Vietnam Academy of Social Sciences **Konrad-Adenauer-Stiftung (KAS)**
Institute for European Studies (IES)

****

EU POLICY ON ARTIFICIAL INTELLIGENCE (AI) ETHICS AND POLICY RECOMENDATIONS FOR VIETNAM

Research Team: **Assoc. Prof. Nguyen Chien Thang**
Ph.D. Candidate Ho Thanh Huong
Ph.D. Hoa Huu Cuong
Ph.D Hoang Vu Linh Chi
MA. Tran Thi Thu Huyen
Ms Nguyen Thi Tham

Hanoi 2021

Table of Contents

ACKNOWLEDGMENTS

The authors would like to thank the Konrad-Adenauer-Stiftung (KAS) Foundation for financial support for this research.

We would like to thank the reviewers, Mr. Nguyen Quang Dong - Institute for Policy Studies and Media Development, Ms Tong Khanh Linh - Institute for Policy Studies and Media Development and Mr Pham Ngoc Vinh, Vietnam E-commerce and Digital Economy Agency, Ministry of Industry and Trade with insightful comments.

We would like to sincerely thank the researchers' contribution of the research institutes of the Vietnam Academy of Social Sciences to the seminar. We sincerely thank Assoc. Prof. Nguyen Chien Thang, Director of Institute for European Studies, who coordinated the research activities effectively to complete the research work on time and meet the objectives and expected results of the project. Finally, we would like to thank the Institute for European Studies for hosting and successfully organizing research activities.

List of Table

List of Figures

Abbreviation

1.	AI	Artificial Intelligence
2.	AIA	Artificial Intelligence Act
3.	ADMS	Automated Decision Making Systems
4.	ASEAN	Association of South East Asian Nations
5.	EU	European Union
6.	HLEG	High-Level Expert Group
7.	R&D	Research & Development

INTRODUCTION

Artificial intelligence (AI) plays a central role in the fourth industrial revolution. AI usually refers to a combination of: techniques that allow searching and analyzing large amounts of data so that machines can learn from data (machine learning); robotics that conceptualize, design, manufacture and operate programmable machines; and automated decision-making systems (ADMs) that can predict human and machine behavior and can make decisions on their own. AI technologies can be extremely beneficial from an economic and social point of view and are already used in areas such as healthcare and transportation, or to effectively manage energy and water consumption. AI is increasingly influencing our daily lives, and its range of potential applications is very wide. However, AI also causes impact to society such as data protection and privacy; bias, prejudice; etc. All of those problems need a new mechanism to regulate. In the context, on April 8, 2019, the EU High-Level Expert Group (HLEG) on AI released an Ethics Guideline for Trustworthy Artificial Intelligence. According to this Guide, a trusted AI should be:

(1) legal - respect all applicable laws and regulations

(2) ethics - respect for ethical principles and values

(3) Robustness - from a technical perspective and taking into account the social environment of the product.

Earlier, a number of EU Member States also have specific regulations in some fields that apply AI. For example, in France, since March 2018, one of the core principles set out in AI Strategy requires AI technology to be explainable so as to be socially acceptable. In Germany, the AI ethics – related debate was basically driven by the interests of specific sectors and led to the action of the The *Ethics Committee* of the Federal *Ministry of Transport* and Infrastructure for Automated and Connected Driving in June 2017 has adopted a code of ethics for transport by vehicle is connected and automatically. In November 2018, a national strategy on AI was born, proposed a series of measures on ethics.

The Vietnamese government has identified artificial intelligence as one of the breakthroughs and spearhead technologies of the industrial revolution 4.0. Since 2014, AI has been listed a high technology prioritized for development investment. For many years, the Government has approved the list of high-tech products prioritized for investment and the list of high-tech products encouraged for development. The Government's will to develop AI is also reflected in a series of documents such as: (i) Directive No. 16/CT-TTg dated May 4, 2017 by the Prime Minister on enhancing accessibility to the 4.0 industry identifies AI technology as one of the breakthroughs and key technologies of Industry 4.0; (ii) Resolution of the Government meeting in March 2018 setting out the task of research, impact assessment and development of the National Strategy on Industry 4.0; In the process of formulating the Strategy, AI technology will be thoroughly evaluated and studied to form a strategy, roadmap and solutions for AI development that best suit the reality and potentiality of Vietnam’s entrepreneurs; and (iii) especially the National Strategy on Research, Development and Application of Artificial Intelligence to 2030, issued on January 26, 2021 with the goal of becoming a center of innovation and development and a leading country in artificial intelligence solutions and applications in the ASEAN and the world. The strategy clearly defines the view that AI is the foundation technology of the fourth industrial revolution, making an important contribution to creating breakthroughs in production capacity and improving the competitiveness of Vietnam,

promoting economic development and sustainable growth. In the National Strategy on Research, Development and Application of AI to 2030, the Prime Minister has outlined 5 groups of strategic orientations, including: Building a system of legal documents and legal corridors; management related to AI; building data and computing infrastructure for AI research, development and application; developing AI ecosystem; promoting AI application; accelerating international cooperation in the field of artificial intelligence.

It can be seen that AI and AI-related policy issues are very new not only to Vietnam - a new comer in this field, but also to the world. Vietnam has just begun to prepare a policy system to regulate this area. In that context, it is very necessary to understand the views, situations and policies on artificial intelligence ethics in the EU, with the hope that the experiences will help Vietnam draw lessons from experience and gain a better perspective in AI policy making process.

Artificial intelligence is at the heart of many technologies, being developed and deployed in our lives. It offers so many benefits in healthcare, improving the efficiency of production systems, increasing security and in many other ways that we can only begin to imagine. However, artificial intelligence also creates some potential risks such as unclear decision making, bias and discrimination, influence on people's personal lives, manipulation of human behavior or used for criminal purposes. The study aimed at understanding the ethics of artificial intelligence and the EU's point of view when designing, developing, implementing or using AI products and services; identifying some challenges in the EU; analyzing cases of Germany and France; and giving out some suggestions for Vietnam's policy makers on this issue.

1. Definitions and ethical issues of artificial intelligence

1.1. Definitions of artificial intelligence, ethics, ethics of artificial intelligence

Definition of AI: a collection of different technologies that can be brought together to allow machines to operate with what seems like a human intelligence. This includes learning the regulations needed to make certain decisions and reasons for coming to certain conclusions, learning from past experiences, and self-correction.

The ultimate goal: It is possible to make these systems work as powerful as the human brain.

Weak AI: Simple, weak or narrow AI designed and trained for specific tasks.

Strong AI: Strong systems should be able to find a solution without human intervention when given an unfamiliar task.

Ethics: According to the Oxford dictionary, ethics includes a set of moral values that guide or control human activities. People acquire moral values from different sources such as religion, culture, environment and family. Some of these moral values are reflected in the rules, regulations, and laws that have been formulated by certain people such as religious leaders and community legislators throughout history. People in communities use these moral values in contexts to interact with each other, make decisions, and take action. These processes include reason, thinking, understanding, logic, intentions, preferences, beliefs, and desires.

Artificial intelligence ethics encompasses a set of values, principles, and techniques that use standards which widely accepted on the right and the wrong to guide ethical behavior in the development and implementation of AI technologies.

AI is the use of machines to do things that would normally require human intelligence. In many areas of human life, AI has rapidly and significantly influenced human society and the way we interact with each other. During the development, AI has presented huge ethical and sociopolitical challenges that require a thorough ethical and philosophical analysis. Its social impact needs to be studied to avoid any negative impact on people and society. AI systems are becoming increasingly autonomous and intelligent. This comprehensive development raises many problems. In addition to the harms and potential impact of AI technology on our privacy, other concerns include ethical and legal status (including moral and legal rights), ethical agents as well as issues related to human dignity. However, AI and its relationship to human society are divided into three different phases: (1) short-term phase (early 21st century): autonomous systems (transportation, weapons), machine bias in the law, privacy and surveillance, black box problems, and AI decision-making; (2) mid-term period (from the 2040s to the end of this century): Governance using AI, ethical and legal status validation of intelligent machines (artificial ethical agents), human interaction and machine, batch automation; (3) long-term period (starting 2100s): technological singularity, mass unemployment, space colonization (Coeckelbergh, 2020).

1.2. Ethical issues of artificial intelligence

1.2.1 Ethical issues of artificial intelligence

Privacy

Privacy allows people to make their own decisions, without coercion, to better account for their behavior, and to be strategic in their social interactions. Privacy has long had a controversial relationship with government viewpoint. Individuals often fight for the right not to be observed, while governments often fight for the right to observe citizens (Tene, O., Polonetsky, 2013).

AI, and specifically machine learning applications that work with big data, often involves the collection and use of personal information. AI can also be used for surveillance, on the street but also in the workplace and via smartphones and social media everywhere. Often people are not even aware that the data is being collected or that the data they provide in one context is then used by third parties in another context. Big data also often means data collected by different organizations and being combined.

Ethical AI requires that data be collected, processed, and shared in a manner that respects the privacy of individuals and their right to know what happens to their data, access to their data, object to collecting or processing their data, and to know that their data is being collected and processed. Many of these problems also arise with other information and communication technologies, and as we shall see, transparency is also an important requirement in such cases. And data protection issues also arise in research ethics, such as the ethics of data collection in social science research.

However, considering the contexts in which AI is used today, the issues of privacy and data protection become increasingly daunting. It is relatively easy to respect these values and rights when conducting a survey as a social scientist: a person can notify respondents and request their consent in an informal manner. It is clear and relatively clear what will happen to the data. But the environments in which AI and data science are used today are often very different. Let’s see social media, although privacy information and apps ask for user consent, it's not clear to users

what happens to their data or even what data is being collected. If they want to use the app and enjoy its benefits, they must agree. Often, users don't even know that AI is powering the apps they use. And often the data provided in one context is then transferred to another domain and used for another purpose (data repositioning), for example when companies sell their data to third parties or move data between different parts of the same company without the users' agreement and knowing about this.

Where will the data end up when we die? Will our data outlive us? What may even last beyond our own expiration date is our data itself, so whereby the accumulated data goes. Even if a person who provides the data dies, will the data about them die with him/her? These are very interesting questions when one considers that our data has a very high viability. People have no control over their personal data if a company refuses to delete our data despite a request. Even our most private data can be in the hands of organizations, companies and other 3rd parties. Without any legal battles, the data is unlikely to be deleted. Furthermore, the data already has the ability to assist in influencing the model, meaning that its deletion is only a formality (Schiff, 2021).

So, in the privacy paradox, even our most private data is not private. Due to the interconnectedness of the Internet, the ability to infer certain qualities from human actions on different websites is increased through the sharing of data between different platforms (such as WhatsApp and Facebook).

Manipulation, Exploitation and Vulnerable Users

The fact that AI is changing every aspect of the way we live and work. Different AI types are active in diverse areas such as vaccine development, environmental management and office administration. And although AI does not have human intelligence and emotions, its capabilities are very strong and develop rapidly. The risks that users are being manipulated and abused are obvious. AI is used to manipulate what we buy, what news we track and reviews of whom we trust, etc. The reality has shown that users' risk being taken advantage of through the use of social networks. As social network users, we may become unpaid labor force, are exploited to produce data for AI then analyze the data we have - and finally for companies using the data, which often includes third party. Companies can use the powerful method to exploit the tendency of human beings for the purpose of improving advertising, marketing and sales. The result is that people's privacy is being eroded.

Several studies have shown how AI can learn to identify vulnerabilities in human habits and behaviors and use them to influence human decision-making (Dezfouli et al., 2020). A team of researchers at CSIRO's Data 61 performed three experiments, which are three games, for humans to play games against a computer. In each experiment, the computer learned from the participants' responses and identified and targeted vulnerable points in the participants' decision-making. The end result is that the computer has learned to direct participants to specific actions.

The danger here is that even in today's democracies, AI could lead to new forms of manipulation, surveillance, and totalitarianism, not necessarily in the form of authoritarian politics but in a hidden and more efficient way: by changing the economy in a way that turns us all into "cattle" that use our smartphones to get our data.

But AI can also be used to manipulate politics more directly, such as by analyzing social media data to help political campaigns (like the famous case of Cambridge Analytica, a used data from

Facebook users without their consent for political purposes in 2016 US presidential election, or by letting bots post political messages on the social network based on analyzing people's data about their political preferences to influence voting. Some also worry that AI, by taking over cognitive tasks from humans, will disable its users by "making them less able to think for themselves or decide what to do."

Box 1: The case of the company Cambridge Analytica

Cambridge Analytica collected information from more than 87 million Facebook users through an external application in 2015. The data came from a personality test; about 270,000 people were paid to take the test. It is called "thisisyourdigitallife", but it not only collected information of participants but also collecting data from the profiles of the participants' friends, and as a result, they had a huge pool of data.

Cambridge Analytica collected personal information about where users lived and what pages they liked, helping to build psychographic profiles that analyzed characteristics and personalities. This type of information was later deployed in political campaigns.

Some AI users are also more vulnerable than others. Privacy and exploitation theories generally assume that the user is a self-sufficient and relatively young and healthy adult with full intellectual capacity. However, the real world is a world filled with children, the elderly, people without "normal" or "adequate" intellectual capacity, etc. These vulnerable users are more at risk. Their privacy can often be easily compromised or they can be easily manipulated; and AI provide new opportunities for such violations and manipulations. Consider a young child conversing with a doll connected to a technological system that includes AI. Possibly, the child does not know that AI is being used or that data is being collected, let alone what is being done with their personal information. Not only can an AI-powered doll or chatbot collect a lot of personal information about the child and its parents in this way; It can also manipulate the child using language and voice interfaces. When AI becomes part of the "internet of toys" (Internet of Toys) and the internet of things, this is an ethical and political issue (Druga et al., 2018)

Equality

The principles state that AI will act in a fair and unbiased manner. Justice is often illustrated by a statue of the Roman Goddess Justitia with a sword, a scale, and a blindfold. The blindfold represents innocence, the scale represents justice, and the sword represents punishment.

Defining fairness at the human level is a major challenge for artificial intelligence. To this day, moral theory is still the subject of controversy, with many schools participating in this never-ending debate. A poll conducted by two scientists to find out which theory is the most supported theory shows that there is not one theory that has the support of the majority. About a quarter of philosophers accept or favor the theory of consequentialism and dentology and about a third lean towards virtue ethics. Therefore, defining "fair" and "ethical" for what machines is a very difficult problem, because there is no consensus on ethical theory (European Parliament, 2020).

In some countries, banks and other credit organization are already using AI systems to pre-sort credit applications on an existing database of applicants. This certainly has a number of

advantages, one of which is being able to arrive at decisions faster and on a more informed basis, making it theoretically more plausible. However, this can also entail disadvantages, namely leading to certain biases. For example, a credit applicant's personal information will in most cases contain information about their area of residence. On that basis, and with the further use of publicly available (or privately collected) data, systematic bias may occur against people from certain residential neighborhood.

There are many examples of bias in AI applications. Does AI have the right to decide who can be released on bail and who can not be? Or it may approve or reject the loan application or your resume? Should we trust the AI making decision rather than human beings or not? That's just a few examples of ethical questions arising from the widespread application of algorithms in multiple processes and operating decisions. This development of AI in replacing humans have triggered an arms race in order to make an assessment of AI capable and efficient.

A feasible way to provide guidance and assessment in these settings is to use control independent audit of third parties. Because independent auditing is widely popular in the evaluation processes and procedures for decision-making that made by human (e.g financial audit), so independent auditing process is taken for granted when we seek to assess the ethics of AI algorithms. The management bodies, the main concern is their assessment on negative impact of the algorithm for the rights and interests of the related parties, with the determination of the respective situations and/or features of algorithm generates this negative impact. Scholars have argued that the approach to classical analysis which was too focused on the technical aspects have ignored the social aspects, such as the algorithms have missed or eliminated some of the hazards relating to the decision or classification, which affect the minorities. Take the example of reviewing loan documents. The negative impact of a tool for loan losses not only depend on the algorithm that is skewed statistically or not. For example, some minority groups with training data bias, however the dangers only appear from the loan officer who has decided to use the lending risks tool in deciding whether to lend or not. Hence, equality needs to be understood that the decision must be made by human, not by machines.

1.2.2 Ethical studies of artificial intelligence

It can be said that the research on AI ethics in the world is still very limited. Although according to Stanford University's AI Index 2021 report, there has been a significant increase in the number of articles with ethics related keywords in headlines submitted to AI conferences since 2015, the average is still low (Human-Centered Artificial Intelligence, 2021). Research institutions and AI development organizations are at the forefront of publishing papers on AI principles. In addition, compared with other institutions, private companies are also at the forefront of publishing publications on AI ethics. Europe and Central Asia have the highest number of publications as of 2020 (44), followed by North America (30), and East Asia and the Pacific (14). Since 2018 technology companies including IBM, Google and Facebook as well as UK, European Union and Australian government agencies have also implemented ethical principles.

Artificial intelligence is already present in all aspects of social, economic, ethical and administrative life, which has created debates about the role of AI in labor mobility, autonomous vehicles, military, disinformation, health care, education. Governments all face these challenges, but AI also creates new opportunities to contribute to social justice and benefits. AI's potential to help solve global challenges and help achieve goals like the United Nations' Sustainable

Development Goals. In fact, whether AI is geared towards social benefits and justice, there are indeed gaps between the expectations and the actual realization of these aspirations.

Governments and technology corporations soon realized the gaps need to be filled in regards to ethical issues when applying technology in their activities, that's why there are many ethical principles. virtue is set forth. There are too many sets of guidelines on the principles of using, operating, and designing AI, and they have certain conflicts with each other. So what are those gaps?

Schiff et al (2021) begin by highlighting several initiatives from corporations to outline their commitments to AI ethics. What they find is that ethical issues are often ambiguous, in particular, without practical guidelines for their implementation and empirical evidence of their effectiveness, ethical claims nothing more than promises without action.

Each document typically addresses a range of social and ethical concerns, proposing principles for ethical AI (responsible or trustworthy) to respond to and in some cases to address specific reforms or internal governance strategies. More importantly, the overall focus of the documents is to present a set of high-level ethical principles that guide an organization's approach to responsible AI. For example, Google's AI guidelines include "Good for society", "Avoid creating or reinforcing AI biases", "Built and tested for safety" and "Responsible". responsibility to everyone" and several other principles (Google, 2018). However, these high-level responsible AI principles are often vague and can contain a multitude of possible interpretations.

They also warn people that companies can enter the ethical arena of AI to enhance their standing with customers and build trust, a tactic known as "shopping ethically, washing moral or moral evasion". Such an approach minimizes accountability on their part while maximizing ethical signaling. Thus, aligning an organization's purpose, mission, and vision with responsible use of AI can help alleviate this challenge, using them as "value leverage".

The effects of AI are really difficult to identify and assess, especially when they have quadratic or third-order effects. We need to approach this problem from an interference point of view to better understand the interdependence of these systems on their surroundings. This is important because harm from an AI system does not arise simply from a product.

Thinking about these intersecting concerns requires working with stakeholders across fields, but they come from different technical and ethical backgrounds. Therefore, it is difficult to make it possible for convergence and shared understanding. Finally, there is now a proliferation of technical tools to address bias, privacy, and other ethical issues. However, a lot of them don't have specific and helpful instructions on how to put them into practice. There is also a lack of guidance on how to customize and troubleshoot for different situations at times, further limiting their applicability.

The ethical issues that arise with AI systems fall into two categories. The first type is these AI systems as objects, i.e. tools created and used by humans. Ethical issues surrounding the first category include: privacy, manipulation, ambiguity, bias, human-robot interaction, impact on autonomy and employment. The second type of ethical problem is AI systems as a subject, ethics for AI systems themselves (Müller, 2020), focusing on the design of machines and the ways in which ethical principles can be incorporated into the design of automated machines/robots.

Research in AI ethics focuses on several key issues including: ethical principles that can be implemented in automated machine/robot decision-making processes; Synthetic study of AI ethics and comprehensive guidance on AI. Regarding the ethical principles that can be

implemented in AI decision-making processes, the studies of Anderson, M. ((Anderson, M., Anderson, 2007; Etzioni, 2016) emphasize the importance of machine ethics , applicable ethical principles and ethics in human-AI interactions. However, Yu et al (2018) propose to divide the problem into four areas: 1) explore ethical dilemmas; 2) individual ethical decision frameworks; 3) collective ethical decision frameworks; and 4) ethics in human-AI interactions (Yu et al., 2018).

There are also a number of synthetic studies on AI ethics. Vakkuri, Ville and Abrahamsson (2018) compiled a lot of articles and keywords from different fields related to AI ethics. The 37 keywords with the highest frequency are divided into 9 groups to find the most popular topics related to AI ethics. This study shows that defining the area of AI ethics remains a challenging task. Prates and et al. provide empirical evidence of the development of AI ethics research. The results of the study show that there has been relatively low attention by the AI community to the ethical consequences of AI over the past several decades. Although awareness among researchers and experts has increased, the authors suggest that research related to AI ethics needs to take into account the technical aspects of today's leading AIs (Prates, M., Avelar, P., Lamb, 2018).

Finally, the study of comprehensive AI guidelines is a very interesting topic for understanding how AI ethical guidelines are implemented. Hagendorff evaluated several recently issued ethics guidelines to address issues that overlap with the shortcomings of 22 ethics guidelines related to AI technology. This study provides a detailed overview of the field of AI ethics and examines to what extent ethical principles and values have been implemented in the research, development, and application of AI systems. and how to improve AI ethical requirements. The author concludes that researches in general has addressed many aspects of AI ethics, but almost no research has been written on the implementation of ethical goals and ethical values (Hagerdorff, 2020).

Ethics is not a technological matter; ethics is a human matter. So that's something we all need to take care of. There are certainly many possibilities in automation and in AI that are emerging that will lead to all the possibilities of relentless human tracking to identify actions and behaviors far beyond what was possible before, and because So we need to be concerned with what's allowed and what's not and how these technologies can be used in an impactful society. According to the human-machine interaction, the system can have different emotional and psychological effects on the human. These include things like uncertainty, anxiety, and damage to one's self-esteem or positive self-image, as well as more obvious harms like hijacking, misleading or damaging one's reputation language.

Trust, accessibility, and accountability are keys to AI ethics. AI provides insight into its own system's reasoning process. Solutions that focus solely on machine learning and deep learning often do not provide insights into their recommendations, thus making them black boxes. That's because it focuses on something that can be easily measured rather than the feel of the technology. It gives fake comfort. Technology must respect ethics and human rights and provide human-centered AI. So, it should have some ethics or regulation.

2. EU policies on AI ethics

2.1. EU Strategy on AI ethics

2.1.1. Timeline of EU policies on AI

Artificial intelligence (AI) is expected to transform the economy and impact almost every aspect of human life over the next few decades. This provides a boost to the growing investments in AI research and development (R&D), as well as the rapid adoption of AI among the public, enterprises, organizations and governments worldwide. By 2030, AI could contribute up to $13 trillion to the global economy, a figure roughly equal to the current annual economic output of China, the world's second-largest economy. Moreover, as AI applications are expanding into many fields, early adopters will benefit in reaping economic benefits and have appropriate strategies in developing this field. The combination of large economic dividends, and social and military benefits, has spurred states to join in a race in this area to rapidly and effectively adapt AI in as many areas as possible.

When mentioned to the global AI startup ecosystem, a study from 2018 notes that the top three players (measured in terms of number of AI startups) is the US with 1,393 startups (40 %), China with 383 startups (11%) and Israel with 362 startups (11%) (Axelle Lemaire et.al., 2018). Four European countries are among the top 10 (UK in fourth, France in seventh, Germany in eighth and Sweden in tenth). However, in total, Europe is second only to the United States, with 769 AI startups (22% of the global total). It can be seen that single European countries will find it difficult to compete globally, but if Europe strengthens its single digital market, the region has the potential to become a major player in AI field, despite Brexit would bring about lasting consequences for the efforts. (Brattberg et. al., 2020)

It can be seen that AI is the inevitable trend, having been and will continue to thrive, dominate in many areas and affect every aspect of human life. In the process of developing that besides the benefits it gives people also raises many ethical issues related to development, applications, using AI as mentioned in Part 1. Being a main player in the field of AI, EU also face problems arising from it and take preparation steps to be ready in AI times.

The following table summarise the EU's steps toward AI.

Table 1: EU timelines on AI

10/04/2018	Member States signed the "Declaration of Cooperation on Artificial Intelligence", agreeing to come together to discuss the most important issues raised by AI, from ensuring competitiveness in R&D to addressing social, economic, ethical and legal questions.
25/04/2018	The European Commission adopts "Communication on Artificial Intelligence", which outlines the EU's approach to AI. This document, with an emphasis on ethical AI, aims to strengthen the EU's technological and industrial capacity, increase public and private sector adoption of AI, and prepare Europeans for the socioeconomic changes brought about by AI and ensure that an ethical and legal framework is in place.
14/06/2018	Set up of the High Level Expert Group (HLEG) on AI

	The European Commission designates AI HLEG. This team includes 52 AI experts from academia, civil society and business, advising the European Commission on the implementation of its AI strategy.
07/12/2018	The European Commission (in collaboration with Member States) presents a "Coordinated Plan on AI" to promote the development and use of AI in Europe, in which noting that the EU is lagging in private investments and may be at risk of losing the opportunities offered by AI if the EU does not have significant efforts.
09/01/2019	Launch of AI4EU Project brings together seventy-nine leading research institutes, SMEs and large enterprises in 21 countries to build a focal point for AI resources, including data warehouses, power calculations, tools and algorithms. The project aims to provide services and provide support to potential users of the technology and help them test and integrate AI solutions in their processes, products and services.
08/04/2019	- HLEG publishes the "Ethics Guidelines for Trustworthy AI", which outlines a human-centered approach to AI and lists seven key requirements that AI systems need to be met in order to be reliable.
	- The European Commission issues "Communication on building trust in human-centered AI", which sets out seven requirements that all AI applications need to comply with in order to be considered trustworthy: human intercessors and supervisors; Strong and safe engineering; privacy and data governance; transparent; diversity, non-discrimination and equity; social welfare and environment; and accountability. The principles identified in the Communications Program are primarily intended to draft AI ethical principles based on an existing regulatory framework, which all AI developers, vendors, and users must apply.
26/06/2019	HLEG issues "Policy and Investment Recommendations for Trusted Artificial Intelligence." makes thirty-three recommendations that can reliably guide AI towards sustainability, growth, and competitiveness, and inclusion — while empowering, benefiting, and protecting people. The recommendations are intended to help the European Commission and Member States update the coordination plan by the end of 2019.
26/06/2019	HLEG kicks off the "Ethical Guidelines Review List for Trusted AI" pilot phase, which will run until December 1, 2019.
19/02/2020	The European Commission issued the "White Paper on Artificial Intelligence - European Approach to Excellence and Trust", "European Strategy on Data" and "Strategy for the Right Europe" with the digital age". The European Commission has been conducting public consultations on the white paper until May 31, 2020, and it plans to present proposals for a regulatory framework by December 2020.

04/2021	The European Commission officially published the draft legal regulation regulating the operation of AI for comments

On February 19, 2020, the Commission published its White Paper on Artificial Intelligence, "AI – towards an ecosystem for excellence and trust". Legal policy requirements for the regulatory framework were developed there. Its goal is not to stifle innovation while adequately addressing risks. The current proposal aims to fulfill the second goal of developing a trusted ecosystem by proposing a regulatory framework for worthtrusty AI. The proposal is based on the EU's fundamental rights and values, and aims to give people and other users the confidence to adopt AI-based solutions, and to encourage companies to develop them.

The draft claims to meet the Council of Europe's requirements to promote funding for AI provided it ensures a high level of data protection, digital rights and ethical standards.

In this political context, the Commission presents a proposed regulatory framework for Artificial Intelligence with the following specific objectives:

• ensure that AI systems placed on the Union market and used are secure and respectful of applicable law regarding the Union fundamental rights and values;

• ensure regulatory certainty to facilitate investment and innovation in field of AI;

• strengthen governance and effective enforcement of existing laws on fundamental rights and security requirements applicable to AI systems;

• facilitate the development of a single market for legitimate, secure and trusted AI applications and prevent market fragmentation.

On 19 February 2020, the European Commission issued the White Paper on AI - European Approach to Achieving Excellent and Trusted AI (European Commission, 2020). The document outlines a risk-based approach to AI and policies to promote adoption of that technology. The European Parliament and Council regulation proposal sets out harmonized rules for Artificial Intelligence and amends a number of the Union legislative acts (Artificial Intelligence Act 21 2021).

To help better define its vision for AI, the European Commission has developed an AI strategy to parallel the European approach to AI. The AI Strategy proposed measures to streamline research, as well as policy options for AI regulation, to be put to work on the AI package.

The Commission announced its AI package in April 2021, proposing new rules and actions to make Europe a global hub for trusted AI. This package includes:

• Communication on Promoting the European Approach to Artificial Intelligence;

• Plan for Collaboration with Member States: Updated for 2021;

• The AI Regulation Proposal sets harmonized rules for the EU (Artificial Intelligence Act).

On April 21, 2021, the European Commission issued the Artificial Intelligence Act (AIA) Draft to codify the standards of its trusted AI system. It sets out core horizontal rules for the development, trade and use of AI-driven products, services and systems within the EU, that apply to all industries. The draft legislation is far more comprehensive in nature than anything

being considered by China or the United States - currently the countries with the most AI research and development globally. This can be seen as an effort by the European Union to influence the development of AI technology to help build a resilient Europe in the digital age where people and businesses get the benefits of AI.

After being passed, the AIA together with GDPR, the Digital Services Act (December 15, 2020) and the Digital Markets Act (December 15, 2020) will tune online platform and service. When completing this "legislative square" along with the Data Governance Act (Data Governance Act, November 25, 2020) will transform into a legislative pentagon, together with European health data Space proposal published — EU will develop a "digital constitutionalism" (Celeste, 2019; De Gregorio, 2021) creating a playing field where its citizens can live and work better and more sustainable.

2.2.2. EU guidelines on AI ethics

The EU's AI Ethical Guidelines were first published on 18 December 2018, the revision after public consultation was published in April 2019. In the view of the Principles, AI is not an end in itself, but a promising means of increasing human development, thereby enhancing the well-being of individuals and society as well as common good, as well as bringing progress and innovation. In particular, AI systems can help facilitate the achievement of the United Nations Sustainable Development Goals, such as promoting gender balance and addressing climate change, rationally using our natural resources, promoting our health, mobility and manufacturing processes, etc.

The revised AI Ethical Guidelines following feedback received from the public consultation on the draft published on December 18, 2018 not only concerns with the reliability of AI systems themselves, but also requires a holistic and systematic approach that includes the reliability of all actors and processes that are part of the system's socio-engineering context in its entire life cycle.

These guidelines address to all AI stakeholders who design, develop, implement, use, or be affected by AI, including but not limited to companies and organizations, researchers, public service, government agencies, civil society organizations, individuals, workers and consumers. Stakeholders committed to achieving reliable AI may voluntarily choose to use the Principles as a method of delivering on their commitments, particularly by using Chapter III's list of factual assessments when develop, implement, or use AI systems. This list of reviews can also supplement - and therefore be included - existing audit processes.

Trustworthy AI has three components, which should be met throughout the entire system lifecycle:

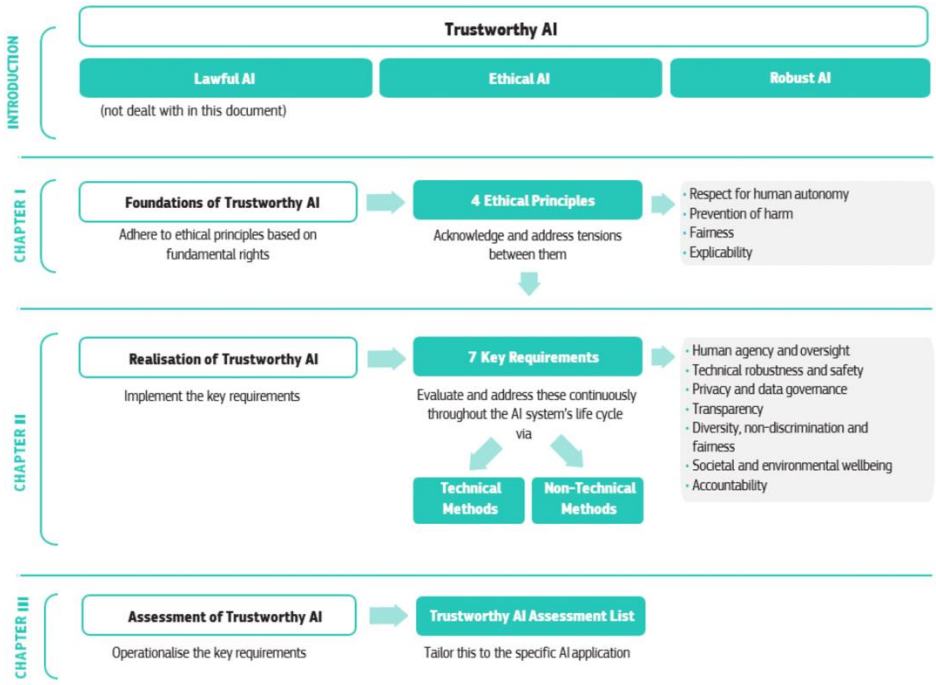
(i) be lawful, comply with all applicable laws and regulations;

(ii) be ethical, ensuring adherence to ethical principles and values; and

(iii) must be robust, both from a technical and social perspective, since even with good intentions, AI systems can cause unintentional harm.

Each of these three components is necessary but not sufficient to achieve reliable AI. Ideally, all three components work in harmony and overlap in their activities.

Figure 1: EU guideline framework for Trustworthy AI



Source: HLEGAI (2019)

The guide outlines a set of seven key requirements that AI systems must meet in order to be considered trustworthy. A list of specific assessments to help verify the application of each key requirement: (HLEGAI, 2019)

(i) *Human agency and oversight:* AI should not trample on human autonomy. Human will not be manipulated or coerced by AI systems and humans will be able to interfere or monitor every decision the software makes.

Fundamental rights: given the reach and capacity of AI systems, they can negatively affect fundamental rights.

Human agency: Users should be provided with tools and knowledge to interact with the AI system, ensuring their autonomy. AI systems can sometimes be deployed to shape and influence human behavior through mechanisms that can be difficult to detect, as they can exploit unconscious processes, including various forms of unfair manipulation, deception, herding, and conditioning, all of which may threaten an individual's autonomy. The overall principle of user autonomy should be central to the system's functionality. The key to this is the right not to be subject to a decision based solely on automated processing where it creates legal effects on users or similarly significantly affects them.

Human oversight: Human oversight helps ensure that an AI system does not undermine human autonomy or cause other adverse effects. Monitoring mechanisms may be required to varying degrees to support other safety and control measures, depending on the area of application of the

AI system and potential risks. "All other things being equal, the less oversight a human can exercise over an AI system, the more extensive testing and stricter governance is required."

(ii) *Technically robustness and safety*: AI must be secure, accurate, not easily compromised by external attacks, and reliable.

Resilience to attack and security: AI systems, like all software systems, should be protected from vulnerabilities that could allow them to be exploited by adversaries.

Fallback plan and general Safety: AI systems should have safeguards that enable fallback plan in cases of problems.

Accuracy: Accuracy refers to an AI system's ability to make accurate judgements, such as correctly classifying information into appropriate categories, or its ability to make accurate predictions, recommendations, or decisions based on data or models.

Reliability and reproducibility: A reliable AI system is one that works properly for a wide range of inputs and in a wide range of situations. Reproducibility describes whether an AI experiment exhibits the same behavior when repeated under the same conditions. This allows scientists and policymakers to describe exactly what AI systems do.

(iii) *Privacy and data governance:* Personal data collected by AI systems must be secured and privacy, should not be accessible to anyone, and not easily stolen.

Privacy and data protection: AI systems must ensure privacy and protect data throughout the entire lifecycle of the system. This includes information originally provided by the user, as well as information generated about the user during his or her interaction with the system (e.g. the output the AI system is generated for the user or how users respond to specific suggestions). Digital records of human behavior could allow an AI system to infer not only an individual's preferences, but also an individual's sexual orientation, age, gender, religious or political views surname. To enable individuals to have confidence in the data collection process, it must ensure that the data collected about them will not be used to unlawfully or unfairly discriminate against them.

Quality and integrity of data: The quality of the data sets used is paramount to the performance of AI systems. When data is collected, it may contain social biases, inaccuracies, errors and confusion. This needs to be resolved before training with any given dataset. In addition, the integrity of the data must be guaranteed. Feeding malicious data into an AI system can change its behavior, especially with self-learning systems. The processes and data sets used should be tested and documented at each step of planning, training, testing, and deployment. This would also apply to AI systems that were not developed in-house but acquired elsewhere.

Access to data: In any particular organization that processes an individual's data (whether someone is a system user or not), data protocols that manage data access rights must be applied. These protocols should outline who can access the data and under what circumstances. Only qualified employees with the competence and need to access an individual's data should be allowed to do so.

(iv) *Transparency:* The data and algorithms used to create an AI system should be accessible, and the decisions made by the software should be understood and traceable by humans. In other words, operators can interpret the decisions their AI systems make.

Traceability: The data sets and decision-making processes of the AI system, including the data collection and data labeling procedures and the algorithms used, must be documented following the best possible standard to enable traceability and increase transparency.

Explainability: involves the ability to explain both the technical processes of an AI system and the related human decisions (e.g., areas of application of the system). Technical explainability requires humans to understand and trace the decisions made by AI systems. Furthermore, there may be a trade-off between enhancing the explainability of the system (which may decrease its accuracy) and increasing its accuracy (at the expense of explainability).

Communication: AI systems should not present themselves as human to users; Humans have the right to be informed that they are interacting with an AI system. This requires AI systems to be recognized as such. In addition, the option to decide against this interaction in favor of human interaction should be provided where necessary to ensure compliance with fundamental rights.

(v) *Diversity, non-discrimination, and fairness:* AI-powered services should be equally available to all, regardless of age, gender, race, or other characteristics.

Avoidance of unfair bias: Data sets used by AI systems (both for training and operations) may be subject to accidental historical bias, incompleteness, and bad governance patterns. Continuing to hold such biases may lead to direct (unintentional) prejudice and discrimination against certain groups or people, potentially exacerbating prejudices and prejudices. Harm can also result from knowingly taking advantage of (consumer) bias or engaging in unfair competition, such as price homogenization by collusion or non-transparent markets.

Accessibility and Universal design: Especially in the business-to-consumer sectors, systems must be user-centric and designed in a way that allows everyone to use AI products or services, regardless of their age, gender, abilities, or characteristics. The accessibility of this technology for persons with disabilities, present in all societal groups, is of particular importance. AI systems should not take a one-size-fits-all approach and should consider Universal Design principles that address the widest possible range of users, subject to relevant accessibility standards. This will allow fair access and active participation by all in existing and emerging computing-human-mediated activities and related assistive technologies.

Stakeholder Participation: To develop reliable AI systems, it is advisable to consult with stakeholders who may be directly or indirectly affected by the system during its life cycle. It is beneficial to solicit regular feedback even after implementation and to establish long-term mechanisms for stakeholder engagement, for example by ensuring information, consultation and participation of employees in the entire process of implementing AI systems in organizations.

(vi) *Societal and environmental well-being*: AI systems must be sustainable (that is, they must be ecologically) and promote positive social change.

Sustainable and eco-friendly AI: Measures to ensure the eco-friendliness of the entire supply chain of the AI system should be encouraged.

Social impact: Regular exposure to social AI systems in all areas of our lives (may be education, work, caregiving, or leisure) can change our perceptions on social authority or influence our social relationships and attachments.

Society and Democracy: In addition to assessing the impact of the development, implementation, and use of AI systems on individuals, this impact should also be assessed from a

societal perspective, taking into account its impact on institutions, democracy, and society at large. The use of AI systems needs careful consideration, especially in situations involving democratic processes, including not only political decision-making but also electoral contexts.

(vii) *Accountability:* AI systems must be audited and protected. Negative effects of the system should be acknowledged and reported in advance.

Auditability: Auditability requires enabling the evaluation of algorithms, data, and design processes, whereby information about business models and intellectual property related to AI systems must be always public. Evaluations by internal and external auditors, and the availability of such audit reports, can contribute to the reliability of the technology. In applications that affect fundamental rights, including safety-critical applications, the AI system must be able to be independently audited.

Mitigation and reporting of negative impacts: Both the ability of a system to report on actions or decisions that contribute to a given outcome, and to react to the consequences of that outcome, must be ensured. tell. Identifying, assessing, profiling, and mitigating the potential negative impacts of AI systems is especially important for those directly impacted. Appropriate safeguards must be in place for whistleblowers, NGOs, trade unions or other organizations when reporting legitimate concerns about AI systems. The use of impact assessments (e.g. red groups or Algorithmic Impact Assessment forms) both before and during the development, implementation, and use of AI systems can be helpful to reduce minimize negative effects. These assessments should be commensurate with the risk posed by the AI system.

Trade-offs: When implementing the above requirements, tension between them can arise, leading to inevitable trade-offs. Such trade-offs need to be addressed rationally and methodically in the modern context. This means that the benefits and values associated with the AI system must be identified and, if conflicts arise, the trade-offs must be clearly acknowledged and assessed in terms of their risks to ethical principles, including fundamental rights. In situations where an ethically acceptable trade-off cannot be determined, the development, implementation, and use of an AI system should not proceed in that manner. Any trade-off decisions need to be properly reasoned and documented. The decision-maker is responsible for the way in which the appropriate trade-offs are being made and must continually review the appropriateness of the resulting decision to ensure that the necessary changes to the system can be made when needed.

Redress: When an unwarranted adverse impact occurs, accessible mechanisms should be foreseen to ensure appropriate remediation. Knowing that you can fix things when things go wrong is key to ensuring trust. Particular attention should be paid to vulnerable people or groups.

2.2.3. AI Act

On April 21, 2021, the European Commission issued the Draft Artificial Intelligence Act (AIA) to codify the standards of its trustworthy AI system. It sets out core horizontal rules for the development, trade and use of AI-driven products, services and systems within the EU, that apply to all industries.

The AIA Draft is divided into a total of twelve “Chapters”, which can be grouped into the following groups:

- Scope of regulation and definition (Title I): defines the broad scope of regulation, relating to the market entry, service entry, and use of AI systems.

- Prohibited Artificial Intelligence Practices (Title II): includes Article 5 AIA - establishes a list of prohibited AIs. As in GDPR, the Commission adopts a risk-based approach. Unlike GDPR, the Commission has actually created a risk classification. A distinction is made between AI systems with unacceptable risk, high risk, and low or minimal risk. Title II covers all AI systems whose use is deemed unacceptable as contravening the Union values.
- High-risk AI systems (Title III): includes special provisions for AI systems that pose a high risk to the health and safety or fundamental rights of natural persons. The classification of an AI system as high risk is mainly based on the intended purpose of the AI system, in addition to its functions. Appendix III specifically identifies several high-risk AI systems. Chapter 3 establishes a set of horizontal obligations on high-risk AI system providers and also imposes commensurate obligations on users and other participants in the AI value chain (e.g., AI value chain). e.g. importer, distributor, agent).
- Transparency Obligation for certain AI Systems (Title IV): Article 52 AIA (Title IV) concerns certain AI systems to account for the specific manipulation risks they pose. Transparency obligations will apply to systems that interact with humans, used to recognize emotions or determine membership in (social) categories based on biometric data or create or manipulate content ("deep fake").
- Measures in support of innovation (Title V): Articles 53 to 55 of the AIA to promote innovation. The goal is to create an innovation-friendly regulatory framework that avoids obsolescence and is resistant to disruption. To this end, national authorities are encouraged to establish regulatory sandboxes and set out a basic framework for governance, oversight and accountability. The AI tuning sandbox creates a controlled environment in which innovative technologies can be tested for a limited time based on an agreed test plan with the relevant authorities. Title V also includes measures to reduce the regulatory burden on SMEs and startups.
- Governance and Implementation: Titles VI, VII, and VIII contain provisions for regulatory oversight and regulatory powers. It is important to ensure that rights are not undermined as the regulation evolves and are consolidated overall.
- Codes of Conduct: Title IX (Article 69 AIA) aims to establish a framework for creating codes of conduct to encourage non-high-risk AI system vendors to voluntarily adopt the requirements required for high-risk AI systems (as outlined in Title III).
- Final Provisions: Among the last clauses (Titles X, XI, XII), Title X should be noted. For companies, Title X (Articles 70 - 72 AIA) is especially important. Among other things, this Title also deals with ensuring effective regulatory implementation. Many of the methods outlined in this section are similar to those in the GDPR. For example, punishments must be effective, proportionate, and inappropriate. The range of fines is more extensive but also revenue oriented, and depending on the violation, ranges from a minimum of 2% to 6% of total annual worldwide sales in the previous financial year or from 10 million to a maximum of 30 million euros, whichever is higher.

On the basis of common awareness and a "balanced" approach that both recognizes the useful effects and identifies the dangers (for regulation) of AI, Europe wants a legal status on new technology generation "encouraging innovation while ensuring the safety of users and consumers." (EU Commission, 2021). Therefore, the key requirement for the new legal rules is to be relevant, flexible and to meet the highest world standards in this regard. AI approach

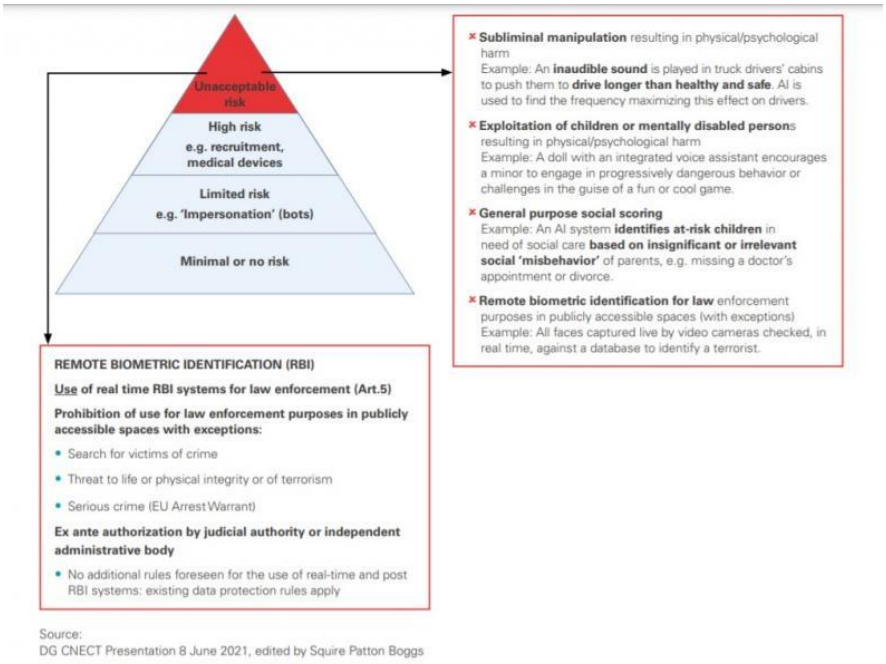
therefore requires legal certainty in the context of flexibility to adapt to future technological developments. (AIA, 2021)

The draft AIA follows a "horizontal risk-based" approach. This means that legally broader requirements exist mainly for AI systems with a higher potential for harm; otherwise, the general minimum requirements will apply. The regulatory approach links general and industry-specific regulatory approaches. The draft regulation should be viewed as a mosaic that will intertwine with other issued regulations (e.g. the Machinery Directive or the General Product Safety Directive). As such, AIA is part of the European Commission's overall digital strategy to create regulatory clarity and develop an "ecosystem of trust in AI in Europe."

The EC proposes a risk-based classification of AI systems with four levels of risk and associated regulatory obligations and restrictions:

(i) Unacceptable risk: AI activities that are particularly harmful, contrary to EU values are prohibited because they violate fundamental rights.

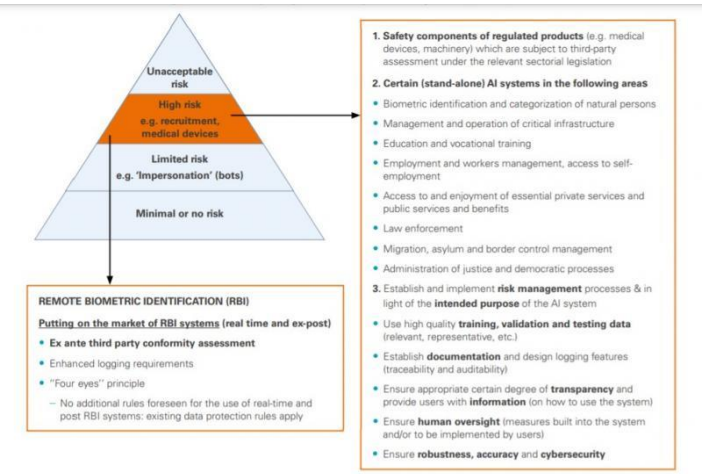
Figure 2: Unacceptable risk



Source: https://www.natlawreview.com/article/proposed-new-eu-regulatory-regime-artificial-intelligence-ai

(ii) High Risk: are systems within an AI system that are intended to be used as a safety component of a product or a product itself, and the product must be assessed for existing quality of third parties (e.g. motor vehicles, trains and aircraft). In addition, the EC reserves the right to directly designate an AI system as high risk by adding it to AIA Annex III, subject to certain criteria.

Figure 3: High Risk AI



Nguồn: https://www.natlawreview.com/article/proposed-new-eu-regulatory-regime-artificial-intelligence-ai

(iii) Limited risk: certain AI systems will only be subject to new transparency requirements, where there is a risk of manipulation (e.g. chatbots) or deception (e.g.: deep fake). Natural persons need to know that they are interacting with an AI system, unless this is obvious in the circumstances and context of use. Law enforcement exceptions exist.

(iv) Minimal Risk: All other AI systems could be developed and used in accordance with existing law without any new legal obligations through AIA. According to the EC, a large number of AI systems currently in use in the EU fall into this category. The EC is recommending voluntary codes of conduct for suppliers of such AI systems.

2.2. Guideline to AI ethics: Cases of Germany and France

2.3.1. Federal Republic of Germany

Germany is a leading country in the development of artificial intelligence, but unlike the US and China, Germany attaches great importance to the ethics of artificial intelligence through the implementation of many measures to implement this issue.

2.3.1.1. Activities for implementing AI ethics

a. Establishment of organizations that study and consult for the government on AI ethics

*** Establishment of an ethics committee on autonomous vehicles**

The "Ethical Committee on Autonomous Vehicles" under the Federal Ministry of Transport and Digital Infrastructure was established in September 2016. The Committee is an interdisciplinary one consisting of senior experts from many fields such as: philosophy, law, social sciences, technology, consumer protection, the automobile industry and the digital economy. It is the first committee in the world to deal with the ethical aspects of autonomous vehicles.

* **Establishment of the commission of inquiry (so-called Enquete Commission) "Artificial intelligence – Social Responsibility and Economic, Social and Ecological Potential"**

This committee is part of the German Federal Assembly and was established in June 2018 with the task of investigating the future effects of AI on social life, economy and employment issues. The committee consists of members of the German Bundestag (as a percentage of the representation of the respective parliamentary group in parliament) and outside experts.

*** Formation of Data Ethics Committee**

The German federal government established the Data Ethics Committee on 18 July 2018 to develop ethical guidelines and recommendations aimed at protecting "individuals, maintaining social cohesion, protecting and promote prosperity in the information age." The 16-member committee is academics, data protection experts and a number of industry representatives. In the view of the Data Ethics Committee, AI is merely one of many possible variations of an algorithmic system, and has much in common regarding the ethical and legal issues it poses. With such a view, the Data Ethics Committee has implemented activities under two issues: data and algorithmic systems (in a broader sense). Since its inception, the Commission has made a series of comments on algorithms and artificial intelligence as well as made recommendations and developed a series of general principles related to the use of data. data and building algorithmic systems.

***Formation of Data Ethics Committee**

The German federal government established the Data Ethics Committee on 18 July 2018 to develop ethical guidelines and recommendations aimed at protecting "individuals, maintaining social cohesion, protecting and promote prosperity in the information age." The 16-member committee is academics, data protection experts and a number of industry representatives. In the view of the Data Ethics Committee, AI is merely one of many possible variations of an algorithmic system, and has much in common regarding the ethical and legal issues it poses. With such a view, the Data Ethics Committee has implemented activities under two issues: data and algorithmic systems (in a broader sense). Since its inception, the Commission has made a series of comments on algorithms and artificial intelligence as well as made recommendations and developed a series of general principles related to the use of data and building algorithmic systems.

b. Developing guidelines and issuing AI ethics code

*** Code of Ethics for Autonomous Vehicle**

Germany has issued the world's first code of ethics for its self-driving car program. These principles were developed by the Ethic Commission on Automated and Connected Driving in June 2017. The code includes rules for software designers, which prioritize issues of “security, human dignity, individual freedom of choice, and individual data autonomy.”

- ✓ Rule 1: the principle of personal autonomy, which means that individuals enjoy freedom of action for which they are themselves responsible.
- ✓ Rule 2: The protection of users should be the most important priority in autonomous vehicle development.
- ✓ Rule 3: State management agencies must be responsible for ensuring the safety of autonomous vehicles when they are licensed to travel on the road.

- ✓ Rule 4: Government regulatory decisions about autonomous vehicles must promote the freedom and protection of individuals as well as ensure the safety of society.
- ✓ Rule 5: While designing autonomous vehicle technology, it is necessary to anticipate all possible emergency situations from the beginning.
- ✓ Rule 6: the protection of human life must be a top priority.
- ✓ Rule 7: When programming for autonomous vehicle systems, programmers should consult an independent state regulatory agency.
- ✓ Rule 8: When programming for autonomous vehicles, it is strictly forbidden to discriminate based on personal characteristics such as: age, gender, physical or mental health.
- ✓ Rule 9: It is necessary to adjust existing legal regulations as well as develop and issue new legal regulations to govern autonomous vehicle systems
- ✓ Rule 10: Liability for damage caused by autonomous vehicle systems must be dealt with according to the same principles as today's liability for manned driving systems.
- ✓ Rule 11: Citizens have the right to be informed about new autonomous vehicle technologies and their implementation.
- ✓ Rule 12: Connectivity and control for all autonomous vehicles must be designed to ensure the safety of road users.
- ✓ Rule 13: Autonomous vehicle systems must be designed to be resistant to cyber attacks, and the software error coefficient must be extremely small, close to zero.
- ✓ Rule 14: For business that use data generated from autonomous vehicle systems, the consent of the vehicle user must be obtained, as well as the provision of personal data will be done voluntarily by the individual.
- ✓ Rule 15: Autonomous vehicle systems should be designed so that accountability can be distinguished between man and autonomous technology in the case of human driving and when autonomous vehicle technology is used.
- ✓ Rule 16: The software and technology on autonomous vehicles must be designed so that, in an emergency, control of the vehicle must be given to the driver.
- ✓ Rule 17: Self-learning systems in autonomous vehicles must be connected to a database in the control center and the control center must be operated by an independent organization based on the standards of reliability. safety, confidentiality….
- ✓ Rule 18: In emergency situations, the vehicle's self-driving software that is activated without human assistance must safely operate the vehicle.
- ✓ Rule 19: It is necessary to teach the use of automated systems in autonomous vehicles to the public, especially in driving schools.

*** Issuing a code of ethics for artificial intelligence**

The Data Ethics Committee of the Federal Republic of Germany offers a view on digital ethics in general and AI ethics in particular as follows: "When designing digital technologies in general and AI in particular, it is necessary to rely on ethical framework in respect of basic human values, rights and freedoms enshrined in the German Constitution and in the Charter of Fundamental Rights of the European Union. According to this approach, the Data Ethics Committee has developed and published an ethical and legal code for digital technology in general and AI in particular [1] as follows:

[1] https://www.bmjv.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_EN.pdf?__blob=publicationFile&v=1

1. "Human Dignity": strictly prohibits acts such as: complete digital surveillance of individuals or activities of deceiving, dominating, insulting... of digital technology towards people.

2. “Self-determination”: Self-determination is a fundamental expression of freedom and includes information self-determination.

3. “Privacy”: potential threats to privacy including collection, evaluation, use of personal data should be prohibited

4.“Protection”: This rule concerns not only the physical and emotional safety of people, but also the protection of the environment and the preservation of extremely important assets. .

5. “Democracy”: digital technologies must prevent manipulation and radicalization from damaging the democracy of society.

6. “Justice and Solidarity”: protecting equitable access to technology and data and equitable distribution of data and technology is an urgent task.

7. “Sustainability”: Digital technology must contribute to the achievement of economic, ecological and social sustainability goals.

*** Ethical principles of artificial intelligence at the company**

Currently, businesses in Germany are also interested in the ethical issues of artificial intelligence, especially those in the field of information technology. The companies have built and implemented the artificial intelligence ethical principles for the products they provide to the market. The artificial intelligence ethical principles address aspects such as:

+ First, fair and accessible to all”.

+ Second, Artificial intelligence must serve society.

+ Third, ethical aspects must be established for AI”.

+ Fourth, the use of data should be transparent and must be protected

+ Fifth, AI must ensure human safety

+ Sixth, humans play a decisive role in the ethical judgment of AI

c. Developing and issuing a national strategy on AI

On 15 November 2018, the Federal Republic of Germany adopted a national strategy on AI jointly implemented by the Federal Ministry of Education and Research, the Federal Ministry of Economy and Energy, and the Federal Ministry of Labor and Social Affairs. The strategy is geared towards ethical AI development, specifically: “Integrating AI with ethical, legal, cultural and institutional conditions based on broad social dialogue and positive political measures.”

To accomplish this goal, the Government has assigned tasks to the Data Ethics Committee so as to research and issue guidelines for the development and use of AI, with particular emphasis on the artificial intellectual ethics aspect. To implement the assigned tasks, the Data Ethics Commission developed a set of guidelines on artificial intelligence ethics and in October 2019. Based on the guidance of the Data Ethics Committee, the government published an interim report presenting the key measures taken by the AI Strategy in November 2019. By October 2020, Research Committee on Artificial Intelligence - Social Responsibility, Economic, Social and Ecological Potential of the German Bundestag presented the final report with specific

recommendations. In December 2020, the Government adopted the AI Updated Strategy. In the strategy, the Federal government advocates the use of an "ethics by design" approach in all stages of the development and use of AI-based applications, as well as participation in dialogue with leading countries and regions on AI development to reach agreement on common guidelines and ethical standards for AI.

d. International cooperation in promoting implementation of artificial intelligence ethics

To promote the artificial intelligence ethics internationally, Germany created an alliance with Canada to develop an ethical AI by establishing the Canada and Germany Conference (GCC) in 2012. The mission of the alliance is to build "trans-Atlantic AI collaborations" with the goal of creating and sustaining the power to advance ethical AI. The cooperation between Germany and Canada differs from China and the US because neither China nor the US intends to create an ethics for AI. Canada and Germany's interest in an AI ethical standard forged an alliance to increase information availability and set a precedent for an international willingness to collaborate.

2.3.1.2. Existing legal bases that can be used to conduct artificial intelligence ethics

Currently, the Federal Republic of Germany has not yet developed and enacted a specific law directly related to the ethics of artificial intelligence. However, some current German legislation may regulate some aspects of artificial intelligence ethics, specifically:

*** Data Protection Law**

The German data protection law (German Bundesdatenschutzgesetz - BDSG) is aimed at ensuring that personal data can only be used with the consent of the individual. The fundamental ethic values protected by law in this respect are the right to privacy and personal autonomy. Several aspects of the law can be used to implement AI ethics as follows[2]:

+ Personal data must be collected directly from the relevant person (Clause 4, Article III of the BDSG).

+ All data processing systems should achieve the goal of not using (or as little as possible) personally identifiable data.

+ If personal data is collected, the responsible entity must inform the affected person of their identity and the purpose of collection, processing or use (clause 4, Article III of the BDSG).

+ Prohibit the collection, processing and use of personal data, unless permitted by law or the relevant person agrees (Clause 4, Article I of the BDSG).

+ Personal data must be collected directly from the relevant person (Clause 4, Article III of the BDSG).

All data processing systems should achieve the goal of not using (or as little as possible) personally identifiable data.

+ If personal data is collected, the responsible entity must inform the affected person of their identity and the purpose of collection, processing or use (clause 4, Article III of the BDSG).

*** Telecommunications Law**

[2] "Begriff und Geschichte des Datenschutzes". 28 May 2014

On February 17, 2017, the German Federal Network Agency banned the sale of a doll named Cayla and ordered the destruction of all devices sold[3]. The legal basis of this decision is No. 2 Article 90, German Telecommunications Law. The reason is that since the doll has a connection to the manufacturer (required because the doll is AI enabled), the doll is actually a spy that watches over the child, recording all the data the child tells to devices, including the child's secrets. The agency warned that these devices could be hacked, exposing children to threats such as pedophilia or ideological influence[4]. Since then, the German regulator has used the telecommunications law to ban devices similar to smart watches. This rigorous method is applied to protect children, one of the most vulnerable ones.

*** Proceedings in taking evidence in court**

Another example of current regulatory applicability for AI is the use of technical applications to obtain evidence in court proceedings. According to Section 244(3)2 of the German Criminal Procedure Code, it is stated that "an application for evidence may be refused if the evidence is completely inconsistent".

On this legal basis, the German Federal Court of Justice has unequivocally decided that the evidence obtained using a polygraph-based method known as "polygraph" is completely inappropriate appropriate and therefore cannot be relied upon for judicial decision-making purposes. This ruling was further confirmed by other German courts that "the "polygraph" method is not generally unequivocally accepted by the experts concerned as an accurate and reliable method of obtaining evidence".

On this legal basis, the German Federal Court of Justice has unequivocally decided that the evidence obtained using a polygraph-based method known as "polygraph" is completely inappropriate appropriate and therefore cannot be relied upon for judicial decision-making purposes. This ruling was further confirmed by other German courts that "the "polygraph" method is not generally unequivocally accepted by the experts concerned as an accurate and reliable method of obtaining evidence".

In addition, "polygraph" is based on statistical data that cannot be extrapolated to individual cases. Finally, tests based on "polygraph" are easy to manipulate [4]. It can therefore be concluded from the existing German case law that at least for the time being AI-driven applications cannot be relied upon for evidence in court proceedings.

2.3.2. Ethical implementation of artificial intelligence in France

2.3.2.1 Some solutions for ethical implementation of artificial intelligence

a. Establishment of organizations to research and advise the government on implementation of AI ethics

[3] Press Release, Bundesnetzagentur Removes Children's Doll "Cayla" From the Market, Bundesnetzagentur [BNetzA] [German Federal Network Agency], (Feb. 2, 2017).

[4] Kay Firth-Butterfield, Generation AI: What happens when your child's friend is an AI toy that talks back?, WORLD ECONOMIC FORUM (May 20, 2018) https://www.weforum.org/agenda/2018/05/generation-ai-whathappens-when-your-childs-invisible-friend-is-an-ai-toy-that-talks-back/. For a legal analysis that was also referred to by the German Federal Network Agency, see Stefan Hessel, "My friend Cayla" - eine nach § 90 TKG verbotene Sendeanlage?, JurPC Web-Dok. 13/2017, Abs. 1–39, http://www.jurpc.de/jurpc/show?id=20170013 (last visited Oct. 6, 2018).

*** Establishment of a digital ethics committee**

In December 2019, the National Committee on Digital Ethics (FNCDE) was established, whose mandate is to "submit comments and recommendations on the ethics of digital techniques, technology, its use and innovating and identifying the relevant equilibrium for holding public debates on digital ethics and artificial intelligence". This committee consists of 27 members who are experts from different fields such as: digital technology, academics, doctors, lawyers and representatives of organizations and government agencies. The mission of the committee is to provide some guidance and opinions on the ethical issues posed by digital technology and artificial intelligence. It will work closely with a number of public agencies and organizations such as: National Commission on Informatics and Liberty, National Institute for Research in Digital Science and Technology, the National Center for Scientific Research (CNRS) and French Digital Council (CNNum).

The Commission has been asked by the French Government to comment on ethical issues related to three specific topics of digital applications such as:

- Simulation of human chats (chatbots) on mobile phones and smart home devices: Ethical issues related to transparency and information in the processing of collected data.

- Self-driving cars: the Commission will analyze specific general liability for manufacturers, insurers and users.

- Medical diagnostics and AI: The committee will discuss the risks when predictive analytics are not followed.

*** Establishment of the National Commission on Informatics and Liberty (CNIL)**

The National Commission on Informatics and Liberty (CNIL) is an independent French administrative agency established in 1978, whose job it is to ensure that data privacy laws are applied to the collection, storage, and retrieval of data storage and use of personal data. In 2017, this committee was given an additional task by the French government to reflect on the good use of AI and the need to monitor these activities. Therefore, CNIL is considered one of the standards and legitimate bodies in terms of regulating engineering and technological development. To carry out the assigned task, CNIL organized a series of ethical debates on AI from January 2017 to October 2017 with 45 debates. These public debates aim to ensure that "the French AI model must address ethical issues such as: respect for privacy, protection of personal data, transparency, and accountability processes of actors and contribute to the welfare of society"[5]

b. Develop and issue guidelines and rules for implementing artificial intelligence

*** Principles issued by the state council**

In 2014 annual report on digital technology and fundamental rights, the State Council called for a "review of the basic principles for the protection of fundamental rights". The first of these concerns the principle of "informational self-determination" to ensure data actors control the communication and use of their personal data. The next concern is the principle of "fairness" which applies, not to all algorithms, but in a more limited way to "platform". According to the French Council of State, "fairness includes the honest guarantee of a search engine optimization

[5] https://blog.advalo.com/en/data-and-ai-the-ethical-implications-and-recommendations-of-cnil-the-national-commission-on-informatics-and-liberty0

(SEO) or ranking service, without seeking to alter or manipulate it for the purposes of purpose not for the benefit of the user"[6]. This means that the information must be made available to the user from the outset.

*** Villani report on "for a meaningful artificial intelligence"**

Cédric Villani is a mathematician, 2010 Fields Prize laureate and member of the French parliament who was commissioned by the Prime Minister to examine the development of an AI development strategy in 2017. Since being appointed. until March 2018 he produced a 147-page report called the "Villani Report" entitled "For a Meaningful Artificial Intelligence". In this report, Villani has laid out some principles for ethical implementation of artificial intelligence.

Principle of "*Opening the black box"*

In algorithmic system, we can observe input data and output data but the inner workings of the system are not well understood (black box). Nowadays, human ignorance is mainly due to changes in the "self-learning" model, so the "black box" needs to be clarified to the user based on the requirements:

+ Accountability of decisions made by machine learning systems: the accountability of this technology is one of the conditions for social acceptance.

Fairness, bias and discrimination: legislation will have to control the ethical performance of AI systems. This is also important in the case of litigation between different parties who oppose the decisions made by the AI system.

+ Development of AI Audits: Provides formal checks by performing audits of algorithms to confirm a party's doubts or claims.

Support for accountability research: there is an urgent need to support research that understands the nature of AI by investing in the same three lines of research: how to make models easier to understand, how to create Smarter users' interface and understand cognitive mechanisms in the workplace to generate satisfactory interpretations. Each of these areas involves a variety of skills, covering not only computer science and mathematics but also design, neuroscience and psychology, and emphasizing the need for interdisciplinary collaboration to understand how things work.

Principle of *"Ethical considerations from the design stage"*

+ Integrating Ethics in Training AI Engineers and Researchers: The purpose of ethics teaching is to pass on to future architects the conceptual tools they need to identify and confront the ethical issues they will encounter in their professional activities. In addition, keeping in mind the practical issues related to the protection of privacy, discrimination and intellectual property, they need practical guidance in order to make connections between normative theories (professional ethics) and application in practice.

Discrimination Impact Assessment: it is the responsibility of data users to self-assess the impact of their activities and to take appropriate remedial action and, in the event of an audit, to can

[6] Il s'agissait de "soumettre [les plateformes] à une obligation de loyauté envers leurs utilisateurs (les non professionnels dans le cadre du droit de la consommation et les professionnels dans le cadre du droit de la concurrence)". Les plateformes apparaissent comme des acteurs classant un contenu qu'il n'a pas lui-même mis en ligne.

demonstrate that all necessary measures have been taken to give them complete control of the process.

Artificial Intelligence Ethical Guidelines promulgated by the National Commission on Freedom and Information Technology (CNIL)

CNIL has introduced two basic principles to "put artificial intelligence at the service of people"[7].

The first principle is "fairness of algorithms", which builds on the principle proposed by the French Council of State in 2014. Accordingly, the principle is based on the idea of fairness for to product users, not only as consumers but also as citizens and even to communities whose lives may be affected by algorithms whether or not these algorithms process data whether personal or not.

The second principle is "attentiveness and vigilance". According to this principle, individuals and those who form the links of the algorithmic chain should be provided with the means to observe, perceive, and always seek answers in the digital society. Besides, it also involves other important actors in society such as: businesses to model algorithmic systems as well as other ethical systems.

c. Formulate and promulgate a national strategy on artificial intelligence

In March 2018, French President Emmanuel Macron presented a five-year AI national vision and strategy. The French AI strategy, titled "AI for people", was developed on the basis of the AI policy report prepared by the famous mathematician Cédric Villani. France has been developing a national AI strategy titled "AI for humanity" for about a year. Ethical issues to ensure fair and transparent use of AI technologies and algorithms are central to France's AI strategy [8].

* The main goals of the French AI strategy are:

- Improve the AI education and training ecosystem, retain and attract world-class AI talents;

- Establish an open data policy to deploy AI applications

- Develop an ethical framework for transparent and fair use of AI applications.

Among the actions outlined in the French AI strategy, there are a number of actions related to the implementation of artificial intelligence ethics.

- Planning the impact of AI on labor. Three specific measures are proposed: to establish a public laboratory on job change with the task of studying changes in the labor market; supporting those affected by the transition of industries using AI; implementing new methods of funding vocational training.

- Opening the AI's black box. Algorithm transparency and mechanisms to test them must be developed to ensure that AI is accepted by society. AI applications must be more explainable, the design of user interfaces must be intuitive and easy to understand and understand. To create a sense of responsibility, ethics training should be a must for AI engineers and researchers. Assessing the impact of discrimination should be part of the development of the algorithms used to develop AI.

2.3.2.2. Some legal regulations can implement artificial intelligence ethics

[7] https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf
[8] https://www.aiforhumanity.fr/en/

**** Data Protection Law***

The Data Protection Act was enacted in 1978, there are certain provisions that can be summarized in three principles that follow a single general principle set forth in Article 1 of the Act "the processing of data will made to serve all people.

First, the law governs the use of personal data necessary for the operation of the algorithms, beyond the rigorous algorithmic processing stage. In other words, it governs the conditions of data collection and retention, as well as the exercise of the data subject's rights (right to information, right to object, right of access, right to rectification) [9] to protect the privacy and freedom of individuals.

Second, the Data Protection Act prohibits machines from making decisions alone without human intervention when there are significant consequences to the subject's data (for example, a court judgment or a loan decision)[10].

Third, the law provides that agents have the right to collect data from information controllers related to algorithm-based processing [11].

**** Data Protection Regulation (GDPR)***

The impacts of AI on the data protection sector should be addressed by referring to article 22 "*Automated personal decision-making, including profiling*" of the General Data Protection Regulation (GDPR)[12], specifically, article 22 has 4 items.

**** Some other laws***

In addition to the above legal provisions, currently in France to regulate the implementation of artificial intelligence ethics can also be applied through a number of laws such as:

The French IP Code, which defines the rights that parties can claim: AI components (e.g., database or document permissions (e.g. photos, videos) used to train AI models); and AI-generated elements (e.g. software, works).

A constitutional bill regarding the Charter of Artificial Intelligence and Algorithms was added to the agenda of the French National Assembly on January 15, 2020.

The Computer Fraud Act 88-19 on January 5, 1988 is about computer frauds, regulating crimes related to any automated data processing system.

Military Programming Act 2013-1168 on December 18, 2013 (from the year 2014 to 2019) and the Military Programming Act of 2018-607 on July 13, 2018 (from the year 2019 to 2025) issues requirements on great importance operators.

[9] Principles of purpose, proportionality, security and limitation of the data storage period.
[10] Article 10 of the 1978 Act
[11] Article 39 of the 1978 Act. Article 15.1 (h) of the General Data Protection Regulation (GDPR) provides that the data subject shall have the right to obtain from the controller the following information: "the existence of automated decision making including profiling referred to in Article 20(1) and (3) and at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject". The legal limits laid down by the GDPR particularly concern "profiling" (no decision based solely on processing, subject to certain exceptions)
[12] Article 22 of the GDPR

The Network and Information Systems Directive (2016/1148) and the Law 2018-133 on February 26, 2018 (and Decree 2018-384), provide specific requirements for providers who exploit essential services and digital service;

Regulation 2019/881 on 17 April 2019 sets out the key requirements for European cyber security certification schemes for information and communication technology products, services and processes.

2.3. Comments

First of all, it must be affirmed that the EU and its members do not consider AI as a subject but only a technique and a foundation to serve people. When it is not the subject, the AI cannot be held responsible for its own errors. The EU approach avoids any sci-fi speculation about AI. Strictly speaking, the EU sees AI as a technology to solve problems and perform tasks, not as some kind of Frankenstein monster. Therefore, the EU regulation excludes the possibility of assigning AI systems any legal status, with rights and obligations, such as the ability to own assets, to enter into contracts, to sue and be sued, etc. The responsibility of any AI system rests solely with those who design, manufacture, market, and use it. Coherently, the EU views emphasizes the importance of human oversight throughout their documents.

The EU Strategy was the first international initiative on AI and supports the strategies of individual Member States. Strategies vary however in the extent to which they address ethical issues. At the European level, public concerns feature prominently in AI initiatives.

In the EU's development strategy on AI, the EU takes a "human-centric" approach to AI. The aim is to establish a framework for the ethical and trustworthy development of AI technologies, and applications in line with European values, and to prepare the groundwork for a global alliance in this area.

Europe's approach to artificial intelligence (AI) aims to help build a resilient Europe in the Digital Decade, where people and businesses can enjoy the benefits of AI. The EU strategy focuses on two areas: AI excellence and AI trustworthy. The European approach to AI will ensure that any AI innovation is based on rules that protect the functioning of markets and the public sector, as well as human safety and fundamental rights.

Trustworthy AI Guide aims to provide guidance for general AI applications, building a horizontal foundation to achieve trustworthy AI. However, different situations present different challenges. Each specific field has different ethical concerns. AI music recommendation systems do not raise the same ethical concerns as medical AI systems. Likewise, different opportunities and challenges arise from AI systems used in the context of different relationships between different actors: business to consumer, business to business, employers versus employees and the public, or generally in different sectors or use cases. Given the specific contextual characteristics of AI systems, the implementation of this Guide should be tailored to the specific AI application. Furthermore, the need for a complementary industry approach should be explored, to complement the more general horizontal AI guideline framework proposed in the Guidelines.

The foundation of trustworthy AI is based on fundamental rights and is reflected by four ethical principles that need to be followed to ensure AI is ethical and trustworthy. Trustworthy AI is heavily focused on the area of ethics. AI ethics is a branch of applied ethics that focuses on the ethical issues posed by the development, implementation, and use of AI. Its primary concern is to

determine how AI can promote or enhance individuals' concerns for the good life, whether in terms of quality of life, or the human autonomy and freedoms is necessary for a democratic society. The guidance outlines four ethical principles, rooted in fundamental rights, that must be respected to ensure that AI systems are reliably developed, deployed, and used.

- Respect for human autonomy
- Prevention of harm caused by AI system
- Fairness
- Explicability

The AIA is the step that codifies the rules of the EU Guidelines on Trustworthy AI. Overall, the aim of the European Commission is to create a harmonized regulatory framework on the basis of the Act to strengthen the internal market. The AIA applies to providers who bring AI systems to market or put them into services, regardless of whether those providers are established inside or outside the EU. Users of AI systems are subject to the new rules if they are in the European Union. The AIA will even apply if the supplier and user are located outside the EU, but the output generated by those systems is used within the EU. The AIA will therefore apply to vendors of AI systems (e.g. a developer of a CV screening tool), as well as users of such AI systems (e.g. a bank purchasing this CV screening tool). It will not apply for private, non-professional use. In addition, AI systems developed and used exclusively for military purposes are exempt.

The draft AIA follows a "horizontal risk-based" approach. This means that legally broader requirements exist mainly for AI systems with a higher potential for harm; otherwise, the general minimum requirements will be applied. The approach links general and industry-specific regulatory approaches. The draft should be viewed as a mosaic that will intertwine with other issued regulations (e.g. the Machinery Directive or the General Product Safety Directive). As such, AIA is part of the European Commission's overall digital strategy to create regulatory clarity and develop an "ecosystem of trust in AI in Europe."

The definition of high-risk AI systems, an important concept of AIA, although such a classification still seems somewhat ambiguous. Some of the AI systems discussed are low-risk or zero-risk, and as such, these risks appear to be outside the scope of strict compliance and subject only to voluntary codes of conduct. However, it is unclear how this classification will work in practice, leaving too much room for uncertainty and loopholes. And in the case of high risk, it is proposed to explicitly combine the two angles from which it would be more appropriate to distinguish. On the other hand, there are high-risk AI systems because critical problems depend on their normal operation. Think of an autonomous driving system: it's a "good thing" that can't be left to work. On the other hand, there are high-risk AI systems because if they are used inappropriately, they can cause significant troubles; think of the abuse of real-time, remote biometric identification for law enforcement purposes, a type of technological surveillance prohibited under the proposal. This is a “bad thing” that should not be put to work. If one fails to distinguish between these two aspects of a high-risk system — something high risk if it doesn't work and something high risk if it works — one can get confused about a “good” AI system must have, with resistance to “bad” AI systems. Note that conceptual confusion and uncertainty about the specific nature of the risks associated with the design, development, and implementation of AI systems will also affect the feasibility of any conformity assessment

(Floridi et al. (2018), Mokander and Floridi (2021), an important element in the AIA and the certification system it proposes.

From an ethical perspective, AIA inherits the same basic approach as in GDPR: based on the protection of human dignity and fundamental rights. However, the AIA seems a bit top-down, less flexible and less focused on protecting citizens and their rights than GDPR.

AIA uses the term "people-centric", whereby the AIA approach puts people at the center of technological development. This sounds very true at first, but is a bit vague. It is clear that any technology, including AI, must serve its people, values, and needs. Yet despite it all, "human-centric" seems to be synonymous with "humanistic," and we know the planet has suffered from humanity's obsession with importance and centrality as if everything should always serve it, including every aspect of the natural world, regardless of cost and loss. Although the terminology is vague, the EU's vision is fundamentally correct, with AIA emphasizing the value of AI as a technology that can be very "green" and provide exceptional support against pollution and climate change. and for the sustainable development of the information society (Cowls et al., 2021). In it, the AIA approach reinforces the idea that environmental protection should be a cross-cutting issue at the EU level.

Regarding ethical approach, AIA explicitly adopts the ethical principles proposed by HLEG and seeks to eliminate or reduce the risks of AI, supporting public trust in innovative technologies while continuing to develop and apply AI in the EU. This risk-based approach seems persuasive (this is the common approach to internal market-based legislation) and is consistent with the view that ethics benefits markets, not must be opposite. But precisely for this reason, one could argue that AIA could do much more to protect consumers' interests and be much more aggressive in providing solutions to remedy possible harm or losses caused by the AI system.

3. Policy suggestions for Vietnam

3.1. AI and AI ethics in Vietnam

The Vietnamese government has identified artificial intelligence as one of the breakthroughs and spearhead technologies of the industrial revolution 4.0. Since 2014, AI has been included in the high-tech list with priority for development investment. Over the years, the Government has approved the list of high-tech products prioritized for investment and the list of high-tech products encouraged for development, in which AI is included in the list of high-tech prioritized for development investment. The Government's will to develop AI is also reflected in a series of documents such as: (i) Directive No. 16/CT-TTg dated May 4, 2017 of the Prime Minister on enhancing accessibility Industry 4.0 identifies AI technology as one of the breakthroughs and spearhead technologies of Industry 4.0; (ii) The Resolution of the Government meeting in March 2018 set out the task of researching, assessing the impact and developing the National Strategy on Industry 4.0.

In the process of formulating the Strategy, AI technology will be carefully evaluated and researched to form specific strategies, roadmaps, solutions for AI development and application that best suit the reality and potential of the company. Vietnam; and (iii) especially the National Strategy on Research, Development and Application of Artificial Intelligence to 2030, issued on January 26, 2021 with the goal of becoming a center of innovation and development solutions and applications of artificial intelligence in the ASEAN region and the world.

It can be said that Vietnam has made initial moves in the process of establishing a legal system for the development and application of artificial intelligence in life. The reality of AI development in Vietnam in recent years shows that AI application has contributed to the creation of new technology products, promoting the development of many fields such as information and communication technology, healthcare, and travel.

3.2 Policy suggestions

With the current development of technology, in the very near future, artificial intelligence systems will affect our lives, becoming more and more popular and pervasive. These breakthroughs seem to come faster than expected, so it is necessary to have urgent and timely policies to keep up with the rapid changes of society. Europe is a region with many leading countries in AI technology, the emergence of new challenges from AI development is inevitable. The question is, how does Europe face these challenges?

The above challenges are controversial within the European Union, raising questions about the ethics of artificial intelligence. The process of designing, applying and implementing AI is considered to have the potential risks to undermine basic human rights. In other words, AI can have effects on the way people perceive the outside world if lacking of transparency, accountability, fairness and honesty. Today, Europe is leading the way in devising ethical principles to ensure that the values of the European Union are respected. Values such as inclusion, tolerance, fairness, solidarity, non-discrimination are indispensable in the European way of life. Besides, there are other values such as dignity, freedom, democracy, justice, the rule of law and human rights.

As such, the Commission has supported dual-targeted investment and management to accelerate the uptake of AI use and address the risks associated with the use of some of these new applications. Automated processing of citizens' health, work, and welfare data can lead to decisions with discriminatory and unfair results. The “dark side” of algorithms in decision-making is addressed through a set of European Union principles. Especially in the case of high-risk systems, these principles will ensure automated decision-making processes compatible with human rights and guarantee democracy. Unlike the two leading countries in AI technology, namely the US with a profit-based approach, China with an approach focused on national security and widespread surveillance, the European Union has the "people-centric" approach to orient development strategies to achieve the above dual goals.

In addition, concepts related to ethics may be understood differently across cultures or communities. Western traditions tend to value individual privacy, but this is not thought to be a matter of great importance in the East. Confucianism emphasizes collective interests more than individual interests, so the concept of individual privacy has traditionally received less attention and is sometimes negative. Currently, the ethical debate is about the desire that AI systems are built on human values. However, who will decide the value of people? In the context of each nation, each country, and region, there are different conceptions of morality and social norms. As a result, countries and companies suggest that a global framework is needed to ensure AI technologies operate on commonly ethical standards.

With the goals set out in the "Summarization Report on the implementation of the 10-year socio-economic development strategy in the period of 2011-2020 and building the 10-year socio-economic development strategy in the period of 2021-2030" of the Central Committee namely

“strongly developing science, technology and innovation to create breakthroughs to improve productivity, quality, efficiency and competitiveness of the economy", the development of the digital economy with the focus of development is that AI technology will be a development step in the coming period. In the process of developing and applying AI technology, Vietnam needs to take into account the ethical issues of artificial intelligence that may arise as mentioned above, as well as need to take into account the creation of its own values shared global values to address the ethics of artificial intelligence. In addition, Vietnam also needs to take steps to prepare for the development of institutions related to AI technology before the technology comes to life and affects people's lives.

The knowledge about artificial intelligence or the delay in investment and development in this technology field can widen the technology gap between Vietnam and the world, especially developed countries. Therefore, Vietnam needs to have solutions to build artificial intelligence technology infrastructure to meet the requirements of science and technology development. It is necessary to have specific goals and plans for regions in order to have directions focused on transforming infrastructure and transforming the structure of the low-value-added labor market into a high-value production market with targeted investments in research and development in artificial intelligence technology.

Only then will Vietnam be able to catch up with AI technology and application with other countries in the region and the world. In addition, institutions need to be able to deal with future ethical problems caused by ideological differences. On the other hand, in a short time, artificial intelligence technology will be integrated in all aspects of life, it is necessary to develop ethical principles right from the design, use and implementation of artificial intelligence technology, avoiding the institutions having to follow technology as well as ensuring the traditional values of Vietnam. Finally, it is necessary to clearly define the role of actors: state or non-state in the management and development of artificial intelligence technology.

As a small country, a technology receiver but not a source technology country, Vietnam does not aspire to be the world leader in this field. However, participating in the global forum on AI and AI ethics from the very beginning is an opportunity for Vietnam to express and put its views and values into building a mechanism for and a global common understanding about AI issues.

AI is new to the world in general and to Vietnam in particular. Therefore, it is very important to build a knowledge base on AI and AI ethical issues, and policy programs for the development, management, application and use of AI. This helps the Government, businesses and people of Vietnam get ready for the global playing field in this field.

Conclusion

Industry 4.0 with the explosive development of new generation of cross-industry technologies has opened up a development era associated with artificial intelligence (AI), the Internet of Things (IoT). Along with that, the explosive development of the digital economy creates both great opportunities and challenges for the development of each country, nation as well as of each community, individual and business. Although being criticized for focusing too much on legal and ethical guidelines, the EU realizes the advantage of being at the forefront as a power union in the field of AI ethics by setting global standards of design and usability as well as clear legal assurance in AI-based applications. As embodied in the General Data Protection Regulation (GDPR), the EU's strategic advantage lies mainly in its market, regulatory and regulatory

powers. However, while the European Commission's digital sovereignty agenda can help foster certain AI developments in Europe, it is equally important for the EU to work closely with like-minded global partners to establish common AI standards and regulations.

The ethical issues of artificial intelligence mentioned in the report are the problems that Europe faces in the process of developing AI technology as well as the digital economy. These ethical concerns are also problems in Vietnam as well as many other countries in the process of the development of AI technology. By highlighting the above challenges of Europe, we hope to be able to provide suggestions for Vietnam in the process of formulating strategies and policies for AI technology development. In addition to having strategies, mechanisms and policies to promote the strong development of AI technology and apply this technology to promote the development and position of Vietnam in the international arena, the Government also needs to anticipate non-technological problems that may arise from the development and application of AI technology to ensure the values that Vietnam wishes to preserve and aim at. Vietnam needs to locate its position in the world and regional AI map, from which there are appropriate strategies and policies to be able to reconcile ethical issues and viewpoints in the AI management strategy with other countries and regions and ensure their own set of principles.

With a human-centric view as the guideline for the European legal system when technology applications focusing on artificial intelligence are absolutely right, in line with the values that Europe only has maintain. The EU position is that artificial intelligence should not be developed for the purpose of creating an artificial moral agent — that is, making machines should not be designed with an evolutionary ethic in mind; machines must remain an Artificial Agent (AA), instead of being upgraded to an Artificial Ethical Agent (AMA). The European Union emphasizes that building responsible, reliable AI is not a must, but an imperative. Black box models are not only “accurate” but must also meet reliability characteristics to facilitate open collaboration and ensure ethical results. There is still a long way to go in implementing responsible AI models in practice when commercial AI applications are in high use and millions of people are facing problems.

Reference

1. Ban Chấp hành trung ương (2020), "Báo cáo Tổng kết thực hiện chiến lược phát triển kinh tế - xã hội 10 năm 2011-2020, xây dựng chiến lược phát triển kinh tế xã hội 10 năm 2021 -2030", https://nhandan.vn/tin-tuc-su-kien/bao-cao-tong-ket-thuc-hien-chien-luoc-phat-trien-kinh-te-xa-hoi-10-nam-2011-2020-xay-dung-chien-luoc-phat-trien-kinh-te-xa-hoi-10-nam-2021-2030-621156/.
2. Thủ tướng chính phủ (2021). *Chiến lược quốc gia về nghiên cứu, phát triển và ứng dụng trí tuệ nhân tạo đến năm 2030.*
3. Hoàng Vũ Linh Chi và Hồ Thanh Hương (2021), Đạo đức trí tuệ nhân tạo ở Liên minh Châu Âu, Tạp chí Nghiên cứu Châu Âu, 2021
4. Artificial Intelligence Act. (21 April 2021). "Proposal for a regulation of the European Parliament and the Council laying down harmonized rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts." *EUR-Lex - 52021PC0206*https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELLAR:e0649735-a372-11eb-9585-01aa75ed71a1.
5. Anderson, M., Anderson, S. L. (2007). Machine ethics: Creating an ethical intelligent agent. AI Magazine, 28(4), 15.
6. Axelle Lemaire, Romain Lucazeau, Tobias Rappers, Fabian Westerheide, and Carly E. Howard (2018), "Artificial Intelligence – A Strategy for European Startups: Recommendations for Policymakers," Roland Berger and Asgard – Human Venture Capital, 2018, 7, https://asgard.vc/wp-content/uploads/2018/05/Artificial-Intelligence-Strategy-for-Europe-2018.pdf.
7. Anderson, M., Anderson, S. L. (2007). Machine ethics: Creating an ethical intelligent agent. AI Magazine, 28(4), 15.
8. Coeckelbergh, M. (2020). AI ethics. The MIT Press.
9. Dezfouli, A., Nock, R., & Dayan, P. (2020). Adversarial vulnerabilities of human decision-making. Proceedings of the National Academy of Sciences of the United States of America, 117(46), 29221–29228. https://doi.org/10.1073/PNAS.2016921117/-/DCSUPPLEMENTAL
10. Druga, S., Williams, R., Park, H. W., & Breazeal, C. (2018). How smart are the smart toys?: children and parents' agent interaction and intelligence attribution. Undefined, 231–240. https://doi.org/10.1145/3202185.3202741
11. Etzioni, A. E. & O. (2016). AI assited ethics. Ethics and Information Technology, 18, 149–156. https://ssrn.com/abstract=2781702
12. European Parliament. (2020). The ethics of artificial intelligence: Issues and initiatives - Think Tank. https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_STU(2020)634452
13. Google. (2018). AI at Google: our principles. https://www.blog.google/technology/ai/ai-principles/
14. General Data Protection Regulation. (27 April 2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). *EUR-Lex - 32016R0679*https://eur-lex.europa.eu/eli/reg/2016/679/oj.

15. HLEGAI. (2019). High-Level Expert Group on Artificial Intelligence, EU - Ethics guidelines for trustworthy AI. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.
16. HLEGAI. (18 December 2018). High-Level Expert Group on Artificial Intelligence, EU - Draft ethics guidelines for trustworthy AI. https://digital-strategy.ec.europa.eu/en/library/draft-ethics-guidelines-trustworthy-ai.
17. HLEGAI. (2019a). High-Level Expert Group on Artificial Intelligence, EU - Policy and investment recommendations for trustworthy artificial intelligence. https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence.
18. Human-Centered Artificial Intelligence, S. U. (2021). Artificial Intelligent Index 2021. https://aiindex.stanford.edu/wp-content/uploads/2021/03/2021-AI-Index-Report_Master.pdf
19. Luciano Floridi (2021), The european Legislation on AI: a Brief Analysis of its Philosophical Approach, Philos Technol. 2021 Jun 3, p 1-8, doi: 10.1007/s13347-021-00460-9, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8174763/
20. Müller, V. C. (2020). Ethics of Artificial Intelligence and Robotics. In The Stanford Encyclopedia of Philosophy.
21. Prates, M., Avelar, P., Lamb, L. C. (2018). On quantifying and understanding the role of ethics in AI research: A historical account of flagship conferences and journals. ArXiv, 1–13.
22. Schiff, D. (2021). Out of the laboratory and into the classroom: the future of artificial intelligence in education. AI and Society, 36(1), 331–348. https://doi.org/10.1007/S00146-020-01033-8
23. Tene, O., Polonetsky, J. (2013). Big Data for All: Privacy and User Control in the Age of Analytics. Northwestern Journal of Technology and Intellectual Property, 11.
24. Yu, H., Shen, Z., Miao, C., Leung, C., Lesser, V. R., & Yang, Q. (2018). Building Ethics into Artificial Intelligence. ArXiv, 1–8.