Journal of Marketing & Social Research

ISSN (Online): 3008-0711

Volume: 02 | Issue 02 | Mar-Apr. | 2025
Journal homepage: https://jmsr-online.com/

Research Article

The ethics of AI business practices: A Literature Review on AI Ethics guidelines and Governance Principles with reference to AI Technology firms (Business Organizations).

Ms. Nidhi Kshirsagar[1], Dr. Ekta Rokade[2] and Dr. Kapil Rokade[3]
[1]Assistant Professor, IPS Academy, Institute of Business Management and Research, Indore, India
[2]Associate Professor, School of Economics, Devi Ahilya Vishwavidyalaya, Indore, India
[3]Visiting Faculty, Devi Ahilya Vishwavidyalaya, Indore, India

Received: 03/03/2025; Revision: 21/03/2025; Accepted: 09/04/2025; Published: 28/04/2025

***Corresponding author:** Ms. Nidhi Kshirsagar (harikumar@miu.edu.in)

Abstract: Today we see AI companies and startups are proliferating world over, AI technology has a promising future and has become part of our daily lives, however the ethical risks associated with AI products and services however have sparked debate amongst the industry, organizations, developers and policy makers. This paper is an attempt to understand the prominent AI ethics guidelines that the organizations and their development team needs to keep in mind throughout the AI lifecycle. This paper uncovers through recent policy documents and academic literature the expectation and challenges faced by the business firms with respect to frameworks, tools and resources available while addressing the ethical issues and concerns of AI Business practice. This review study has examined 5 global guidelines highlighting the major AI ethics principles and 24 academic articles have given insights into how firms wish to strike a balance between their business goals and ethical responsibilities by including long-term social impact measures alongside short-term performance indicators.

Keywords: AI Ethics, AI ethics business practice, AI Ethics framework, Ethical AI guidelines.

INTRODUCTION

The proliferation of AI technology has sparked significant discussions about the ethical and governance principles that should guide its development and deployment. The much talked about ethical risks such as Biasness, Data Privacy breaches, Deepfakes, Lack of transparency, accountability and safety concerns are all under consideration under various frameworks. Instances of chatbots using abusive language with people, selective selection of specific gender/cast/creed for job opportunities or financial loans, or even criminal judgements are only a few examples of AI assisted systems that are created by Giant firms such as Google and Microsoft. These business organizations and firms faced a major backlash due to the AI products they created hampering their image and costed them hefty legal compensations. Thus, it would be apt to understand which ethical lens and ethical frameworks these companies uses of should use to make this upcoming and expanding technology safe and socially acceptable .

IMPORTANT THEORETICAL FRAMEWORKS FOR ETHICAL AI

Many organisations have produced statements of the values or principles that should guide the development and deployment of AI products and services in society. We here have discussed the 5 important theoretical frameworks that have been proposed to address the ethical challenges posed by AI at global level

1. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations (AI4People): "AI4People Institute was launched in February 2018 as a pioneering research/policy project by Professor Luciano Floridi, Michelangelo Baracchi Bonvicini, Robert Madelin and Tony Blair to shape the debate on AI Ethics in the European Union and prompt European institutions to act quickly to stem future AI risks in years where no government had yet addressed this crucial issue. Its action is at the origin of the regulatory process that led to the "AI Act in Europe", the world's first AI regulation. The mission is to ensure that artificial intelligence (AI) benefits and serves humanity as a whole. AI4People offer 20 concrete recommendations tailored to the European context which, if adopted, would facilitate the development and adoption of AI that maximises its opportunities, minimises its risks, and respects the core ethical principles identified." (https://eismd.eu/featured/ai4peoples-ethical-framework-for-a-good-ai-society/). The framework proposes following principles :

   (a) Beneficence: promoting well-being, preserving dignity, and sustaining the planet. This principle proposes that AI technology must be in line with the basic preconditions for life on our planet, continued prospering for mankind and the preservation of a good environment for future generations. The principles of beneficence firmly underlines the central importance of promoting the well-being of people and the

planet.

(b) Non-maleficence: privacy, security and "capability caution". This principle proposes of not doing harm to the people at the same time doing only good for the people and society should be the underlying principle. Privacy is characterised as being intimately linked to individuals' access to, and control over, how personal data is used. This infact is an important part of human rights. This principle focuses on avoiding accidental overuse and deliberate misuse that can cause harm arising from whether from the intent of humans or the unpredicted behaviour of machines.

(c) Autonomy: the power to decide (whether to decide). Autonomy means Individuals have a right to make decisions for themselves about the treatment they do or not receive. In the context of AI means striking a balance between the decision-making power we retain for ourselves and that which we delegate to artificial agents. This principle suggests that humans should always retain the power to decide which decisions to take, exercising the freedom to choose where necessary, and ceding it in cases where overriding reasons, such as efficacy, may outweigh the loss of control over decision-making.

(d) Justice: promoting prosperity and preserving solidarity. This principle is based on bioethics and explains that using AI should be to correct past wrongs such as eliminating unfair discrimination, ensuring that the use of AI creates benefits that are shared (or at least shareable) and preventing the creation of new harms, such as the undermining of existing social structures. AI technologies themselves should benefit and empower as many people as possible and also contribute to global justice.

(e) Explicability: enabling the other principles through intelligibility and accountability. Principles such as transparency, accountability, interpretability, explainability are for AI to be beneficent and non-maleficent, we must be able to understand the good or harm it is actually doing to society, and in which ways; for AI to promote and not constrain human autonomy, our "decision about who should decide" must be informed by knowledge of how AI would act instead of us; and for AI to be just, we must ensure that the technology – or, more accurately, the people and organisations developing and deploying it – are held accountable in the event of a negative outcome, which would require in turn some understanding of why this outcome arose.

2. OECD Ethical Principles : "The OECD AI Principles were initially adopted in 2019 and updated in May 2024. Adherents updated them to consider new technological and policy developments, ensuring they remain robust and fit for purpose. The Principles guide AI actors in their efforts to develop trustworthy AI and provide policymakers with recommendations for effective AI policies." (https://oecd.ai/en/ai-principles)

(a) "Inclusive growth, sustainable development and well being : Stakeholders should proactively engage in responsible stewardship of trustworthy AI in pursuit of beneficial outcomes for people and the planet, such as augmenting human capabilities and enhancing creativity, advancing inclusion of underrepresented populations, reducing economic, social, gender and other inequalities, and protecting natural environments, thus invigorating inclusive growth, well-being, sustainable development and environmental sustainability.

(b) Respect for the rule of law, human rights and democratic values, including fairness and privacy : AI actors should respect the rule of law, human rights, democratic and human-centred values throughout the AI system lifecycle. These include non-discrimination and equality, freedom, dignity, autonomy of individuals, privacy and data protection, diversity, fairness, social justice, and internationally recognised labour rights. This also includes addressing misinformation and disinformation amplified by AI, while respecting freedom of expression and other rights and freedoms protected by applicable international law. To this end, AI actors should implement mechanisms and safeguards, such as capacity for human agency and oversight, including to address risks arising from uses outside of intended purpose, intentional misuse, or unintentional misuse in a manner appropriate to the context and consistent with the state of the art.

(c) Transparency and explainability : AI Actors should commit to transparency and responsible disclosure regarding AI systems. To this end, they should provide meaningful information, appropriate to the context, and consistent with the state of art, to foster a general understanding of AI systems, including their capabilities and limitations, to make stakeholders aware of their interactions with AI systems, including in the workplace, where feasible and useful, to provide plain and easy-to-understand information on the sources of data/input, factors, processes and/or logic that led to the prediction, content, recommendation or decision, to enable those affected by an AI system to understand the output, and, to provide information that enable those adversely affected by an AI system to challenge its output.

(d) Robustness, security and safety : AI systems should be robust, secure and safe throughout their entire lifecycle so that, in conditions of normal use, foreseeable use or misuse, or other adverse conditions, they function appropriately

and do not pose unreasonable safety and/or security risks. Mechanisms should be in place, as appropriate, to ensure that if AI systems risk causing undue harm or exhibit undesired behaviour, they can be overridden, repaired, and/or decommissioned safely as needed. Mechanisms should also, where technically feasible, be in place to bolster information integrity while ensuring respect for freedom of expression.

(e) Accountability: AI actors should be accountable for the proper functioning of AI systems and for the respect of the above principles, based on their roles, the context, and consistent with the state of the art. To this end, AI actors should ensure traceability, including in relation to datasets, processes and decisions made during the AI system lifecycle, to enable analysis of the AI system's outputs and responses to inquiry, appropriate to the context and consistent with the state of the art. AI actors, should, based on their roles, the context, and their ability to act, apply a systematic risk management approach to each phase of the AI system lifecycle on an ongoing basis and adopt responsible business conduct to address risks related to AI systems, including, as appropriate, via co-operation between different AI actors, suppliers of AI knowledge and AI resources, AI

(a) system users, and other stakeholders. Risks include those related to harmful bias, human rights including safety, security, and privacy, as well as labour and intellectual property rights."[1]

(b) **3.Ethics for AI:** UNESCO produced the first-ever global standard on AI ethics – the 'Recommendation on the Ethics of Artificial Intelligence' in November 2021. It is applicable to all 194 member states of UNESCO. The protection of human rights and dignity is the cornerstone of the Recommendation, based on the advancement of fundamental principles such as transparency and fairness, always remembering the importance of human oversight of AI systems.[2]

(a) Respect, protection and promotion of human rights and fundamental freedoms and human dignity : protection and respect to human dignity at all times, non discrimination, enhance quality of life for all.

(b) Environment and Ecosystem Flourishing : AI must support environmental sustainability, internatopnal and national laws, reduce carbon footprints, and promote sustainable use of resources.

(c) Ensuring Diversity & inclusion

(d) living in peaceful, just and interconnected societies

(e) Proportionality and do no harm

(f) safety & security

(g) fairness & non-discrimination

(h) sustainability

(i) Right to privacy and data protection

(j) human oversight and determination

(k) transparency & explainability

(l) responsibility & accountability

(m) Awareness & literacy

(n) Multi-stakeholder and adaptive governance and collaboration

3. Ethics Guidelines for Trustworthy AI (High-Level Expert Group on Artificial Intelligence)[3]:

(o) Human agency and oversight: AI systems should empower human beings, allowing them to make informed decisions and fostering their fundamental rights. At the same time, proper oversight mechanisms need to be ensured, which can be achieved through human-in-the-loop, human-on-the-loop, and human-in-command approaches

(p) Technical Robustness and safety: AI systems need to be resilient and secure. They need to be safe, ensuring a fall back plan in case something goes wrong, as well as being accurate, reliable and reproducible. That is the only way to ensure that also unintentional harm can be minimized and prevented.

(q) Privacy and data governance: besides ensuring full respect for privacy and data protection, adequate data governance mechanisms must also be ensured, taking into account the quality and integrity of the data, and ensuring legitimised access to data.

(r) Transparency: the data, system and AI business models should be transparent. Traceability mechanisms can help achieving this. Moreover, AI systems and their decisions should be explained in a manner adapted to the stakeholder concerned. Humans need to be aware that they are interacting with an AI system, and must be informed of the system's capabilities and limitations.

(s) Diversity, non-discrimination and fairness: Unfair bias must be avoided, as it could could have multiple negative implications, from the marginalization of vulnerable groups, to the exacerbation of prejudice and discrimination. Fostering diversity, AI systems should be accessible to all, regardless of any disability, and involve relevant stakeholders throughout their entire life circle.

(t) Societal and environmental well-being: AI systems should benefit all human beings, including future generations. It must hence be ensured that they are sustainable and environmentally friendly. Moreover, they should take into account the environment, including other living beings, and their social and societal impact should be carefully considered.

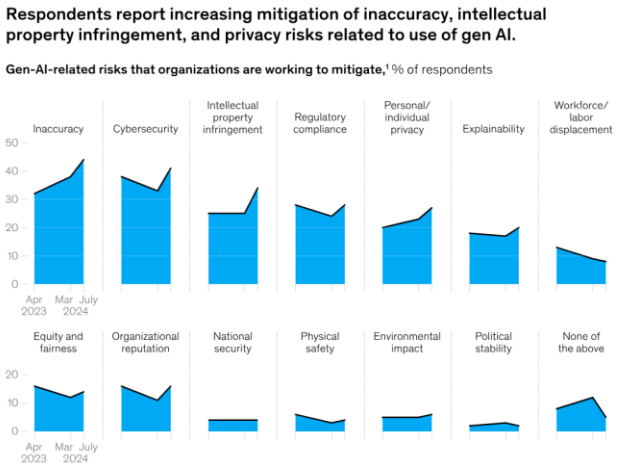
(u) Accountability: Mechanisms should be put in place to ensure responsibility and accountability for AI systems and their outcomes. Auditability, which enables the assessment of algorithms, data and design processes plays a key role therein, especially in critical applications. Moreover, adequate an accessible redress should be ensured.

4. **Ethically Aligned Design. A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, Version 2[4] :**

i. Human Rights : A/IS must respect, fulfill, ensure freedom, human dignity, respect cultural diversity, be safe and secure and bring no harm to humans
ii. Prioriting Wellbeing : For A/IS technologies to provably advance benefit for humanity define and measure the benefit that is to be increased whereas the negative unintended consequences must be avoided.
iii. Accountability : The programming, output, and purpose of A/IS are often not discernible by the general public. Based on the cultural context, application, and use of A/IS, people and institutions need clarity around the manufacture and deployment of these systems to establish responsibility and accountability, and avoid potential harm. Additionally, manufacturers of these systems must be able to provide programmatic-level accountability, if necessary apportion culpability among several responsible designers, manufacturers, owners, and/or operators, to avoid confusion or fear within the general public.
iv. Transparency : transparent A/IS are ones in which it is possible to discover how and why a system made a particular decision, or in the case of a robot, acted the way it did. Note that here the term transparency also addresses the concepts of traceability, explicability, and interpretability.
v. A/IS Technology Misuse and Awareness of It : designers to anticipate, reflect, and engage with users of A/IS thus, through education and awareness, citizens, lawyers, governments, etc. have a role to play in developing accountability structures.

CURRENT STATE ON AI FIRMS' ETHICAL PRACTICES

The online survey was conducted by McKinsey between July 16 to July 31, 2024, and garnered responses from 1,491 participants in 101 nations representing the full range of regions, industries, company sizes, functional specialties, and tenures. Forty-two percent of respondents say they work for organizations with more than $500 million in annual revenues. McKinsey's this survey report "The State of AI 2025" report suggests that the organizations are all up and willing to implement the Gen AI technologies to push their bottom lines, also they are more than willing to incorporate the redesigned workflows, hierarchy channels, additional critical roles in order to mitigate AI related risks, according to report companies with more than $500 annual revenue are more interested to take the leap than the smaller ones. Business have now started to feel that better AI governance practices would assure better profit lines. The report illustrates that the companies are actually making efforts to cut down the ethical risks that associate with their Gen AI outputs



Source:
https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

REVIEW OF LITERATURE

Vakkuri, V., Kemell, K. K., & Abrahamsson, P. (2019). This paper aims to develop a framework that would bring the gap between AI ethics practical implementation with that of industry there are three key principles transparency accountability and responsibility which needs to be embedded in the AI life cycle by the software companies in the paper the emphasis is made on health care sector how they are applying these principles for a dressing social ethical impact the finding of the study reviews that for bringing transparency developers can incorporate tresabel feature implementation and clear documentation for decision making purposes for accountability practices related to error handling and data security measures are adopted for responsibility aspects of data sensitivity and fairness are considered though they are inconsistently applied. The major challenges that the practitioners face are available as tools and resources to formalise the ethical principles de-prioritisation of
ethical goals because of business freshers and resource constraint and limited mechanisms to cater to socio-ethical impacts of AI systems. Authors recommend that their should be structured ethical guidelines which incorporate accountable t responsibility and transparently transparency principle within the work flow of the organisation also regular reviews should be conducted by the organisation along with bringing ethical awareness and training programs for the development teams they should also allocate enough resources so that ethical practices can be formalised and monitored Canca, C. (2020). In this paper

the authors have criticized the qualification of ai ethics principle however these are important and valuable considerations but are insufficient for solving complex ethical dilemmas does the propose of framework through which these principles can be operationalise and distinguished on the basis of core principles such as autonomy beneficial and justice and the instrumental principles which are to protect the values such as transparency and accountability the authors emphasize the importance of context sensitive tools and to develop a checklist for guiding the developers and organisations to address ethical challenges posted by the AI systems. Author suggest that these core principles autonomy refers to giving the respect to individual freedom beneficent means maximizing the benefits and minimising the harm to the society and justice which ensures fairness and equity among all stakeholders the instrumental principles of transparency suggest that their needs to be a clarity in decision making process accountability with respect to responsibility of outcomes of these systems and privacy with respect to save garden the personal information of users. Author suggests that it is important that companies must create customise principles that align with their values and priorities and also develop ethical frameworks which adaptor specific contacts such as healthcare or criminal justice where priorities like safety and fairness vary, they also recommend that organisations develop systems to guide ethical trade off and they invest in ethics activities to resolve complex dilemmas that principle based frameworks cannot address alone.

Kim, T. W., Hooker, J., & Donaldson, T. (2021). The authors have referred here and recommended a hybrid approach where machine learning and logic both should be used for incorporating ethical reasoning into the AI systems. A hybrid framework would bring empirical flexibility and normative consistency also author suggest that there is a greater need for interdisciplinary collaboration between the computer scientist and ethicist to refine the ethical principles this will make a systems context sensitive as well as ethically consistent.

Mayer, A. S., Haimerl, A., Strich, F., & Fiedler, M. (2021) .As AI technologies rapidly evolve, there is growing pressure on private companies to act responsibly. This has led to the rise of AI ethics, a field still developing in both theory and practice. While many organizations publish AI ethics guidelines, it's unclear how effective these are in shaping employee behavior. This study explores how companies in Germany encourage ethical AI practices among employees. Based on expert interviews, it identifies five key implementation strategies: ethics guidelines, employee involvement, organizational integration, practical support, and risk management. The findings offer a useful framework for companies and future research on applying AI ethics in real-world settings.

Miller, G. J. (2021, September). Authors in this paper have identified 5 he success factors for an AI project which includes project governance product quality usage quality benefits and protections and societal impacts full stop the objective of the paper is to translate the ai ethical principles into actionable project deliverables with success the authors have tried to understand how project management and ai ethics can be interwind to achieve sustainable and equitable outcomes. There is a framework which will help moral decisions of air decision making across development usage and consequences. The ethical principles of beneficence fairness and privacy project management practices along with stakeholder involvement have been emphasized to draw project success stakeholders' satisfaction positive societal impacts and ethical compliance. In this paper the authors have conducted a systematic literature review of 144 articles across multiple data bases and conferences. The success factors of project governance includes scope definition risk assessment ethics policies and divers team compositions the product quality includes data transparency equitable representation algorithmic fairness and system security usage quality includes stakeholders centric communication human intervention and decision accountability benefits and protection would include legal safeguards cost efficiency intellectual property Management with positive societal impact such as production of civil freedom sustainability and positive environmental considerations. The authors recommend that project managers should infuse ethical consideration into the project scope and deliverable they should include responsibility matrices to mitigate risk and align rolls with ethical goals for developers authors recommend building a transparent interpretations and equitable air systems with divorce perspective and design and testing further for policy makers it is recommended that the developed regulations that balances both innovation as well as ethical compliance and ensure the mechanism that monitors and audits AI system post deployment for positive socio ethical impacts.

Morley, J., Elhalal, A., Garcia, F., Kinsey, L., Mökander, J., & Floridi, L. (2021). In this paper the current gap between abstract AI principles and their practical implementation is assessed this paper proposes ethics as a service a framework which is inspired by clay cloud computings as a service model in order to operationalise ethical principles of AI practically in this model the responsibility of ethical AI implement distributed among independent ethics board as a responsibility AI practice nurse use translation tool to incorporate ethics guidelines. This paper highlights limitations of current ai ethics his frameworks and how ethical air governance be distributed among the stakeholders. The paper suggests that government structure flexibility of ethical principles and contextual adaptation results into ethical compliance system accountability and society impact. The key findings suggest that currently the translation tools like adaptability or our insufficient and providing practical guidance for AI implementation further there is a need for taylormade Ethical principles with respect to context implementation tools and document decisions. This model provide the structure and flexible approach for embedding ethics across the AI life cycle and ensure that there is enough practices autonomy for an option of ai ethical principles.

Bessen, J., Impink, S. M., & Seamans, R. (2022, July). In this paper a survey of 225 AI startups was conducted between January to March 2021 where investigation was made to understand ethical policy adoption, the various costs associated with adoption and specific action taken in order to align the organisation with the ethical principles. The findings of the study show that 58% of Tata have incorporated and adopted ethical AI principals also where the firm's are more likely to have data sharing relationship or have prior GDP are experience would easily adopt to ethical air principles. The business cost associated with ethical AI includes dropping of bios data turning down the business proposition and even firing the employees however understood in bias training, hiring of diverse talent and sourcing represented data sets include additional cost further, resource limitation often poses a constraint in adhering to ethical guidelines. Study also review that larger and better funded forms are more likely to adopt costly ethical practices hence the recommendation is made that a scalable ethical framework that balance is the business viability along with the others of the principles be made also the policy maker should support startups and companies which want to other practices investors should encourage the startup to priority development as they part of long term business strategies and provide funding and mentally start of navigate the cost of ethical adherence.

Ciobanu, A. C., & Mesnita, G. (2022, March). AI Ethics for Industry 5.0-From Principles to Practice. In I-ESA Workshops. This paper has identify the framework that would be usable for operationalising AI ethics in the context of industry 5.0. the idea is to emphasize human centred designs which emphasis on human machine cooperation. The proposed framework has two different layer first being AI embedded ethics by design and the second is AI desired state configuration which ensure continuous ethical alignment throughout the AI life cycle. The objective of the paper are to explore the role of AI and how it can support human centric balls in industry 5.0 also to address challenges that are occurring in operationalising ethical principles in ai systems. Principles such as autonomy fairness explicability that are based on European commission's ethics guidelines for trustworthy AI have to be embedded into the AI systems in this paper authors have proposed the framework which is AI embedded ethics by design where the developers use dashboards to train the AI models with ethical principles from the starting of the AI software life cycle they continuously test to mitigate the risk and ensure compliance before the deployment of the final AI products also AI desire state configuration framework is propose which should be used for deployment for monitoring and feedback to loop in the continuous feedback and updates into the non-technical stakeholders can contribute to assessing and refining the AI systems. The paper provides valuable findings that human machine collaboration is the key to innovation and creativity and to achieve sustainable Technology advancements however the systems are context dependent and its framework should be Taylor for specific environment and culture social dimension should be integrated with ethical principles. There is a need for continuous feedback mechanism for bringing in changes as per the context all these will benefit organisations in industry 5.0 to align themselves with ethical principles write from development to deployment and post deployment phase and would most collaboration among various stakeholders

Deshpande, A., & Sharp, H. (2022, July). In this paper several complexities have been identified which the practitioners face while trying to engage stakeholders and building responsible air system the authors have referred to more than 178 guidelines which normally the builders struggle to translate in their practices for the this suggest that applying ISO 26000:2010 framework for social responsibility a structured and indicator long list of stakeholders can be considered throughout the air life cycle full stop authors have referred to 45 Academy articles that guide developers policy makers and organisations in better integrating their stakeholders into ai ethics design and governance the research questions a rest in this paper are who are the stakeholders which are impacted by responsible asy systems and which stakeholders should address this impacts. The key finding suggests that there are three levels of stakeholders: individual stakeholders, organisational stakeholders and national or international stakeholders. Under individual stakeholders, users, non-users, developers' engineers, designer, non-AI experts, data subjects can be considered whereas at organisation level technology companies Research Institute professional bodies and NGOs that advocate ethical AI should be incorporated at National and international level government National agency International regulatory bodies and standardization organisations can be incorporated.

Sloane, M., & Zakrzewski, J. (2022, June) .Authors suggest that the social practice theory which has five interlinked elements principles, needs, narratives, ethics material and cultural genealogy are deeply embedded in organizational routine and are shaped by cultural norms in German AI technology startups. They moreover take this as a collective responsibility. The authors have conducted semi structured interviews with 64 professionals with AI startups. The key finds reveal that AI ethics is just not understood in words but are practiced with responsibility and codetermination which are enacted through concrete organizational behaviours. Ethics is shared concept and is translated into practice by actionable routines, internal ethics councils. Cultural roots, collective decisions and actions reinforce the conviction for ethical development and deployment. Organizations believe that ethics actually has functional benefits, it is able to fulfil the organizational needs of attracting talent, navigating regulations and also is a big agent to build client trust.

Vakkuri, V., Kemell, K. K., Tolvanen, J., Jantunen, M., Halme, E., & Abrahamsson, P. (2022, June) . In this paper the authors have suggested that there is significant gaps between the EU ethics guidelines for trustworthy AI and their implementation in 39 companies. Further they have tried to understand how is the gap addressed by these organizations through AI ethics governance, regulations

and internal policies, the survey data is collected by the way of structured interview and analysed through thematic analysis to understand the level of compliance. The Findings reveal that most companies do not follow the principles but rather relay on internal policies and general regulations, also broader stakeholder viewpoints are ignored but direct customer expectations are only given importance, with special reference to the principles of transparency and accountability. Many companies think GDPR and the legal frameworks more than sufficient for AI ethics and lastly ethical principles such as fairness, diversity and environmental impact are not given any priority. Hence the authors suggest that integration of Ethical framework into AI development phase is crucial and must be in an actionable format.

Ali, S. J., Christin, A., Smart, A., & Katila, R. (2023, June) . In this paper the authors have explode the experience workers or ethics entrepreneurs and Technology firms. With the help of interviews and observations the practical challenges while implementation of ethical air practices is uncovered. The data is collected from 25 ai ethics workers engineers and managers with the help of observation and interviews thematic coding of interview is done and observational notes have been prepared the finding suggest that ethics is not prioritized when comes to product launchers environment well-being matrix are difficult to quantify and justify also frequent reorganisations at institutional knowledge and relationships level the syrups the ethics implementation within AI systems. However ethics workers are made to act as institutional entrepreneurs and must advocate in ai there is seemingly lack of formal authority at their level and mid management level so they are challenges in persuasion for Ethical implementation there has been instances were individual workers face personal risk when there is ethical concerns or whistle blow protection hence there is a need for early stage review of ethics alignment in the organisation and AI systems ethics workers need protection for whistle blower policies.

Attard-Frost, B., De los Ríos, A., & Walters, D. R. (2023) . The paper is an attempt to evaluate treatment of AI business practices in existing AI ethics guidelines simultaneously propose a framework for integrating business practice into AI ethics. Authors have referred and reviewed 47 guidelines on AI ethics. They feel that mostly guidelines are narrow in nature and does not incorporate business practices. The ethics of business decisions for data sharing, marketing strategies, labour practices are ignored. FAST principles i.e; fairness, accountability, sustainability and transparency are not fully assimilated into the overall business workflow and organizations often getting into 'ethics washing'. Authors recommend that ethical framework should have involvement of political, economic and business organization dynamics into AI guidelines, clear standards for business practices be made with stakeholder engagement. Interdisciplinary collaboration is needed for crafting robust and holistic Ai ethics guidelines. Businesses needs to be sensitized for addressing the gaps in accountability, transparency and sustainability in their AI business practice.

Bevilacqua, M., Berente, N., Domin, H., Goehring, B., & Rossi, F. (2023) . In this paper the authors have come up with a holistic return on ethics (HROE)framework which evaluate the returns on investments made by a form and its AI ethics initiators this framework has three categories of economic impacts which is derived from direct financial benefits in tangible impact which comes from reputation and stakeholder trust and real options analysing the capability is for future adaptability put together this framework helps organisations to justify and evaluate their investment initiatives in ethical AI practices. The HROE framework offers a novel way to evaluate the comprehensive benefits of AI ethics investments. By integrating financial, reputational, and capability-based metrics, it empowers organizations to make informed decisions about ethical practices and long-term strategies. This approach also underscores the societal and organizational significance of responsible AI.

Corvite, S., Roemmich, K., Rosenberg, T. I., & Andalibi, N. (2023) . This qualitative study explores dimensions of emotion AI at a workplace which contributes positively to the overall wellbeing, fairness and performance or do they instigate further bias, discrimination and undue surveillance. 395 US adults were surveyed, those got recruited by Prolific. The employees were made to imagine if EAI was implemented in the workplace and then responses were collected to perceive benefits and risks of the same. The data findings of the paper suggest that only few participants considered the potential benefits of emotion AI in respect of wellbeing, and reduced bias, whereas about 32% categorically found no personal benefits in fact they highlight it several risk with emotion AI like privacy invasion greater labour demands discrimination as well as employer control. Further author suggests that in spite of accuracy and biases been taken care of still there are less there are inherent in EAI this can affect the decision-making process regarding the deployment of EAI in the organisation

Lu, Q., Zhu, L., Xu, X., Whittle, J., Zowghi, D., & Jacquet, A. (2023) . This research paper focus is an identifying governance process and product patterns which can be customised for different stakeholder needs so that responsible AI can be operationalised also this paper provides practical guidance for implementing responsible AI practices at both organisational as well as industry levels. The paper has developed of framework with help of analysis of real-world case studies and best practices implemented in the industry. The findings of the study reviews that lack of connectedness and reusable solution post challenges for implementing responsible. There is interest of stakeholders regarding risk management priorities, communication gap between the AI developers and management teams and expert is shortage and responsible AI principles across organisations is observed. Responsible AI pattern catalogue suggested in the paper provides reusable solutions for governance processes and products and it gives hand on tools for risk assessment question bank and verifyable claims for AI art of facts. The

author suggest there is a multi-level governance framework needed which interlinks industries organisations and internal team.

Morley, J., Kinsey, L., Elhalal, A., Garcia, F., Ziosi, M., & Floridi, L. (2023) . The paper explore the challenge is an opportunities in translating the abstract AI principals to the practical implementation the authors suggest they are needs for tools frameworks and cultural shifts to make aih6 actionable study also explorers barriers such as lack of conceptual clarity lack of resources and inadequate countability mechanism because of which practical implementation of ethical principles is lacking authors suggest that there is need for effective implementation of ethical AI practices as this would lead to wider adoption of ethical practices build system trustworthiness and bring societal impacts that are positive. Authors have conducted mixed method research approach by surveying through semi structure interviews involving 54 practitioners from startup and business corporation's and public sector organisations. The data is analysed two thematic analyses to identify the common barriers unable hours and perceptions towards ai ethics. Authors recommend integrating ethics into their regular existing organisational work flow and embedding prompts for Ethical reflection and development tools at both design and development and deployment phases at organisation level it is needed that the top leadership create support system such as establishment of ethics committees and vessel blower protections to enhance ethical decision making also proper reward system should be implemented within the teams who pursue for Ethical AI design development and employment. Policy makers are suggested to create a collaborative ecosystem with Academy industry and government co developing ethical standards.

Baldassarre, M. T., Gigante, D., Kalinowski, M., & Ragone, A. (2024, April). In this paper the authors have referred to Polaris framework which aims to include 4 key principles of explain ability fairness security and privacy as the guidelines and tools to operationalized trust to the AI. The objectives of the paper are to develop a practical and adaptable framework so that trust in the AI principles can be developed and incorporated at software development life cycle stage also it provides guidelines for various stakeholders and to validate the framework for real world industrial applications. In this research paper a systematic review of 138 frameworks to assess trustworthy AI principals are acknowledged with an interview conducted with 34 professionals who are associated with AI development. The key findings of the paper include a strong demand for actionable tools and guidelines for stakeholder's gap identification at monitoring and development phases and incorporation of practitioner's recommendation at development and development phase for the author suggest that organisation should include and incorporate Polaris into organisational work flows to ensure the alignment and development of Trustworthy AI.

Evers, C. (2024, June). This study examines the discourse surrounding ethical AI by the European Commission (EC) and Big Tech companies (Google and Microsoft), using Critical Discourse Analysis and Gramsci's concept of hegemonic discourse. It finds that while both actors appeared to share common ethical AI principles in non-legally binding settings (2018–2021), their convergence was largely superficial and depoliticized. The introduction of the 2021 EU AI Act (AIA) proposal marks a dislocation, triggering depoliticization of the discourse by revealing deeper conflicts in regulatory preferences, particularly regarding fairness and transparency. Findings of the study suggests that Big Tech companies adopt performance-based, risk-limiting, and self-regulatory strategies motivated by corporate risk and efficiency. Conflicts emerge around concepts of Transparency where EC demands source code and dataset access whereas Big Tech raises concerns over trade secrets and privacy. In case of Fairness EC emphasizes bias and discrimination reduction whereas Big Tech frames fairness through technical efficiency and minimal disruption to business models.

Lindberg, S., Rossitto, C., Knutsson, O., Karlström, P., & Männikkö Barbutiu, S. (2024) . The paper aims to comprehend the perception of the design leaders on what kind of challenges they face while designing and developing AI systems, also what strategies can be adopted by the organization to cultivate and nurture ethically aligned designs. For the same the authors conducted three co-design workshops with ten design leaders from Sweden, who discussed and brainstormed on ethical challenges. The key challenges highlighted in the paper include lack of awareness, knowledge and habits of AI practioners. Authors suggest that ethical awareness, along with organizational culture that advocates and prioritizes ethics and provides sufficient tools and resources for practicing ethics, and stakeholder engagement plays a vital role in infusing adoption practices, diversity as well as user satisfaction, it is essential that the organization makes efforts for raising awareness and brings business opportunity as the narrative for the designers, also creating safe and inclusive space for ethical discussions and collaborative learnings and making sincere efforts for embedding ethics into design critiques, developments and research initiatives.

Laine, J., Minkkinen, M., & Mäntymäki, M. (2024) . Authors in this paper have conducted a systematic review of 93 studies on ethics based AI auditing which have been published up to March 2022 in journals conferences in other academic repositories they have made a use of ethical principles based on you high level expert group on AI and jubin at all empirical analysis. Authors have mamp stakeholder contributions and gaps an auditing literature. The key findings suggest that there are important ethical principles which relates to ai auditing such as fairness transparency non-malefficience responsibility privacy trust beneficence. Stakeholders such a system developers auditors regulators and users contribute in implementation of AI ethics principles however important challenges such as lack of standardised definition and specify tools for Ethical principles and discrepancy between the abstract principles and practical applications is seen as a barrier. Author suggest that its needed to develop a standardize framework for ethics base AI auditing that would address

ambiguity in an implementation of principles also there is a need for stakeholder engagement with AI audit systems a strong collaboration between developers auditors and regulators would ensure alignment of ethical practices with legal requirements within the organization

Nidhi Kshirsagar, Ekta Rokade. (2024) . The purpose of this review paper is to understand the perception and attitude of the AI developers and practitioners towards ethical and governance principles in current scenario, those that are prioritized over others, the challenges and the barriers faced and the ways and means by which practical implementation of these principles can be done as explored in the existing literature and prior studies. Authors have refereed to 20 academic articles that provide various inputs on ai practioners perspective, the major findings include AI professionals have a consensus on the importance relating to AI ethics principles like transparency, accountability, fairness, robustness and justice etc. but putting them for practical application, is a difficult task. The challenges come from guidelines which are ambiguous in nature, limitations of resources and also organisations resistance to change post major barriers. Organisations also think that time and cost implications for incorporating AI ethics make it difficult for them and professionals to incorporate ethics in practical applications. Most of the researchers believe that there is a greater need for interdisciplinary collaborations, ethics trainings and stakeholder engagement as important tools for in depth alignment of ethical AI principles, further these should be intertwined with the organisational goals to foster trust and long-term society benefits. The authors recommend d that the organisations should established dedicated ethics committees and boards that would integrate ethics reviews into their existing work flows also they need to invest in ethics training programs to enhance air professionals awareness and their capabilities in ethical AI.

Vincenzi, B., Stumpf, S., Taylor, A. S., & Nakao, Y. (2024) , this paper suggests that the lay users can be made a part of and be involved in development of human centric and responsible AI systems, authors have conducted a survey of 1121 participants from UK to asses their interest & involvement in the entire AI development lifecycle participation. The Mixed method approach for research is used and the data is then analysed with the help of thematic analysis for open ended responses. According to the responses they prefer involvement of HR teams, Ethics committee, AI technical experts and external regulator for better accountability. Lay users suggest that they be involved during discussion of Business case stage, data collection and database building with customer/lay users data with fairness at both collection and data labelling stage, investigation regarding fairness during model training and evaluation and also vetting AI system performance post deployment. The respondents have raised concerns regarding fairness, transparency, hence the organizations need to build strong internal Ethics committee, include external regulator, developers to create more user friendly data sets that are comprehensible by lay users so that they can identify and flag the ethical concerns while data collection, model creation, model evaluation and deployment stages. They also recommend that the standardization bodies need to add strong points for lay users/customers involvement in AI development and deployment stages.

CONCLUSION

The literature review has significant implications for the development and deployment of AI technology (Bostrom & Yudkowsky, 2014) (Hermann, 2021) (Corrêa et al., 2023). The findings suggest that while there is growing awareness of the need for ethical AI principles, there remains a lack of consistency and comprehensive implementation across the industry (Shams et al., 2023).

The collective body of research on AI ethics highlights a widespread acknowledgment of ethical principles—such as transparency, accountability, fairness, privacy, and inclusivity—yet reveals a persistent gap between these high-level ideals and their practical implementation. Across industry, academia, and policy domains, efforts to operationalize AI ethics often encounter barriers including vague guidelines, limited tools, resource constraints, organizational inertia, and a lack of stakeholder alignment. While many frameworks and tools (e.g., Ethics-as-a-Service, POLARIS, Responsible AI Pattern Catalogues, Holistic Return on Ethics HROE) have been proposed to bridge these gaps, their success depends heavily on contextual adaptation, stakeholder engagement, and continuous governance. Moreover, the ethical discourse is frequently shaped by power dynamics, with large tech firms often promoting self-regulatory models that contrast with the more prescriptive approaches favored by policymakers. Notably, cultural, organizational, and sector-specific factors play a crucial role in how ethics is interpreted and enacted. Overall, the research points toward the urgent need for enforceable regulations, interdisciplinary collaboration, and the integration of ethics into every stage of AI design, development, and deployment. Only through systemic change and inclusive, context-aware practices can responsible, trustworthy, and socially beneficial AI be achieved.

Recommendations for Developers, Business Organizations and Policy makers :

To build responsible AI, it's essential that developers and practioners integrate ethics throughout the development process using practical tools like checklists, bias detection systems, and explainable AI features. Involving a wide range of stakeholders—including everyday users, experts, and impacted communities—at every stage ensures more inclusive and thoughtful outcomes. Promoting ethical awareness through regular training and collaboration across different disciplines helps teams better understand the broader social effects of their work. Additionally, keeping clear records of ethical decisions supports transparency, accountability, and continuous improvement in future projects.

To ensure ethical AI practices, business organizations should establish ethics governance through cross-functional boards or review panels that have real decision-

making authority. It's important to strike a balance between business goals and ethical responsibilities by including long-term social impact measures alongside short-term performance indicators. Companies should also invest in practical tools and frameworks—like Responsible AI Pattern Catalogues or Ethics-as-a-Service models—that help apply ethical principles in everyday work. Lastly, fostering a culture of open discussion, inclusive decision-making, and safe spaces for raising concerns can strengthen ethical awareness and accountability across all levels of the organization.

Policy makers and regulators needs to ensure trustworthy and fair use of AI, there is a need for strong, legally enforceable governance frameworks that move beyond voluntary guidelines. High-risk AI systems should be required to maintain transparency and allow for audits, including access to data sources, explanations of decisions, and fairness checks. Regulations should also be flexible and tailored to fit the unique needs and risks of different sectors like healthcare, finance, and recruitment. Importantly, the voices of marginalized groups and civil society must be included in shaping these policies to make AI systems more inclusive and equitable for all.

REFERENCE:

1. https://eismd.eu/featured/ai4peoples-ethical-framework-for-a-good-ai-society/
2. https://oecd.ai/en/ai-principles
3. https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html
4. The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, Version 2. IEEE, 2017. http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html.
5. https://www.unesco.org/en/artificial-intelligence/recommendation-ethics
6. Vakkuri, V., Kemell, K. K., & Abrahamsson, P. (2019). AI ethics in industry: a research framework. *arXiv preprint arXiv:1910.12695*.
7. Canca, C. (2020). Operationalizing AI ethics principles. *Communications of the ACM*, *63*(12), 18-21.
8. Kim, T. W., Hooker, J., & Donaldson, T. (2021). Taking principles seriously: A hybrid approach to value alignment in artificial intelligence. *Journal of Artificial Intelligence Research*, *70*, 871-890.
9. Mayer, A. S., Haimerl, A., Strich, F., & Fiedler, M. (2021). How corporations encourage the implementation of AI ethics.
10. Miller, G. J. (2021, September). Artificial intelligence project success factors—beyond the ethical principles. In *Special Sessions in the Advances in Information Systems and Technologies Track of the Conference on Computer Science and Intelligence Systems* (pp. 65-96). Cham: Springer International Publishing.
11. Morley, J., Elhalal, A., Garcia, F., Kinsey, L., Mökander, J., & Floridi, L. (2021). Ethics as a service: a pragmatic operationalisation of AI ethics. *Minds and Machines*, *31*(2), 239-256.
12. Bessen, J., Impink, S. M., & Seamans, R. (2022, July). The cost of ethical AI development for AI startups. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 92-106).
13. Deshpande, A., & Sharp, H. (2022, July). Responsible ai systems: who are the stakeholders?. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 227-236).
14. Sloane, M., & Zakrzewski, J. (2022, June). German ai start-ups and "ai ethics": Using a social practice lens for assessing and implementing socio-technical innovation. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 935-947).
15. Vakkuri, V., Kemell, K. K., Tolvanen, J., Jantunen, M., Halme, E., & Abrahamsson, P. (2022, June). How do software companies deal with artificial intelligence ethics? A gap analysis. In *Proceedings of the 26th International Conference on Evaluation and Assessment in Software Engineering* (pp. 100-109).
16. Ali, S. J., Christin, A., Smart, A., & Katila, R. (2023, June). Walking the walk of AI ethics: Organizational challenges and the individualization of risk among ethics entrepreneurs. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (pp. 217-226).
17. Attard-Frost, B., De los Ríos, A., & Walters, D. R. (2023). The ethics of AI business practices: a review of 47 AI ethics guidelines. *AI and Ethics*, *3*(2), 389-406.
18. Bevilacqua, M., Berente, N., Domin, H., Goehring, B., & Rossi, F. (2023). The Return on Investment in AI Ethics: A Holistic Framework. *arXiv preprint arXiv:2309.13057*.
19. Corvite, S., Roemmich, K., Rosenberg, T. I., & Andalibi, N. (2023). *Data Subjects' Perspectives on Emotion Artificial Intelligence Use in the Workplace: A Relational Ethics Lens. Proceed-ings of the ACM on Human-Computer Interac-tion, 7 (CSCW1), 124. Documento en línea*.
20. Lu, Q., Zhu, L., Xu, X., Whittle, J., Zowghi, D., & Jacquet, A. (2023). Operationalizing Responsible AI at Scale: CSIRO Data61's Pattern-Oriented Responsible AI Engineering Approach. *Communications of the ACM*, *66*(7), 64-66.
21. Morley, J., Kinsey, L., Elhalal, A., Garcia, F., Ziosi, M., & Floridi, L. (2023). Operationalising AI ethics: barriers, enablers and next steps. *AI & SOCIETY*, 1-13.
22. Baldassarre, M. T., Gigante, D., Kalinowski, M., & Ragone, A. (2024, April). POLARIS: A framework to guide the development of Trustworthy AI systems. In *Proceedings of the IEEE/ACM 3rd International Conference on AI Engineering-Software Engineering for AI* (pp. 200-210).
23. Evers, C. (2024, June). Talking past each other? Navigating discourse on ethical AI: Comparing the discourse on ethical AI policy by Big Tech companies and the European Commission. In *The 2024 ACM*

Conference on Fairness, Accountability, and Transparency (pp. 1885-1896).

24. Lindberg, S., Rossitto, C., Knutsson, O., Karlström, P., & Männikkö Barbutiu, S. (2024). Doing Good Business? Design Leaders' Perspectives on Ethics in Design. *Proceedings of the ACM on Human-Computer Interaction*, *8*(GROUP), 1-22.
25. Laine, J., Minkkinen, M., & Mäntymäki, M. (2024). Ethics-based AI auditing: A systematic literature review on conceptualizations of ethical principles and knowledge contributions to stakeholders. *Information & Management*, 103969.
26. Nidhi Kshirsagar, Ekta Rokade. (2024). AI Ethics in Practice: A Literature Review on AI Professional's perception and attitude towards Ethical and Governance principles of AI. *European Economic Letters (EEL)*, *14*(4), 1035–1046. https://doi.org/10.52783/eel.v14i4.2232.
27. Vincenzi, B., Stumpf, S., Taylor, A. S., & Nakao, Y. (2024). Lay User Involvement in Developing Human-Centric Responsible AI Systems: When and How?. *ACM Journal on Responsible Computing*.