Regulatory Practices Lab

THE CENTRE FOR internet & society

Artificial Intelligence: a Full-Spectrum Regulatory Challenge

A Policy Brief By **SUNIL ABRAHAM**

RESEARCH TEAM

Programme Officers | **SHWETA MOHANDAS, MIRA SWAMINATHAN AND SHWETA REDDY**

Interns | **VISHNU RAMACHANDRAN, CHANDRAMANI SIROTHIA, AND GURSIMAR SETIA**

Design | **SAUMYAA NAIDU**

The Centre for Internet and Society, India

WORKING DRAFT

(Formatting Credits: Akash Sheshadri)

Shared under

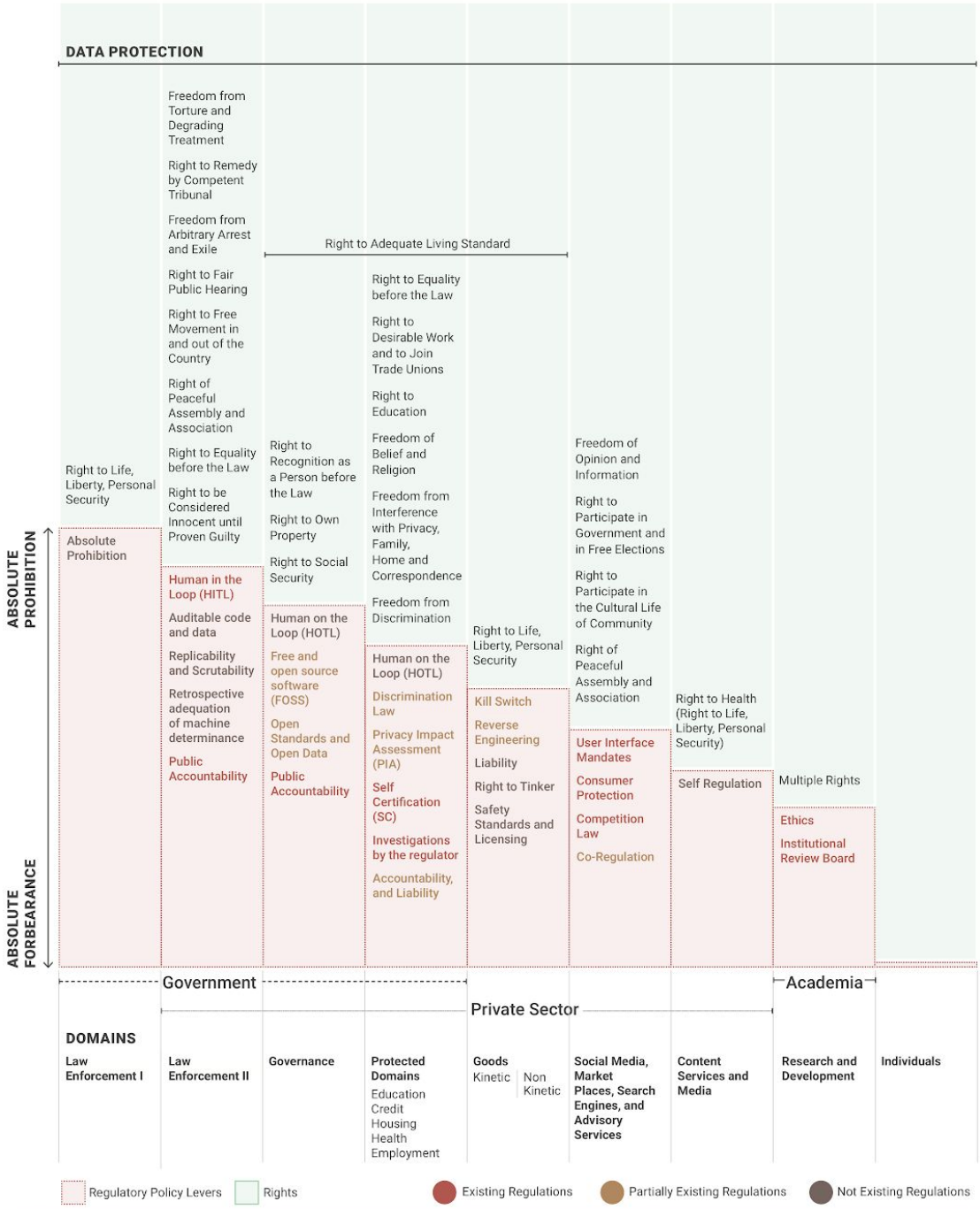
Creative Commons Attribution 4.0 International license

Contents

Acknowledgement

The author would like to thank Nalini Bharatula, Malavika Raghavan, Beni Chug, Mansi Kedia, Ujjwal Krishna, Isha Suri, Alok Prasanna Kumar and Divij Joshi for their inputs on the infographic and the policy brief. The authors would also thank Arindrajit Basu, Amber Sinha and Anubha Sinha for insights that helped modify the policy brief. The policy brief immensely benefited from all these insights but any errors and inaccuracies remain the authors' alone.

Artificial Intelligence Regulatory Spectrum



Designed by **Saumyaa Naidu**
Shared under
Creative Commons Attribution 4.0 International license

Regulatory Practices Lab

THE CENTRE FOR internet & SOCIETY

Introduction

The above infographic is a visual representation of the AI Regulatory Spectrum. The infographic highlights the existing regulation, partially existing regulations and non existing regulations between the government and private sector ranging on a scale from absolute prohibition to absolute forbearance. Each domain has various policy levers within the same and are also overlooked by the various rights of adequate living standards. The various domains also are regulated by the data protection and fundamental rights regimes.

Scope: Definition of AI, any unlawful use of AI, regulation of production and development of AI, societal and environmental impacts of AI and other negative externalities of AI are outside the scope of this policy brief. Use of AI by the military is not included in the scope of the policy brief as the author does not possess the necessary subject matter expertise.The author intends for the policy brief to be considered as a straw man framework to promote more granular discussions around the topics outlined in the policy brief. Even though this brief focussing on the hard regulatory agenda it does not discount the role of ethics when it is adopted in good faith by the private sector to go beyond the regulatory minimum.

Data Protection: AI systems are mandated to comply with the data protection provisions in accordance with their jurisdiction of operation. Due to the nature of the AI systems and their ability to learn from training data and subsequently address certain future behaviours may cause compliance with principles related to purpose limitation, data minimisation and storage limitation problematic. However, data governance model incorporating anonymisation techniques, comprehensive and clear privacy notices, privacy by design and privacy impact assessments[1] need to be adopted for an AI system to function within the current data protection frameworks.

[1] Information Commissioner's Office(2017), Big Data, Artificial Intelligence, Machine Learning and Data Protection, ICO, Retrieved From
https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf

Law Enforcement - I

Absolute Prohibition

In case of Facial recognition in the Government's case, absolute prohibition of centralised mass surveillance systems is a more efficient policy, rather than expecting them to comply with stringent regulations. Some of the biggest companies in the world have entered into contracts with governments to implement facial recognition amidst pushback from activist groups.[2]However, it is not practical to give the military, intelligence and law enforcement wings of the government powerful technologies and then expect them to comply with stringent regulations with respect to their application. Hence, a more elegant policy solution would be an absolute prohibition on centralised mass surveillance systems. Facial recognition can only be used in a democratised fashion when people are using in their private capacity.

[2] Abril Daniellel (2019) Coalition Pressures Amazon, Microsoft, and Google to Keep Facial recognition Surveillance away from Government. Fortune. Retrieved from: **https://fortune.com/2019/01/15/coalition-pressures-amazon-microsoft-google-facial-recognition-surveillance-government/**

Law Enforcement - II

Human In The Loop (HITL), Auditable code and data, Replicability, Scrutability, Retrospective Adequation,and Public Accountability

For all other uses of AI in law enforcement[3] where an absolute prohibition would be over regulation - Human In The Loop(HITL) mandates would be required as the potential for rights infringement is high and it will enable holding officials accountable for their actions. For example for and the authorization for a kill shot in a weapons systems, HITL would require the human commands to flow to the weapon for specific actions.[4] A similar approach called "The Human in Command" has been suggested by the European Economic and Social Committee[5] which requires individuals to be able to retain control over the machines at all times.[6] Given national security concerns[7] and also given potential for misuse - auditable code

[3] For the purposes of this policy brief, law enforcement includes those agencies that use non lethal force in response to internal disturbances and tensions or riots.

[4] Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the use of Certain Conventional Weapons (2018) . Report of the 2018 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System. United Nations Office At Geneva . Retrieved From **https://www.unog.ch/80256EDD006B8954/(httpAssets)/20092911F6495FA7C125830E003F9A5B/$file/CCW_GGE.1_2018_3_final.pdf**. Pg.12
Caution must be exercised while using the terms "in the loop" or "on the loop" due to the lack of clarity in the definitions. It has been suggested that apart from the final decision, there are a number of micro level decisions where the human operator may have an opportunity to question the command. The time delay in fielding the autonomous system, the decision to engage a weapon against a target and the impact time are few of the important factors that needs to be considered. It is recommended that the hardware that is used to implement the human in the loop should be as similar to the operational hardware as possible. HITL should be implemented in cases where decisions of the government are dependent on human accountability for proper functioning to ensure that the liability of the decision remains with the responsible individuals and not machines.
IEEE. (2017). Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems,The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Version 2.Retrieved from **http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html**, p.115, p.123, p.125

[5] European Economic and Social Committee(2017) Artificial Intelligence - The Consequences of Artificial Intelligence on the (digital) single market, production, consumption, employment and society , European Economic and Social Committee Retrieved from **https://www.eesc.europa.eu/our-work/opinions-information-reports/opinions/artificial-intelligence**

[6] "Human in Command" - This refers to the capability of the human to oversee the overall activity of the AI system (including its broader economic, societal, legal and ethical impact) and the ability to decide to and when to use these systems. This would not only include the ability to oversee the general functioning of the system but also to override the decisions made by the system.

[7] On May 7, 2019, the city of Baltimore was targeted with a ransomware attack, demanding approximately 100,000 USD in Bitcoin to decrypt the malware. The attack impeded access to government email accounts (approximately 10,000) and impacted online payments to city departments. Citizens have not been able to pay utilities, parking tickets, and other related municipal taxes.
Sullivan Emily (2019) Ransomware Cyberattacks on Baltimore Put City Services Offline NPR Retrieved From **https://www.npr.org/2019/05/21/725118702/ransomware-cyberattacks-on-baltimore-put-city-services-offline**

and data mandates will be an inappropriate regulatory response and therefore should be replaced by an internal audit requirements for both code and data.

While the FOSS, open data and open standards mandates in the welfare section below are meant to prevent discriminations and exclusion harms - here the harms can impact a plurality of rights therefore the audits must be able to guarantee replicability, and scrutability. For an AI system to be replicable, it must return exactly the same answer if it is asked the same question twice. For an AI system to be scrutable, a human must be able to explain how the AI works.[8] There are different proposals around scrutability, explainable AI, retrospective adequation of machine inferences to human reasoning, etc.[9]

Explainable AI can be understood as the ability of an AI to formulate for the user of the AI a line of reasoning that explains the decision making process of a model using human-understandable features of the input data.[10] Holzinger et al classifies two types of explainability in AI post-hoc explainability - occurring after the event in question and ante-hoc explainability - explainability occurring before the event in question, in other ways by incorporating explainability directly into the structure of an AI-model, explainability by design.[11] Retrospective adequation of machine inferences to human reasoning is a proposal from CIS which proposes that the AI provide a human readable explanation along with the decision produced through machine learning, so that the human is able to apply her mind and review the decison of the machine.[12]

[8] IEEE. (2017). Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems,The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Version 2.Retrieved from **http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html**, pg.71
[9] Sinha, A. (2019). Use of Sentiment Analysis by Law Enforcement. Unpublished manuscript
[10] Doran, D., Schulz, S., & Besold, T. R. (2017). What Does Explainable AI Really Mean? A New Conceptualization of Perspectives. Retrieved from **https://arxiv.org/pdf/1710.00794.pdf**, p. 2.
[11] Holzinger, A., Biemann, C., Pattichis, C., & Kell, D. (2017, December). What do we need to build explainable AI systems for the medical domain?. Retrieved from **https://arxiv.org/pdf/1712.09923.pdf**, p. 4.
[12] Sinha, A. (2019). Use of Sentiment Analysis by Law Enforcement. Unpublished Manuscript
Transparency in the predictive policing tool can be enhanced by posting the results of such policing after they have been utilized by the police. McCarthy James Odhran (2019) AI & Global Governance: Turning the Tide on Crime with Predictive Policing Centre for Policy Research, United Nations University Retrieved from **https://cpr.unu.edu/ai-global-governance-turning-the-tide-on-crime-with-predictive-policing.html**)
Similar transparency mechanisms for algorithmic outputs in recidivist assessments used in the criminal justice system shall aid the judges in determining the level of reliance on the outputs predicted.
Angwin Julia, Larson Jeff, Mattu Surya and Lirchner Lauren (2016). Machine Bias Pro Publica Retrieved from **https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing**

Governance

Human On The Loop, Free and Open Source Software, Open Standards and Open Data; and Public Accountability

This section covers governance excluding law and order functions of the government. Given that welfare delivery happens at scale and that the power differential between citizens and welfare officials is less that between citizens and police/military, HITL would be an inappropriate and onerous requirement. Human On The Loop [HOTL] is lower accountability standard where unlike HITL, the AI will produce outcomes without requiring action on the part of the human. However, the human will be kept informed and will have the option to override machine decisions and actions.[13] This will enable citizens to pin responsibility for exclusion or discrimination in welfare delivery on specific government officials ensuring "skin in the game". Citizens should be able to request for human intervention and an explanation for the decisions undertaken[14]. If welfare is being provided by private sector contractors because of public private partnerships, then again it is critical that citizens be able to blame specific individuals. In other words, AI should be used to make the welfare state more transparent to the citizen and not vice versa.

Since the early 2000s, the global free software movement has been calling for all software used in e-governance including welfare be made available under a FOSS license. Separately to prevent vendor lock-in, governments have realized that e-governance software must comply with relevant open standards, and the training data should be available under an open data license (OD). These mandates will enable both public and independent audits. In many countries across the globe there are national FOSS and open data policies notified or enacted in the 2000s.[15] However the FOSS policies were rendered ineffective as the software model shifted from COTS to SAAS. These policies need to be updated requiring the GNU Affero General Public License so the software paradigm is not used to circumvent regulation.

[13] IEEE. (2017). Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems,The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Version 2.Retrieved from **http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html**,pg.115
[14] Recital 71 of General Data Protection Regulation. (2016). Retrieved from **https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN**
[15]Lewis, J. A. (2010). Government open source policies. Center for Strategic and International Studies. Retrieved from **https://www.csis.org/analysis/government-open-source-policies**

All the regulatory obligations described above should also extend to welfare programmes that are being run using private public partnerships.[16]

Open data is both required to ensure that the training data sets that were used to create the AI models are accurate, verified, free from bias and free from exclusion errors - however there is a completely distinct role that open data can provide to contribute to greater public transparency and accountability.

Protected Domains

Human on the Loop, Discrimination Law, Privacy Impact Assessment, Self Certification, Regulatory Investigations in Protected Domains, Accountability and Liability

The large number of people and transactions within the protected domains make it impossible to implement HITL, therefore Human on The Loop should be used to provide citizens with a person to hold accountable and liable.

Unlike the government, private-sector actors should not be required to make their AI source code or training data publicly available under FOSS and OD licenses, respectively. This is because trade secrets are necessary for competitive advantages in the market. Firms should therefore not be required to meet the replicability and scrutability standard. AI uses in these areas should require Discrimination Law, Privacy Impact Assessment (PIA), subsequent Self-Certification (SC) and investigations by the regulator if discrimination is strongly suspected[17]

Discrimation Law

Anti-discrimination laws differ widely in various regions of the world[18] with additional

[16] Citron, D. K. (2007). Technological due process. Wash. UL Rev., 85, 1249. Retrieved from **https://openscholarship.wustl.edu/cgi/viewcontent.cgi?article=1166&context=law_lawreview**

[17] IEEE. (2017). Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems,The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Version 2.Retrieved from **http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html**, p. 70.

[18] For Instance: In the US, there is an expansive legislation of anti-discrimination provisions, which are backed by strong precedents. The first major anti-discrimination law in the US was the Civil Rights Act of 1964 added protections as to color, religion, sex, or national origin in the areas of voting, education, employment, and public accommodations. Altman, A. (2017) Civil Rights. The Stanford Encyclopedia of Philosophy (Winter 2017 Edition). Retrieved from **https://plato.stanford.edu/archives/win2017/entries/civil-rights/**.

protection available in some countries based on national priorities.[19] The enforcement of anti-discrimination laws vary widely due to the cultural barriers, the predominance of personal laws, and at times the laws[20] itself promote illegal discrimination in violation of constitutional guarantees. Where missing, sectoral laws should be enacted for protected domains which can then be enforced by existing or new sector specific regulators.[21]Self certification is usually the bare minimum required by the regulator.[22] Regulated entities wishing to go beyond the legal minimum and set higher ethical standards for themselves can adopt an intersectional approach[23] Against this, there are proposed methods for identifying various combinations of protected attributes and certifying fairness across 'exponentially many' subgroups. Under special circumstances the regulator or regulations may require evaluation, monitoring and auditing by an independent third party.[24]

Privacy Impact Assessment

Any non-compliance with regulations and risks to privacy rights can be modeled and addressed by conducting privacy impact assessments of the complete life cycle of data processing.[25] Mitigating measures to reduce such risks can be implemented post the result of the assessment.

Regulatory investigations in protected domains

The private sector applications of AI should be regulated depending on how they may discriminate against protected classes, create physical harm or impact political organizing

[19] For Instance: There are some countries whose national laws fail to define different forms of discrimination which leaves it to the courts to set the law with respect to EU directives. There are also problems with the effective implementation and application of the anti-discrimination law, as well as the procedural barriers to enforce these laws. Chopin, I., & Germaine, C. (2017, November). A Comparative Analysis of Non-Discrimination Law in Europe. Europe: European Commission. Retrieved from **https://publications.europa.eu/en/publication-detail/-/publication/36c9bb78-db01-11e7-a506-01aa75ed71a1**.

[20]In some of these countries personal laws and statutory laws are discriminatory towards homosexuals and women, disabled people and ethnic minorities. For the study cited the global south is considered to be these countries: Kenya, Bangladesh and Nepal; among lower middle income countries, it examines India, the Philippines and Zambia; and among upper middle income countries, it examines South Africa, Botswana, Brazil, Jamaica and the Czech Republic. Fredman, S. (2012). Anti-discrimination laws and work in the developing world: A thematic overview. Retrieved from **https://openknowledge.worldbank.org/bitstream/handle/10986/12129/WDR2013_bp_Anti-Discrimination_Laws.pdf?sequence=1&isAllowed=y.**

[21] For example, Under S. 1691c of the Equal Credit Opportunities Act, 1974, the administrative bodies that regulate the sector are stated in order for them to enforce the requirements of this legislation.The Federal Trade Commission has the overall enforcement authority with respect to the provisions of the Act.

[22] An example of Employment Agency Self Certification which is in compliance with Employment Agency Laws of USA **https://www1.nyc.gov/assets/dca/downloads/pdf/businesses/Employment-Agency-Self-Certification.pdf**

[23] "Intersectionality" - Intersectionality here is not a matter of randomly combining infinite variables to see what 'disadvantages' fall out; rather, it is about mapping the production and contingency of social categories. Hoffmann, A. L. (2019). Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse. Information, Communication & Society, 22(7), 900-915.p. 906.

[24] Fredman, S. (2012). Anti-discrimination laws and work in the developing world: A thematic overview. Retrieved from **http://siteresources.worldbank.org/EXTNWDR2013/Resources/8258024-1320950747192/8260293-1320956712276/8261091-1348683883703/WDR2013_bp_Anti-Discrimination_Laws.pdf**

[25] Article 35 General Data Protection Regulation (2016) Retrieved from **https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN**

and public safety. It may be possible to demand key subsets of information about the algorithms such as the variables that are in use, the training data, standard deviations of the results produced. [26]In some situations, regulators may lift the corporate veil if misuse or harm become clear.[27] In such cases, the results of the algorithmic impact assessment of the application can be shared with the regulators who shall be bound by the principle of confidentiality[28]

Accountability

The government in consultation with industry groups should establish standards for the creation, and retention of logs for a specified period of time and indirect public disclosure using privacy protecting mechanism like zero knowledge proofs to enable higher degree of trust.[29] AI has to be traceable in order to be trustworthy. Granular logging will help assign accountability when harms are caused. Traceability allows the identification of reasons for errors as well as facilitates audits and investigations.[30]This is similar to black boxes that have been installed in planes which helps hold various stakeholders accountable during the use of complex technological systems with multiple points of possible failure.[31] Liability

The principle of strict liability will apply in this domain as the injurer, in this case any of the service provided by the protected domain, is at a better state to determine the costs of risk associated with their actions.[32] Further, there is a sense of responsibility expected out of the members in this domain.

[26] Committee Of Experts on Internet Intermediaries(MSI-NET) (2018), Algorithms and Human Rights Council of Europe Retrieved from **https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5** Pg. 38

[27] For Instance: Traditionally, protected domains have included education, credit, housing, health, and employment in which people should not be discriminated against. Protected classes in these domains may include race, gender, sexual orientation, and other such categories.
World Economic Forum Global Future Council on Human Rights (2016- 2018). (2018, March). How to Prevent Discriminatory Outcomes in Machine Learning. World Economic Forum. Retrieved from **http://www3.weforum.org/docs/WEF_40065_White_Paper_How_to_Prevent_Discriminatory_Outcomes_in_Machine_Learning.pdf**, p. 8.

[28] Koene Ansgar, Clifton Chris, Hatada Yohko, Webb Helena, Patel Menisha, Machado Caio, LaViolette Jack, Richardson Rashida, Reisman Dillon (2019) Governance framework for algorithmic accountability . Panel for the Future of Science and Technology ,European Parliament Retrieved from **https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf** p. 73

[29] IEEE. (2017). Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems,The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Version 2.Retrieved from **http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html**, p. 159, p. 160.

[30]IEEE. (2017). Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems,The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Version 2.Retrieved from **http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html**, p. 44.

[31] Jourova, V. (2019, January 21). [Tweet] VeraJourova. Retrieved from **https://twitter.com/VeraJourova/status/1087376736654893057**.

[32] Omri Rachum-Twaig, 'Whose Robot Is it anyway?: Liability For Artificial Intelligence Based Robots' University of Illinois Law Review 2020 U. ILL. L. REV

Goods (Kinetic and Non Kinetic)

Kill Switches, Reverse Engineering, Liability, Right to Tinker,Safety Standards and Licensing

Kill Switches

How do we deal with AI based robots that have gone rogue? The European Union is considering mandatory implementation of kill switches in kinetic AI[33] goods. Researchers recommend that kill switches be thoroughly debugged after it is built and not be tampered with while adding new features to the machine.[34] Additionally, the kill switches should be placed in such a way that the operators can engage it to shut down the system.[35]

Liability

There are many hands involved in the creation and implementation of AI.[36] They include the programmer the practitioner and the user. Therefore from an enforcement perspective it is simplest for government to retain a strict liability regime that is common for all other non-AI goods. In the case of harm[37] a strict liability regime holds the manufacturer solely liable if the defences[38] do not stand legal scrutiny. Further, providing insurance to the users by the manufacturers reduces the chilling effect on innovation that such a liability might have by derisking compensation payouts for the possible damages.[39] Once, technologies mature,

---

[33] "Kinetic AI" - It can be described as a physical instantiation of AI that may produce physical, kinetic damage, such as an automated vehicle, movable robot.

[34] Li, H., & Yang, S. X. (2003). A behavior-based mobile robot with a visual landmark-recognition system. IEEE/ASME transactions on mechatronics, Retrieved from **https://www.researchgate.net/publication/3414006_A_behavior-based_mobile_robot_with_a_visual_landmark-recognition_system** pg. 395

[35] Christen, M. B. (2017). An Evaluation Schema for the Ethical Use of Autonomous Robotic Systems in Security Applications. University of Zurich. University of Zurich Digital Society Initiative White Paper Series. Retrieved from **https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3063617**, p. 48.

[36] Nissenbaum, H. (1996). Accountability in a Computerized Society. Retrieved from **http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.40.77**, p. 4.

[37] For Instance: Self-driven cars, automated factory robots, automated drones. These AI enabled products can be potentially dangerous to the public if they malfunction. For example, if an automated assembly line robot in a factory misbehaves due to an AI failure and ends up killing a factory worker.
IFLScience. (n.d.). Robot Kills Worker at Volkswagen Plant in Germany. IFLScience. Retrieved from **https://www.iflscience.com/technology/robot-kills-worker-volkswagen-plant-germany/**.

[38] Act of God, Negligence from the side of the user, Assumption of risk etc.

[39] For Instance: California's Assemblyman Mike Gatto announced a comprehensive solution to the challenges that drones pose, The Drone Registration/ Omnibus Negligence-prevention Enactment (DRONE) Act of 2016. This proposed bill requires physical or electronic license plates for drones to hold them responsible for any accident, insurance for the purpose of compensating the victim in case of an accident and GPS capability that features automatic shut-off technology that would activate if approaching an airport.

more graduated regimes of liability can be developed that distribute liability between manufacturers and sub-contractors based on predetermined standards of care.[40] For high risk AI goods, victims could get compensated in an expedited fashion if law requires the private manufacturers to advance the funds to government in an escrow arrangement.

Reverse Engineering

The right to reverse engineering for the purposes of security audits, interoperability and for the purposes of building competing devices is guaranteed in jurisdictions such as India[41]. However, once the this right and the right below have been exercised, the provider and subcontractor can no longer be held liable or culpable in case of harms.

Right to Tinker and Repair

The right to reverse engineer under copyright law is limited to software[42]. This does not provide the consumer or her agent with the right to study and modify hardware. Therefore, the concept of freedom to tinker[43] should be extended to all AI machines and devices to enable individuals, academicians and civil society organisations to audit machines thoroughly. However, like with reverse engineering the exercise of this right voids liability and culpability guarantees from the manufacturer.

Safety Standards

AI in the physical world should "meet specific internal and external safety requirements" that are "determined by AI and safety specialists, businesses and civil society organisations collectively."[44]Another method of ensuring safety has been stated to be the use of

---

CBS Local, Los Angeles. (2016, January 14). Drone Legislation Would Require Owners To Buy Insurance, Get UAV 'License Plates'. Retrieved July 5, 2019 from **https://losangeles.cbslocal.com/2016/01/14/drone-legislation-would-require-owners-to-buy-insurance-get-uav-license-plates/**.

[40] Rachum-Twaig, Omri, Whose Robot Is It Anyway?: Liability for Artificial-Intelligence-Based Robots (February 21, 2019). University of Illinois Law Review, Vol. 2020, Forthcoming . Available at SSRN: **https://ssrn.com/abstract=3339230**

[41] Copyright Office, Government of India. (1957, June). Section 52 (ab and ac) of The Copyright Act, 1957. Retrieved from **http://www.copyright.gov.in/Documents/CopyrightRules1957.pdf**, Chapter XI.

[42] The Indian Copyright Act is the only legislation that allows for reverse engineering of software. The Indian Copyright Act under Section 52 (ab and ac) exempts reverse engineering of software for operating interoperability, or for the purpose of observing, studying or testing the functioning a computer programme in order to determine the ideas and principles which underlie any elements of the programme while performing such acts necessary for the functions for which the computer programme was supplied.
Copyright Office, Government of India. (1957, June). Section 52 (ab and ac) of The Copyright Act, 1957. Retrieved from **http://www.copyright.gov.in/Documents/CopyrightRules1957.pdf**, Chapter XI.

[43] "Freedom to Tinker" - Computer scientist Edward Felten defines Freedom to Tinker "your freedom to understand, discuss, repair, and modify the technological devices you own. He also states that "this freedom is more than just an exercise of property rights but also helps to define our relationship with the world as more and more of our experience is mediated through these devices.
Samuelson, P. (2016, January 07). Freedom to Tinker. Berkeley Law (Scholarship Repository). Retrieved from **https://scholarship.law.berkeley.edu/cgi/viewcontent.cgi?article=3633&context=facpubs**, p. 565.

[44] "Internal Safety" & "External Safety" - Internal safety entails the need to ensure that AI systems use robust and well programmed algorithms which are resistant to hacking. External safety entails the need to ensure that the system operates safely in both normal and unexpected situations, before the system is deployed. The EESC states that the AI systems should only be used if they conform to internal and external safety requirements.

geo-fencing and automatic shut down functions in case of the machine proceeds to harm humans or property.[45]

Licensing

AI based civilian drones could be regulated by requiring physical or electronic license plates to hold details about the owner so that they can be held responsible for any accident. Similar to other vehicles, It should be mandatory for the owners to have an insurance policy covering damage to life, injuries, interference, or property damage.[46]

European Economic and Social Committee. (2017). Opinion of the European Economic and Social Committee on 'Artificial intelligence — The consequences of artificial intelligence on the (digital) single market, production, consumption, employment and society'. 526th EESC Plenary Session of 31 May and 1 June 2017. Retrieved from **https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52016IE5369&from=EN**, p. C 288/4.

[45] Bot Disclosure and Accountability Act of 2018, S.3127, 115th Congress, 2nd Session. (2018). Retrieved from **https://www.congress.gov/bill/115th-congress/senate-bill/3127/text**.

[46] For Instance: California's Assemblyman Mike Gatto announced a comprehensive solution to the challenges that drones pose, The Drone Registration/ Omnibus Negligence-prevention Enactment (DRONE) Act of 2016. This proposed bill requires physical or electronic license plates for drones to hold them responsible for any accident, insurance for the purpose of compensating the victim in case of an accident and GPS capability that features automatic shut-off technology that would activate if approaching an airport.
CBS Local, Los Angeles. (2016, January 14). Drone Legislation Would Require Owners To Buy Insurance, Get UAV 'License Plates'. Retrieved July 5, 2019 from **https://losangeles.cbslocal.com/2016/01/14/drone-legislation-would-require-owners-to-buy-insurance-get-uav-license-plates/**.

Social Media, Market Places,Search Engines and Advisory Services

User Interface Mandates - Data Protection Law, Consumer Protection Law, Competition Law and Co-regulation

User Interface Mandates should be introduced under Data Protection laws, some harms to the integrity of our networked public sphere thanks to “echo chambers” and “filter bubbles” can be addressed by providing interfaces that empower the user to understand the objectives that the AI is maximizing for and provides them with options to change it.[47] This could be understood as statutorily mandated personalization interfaces. For example, a heterosexual person whose wall is prioritizing hetereonormative content should be notified that this is the case and provided with the option to prioritize content preferred by sexual minorities. There are different user empowerment mandates when it concerns public sphere [48] and private sphere.[49]

Several changes may be required to Consumer Protection laws. To begin with non-paying users should get all the protections available under these laws. This will allow that all traditional consumer rights can also be exercised against Social Media, Market Places, and Search Engines. Finally some additional rules/regulations may be required for AI services. AI services should disclose to the user when they are interacting with an AI, be it through voice or text [50] to ensure that the consumers are not mislead.

[47] A 2013 study of Facebook and LinkedIn found that users were no longer able to fully control the production, selection and distribution of Personal Identifiable Information.
Heyman, R., & Pierson, J. (2013). Blending mass self-communication with advertising on Facebook and LinkedIn: Challenges for social media and user empowerment. International Journal of Media & Cultural Politics, 9(3), pp. 229-245.
The right to have the conceptual model of the AI exposed to the user with the power to change the optimisation configuration. This means that the user will be able to know the metrics that are used to provide the information to his feed, as well as tweak the metrics to see how his feed would be with different metrics.This would provide the user with the option to break through the filter bubble that the algorithm has personalised for them.
[48] A platform like Twitter could be a public sphere if the assessment metrics of a public sphere is applied. Liu Z., Weber I. (2014) Is Twitter a Public Sphere for Online Conflicts? A Cross-Ideological and Cross-Hierarchical Look. In: Aiello L.M., McFarland D. (eds) Social Informatics. SocInfo 2014. Lecture Notes in Computer Science, vol 8851. Springer, Cham)
[49] The paper makes the assumption that platforms like Netflix should be considered as a private sphere as the users cannot engage with each other on any public discussion.i
[50] In the United States the Bot Disclosure and Accountability Act of 2018 mandates public disclosure of software programs intended to impersonate or replicate human activity. The draft legislation also requires that users of social media that employ automated software intended to impersonate or replicate human activity to provide

Emerging competitors are unable to take on incumbent e-commerce and social media monopolies because they don't have an essential facility - user data. Lack of competition in this area can result in monopoly power exacerbating harms. Competition Law can be used to prevent the "winner takes all" phenomenon which is a consequence of the network effect. Competition law rules and regulations for AI can mandate interoperability[51]so that the alternative service provider can also feature data and content from the incumbent service provider. Competition law enforcement authorities will need to be vigilant to scrutinise emerging forms of anti competitive behaviour such as algorithmic collusion.[52]

Co-regulation: Possible harms to public health, democracy and social well being can be prevented using co-regulation of Artificial Intelligence that works on algorithms in user's social media use. Thus a system wherein the intermediaries and the supervisors create a code of conduct for their functioning which can then be accredited by an independent regulator.[53] In this system a punitive action can be taken against the intermediaries for not following the code of ethics as decided upon by them and accredited by the regulator.

Co-Regulation with Advisory Services. Artificial intelligence has changed the way the users gather information as they now seek advice through various services.[54] In order to ensure the timely, comprehensive, and bias-free insights, a system of co-regulation can be

---

clear and conspicuous notice of the automated program in clear and plain language to the user of the social media interacting with the bot.
Bot Disclosure and Accountability Act of 2018, S.3127, 115th Congress, 2nd Session. (2018). Retrieved from **https://www.congress.gov/bill/115th-congress/senate-bill/3127/text**.
Similarly, The California bill, B.O.T. Act of 2018 proposes to make it unlawful "for any person to use a bot to communicate or interact with another person in California online with the intent to mislead the other person about its artificial identity for the purpose of knowingly deceiving the person about the content of the communication in order to incentivize a purchase or sale of goods or services in a commercial transaction or to influence a vote in an election." The bill requires the disclosure to be "clear, conspicuous, and reasonably designed to inform persons with whom the bot communicates or interacts that it is a bot". California Senate Bill No. 1001 Bots: Disclosure. (2018). Retrieved from **https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180SB1001**.

[51] "Interoperability" - When two platforms become interoperable, they become more open: users of platform A can interact with platform B's users.
Eisenmann, T. R., Parker, G., & Alstyne, M. V. (2008, August 31). Opening Platforms: How, When and Why? Harvard Business School. Retrieved from **https://pdfs.semanticscholar.org/34ff/d7516eafd11553a2d3b40f33c71b965f3e84.pdf**, p. 6.

[52] Deng, Ai (September 16, 2018), What Do We Know About Algorithmic Tacit Collusion? . Antitrust, Retrieved from: **https://ssrn.com/abstract=3171315** or **http://dx.doi.org/10.2139/ssrn.3171315**

[53] Article 40 of GDPR states The Member States, the supervisory authorities, the Board and the Commission shall encourage the drawing up of codes of conduct intended to contribute to the proper application of this Regulation, taking account of the specific features of the various processing sectors and the specific needs of micro, small and medium-sized enterprises.
Further Article 41 says: The monitoring of compliance with a code of conduct pursuant to Article 40 may be carried out by a body which has an appropriate level of expertise in relation to the subject-matter of the code and is accredited for that purpose by the competent supervisory authority.

[54] The consultancy services include everything from legal advice to medical advice and these affect our everyday lives as customers, employees, partners, and investors. Further, leaders are starting to use artificial intelligence to automate mundane tasks such as calendar maintenance and making phone calls **https://www.ibm.com/in-en/services/artificial-intelligence**

implemented[55] The businesses that deploy chatbots should have internal policies in place that govern the permitted activities of the chatbot, the extent of information that is fed to the chatbot and the methods by which information is provided, data collection and processing by the chatbot, the maintenance of the chatbot and the monitoring of chatbot activities and the possible triggers for human intervention.[56] Further, any other action which is beyond these internal policies can be governed by the concerned professional body.

Content Services and Media

Self Regulation

Due to the minimal degree of harm in the sphere of passive data consumption, a system of self regulation is sufficient . The framework should consist of a set of guiding procedures that address the mental and physical health repercussions caused due to addictive behaviour of the individual, issues related to social manipulability and polarization of view points[57] and potential claims of discrimination[58] a consequence of the recommendations provided. Some of the key best practices that may be considered to determine the appropriateness of the recommendation are periodic analysis of the training data requirements to determine the nexus between the user data being processed and recommendations provided by the system [59], providing consumers an option to submit feedback regarding the recommendation and maintain logs of the same to determine the overall efficiency of the system.

[55] Susan G. Hadden (Nov., 1986), Intelligent Advisory Systems for Managing and Disseminating Information, Public Administration Review, Vol. 46, Special Issue: Public Management Information Systems , pp. 572-578

[56] Kherk Ying Chew AI and its legal impacts- Chatbots Legal Bytes **https://s3.amazonaws.com/documents.lexology.com/de0a15dd-6ff1-4273-be1e-6826b196f125.pdf?AWSAccessKeyId=AKIAVYILUYJ754JTDY6T&Expires=1569307580&Signature=dsVcNvF5ovKxuQsoglmO%2FTYgYB8%3D**

[57] Milano, Silvia & Taddeo, Mariarosaria & Floridi, Luciano. (2019). Recommender Systems and their Ethical Challenges. SSRN Retrieved from **https://ssrn.com/abstract=3378581**

[58] Lara Zarum (2018) Some Viewers Think Netflix Is Targeting Them by Race. Here's What to Know New York Times Retrieved From **https://www.nytimes.com/2018/10/23/arts/television/netflix-race-targeting-personalization.html**

[59] Larson, M., Zito, A., Loni, B., & Cremonesi, P. (2017). Towards Minimal Necessary Data: The Case for Analyzing Training Data Requirements of Recommender Algorithms. In FATREC Workshop on Responsible Recommendation Proceedings (pp. 1-6) **https://doi.org/10.18122/B2VX12**

Research and Development

Ethics

Institutional Review Boards

Unlike medical and behavioral research involving humans, AI research affects a large number of people and at times the harms might not be apparent immediately on the use. The existing regulation pertaining to the functioning of the Institutional Review Boards excludes AI research since the vast majority of such research uses data that is either considered to be a part of the public available or de-identified data set and the lack of direct interaction with the individuals.[60] The common privacy and anonymization safeguards might be insufficient when multiple disconnected public or de-identified data sets are analyzed together. The nature of AI research is such that measuring harm caused on the basis of a direct intervention in the lives of the individual may result in a regulatory gap that could undermine public trust. [61] Approval of the Board must be sought annually throughout the duration of the project, in addition to the approval sought prior to the initiation of the project, to account for a fresh analysis of the harms caused at the various stages of the project. IRBs should engage in discussion on substantive ethics and regulations while approving or rejecting a research project and not just rely on boilerplate language replacements or procedural instructions in the proposal stage[62]. Disclosing the decision to the general public at large will provide valuable guidance to the researchers in the field.

Individuals

[60] Department of Health and Human Services(2015) Notice of proposed Rule Making: Federal Policy for the Protection of Human Subjects Federal Register Retrieved from **https://s3.amazonaws.com/public-inspection.federalregister.gov/2015-21756.pdf**

[61] Metcalf Jacob, Crawford Kate (2016) Where are human subjects in Big Data research? The emerging ethics divide Big Data and Society Retrieved from **https://journals.sagepub.com/doi/full/10.1177/2053951716650211**

[62] Justin T.Clapp Katharine A.Gleason StevenJoffe (2017) Justification and Authority in Institutional Review Board Letters Social Science and Medicine Retrieved from **https://www.sciencedirect.com/science/article/pii/S0277953617306184**