social world carrying with it inherent assumptions, biases and context. These simplified labels are overly reductive—man/woman, white/non-white, threat/non-threat—and serve to perpetuate the divisions or binaries that the labels themselves inhabit, attributing a particular meaning and applying the characteristics to these arbitrary and simplistic divisions.<sup>72</sup> Facial recognition, for example, can involve identifying an individual’s social attribute from their physical appearance, including their race, ethnicity, gender, class, social behaviour, attitude or sexual orientation. This

<sup>66</sup> Oscar H. Gandy, Jr, *The panoptic sort: a political economy of personal information* (Boulder, CO: Westview, 1993).

<sup>67</sup> Simone Browne, *Dark matters: on the surveillance of blackness* (Durham, NC: Duke University Press, 2015), p. 7.

<sup>68</sup> Luke Stark, ‘Facial recognition is the plutonium of AI’, *XRDS: Crossroads, The ACM Magazine for Students* 25: 3, 2019, pp. 50–55 at p. 52, <https://doi.org/10.1145/3313129>.

<sup>69</sup> Marta Maria Maldonado and Adela C. Licona, ‘Re-thinking integration as reciprocal and spatialized process’, *Journal of Latino and Latin American Studies* 2: 4, 2007, p. 129.

<sup>70</sup> Kimberlé Crenshaw, ‘Mapping the margins: intersectionality, identity politics, and violence against women of color’, *Stanford Law Review* 43: 6, 1991, pp. 1241–99, <https://doi.org/10.2307/1229039>; Michel Foucault, *Security, territory, population: lectures at the Collège de France 1977–1978* (New York: Picador, 2009), pp. 239–63; Achille Mbembe, *On the postcolony* (Berkeley, CA: University of California Press, 2001).

<sup>71</sup> Mark Andrejevic and Neil Selwyn, ‘Facial recognition technology in schools: critical questions and concerns’, *Learning, Media and Technology* 45: 2, 2020, pp. 115–28 at p. 121, <https://doi.org/10.1080/17439884.2020.1686014>.

<sup>72</sup> Stark, ‘Facial recognition is the plutonium of AI’.

processing is then made more problematic as the intelligence-AI then attributes a particular meaning to those classifications. It is not only that the target is quantified under some predefined label, but that this is given status or meaning through its connection to other data points and how they are themselves categorized by the AI, which is then used and exported onto another individual in other scenarios. For example, 'facial recognition involves identifying, extracting, and selecting contrasting patterns in an image, and then classifying and comparing them to a previously compiled database of other patterns'.<sup>73</sup> Those who fall under this label are then treated as the label dictates, rather than as the individuals they are. As Spiros Simitis argues, a profiled individual is 'necessarily labeled and henceforth seen as a member of a group, the peculiar features of which are assumed to constitute [their] personal characteristics'.<sup>74</sup> This practice of intelligence-AI ends up 'conflating biological characteristics with social attributes' as it formalizes 'phenotypic differences' with the tendency 'seemingly irresistibly, to spill over into claims about genetic capabilities and aptitudes'.<sup>75</sup> Furthermore, labeling people in this way can cause a tendency to create self-fulfilling prophecies, when individuals feel they must act to meet certain expectations of them because of how they are treated.<sup>76</sup>

### **Application: the coercive state**

This intelligence-AI processing, analysis and attaching of meaning to the data leads to a set of practices, beliefs and positions that encourage, rationalize and exacerbate existing social biases by embedding feedback loops, promoting self-fulfilling prophecies and falling foul of tech-washing. Intelligence-AI shapes security practice, and in doing so manifests the biases seen at the processing and analysis phases, taking the characteristics of one target and overlaying them over their group to identify and classify suspect populations.<sup>77</sup> Singular attributes can be used to create a profiling that locates pre-threats even though other individuals do not have any of the other 'threatening' attributes seen in the original offender.<sup>78</sup> Simply having similar patterns to previous threats is not sufficient to count as a legitimate reason for targeting someone, and is more about guilt by proximity rather than representing some form of threat. This is indicative of a larger move in security towards pre-emptive risk assessment, as security is 'not based on individualised suspicion, but on the probability that an individual might be an offender'.<sup>79</sup>

<sup>73</sup> Stark, 'Facial recognition is the plutonium of AI', p. 53.

<sup>74</sup> Spiros Simitis, 'Reviewing privacy in an information society', *University of Pennsylvania Law Review*, 135: 3, 1987, pp. 707–46 at p. 719.

<sup>75</sup> Andrejevic and Selwyn, 'Facial recognition technology in schools', p. 122.

<sup>76</sup> Robert K. Merton, *Social theory and social structure* [1949] (New York: Free Press, 1968), p. 477.

<sup>77</sup> Lucia Zedner, 'Pre-crime and post criminology', *Theoretical Criminology* 11: 2, 2007, pp. 261–81 at p. 265, <https://doi.org/10.1177/1362480607075851>.

<sup>78</sup> Valeria Ferraris et al., *Defining profiling*, working paper (Protecting citizens' rights fighting illicit profiling, 2013), <https://doi.org/10.2139/ssrn.2366564>, p. 5.

<sup>79</sup> Clive Norris and Michael McCahill, 'CCTV: beyond penal modernism?', *British Journal of Criminology* 46: 1, 2006, pp. 97–118 at p. 98, <https://doi.org/10.1093/bjc/azio47>.

The trend described above is problematic, as it has been reported that the AI algorithms overly predict individuals from certain racial groups, or from particular neighbourhoods, as a 'proxy variable' for race. One investigation demonstrated that the predictive technology used by the policing software PredPol disproportionately predicted threats would be committed in neighbourhoods by working-class people, people of colour and Black people in particular.<sup>80</sup> Moreover, even if the data on which the AI was trained was made less biased, the system would still face a feedback issue where the focus of resources in any given area created a self-fulfilling prophecy, while tech-washing means that the outcomes are assumed correct.<sup>81</sup> Intelligence–AI will therefore not only miss problems from unmonitored or emerging arenas, but will reinforce the idea that a particular group or area is the right target for intelligence surveillance and interventions.

These intelligence–AI decisions and conclusions are then given coercive weight and force, receiving official recognition from the state and its security apparatus. This results in security activity that can represent both a material and a symbolic threat to people's physical well-being, privacy, autonomy and liberty. This can include incorrect, disproportionate, biased or arbitrary application of the direct force of the state, such as the power to arrest and detain, physical attacks, punishment, control, intimidation, subjugation, coercion, agenda-setting, domination, marginalization or the maintenance of existing power dynamics. It can also include symbolic threats, often referred to as the 'chilling' effect of state surveillance, which can alter how people decide to act or limit their options, thus violating their autonomy. If, at the core of autonomy, there is the capacity to make decisions freely, without undue influence or control, then distorting influences both overtly and covertly violate an individual's autonomy. This includes cases where an individual's decision-making process is altered through a self-imposed pressure caused by how people think others are viewing, evaluating and possibly intervening in their life, a force which is increased when it comes from coercive actors such as the state and its security apparatus.<sup>82</sup>

Indeed, given intelligence–AI's reach and ability to transcend space and time limitations and bring the asymmetric gaze of the panopticon across social spaces, both online and in the physical world, it can detrimentally affect an individual's autonomy as they start to 'self-discipline' their actions and surrender to the wishes of the observers as the individual 'becomes the principle of his own subjection'.<sup>83</sup>

<sup>80</sup> Dhruv Mehrotra et al., 'How we determined predictive policing software disproportionately targeted low-income, Black, and Latino neighborhoods', *Gizmodo*, 2 Dec. 2021, <https://gizmodo.com/how-we-determined-predictive-policing-software-dispropo-1848139456>.

<sup>81</sup> Will Douglas Heaven, 'Predictive policing algorithms are racist. They need to be dismantled', *MIT Technology Review*, 17 July 2020, <https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice>.

<sup>82</sup> Herman, *The practice of moral judgement*, p. 228; Frankfurt, 'Freedom of the will and the concept of the person', p. 7.

<sup>83</sup> Michel Foucault, *Discipline and punish: the birth of the prison* [1975] (Harmondsworth: Penguin, 1979), pp. 202–3. Also see Gandy, *The panoptic sort*; David Lyon, *The electronic eye: the rise of surveillance society* [1994] (Cambridge, UK: Polity Press, 2013), p. 65; Michael McCahill, 'Beyond Foucault: towards a contemporary theory of surveillance', in Clive Norris, Jade Moran and Gary Armstrong, eds, *Surveillance, closed-circuit television and social control* (Aldershot: Ashgate, 1998), pp. 41–65.

Even when the impact of intelligence–AI is not felt materially, its increasing use can establish social and political environments that then threaten people's (political) autonomy as a result of the new types of surveillance AI enables, creating direct implications for the individual's (and that individual's social group's) liberty, choices, opportunities and social acceptance. The power of the panoptic gaze means that people do not have to experience direct intervention such as physical detainment; however, the sensation of being monitored is sufficient to influence their decision-making.

Moreover, this can threaten people as politically autonomous agents, as maintaining one's autonomy forms the foundation of a bundle of civil rights including freedoms of expression, association and political engagement. Political autonomy means being able to create one's own ideas and to share, shape and critically reflect on them through access to other individuals and other information collections; to engage with the wider political community; and to use this autonomy to review and check on political authority.<sup>84</sup> Intelligence–AI's panoptic gaze focuses on political spaces, both in cyberspace and the real world, and so its (perceived) surveillance and intervention affect how people decide to act as political agents. This can include monitoring what information people share and consume, how they congregate and travel, and how they carry out political activity such as protesting. For example, Israeli security forces' use of intelligence–AI to scrape data from social media in 2017 resulted in the arrest of 300 Palestinians in the West Bank and East Jerusalem on charges related to posts on Facebook.<sup>85</sup> The intelligence–AI analysis is using social profiling to classify and securitize people's political expression.

## Conclusion

Intelligence–AI has the potential to increase the speed, reach and analytical power of intelligence actors in ways that are still to be fully realized. However, intelligence–AI also needs to be placed under a more critical lens than other forms of AI development. Intelligence has a reach unlike any other organization, and AI now offers a realistic and timely way to process and analyse that information, solving the bottleneck problem and driving ever further a mandate to collect more information. The concern over the processing carried out by AI is not just limited to how accurate the system can be made, as being more accurate does not undo

<sup>84</sup> See Ross W. Bellaby, *The ethics of hacking* (Bristol: Bristol University Press, 2023), pp. 55–62.

<sup>85</sup> Ruth Eglash and Loveday Morris, 'Israel says monitoring social media has cut lone wolf attacks. Palestinians are crying foul', *Washington Post*, 9 July 2019, [https://www.washingtonpost.com/world/middle\\_east/israel-says-that-monitoring-social-media-has-cut-lone-wolf-attacks-palestinians-are-crying-foul/2018/07/08/bfe9ece2-7491-11e8-bda1-18e53a448a14\\_story.html](https://www.washingtonpost.com/world/middle_east/israel-says-that-monitoring-social-media-has-cut-lone-wolf-attacks-palestinians-are-crying-foul/2018/07/08/bfe9ece2-7491-11e8-bda1-18e53a448a14_story.html). Also see Omer Benjakob and Phineas Rueckert, 'Fake friends: leak reveals Israeli firms turning social media into spy tech', *Haaretz*, 28 Feb. 2023, <https://www.haaretz.com/israel-news/security-aviation/2023-02-28/ty-article/fake-friends-leak-reveals-israeli-firms-turning-social-media-into-spy-tech/00000186-7f75-d079-ade7-ff7507970000>; Josef Federman, 'Israel: social media monitoring nabs would-be attackers', *AP News*, 12 June 2018, <https://apnews.com/article/c573e9c93d8544209a52baed19be7984>; Olivia Solon, 'Why did Microsoft fund an Israeli firm that surveils West Bank Palestinians?', *NBC News*, 28 Oct. 2019, <https://www.nbcnews.com/news/all/why-did-microsoft-fund-israeli-firm-surveils-west-bank-palestinians-n1072116>.

the problem of making arbitrary, politically and socially contingent conclusions about people, which become unchallengeable due to the secretive nature of the intelligence community and the impenetrable nature of AI. Conclusions are then given force, as they shape and inform security responses which—without the ability to interrogate its use, can unknowingly distort intelligence policy application, promoting distrust between individuals and the state as well as between different social groups, having real repercussions for individuals in terms of social mobility and treatment. In terms of intelligence, this can result in harsher tactics, including escalating interrogation techniques, increasingly intrusive collection methods or unequal treatment based on race or ethnicity.

This analysis therefore has important implications for how intelligence–AI should be understood and practised. However, the existing guidelines, principles, regulations and agreements are not sufficient to tackle the unique ethical concerns that intelligence–AI creates. Not only are they often vague, abstract and lacking in a usable process to direct ethical behaviour, but in the case of intelligence they are insufficient to tackle the challenges raised by its coercive power, secretive environment and the lack of understanding or consent from the general population. Looking forward, therefore, there is a need to develop a specialized ethical framework that provides practical guidance on how AI technology should be used and how the population should be informed. This includes creating a more explicit set of targeting guidelines on what data should be collected and from whom; the types of analysis allowed, moving away from making determinations based on characteristics such as race, religion, location, ethnicity, gender and physical appearance; and introducing post-bellum mechanisms to explicitly inform people on how their data has been used.