Digital Transformation of the World Economy by AI: Some Moral, Ethical and Sharīʿah Concerns

Volker Nienhaus

Emeritus Professor of Economics, University of Bochum, Germany

volker.nienhaus@gmx.net

Abstract:

The digital future of the world economy in the post-Covid-19 era will be shaped by an accelerated deployment of digital technologies in the private and public sectors with more active involvement of regulators and legislators. In recent years, ethical concerns have become more prominent and may limit machine learning models' deployment in consumer-facing businesses, including financial services. A widely shared Islamic perspective is still to emerge. The paper draws attention to ethical issues embedded in data and autonomous decision models. This debate in the West could be enriched by more contributions from an Islamic perspective.

Keywords: Personal Data Monetisation, Big Data Ethics, Geodemographic profiling, Ethical AI

التحول الرقمي للاقتصاد العالمي من خلال الذكاء الاصطناعي: قضايا أخلاقية وسلوكية و شرعية

فولكر نينهاؤس

الأستاذ الفخري للاقتصاد، جامعة بوخام - ألمانيا

volker.nienhaus@gmx.net

(سلّم البحث للنشر في 22/ 2/2021م، واعتمد للنشر في 12/ 3/ 2021م)

https://doi.org/10.33001/M011020211591

Check for updates

الملخص:

إن التطبيقات المتسارعة للتكنولوجيا الرقمية ستشكل مستقبل الاقتصاد العالمي بعد أزمة كوفيد 19 في كلا القطاعين العام والخاص، وذلك بالتزامن مع مشاركة نشطة من قبل الجهات التنظيمية والتشريعية. في السنوات الأخيرة، أصبحت المخاوف الأخلاقية أكثر بروزًا وربما قد تؤدي إلى الحد من انتشار نماذج التعلم الآلي في بيئة الأعمال التي تتعامل مع العملاء بشكل مباشر بما في ذلك الخدمات المالية. لا يزال بروز نظرة إسلامية واسعة النطاق بهذا الخصوص في بداية الظهور. إن هذا البحث يلفت النظر إلى القضايا الأخلاقية المدمجة في البيانات ونماذج القرارات المستقلة. يمكن إثراء هذا النقاش في الغرب من خلال إسهامات أكثر من منظور إسلامي.

الكلمات المفتاحية: تسييل البيانات الشخصية، أخلاقيات البيانات الضخمة، التصنيف الجيوديموغرافي، الذكاء الاصطناعي الأخلاقي

Introduction

Ethical issues related to AI and its application in different contexts have attracted much attention from data and social scientists, philosophers, lawyers, human rights activists, regulators, and policymakers in the West.(1) Many of these issues could or should also be addressed from a Sharīʿah perspective. As the Sharīʿah is a comprehensive legal and moral code, a Sharīʿah perspective implicates a systemic dimension: laws, regulations and behavioural guidances should be derived from or consistent with the norms and values of the Islamic worldview. A structurally similar approach can be perceived in the evolution of a legal and political governance system for personal data and AI in the European Union in the 2000s. From (rather abstract) common values fundamental rights and responsibilities are derived, then specified, and subsequently 'translated' into legal acts, regulations, and guidance.

The Emerging Data Governance and AI Regime of the European Union

In 2018, the European Commission released a communication on "Artificial Intelligence for Europe" that underlines that new technologies are based on values. "The EU must … ensure that AI is developed and applied in an appropriate framework which promotes innovation and respects the Union's values(2) and fundamental rights as well as ethical principles such as accountability and transparency. The EU is also well placed to lead this debate on the global stage."(3) A High-Level Expert Group on Artificial Intelligence (AI HLEG) was set up to submit "Ethics Guidelines for Trustworthy AI" in 2019(4). Before, the three governing bodies of the EU – the European Parliament, the Council of the European Union, and the European Commission – had jointly released the "Charter of Fundamental Rights of the European Union" in 2000.(5) The preamble mentions

(1) Many issues that are presented here from an ethical perspective have also be analysed from a legal perspective, for example the issues of fairness, (non-)discrimination, or accountability; see Ebers and Navas 2020, Ebers and Gamito 2021, Barfield 2021.

(2) These values are laid down in article 2 of the Treaty on European Union (so-called Maastricht Treaty): "The Union is founded on the values of respect for human dignity, freedom, democracy, equality, the rule of law and respect for human rights, including the rights of persons belonging to minorities. These values are common to the Member States in a society in which pluralism, non-discrimination, tolerance, justice, solidarity and equality between women and men prevail." European Union 1992.

(3) European Commission 2018.

(4) High-Level Expert Group on Artificial Intelligence 2019.

(5) European Union 2000.

the “indivisible, universal values of human dignity, freedom, equality and solidarity” and “the principles of democracy and the rule of law” as the normative basis of the fundamental rights outlined in the Charter. Article 8 covers the protection of personal data.(6) To give this fundamental right a more concrete shape and legal underpinning, the EU started working on a “General Data Protection Regulation (GDPR)” in 2012. It was adopted in 2016 and became effective in 2018.(7) The GDPR shall be supplemented by a “Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)”, for which the European Commission released a proposal in 2021.(8) The AI Act supports the rather ambitious “objective of the Union being a global leader in the development of secure, trustworthy and ethical artificial intelligence”.(9) Another international body concerned with values-based AI principles is the OECD(10), but “the EU was closely involved in developing the OECD’s ethical principles for AI”.(11)

The structural similarities between the emerging AI regime of the EU and a Sharīʿah-based system suggest that it might be worthwhile to take a closer look at the European developments.

- First, there are initiatives towards regulating AI in several countries, but they are often merely technical and not embedded in an elaborate normative framework.
- Second, an AI regime for the EU must be acceptable to 27 member states with different interpretations of the common values, disparate political agendas, various legal systems, divergent economic interests, heterogeneous levels of technical and socio-economic development, and so forth. Diversity in many respects is also a characteristic feature

(6) “1. Everyone has the right to the protection of personal data concerning him or her. 2. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified. 3. Compliance with these rules shall be subject to control by an independent authority.”

(7) European Union 2016.

(8) European Commission 2021a. Important annexes have been compiled in a separate document, see European Commission 2021b. The European Parliament and the 27 member states as well various stakeholders such as lobby groups, civil society organisations, political parties, and industry bodies will articulate their views on the proposal, and it will take months if not years before a final version will enter into force.

(9) European Commission 2021a, 1-2.

(10) See https://www.oecd.org/going-digital/ai/principles/.

(11) European Commission 2020, 8. In contrast to the EU, the OECD as an intergovernmental organisation has no legislative power. Out of 38 members of the OECD (https://www.oecd.org/about/) only Turkey is a Muslim country.

of the Muslim world. It can be illuminating to study how consensus has been reached in the EU despite the many divergencies at the outset.

- Third, secular and Islamic values differ in their origin, but there are many similarities when it comes to operationalisation and implementation. Most (if not all) issues that have to be tackled in Europe will also pop up in the Muslim world and should be addressed from an Islamic perspective.
- Finally, the aim of the EU to become the global leader in the development of ethical AI might be seen as a challenge in other parts of the world. It could induce Muslim scholars and scientists to devise an ethical AI regime from an Islamic perspective (similar or radically different).

While the EU can serve as a reference for the emergence of a legal and regulatory AI regime with an explicit ethical underpinning, initiatives from a Sharīʿah perspective are still in a more fragmented early stage. In addition to a few individual contributions, the Islamic Development Bank Institute (IsDBI) has recently launched a study on AI and Islamic finance.[12] Because its focus is on the role of AI for financial inclusion, it touches only upon ethical issues of AI with relevance for finance. It does not intend to develop a general and comprehensive AI regime from an Islamic perspective. Nevertheless, this study reflects the 'state of the thinking' in a leading global Islamic institution. It shall be used as a reference for the Islamic view on finance-related ethical data and AI issues.

1. Ethical Issues with Data

The processing of data is at the core of the global digital transformation. Personal data are of particular significance for the transformation of economies and societies. Their processing is not a value-neutral technical procedure but may pose ethical challenges related to, for example, privacy, commercial exploitation, discrimination, and unfair treatment.

1.1 Personal Data and Human Rights

The GDPR (article 1(1)) "lays down rules relating to the protection of natural persons with regard to the processing of personal data and rules relating to the free movement of personal data". It "protects fundamental rights and freedoms of natural persons and in particular their right to the protection of

(12) Ashraf et al. 2021.

personal data" (article 1(2)). This is understood to mean that data privacy is a basic human right.(13) Article 4(1) gives a wide definition of personal data as "any information relating to an identified or identifiable natural person ('data subject')".[14] The European Commission explains that "[d]ifferent pieces of information, which collected together can lead to the identification of a particular person, also constitute personal data. Personal data that has been de-identified, encrypted or pseudonymised but can be used to re-identify a person remains personal data and falls within the scope of the GDPR."

It is expected that other countries will follow the European example such that "[d]ata privacy will soon be a human right around the world. California, India, Singapore and Japan have led the way, with other countries actively pursuing the idea."[15]

The EU has translated the human right of data privacy into an elaborate data regulation system with, among others, the obligation of the data operator to ask the data subject for consent to the processing of personal data and restrictions on data transfers to recipients located in jurisdictions outside the EU where the level of protection is not considered adequate. When assessing the adequacy, the EU Commission shall take account of, for example, the rule of law, respect for human rights, the access of public authorities to personal data, and data protection rules (article 45(2)).[16]

The high importance that Sharīʿah attaches to the protection of privacy raises the obvious question of whether data privacy can have a similar elevated status in an Islamic context as it has in a secular legal and regulatory system. Data protection could be justified from a Sharīʿah perspective. Still, a cursory look at data protection laws of a few other Muslim countries does not indicate that personal data protection is justified by reference to Sharīʿah principles.[17]

(13) See Vincent 2018. However, the recital number 4 of the GDPR (European Union 2016) points out: "The processing of personal data should be designed to serve mankind. The right to the protection of personal data is not an absolute right; it must be considered in relation to its function in society and be balanced against other fundamental rights, in accordance with the principle of proportionality."

(14) Article 4(1) also gives a wide definition of an identifiable natural person as "one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person". Different pieces of information, which collected together can lead to the identification of a particular person, also constitute personal data..

(15) Raval 2019.

(16) The list of countries for which the EU Commission (2017) has made adequacy decisions is rather short: Andorra, Argentina, Faeroe Islands, Guernsey, Israel, Isle of Man, Jersey, New Zealand, Switzerland, and Uruguay plus private entities in Canada falling under the scope of the Canadian Personal Information Protection and Electronic Documents Act and companies in the US abiding by the binding Privacy Shield principles. However, the European Court of Justice (2020) declared the Privacy Shield invalid on 16 July 2020.

(17) See Caruana and Cannataci 2007, Guseyva 2020. Even the government of the Islamic Republic of Pakistan did not refer to Islamic principles when drafting the Pakistani Data Protection Law in the early 2000s. "[T]he main objective of the draft … is not to enshrine the principles of Islam on privacy, but to satisfy the requirements of EU Directive 95/46 … with the hope of ensuring that data will be allowed to flow freely between the EU and Pakistan, thus making Pakistan an attractive market for outsourcing." Hayat 2007, 146. The EU directive 95/46 was the predecessor of the GDPR and comprised restrictions of personal data flows.

Nevertheless, it seems to be a consensus that Sharīʿah supports personal data protection. What is less clear is the appropriate design of a data protection law – from the definition of personal data to data flow restrictions and the (limitation of) intervention rights of governments and freedom of information rights of citizens towards the state. The IsDB study mentions data protection and privacy concerns several times but does no add a specific Sharīʿah dimension to the conventional arguments.(18)

1.2. Data Ownership

One set of personal data is of great interest for businesses in the financial industry, namely financial transaction data (bank transfer, credit card payments, use of ATMs, payment services, etc.). In the past, information about bank customers' financial transactions was effectively 'owned' by the bank and generally not shared with other (competing) financial firms.(19) The government of the UK took a game-changing decision when it mandated the UK's nine largest banks to install an Open Banking Application Programming Interface (API) by January 2018.(20) Through this API, other (licensed and regulated) financial service providers get access to current accounts' transaction data when the account holders allow access. The objectives of Open Banking are the intensification of competition in retail banking and the support of newcomers (FinTechs) that can provide cheaper and better services to bank customers and small and medium-sized enterprises (SMEs).(21)

On a more fundamental level, privacy and data protection legislation and Open Banking initiatives in the UK and elsewhere(22) are manifestations of a rethinking of personal data ownership (or the "individual's sovereignty

(18) A specific Islamic perspective is expounded only with respect to the Sharīʿah-compliant financing techniques in general and Islamic social finance in particular (which is in line with the focus of the IsDBI study). Other than that, the study provides only a rather general statement on the normative basis of an Islamic system: "Islam promotes markets based on moral principles: seeking mutual gain and win-win outcomes that make the two parties of trade better-off. The same principle prevents gambling and interest-based finance that tends to be a win-lose situation and thus morally damaging. Islamic finance emphasizes respect of property rights, social and economic justice, observance of the rights about earnings and distribution of wealth, governance, mutual agreements etc. The sanctity of contracts and ethics in business conduct would be inculcated at the individual level and enforced by society and law." Ashraf et al 2021, 26.

(19) An exception were notifications about loan defaults which were submitted to credit bureaus and shared within the banking industry.

(20) Open banking initiatives for sharing transaction data with approved third parties have also been launched in other jurisdictions such as the EU (https://www.openbankingeurope.eu/), New Zealand (https://www.moneyhub.co.nz/open-banking.html), or Nigeria (https://www.openbanking.ng/). For a global survey on open banking see: The Paypers 2020.

(21) By the end of 2020, 79 account providers (banks, building societies and payment companies) and 215 third party providers of account information and payment initiation services participate in the UK's Open Banking system, all regulated by the Financial Conduct Authority, https://www.openbanking.org.uk/about-us/latest-news/obie-highlights-december-2020/.

(22) Legal initiatives similar to the GDPR have also been launched in a (relatively small) number of Muslim countries; see the reports especially on the UAE, Qatar, and Bahrain in Manda and Eskhita 2019; less stringent personal data protection laws are reported for Indonesia, Saudi Arabia, and Malaysia.

over cyberassets such as personal data"[23]). Financial institutions and other data-collecting businesses, mainly social media platforms and e-commerce ventures, store collected customer data in databases with a closed architecture.[24] The operators of the databases act like the owners of the collected data and monetise them: Businesses of all branches of industry are willing to pay significant amounts for access to collected personal data for use in marketing campaigns, product customisation, brand development, etc. Collected personal data have become valuable assets for which a multi-billion market has emerged.[25] Personal data are treated as a kind of commodity ('data as the new oil')[26], and the authors of the IsDB study criticise the treatment of big data as a natural resource. This, however, is not substantiated by Sharīʿah specifics of ownership but by the (disputable) argument that "[t]roughout human history, the privatization of natural resources – oil, coal, natural gas, forests and timber, minerals – have created large monopolies and contributed to massive wealth creation for a privileged few. A natural outcome of this is gross and ever-increasing economic inequalities. The idea of a monopoly is alien to Islamic economic ethics."[27]

Personal data are an immaterial commodity that has some odd features. For example, personal data are inseparable from the person (= data subject) and can be used (with or without monetary compensation) by different users, i.e. multiple times without being depleted. Not the data subjects but the data collecting business (e.g. Facebook, Twitter, Instagram, Google, Amazon) monetise personal data – often without authorisation by the data subjects.[28] The monetisation by data collectors is like a lease or sale of usufruct of their database. It would be interesting to learn how the advertising-based business model of social media, search engines and e-commerce is assessed from a Sharīʿah perspective (legally as well as ethically). For example, it is unclear how Sharīʿah rules of ownership and nominate exchange contracts should be applied when Muslims want to monetise their personal 'data assets'.

(23) Bryson 2020, 18.

(24) "A closed architecture like Facebook's or Twitter's puts all the information about its users - their handles, their likes and photos, the map of connections they have to other individuals on the network - into a private database that is maintained by the company." Johnson 2018.

(25) Data brokers collect information about consumers from all kinds of public and non-public sources (from official registers and social media profiles to loyalty programmes and cookies on websites) and create individual profiles that are sold to businesses and other users such as NGOs or political parties, for targeted marketing, fundraising or election campaigns, respectively. The size of the data brokering industry is estimated at USD 200 bn (WebFX Team 2020). Thirani and Gupta 2017 reported that the EU Commission expected the value of personalised data to reach EUR 1 trillion by 2020.

(26) This catchy phrase does not pay due attention to specific qualities of data as pointed out by French, Carr, and Lowery 2020, 2.

(27) Ashraf et al. 2021, 21.

(28) A kind of authorisation became obligatory in Europe only recently after the GDPR entered into force in 2018.

In recent years, Open Banking and the GDPR can be seen as first or preparatory steps for revising the current monetisation practices with the long-term perspective of giving the effective ownership of personal data (back) to the people (data subjects). Strategies for effective monetisation of personal data by the data subjects are still in an early stage of development. One approach requires collective action and a platform organisation to tackle scale issues, network effects, and prohibitive switching costs in (un)competitive markets.(29) Another approach focuses on open, decentralised blockchain architecture and non-fungible tokens that contain people's social identity.(30) The tokens make the social identity portable and eliminate switching costs.

Open Banking regulations and data privacy laws of the GDPR type can have far-reaching implications for business models of (Islamic) FinTechs and data brokers if data subjects become less 'generous' in consenting to the exploitation of their data. In any case, competition within the (conventional and Islamic) financial industry will be intensified by revisions of the legal and regulatory framework.(31)

1.3. Proxies for Prohibited Data

Article 9(1) of the GDPR prohibits the processing(32) of certain types of data ('prohibited data'): "Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited."

Restrictions in the processing of prohibited data have reinforced the interest in a type of data – geodemographic data – that can be used for profiling without people's consent because they are not collected from them but. "Geodemographic systems estimate the most probable characteristics of people based on the pooled profile of all people living in a small area near a particular address."(33) Sources for geodemographic profiling are, among

(29) For a summary of (1) some implementations of Personal Information Management Services (PIMS) that enable people to accrue and trade data about themselves and (2) a selection of Personal Data Projects by telcos, social networks, governments and digital agencies see Mobile Ecosystem Forum and Juniper Research 2016. For a theoretical analysis see Bataineh, Mizouni, Bentahar, and El Barachi 2020.

(30) "You should own your digital identity - which could include everything from your date of birth to your friend networks to your purchasing history - and you should be free to lend parts of that identity out to services as you see fit." Johnson 2018.

(31) The IsDBI study does not refer to these – possibly game-changing – revisions of the legal and regulatory framework.

(32) Article 4(2) GDPR gives a wide definition of processing that includes, among others, the collection, storage, structuring, and disclosure of personal data.

(33) https://en.wikipedia.org/wiki/Geodemography.

others, census data, public registers, satellite and Google Street View images, crime statistics, etc., which are publicly available with sufficient geographic granularity, i.e. for small areas, for instance, defined by a zip code. Big data algorithms process the collected geodemographic data to identify local clusters (neighbourhoods)(34) of people with similar features such as lifestyle, housing conditions, car ownership, or social class correlated with, for example, disposable income, wealth, etc. or the probability of (un)employment.

Given that migrants from one country often live together and people with a common religion congregate in neighbourhoods, geodemographic clusters can become de facto proxies for prohibited data such as ethnicity or religion.

The purpose of using geodemographic data is to find out more about a loan applicant's creditworthiness beyond what can be seen from the personal data submitted by a loan application. This 'more' will have a weight in the final decision because the inclusion of the geodemographic data into the decision model improves its predictive accuracy. If this were not the case, the data would be ignored. Suppose it is highly probable that the geodemographic data are in substance proxies for variables that must not be considered in the decision making. In that case, their use in life-altering decisions raises ethical and legal questions.

From an Islamic perspective, a first question could be whether considering geodemographic data in a decision about an application for a loan, a job, etc., is like spying on the applicant or a breach of privacy. Even if this is answered negatively, a second question is about the morality of using variables in decision making that are proxies for data banned by law and whether the answer would be different for a country with no law to ban prohibited data. Given the concept of brotherhood in Islam, anti-discrimination measures should be welcomed from a Sharīʿah perspective. If this is so, the use of data that are possible entry points for discrimination should be thoroughly reviewed. This requires some statistical expertise of those authorities that certify the Sharīʿah-compliance of FinTechs who operate decision models based on big and geodemographic data.(35)

(34) As an xample for a socio-ethnic clustering of neighbourhoods in London see Singleton and Longley 2015.

(35) The IsDBI study refers to the use of a number of ethically questionable techniques by FinTechs without any critical comments: Several FinTechs "are using different types of data extracted from mobile phones to feed their behavioral analysis algorithms to predict users' lifestyles, social networking, and economic activity. … [Some FinTechs evaluate] call detail records such as the number of calls/text messages made/received, the total duration of incoming/outgoing calls, number and duration of calls sorted by time of day and daily number call sorted by duration. Mobile phone data also shows location movements, which can be used to infer the users' weekly and monthly cycle of the previously visited locations, such as a salaried income and settled house and family. More complex models have recently started analyzing the categories of apps on a loan applicant's phone and the patterns of the phone's calendar, internet browsing history, or mobile wallet transactions." Ashraf et al. 2021, 13-14. "[C]ompanies working on models to predict creditworthiness have developed AI models … [that] extract features from facial expressions to calculate a trustworthiness judgment rate that predicts the trustworthiness of the applicant's face collected after the imaging session … The attributed trustworthiness judgment rate will then be included in the overall creditworthiness score." Ibid., 15.

1.4. Verification of Big Data by Small People?

FinTechs can combine data from different sources – e.g. cash flow data from Open Banking, lifestyle data from social media, and geodemographic data on the socio-economic environment - as inputs for their proprietary financial services and decision models. “Big-data scoring tools may now base credit decisions on where people shop, the purchases they make, their online social media networks, and various other factors that are not intuitively related to creditworthiness.”(36)

Data transparency laws or regulations give applicants the right to learn from the financial service provider what data have been collected and how they have been handled. The information rights alone do not restrict FinTechs in the scope of the information they may use for their credit scoring; prohibited data excluded. Loan applicants may have to consent to the collection and processing of personal information from crime registers, social media, e-commerce, telecommunication and utility providers, and other sources when they sign the financing application. Particularly for unbanked applicants with no credit histories (‘thin-file applicants), the consent to the use of non-traditional data is a take-it-or-leave-it decision. In jurisdictions where transparency laws do not exist, it is left to the firms’ discretion to disclose to consumers the data basis of their decisions. FinTechs very often remain opaque about the automatically processed data.

Where transparency laws and data protection rights exist, the data subjects must be put in a position to verify the information sourced from third parties. The challenge for data subjects is that the volume of processed information might be so large that it is factually impossible to validate all data within a reasonable time. When a data subject detects inaccuracies in his or her personal data,(37) he or she must have a right to rectification. Article 16 of the GDPR grants this right but goes further by also giving “the right to have incomplete personal data completed”.

Since Islamic contract law attaches great importance to transparency and unambiguousness, opaqueness in (pre-)contractual relations should raise at least ethical (if not legal) concerns. If Sharīʿah authorities consider data opacity reprehensible, they could (1) lobby for a transparency (and dispute settlement)

(36) Hurley and Adebayo 2016, 148.

(37) The Federal Trade Commission (FTC) of the US conducted a study of the accuracy and completeness of consumer credit reports. The overall finding was that 21% of nearly 3,000 reviewed credit reports had material errors that had to be corrected after a dispute process. The correction had an impact on the credit score in 13% of the cases; see FTC 2012.

regulation in jurisdictions where it does not yet exist, (2) urge opaque firms to become more transparent and give affected people an opportunity to inspect, correct, and complete the personal data used for processing, and (3) give the public the general advice to shun firms with opaque data sources and procedures. It is questionable whether FinTechs with opaque data practices can be certified as Sharīʿah-compliant.(38) The lack of transparency and opacity of FinTechs is only briefly mentioned in the IsDBI study, and the issue of Sharīʿah-compliance is not discussed.

2. Ethical Issues in Models

"Among many other applications, machine learning is used today to make life-altering decisions about employment, bail, parole, and lending."(39) The proposed AI Act classifies AI systems applied for life-altering decisions as "high-risk" (article 6). Annex III specifies areas in which AI systems are considered high-risk, including (but not restricted to) biometric identification, access to educational and vocational training institutions, filtering job applications, evaluation of job performance, and access to and enjoyment of essential private and public services.(40) High-risk AI systems have to comply with several provisions summarised below.

2.1. Credit Scoring Models as High-Risk AI Systems

Ethical and (potentially) Sharīʿah issues of high-risk AI systems can be exemplified by credit scoring models (CSMs). CSMs assess the capacity of loan applicants to meet future payment obligations to minimise the financier's risk of capital loss. "The credit-scoring industry has experienced a recent explosion of start-ups that take an 'all data is credit data' approach, combining conventional credit information with thousands of data points mined from consumers' offline and online activities."(41) CSMs that FinTechs developed are meanwhile also applied by incumbent banks. CSMs can be used for decision support or – as autonomous CSMs – decision making without human intervention.

When FinTechs introduced CSMs that did not rely on banking data and credit histories but used big data and AI to provide credit to thin-file borrowers,

---

(38) "If users did not know how their scores been calculated and even cannot complain about the unfair scores, the fairness of credit reporting can be questionable." Chang and Lin 2019, 351-352.

(39) Hall and Gill 2019, 6.

(40) For more details and the full list of high-risk AI systems see annex III in European Commission 2021b.

(41) Hurley and Adebayo 2016, 148.

their market entry was widely hailed as a significant push towards financial inclusion. The authors of the IsDBI study are optimistic about CSMs: "The adoption of AI results in an objective, better-informed faster (backed by data), more accurate credit risk assessment of credit-worthy MSMEs [micro, small, and medium enterprises] lacking credit history."(42)

However, the IsDBI study also points out that many conventional FinTech lending platforms "offer a short-term loan on a predatory interest rate, more often as a consumption loan. Furthermore, AI is mostly about debt creation and credit extension, leading to higher indebtedness in society faster in the name of financial inclusion in developing countries and efficiency enhancement in developed countries. There is a need to critically review such platforms' practices to avoid the financialization of fringe borrowers."(43) FinTechs that observe Islamic finance principles should be immune to this critique. Their use of CSMs does not only serve a legitimate purpose (to protect the capital of the financier) but has the additional social benefit of financial inclusion. Hence, Islamic finance scholars who consider technologies as ethically neutral do not see any reasons to scrutinise the underlying technology of CSMs for ethical issues.

In contrast, consumer protection and data privacy advocates in secular societies drew attention to ethical issues that are not only related to the application of CSMs but inherent in the CSM technologies themselves, ranging from unavoidable prediction errors and fairness dilemmas to unintuitive data and uninterpretable models (all detailed below). The discussion about such inherent deficits of AI models in general and CSM in particular has led to a more cautious attitude in Europe. The proposed European AI Act (annex III(5) b) classifies AI systems for the evaluation of the creditworthiness of natural persons and the calculation of a credit score as high-risk.

2.2. Prediction Errors and Fairness

Personal data used to determine eligibility for credit, insurance, employment, and other similar purposes can be incorrect or incomplete. "Errors in consumer reports … can cause consumers to be denied credit or other benefits or pay a

---

(42) Ashraf et al. 2021, XII. The study also notes that the adoption of AI "brings challenges related to … possible consequences of the adoption of AI regarding fairness that is discrimination among borrowers, data privacy and security." Ibid., XII.

(43) Ibid., XII and 22. As a possible solution, the IsDBI study suggests to expand the scope of AI models: "There is a need to develop such AI models that can help MSMEs build capital while financiers reap the benefit of their investment. This requires a paradigm shift from risk transfer to risk sharing, from assessing creditworthiness to business worthiness" (p. 22), i.e. "to assess the experience, business acumen, entrepreneurial traits, and ability to network. In such a situation, a psychometric analysis, network analysis, credit screening (financial management history), and business worthiness of microenterprise can be used to avoid adverse selection" (p. 35). This concept has not yet been tested in practice.

higher price for them and may lead credit issuers to make inaccurate decisions that result in declining credit to a potentially valuable customer or issuing credit to a riskier customer than intended."[44] Even with thousands of data points, prediction errors can never be ruled out. Suppose a CSM predicts whether an applicant will meet the future payment obligations and decides autonomously on accepting or rejecting a loan application. The model can make two types of errors:

- accept an application and give credit to a borrower who later defaults (= false positive error),
- reject an application of a person who would have met all payment obligations (= false negative error).

There is a fundamental difference between these types of possible errors. While the false positives can be observed, false negatives are counterfactual and thus unobservable and unmeasurable (except in experimental settings).[45]

Suppose that experiments with the CSM have shown a trade-off between false positives and false negatives. A decrease of false positives leads to an increase of false negatives, given that the terms for the financing (cost and tenure, collateral) do not change.[46] As only the percentage of false positives can be measured, the numerical value for the trade-off cannot be calculated. Yet, it is safe to say that a specified level of false positives implies some level of false negatives, and a decrease of the false positives most probably increases the level of false negatives. In a more substantive wording: measures to reduce the level of credit defaults (to increase the protection of existing wealth of the lenders) will increase the level of lost opportunities for the generation of new wealth by borrowers. When the CSM is based on machine learning (see below para 2.4), its operator will set the target level of false positives. The AI makes the necessary adjustments of parameters on its own, i.e. it will be 'responsible' for the level of false negatives. The AI co-determines the relation between wealth protection and lost opportunities for wealth creation. This can be understood as the participation of the AI in an intrinsically ethical decision.

(44) FTC 2012b, 1.

(45) "Determining whether an action causes an outcome requires two predictions: first, what outcome will happen after the action is taken, and second, what outcome would have happened had a different action been taken. But that's impossible. You will never have data on the action not taken." Agrawal, Gans, and Goldfarb 2018, 25%.

(46) The intuitive rationale is that a reduction of false positives (i.e. of the overall default rate) results from a tightening of the creditworthiness requirements. Fewer loan applicants will meet the tighter requirements and the number of declined applications increases. Suppose the candidates for a false negative decision are randomly distributed among all applicants, and their number is independent of the tightness of the credit scoring criteria (as the criteria do not capture all factors that can cause the failure or success of a borrower). Then the increase in the number of declined applicants increases most probably the number of declined false negative candidates.

The ethical quality becomes even more apparent when differences in false positives emerge between groups of applicants that are defined by a ‘sensitive’ differentiator such as gender, ethnicity, or religion. Issues of discrimination and fairness immediately pop up, and a spontaneous reaction would be a call for the equalisation of the false positive ratios. However, there is a high probability that the false positive/negative trade-off implies that equalised false positive ratios produce unequal false negative ratios between the groups. This would shift a possible discrimination from wealth protection to opportunities for wealth creation.

Without further details, the example shows that AI is not value-neutral but inextricably enmeshed in ethically loaded decisions. This triggers a host of questions such as the following: is the equalisation of false positives an appropriate fairness criterion, or should it be the equalisation of false negatives, and how could this criterion be operationalised?(47) Can decisions that have to meet fairness criteria be left to autonomous AI systems, or should a human operator always have the final say?(48) If the latter applies, does this mean that only interpretable models (see 2.4. below) will be allowed in specified areas?

2.3. Unintuitive Data

Consumers can learn about CSMs and the thousands of data points they process. Many of these data points may look unrelated to creditworthiness (= unintuitive data). Unintuitive data play a role in prediction models based on machine learning (ML) algorithms and big data: the algorithm decides on their relevance, and unintuitive data can be relevant if correlated with the target variable. Their inclusion improves the model’s predictive quality.(49)

This formal argument does not explain how unintuitive data are causally linked to creditworthiness. It may well be that causality cannot be identified for a single unintuitive variable in isolation. However, in combination with other input variables, unintuitive data may detect behavioural patterns that are statistically relevant and enhance the quality of predictions. Explanations for such ‘hidden causalities’ require a deeper understanding and closer

(47) On fairness criteria and AI see Mehrabi et al. 2019, Binns 2020.

(48) The proposed AI Act of the EU is quite restrictive in this regard: article 14 requires that high-risk AI systems shall be effectively overseen by natural persons. The individual overseeing the system should, among others, “fully understand the capacities and limitations” of the system, be able “to override or reverse the output” and to “interrupt the system through a ‘stop’ button or a similar procedure”.

(49) “Analysts built their regression models on hypotheses of what they believed mattered and how … Machine learning models are particularly good at determining which of many possible variables will work best and recognizing that some things don’t matter and others, perhaps surprisingly, do. … In this way, machine learning enables predictions based on unanticipated correlations” Agrawal, Gans, and Goldfarb 2018, 15%.

examination of the underlying model.

When unintuitive data become crucial for the decision of a CSM model on a loan application, they are most probably indicators for a group membership. Big data tools create “a system of ‘creditworthiness by association’ in which consumers’ familial, religious, social, and other affiliations determine their eligibility for an affordable loan. These tools may furthermore obscure discriminatory and subjective lending policies behind a single ‘objective’ score. Such discriminatory scoring may not be intentional; instead, sophisticated algorithms may combine facially neutral data points and treat them as proxies for immutable characteristics such as race or gender, thereby circumventing existing non-discrimination laws and systematically denying credit access to certain groups.”(50)

Discriminatory or biased scoring implies an unfair treatment of people, which is a feature of the scoring technology (applied in the context for which it was developed). This technology is not ethically neutral and should be judged by ethical standards comparable to those applied to human decision-makers.

2.4. Model Interpretability

A person whose credit application has been rejected may feel being mistreated. People negatively affected by a model-based decision in potentially life-altering contexts must have a right to challenge the decision. To enable people to challenge adverse decisions, it is a matter of fairness to give them substantive reasons why the adverse decision was taken. The mere reference to the output of an algorithmic model is not sufficient – especially not when the output refers to or is based on unintuitive data with no apparent causal relationship to the matter of decision.

In several countries, consumer protection laws give consumers in areas such as finance or employment a right to learn why an application (for a loan, insurance, or job) has been rejected. For example, the Equal Credit Opportunity Act (ECOA) and the Fair Credit Reporting Act (FCRA) give consumers and businesses in the US the right to receive a notice of the reasons for an adverse action, such as the decline of a credit application.(51) The notice shall disclose the specific reasons for the rejection. If the adverse

(50) Hurley and Adebayo 2016, 149. “[S]ome data may not appear correlated with race or national origin when used alone but may be highly correlated with prohibited characteristics when evaluated in conjunction with other fields.” Evans 2017, 5.

(51) Both laws were enacted in the 1970s. See with details Ammermann 2013; on gaps and inadequacies of these laws see Hurley and Adebayo 2016, 183-195.

action relied on information obtained from third parties, the consumer has the right to request the information from them. The ECOA and FCRA use broad definitions, including banks, credit bureaus, and FinTechs that operate as marketplace lenders or data aggregators (who provide data services for consumer reports and credit scorings).(52) In the EU, articles 13 to 15 of the GDPR grant consumers equal or even stronger information rights.

As the laws of the US do not require FinTechs to disclose their prediction model in detail, trade secrets are preserved. Still, FinTechs have to open their black box wide enough to let each applicant know why the credit application was rejected and how he or she might be able to improve the probability of a positive decision.(53) In the EU, the proposed AI Act requires that high-risk AI systems must be "sufficiently transparent to enable users to interpret the system's output" (article 13(1)).

Explaining the reasons for a negative action or the interpretation of an AI system's output and its oversight requires knowledge of the deployed decision model,(54) particularly an understanding of the link between the input data and the output of the model, technically called the 'response function'. There is a relationship between the linearity(55) and monotonicity(56) of the response function and the interpretability of a model.

- The degree of interpretability of models with linear and monotonic functions is high.(57) Humans "could go through each step of the algorithm and check if each step is reasonable to them."(58)
- The degree of interpretability of models with nonlinear and monotonic functions is medium. Nonlinearity means that there is no clear

(52) „Some fintech products are designed to avoid consumer protection laws while others claim that existing rules do not apply to them. … Products that claim not to be a loan may be designed to avoid credit laws. Companies that collect and distribute information about consumers may not follow the Fair Credit Reporting Act. Some regulators are rushing to exempt new products from consumer protection laws through regulatory 'sandboxes'." Saunders 2019, 4.

(53) "People do not expect explanations that cover the actual and complete list of causes of an event. We are used to selecting one or two causes from a variety of possible causes as THE explanation." Molnar 2020, 37. For insight from philosophy, cognitive and social sciences for appropriate explanations of models based on AI see Miller 2018.

(54) However: "We do not need to completely understand exactly how a machine-learning algorithm works any more than we need to completely understand the physics of torque to regulate bicycle riding in traffic." Bryson 2020, 8.

(55) Linearity means that the relation between input and output is linear, i.e. an increase of the input value will lead to a proportionate increase of the output value.

(56) Monotonicity means that the relation between input and output is always in one (positive or negative) direction. For example, an increase in the input value will always result in either an increasing or a decreasing output value and no flipping from increasing in some to decreasing in other situations (and back). A monotonic function has no turning points.

(57) "[F]or a change in any given input variable … the output of the response function changes at a defined rate, in only one direction, and at a magnitude represented by a readily available coefficient." Hall and Gill 2019, 15. For example, linear regression algorithms fall into this category. Decision trees are another example for machine learning models with a high interpretability.

(58) Shen 2020.

proportionality between changes in the value of input and output variables. Most machine learning algorithms generate nonlinear functions (for a better fit to nonlinear data). Still, as long as a function is monotonic, the changes of an input variable will always go in only one direction. Models with nonlinear monotonic response functions are less determined but still directional so that they are "fairly interpretable and potentially suitable for use in regulated applications."(59)

- However, the response functions generated by most machine learning algorithms are nonlinear and nonmonotonic. "This class of functions is the most difficult to interpret, as they can change in a positive and negative direction and at a varying rate for any change in an input variable."(60) Models based on deep learning algorithms that autonomously adjust response functions in reaction to the continuous inflow of data in an unpredictable and unobservable way are uninterpretable.

Life-altering decisions of high-risk AI systems in areas such as finance, employment, healthcare, or parole should be based on algorithmic models with high or at least medium interpretability. It seems that many CSMs of FinTechs have difficulties meeting the interpretability requirement. Interpretability can be achieved "by restricting the complexity of the machine learning model (intrinsic) or by applying methods that analyse the model after training (post hoc)."(61) If this is impossible, then life-altering decisions in areas such as finance, employment, or healthcare should not be taken by non-interpretable algorithmic models. The scope of permissible prediction/ decision models would effectively be limited in these areas, but without a ban of the technologies as such.

Regardless of the legal framework, secular ethics considers it a matter of fairness that people have a right to get substantive reasons for an adverse action and challenge model-based decisions in critical (life-altering or high-risk) areas. Contributions from an Islamic perspective could enrich the debates. It could be clarified whether general fairness criteria of the West are compatible with Islamic concepts of fairness or justice. Another Sharīʿah topic is whether the degree of uncertainty of nonlinear and nonmonotonic response functions in decision models violates the prohibition of gharar. A more general question is whether FinTechs that apply Islamic contracts in their financing

(59) Hall and Gill 2019, 15.
(60) Ibid., 16.
(61) Molnar 2020, 25.

business but deploy black-box technologies for decisions on loan applications could be certified as Sharīʿah-compliant from an ethical point of view: black-box decision models are not neutral tools. They may be used (e.g., by FinTechs) with the best of all intentions, but ethical issues still arise because of the inherent lack of interpretability of nonlinear nonmonotonic response functions.

2.5. Autonomous Vehicles

The purpose of an autonomous vehicle is to transport passengers from A to B without interventions by the passenger. Traditional vehicles for transportation without passenger intervention are, for example, taxis and busses. Self-driving cars will be an additional option in the future. To be transported from A to B without own interventions is an intention that does not raise ethical issues. If technology is ethically neutral, there is no reason to raise ethical objections against the use of autonomous vehicles from an Islamic perspective.(62)

However, the assumption of neutrality of technology has been challenged in the West. A transport system composed of a vehicle and a driver can get into situations where he or she must solve a moral dilemma. A widely discussed dilemma is structured as follows: a car's brakes fail, and if it stays on course, pedestrians crossing the street would be killed. Swerving would be possible but crash the vehicle into a wall and kill the passengers. The driver must take the moral decision on whom to save and whom to give up. In traditional vehicles, the moral decision is taken by the driver, who is a human being. In an autonomous vehicle, the driver is a machine. Thus, technology has to make a decision classified as moral when taken by a human being. It may be too strong to say that the machine acts morally, but it is safe to say that the machine's decision must consider ethical principles. These principles have to be embedded in the programming of the autonomous vehicle.

The German government appointed an Ethics Commission on Automated and Connected Driving to develop such principles. The commission produced an Ethics Code with 20 ethical guidelines(63) that shall give orientation to producers and, in particular, programmers of self-driving vehicles as well as legislators and regulators. The common normative ground was humanistic and liberal principles. Nevertheless, there was no complete consensus on all

(62) Legal issues such as the tortious liability of autonomous cars, respectively, their manufacturers/programmers and owners/users have been discussed from an Islamic perspective, e.g. Olalekan 2018.

(63) Ethics Commission 2017, 7: "Its members are drawn from the fields of philosophy, jurisprudence, social sciences, technology impact assessment, the automotive industry and software development". See also Lütge 2017.

guidelines, and ethical discussions continue.

Principles are essential but not sufficient for a judgement on the moral quality of a life-and-death decision. The concrete alternatives in a specific dilemma situation have to be taken into consideration. A massive database with different (historical or engineered) crash scenarios can be built, but it is virtually impossible to anticipate every possible constellation for a dilemma situation. Furthermore, the available sensor technology and processing capacities of the AI of self-driving cars may not be able to identify the programmed scenarios clearly and timely. Therefore, the behavioural guidelines programmed into the control software must be more general.(64)

The bottom line of the ethical principles will be deciding who will be saved and who will be killed - either the passengers or third parties. Concerning the outcome, the self-driving car's software can operate basically in three alternative modes:

- an altruistic mode that will save third parties,
- an egoistic mode that will save the passengers,
- an impartial mode that gives equal importance to third parties and passengers and decides by the number of lives saved.

There are different options for who sets the mode for the car. It could be (1) the government that prescribes, for example, the altruistic mode,(65) or (2) the manufacturer of the car who programs the egoistic mode, assuming that customers would not be willing to buy a car that gives the survival of third parties priority over the passengers.(66) A third alternative has been proposed: to let the driver decide on the operational mode. The driver would have to select one of the three modes before starting the vehicle, and that mode cannot be changed during the trip.(67) This alternative (which was not considered by the German Ethics Commission) should be appealing in a society where the personal freedom of choice has a high value.

(64) This is considered in the guidelines 8 and 9 of the German Ethics Code for Automated and Connected Driving; see Ethics Commission 2017 and Lütge 2017.

(65) The government could be inspired by a sentence in the Guideline 9: "Those parties involved in the generation of mobility risks must not sacrifice non-involved parties." Lütgte (2017) clarifies that this "implies that it cannot be a general rule for software code to unconditionally save the driver. However, the driver's well-being cannot be put at last, either." This points more towards the impartial mode. An even stronger argument against the altruistic mode is given in the explanation of guideline 9 by the Ethics Commission itself: "No obligations of solidarity must be imposed on individuals requiring them to sacrifice themselves for others, even if this is the only way to save other people" (2017, p. 18).

(66) Guideline 12 of the German Ethics Code calls for transparency: "The public is entitled to be informed about new technologies and their deployment in a sufficiently differentiated manner."

(67) Contissa, Lagioia, and Sartor 2017.

The contradicting modes of operation have fundamentally different implications for individual technology users and society at large. Therefore, programming individually and socially acceptable decision rules for dilemma situations cannot be left to programmers alone. The deployment of morally charged technologies requires a thorough multidisciplinary ethical debate. For Muslim countries, participation of Sharīʿah scholars with a deeper understanding of the relevant technologies seems desirable.

The German Ethics Commission could not produce a set of unambiguous and implementable behavioural rules for autonomous vehicles. People may have different preferences for how autonomous vehicles should behave in moral dilemma situations. To assess the individual and social acceptance of laws and regulations of operational modes of self-driving cars, it would be conducive to know more about people's ethical preferences. The following are two examples of studies that have addressed this question.

- A survey in Texas asked students and faculty of a faith-based private university about their preferences for the operational mode of self-driving cars primarily.(68) The researchers added the 'random setting' as a fourth mode by which the altruistic, egoistic, or impartial mode is randomly selected by the car. They found that more than three-quarters of the 284 respondents identified the impartial mode as the most moral of the four alternatives. Being asked whether they would buy a self-driving car that requires the driver to choose the operational mode, the majority of the respondents would not be comfortable with this feature.
- A group of researchers from Harvard and MIT created a kind of online game (the "Moral Machine experiment") "for collecting large-scale data on how citizens would want autonomous vehicles to solve moral dilemmas in the context of unavoidable accidents. The Moral Machine attracted worldwide attention, and allowed us to collect 39.61 million decisions from 233 countries, dependencies, or territories".(69) The players are confronted with different specifications of the moral dilemma (accident scenarios) and have to decide whether the car shall stay on course (which would save the passengers) or swerve (which would save the pedestrians). The scenarios differed, among others, by the number of pedestrians and passengers, their age, gender, and social status, and whether the pedestrians crossed the street legally or

(68) Stephen and Jones 2020.

(69) Awad et al. 2018. The Moral Machine is accessible at https://www.moralmachine.net (as of 28 August 2021).

disregarded a red traffic light.[70] From the decisions in the different scenarios nine preferences for sparing other groups (such as passengers/pedestrians, males/females, elderly/young people, more/fewer people) have been deduced. The strongest preferences include sparing humans over animals and more over fewer lives. For a more differentiated analysis, three regional clusters (Western, Eastern, Southern) were identified with similar preferences of the players within each cluster but significantly different preferences between clusters. For example, players from countries in the Western cluster had a much stronger preference for inaction (i.e. to stay on course) than the players in the Eastern and particularly the Southern cluster.

The Harvard/MIT study identified preferences that are in apparent conflict with the verdict of guideline 9 of the German Ethics Commission: “In the event of unavoidable accident situations, any distinction based on personal features (age, gender, physical, or mental constitution) is strictly prohibited.” However, in many countries, players preferred the young over the old and (especially in countries of the Southern cluster) female over male. This means that people’s moral sentiments can be in conflict with guidelines (or maybe even laws or regulations) based on sophisticated but abstract ethical reasoning.

The Eastern cluster comprises 28 jurisdictions, of which 14 are member countries of the OIC: Iran, Pakistan, Jordan, Palestine, Oman, United Arab Emirates, Egypt, Lebanon, Brunei, Bahrain, Kuwait, Saudi Arabia, Indonesia, Malaysia.[71] The differences between the Western, Eastern, and Southern groups of countries are considerable. One component of an explanation could be religious differences between the clusters. A closer examination of the raw data could be worthwhile but is beyond the scope of this paper. Nevertheless, the findings suggest, as a general conclusion, that religion is an essential source of moral sentiments, but not the only one – especially not in modernising societies where people are exposed to evolving technologies while religious authorities draw their inspirations from historical texts and cases. Many examples can be found in the secularising societies of the Western world. It is time to assess the impact of Sharīʿah authorities on the global

(70) Critics of the Moral Machine (e.g. McDermid 2019) argue that the researchers grossly overestimated the capacities of the (weak) artificial intelligence of a self-driving car: Its AI will never be able to recognise all the specificities that have been loaded into the descriptions of the dilemmas (scenarios) for which a decision has to be taken (for example, the age or social status of passengers and pedestrians). Although this critique is technically correct, it does not address the main purpose of the thought experiment, which is to better understand how citizens want self-driving cars to behave in dilemma situations.

(71) The other 14 countries are Nepal, Armenia, Macedonia, India, Mauritius, Andorra, Cambodia, Japan, Macau, China, South Korea, Taiwan, Thailand, Hong Kong.

discourse on technology ethics (in particular AI ethics) and the evolution of moral sentiments in Muslim societies.

3. Conclusion: Towards Ethical AI

AI is based on statistics and uses data as an input for the creation of models that allow predictions for individual cases. Data are abstractions that reduce the highly complex reality to a manageable dimension. Abstractions – and hence data – serve a particular purpose for which they have been structured and selected. Since selection is linked to a purpose or objective, data sets are not value-neutral. The choice of some and the neglect of other data for the creation of an AI model is a conscious decision for which human beings are responsible. "An ethical use of AI requires that data are collected, processed, and shared in a way that respects the privacy of individuals and their right to know what happens to their data, to access their data, to object to the collection or processing of their data, and to know that their data are being collected and processed and (if applicable) that they are then subject to a decision made by an AI."(72) Some jurisdictions, notably the EU, have added a "right to be forgotten".(73)

Data capture some and ignore other dimensions of reality. Engineers of AI models have to decide which data to use for the training and calibration of models. The selected data sets can have many kinds of flaws and biases. For example, the training data may not be representative of the population to which the AI model will be applied, or they are too old to capture recent behavioural changes (e.g. due to Covid-19). The biases may go unnoticed by the engineers or the model users, but biased input will generate biased output. Furthermore, new faults and still unobserved biases can emerge when unrelated data sets from big data pools are combined in AI models. Biased model outputs can be discriminatory and violate principles of fairness and justice concerning individuals or groups (defined, for example, by gender, ethnicity, religion, age, or disability). Biases are not always obvious, and a systematic check for biases should be good practice before an AI model is deployed. In the EU, the proposed AI Act goes beyond good practice recommendations. It details rather strict criteria regarding the data quality and governance for high-risk AI systems (article 10) and stipulates comprehensive record-keeping (article 12).

(72) Coeckelbergh 2020, 40%. "AI may lead to new forms of manipulation, surveillance, and totalitarianism, not necessarily in the form of authoritarian politics but in a more hidden and highly effective way: by changing the economy in a way that turns us all into smartphone cattle milked for our data." Ibid., 42%.

(73) Article 17 GDPR gives data subjects the right to obtain the erasure of personal data when, among others, the data are no longer necessary for the purpose for which they were collected or the consent to their processing is withdrawn.

To avoid discrimination in areas where it has been observed, laws and regulations can prohibit the use of sensitive data for decisions (e.g. on employment or financing). However, prohibited sensitive data correlate with so many other variables in a big data pool that they can be proxied with high accuracy. Therefore, some regulators restrict or ban the use of autonomous AI models in life-altering or high-risk areas.

A widely accepted fundamental principle for AI systems is fairness in the treatment of affected individuals or groups. This requires the translation of fairness criteria from the verbal into the formal language of statistics to integrate them into the optimisation algorithms of AI models. The operationalisation of fairness needs people who have expertise and authority in ethics (such as ethicists or Sharīʿah scholars) and sufficient knowledge of statistics and AI algorithms - either by themselves or in a team with technical experts.

Programmers define the response function and the restrictions for the optimisation of an AI model, but the optimisation process often takes place in a black box and remains opaque for outside observers. It “is no longer transparent how the AI comes to its decision, and humans cannot fully explain the decision. They know how their system works, in general, but cannot explain a particular decision.”(74) Many people who use AI do not know what it is doing and what effects it has. This “is a problem for responsibility and hence a serious ethical problem.”(75) Ethicists argue that the explanation of decisions of AI systems is “a moral requirement. Explainability is a necessary condition for responsible and accountable behavior and decisions.”(76) This implies that unexplainable (results of) black-box models produced by neural networks should not be deployed by private companies and public bodies that claim to meet ethical standards for decisions that could be life-altering.

The European Commission has taken up such concerns and warnings. It intends to introduce legal provisions regarding the transparency and interpretability of high-risk AI systems (article 13 of the proposed AI Act) with potentially far-reaching implications for the use of AI models of the machine learning black box type

The concerns raised by ethicists also apply, among others, to Islamic financial institutions and FinTechs that deploy opaque AI models. The IsDBI study briefly touches on the ethical dimension of AI models. “First, AI has ethical

(74) Coeckelbergh 2020, 49%.
(75) Ibid., 49%.
(76) Ibid., 52%.

constraints programmed into it. Second, AI weighs inputs in a given ethical framework to choose an action. At the highest level, AI makes ethical judgments and defends the reasoning."(77) When Sharīʿah scholars form an opinion on these issues, they should overcome the view that AI models are just neutral tools like a hammer and ethical issues only follow from the intentions of the human being who wields the hammer. "[M]odels derived directly from data via machine learning are different. They are … often so complex and opaque that even their designers cannot anticipate how they will behave in many situations." (78)

Dealing with ethical issues in AI began in an anecdotal manner. However, in recent years AI ethics(79) began to emerge as a (sub)discipline of applied or practical ethics (such as bioethics, medical ethics, or ethics of public policy).(80) Furthermore, it has attracted the interest of regulators and legislators and intergovernmental institutions who started initiatives to put insights and principles of ethical AI into practice with particular attention paid to the finance industry. Prominent examples are the EU's proposed AI Act and the policy considerations in a recent OECD study.(81)

The EU as an international body aims at global leadership in the debate on values, ethics, and AI. The AI HLEG has articulated the ambitious goal "to make ethics a core pillar for developing a unique approach to AI, one that aims to benefit, empower and protect both individual human flourishing and the common good of society. We believe that this will enable Europe to position itself as a global leader in cutting-edge AI worthy of our individual and collective trust."(82) This global aim should trigger an international debate that would benefit from a response from an Islamic perspective. There may be differences regarding the underlying values as well as the application of AI in particular fields.(83) Since ethical AI in finance is highly relevant both for the West (e.g. the EU and OECD) and the Muslim world (as indicated by the

---

(77) Ashraf et al. 2021, 21. As the study does not focus on AI ethics, it does not elaborate further on how AI makes ethical judgements and defends the reasoning.

(78) Kearns and Roth 2019, 7.

(79) "'AI ethics' is now used to encapsulate a multiplicity of value-based, societal concerns associated with the use of AI applications across an increasingly extensive and diverse range of social and economic activities." Yeung, Howes, and Pogrebna 2020, 78.

(80) See Leben 2019, Kearns and Roth 2019, Coeckelbergh 2020.

(81) See OECD 2021, 56-58. This study mentions roughly the same topics that have already been addressed in the context of the proposed AI Act, including disclosure, transparency and interpretability, and human primacy in decision making.

(82) High-Level Expert Group on Artificial Intelligence 2019, 5.

(83) Considering that the majority of Muslim countries constitute a significant portion of the developing countries of the global South, a debate on AI principles can also lead to political controversies. It has been criticised that AI principles of institutions such as the World Economic Forum or the OECD fail to protect the people of the South and expose them to a "commodification of citizens" and "data colonialism"; see Arun 2020; see also Ashraf et al. 2021, 21-22.

IsDBI study[84]), this could be a good starting point for an in-depth analysis.

- There may be consensus on the wording of ethical principles such as respect for human autonomy, prevention of harm, and fairness, and on requirements for trustworthy AI such as privacy and data governance, diversity and non-discrimination, environmental and societal well-being, and accountability.
- But subtle or even fundamental differences in the interpretation, importance, and implications for policies might become apparent when it comes to the substance of these terms.[85] These differences will be found within the West, the West and the Islamic world, and within individual countries between advocates of ethical AI and proponents of widely unrestricted use of data.[86]

"Codes of ethics are embedded within far wider questions of value, values which may not be explicitly included in the codes themselves, but which are assumed or referenced within wider societal values and norms within which the codes are nested"[87] such as cultural, ideological, or religious value systems. AI will be deployed in more and more domains, and there seems to be a growing demand for ethical guidance and competition for leadership in ethical AI. For the time being, the Islamic perspective is not very well articulated.

(84) See IsDBI 2021, 12 (table 1) and 21-22.

(85) "[T]here is no clear, agreed set of ethical standards within the tech industry. This has resulted in conceptual incoherence, particularly because the norms identified in any given "ethics code" have not typically been rooted in any explicit vision of the kind of political community which those norms are intended to nurture and maintain." Yeung, Howes and Pogrebna 2020, 101. The lamented incoherence could motivate tech-savvy Sharīʿah scholars (or Sharīʿah-savvy technologists) to develop a consistent "AI ethics from an Islamic perspective".

(86) For example, governments in some Muslim countries may have to reconsider whether prevalent American business models – which are based on weak data privacy laws – can be reconciled with Islamic perspectives on fairness and social justice.

(87) Boddington 2020, 139.

References

Data format: Year-Month-Day.
All weblinks last accessed 2021-08-30.

1. Agrawal, Ajay; Gans, Joshua; Goldfarb, Avi (2018): Predictive Machines – The Simple Economics of Artificial Intelligence. Boston, Harvard Business Review Press (eBook).
2. Almutairi, Sattan Eid (2019): The Islamic and Western Cultures and Values of Privacy, in: Muslim World Journal of Human Rights, 16(1), 51-80, https://doi.org/10.1515/mwjhr-2019-0004.
3. Ammermann, Sarah (2013): Adverse Action Notice Requirements Under the ECOA and the FCRA, in: Consumer Compliance Outlook, Second Quarter 2013, https://consumercomplianceoutlook.org/2013/second-quarter/adverse-action-notice-requirements-under-ecoa-fcra.
4. Arun, Chinmayi (2020): AI and the Global South, in: Dubber, Pasquale, and Das (2020), 589-606.
5. Ashraf, Dawood; Khedher, Anis Ben; Moinnuddin, Mohammad; Obaidullah, Mohammed; Ali, Salman Syed (2021): Artificial Intelligence and Islamic Finance – A Catalyst for Financial Inclusion. Jeddah, Islamic Development Bank Institute, https://irti.org/product/artificial-intelligence-and-islamic-finance.
6. Awad, Edmond, et al. (2018): The Moral Machine Experiment, in: Nature 563, 59–64. https://doi.org/10.1038/s41586-018-0637-6.
7. Barfield, Woodrow (ed.) (2021): The Cambridge Handbook of the Law of Algorithms. Cambridge: Cambridge University Press.
8. Bataine, Ahmed Saleh et al. (2020): Toward Monetizing Personal Data: A Two-sided Market Analysis, in: Future Generation Computer Systems, 111, 435-459, https://doi.org/10.1016/j.future.2019.11.009
9. Binns, Reuben (2020): On the Apparent Conflict Between Individual and Group Fairness, in: Conference on Fairness, Accountability, and Transparency, Barcelona, January 27–30, 2020, https://arxiv.org/pdf/1912.06883.
10. Boddington, Paula (2020): Normative Modes – Codes and Standards, in: Dubber, Pasquale, and Das (2020), 125-140.
11. Bryson, Joanna J. (2020): The Artificial Intelligence of the Ethics

of Artificial Intelligence – An Introductory Overview for Law and Regulation, in: Dubber, Pasquale, and Das (2020), 3-25.

12.Caruana, Mireille M.; Cannataci , Joseph A. (2007) European Union Privacy and Data Protection Principles - Compatibility with Culture and Legal Frameworks in Islamic States, in: Information & Communications Technology Law, 16(2), 99-124, http://dx.doi.org/10.1080/13600830701531953.

13.Chang, Victor; Lin, Jing (2018): A Discussion Paper on the Grey Area – The Ethical Problems Related to Big Data Credit Reporting, in: Proceedings of the 3rd International Conference on Internet of Things, Big Data and Security (IoTBDS 2018), pages 348-354, https://www.scitepress.org/papers/2018/68236/68236.pdf.

14.Coeckelbergh, Mark (2020): AI Ethics. Cambridge, MA: MIT Press.

15.Contissa, Giuseppe; Lagioia, Francesca; Sartor, Giovanni (2017): The Ethical Knob - Ethically-Customisable Automated Vehicles and the Law. https://ssrn.com/abstract=2881280.

16.Dubber, Markus D.; Pasquale, Frank; Das, Sunit (eds.) (2020): The Oxford Handbook of Ethics of AI. Oxford and New York: Oxford University Press.

17.Ebers, Martin; Gamito, Marta Cantero (eds.) (2021): Algorithmic Governance and Governance of Algorithms – Legal and Ethical Challenges. Cham, Springer.

18.Ebers, Martin; Navas, Susana (eds.) (2020): Algorithms and Law. Cambridge: Cambridge University Press.

19.Ethics Commission (2017): Automated and Connected Driving – Report, June 2017. Berlin: Federal Ministry of Transport and Digital Infrastructure, https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf.

20.European Commission (2017): Digital Single Market – Communication on Exchanging and Protecting Personal Data in a Globalised World – Questions and Answers, 2017-01-10, https://ec.europa.eu/commission/presscorner/detail/en/MEMO_17_15.

21.European Commission (2018): Communication Artificial Intelligence for Europe. COM(2018)237, 2018-06-26, https://ec.europa.eu/transparency/documents-register/api/files/COM(2018)237_0/de00000

000142394?rendition=false.
22.European Commission (2020): White Paper on Artificial Intelligence – A European Approach to Excellence and Trust. COM(2020) 65 final, 2020-02-19, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri =CELEX:52020DC0065&qid=1629889081836&from=EN.
23.European Commission (2021a): 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, 2021-04-21, https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF.
24.European Commission (2021b): Annexes to the 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, 2021-04-21, https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_2&format=PDF.
25.European Court of Justice (2020): The Court of Justice invalidates Decision 2016/1250 on the adequacy of the protection provided by the EU-US Data Protection Shield, Press Release 91/20, Luxembourg, 2020-07-16, https://curia.europa.eu/jcms/upload/docs/application/pdf/2020-07/cp200091en.pdf.
26.European Union (1992): The treaty of European Union – Consolidated Version, https://eur-lex.europa.eu/resource.html?uri=cellar:2bf140bf-a3f8-4ab2-b506-fd71826e6da6.0023.02/DOC_1&format=PDF.
27.European Union (2000): Charter of Fundamental Rights of the European Union. Official Journal of the European Communities C 364/1, 2000-12-18, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:12012P/TXT&from=EN.
28.European Union (2016): Regulation (EU) 2016/679 of 27 April 2016 [General Data Protection Regulation], Official Journal of the European Union L 119/1, 2016-05-04, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679.
29.Evans, Carol A. (2017): Keeping Fintech Fair- Thinking about Fair Lending and UDAP Risks, in: Consumer Compliance Outlook, Second

Issue 2017, https://www.consumercomplianceoutlook.org/2017/second-issue/keeping-fintech-fair-thinking-about-fair-lending-and-udap-risks/.

30. FTC - Federal Trade Commission (2012): Report to Congress Under Section 319 of the Fair and Accurate Credit Transactions Act of 2003, https://www.ftc.gov/sites/default/files/documents/reports/section-319-fair-and-accurate-credit-transactions-act-2003-fifth-interim-federal-trade-commission/130211factareport.pdf.
31. Federal Trade Commission (2012b): Prepared Statement of the Federal Trade Commission Entitled “Examining the Uses of Consumer Credit Data, https://www.ftc.gov/sites/default/files/documents/public_statements/prepared-statement-federal-trade-commission-entitled-examining-uses-consumer-credit-data/120913creditscorestestimony.pdf.
32. FTC – see Federal Trade Commission
33. Guseyva, Viktoriya (2020): Data Residency Laws by Country - An Overview. InCountry, https://incountry.com/blog/data-residency-laws-by-country-overview/.
34. Hall, Patrick; Gill, Navdeep (2019): An Introduction to Machine Learning Interpretability. 2nd ed., Sebastopol, CA: O’Reilly.
35. Hayat, Muhammad Aslam (2007): Privacy and Islam - From the Quran to Data Protection in Pakistan, in: Information & Communications Technology Law, 16:2, 137-148. http://dx.doi.org/10.1080/13600830701532043.
36. High-Level Expert Group on Artificial Intelligence (2019): Ethics Guidelines for Trustworthy AI. Brussels: European Commission. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419.
37. Hurley, Mikella; Adebayo, Julius (2016): Credit Scoring in the Era of Big Data, in: Yale Journal of Law and Technology, 18 (1), 148-216.
38. Johnson, Steven (2018): Beyond the Bitcoin Bubble, in: The New York Times Magazine, 2018-01-16, https://www.nytimes.com/2018/01/16/magazine/beyond-the-bitcoin-bubble.html.
39. Kearns, Michael; Roth, Aaron (2019): The Ethical Algorithm – The Science of Socially Aware Algorithm Design. New York: Cambridge University Press.

40.Leben, Derek (2019): Ethics for Robots – How to Design a Moral Algorithm. Abingdon: Routledge.
41.Lubis, Muharmaan; Kartiwi, Mira: Privacy and Trust in the Islamic Perspective - Implication of the Digital Age. Conference Paper: 5th International Conference on Information and Communication Technology for the Muslim World. http://doi.org/10.1109/ICT4M.2013.6518898.
42.Lütge, Christoph (2017): The German Ethics Code for Automated and Connected Driving, in: Philosophy & Technology 30, 547–558, https://www.researchgate.net/profile/Christoph-Luetge/publication/320011270_The_German_Ethics_Code_for_Automated_and_Connected_Driving.
43.Manda, Vijaya Kittu; Eskhita, Radwan (2019): Should Islamic Banking & Financial Institutions go with General Data Protection Regulation Compliance?, in: International Journal of Islamic Economics and Finance, 2(1), 109-130, https://journal.umy.ac.id/index.php/ijief/article/download/6387/4434.
44.McDermid, John (2019): Self-driving Cars - Why We Can't Expect Them to be 'Moral'. The Conversation, 2019-01-24, https://theconversation.com/self-driving-cars-why-we-cant-expect-them-to-be-moral-108299.
45.Mehrabi, Ninareh et al. (2019): A Survey on Bias and Fairness in Machine Learning. https://arxiv.org/pdf/1908.09635.pdf.
46.Miller, Tim (2018): Explanation in Artificial Intelligence – Insights from the Social Sciences, https://arxiv.org/pdf/1706.07269.pdf.
47.Minoli, Daniel; Occhiogrosso, Benedict (2018): Internet of Things Applications for Smart Cities, in: Hassan (2018), 319-358.
48.Mobile Ecosystem Forum and Juniper Research (2016): Whitepaper – Understanding the Personal Data Economy. https://mobileecosystemforum.com/wp-content/uploads/2016/11/Understanding-the-Personal-Data-Economy-Whitepaper.pdf.
49.Molnar, Christoph (2020): Interpretable Machine Learning - A Guide for Making Black Box Models Explainable. https://leanpub.com/interpretable-machine-learning.
50.OECD (2021): Artificial Intelligence, Machine Learning and Big Data in Finance: Opportunities, Challenges, and Implications for Policy

Makers, https://www.oecd.org/finance/artificial-intelligence-machine-learningbig-data-in-finance.htm.

51. Olalekan, Omoola Sodiq (2018): Autonomous Vehicles and Tortious Liability – An Islamic Perspective, in: Jurnal Syariah 26(1), 99-122, https://ejournal.um.edu.my/index.php/JS/article/view/13258/8362.
52. Raval, Tony (2019): Data Privacy as a Basic Human Right. Forbes, 2019-11-12, https://www.forbes.com/sites/forbestechcouncil/2019/11/12/data-privacy-as-a-basic-human-right/.
53. Renda, Andrea (2021): Moral Machines - The Emerging EU Policy on "Trustworthy AI", in: Barfield 2021, 667-690.
54. Saunders, Lauren (2019): Fintech and Consumer Protection – A Snapshot. National Consumer Law Center, https://www.nclc.org/images/pdf/cons-protection/rpt-fintech-and-consumer-protection-a-snapshot-march2019.pdf
55. Shen, Owen (2020): Interpretability in ML - A Broad Overview, The Gradient, 2020-11-21, https://thegradient.pub/interpretability-in-ml-a-broad-overview/.
56. Singleton, Alex David; Longley, Paul (2015): The Internal Structure of Greater London - A Comparison of National and Regional Geodemographic Models, in: Geo 2(1), 69-87, https://rgs-ibg.onlinelibrary.wiley.com/doi/epdf/10.1002/geo2.7.
57. Stephen, Kiana; Jones, Beata M. (2020): The Effect of Religiosity on Decision Making in Self-Driving Cars – The Case of "The Ethical Knob", in: Issues in Information Systems, 21, (1), 74-90, https://iacis.org/iis/2020/1_iis_2020_74-90.pdf.
58. The Paypers (2020): Global Open Banking Report 2020. Amsterdam, https://thepaypers.com/reports/reportdownload/the-global-open-banking-report-2020-beyond-open-banking-into-the-open-finance-and-open-data-economy/cid=1244913.
59. Thirani, Vasudha; Gupta, Arvind (2017): The Value of Data. World Economic Forum, 2017-09-22, https://www.weforum.org/agenda/2017/09/the-value-of-data/.
60. Vincent, Sarah St. (2018): Data Privacy Is a Human Right – Europe Is Moving Toward Recognizing That. Foreign Policy in Focus, 2018-04-19, https://fpif.org/data-privacy-is-a-human-right-europe-is-moving-

toward-recognizing-that/.

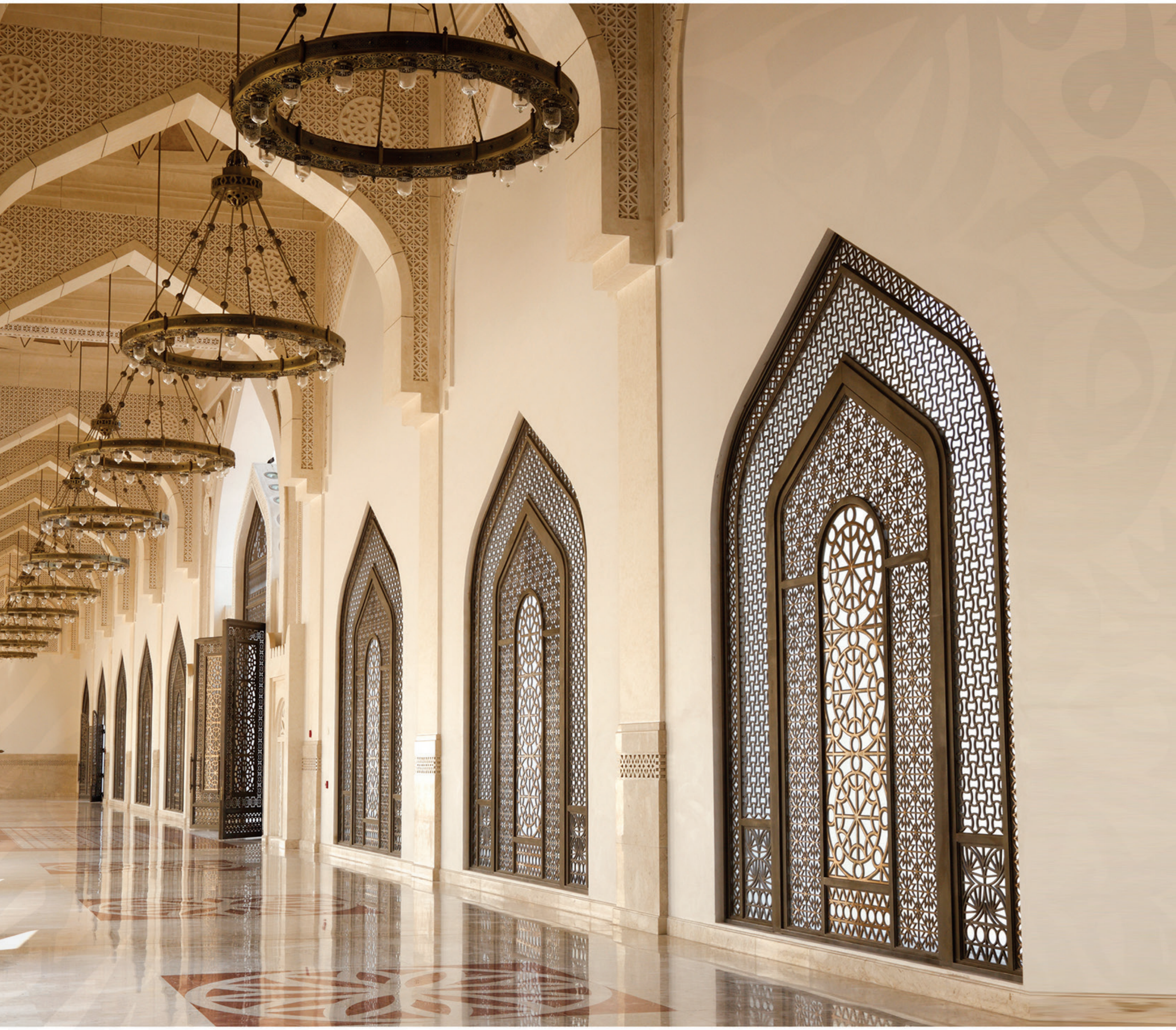
61. WebFX Team (2020): What Are Data Brokers – and What Is Your Data Worth?, 2020,03,16, https://www.webfx.com/blog/internet/what-are-data-brokers-and-what-is-your-data-worth-infographic/)
62. Yeung, Karen; Howes, Andrew; Pogrebna, Ganna (2020): AI Governance by Human Rights-Centered Design, Deliberation, and Oversight – An End to Ethics Washing, in: Dubber, Pasquale, and Das (2020), 77-106.
63. Zhong, Ziyuan (2018): A Tutorial on Fairness in Machine Learning. Towards Data Science, 2018-10-22, https://towardsdatascience.com/a-tutorial-on-fairness-in-machine-learning-3ff8ba1040cb.

Bait Al-Mashura Journal

مجلة بيت المشورة

International Academic Refereed Journal On Islamic Economics and Finance

Issue (16) State of Qatar - October 2021



Published by

Bait Al-Mashura Finance Consultations

ISSN : 2409-0867 إلكتروني

ISSN : 2410-6836 ورقـــي

https://doi.org/10.33001/M01102021issue16

mashurajournal.com