Subcommittee of the Israeli National Intelligent Systems Project on Artificial Intelligence Ethics & Regulation

REPORT

November 2019

* The following report was written by the ethics and regulation team, which was part of the Israeli National Intelligent Systems Project headed by Prof. Ben-Israel and Prof. Matania. The final project report and its conclusion are currently in their final work stages and will also include some of the conclusions of the present report, following relevant adjustments.

Committee Members

Chair: Prof. Karine Nahon, Interdisciplinary Center (IDC) Herzliya and President of the Israel Internet Association (ISOC-IL)

Secretary: Dr. Dalit Ken-Dror Feldman, Legal Supervisor, Law Technology and Cyber Clinic, University of Haifa, lecturer at the Zefat Academic College

Authors: Prof. Karine Nahon, Dr. Tehilla Shwartz Altshuler, Adv. Amit Ashkenazi, Dr. Ran Gilad Bachrach, Dr. Dalit Ken-Dror Feldman, Mr. Amit Keren

Members

1. Dr. Tehilla Shwartz Altshuler, Head of Democracy in the Information Age Program, Israel Democracy Institute
2. Adv. Amit Ashkenazi, Head of Legal Dept., National Cyber Directorate, Prime Minister's Office
3. Dr. Ran Gilad Bachrach, Microsoft Research
4. Adv. Yaara Ben Shachar Tyk, Legal Advisor, ICT Authority, Prime Minister's Office
5. Oleg Brodt, Chief Innovation Office, Cyber@ Ben-Gurion University; R&D Director at Deutsche Telekom Innovation Laboratories Israel
6. Uri Eliabayev, AI business consultant & founder of the Machine & Deep Learning Israel Community
7. Prof. Niva Elkin-Koren, Director of the Haifa Center for Law & Technology and Center for Cyber Law & Policy
8. Yoav Evenstein, Head of the Israeli delegation and expert member of the ISO/IEC AI International Standardization Committee; as well as expert member of the IEC Meta Group for AI Ethics
9. Adv. Yoram Hacohen, CEO, Israel Internet Association (ISOC-IL)
10. Shlomi Hod, data scientist
11. Amit Keren, Managing Director of Deutsche Telekom Israel
12. Dan Kotliar, Dept. of Sociology & Anthropology, Hebrew University of Jerusalem
13. Dr. Neal Naimer, R&D Policy Advisor, National Council of R&D at the Ministry of Science & Technology
14. Inbar Naor, data scientist, Taboola
15. Zafrir Neuman, Chief Legal Counsel, National Technological Innovation Authority
16. Shay Palachy, data science & computational learning consultant
17. Dr. Roy Schöndorf, Deputy Attorney General (International Law)
18. Idan Shaer, cyber & AI consultant

Observers:

Shira Rivnai Bahir, Dept. of Sociology & Anthropology, Ben-Gurion University

Limor Shmerling Magazanik, Managing Director, Israel Tech Policy Institute

Disclosure: The opinions expressed in the following report do not necessarily represent those of the organizations to which the authors and other committee members belong.

Chapter 1: Introduction

This Committee is a subcommittee of the National Intelligent Systems Project, dedicated to the ethical, legal and regulatory aspects of artificial intelligence (AI). Specifically, the Committee has been commissioned to suggest guiding principles in the Israeli context that would be taken into account as part of the national plan to turn Israel into an AI leader.

The ethical discourse around AI has been growing recently, given technological advances that raise new questions regarding the risks, responsibilities and social benefits involved in the development and use of AI. Ethical, legal and regulatory aspects affect the freedom of action of various actors, and are related to the need to protect "familiar" social values, including the right to privacy, innovation and fairness. At the same time, we expect the new technological advances to contribute substantially to overall human wellbeing.

Scope and Working Assumptions

1. This document is designated for two target audiences: decision makers in the ethical and legal/regulatory areas; and AI developers required to make decisions with ethical and legal/regulatory implications.
2. The report focuses on technologies and uses that are common today and will be common in the next few years. It therefore focuses on narrow AI and machine learning.[1]
3. Over the long run, including ethical considerations in the development and maintenance stages of AI products will benefit corporations and countries that have done so.
4. Any intervention must be *proportional* and based on a review of all relevant considerations, such as individual rights, commercial considerations, innovation, etc. The Committee also assumes that potential problems must be tackled in advance.
5. The Committee did not discuss ethical and regulatory considerations related to the laws of war, given the dearth of literature on this new issue, and recommends discussing it in the future in dedicated forums.

[1] *Narrow AI* refers to tools capable of solving specific or a narrow range of problems. They may demonstrate humanlike or even superhuman capabilities while performing the task, but have a limited range of capabilities and particularly lack self-awareness and cognition. Broad AI systems, on the other hand, can deal with a wide variety of tasks, similarly to humans.
Machine learning is the leading approach to realizing AI. The machine learns by observing demonstrations of the task, and produces a policy to be followed when required to perform. For the purpose of this document, "deep learning" is a way of realizing machine learning.

What Is New and Special about AI?

1. Radicalization of existing social relations

AI systems tend to radicalize existing social relations. For example, if there is inequality between different social groups, AI systems can reproduce and even exacerbate it. This is true of discrimination, stereotype, rights violations, political extremism, etc. For the sake of convenience, we will demonstrate that claim with regard to inequality. There are several main reasons for that phenomenon:

a. Since AI systems depend on the information provided to them, their input can reflect inequality that already exists, and if the data entered have been manipulated, the system will learn that manipulation.

b. AI systems are becoming increasingly common in a growing number of social contexts. Therefore, their impact – and potential biases – affect larger audiences.

c. There is an erroneous tendency to treat the products of AI systems, which analyze data quickly and on a large scale, as scientific truth. Consequently, there is the danger that such systems would not be subject to the controls applied to equivalent human decisions, when a bias is suspected.

d. Due to the systems' complexity, it is difficult to anticipate and validate their behavior in advance. Consequently, it is often hard to distinguish between "true" diagnosis based on a valid review and monitoring process, as done with regard classical algorithms or human decision-making, and a biased diagnosis.

2. The procedural challenge: How to "engineer" values

AI systems are also active in areas where decision-making and discretion have traditionally been the purview of humans. Often, the professional charged with making the decision in question is also skilled and authorized to apply normative considerations. When developing AI systems that replace human discretion, the responsibility for these normative considerations is transferred from professionals such as doctors and lawyers to engineers and information scientists, which does not occur as often when dealing with classical algorithms.

3. Privacy risks

Information privacy is the person's right to control or have a say on the uses made of his or her information. Often, the violation of the right to privacy can lead to violation of other rights, such as the right to equality. The information age poses new challenges to the effective protection of the right to privacy. These are expressed in the ability to gather information almost unlimitedly, process it to an extent never known before, and produce new insights about an individual that may not be related to the original context in which the information was collected. Another challenge is that these huge amounts of information make it possible to analyze individuals' characteristics in a way that enables others to influence them, often unfairly.

AI systems are based on the processing of big data from various sources, including the integration and cross-referencing of information, sometimes – again – without any relation to the original objective and context for which and in which the information was collected. Finally, AI systems enable to derive conclusions and carry out actions based on information, in a way that produces privacy and autonomy risks of a scope and type that are without precedent.

4. Complexity that erodes public trust

The complexity and ambiguity inherent to AI-integrated products and services make it difficult to promote public trust in the technology and public understanding of how it operates and how it affects our lives. This lack of clarity often leads to distrust, even in areas where AI systems offer a clear business – and social – benefit. For example, in the area of autonomous vehicles, it is estimated that for the public to trust a driverless car, the rate of fatal accidents involving such cars must be 1,000 times lower than the rate involving traditional cars.[2]

Public mistrust of a particular AI application can affect the entire field. Experience has shown that disappointment has led to years-long "hibernation periods", where advances in the AI area have stopped almost completely. The EU's High-Level Expert Group on Artificial Intelligence, for example, argued that if the public is unable to trust AI systems, they will not be adopted, resulting in the loss of the considerable added value they have to offer.[3]

5. Unfair economies of scale

One of the main strengths of AI systems is their ability to constantly improve. For example, a system for recommending contents receives feedback from the users on the quality of the contents recommended, thereby enabling it to make better recommendations in the future. Accordingly, a large company able to collect feedback from a large number of users will improve its system faster than its smaller competitors. Consequently, powerful players with the big data required to develop AI systems take advantage of internet economies of scale to shape the way new players enter the market, with a negative effect on competitiveness. When it comes to completely new players, the fact that they lack the amount of data required could mean they are in effect barred from the AI market.

6. Changes in familiar warranty categories

Many products currently combine the abilities to collect and process information enabled by connecting the physical product to a remote processing capability, following the growing trend known as Internet of Things (IoT). The ability to collect and process data through products enables companies to offer new related services, but also raises new questions about

[2] Shai Shalev-Shwartz, Shaked Shanmmah and Amnon Shatua, 2017, On a Formal Model of Safe and Scalable Self-driving Cars, ArXiv, https://arxiv.org/pdf/1708.06374.pdf.
[3] Ethical Guidelines for Trustworthy AI, The High-Level Expert Group on Artificial Intelligence, EU, 2019, https://ec.europa.eu/futurium/en/ai-alliance-consultation. (Hereafter, EU)

the warranty for these services, and the division of responsibility between the producer and those providing the services in practice. AI-integrated products, in particular, also include the combination of a physical product and remote computability and operability. Thus, the classical division between product and service and product warranty and service warranty needs to be reexamined.

Things become even more complex when such products and services are used by other business entities. For example, when a grocery chain uses a drone for deliveries. The drone is capable of flying, navigating and dealing with the environment. In addition, it provides mapping and weather forecast services. All these are acquired by a grocery chain, for the modest purpose of delivering groceries.

Chapter 2: Ethics and Artificial Intelligence

Given the unique challenges of AI technologies, various groups of experts and companies have recently issued ethical policy documents and proposals. For example, HLEG stated that it would be impossible to promote trust in AI systems through regulatory measures only, and that trustworthy AI can only be achieved by selecting the appropriate ethical framework and implementing it appropriately.[4]

Recall, however, that many countries are currently racing for the lead in the AI area.[5] The challenge of maintaining the competitiveness of democratic countries facing nondemocratic countries with different values is exacerbated in the AI area. These moves are also promoted in various international forums.[6] In Israel, the Innovation Authority points to the importance of infrastructural investment in this area as enabling growth and leadership, and suggests that awareness of the importance of a clear ethical and legal framework would prevent a chilling effect on innovation.[7] Given these global processes:

> *the Committee believes that maintaining a system of ethical principles that would guide both private and public sector organizations is critical for the long term.*

The Committee also calls to develop ethical programs for professionals developing IA systems, so that they are able to raise the ethical problems and address them at the early design R&D stages. Training is not within the scope of this Committee, so we have not provided further details beyond calling for the ethical education of professionals.

*

Given global developments and the challenges posed by AI, the Committee has identified the following ethical principles for the AI area, as part of the instrument we will submit to policymakers for the design and use of the technology. Clearly, ongoing discussions between technological experts, social scientists and legal experts is required in order to update and adjust the ethical framework. Therefore, the recommendations listed below should be treated as reflecting the Committee's understanding of the matter at the time of writing.

---

[4] EU.

[5] See, e.g. Executive Order on Maintaining American Leadership in Artificial Intelligence, February 11, 2019, https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/ (hereafter, US). See also a 2018 French government report, https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf, and a British government report from the same year, https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf. Non-democratic countries such as China, https://flia.org/notice-state-council-issuing-new-generation-artificial-intelligence, and Russia, https://futureoflife.org/ai-policy-russia/ also addressed the issue.

[6] See the 2019 OECD report that recommends that member states develop AI technology in a way that promotes social welfare, https://www.oecd.org/going-digital/ai/principles/.

[7] Innovation Report, 2018-19, https://innovationisrael.org.il/en/report/innovation-report-2018.

Ethical Principles for AI

1. **Fairness:** Striving for substantial equality, prevention of biases (in information, in the process and in the product), prevention of discrimination, and avoidance of widening socioeconomic and educational gaps.
2. **Accountability:**
    a. *Transparency:* Providing information about the process and related decision making.
    b. *Explainability:* Being able to explain the system's decision-making process (on the level of individual users, as well as on a collective level if the system affects group, as well as for the system operators themselves).
    c. *Ethical and legal responsibility* – to be divided among the relevant actors in the value chain, together with risk management. Determining the responsibilities for setting rules for reasonable measures to prevent the risk according to the context and the estimated severity of the risk, for managing the risks and for appointing an employee in charge of risk management.
3. **Protecting human rights:**
    a. *Bodily integrity:* Preventing any harm to life or limb.
    b. *Privacy:* Preventing damage to privacy due to collecting, analyzing and processing information, sharing the information and making new and different uses of the information.
    c. *Autonomy:* Maintaining the individual's ability to make intelligent decisions, including the prevention of unfair or unconscious influence on individual behavior.
    d. *Civil and political rights:* Including the right to elect, freedom of speech and freedom of conscience religion.
4. **Cyber and information security:** Maintaining the systems in working order, protecting the information they use, and preventing misuse by a malicious actor.
5. **Safety:** Preventing danger to individuals and to society and mitigating any damage.
    a. *Internal safety:* In developing the AI tool.
    b. *External safety:* For the environments and clients, in using the tool.
6. **Maintaining a competitive market** and rules of conduct that facilitate competition.

Examples for Ethical Risks in AI Systems

1. **Fairness:**
    - The system decides on allocating resources such as funds and medical treatments.
    - The system evaluates candidates for a workplace or higher education.
    - The system evaluates people for the purpose of criminal punishment or the mitigation thereof.
    - The system makes decisions that threaten users' property and financial interests.
2. **Accountability:**
    - The system does not provide the user with information about how it operates with relation to the user.
    - The division of responsibilities in case of an unresolved malfunction.
3. **Protecting human rights:**
    - The system collects personal information with or without the users' consent.
    - Personal information is used to develop the system.
    - The system produces personal information (e.g. a face recognition system that enables surveillance).
    - The system can affect the users' worldview.
    - The system filters or produces information so that it is customized for the users based on their personal data (targeting), or uniquely tailored for the user without the latter being able to know what information is provided to others.
4. **Cyber and information security:**
    - If malicious actors manage to hack the system they will be able to cause physical, mental or financial damage, or threaten national security.
5. **Safety:**
    - The system can cause death or physical harm to the users or others.
    - A malfunction can cause death or physical harm to the users or others.
6. **Maintaining a competitive market**
    - The system produces an advantage for competitors with big data.
    - The system is based on a large database accessible to only few market players.
    - In the course of its operations, the system produces a large and unique database that is inaccessible to competitors.
    - Non-competition agreements and automatic coordination between companies based on AI systems.

We will now provide further details on these principles, as arising from various policy documents in both the public and private sectors. Some of these principles are detailed in policy documents of technological companies as part of self-imposed limitations. This suggests that in the conflict between values and innovation, these companies are aware of the need for balance, and even propose the desirable balance levels at their own initiative.

1. Fairness

Technology is not neutral, as it is based on human programming and various commercial interests. Moreover, the AI systems are based on information related to human behavior, which may reflect and exacerbate various types of social biases.

According to Microsoft, for example, "AI systems should treat everyone in a fair and balanced manner and not affect similarly situated groups of people in different ways". [8] Unfairness can occur in various stages of development. For example, all populations must be represented in databases, without gender or racial bias. Unfairness can also occur if the public is unaware of the limitations of AI and assumes AI systems always make better decisions. Another aspect of fairness is for AI systems to empower all social groups. Finally, Microsoft recommends diversifying the R&D team itself.[9]

The French government report suggests that AI

> must not become a new way of excluding parts of the population. At a time when these technologies are becoming the keys to opening the world of the future, this is a democratic requirement. AI creates vast opportunities for value creation and the development of our societies and individuals, but these opportunities must benefit everyone across the board.[10]

Finally, IBM suggests that

> Real-time analysis of AI brings to light both intentional and unintentional biases. When bias in data becomes apparent, the team must investigate and understand where it originated and how it can be mitigated. Design and develop without intentional biases and schedule team reviews to avoid unintentional biases. […] Instill a feedback mechanism or open dialogue with users to raise awareness of user-identified biases or issues.[11]

To conclude, in order to minimize biased outputs of AI technology, the target population must be proactively studied, and groups liable to be mis- or underrepresented must be identified in advanced. Accordingly, the Committee recommends consulting with domain experts specializing in the particular social spheres where an AI system is to be integrated.

---

[8] The Future Computed - Artificial Intelligence and Its Role in Society, Microsoft, 2018, p. 58. https://blogs.microsoft.com/uploads/2018/02/The-Future-Computed_2.8.18.pdf. (Hereafter, Microsoft)
[9] Microsoft, 68.
[10] Cedric Villani, For a Meaningful Artificial Intelligence, https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf, p. 16. (Hereafter, France)
[11] Adam Cutler et al., 2018, Everyday Ethics for Artificial Intelligence, IBM. https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf, p. 36. (Hereafter, IBM)

Moreover, consulting with representatives of the target users themselves (through short-term research) can help produce fairer systems. The idea is to gather feedback not on the installation of the system in question, but on their needs and challenges such installation can pose.

2. Accountability

a. Transparency

The Committee considers transparency a key value in technological development and in developing AI products in particular. Transparency is a value in its own right, particularly in a democracy, and an aspect of accountability as well. It is also a tool for assuring other values such as fairness. Finally, it is essential to public trust.

In the EU Expert Group Report, there is no consensus regarding the meaning of transparency. Sometimes, it is treated as part of the development stages, where transparency varies and depends on the target audience. However, sometimes transparency was treated as a simple for explanation and interpretation only. We distinguish between transparency and explainability. Whereas the former refers to access to information, the way the process unfolds and its various elements, explainability refers to explaining the operations themselves and the process leading up to them.

The EU report recommended three reporting levels: (1) the *micro level* of users, researchers and professionals; (2) the *meso level* of public institutes, corporations and universities; and (3) the *macro level* of politicians and the way they report to civilians. The authors also recommended considering transparency for special groups, such as people with disabilities, who may be expected to make frequent use of AI.[12]

According to the EU Report, understanding how the system works (transparency and explainability) is important for cultivating trust. The level of explanation provided should be similar to that received from service providers. The users need to be explained why the systems are safe, and after gaining some understanding of how the technology works, public trust in the product and its developers could be gained. This trust is important to maintain throughout the system lifecycle and not only at the production stage, as the system is expected to change with time.

b. Explainability

The IBM Report argues that the more AI becomes integrated in various systems that are part of our lives, the decision making process of AI systems would have to be explained in a way that is clear to the end-users. Moreover, the users need to know that they have contacted a machine and not a person, and that at any stage, they can ask the system why it works the way it does. The interface for asking those questions and receiving answers should be

[12] EU

accessible to the user at any stage, and the system's processing should be reviewable throughout, particularly when the information processed is sensitive. Whenever the system helps a person make sensitive decisions, it must be able to explain and substantiate its recommendation.[13] Microsoft offers similar recommendations.[14]

The EU Report also discusses the balanced information principle, according to which sufficient but not too detailed information needs to be provided. The fear is that detailed information would be provided in a professional language inaccessible to the ordinary user, thereby achieving the opposite result.[15]

c. Responsibility

Responsibility is the third facet of accountability, referring to the regulation and division of ethical and legal responsibility given various levels of risk or damage. Such regulation will refer, among other things, to circumstances in which responsibility is imposed and on whom, with reference to the stages from planning and development to actual use. The issue of responsibility is complex due to the interactions between various stakeholders, including autonomous systems, developers, users and manufacturers.

The complexity grows particularly when AI takes part in decision making. How is liability to be divided among domain experts and AI systems when the latter replace or assist doctors in the hospital, for example?[16] According to the EU, the system must be designed so as to enable careful monitoring of decisions made.[17]

IBM decided to take a clear stand and hold AI developers responsible for technological design, but also for the decision making process *and* its outcomes. The report also stresses the need to document the development and decision making process, since documentation encourages compliance with legal and ethical rules.[18] According to Microsoft, whoever develops an AI system must be responsible for the way it operates. Appointing an internal review unit would make sure that the issue of responsibility is examined from time to time and that the company receives proper guidance.[19]

The EU recommends a system of reliefs and compensations, and appointing a dedicated official who would receive complaints. Their basic assumption is that trustworthiness depends on the certainty that should someone break the rules, they would be held accountable.[20]

---

[13] IBM, 28-33.
[14] Microsoft, 58-59.
[15] EU
[16] Karni Chagal, 2018, The Reasonable Algorithm, *Journal of Law, Technology and Policy*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3095436.
[17] EU
[18] IBM
[19] Microsoft, 73.
[20] EU

3. Protecting Human Rights

The Committee views the protection of all human rights an obligation of the highest importance, particularly rights that are more likely to be affected in the AI era – privacy, autonomy, and political and civil rights – and in ways we cannot anticipate fully. The February 2019 Executive Order states that "The United States must foster public trust and confidence in AI technologies and protect civil liberties, privacy, and American values in their application in order to fully realize the potential of AI technologies for the American people".[21]

The emphasis in this section is on the importance of the awareness of the potential impact of AI products and maintaining continuous, value- and risk-oriented dialogue between the human rights defenders and product developers. The Committee believes that such a dialogue, in itself, promotes a balanced incorporation of values in decision making, in a way that meets corporations' general accountability requirements.

According to the EU, AI trustworthiness also depends on a constant ethical intention to do good and on the technology being under control, because good intentions can also cause unpredictable damage. Attention to human rights must be devoted in the development and design of AI systems from the earliest stages, including when selecting the development team. Particular attention must be given to situations involving vulnerable populations, such as children, people with disabilities, minorities and any situation where there is a power gap between the developers and users.[22]

In order to ensure it matches the norms and values of its potential users, the AI system – argues IBM – must be developed out of due consideration for them. To do so, it recommends that the development team gain better understanding of the target culture by consulting interdisciplinary academic, business and technological team, as well as software design researchers.[23] Nathan Matias of the MIT Media Lab also suggests pretesting technologies that may impact life or liberty and making test results public.[24]

Google has undertaken not to take part in developing AI applications likely to cause overall harm; weapons or other technologies intended to cause injury; technologies that gather or use information for illegal surveillance; and technologies whose purpose contravenes principles of international law and human rights. Google will only develop AI technologies when their accumulated benefits outweigh the potential risks and disadvantages. Finally, it undertakes to test AI applications, when applicable, in a controlled and restricted environment, and continue monitoring every step in their implementation.[25] Such an environment is provided, for example, in TeraLab, at Institut Mines-Télécom (IMT) in Paris, and the French similarly

[21] US, Sec. 1.
[22] EU
[23] IBM
[24] Nathan Matias, MIT Media Lab, The Obligation to Experiment, 2016, https://medium.com/mit-media-lab/the-obligation-to-experiment-83092256c3e9.
[25] Sundar Pichai, AI at Google: Our Principles, 2018, https://www.blog.google/technology/ai/ai-principles/. (Hereafter, Google).

recommend establishing innovation sandboxes for the same purpose.[26] In Israel, the first report Samuel Neaman Institute for National Policy recommended establishing testbeds.[27]

a. Privacy

Privacy is often discussed in the context of AI applications, since these are largely based on information about individuals or on deriving conclusions about them from personally identifiable information (PII). PII is defined as information about an identified person or information that can be used to identify a person. Legal principles for the protection of privacy are designed to enable the use of PII for legitimate purposes while minimizing the risk to privacy.

In Israel, the right to privacy is protected in Basic Law: Human Dignity and Liberty (1992).[28] The regulation of privacy in the information processing area is led by the new European law widely known as GDPR (General Data Protection Regulation, 2018). Accepted principles of information processing include the need for legal cause for collecting and processing information (usually informed consent), limiting the use of the information to its stated purpose, the right to review and revise the information, transparency vis-à-vis the information owner and the obligation to protect the information. Under more advanced regulatory regimes such as GDPR, the rules are more detailed and include the requirement to design technological systems in a way that protects the right to privacy, the right to transfer personal information to a different service provider, and the right to have information about yourself be erased ("the right to be forgotten").

Privacy protection regimes are currently facing a significant gap between the principled importance of consent to collect and use information and a reality where this agreement is based on standard forms that often do not serve the purpose of agreement. This complexity also affects the AI areas, as it is based on the processing of personal information.

Tech giants refer extensively to privacy in the AI context. For example, Google has undertaken to develop its AI applications while protecting user privacy, including transparency and the user's ability to control the use made of the information.[29] Intel has gone further and suggested that privacy should be rethought in the present age, including reviewing whether principles such as privacy by design are at all relevant in the AI era.[30] Microsoft similarly believes that the AI privacy realm is bound to evoke new questions never discussed before and the need to update laws, and that therefore "One goal should be to ensure that governments work with businesses and other stakeholders to strike the balance that is needed to maximize the potential of AI […] and address new challenges as they arise".[31]

---

[26] France, 9.
[27] Getz et al., 91.
[28] http://knesset.gov.il/laws/special/eng/BasicLawLiberty.pdf
[29] Google.
[30] Intel, 2017, Artificial Intelligence: The Public Policy Opportunity, https://www.intel.ai/ai/wp-content/uploads/sites/69/Intel-AI-Public-Policy-WP-2017.pdf, p. 3.
[31] Microsoft, 75.

Note that privacy and fairness may conflict. For example, if many members of a certain group avoid sharing their personal information with AI systems, the quality of service these systems provide to this group would be compromised.

Finally, with regard to the legal situation in Israel, the question of integrating data with AI is not covered by the existing legislation – an example for a "new challenge" as suggested by Microsoft. Overall, the Israeli Protection of Privacy Law (1981) is outdated and must be updated according to regulatory developments in the West.

b. Autonomy

Autonomy is based not only on an individual's ability to choose among options, but also on the availability of the information allowing cogent choice and assessing its reliability. These issues cannot be taken for granted in the AI era. Moreover, the ability to conduct in-depth analysis of information about a person enabled by AI makes it possible to devise highly intrusive persuasion attempts, again with potential implications that are not fully understood as yet.

Autonomy is also related to the range of human decisions involved in interaction with technology, which technology might narrow. We must therefore always examine whether a given application affects autonomy and how. Note that within this discussion, there may be cases where autonomy is narrower to begin with (due to certain socioeconomic or normative characteristics), or where narrower autonomy is seen as more appropriate normatively, making the special steps to protect freedom of choice may not be necessarily required.

Some AI technologies, such as "deep fake", are designed to produce unreliable information that can hardly be distinguished from reliable one. These technologies have the potential of reducing the ability of individuals to understand reality and make autonomous, informed decisions, and of eroding the trust between people and between them and their government. For example, we are not far from the day when it would be possible to artificially produce a film where a leader declares war, leading to catastrophic results. The Committee believes that the State of Israel should examine ways of dealing with these technologies in a separate report.

One final area relevant to autonomy is the penetration of AI tools into the news media. Many communication channels use AI to produce individually customized news. This tool has many advantages, but also poses the danger of selective exposure: certain groups in the population are exposed to standardized information and are unaware of evidence and arguments that are inconsistent with their worldview. This would deny such a population the freedom of choice or the freedom to be exposed to a diversity of opinions, and make them vulnerable to unfair and highly effective influence campaigns by interested parties. In particular, this could enable foreign governments to intervene in elections.

c. Civil and political rights

Violating rights such as freedom of expression, equality and freedom of conscience and religion could occur when the discourse is manipulated automatically, such as over-echoing certain views and silencing others, radicalizing the discourse and legitimizing opinions that could be offensive for certain groups, as well as influencing persuasion processes. All these can affect the ability to exercise certain civil and political rights, raising the specter of massive dissemination of lies and severe damage to the democratic process.

4. Cyber and Information Security

Cyber and information security are about protecting computers and networks and/or the information stored in or transferred through them against abuse. Accordingly, it is a fundamental requirement for safe and effective development and implementation of AI technology. Information is the energy that fuels the current wave of AI. Therefore, AI professionals collect huge amounts of information both while building the tools and while operating the systems. This can include personal, medical, economic and other sensitive information. Note that non-sensitive information can become sensitive when cross-referenced with other information sources. Sometimes, even the amount of data can make information sensitive. Therefore, all those involved in AI must take care to secure the information in order to avoid leaks and breaches and prevent malicious actors from accessing it.[32] Algorithmic systems that support or replace human decision making are themselves in need of protection in order to prevent the materialization of exploitation scenarios.

5. Safety

AI systems may be integrated as decision makers or decision-making aids in such ways that they can effect safety either directly or indirectly. Thus, for example, there have been several cases where autonomous vehicles have been involved in lethal accidents. The risks and damages are due to several possible factors. First, failing to identify bias or other limitations in the data can lead the system to make erroneous decisions (for example, with regard to recommendations on caring for pneumonia patients).[33] Another challenge is handling extreme cases. In such cases, one needs to develop mechanisms to limit the degree of damage. Moreover, system performance must be monitored to identify points of failures and address them quickly.[34]

Another issue that needs to be considered is safety at the stage of designing the AI application. As mentioned, these applications require large amounts of information, so that care must be taken that the race to data does not involve unnecessary risks. For example, in order to build an autonomous car, data must be collected in diverse driving conditions, some dangerous, so that the system could learn how to respond in such situations. Therefore, the

[32] Microsoft
[33] Microsoft, 64.
[34] Microsoft, 61-65.

development stage could involve placing people in risk situations and even encouraging them to take risks in order to gather information that will help the system.

6. Maintaining a Competitive Market

The Committee considers competition essential for innovation and social welfare. Thus, maintaining a free market with fair competition would allow all actors in the value chain, particularly small-to-medium enterprises and startups to benefit and profit from the activity.

In the AI area, there are players with significant market power in various segments of the value chain. This power derives, among other things, from the characteristics of network-economy, bilateral markets, and broad access to the information used for processing. Concentrations of economic power can also lead to concentrations of political power, allowing tech giants to dictate the rules of the game in the market. The fear is that the influence of these mega-players on the market could make it difficult for new technologies or applications to enter the market, and compromise the innovation so critical for AI. For example, the power of information platforms has been translated to power in the advertising market (Google and Facebook are obvious examples), and subsequently in the retail sector (Amazon). Further increase in that power may affect other economical sectors.

Accordingly, we need to examine whether the exiting competition laws, standardized contracts and consumer protections are prepared to meet the anticipated challenges. To that we must add the international challenge, resulting from the fact that some of the key players are based on the US. According to the EU experts, we need to consider changing the rules of competition in order to enable greater competition in the AI area. To do so, the private sector must be mapped according to its readiness to assimilate AI, and a different response given to each sector according to the different levels of readiness found.[35] A 2016 UK report recommended limiting the power of hi-tech giants using antitrust tools.[36]

In this context, a major challenge in the AI area is the need for access to large databases, owned by a small number of entities, including private ones. Thus, companies that hold more information might prevent smaller competitors from competing and could be strongly biased in their favor in the service they provide.[37] Reports in some countries, such as France, called upon the government to create open databases as a public good and ground their openness in a copyright law.[38] In Israel for example, there is no copyright protection for several works listed in Art. 6 of the 2007 Copyright Law. This recommendation is designed to prevent bias and maintain the system's fairness by training the AI systems on fair and comprehensive databases that represent the entire population. In addition, this call was designed to allow open and fair competition between small companies with small databases and large companies with large databases on which AI systems can be trained.

---

[35] EU
[36] House of Commons Science and Technology Committee, Robotics and Artificial Intelligence, 2016, https://publications.parliament.uk/pa/cm201617/cmselect/cmsctech/145/145.pdf
[37] Getz et al., 90
[38] France

Another solution is to prepare open databases and making them publicly accessible through AI technology. This was done, for example, in a study where AI created faces.[39] This way developers could also ensure that the database represents the entire population while protecting individual privacy and using the computer-generated contents. Microsoft also called for opening databases, particularly public ones, but warned that protecting privacy is an issue and suggested investing in research on "de-identification" techniques.[40]

Decision-Maker Instrument for Assessing Ethical Challenges

The Committee recommends viewing all those involved in AI as responsible for acting legally and ethically. They are also expected to keep up-to-date with risks and ways of dealing with them, and may be considered negligent if they fail to do so.

To enable AI professionals to identify and respond properly to ethical risks, we have developed a dedicated instrument. Note that given the dynamic nature of the area, that instrument also needs to be kept up to date. The instrument is comprised of two parts:

1. A set of preliminary questions that should be asked when developing an AI product, designed to assess its influence. These questions are addressed to product developers throughout the development and production change.[41]
2. Use and maintenance of a frequency map, that helps locate the challenging areas in term of applying ethical values to the system's development. The frequency map presented in the report (see below) assesses ten representative case study for points of ethical conflict according to the passage of time and the development stages. Through the test cases, the map raises awareness of areas where other organizations had trouble in the past, and areas for particular attention by decision makers. To remain relevant, the map needs to be updated with new test cases on an ongoing basis.

In the following pages, we describe how such an instrument can be designed for an IA organization.

[39] Tero Karras, Samuli Laine and Aila Timo, 2018, "A Style-Based Generator Architecture for Generative Adversarial Networks", *Neural and Evolutionary Computing*, https://arxiv.org/abs/1812.04948.
[40] Microsoft, 78
[41] The Canadian government developed an algorithmic impact assessment designed to evaluate and mitigate risks when developing and implementing AI systems: https://www.canada.ca/en/government/system/digital-government/modern-emerging-technologies/responsible-use-ai/algorithmic-impact-assessment.html

2. Preliminary questions

(1) What is the level of potential individual harm? The more the potential individual harm increases, the more values such as fairness, human rights, accountability, autonomy and information security need to be attended to. If the company has a large market share, even a small degree of harm could affect a large number of customers. In the privacy area, for example, various conventions currently refer to the importance of consent. When consent for collecting and processing the information has not been given, greater care must be shown not only at the stage of information collection – currently regulated in various countries – but also in the stage of using the information to train AI systems. The same goes for personal information. This issue is also mostly regulated today, but if personal information is used, greater sensitivity must be shown in collecting the information, in training the system, as well as in sharing the information with others.

(2) What is the extent of potential perceptual impact? Although this issue is closely related to (1), the Committee believes it requires separate attention. If the system can affect human perception greater sensitivity must be shown, particularly in cases of potential misuse of the AI system to manipulate minds (e.g. deep fake).

(3) What is the degree of potential damage to the public? Here, too, the number of customers or the market power of the company developing the AI technology affect the damage potential. Also of interest is the question whether a *particular* social group is liable to be damaged, in which case greater sensitivity is in order.

(4) Is there any impact on the allocation of public resources? This issue is closely related to (3), but requires separate attention according to the Committee. Can the system affect the allocation of financial or other resources? If it can, greater sensitivity is required.

(5) Is the development team diverse enough? In particular, does it include representatives of groups liable to be adversely affected? If not, greater care must be given to the normative assessment of the system, including examining whether the audiences at risk may be identified.

(6) What is the expected extent of damage due to misuse of or loss of control over the product? The greater the individual, public or national economic damage, greater care must be given to the ability to deactivate the system when it goes out of control or when it is being abused.

(7) Is there a fast way to identify unpredicted ethical failures? If significant damage could accrue until the failure is identified, a particularly high testing threshold should be applied before the system is launched for public use.

3. Frequency map

The frequency map indicates the frequency of ethical issues along the product's development chain. It pinpoints areas where failures have been found in the past and provides information about their rate of incidence. As the frequency can change with time and new events found, we recommend updating the map on a regular basis, as also demonstrated below.

In order to create the frequency map, we used ten test cases selected out of real-life past cases that represent various challenges. The map illustrates all the ethical principles listed under "Ethical Principles for AI" on p.8 above.

Table 1: Prototypical Test Cases of Ethical Challenges

1	**AI system for screening workplace candidates** Companies are contacted by multiple candidates wishing to work for them. In order to select the best candidates, several companies have developed AI-based tools trained based on past decisions by the companies. When one such system developed by Amazon was tested, it was found to discriminate against women candidates for technical job. It is assumed that in the past company executives used to discriminate this way, and the system learned to emulate this behavior[42].
2	**Using AI for political influence** Cambridge Analytica collected personal data of millions of Facebook profiles without the users' agreement or knowledge, and used them to influence the users for political purposes. There was probably use of AI technology to manipulate minds. This activity went on for several years[43].
3	**Predicting disease risk** During the 1990s, several research centers joined hands to develop a system that would estimate the degree to which pneumonia represents a life risk for specific patients. This was designed to help doctors decide which patients to hospitalize and which can be treated in the community. Shortly before the system's launch, it was found that its recommendations for asthmatics could risk their lives, because the information used to build the system was biased: asthmatics with pneumonia had received preliminary intensive care that saved their lives, and the system deduced that pneumonia was not risky for asthmatics.[44]

[42] https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine
[43] https://www.theguardian.com/us-news/2015/dec/11/senator-ted-cruz-president-campaign-facebook-user-data, https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election
[44] http://people.dbmi.columbia.edu/noemie/papers/15kdd.pdf

4	**System for assessing detainee dangerousness** When deciding whether to remand a detainee, one of the considerations is the danger he poses to others. The decision is based on multiple parameters, such as criminal history. Several US districts have adopted an AI system called Compas to help judges assess suspects' dangerousness. The system was tested and was found to assess white detainees as less dangerous than black ones.[45]
5	**Virtual AI-guided players accumulate tie-breaking weapons** In a game called Elite Dangerous, human players compete against AI-guided players. To make the game more interested, restrictions on the virtual players were changed in Version 2.1, to enable them to fly and fight better. The AI mechanisms found a way of taking advantage of those changes to accumulate weapons in a way that prevented human users from being able to match them.[46]
6	**The racist bot** Microsoft launched a bot in order to teach it to correspond freely with Twitter users. The idea was that the bot would engage in conversation and learn to improve its dialogue skills in the process. Less than 24 hours after the launch, it was found that since it emulated the users, several users chose to turn it into a racist bot by using racist comments themselves. [47]
7	**The impersonator bot** Google Duplex enables a bot to hold a conversation in a manner that made it difficult for its interlocutors to determine whether it was human. Building this tool required access to huge amounts of data available to only very few knowledge-intensive companies. [48]

[45] https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
[46] https://futurism.com/this-video-games-artificial-intelligence-turned-on-players-using-super-weapons
[47] https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist
[48] https://www.androidcentral.com/google-duplex-will-let-people-know-its-not-human

8	**Autonomous car runs over pedestrian** A pedestrian that crossed the street in a dark area was killed in Arizona by an Uber autonomous vehicle. Apparently, the vehicle identified an "obstacle" and could have avoid crashing into it. Nevertheless, since the engineers had previously lowered the software's sensitivity to barriers, the vehicle did not stop and the woman was killed. The human driver in the vehicle was not alert enough to prevent the accident.[49]
9	**Face recognition bias** Amazon developed a tool for engineers enabling them to add face recognition capability to the system they were developing. The system was designed, among other things, to be used by law enforcement, border police, etc. A test revealed that the system erred much more frequently when activated on people with a dark skin than on people with a light skin.[50]
10	**Content recommendation systems show different information to different groups** Various companies use AI to offer more personally relevant information for users. It was found, however, that Google's ad system presents ads seeking information related to criminal acts when a user searches for information under a name more common in minority populations. [51]

[49] https://www.cbc.ca/news/business/uber-arizona-crash-1.4594939
[50] https://www.theverge.com/2019/1/25/18197137/amazon-rekognition-facial-recognition-bias-race-gender
[51] https://www.bostonglobe.com/business/2013/02/06/harvard-professor-spots-web-search-bias/PtOgSh1ivTZMfyEGj00X4I/story.html

Ethical milestones along the development process

Below, we present examples for ethical issues arising during the development process and follow up on them as they unfold, in order to identify particularly sensitive development milestones. To do so, we present a typical AI development process.

1. Product definition
    a. Understanding the business need or problem the system is trying to solve and creating the R&D organization
    b. Data collection – identifying information sources from within and outside the organization to be used for building the system and assessing its performance
2. Product training
    a. Processing and filtering the raw data into a form that would enable the AI algorithms to receive the data and perform calculations with them
    b. Modelling – applying an AI algorithm to the information processing in an attempt to identify generalizable patterns
3. Integration
    a. Evaluating the model for accuracy
    b. Connecting the AI components with the rest of the system and distributing it for wide use
4. Market management
    a. Performance monitoring to make sure the system works as expected
    b. Ecosystem – together with the process within the organization, there is need to also address the ethical considerations arising out of the fact that the process takes place in the Israeli ecosystem. Integrating AI could affect the socioeconomic, regulatory and other systems, and this should be continuously monitored after launch.

Creating the frequency map

Review the list of test cases and the implications and reported events considering the list of ethical values on p.8 above. Fill in the table according to the emerging ethical challenges. The numbers within the table cells refer to the event number. Next, check the accumulated number of events. Cells with low, medium and high event frequencies are colored beige, yellow, and red, respectively. Note that this table does not indicate the degree and scope of the potential harm. A more sophisticated tool can take these factors also into account. The decision regarding what constitutes low or high frequency should be taken when selecting the number of events the organization refers to. In Table 2, we have ten events, and the frequencies have been determined accordingly.

The Committee recommends that decision makers discuss and offer solutions for emerging challenges according to the frequency map throughout their development process. Since the map depends on a list of test cases, each organization needs to choose a set of test cases relevant to the product under development, assuming that this set changes in time.

Table 2: Frequency Map of Ethical Challenges in the AI Development Process

	Business need	Data collection	Data organization	Modelling	Model evaluation	Distribution	Performance monitoring	Ecosystem
Fairness		1,3,4		3,4	1,3,4	1,4,9	1,3,4,9	1,4,9
Transparency	4			3,4				4
Explainability	4			3,4				9
Accountability	1,2,3,4				1,3,4	3,4	3,4,5,6	2,5,6,9
Privacy	2,9	1	1,2	1	1			2,9
Freedom of choice	7,10						10	6,7,10
Infosecurity			2					2,9
Human rights	4,9			4		4	4	4,9
Safety	3,4	3,4		3,4	3,4	3,4,5,8	3,4,5	3,5,8
Free market	5					5,6	5	

Legend

1	Job candidate screening
2	Political influence
3	Predicting disease risk
4	Assessing detainee dangerousness
5	AI-guided players gain tie-breaking weapons
6	Racist bot
7	Impersonator bot
8	Autonomous vehicle runs over pedestrian
9	Face recognition bias
10	Content recommendation systems present different information to different groups

Low frequency of problematic cases (single case)
Medium frequency of problematic cases (two cases)
High frequency of problematic cases (three cases or more)

Chapter 3: Regulation

This chapter addresses the unique characteristics of AI technology, its development methods and various uses in the regulatory context. Its basic assumption is that balanced intervention is required: not only retrospective intervention, but in any case one that does not exceed certain minimal ethical and regulatory requirements, in order to enable Israeli innovation and continued leadership in the area.

AI Regulation: Private Activities

Regulation is designed to lead to value internalization and behavioral change among members of society. The AI area encompasses a variety of technology and economical uses, with the private sector leading much of the advances.[52] This section focuses on the regulation of private AI activities, with relation to the ethical principles and values identified above.[53] These activities are examined in several spheres: the general sphere of societal influence; the sphere of AI developers and users; and the national sphere of Israel as a producer and user of AI technologies.

With regard to the regulation of innovative information technologies, several characteristics of those technologies – also applicable to AI – must be taken into account.

1. The development of AI technologies is highly dynamic in terms of adoption by society. The rapid development and adoption often prevent regulators from analyzing and the technology's impacts and understanding them in depth for the purpose of imposing a normative framework. Finally, a given technological phenomenon often transforms ethical and societal views in a way that affects the normative framework itself, adding to the difficulty in providing a timely regulatory response.
2. The AI area is global which crosses borders, affected by international technological developments. The global economy also contributes to that characteristic by enabling the movement of information, capital and services across borders, in a way that challenges national regulatory regimes.
3. The rise of corporate giants with multinational market power, platforms of information transfer, commerce and mediation in services and their related data. These information and capital giants have tremendous local and global impact on the technological, business and even political environment.

The various branches of the legal discipline may affect the AI area both directly and indirectly, both legally and non-legally (ethically). Some regulators' determinations may directly affect the development of AI, as in decisions regarding the implementation of privacy

[52] Dr. Roy Schöndorf, Formulating an Israeli Artificial Intelligence Strategy – International Legal Aspects. Ministry of Justice Memorandum prepared for the committee. (Hebrew) (hereafter, Israeli AI Strategy)

[53] This section focuses on the regulation of administrative authority over private activities. With regard to activities by those subject to administrative law (e.g. government development and usage of AI), administrative law itself is the applicable regulatory law. Ways of promoting the internalization and enforcement of substantive principles by entities subject to administrative law are legally, institutionally and procedurally distinct from what the present document suggests, requiring a separate discussion.

protections in AI systems. On the other hand, some regulators' determinations may affect the implementation environment of AI, such as decisions regarding places where autonomous vehicles would be allowed or not allowed. The various branches of the legal discipline can also affect the incentives of the various actors for creating self-regulatory mechanisms, by legislation that would hold them liable or exempt them from liability. Thus, we can even thing of financial incentives for implementing ethics throughout a product's development and maintenance stages, and across the entire value chain.

These insights are based on recent experience in the interrelations between the law and the internet. For example, Section 230 of the 1996 US Communication Decency Act, that exempts intermediaries from liability for information passed through them, has had a remarkable effect on the rise and success of new platform. Conversely, this act has enabled such platforms to determine the "rules of the game" involved in the transfer of information through them by themselves, usually contractually.

In the specific AI context, the tendency of the EU has been to adjust liability rules of private law to the reality of AI-based products and services, based on the "strict liability" model", which represents an active legal intervention where the manufacturer or user (according to the context) are held liable, unless they can prove that certain conditions have prevailed that relieve or exempt them from liability.[54]

Following that approach, we propose a regulatory model that also combines ethical aspects, i.e., that holds companies liable as long as they do not meet industry standards that the regulator would have to determine. Such a model combines the desire to impose liability for preventing damage on the effective cause with the need to avoid chilling out innovation. We can also think of a "carrot and stick" incentive similar to those in the environmental area – reduced punishment in return for adopting an ethical code, and vice versa.[55]

To conclude, the Committee has identified several legal and extra-legal regulatory approaches for addressing the challenges posed by AI (see Table 3):

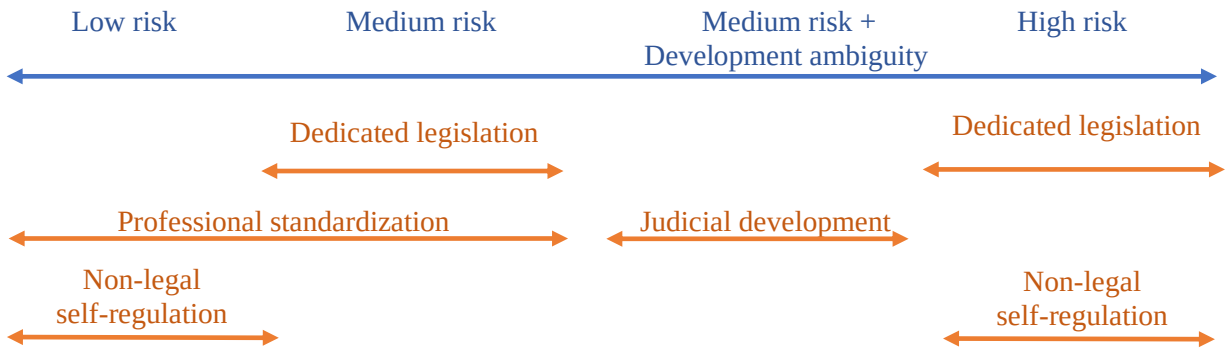
1. Regulation by way of *a dedicated law*, such as legislation designed to protect computers and their information, or a warranty law for damaged products.
2. Regulation by way of *judicial development*, such as of tort or contract laws.
3. Regulation by way of *professional standardization* by institutes that include government, industry, academic and civil society representatives.
4. *Non-legal self-regulation* by ethical rules or professional standards usually developed by the relevant professional community. For example, establishing internal, public advisory and review boards, that include government, industry, academic, and civil society representatives that can make precedential decisions in the development and implementation processes.

[54] European Commission, 2018, Liability for Emerging Digital Technologies, SWD (2018) 137 final. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51633
[55] See, e.g., US Federal Sentencing Guidelines for Organizations, https://www.ussc.gov/guidelines/organizational-guidelines.

Table 3: AI Regulation Options

Type of Regulation	Characteristics	Strengths	Weaknesses	Committee Recommendations
Dedicated legislation	Dedicated law or amendment enforced by a state authority or private entities	• Increased clarity about protected values • Allows concrete judicial development based on legislator guidelines • Partial flexibility	• Lack of professional expertise in a single organization • Retroactive enforcement only • Potential for increased uncertainty • Lack of involvement in present power relations that may privilege certain players	Suitable mainly for medium & high risk areas
Judicial development	No specific law	• No direct regulatory or legal friction • Flexibility • Enables judicial development	• Usually applicable to more obvious cases of harm, and may therefore fail to meet the entire range of harm risks • Lack of professional expertise in a single organization • Uncertainty • Advantage for strong players	Suitable for medium risk situations with development ambiguity
Professional standardization	Allows future adoption by the legal system	• Flexibility • High legitimacy in the professional community • Participatory process	• Risks excluding the law and its values • Dependency on the law for binding validity, oversight & enforcement • Advantage for strong players	Suitable for medium and low risk situations + as a framework for developing & reviewing the application of ethical values
Non-legal regulation	No legal norm (e.g. applying ethical principles)	• Flexibility • High legitimacy in the professional community	• Risks excluding the law and its basic values (equality, fairness, human rights) • Dependency on the professional community for development • Lack of reliable enforcement mechanism • Advantage for strong players	Suitable for low risk situations, where non-legal regulation is sufficient, and for high risk situations, where technological development is relatively rapid for the legal channel

Figure 1: Options for AI Regulation and Their Suitability for Different Risk Levels



Each of the four options has strengths and weaknesses in the balancing of risk management, certainty, side effects, flexibility and innovation. Figure 1 below depicts what the Committee considers the *optimal* matching of regulatory options according to the risk areas of AI products or services. In addition, Table 3 provides information on the advantages and disadvantages of the various types of regulation in order to enable decision makers to hold a productive discussion about those options in government and regulatory bodies. Note that often, regulation is implemented using a combination of approaches. For example, in the privacy area, there is a Privacy Protection Authority, next to laws, bylaws, judicial development and standardization.

The table focuses on the various options in light of the question, what is regulated? The institutional aspect must also be taken into account, however. Regulation by dedicated legislation may be promoted by a dedicated authority established by law that specializes in AI, or without one. The advantage of such an authority would be its high specialization and ability to serve as a hub for rapid response to communications and various challenges. On the other hand, such an authority could cause over-regulation and chill out innovation.

*

In Chapter 2, the Committee proposed an ethical decision-making tools, and presented a series of questions for decision makers, as well as parameters the Committee believes must be taken into account (see p.19). Beyond the ethical aspects that decision makers need to consider, these parameters affect the choice of regulatory approach. Together with these parameters, that justify a more interventionist regulatory approach, the Committee referred to the importance of raising public awareness of the positive and challenging aspects of AI and ethical education, the discussion of which is beyond the Committee's scope.

Regulatory Guidelines Proposed by the Committee

1. Alignment with international legislation and standardization, and promoting Israeli policy in global arenas

Every proposed model has to meet this requirement to enable Israel to remain in the forefront of AI development and make sure the Israeli regulator does not obstruct the activities of local development and industrial centers relative to other developed countries. Note that it is necessary to refer not only to the policies adopted by other countries, but also to regional and international decisions, including by the OECD and the World Economic Forum (WEF).[56] In following this guideline, the government is also required to proactively promote the interests and values important for Israeli policymakers in the international arenas where this discourse is held, in order to have a voice with regard to arrangements that have not yet been finalized.

Moreover, if the State of Israel wishes to position itself as a global opinion leader, the Committee recommends that it formulate a clear position and present it in important international forums. A memorandum written for the community by Dr. Roy Schöndorf, Deputy Attorney General (International Law) mapped the important international arenas involved.[57]

2. Mapping the actors to create a adapted responsibility and incentive framework

Promoting and implementing a regulatory policy also requires adaptability to the various incentives available to the players active in the regulated field. Accordingly, the Committee recommends mapping the various players along the value chain and the product and service value chain and their incentives for internalizing the ethical and substantive principles or guidelines described in the previous section. This mapping needs to also include the entities involved in basic, practical and industrial research, those responsible for implementation, etc.

Such mapping will enable determining the responsibility of each player for implementing the guidelines. The Committee suggests that the responsibility be determined based on the relative role in the chain and out of understanding of the implications of holding an individual or entity responsible versus the ability to use AI technology in a socially efficient manner. Note that promoting incentives for the various players can be done under each of the regulatory models summarized in Table 3 above, that is, by way of imposing responsibility, clarifying responsibility, imposing an insurance obligation or advance regulation. Finally, it seems advisable to examine whether civil law provides an up-to-date response for the types of risks and activities entailed in AI.

[56] For example, following Government Resolution 4481, the Israeli Innovation Authority collaborates with the WEF and participates in an international network of regulating innovative technologies, including AI.
[57] Israeli AI Strategy.

3. Adjusting the accountability principle to the dynamism of the AI area

Modern regulation of technology (such as privacy regulation) faces the challenge of adjusting a normative framework to a certain technological and economic context. Accordingly, a key element in privacy laws is the principle of accountability – the management's responsibility to concretely review the risk to the protected value arising from the proposed activity, and adjust a risk management framework to it. Applying this concept in the AI area requires dealing with the technology's unique characteristics, including the difficulty of anticipating precisely how it would operate under certain circumstances, and explain it precisely.

In order to apply the accountability principle in AI contexts, a regulatory principle needs to be added, that requires any organization implementing AI technologies to implement a testing environment and control perimeters prior to implementing the technology. Obviously, the level and depth of the testing required would be related to the substantive principles described above and to the anticipated effects of the technology in question.[58]

4. Promoting normative clarity in critical stages of the AI product value chain

The first stages of developing an AI system – understanding the business need, collecting and organizing the data, building and evaluating the model, distribution and performance monitoring – can be significant for promoting the values and implementing the principles. Moreover, the Committee believes that promoting regulation that can support these stages could contribute significantly to implementing those values and principles. We have seen that there is shortage in databases for open use, in order to enable small companies to succeed in the AI area as well.

5. Constant review of the regulatory policy by the regulator

Given the unique characteristics of the AI area, we may assume that reviewing and enforcing a regulatory policy in the context of certain technological conditions will require more frequent updates than in ordinary regulatory review cycles. In addition, we propose that the regulatory authorities apply a procedural principle according to which controlled testing of the regulatory policy applied in a certain context be made possible, among other things with particular attention to its innovativeness or the risks involved in implementing it. Finally, this principle addresses the challenge of balancing the desire for innovation and the uncertainty with regard to its ethical impact.

6. Regulatory sandbox

The idea of creating a controlled testing environment for more innovative application with a high potential risk has been gaining traction recently. As part of the attempt to calibrate regulatory policies so as to ensure they are socially effective without affecting innovation,

[58] See US for the need to promote technical standards for the testing and safe deployment of technologies.

one of the ideas was to create a regulatory "sandbox", allowing those operating within it to test new activities that could violate the law, under controlled conditions. This idea is particularly valuable in the AI area because of the need to allow innovation on the one hand and address unpredictable risks to social interests on the other. Given the Israeli government policy on this issue and similar recommendations currently being developed in the OECD, the Committee recommends to develop appropriate regulation that would allow for such an environment.

7. The interface between the proposed principles and existing regulations

Chapters 1 and 2 discussed the unique aspects of AI systems. Since the technology is general, its applications and related risks are context-dependent. Many areas are already regulated. It is therefore important to understand the considerations, values, interests and potential social benefits of the regulation in those areas before deciding on new or adjusted regulation. These parameters also affect the desirable normative outcome and the way to realize it.

Given the characteristics of AI systems, and after presenting the critical ethical milestones in system development, the Committee recommends to begin by mapping the authorities regulating the issue of information resources used for AI technology processing, as well as the regulatory authorities responsible for the market of the product in question. This will help remove barriers to the use of AI and for adjusting the regulatory framework to the risks inherent in the system.

In particular, every authority is responsible for examining the implications of AI systems for the area it regulates. For example, the Ministry of Health is responsible for examining the health area and integrating AI within it. It would be appropriate for the Cyber Directorate to examine the implications of integrating AI systems both in defending Israel and given efforts to integrate AI systems in other countries.

8. Existing general institutes – Privacy Protection Authority

As described in the previous chapter, information privacy protection laws regulate the use of personal information and protect individual autonomy. Accordingly, the Privacy Protection Authority has a broad, general and infrastructural role in regulating uses of personal information in AI contexts. Accordingly,

a. The Committee recommends that with relation to AI applications involving the processing of personal information, or having implications for decision making that is based on personal information regulated by the privacy laws, the leading regulatory authority would be the Privacy Protection Authority, and that it will act in synch with any dedicated regulatory authorities yet to be established.

b. Accordingly, the Committee recommends that the Privacy Protection Authority examine the guidelines presented above and form a plan for implementing regulation in the personal information area in the AI context.

c. It is important to make sure that the Privacy Protection Authority has enough technological and material resources to develop an up-to-date technological and legal framework for the area of information anonymization. The ability to anonymize personal data, at a reasonable confidence level, is fundamental to the development and promotion of AI.[59]

9. Existing general institutes – Competition Authority

AI technology is described in certain contexts as infrastructural technology that affects or dictates the way additional technologies that rely on it are developed and used. Note, however, that part of the development and implementation of those technologies in contexts where the information resources required to develop or implement the technology are in the hand of powerful market players could result in unfair competition, and the maximization of the value of this technology for the powerful players. This is particularly true of information platform whose market power derives from their status in web economy in bilateral markets. Recall that the Committee has defined the maintenance of a competitive market as one of the ethical values in this report. Accordingly, the Committee recommends that the Competition Authority (formerly, Antitrust Authority) formulate regulations designed to maintain fair competition in the AI area; protect the consumer and ensure the accessibility of technology; and prevent technological risks and costs from being rolled over to weaker players at the bottom of the value chain, in a way that is socially inefficient. Note that to the best of the Committee's knowledge, the Competition Authority is currently looking into the need to create regulatory intervention tools in the information technology area.[60]

10. The need for interministerial coordination

Since AI technologies are expected to have significant effects on various regulatory aspects, the Committee recommends establishing an interministerial coordination mechanism to create a uniform and coherent policy shared by all government ministries. Since every decision in the AI area has immediate horizontal effects, all relevant influences of every decision on the various regulators and Israeli society as a whole must be examined carefully.

In addition to the overall and guiding coordination, we expect an occasional need for coordinating various interministerial projects, as in smart transportation, which involves the Ministries of Transportation, the Interior, Public Security, etc. Accordingly, we propose establishing an effective government coordination mechanism for sharing professional and regulatory information among the various entities, thereby enabling them to make better-informed decisions.

---

[59] Jules Polonetsky, Omer Tene and Evan Selinger, 2018, Consumer Privacy and the Future of Society. In: *The Cambridge Handbook of Consumer Privacy*, eds. Evan Selinger, Jules Polonetsky and Omer Tene, https://ssrn.com/abstract=3158885.

[60] See the Competition Authority's call for innovation: https://www.gov.il/en/departments/competition.

Finally, the Committee recommends creating a government coordination mechanism and knowledge hub to help the government and regulatory authorities promote the issue coherently. In the Committee discussions, some members feared the outcomes of lack of coordination or conflicts between various entities, due to the proposal to rely on sectorial regulatory authorities. The Committee has not discussed the characteristics of the proposed mechanism, but recommends creating such a function as part of the regulatory reform.

11. Authorities responsible for information resources

Authorities that are responsibile on substantive information resources used for AI technologies have a key role in examining whether the regulatory framework they apply is suitable for achieving societal benefit in this field, while maintaining a fair and free competitive market and protecting human rights. Consideration must be given in this regard not only to risks but also to innovation spaces and promoting societal interests.

The Committee therefore recommends that authorities responsible for areas of activity affected by the products of information processing will be required to undergo evaluation in light of the principles detailed above. Specifically, the authorities need to examine whether, when deploying AI technologies or using them in the activity areas regulated by them there is need for adjusting the applicable framework in order to promote the protection of the regulated interests.