Harnessing Innovation:
Israeli Perspectives on AI Ethics and Governance

Report for CAHAI

Authors:

Prof. Isaac Ben-Israel

Prof. Eviatar Matania

Leehe Friedman

October, 2020

Executive Summary

This article sets forth the current state of play in Israel's policy development, with respect to the opportunities and challenges presented by artificial intelligence (AI) in relation to human rights and ethics. It is based, to a large extent, on the report of Israel's National Initiative for Secured Intelligent Systems, which has been recently submitted to the Israeli government. The present survey describes Israel's unique approach in attempting to leverage opportunities presented by AI while addressing the challenges that it poses. This article outlines how Israel's governance approach thus far seeks to balance the need to enable innovation, both in the public and private sectors, with moral and human rights imperatives which are omnipresent in AI developments.

Israeli policy-makers tend to view AI developments not just as a disruptive but as a transformative: AI technology is seen as critical to the welfare, economy and security of Israel's citizens. Taking this as the starting point, the priority for Israel has been to establish a holistic and sustainable seccured AI ecosystem, driven by the private sector but in which government, private industry and academia all participate, and which supports the use of AI at all levels. Bearing this in mind, this article highlights the key challenges that have been identified by policy makers, in Israel and abroad, in the fields of human rights, democracy and the rule of law – security, privacy, autonomy, civil and political rights, safety, fairness – including fair competition – and accountability. Israel's proposed approach in response to these challenges, according to the National Initiative's Report, breaks new ground. While it is firmly anchored in established governance principles and international AI policy best practices, it nonetheless represents a novel governance approach, focusing on balanced regulation to foster innovation. To that effect, it proposes original policy tools, such as risk assessment tool that match different regulatory approaches based on the risk level associated with a particular activity, and a dynamic frequency map that helps locate challenging areas in term of applying ethical values to the a particular AI system’s development.

It signals also the need for engagement with countries and international forums, to learn from and contribute to international processes involving questions of AI, ethics, law and governance.

This article is not intended as an official government paper, and does not necessarily reflect Israeli government policy. Its authors are writing in their personal capacity, though they received relevant information from various government officials.

Acknowledgments

The authors would like to thank Cedric Sabbah, Director of International Cybersecurity & IT Law at the Office of the Deputy Attorney General (International Law), within Israel's Ministry of Jujstice, Prof. Karin Nahon, Head of the Ethics and Regulation working group of the National Initiative for Secured Intelligent Systems, Dr. Roy Schöndorf, Deputy Attorney General (International Law), and Dror Ben Moshe, Big Data Department Manager at the Ministry of Health, for their valuable contribution, professional guidance and insightful comments.

About the Authors

Prof. Isaac Ben-Israel

Professor Isaac Ben-Israel is a retired Major General who joined Tel-Aviv University in 2002, where currently he is the Director of the Blavatnik Interdisciplinary Cyber Research Centre and the Yuval Ne'eman's Workshop for Science, Technology & Security.

Outside the university he is the Chairman of Israel Space Agency.

In 2010, he was appointed by the PM to lead a task force that led the Cyber Revolution in Israel.

During the last two years Prof. Ben-Israel has been co-chairing Israel's National Initiative for Secured Intelligent Systems (AI) to recommend the PM and the government about a national plan to promote Israel as a global power in AI.

Prof. Eviatar Matania

Eviatar Matania is a professor at the School of Political Sciences, Government and International Affairs at Tel-Aviv University, where he heads the MA program of Security Studies and the MA program of cyber politics and government.

Matania is also an Adjunct Professor at Oxford's Blavatnik School of Government, where he convenes the Cyber Module.

Matania was the founding head and former Director General of the Israel National Cyber Directorate (INCD) in the Israeli Prime Minister office, where he reported directly to the Prime Minister, and was responsible for Israel's overall cyber strategy, policy and its implementation to defend the Israeli civilian sector.

During the last two years Prof. Matania has been co-chairing Israel's National Initiative for Secured Intelligent Systems (AI) to recommend the PM and the government about a national plan to promote Israel as a global power in AI.

Leehe Friedman

Leehe Friedman is Adjunct Professor and the Director of the Honors Track in Strategy & Decision Making at the Lauder School of Government, Diplomacy and Strategy at IDC Herzliya.

During the last two years Ms. Friedman has been the Coordinator of Israel's National Initiative for Secured Intelligent Systems (AI) to recommend the PM and the government about a national plan to promote Israel as a global power in AI.

Contents

Introduction

Israel perceives AI as a core emerging technology and as "*an infrastructure of infrastructures, one that is critical to the future of the State of Israel – to its security, its economy and to the welfare of its population"*.[1] AI applications, due to their potential to enhance availability, reliability and efficiency of national infrastructures, services and systems, at lower costs to the state and its citizens, hold key roles in Israel's capacity to meet some of its national challenges in the 21st century.

Looking ahead, AI is likely to fundamentally transform all aspects of private and public life. In order to harness the positive potential of AI technologies, Israel strives to establish a holistic and sustainable AI ecosystem that includes the government, private industry and academia. A feedback loop involving these three sectors would benefit society as a whole by: (a) increasing the use of AI applications; (b) enhancing the work of the government and the services it provides; (c) fostering the economy and innovation of new techno-scientific developments which in turn would increase again the demand for new AI applications.

Accordingly, the Israeli approach towards AI is based on two complementary efforts:

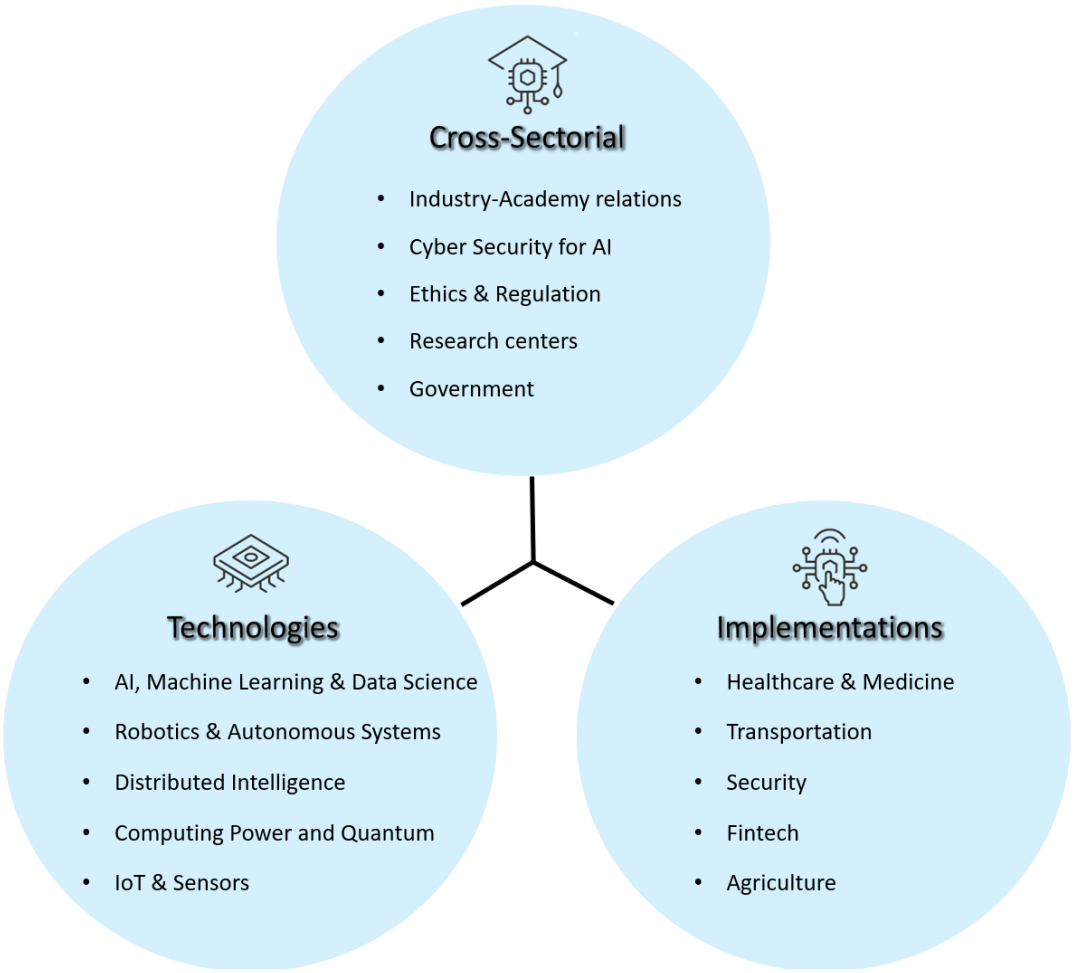
I. Promoting a wide and fair use of AI applications both in the public and private sectors.
II. Fostering a leading technological industry that would develop AI-based solutions for emerging challenges in Israel and around the globe.

Israel's National Initiative for Secured Intelligent Systems

In order to understand Israel's current AI landscape and examine how its broad policy goals can be achieved given the characteristics of Israel society, Israel's Prime Minister launched in 2018 the National Initiative for Secured Intelligent Systems, and appointed two of the authors of this article, Prof. Isaac Ben-Israel and Prof. Eviatar Matania, to co-lead the initiative. Its mandate was to generate a national plan in the field of AI and related intelligent technologies. The work initiative used a multistakeholder approach: hundreds

[1] Ben-Israel, I., Matania, E. & Friedman, L. (Eds.) (Sep. 2020). *The National Initiative for Secured Intelligent Systems to Empower the National Security and Techno-Scientific Resilience: A National Strategy for Israel. Special Report to the Prime Minister*. (Hebrew) p. 3.

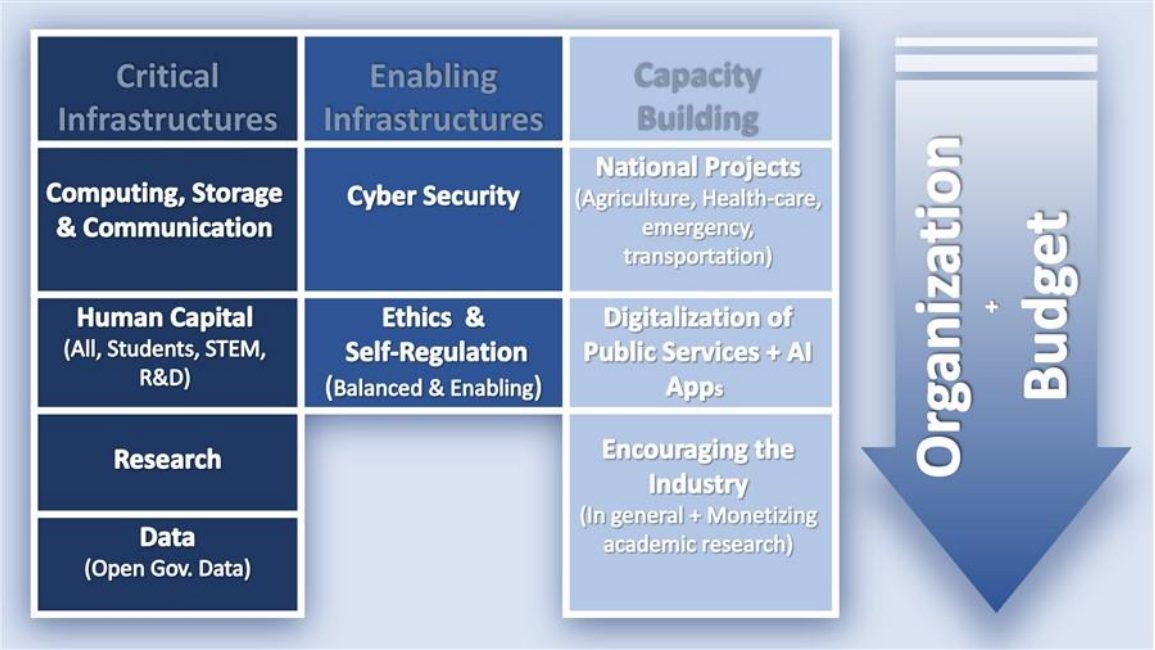
of Israeli experts in various domains and from the academic, industrial and governmental sectors volunteered to take part in this endeavor. The experts were divided into 15 working groups dealing with various technological, sectorial and cross-sectorial aspects of intelligent systems, according to the following model[2]:



Each working group analyzed the Israeli AI environment according to its thematic perspectives. The working groups used as a comparative point previous and cutting-edge work that has been conducted by other jurisdictions, in order to present a national plan customized to the specific characteristics of Israel. Conclusions and recommendations were integrated into a final report which proposes a National Strategy for Israel in the field of Secured Intelligent Systems (the "**National Initiative Report**"). It defines intelligent technologies as a national priority and draws an operative national plan for the establishment of a sustainable eco-system in the field of secured intelligent systems. The

[2] Ibid. p.15.

national plan is based on three layers: (1) critical infrastructures; (2) enabling infrastructures; (3) capacity building; and consists of the following building blocks:[3]



The National Initiative Report has been recently submitted to the Israeli Prime Minister.

Part I: AI applications in Israel – A public policy opportunity

"Israel is now number three in the world for AI solutions. With only 8.5 million citizens, Israel has a market share of 11% and is equal to China. Israel has 40x more AI companies per capita than the market leader USA, and that makes Israel the clear hidden champion of Artificial Intelligence".[4]

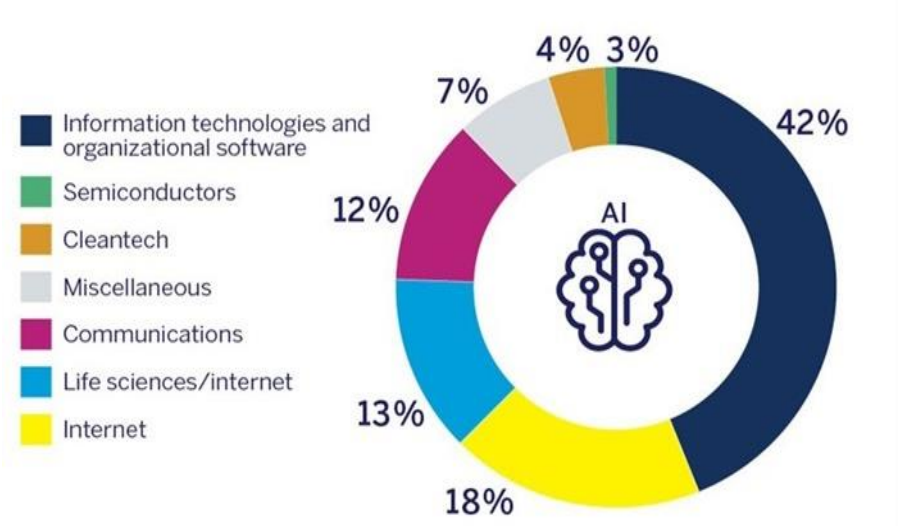
Israel has a strong high-tech and innovation ecosystem coupled with a culture that embraces and adapts to technological developments. The prevailing atmosphere in the "Startup Nation" is one that encourages both the public and the private sectors to explore and use AI applications in various fields. However, the AI applications landscape in Israel is shaped, first and foremost, by the private market.

[3] Ibid. p.23.
[4] ASGARD. *The Global Artificial Intelligence Landscape*. Retrieved from https://asgard.vc/global-ai/7

1.1 The private sector

According to the Israel Innovation Authority, investments in Israeli high-tech AI projects increased in 2011-2019 by a factor of 12.5, from 305 million dollars to 4 billion dollars. In 2019, 42% of the total sum invested in Israeli high-tech went towards AI technologies.[5] Around 1,400 AI companies are currently operating in Israel, developing and utilizing AI technologies in various domains such as business analysis, cyber and healthcare applications and more. Over 40% of the companies deal with information technologies and organizational software, while 30% focus on internet services and communications[6].

Distribution of AI companies by sector

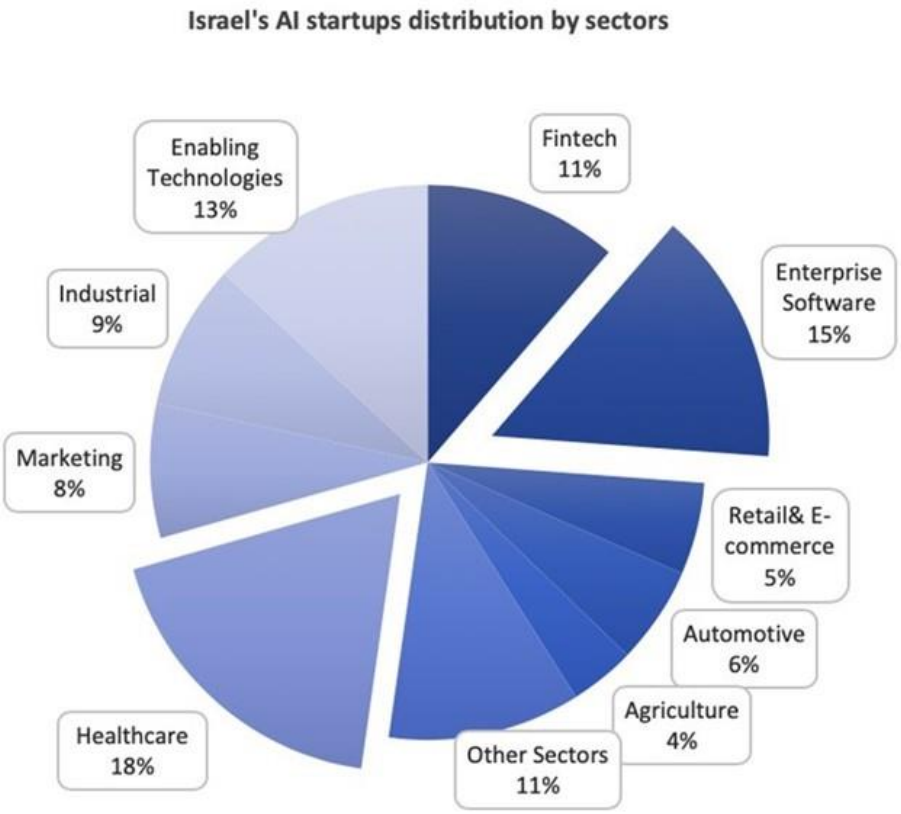


Source: Israel Innovation Authority (2020). *Bolstering Artificial Intelligence*

1,024 of these companies are startups. Despite Israel's small size and limited resources, it ranks third in the world in terms of the number of AI startups, after the United States and China, and first in terms of the number of AI companies per capita.[7] In the past five years, an average of 140 new startups have emerged annually, offering applications and products which cover all sectors and areas of life. However, the leading sector is healthcare with 188 startups (18%) offering AI solutions in the fields such as diagnostics, monitoring, disease management, personalization and clinical workflow. Enterprise software closely follows with 152 startups (15%) developing and utilizing AI products and services in the fields of

[5] Israel Innovation Authority. (2020). *Bolstering Artificial Intelligence*. Retrieved from https://innovationisrael.org.il/en/reportchapter/bolstering-artificial-intelligence-0#footnoteref2_bscx727
[6] Ibid; Israel Innovation Authority. (2020). *Innovation in Israel - 2019 Innovation Report*. p.62 (Hebrew).
[7] ROLAND BERGER GMBH & ASGARD. (2018). Artificial Intelligence – A strategy for European startups. P. 17

sales and training, HR, data and intel, customer and support, development and IT, management and teamwork, security, privacy and finance.[8]



Source: Cardumen Capital. Israel's Artificial Intelligence Startups, June 2020. *Medium*

Israel also ranks high in the number of companies that develop infrastructure technologies for AI such as special-purpose chips, infrastructure algorithms, and complex systems for the acceleration of computing.[9]

For a full map of Israel's AI startup landscape in 2020 segmented by sectors and applications, see Appendix I.

1.2 Government initiatives and policy

In addition to the activity of the private sector, the Israeli government has important role in promoting AI applications. It can do so by initiating projects itself, or by creating an

[8] Cardumen Capital (June 9, 2020) Israel's Artificial Intelligence Startups, June 2020. *Medium*. retrieved from https://medium.com/@cardumencapital/israels-artificial-intelligence-startups-june-2020-81e27d9332d8

[9] Israel Innovation Authority. (2020). *Bolstering Artificial Intelligence*. Retrieved from https://innovationisrael.org.il/en/reportchapter/bolstering-artificial-intelligence-0#footnoteref2_bscx727

encouraging environment for the private sector to further develop and use AI technologies, for example through enabling regulation, incentives for the industry, etc. The great potential which the government attributes to AI technologies and their applications originates in two characteristics of Israeli society:

a. **High population growth rate –** Israel's demographics are exceptional for developed economies and as such they present its government with some unique challenges. With 3.1 children per women, Israel has the highest fertility rate among the countries of the OECD, of which the average fertility rate is 1.6 children.[10] Furthermore, the average life expectancy in Israel is of 82.9 years, the fifth highest within the OECD.[11] Consequently, the annual population growth rate in Israel – 1.9%[12] – is almost four times higher than the average population growth rate of OECD member countries (0.54%).[13] The consistent fast growth of the population requires the State of Israel to adjust its public and social services and to maintain and increase accordingly its infrastructures in all fields of life (e.g. healthcare, transportation, education, energy, etc.).
b. **Population density and overloads on infrastructures –** the majority of the Israeli population lives in the center of the country, and over 40% of the population is spread over less than 7% of the country's territory.[14] This leads to severe overloads on the infrastructures and services in highly populated areas. One prominent example is the growing traffic congestion in the center of the country, which have negative ramifications on productivity, the environment and the number of accidents and casualties

The high population growth rate and the population density shape Israel's approach to AI, as they significantly increase the make the Israeli demand for national infrastructures and services. The government's motivation to increase the use of AI, beyond the areas that are already covered by the private sector, lays in the technology's potential to answer the growing need to enhance availability, reliability and efficiency of public services and national infrastructures, at lower costs to the state and its citizens. For this reason, among

[10] OECD (2020), Fertility rates (indicator). Retrieved from https://data.oecd.org/pop/fertility-rates.htm#indicator-chart
[11] OECD (2020), Life expectancy at birth (indicator). Retrieved from https://data.oecd.org/healthstat/life-expectancy-at-birth.htm
[12] Central Bureau of Statistics. (2020). Israel in Figures Selected Data From the Statistical Abstract of Israel 2019. Retrieved from https://www.cbs.gov.il/he/publications/DocLib/isr_in_n/isr_in_n19e.pdf p.6.
[13] World Bank, Population Growth for OECD Members [SPPOPGROWOED]. Retrieved from FRED, Federal Reserve Bank of St. Louis https://fred.stlouisfed.org/series/SPPOPGROWOED
[14] Central Bureau of Statistics. (2020). *Statistical Abstract of Israel 2019 - No.70.* Retrieved from https://www.cbs.gov.il/he/publications/DocLib/2019/Shnaton70_mun.pdf (Hebrew). p.21.

the first sectors for which the government promotes AI solutions are healthcare and transportation, where the overload on current infrastructures is most acute.

However, governmental ministries in Israel differ in their readiness to embrace AI applications due to variance in digital maturity and in some cases even digital gaps. Israel's National Digital Initiative, also called "Digital Israel", is the government body responsible for e-government services. Digital Israel spearheads government efforts for digital transformation, to reduce socioeconomic gaps, promote economic prosperity, and create a smarter, friendlier government. Its scope of activity encompasses a broad array of e-government services, at all levels of government (including municipalities). Thus, Digital Israel works with other government ministries, assisting them in developing and deploying digitization plans. Digital Israel also leads the government initiatives plans for Smart Cities as well as the National Plan for Digital Literacy.

To a very large extent, the fulfillment of Digital Israel's mandate depends on the availability and transferability of data. Indeed, in order to maximize the full potential of digital transformation, government bodies must be able to collect large amounts of information, combine it with data from other sources, and deploy technical tools to analyze the data and draw conclusions. In many cases, the data that must be collected and shared includes personally identifiable information ("PII"). In addition, the software tools that can be used include big data analysis, much of which can be enhanced by machine learning. Thus, the projects that Digital Israel wishes to implement face a key challenge, namely, how to balance between the data needs, on the one hand, and the legal and ethical considerations on the other.

To date, there is no all-encompassing government policy to address this challenge. Such a policy is currently being finalized. Pending its adoption, Digital Israel's activities are informed by the existing legal framework, which includes constitutional human rights protections (privacy, non-discrimination, freedom of expression) as well as administrative law rules and principles applicable generally to all government bodies (transparency, accountability, fairness, due process and reasonableness).

Below is a description of the main projects involving AI that are in the process of development (in each case, in conjunction with the relevant government ministry). In the development of each of those projects, Digital Israel has worked with in-house counsel and Ministry of Justice constitutional counsel, to ensure that the development and deployment of each project complies with applicable constitutional and administrative law limitations

1.1.1. Digital health

The project has ambitious goals, including:[15]

- Customized treatment: promoting research, development and implementation of tools that allow the patient to receive the best and most personalized treatment;
- Promoting health and patient prevention through use of digital tools in a way that shifts the focus from patient care to preventive medicine;
- Sustainable health: promoting the development and implementation of systems that increase the operational and managerial effectiveness of the health system, in a way that frees up existing resources;
- Development and implementation of digital tools that streamline communication between the Ministry of Health and those it serves;
- Delivery of emergency treatment services through an appointment management system and an application informing the patient on the progress of the treatment. The information collected will enable better management of resources to avoid congestion in emergency rooms;
- Sharing clinical information across service provider platforms has been expanded, to connect different service providers and allows them to view treatments and diagnoses made by other health professionals in different organizations.

Some of these goals are already being implemented. By a government resolution,[16] the Ministry of Health has created a platform called "TIMNA", which grants third-parties controlled access to health data in order to promote applied research. The data includes vast quantities of health records, gathered by hospitals and clinics from around the country, providing an invaluable resource. Access to the data is subject to strict privacy and ethical restrictions. First, the institutions, researchers and start-ups seeking access to this data must provide Helsinki committee approval for their project. They are required to identify the specific types of data that they need, and only that data is provided. They must sign privacy commitments. All personal data is anonymized. The research takes place entirely within the digitized platform – no personal data can be extracted from the platform, further protecting confidentiality. Furthermore, before the research is published and an algorithm

[15] Ministry of Health digital services home page (Hebrew) https://www.health.gov.il/About/projects/DigitalHealth/Pages/default.aspx.
[16] Israel Government Resolution 3709, "National plan to advance digital health as a means to improve health and foster growth" August 23, 2018, https://www.gov.il/he/Departments/policies/des3709_2018 (Hebrew).

is used, the Ministry of Health reviews it to ensure that no personal information is used or disclosed. Thus far, the TIMNA projects making use of AI are as follows:

- The Israel Center for Disease Control applies an AI algorithm to review diagnostic forms of patients and verify their cancer diagnosis. This saves significant amounts of time, as it automates the process of reviewing over 100,000 forms a year. Audits are conducted to ensure that there are no false negatives.
- Similarly, an algorithm is being developed to assist with medical follow-ups in two areas: child development and pregnancy. In both cases, the algorithm analyzes in real time status reports, diagnoses and notes and recommendations of doctors and nurses, comparing them against standard protocols. It then alerts the hospital or clinic of potential errors, misdiagnosis, or issues that might require additional testing or follow-up. The system is geared towards assisting health professionals in catching mistakes and does not entail significant ethical risks to the patient.
- Another field of study is the use of AI to analyze of medical images (MRI, CT etc.). Thus far the results have been promising, in that the algorithms have been able to detect cases that were missed by doctors. The intent is not to supplant the doctor's decision-making but rather to streamline the process and assist him/her in analyzing the images.
- Finally, the Ministry of Health has deployed AI algorithms to assist with its efforts in slowing the spread of COVID-19. Often, the epidemiological study based on discussions with an infected individual are incomplete, due to failure to remember all locations visited, the interviewer's failure to enter all the information correctly, or other human error. The algorithms are used to form a more complete picture of the likely course of previous infections and predict future infections. This information is then used to inform government policy with respect to closure measures at a general scale. It should be noted that the information is not used to make decisions about specific individuals or communities.

1.1.2. Transportation

The Ministry of Transportation is establishing a pilot project to enable testing of autonomous vehicles. The project would allow manufacturers to apply for a special license, under which they may test their product in real-world conditions, in low-risk driving environments. To that effect, the Ministry published a draft bill, which is open for public

comments.[17] The draft bill does not yet contain all the rules that will apply to trials or to the requirements of an autonomous driving system (capabilities, safety, and oversight). These issues, as well as ethical issues that could arise, will be addressed at a later stage. This approach reflects a cautious and incremental innovation philosophy: in order to understand the impacts of a technology, and given the risks to human life, the trials are permitted in an environment that is not "controlled" but that presents a relatively low risk. This should enable policy-makers to make adjustments before moving forward with larger scale experimentation.

Progress is also being made in the field of public transportation as innovative solutions are being developed to reduce traffic congestion[18].

1.1.3. Taxation

Israel's Tax Authority launched a project to assist investigators in detecting tax fraud. The project uses AI tools to predict the likelihood of tax fraud, based on certain indicators. Privacy concerns were central to how the project was designed. Indeed, the project is based on a layered approach for access to information: initially, few indicators and little information is used to flag risks of fraud; if the initial investigation suggests a higher risk of fraud, only then is personal information required in order to determine with greater certainty the identity of the potential offender.

1.1.4. Proposals for national projects

In addition to the aforementioned unfolding projects, the National Initiative Report recommends that the government launch, in cooperation with the industrial and the academic sectors, four more national projects in the fields of healthcare, transportation, security and agriculture. All four were conceived to answer genuine national needs deriving from developments withing Israeli society which trigger demands for improved and novel infrastructures and services.

[17] The text of the draft bill can be found here: https://www.nevo.co.il/law_word/law11/200820-2.doc (Hebrew).
[18] Ben Dror, M. and Azaria, M. (July 24, 2020). Israel's 'smart commuting' shows what public transport could be like after COVID-19. *World Economic Forum*. Retrieved from https://www.weforum.org/agenda/2020/07/israel-smart-commuting-after-covid-public-transport-innovation/

- **Healthcare** – Reforming the national healthcare system by improving the quality and availability of medical services, and relieving the overloads on hospitals by launching a national system based on intelligent technologies for: (i) remote patient management; (ii) more efficient triage and treatment in emergency medicine; (iii) generating comparative quality indices for measuring clinical outcomes.[19]
- **Transportation** – *"Of the many ways in which intelligent systems can solve acute problems in the field of transportation, we chose to recommend, at the first stage, the installation of Smart Traffic Lights in an entire pilot metropolitan area, with the intent to address traffic congestion, which is the most severe and acute transportation problem in Israel."*[20]
- **Security** – Creating a national dual-system that will harness the potential of intelligent technologies to improve predicting abilities and decision-making processes, for better management on the national level. By collecting and analyzing data for civilian applications of command and control in normal times, it will enhance the national capacity to prepare for times of emergency (natural disasters, epidemics and security threats from enemies) and to make decisions during crises.[21]
- **Agriculture** – "*In an age when food security, water management and other areas in agriculture become acute global challenges, […] We recommend promoting a national project to develop an intelligent technologies based system for early detection of pests and diseases in agricultural crops; alongside integration of intelligent systems into the agricultural sector for optimizing the use of nature resources and [other] inputs to ensure optimal food production."*[22]

Finally, acknowledging the variance in digital maturity within the government, the National Initiative included a working group dedicated to the government sector, in order to set guidelines for preparing the entire government to the age of AI. The working group assessed the required organizational, technological and regulative measures to foster implementations of AI applications within the government, in order to improve both the inter-ministerial work and the interactions between the government and the citizens.[23]

[19] Ben-Israel, I., Matania, E. & Friedman, L. (Eds.) (Sep. 2020). *The National Initiative for Secured Intelligent Systems to Empower the National Security and Techno-Scientific Resilience: A National Strategy for Israel. Special Report to the Prime Minister*. (Hebrew) p.35.
[20] Ibid. P.37
[21] Ibid.
[22] Ibid. P.36
[23] Sharvit, S. et al. (2020). Government Working Group Report. In Ben-Israel, I., Matania, E. & Friedman, L. (Eds.). *The National Initiative for Secured Intelligent Systems to Empower the National Security and*

Part II: Risks and challenges posed by AI in the fields of human rights, democracy and the rule of law

The topics of human rights and ethics were addressed by the National Initiative, through a dedicated working group that studied the issue in depth from a regulatory and ethical perspectives. The working group's report ("**Ethics Report**")[24] identifies issues that are novel and unique to AI and the ways in which it is expected to affect our lives. In light of these issues, it sets forth a series of ethical challenges posed by the technology regarding human rights, democracy and the rule of law.

Below is an outline of the issues identified by the Ethics Report, followed by an overview of the risks and ethical challenges that it suggests addressing.

2.1 What Is New and Special about AI?

a. *"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:*

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

Techno-Scientific Resilience: A National Strategy for Israel. Special Report to the Prime Minister. pp. 214-227. (Hebrew).

[24] Nahon K., Ashkenazi A., Gilad Bachrach R., Ken-Dror Feldman D., Keren A. and Shwartz Altshuler T. (2020). Working Group on Artificial Intelligence Ethics & Regulation Report. In Ben-Israel, I., Matania, E. & Friedman, L. (Eds.). *The National Initiative for Secured Intelligent Systems to Empower the National Security and Techno-Scientific Resilience: A National Strategy for Israel. Special Report to the Prime Minister*. pp. 172-119. (Hebrew).

- *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."* [25]

b. *The procedural challenge: How to "engineer" values*. This issue arises in areas where AI systems are developed to replace human decision-makers who are 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*." [26]

c. Privacy and autonomy risks of unprecedented scope and scale. see sections 2.2 and 2.3 below.

d. *Complexity that erodes public trust*. Lack of clarity and public understanding of how AI systems operates and how it affects our lives often leads to distrust, which may result in reluctance to embrace the technology, even in areas where AI systems offer a clear business – and social – benefit. The report mentions in this regard, the assessment of the EU's High-Level Expert Group on Artificial Intelligence.[27]

e. *"Unfair economies of scale. […] 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*." [28]

f. *"Changes in familiar warranty categories. The ability to collect and process data through products* [IoTs] *enables companies to offer new related services, but also raises new questions about 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*

[25] Ibid. p. 179.
[26] Ibid. p. 180.
[27] 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)
[28] Nahon K. et al. (2020). p.180.

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." [29]

2.2 Ethical risks and Challenges

The Ethics Report address the following ethical challenges. It should be noted that it relies extensively on the EU's High-Level Expert Group on Artificial Intelligence in elaborating upon the different ethical principles.

2.2.1 Security

The right to life and security is often overlooked in global discussions involving human rights. And yet, it remains the most fundamental right of all, absent which other human rights cannot be applied. At a basic level, with respect to AI, the Ethics Report underlines the need to secure AI applications and AI-enabling networks and computers. The report makes several important observations in that regard. It notes that information is the "energy that fuels the current wave of AI", such that security of AI applications and networks is a precondition to development and implementation of AI technology[30]. It further notes that the information that can be collected by AI to build and deploy AI tools includes vast amounts of personal and commercial information, including personal, medical, economic and other sensitive information. It notes that even information that appears "non-sensitive" can become sensitive when cross-referenced with other information.

From a policy perspective, this ties security directly with human rights. For example, protecting privacy requires, at a fundamental level, securing private information from malicious cyber operations. Similarly, beyond commercial and performance considerations, protecting data integrity is also a human rights imperative: in order to protect against bias in a particular AI application, the data as well as the algorithms upon

[29] Ibid. p.181.
[30] Ibid. p.190.

which the application is based must not be tampered with. Freedom of expression and access to information are also highly dependent on the security of AI-related applications and networks.

The use of AI further deepens the reliance upon computers, hence creating new vulnerabilities for cyber-attacks. As part of the national initiative, a dedicated working group for cybersecurity in the age of intelligent systems identified new cyber threats presented by AI technologies. First, attacks against AI systems, which can result in damage to the decision-making mechanism, thus leading to false, misleading or biased decisions, or to threats against the system's IP. Second, malicious exploitation of AI capabilities as cyber weapon to launch sophisticated "intelligent attacks". These vulnerabilities of AI systems raise the question – whether and how an AI system can be authenticated as secured and reliable.[31] AI security is thus a basic layer over which richer interactions can take place. Its importance is indeed reflected in the title of the Israeli national AI initiative - the National Initiative for *Secured* Intelligent Systems.

2.2.2 Privacy

Tellingly, privacy is the first and foremost of human rights addressed by the Ethics Report. The report underlines that AI applications "are largely based on information about individuals or on deriving conclusions about them from personally identifiable information".[32] Protection of privacy is largely dependent upon a robust legislative framework. In Israel, this framework consists mainly of Israel's Basic Law: Human Dignity and Liberty (1992),[33] its Privacy Law (1981), several privacy regulations including data transfer regulations[34] and the comprehensive 2017 Protection of Privacy Regulations (Data Security).[35] The legal regime is complemented by extensive case law and a robust judiciary. Over time, a number of privacy protection principles have emerged: the need for legal cause for collecting and processing information (e.g. informed consent), usage limitations,

[31] Zack, H. et al. (2020). Working Group on Cyber and Intelligent Systems Report. In Ben-Israel, I., Matania, E. & Friedman, L. (Eds.). *The National Initiative for Secured Intelligent Systems to Empower the National Security and Techno-Scientific Resilience: A National Strategy for Israel. Special Report to the Prime Minister*. pp. 168-171. (Hebrew). p.168.
[32] Nahon K. et al. (2020). p.188.
[33] http://knesset.gov.il/laws/special/eng/BasicLawLiberty.pdf. This is a quasi-constitutional law, whose underlying principles are seen as constitutionally mandated, even in the absence of a formal written constitution.
[34] Privacy Protection (Transfer of Data to Databases Abroad) Regulations (2001). See unofficial translation here: https://www.gov.il/BlobFolder/legalinfo/legislation/en/PrivacyProtectionTransferofDataabroadRegulationsun.pdf.
[35] See for unofficial translation here: https://www.gov.il/BlobFolder/legalinfo/data_security_regulation/en/PROTECTION%20OF%20PRIVACY%20REGULATIONS.pdf

the right to review and correct one's personal information, transparency vis-à-vis the information owner and the obligation to protect the information

Against this backdrop, the Ethics Report notes that there remain gaps between traditional conceptions of privacy protection and the challenges raised by AI. Indeed, to the extent that AI relies on the collection and processing of PII, it can be expected that novel privacy issues will arise, which may require adapting existing privacy laws further down the road. The Ethics Report also observes that in certain cases, there might arise a conflict between privacy and fairness. If individuals belonging to a certain group refrain from sharing their personal information with an AI application, that application will not be able to draw from data that takes this group into account, potentially leading to greater discrimination.[36] There is thus a policy imperative to enabling the collection of PII while ensuring that such data will be both secured and subject to robust privacy protections.

2.2.3 Autonomy

The Ethics Report defines autonomy as "the individual's ability to make intelligent decisions, including the prevention of unfair or unconscious influence on individual behavior."[37] In human rights terms, this can refer to concepts such as human dignity and the right of access to information. The Ethics Report states[38]:

> *"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*

[36] Nahon K. et al. (2020). p.188.
[37] Ibid. p.182.
[38] Ibid. p.189.

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."

2.2.4 Civil and political rights

The Ethics Report defines civil and political rights as including the "right to elect, freedom of speech and freedom of conscience religion."[39] These go to the core of democratic values and warrant special protection. In that respect, the Ethics Report notes with concern how the automated manipulation of global discourse is manipulated, for example by over-amplifying certain views while silencing others, polarizing the discourse and giving legitimacy to views that could be offensive to certain groups, and disseminating false information on a large scale.[40]

All these can harm the democratic process itself, creating rifts within society and undermining faith in the democratic process, and produce.

[39] Ibid. p.182.
[40] Ibid. p.190.

2.2.5 Fairness

This is a broad ethical principle, that refers to the need to achieve substantial equality, to prevent of biases (in information, in the process and in the product), prevent discrimination, and avoid widening socioeconomic and educational gaps. The Ethics Report notes: *"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*".[41] It provides the following examples of AI systems that raise fairness questions:

- 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.

To address these risks, the Ethics Report underlines the importance of proactively studying the target population and identifying in advance groups that are liable to be misrepresentation or underrepresentation. In addition, it emphasizes the need to consult with representatives of the target users themselves to help produce fairer systems.[42]

2.2.6 Accountability

The Ethics Report separates accountability into three categories: *transparency, explainability and responsibility.*

Transparency is about *"Providing information about the process and related decision making"*[43] and is referred to as a *"key value in technological development and in developing AI products in particular"*.[44] It is both a value that stands alone, and an aspect of accountability as well. It is enables the monitoring and realization of other values such as fairness. Transparency is a core component of public trust.

Explainability is an AI system's capability of explaining its decision-making process, whether to the individual end-user, or on a collective level if the decision affects group. It

[41] Ibid. p.184.
[42] Ibid.
[43] Ibid. p.182.
[44] Ibid. p.185.

also includes a system's capability of providing meaningful explanation to the operators of the system themselves. The Ethics Report adopts the EU Experts Group Report's position that explainability includes a principle of "meaningful information", that is that the level of information provided should be sufficient without being exceedingly technical or detailed.

Responsibility, involves making appropriate rules to prevent risk, based on the the context and the estimated severity of the risk, managing the risks and appointing an employee in charge of risk management. The Ethics Report notes that the diversity of stakeholders and the complexity of AI systems make this a particularly challenging task. This is compounded by the fact that AI systems also make their own "decisions".

2.2.7 Safety

The Ethics Report recognizes the need to address safety risks that arise from AI systems. Indeed, the more an AI system is empowered to make decisions with a direct impact on human life, the riskier is it to use. The risk arises both in ordinary operation of the AI system, as well as in extreme situations. The Ethics Report thus distinguishes between safety risks occurring as part of a malfunction, and those that occur when the system operated without malfunction but in a manner that nonetheless causes death or physical harm.

Safety risks can be mitigated by implementing a number of measures. For example, in order to prevent incorrect decisions based on faulty bias, a diverse dataset should be used. Similarly, safety considerations must be borne in mind at the design stage. The Ethics Report provides an interesting – and perhaps counter-intuitive – example. In the design of an autonomous car, it is important for the system to have trained on diverse conditions, including conditions where harm could be imminent. For this training process to occur, it would be necessary to place individuals in risk situations, which can then form part of the data set.[45] Of course, this could at least in part be done through simulations not involving physical human beings, but it highlights the tensions that exist in order between different ethical and human rights principles, and the trade-offs that are sometimes necessary in order for AI systems to be as "good" as possible.

[45] Ibid. p.190.

2.2.8 Fair competition

Among the different ethical principles that are commonly referred to in various AI ethics documents and standards, fair competition is probably the least often quoted. By contrast, it is a standalone ethical principle in the Ethics Report. Fair competition refers to the need to enable innovators, entrepreneurs, software engineers and other stakeholders in the supply chain, to benefit from equitable access to data and opportunities to create and deploy AI systems. This principle may, at first glance, be characterized as an economics goal in disguise, but the Ethics Report develops it on distinctly human rights and ethics-based grounds. Indeed, the report notes: that fair competition is needed "*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*."[46] The ethical imperative is societal as opposed to individual: in order for society to reap the benefits of AI transformation, greater innovation is needed, by a diversity of actors.

As noted by the report, examples of the challenges posed by unfair competition include:

- 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.

The Ethics Report notes: "*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*."[47] It further suggests that competition laws, standardized contracts and consumer protections be updated 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 in the United States*."[48] In short, a major part of the challenge is to enable SMEs to have access to large databases.

[46] Ibid.
[47] Ibid. pp.190-191.
[48] Ibid.

Part III: Israel's approach to address the challenges

Israel is aware of the potential risks and challenges presented by the growing use of AI applications and is determined to address them. At the same time, as noted by the National Initiative Report, experience shows that over-regulation can stifle innovation, particularly when dealing with emerging technologies[49]. The National Initiative included two working groups dedicated respectively to ethics and regulation, and to cybersecurity for AI systems. Their main challenge was to establish a model that would **balance** the need to: (1) ensure ethical and secured development and deployment of AI applications in accordance with the values of Israel as a democracy; (2) foster technological innovation and scientific research and development which are fundamental to the Israeli economy and national security. As will be elaborated below, the approach of the National Initiative Report is novel in its manner of combining ethics, human rights and innovation.

3.1 Six Ethical Principles for AI

Acknowledging that human rights and ethical considerations remain paramount, the Ethics Report's lists "6 Ethical Principles for AI" that should inform public policy making:[50]

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.*

[49] Ben-Israel, I., Matania, E. & Friedman, L. (Eds.) (Sep. 2020). p.32.
[50] Nahon K. et al. (2020). p.182.

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.*

In light of these guiding ethical principles, the National Initiative Report suggests a balanced regulatory model based on applying the minimal regulatory intervention required for maintaining adequate ethical environment, on the one hand, while refraining from any unnecessary restrains on innovation and scientific progress, on the other hand. Accordingly, the National Initiative Report calls "*to encourage self-regulation through the use of the tools developed within the framework of the [national] initiative to assess risks and identify in advance ethical challenges in the stages of development and production. Ethical limitations should be integrated into the intelligent systems, forbidden conducts should be defined, and the ethical principles should be implemented during the learning and training process of those who deal with AI systems".*[51]

3.2 Balanced regulation to foster innovation

One of the key features of the Ethics Report is its approach to regulation. Rather than setting out a list of activities that must be regulated, it takes a systematic approach to the question, comprising three steps: (1) mapping of different types of regulatory approaches, along with their respective advantages and drawbacks; (2) identifying the main areas and activities of AI that could benefit from some level of regulation, and the risks associated with each of them; (3) matching different regulatory approaches to the various AI activities. This provides a roadmap for the government to craft tailored, sector-specific regulations.

Details of the approach are provided below.

[51] Ben-Israel, I., Matania, E. & Friedman, L. (Eds.) (Sep. 2020). p.31.

The Ethics Report identifies the following broad regulatory approaches:

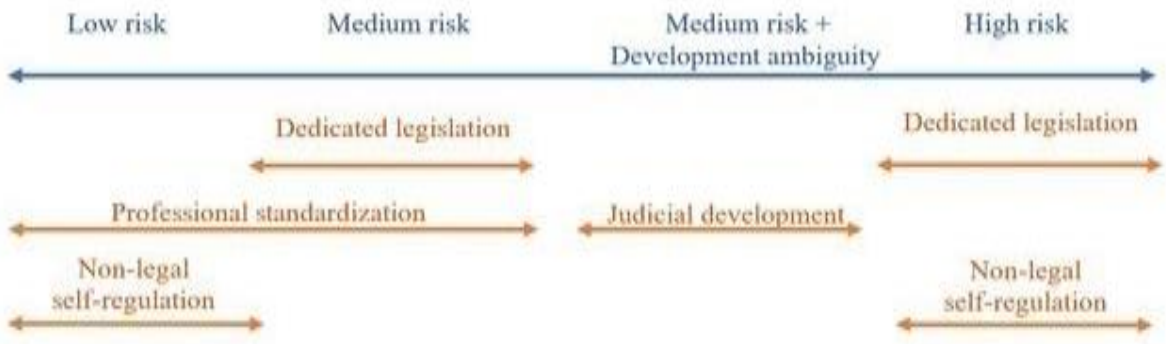
1. Legislation or regulation
2. Judicial decision making to interpret existing legislating or fill the gaps
3. Professional standards (by government, industry, academia or civil society)
4. Self-regulation by ethical rules or professional standards usually developed by the relevant professional community.

The report then highlights the advantages and drawbacks of each approach. The following table is reproduced as-is from the Ethics Report[52]:

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

[52] Nahon K. et al. (2020). p.200.

The Ethics Report then proposes the following model, to match different regulatory approaches based on the risk level associated with a particular activity:



Thus, for example, high-risk activities are better addressed by legislation and self-regulation ex ante, than by post facto judicial intervention. At the other end, low risk activities do not necessarily require dedicated legislation, and can be addressed through standards and self-regulation.

This model, of course, is not meant to apply in a rigid fashion. Rather, it presents a framework that enables policymakers and regulators to gauge the appropriate means of an activity, factoring in a multitude of variables. It further notes that the question of "who regulates" is no less important: regulation by a central AI body enables the development of consistent policies; however, there is a risk of over-regulation and chilling innovation if a regulation is adopted across the board. Conversely, regulation could be left to different sector-based bodies, which would allow for greater experimentation, at the expense of uniformity of rules.

In light of the foregoing, the Ethics Working Group proposes 11 regulatory guidelines:

1. **Alignment of Israel's regulation with international legislation and standardization, and promoting Israeli policy in global arenas** – this is essentially about participating in the international discussion around AI regulation, to be attuned to emerging international standards, while taking part in the shaping of those standards going forward.
2. **Mapping the actors to create an adapted responsibility and incentive framework** – this requires a multi-stakeholder approach, to enable policymakers to understand their respective roles in the value chain, their incentives, and their responsibility.

3. **Adjusting the accountability principle to the dynamism of the AI area** – the suggestion here is to require that organizations implementing AI technologies implement a testing environment and control perimeters prior to implementing the technology, in order to determine how to best apply the accountability principle in a given case taking account the anticipated effects of the technology.
4. **Promoting normative clarity in critical stages of the AI product value chain** - this emphasizes the importance of guidance in the early stages of AI development. An AI risk assessment tool, and perhaps in certain cases a regulatory requirement for AI impact assessment, would be useful in mitigating risks and in enabling developers to implement the various ethical principles and legal rules.
5. **Constant review of the regulatory policy by the regulator** – beyond monitoring the implementation of existing regulation and updating legal texts, this principle calls for regulatory experimentation. It requires regulators to take an agile, flexible approach, promote innovation while factoring in risks.
6. **Regulatory sandboxes** – the concept of regulatory sandboxes is well known. Controlled testing is particularly useful in an AI context "*because of the need to allow innovation on the one hand and address unpredictable risks to social interests on the other*".[53]
7. **The interface between the proposed principles and existing regulations** – given that laws and regulations already apply in many fields of activity (health, transportation, finance, education, etc.), the existing legal landscape must be borne in mind, along with the specific values, interests and potential social benefits of regulation, in determining whether new regulation is needed and what ought to be its focus and scope. At a basic level, every government body is already responsible to undertake this examination within the scope of the field it regulates.
8. **The role of the Privacy Protection Authority** - Privacy is a cross-sectoral issue, such that the Privacy Protection Authority has a predominant role to play in assessing the privacy implications of AI systems, and making regulation as needed, in coordination with other government bodies. Furthermore, it is important that the Privacy Protection Authority obtain the resources required for developing an up-to-date legal and

[53] Ibid. p.203.

technological framework for the area of information anonymization, as it is a fundamental to the development of AI.

9. **The role of the Competition Authority** – as mentioned above, competition and a fair market is not just about economy – it is also an ethical matter. Thus, the Competition Authority should be tasked with *"formulating regulations designed to maintain fair competition in the AI area, protect consumers 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."*[54]
10. **The need for interministerial coordination** - to ensure coherent policy and regulation development, an interministerial coordination mechanism should be implemented.
11. **Authorities responsible for information resources**

    Authorities that are responsible 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."[55]

3.3 Original Ethical Risk Assessment Tool

As the Israeli approach encourages self-regulation, the Ethics Report stresses the responsibility of all those involved in AI to remain up-to-date with the risks of the dynamic technology. To assist them in this demanding duty, the Ethics and Regulation working group developed an original *Decision-Maker Instrument for Assessing Ethical Challenges.*

[54] Ibid. p.204.
[55] Ibid. p.205.

The instrument is designed to enable AI professionals to identify ethical risks throughout the development and production change and to respond properly. It consists of two parts:

1. A set of preliminary questions that should be addressed to AI product developers in order to assess the influence of the product:[56]

1. *What is the level of potential individual harm?*
2. *What is the extent of potential perceptual impact?*
3. *What is the degree of potential damage to the public?*
4. *Is there any impact on the allocation of public resources?*
5. *Is the development team diverse enough?*
6. *What is the expected extent of damage due to misuse of or loss of control over the product?*
7. *Is there a fast way to identify unpredicted ethical failures?*

2. A dynamic frequency map that helps locate challenging areas in terms of applying ethical values to the system's development. The map presents the six ethical principles juxtaposed cardinal milestones along the development process. It indicates the frequency of ethical issues along the product's development chain by highlighting areas where failures have been found in the past and providing information about their rate of incidence (See Appendix II for a sample frequency map). The map is based on assessment of real-life past cases of AI systems which presented ethical challenges or conflicts, thus raising awareness to areas where AI organizations experienced trouble in the past, and areas for particular attention by decision makers. It is important to note that as the map is shaped by the test cases used to create it, each organization is expected to select a set of cases that are relevant to the product it develops. Furthermore, to remain relevant, the map needs to be frequently updated with new test cases. The Ethics Report explains in detail how an AI organization can create and update a frequency map relevant to its product. See Appendix II for further information.

3.4 International activity and cooperation

Israel has been involved in international forums dealing with AI ethics and human rights. Israeli representatives were active in the drafting of the OECD's AI Recommendations and guiding principles. In addition, Israel is a member of the "Digital Nations" ("DN"),

[56] Ibid. pp.192-193.

regrouping 10 of the world's leading digital economics. In 2018, Israel hosted the annual DN meeting, in which a declaration on responsible AI was adopted.[57] In 2019, the DN also adopted a declaration on data governance.[58] While these declarations are not legally binding, they reflect the Digital Nations' commitment to abide by high standards of human rights, ethics and accountability in they use of digitization. Israel has also partnered with the World Economic Forum's C4IR project, in conducting research projects in the fields of transportation and health.[59]

[57] Shared Approach on the Responsible Use of AI, https://fdfd812d-4234-49d8-8755-ff45ad565157.filesusr.com/ugd/189d02_ef802d92ba5147d2901bde25c6e954a3.pdf, November 2018, Written by the Artificial Intelligence working group, this framework was adopted at the D9 Ministerial Summit in Israel in 2018.

[58] Data 360 Declaration, https://fdfd812d-4234-49d8-8755-ff45ad565157.filesusr.com/ugd/189d02_abce8f2b8cc140e4baeec7dcab7bee97.pdf, November 2019, Drafted by the Data 360 working group, this shared declaration was presented at the D9 Ministerial Summit in Uruguay in 2019.

[59] Israel Innovation Authority. *Establishment of the Israeli Center for the Fourth Industrial Revolution – World Economic Forum*. Retrieved from https://innovationisrael.org.il/en/contentpage/establishment-israeli-center-fourth-industrial-revolution-world-economic-forum .

Appendix I – Israel's AI startup landscape segmented by sectors and applications

CARDUMEN CAPITAL

Cardumen Capital updates this map quarterly. If you would like to add your startup or reach out to us regarding our expertise in artificial intelligence, please email us to **AI@cardumencapital.com**

CARDUMEN CAPITAL

Source: Cardumen Capital (June 2020). https://www.cardumencapital.com/ai-israel-landscape

Appendix II – Frequency Map of Ethical Challenges in the AI Development Process

The following part is reproduced as is from pages 194-197 of the Ethics Report

1. 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[60].
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[61].
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.[62]
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'

[60] https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine
[61] 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
[62] http://people.dbmi.columbia.edu/noemie/papers/15kdd.pdf

	dangerousness. The system was tested and was found to assess white detainees as less dangerous than black ones.[63]
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.[64]
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. [65]
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. [66]
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.[67]
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.[68]
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. [69]

[63] https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
[64] https://futurism.com/this-video-games-artificial-intelligence-turned-on-players-using-super-weapons
[65] https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist
[66] https://www.androidcentral.com/google-duplex-will-let-people-know-its-not-human
[67] https://www.cbc.ca/news/business/uber-arizona-crash-1.4594939
[68] https://www.theverge.com/2019/1/25/18197137/amazon-rekognition-facial-recognition-bias-race-gender
[69] 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)