Stato, Chiese
e pluralismo confessionale

VV. AA. (Luca Pietro Vanoni, Andrea Cesarini, Federico Colombo, Alessandro Negri, Tania Pagotto, Greta Pavesi, Giada Ragone)

The spatial ramifications of religion: new and traditional legal challenges

Papers presented at the panel "*The spatial ramifications of religion: new and traditional legal challenges* " held at the Annual Conference 2022 of EuARe (European Academy of Religion), which took place in Bologna from 20th to 23rd June 2022 and was organised by FSCIRE.
Papers selected by the organizing Committee.

Interventi presentati il 20 giugno 2022 al panel *"The spatial ramifications of religion: new and traditional legal challenges"* in occasione della Annual Conference 2022 di EuARe (European Academy of Religion) tenutasi a Bologna dal 20 al 23 giugno 2022 e organizzata da FSCIRE.
Interventi selezionati a cura del Comitato organizzativo.

Stato, Chiese
e pluralismo confessionale

Giada Ragone
(ricercatrice di Diritto costituzionale nell'Università degli Studi di Milano,
Dipartimento di Diritto pubblico italiano e sovranazionale)

Artificial Intelligence and New Scenarios of Religious Discrimination in Virtual and Real Space

SUMMARY: 1. Introduction: does AI discriminate? - 2. AI and religious discrimination: an overlooked but very real relationship - 3. Is non-discrimination law ready for the challenge of AI? - 3.1 AI bias at the bar: the Bridges case - 4. Conclusion.

1 - Introduction: does AI discriminate?

Applications of Artificial Intelligence (AI)[245] are currently diverse and unforeseeable. They increasingly affect people's day-to-day lives and routines in many fields[246]. In the public and private sectors, AI systems, especially those based on machine learning, not only assist human activities, but also make many choices and selections previously carried out by people[247].

---

[245] A tentative definition of AI is provided by the Glossary of the European Ethical Charter on the use of Artificial Intelligence in judicial systems and their environment, p. 69-70. According to this document, AI is a "set of scientific methods, theories and techniques whose aim is to reproduce, by a machine, the cognitive abilities of human beings. Current developments seek to have machines perform complex tasks previously carried out by humans. However, the term artificial intelligence is criticised by experts who distinguish between "strong" AIs (yet able to contextualise specialised and varied problems in a completely autonomous manner) and "weak" or "moderate" AIs (high performance in their field of training). Some experts argue that "strong" AIs would require significant advances in basic research, and not just simple improvements in the performance of existing systems, to be able to model the world as a whole".

[246] As effectively described in the position paper "Religion & Innovation" 2019 by the Centre for Religious Studies at the Bruno Kessler Foundation (available here: *https://isr.fbk.eu/en/about-us/position-paper/*). **J.F. BALKIN**, *The Three Laws of Robotics in the Age of Big Bata*, in *Ohio State Law Journal*, vol. 78, n. 5/2017, pp. 1271-1241 even foresees the rise of an "Algorithmic Society".

[247] On the ethical implications of this substitution, see **A. CELOTTO**, *Come regolare gli algoritmi. Il difficile bilanciamento fra scienza, etica e diritto*, in *Analisi Giuridica dell'Economia*, n. 1/2019, p. 2.

The passage from human to AI decision-making meets the need for efficiency and speed. AI can analyse large datasets, discover patterns in massive amounts of data, and develop profiles that can be used to make decisions about people. However, despite their efficiency and their illusory lack of human attributes[248], AI models are not purely impartial machines. Indeed, it is not unusual for a computer process to show algorithmic bias: a lack of justice that "can be interpreted as one group's prejudice based on a particular categorical distinction"[249]. From a legal perspective, it is undisputed that besides privacy, liability and other concerns[250], a major challenge inherent to AI is the risk of discrimination: i.e. illegal differentiation on the basis of protected characteristics (such as gender, race, sexual orientation and also religion), not justified by a legitimate aim.

Regarding human behaviours, AI-driven decisions can also lead to direct and indirect discrimination[251]: "direct" meaning that people are directly discriminated against on the basis of protected characteristics; "indirect" or "disparate impact" meaning that a practice is neutral at first glance but in the end has a discriminatory effect. The discriminatory effects of AI are debated worldwide, and an Algorithmic Justice League has even been founded at the Massachusetts Institute of Technology to identify, mitigate and highlight algorithmic bias.

The origin of algorithmic discrimination may or may not be voluntary, i.e. intentional[252]. In most cases, the discriminatory effect and bias of AI are products of limited data, insufficient training or poorly written algorithms. The loopholes of AI can depend on where the system

---

[248] **B. PLOMION**, *Does Artificial Intelligence Discriminate?* in *Forbes*, 2nd May 2017. See also **K. CRAWFORD**, *The Hidden Biases in Big Data*, in *Harvard Business Review*, 1st April 2013

[249] **S. NUORI**, *The Role of Bias in Artificial Intelligence*, in *Forbes*, 4th February 2021.

[250] An interesting analysis of the main constitutional concerns of AI is provided by **A. SIMONCINI**, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in *Biolaw Journal*, n. 1/2019, p. 63 et seq. On AI and fundamental rights, *ex multis*: **O. POLLICINO**, "*Getting the Future Right - Artificial Intelligence and Fundamental Rights". A view from the European Union Agency for Fundamental Rights*, in *Biolaw Journal*, n. 1/2021, p. 7 et seq.

[251] Although, it was noted that the AI "has blurred the distinction between direct and indirect discrimination" (**C. NARDOCCI**, *Intelligenza artificiale e discriminazioni*, in *Rivista del Gruppo di Pisa*, n. 3/2021, p. 59).

[252] For a taxonomy of ways in which AI can discriminate, see **C. NARDOCCI**, *Intelligenza artificiale e discriminazioni*, pp. 16-17.

is designed and on the data used to test and refine it. Moreover, if AI learns through observation of human practice driven by ignorance or intolerance, it will reflect these things. Lastly, it should be borne in mind that algorithms do not always behave according to understandable logic[253].

However, there is also the possibility that AI discrimination is intentional: an organization could, for example, use proxies[254] to discriminate on the basis of sex, age, ethnicity or religion, relying on the fact that discrimination through algorithms is difficult to discover[255].

AI discrimination is a potential threat in the real world and in virtual space. Indeed, as Alessandro Negri[256] and Andrea Cesarini[257] show in this Symposium, the problematic realm of internet platforms should not be underestimated[258]. The latter use complex AI systems to control online cyberspace, and do not necessarily observe anti-discrimination law, being "unbound by any state dimension (or control)"[259]. Briefly, the potential discriminatory effects of AI are diverse and unpredictable, as well as its applications.

2 - AI and religious discrimination: an overlooked but very real relationship

[253] See **F. PASQUALE**, *The black box society: The secret algorithms that control money and information*, Harvard University Press, Cambridge-London, 2015.

[254] I.e. "intentional proxy discrimination", see **C. NARDOCCI**, *Intelligenza artificiale e discriminazioni*, p. 28.

[255] This scenario is a negative example of human-machine cooperation, also known as "automation bias". However, there are also positive examples of collaboration between human and artificial intelligence: in particular when humans have the last word in AI-driven decisions, which could be a sort of remedial action against unintentional discrimination.

[256] Focusing on the space for censorship and freedom of art in social media.

[257] Providing a "spatial" analysis of the principle of non-discrimination on the basis of religion.

[258] As noted by **L.P. VANONI** in the *Introduction* to this Symposium, "The new technologies are digitalizing the traditional public square, not only by replacing it with the social networks, but also by reshaping the boundaries of classical institutional places."

[259] **F. BIONDI**, *Intelligenza artificiale: coordinate costituzionali*, in *Diritto e valutazioni scientifiche* edited by B. LIBERALI, L. DEL CORONA, Giappichelli, Torino, 2022, p. 479.

As already mentioned, discrimination through AI may concern different human characteristics. For instance, Recital 71 of EU GDPR[260] provides a varied list of grounds for discrimination that controllers of algorithmic profiling procedures should avoid. The list includes "racial or ethnic origin, political opinion, *religion or beliefs*, trade union membership, genetic or health status or sexual orientation".

Although beliefs are a pivotal aspect of human coexistence, there is no case law on AI and religious bias and most of the literature and studies[261] focus on other grounds of discrimination. Still, this does not mean that AI is necessarily without potential direct or indirect discriminatory effects on the grounds of religion.

As an example of indirect discrimination, let us consider how some AI recruitment systems work. When an AI system has to assess job applications, the employer usually asks it to focus on and detect certain features. By selecting specific features, the employer may introduce bias against certain groups, voluntarily or otherwise. For example, many employers in the USA look for people who studied at certain famous universities (Harvard, Stanford, NYU, Columbia, Berkeley). This among others excludes students who due to their belief or faith, attended religion-oriented universities (Notre Dame for Catholics, BYU for Mormons etc.), which are also excellent universities but less famous. The choice of selecting job applicants according to the university they attended is apparently neutral but can discriminate against certain religious groups.

The same is true if the criterion to select a good employee is proximity to the company. If the company office is in the city centre, poorer people living far from the centre are at a disadvantage. In our societies, since poorer people often have an immigrant background and belong to religious minorities, the proximity criterion could be a multiple discrimination, also on religious grounds.

Focusing on direct discrimination, an interesting case concerned Churchix, a facial recognition software used by churches[262] to track the

---

[260] Regulation EU n. 2016/679 of 27 April 2016.

[261] Including the 2018 report by the Council of Europe on "Discrimination, artificial intelligence and algorithmic decision-making".

For an innovative contribution on AI and religious freedom, see **L.P. VANONI**, Deus ex machina. *Intelligenza artificiale e libertà religiosa nel sistema costituzionale degli Stati Uniti*, in *Stato, Chiese e pluralismo confessionale*, (*https://www.statoechiese.it)* , n. 15/2020, p. 87 et seq.

[262] In 2015, according to the Washington Post (see "*Skipping church? Facial recognition software could be tracking you*" of 14 July 2015), 40 churches all over the world used this

regular attendance of believers, see who is missing and even make security checks. Churchix can tell the church authorities when a person who is not in its face database is attending a religious event. The authorities can therefore quickly deal with a misbehaving visitor.

This tool raises tricky constitutional issues, ranging from ensuring privacy to guaranteeing religious freedom (AI tracking of churchgoer attendance could be seen as compulsion and a violation of believers' freedom). There is the concrete risk of direct discrimination on the basis of religious habits: Churchix labels those who are not members of a specific religious community or congregation as potential threats, persons to monitor and who arouse suspicion.

Let us now consider virtual space. Another area where AI can easily lead to religious bias is that of automatic writing prompts. For example, in a paper published in Nature Machine Intelligence[263], a group of researchers found that large language models - increasingly used in algorithmic applications for mobiles and other devices - associate Muslims with violence and bad behaviour.

The occasions where algorithms could give rise to religious bias are many and underline why scholars should not overlook such grounds for discrimination.

3 - Is non-discrimination law ready for the challenge of AI?

As we know, direct and indirect discrimination on the grounds of certain human characteristic are prohibited in many treaties, fundamental rights charters and constitutions[264]. Non-discrimination law has various tools and extensive case-law that can be interpreted for application to

---

facial recognition system.

[263] **A. ABID, M. FAROOQI, J. ZOU**, *Large language models associate Muslims with violence*, in *Nature Machine Intelligence*, n. 3/2021, p. 461-463.

[264] Considering only the Italian case, Article 3 of the Constitution states that "All citizens have equal social dignity and are equal before the law, without distinction of sex, race, language, religion, political opinion, personal or social condition". In addition, legislative power shall be vested in compliance with constraints deriving from EU legislation and international obligations, that also include non-discrimination provisions. For example, under Article 14 of the ECHR "The enjoyment of the rights and freedoms set forth in [the] Convention shall be secured without discrimination on any ground such as sex, race, colour, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth or other status".

discrimination through AI. Nonetheless, the application of traditional instruments of current discrimination law to algorithmic bias encounters particular issues, three of which are worth mentioning.

First, this branch of law has different sources, often concerning specific fields (employment, education etc.), and since AI is used for many purposes, some grey areas can remain unregulated.

Second, proxies can produce unreasonable differentiations based on "new" features not listed in non-discrimination acts. This means that new groups or minorities can suffer bias or discrimination unforeseen by any law.

Third, regarding the issue of liability, it can be extremely challenging to determine whether the source of AI bias stems from human decision makers or automated AI systems. Judges must try to understand what happens in the black box of the algorithm and assess whether the discriminatory effect is intentional or due to poorly written algorithms. Where the illegal effect stems from a malfunctioning algorithm, it is necessary to establish whether AI users are held to be aware of (and prevent) the discriminatory attitude of the machine. This issue emerges clearly from the first court case on discrimination through a facial recognition system.

3.1 - AI bias at the bar: the Bridges case

The first case ever decided by a court on the discriminatory effect of a facial recognition system was the *Bridges* case in Wales. It concerned use of this instrument by police in public places to identify people for whom warrants had been issued for suspected offences[265]. Wherever it is installed, the system (called "AFR Locate") looks for face matches with a police database of photographs.

A Cardiff citizen, Mr. Bridges, brought suit in the High Court of Justice of Wales, alleging that he was caught on AFR Locate cameras and that this violated the law in several respects. Among the other groups of claims, the appellant asserted that the South Wales Police Force had failed to comply with its obligation under the Equality Act 2010 and the Public Sector Equality Duty (PSED) of 2011. Indeed, the police did not consider the possibility that AFR Locate might produce results that were indirectly

[265] "Between late 2019 and mid-2020, an unprecedented controversy reached courts in Europe" (**A. PIN**, *'A Novel and Controversial Technology.' Artificial Face Recognition, Privacy Protection, and Algorithm Bias in Europe*, in *William & Mary Bill of Rights Journal*, Vol. 30, n. 2/2021, p. 291).

discriminatory on the grounds of sex and race, for example producing a higher rate of false positive matches for female faces and for black and minority ethnic faces than for other groups.

On 4th September 2019, the High Court of Justice of Wales held that it was lawful for the police to use the AFR Locator, since it met the fundamental need for public security and passed the proportionality test[266]. In particular, it offered the following guarantees: after matching images, it destroys all pictures collected; people are informed that the system is watching them in a given area; when the AFR detects a suspect, the last word for identification is that of a policeman (a human). Regarding the discrimination argument, the Court simply said that there was no firm evidence that the Locator produces indirectly discriminatory results[267].

This judgement was overturned on 11th August 2020 by the Court of Appeal of England and Wales[268]. In the judges' opinion, the police forces had never sought to verify, directly or otherwise, whether the software showed unacceptable bias on the grounds of race or sex. Indeed, nobody had access to the datasets on which the system was trained and therefore could not analyse those datasets for bias. In the Court's view: "[a]s a minimum for confirming whether an AFR system is biased, the database statistics, such as the number of males to females, and different races considered, would need to be known"[269].

It was not proved that the facial recognition system used by the South Wales Police Force was discriminatory, but due to the lack of proper controls, the Court established that the police had not done all that they reasonably could to comply with the non-discrimination law.

The police is just a final user of the facial recognition method: it did not write the algorithm, nor train it with the necessary datasets. However, since AFR is "a novel and controversial technology"[270], the Courts established that all police forces that intend to use it in future should ensure that every reasonable measure is taken to verify that the software does not have a racial or gender bias.

---

[266] R (Bridges) v. Chief Constable of South Wales Police [2019] EWHC (Admin) 2341, hereinafter *Bridges 2019*.

[267] See *Bridges 2019*, par. 153.

[268] R (Bridges) v. Chief Constable of South Wales Police [2020] EWCA (Civ) 1058, hereinafter *Bridges 2020*.

[269] See *Bridges 2020*, par. 193.

[270] See *Bridges 2020*, par. 201.

As pointed out in the literature, requiring that AI users (in this case the police) demonstrate that they have taken into proper account the risk of discrimination while deploying facial recognition systems, as the Court of Appeal has done, "shifts part of the burden of proof from the subject claiming to have been discriminated against to the respondent"[271].

The Bridges case suggests that AI users must be informed about how the instruments they use work and can incur sanctions in the case of lack of proper assessments. Whether this jurisprudence will be upheld in the future, including in other jurisdictions, is currently unknown. It is certainly a very cautious approach to AI recognition systems and can be extended to all types of discrimination, including religious discrimination. Indeed, it cannot be excluded that facial recognition algorithms could have discriminatory effects not only against coloured persons or women, but also against people who wear headdresses, veils, beards or hairstyles related to certain religious beliefs.

4 - Conclusion

The debate on the regulation of AI is ongoing and there is a need for specific legal tools, tailored to the features of this technology[272]. Although it has been argued that the legal categories we already have can frame the phenomenon, the need for new rules cannot be denied[273]. Many voices, especially in Europe, claim that it is time to negotiate international or global treaties on artificial intelligence.

Regarding the discriminatory effects of AI, it stands to reason that their mitigation is not only a legal issue, but also a question of technological development: the creators of AI, engineers and computer scientists, are constantly developing corrective measures and eliminating bias-producing flaws from their systems. A pivotal role is also played by humans: people cannot completely abdicate to machines for decision-making. Their role in mitigating and smoothing AI errors and biases is irreplaceable.

---

[271] **A. PIN**, *'A Novel and Controversial Technology.' Artificial Face Recognition, Privacy Protection, and Algorithm Bias in Europe*, p. 309.

[272] See **A. SIMONCINI**, *Verso la regolamentazione della Intelligenza Artificiale. Dimensioni e governo*, in *Biolaw Journal*, n. 2/2021, p. 411 et seq.

[273] **M. LUCIANI**, *Forum AI and Law*, in *Biolaw Journal*, n. 1/2020, p. 489.

Nonetheless, as the Bridges case demonstrates, jurists too have been refining their tools and framing the problem to meet the challenge. Not by chance, the strategy on AI elaborated by the European Commission also names observation of the principles of non-discrimination and equity among the fundamental requirements for a reliable AI, aware that "the [d]ata sets used by AI systems (both for training and operation) may suffer from the inclusion of inadvertent historic bias, incompleteness and bad governance models"[274].

Although a start has been made toward mitigating algorithmic discrimination, two final remarks seem necessary.

First, it should not be forgotten that the development of AI can also be useful for fighting discrimination. AI and algorithms can be used to detect unfair discrimination, and software for this purpose has already been written. So it is conceivable that judges will soon use AI to detect discriminatory choices and decisions by humans or even other AI systems. Returning to the example of recruitment policies: in order to verify whether or not an automated selection processes produces discriminative effects on the grounds of religion, an algorithm can be used to analyse the data of people hired.

Second, it is extremely important that jurists' attention to the phenomenon of AI bias should not be limited to certain grounds of discrimination but should also extend to less explored areas, including religious discrimination. Indeed, due to the expansion and development of AI in everyday life, religious and other discrimination could easily find its way into new virtual and real spaces.

[274] *Ethics guidelines for trustworthy AI*, High-level expert group on Artificial Intelligence set up by the European Commission, 8 April 2019, p. 18. Among the other relevant EU documents, see: "Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, COM (2021) 206 final (21 April 2021). See also "An EU Artificial Intelligence Act for Fundamental Rights - A Civil Society Statement", 30 November 2021).