Digital Security Lab

INTERNATIONAL RENAISSANCE FOUNDATION

MOVING FORWARD TOGETHER

Running Up That Hill: Artificial Intelligence in Ukrainian Public Sector

Analytical study

Digital Security Lab Ukraine
2024

Running Up That Hill: Artificial Intelligence in Ukrainian Public Sector – NGO Digital Security Lab Ukraine, 2024.

Author: **Tetiana Avdieieva**

The development of technologies did not leave the public sector aside, contributing to the automation of processes, invention of innovative solutions, and improvement of the quality of public services. As a result, more and more initiatives are there to implement artificial intelligence (AI) technologies in the public sector. Some of them aim at facilitating processes within state bodies, others, at improving services and processes, creating a convenient and accessible digital environment. These innovations are implemented in the context of active regulatory processes within the European Union, the Council of Europe and many foreign countries. This analytical study outlines key initiatives of the AI regulation in the international arena, at the national level in Ukraine and in other states, and also provides an overview of the AI projects that are already in practice or in the pipeline for the near future, analyzing their safety for human rights and democratic principles. Finally, Digital Security Lab Ukraine provides recommendations for improving the regulatory environment and ensuring the ethical use of AI in the public sector.

Contacts:
https://dslua.org/
dslua@dslua.org
FB: https://www.facebook.com/dslua
Twitter: @DSLab_Ukraine

This analytical study was compiled with the support of the European Union and the International Renaissance Foundation within the framework «European Renaissance of Ukraine» project. Its content is the exclusive responsibility of the authors and does not necessarily reflect the views of the European Union and the International Renaissance Foundation.

The International Renaissance Foundation is one of the largest charitable foundations in Ukraine. Since 1990 we have been helping to develop an open society based on democratic values in Ukraine. During its activity, the Foundation has supported about 20 thousand projects. The funding amounted to over $350 million.

Site: www.irf.ua
Facebook: www.fb.com/irf.ukraine

The European Union is made up of 27 Member States who have decided to gradually link together their know-how, resources and destinies. Together, during a period of enlargement of 50 years, they have built a zone of stability, democracy and sustainable development whilst maintaining cultural diversity, tolerance and individual freedoms. The European Union is committed to sharing its achievements and its values with countries and peoples beyond its borders.

Contents

Introduction 5

I. International regulation: reconcile the inconsistent? 6

II. Foreign regulation: to regulate or not to regulate – that is the question 19

III. Ukrainian standards: in-between ethics and law 25

IV. Public administration: Diia in action 29

V. The information sphere: on top of the news mountain 32

VI. Educational sector: time to fulfill the Dream (Mriia) 33

VII. Healthcare: who keeps a finger on the pulse? 35

VIII. Social sector: how to support a support system? 37

IX. Tax and customs: the fate of high-profile announcements 39

X. Justice: looking for an electronic Themis 41

XI. About the law enforcement sector, or where military technologies migrate 45

XII. Defense sector: industry as a united front 52

Recommendations 57

Running Up That Hill: Artificial Intelligence in Ukrainian Public Sector

AI will probably most likely lead to the end of the world,
but in the meantime, there'll be great companies.

(c) Sam Altman

In the spring of 2023, ChatGPT accidentally created a fake biography of Oles Honchar, ascribing him the status of a prisoner of war in the USA, forced labor in Germany, and even confusing his birthplace with Dovzhenko. This became clear thanks to the fact that on the poet's anniversary, the Ukrainian media decided to remind readers of facts from Honchar's life – and they made quite a mistake! Such a failure is not the first one – it happened earlier with Shevchenko's biography. And, probably, ChatGPT has already twisted a lot of biographies, because it was officially banned in a number of countries. In addition, Italy emphasized that it suspects the chatbot in violating the rules of personal data protection. It would seem that fictional biographies are not so dangerous, and if someone does not like the possibility of processing sensitive information, such applications can always be avoided. However, such errors are a wake-up call that the AI tends to make mistakes more often than we imagined. And while in entertainment applications, an error does not mean too much, mistakes during the provision of public services, the investigation of crimes or medical manipulations can cause much more serious consequences.

In Ukraine, the practices of using and regulating AI are becoming more popular almost every minute. Back in April 2023, the Ministry of Digital Transformation of Ukraine (MDT) emphasized that we have about 4,200 developers of the AI systems. Obviously, this number has only increased over the past year. Currently, Ukrainian Tech Ecosystem Overview – a national project for monitoring the IT companies, developers and technical ecosystems of the Ukrainian market – includes 1,659 companies that create products and 550 service companies. Many work with automated systems or directly develop them. For example, the DeepGreen Ukraine system – a forestry monitoring service that uses open satellite images and data from the State Forestry Agency to detect illegal logging – was recently developed. And at the end of December 2023, even the use of AI in Ukrainian cinema was reported, which had never happened before. Moreover, over the past year, news columns have been full of headlines about the widespread implementation of AI in the public sector – from the creation of training courses for developers to the automation of processes in most ministries (and even diplomatic missions and embassies).

New state initiatives are announced with enviable frequency – and digitization, which is in full swing, cannot but please. At the same time, the lack of strict standards for both public projects and private developments is somewhat frightening. It is not only about the Ukrainian context – the Brownian movement of international organizations in attempts to balance the interests of industry with human rights and public good makes us wonder if we should even expect adequate regulation in the near future?

I. International regulation: reconcile the inconsistent?

Amid the debate over the regulatory framework, leading tech companies (such as OpenAI, Google, Microsoft and Anthropic) have come together to create a collaborative research hub. The tech giants emphasized that they are exploring the possibilities of creating ethical and safe AI products. At first, the initiative caused concern – human rights defenders considered this move an attempt to prove the ability to develop standards at the level of self-regulation and a signal about 'having no need in state standards.' The public statements of the executive directors of leading companies at the British AI Safety Summit 2023 became a kind of a response. They highlighted the need to develop universal international standards for all developers. In turn, this has caused a wave of discontent among small startups, since some of the proposed requirements may become an excessive financial burden for them. It would seem that no matter what approaches you offer, someone will definitely be dissatisfied. However, is it true that key framework is really lacking? Actually, no.

According to statistics from the Organization for Economic Cooperation and Development (OECD), there are currently about 1,600 regulatory initiatives in the world regarding standards for the development and use of AI. Most of them are available in the US (81 initiatives), the United Kingdom (61), Australia (43), Germany (37), Turkey and Portugal (36 each). These statistics do not include the initiatives that are currently being discussed at the EU level in the national regulation of the respective member states. However, even without international acts, the number of regulatory proposals, strategies, policies and financial programs is currently impressive. Moreover, the requirements for the AI systems that are proposed to be approved at the national level in the 'mother countries' of tech giants are similar in principle and values (requirements for transparency, data protection, labelling of content generated by AI, etc.). So, what is currently happening in the world of regulatory initiatives?

The Draft Act on Artificial Intelligence and the White Paper (EU). These are two key documents that are the fundamental pillars of the AI regulation in the

EU and that have drawn the fire of almost every critic who are anyhow related to digital rights and regulation of the digital environment. The first, of course, was **the White Paper** – a strategic document outlining the AI development priorities and key principles of systems development and application. In a nutshell, the White Paper was an attempt to combine three areas – business, people's welfare, and public interest (national priorities). Although the document was not a draft of a legally binding act, the EU in every way encouraged public participation in its development – providing comments, participation in working groups of representatives of various interested parties, etc. As a result, the White Paper received rather favorable reviews, being considered a positive and balanced step in the AI regulation and a good road map for further initiatives. What was the White Paper about?

- investing in the development of AI with an analysis of market needs, the readiness of the transition to automated systems in certain industries and, in general, the definition of AI as a priority stream for the development of the economy;
- development of the technical industry, increase of capabilities (including educational) and skills in the field of creating the AI systems and their integration into public life – notably, to promote such goals, it was proposed to create digital hubs in each EU member state (at least one hub in each country);
- partnership with the private sector (in the form of public-private partnerships) and the general implementation of AI technologies in the public sector, introducing the infrastructure that facilitates the development of AI – such as access to data, international cooperation, clear national regulations;
- creating a clear regulatory system (notably, making sure that fundamental principles such as oversight, safety, data protection, transparency, non-discrimination, social and environmental welfare, responsibility are complied with);
- general mapping of risks (human rights, security risks, reconsidering the lack of responsibility mechanism, etc.) and framing of AI within the existing EU acts for the period until special regulation is developed and adopted;
- 'spoilers' regarding the future regulation – setting milestones and defining problems that should be solved by the AI Act, providing a list of principles regarding safe development and use of AI, outlining data protection and transfer when developing systems in view of the General Data Protection Regulation (GDPR), and what is important is the establishment of the idea of marking AI-generated content to ensure transparency and quality of high-risk systems.

The Draft Artificial Intelligence Act is not so successful in terms of rapid adoption, but it is clearly a more popular document. And the first reason is, obviously, its mandatory nature. As soon as the Act is adopted, many companies will have to significantly revise their policies, and some will probably even leave the EU market. Naturally, this seriously affects the economy of the member states, which is a partial reason for slowing down the process of Act's adoption. Another reason, indeed, was fundamental disagreement between key stakeholders – just look at the discussion regarding the regulation of generative AI and the foundation models. In particular, the dilemma of what should be regulated – technology or its application – has remained a key issue for a long time? Another challenge was the development of a universally acceptable human rights impact assessment system.

However, at the end of 2023, media reported about a political agreement that was reached as part of negotiations between the European Parliament and the European Council. The changes, though, still to be formally approved to enter into force. The AI Act will become applicable 2 years after adoption (and some restrictions will be already active in 6 months). The text updated after the political agreement has quite a few cosmetic changes, while the general principled approaches were outlined in the press releases of the European Commission and the European Council. For example, the general purpose AI systems – that is, technical solutions that can be used in dozens of different systems and are basic for any product – are still planned to be regulated within the framework of the AI Act.

Adoption of the Act in the near future is a matter of high relevance, because soon the EU will face a change of leadership and re-elections to the European Parliament. It is quite difficult to predict the future attitude of body to be formed and the newly appointed European Commission, so there is hope for the adoption of the Act before the turning point of the redistribution of political forces within the EU. What does the AI Act offer and what political compromise was achieved?

- First of all, the widely-debated model of classification of the AI systems depending on the level of risk. The Act proposes to divide systems into four groups: prohibited, high-risk, limited-risk, and low-risk systems. The future Act contains the biggest number of requirements for high-risk systems (notifying the users, human oversight, registration of activities, high quality of data sets, detailed documentation for authorities carrying out public monitoring, development of adequate assessment systems, high level of reliability, security and accuracy). At the same time, no less important is the list of prohibited systems, which includes, for example, social ranking technologies (similar to the Chinese model), technologies used to persecute or manipulate people. In this case, it is difficult to assess

which category a particular system belongs to, because many technologies may well have dual uses (the facial recognition is a great example). The list of prohibited systems became even more extensive after the provisional agreement which supplemented it by prohibiting the emotion recognition and data scraping tools);

- Mandatory certification for certain AI systems (high-risk) – notably, biometric identification systems, systems created to maintain the operation of critical infrastructure, systems used by the government to provide public services (which is especially relevant in the context of this analytical study), training or professional evaluation systems, etc.;
- Requirements regarding the need to notify users that they are interacting with an AI system, not a real person, as well as the ability to refuse such interaction in cases concerning the application of such systems as biometric identification (except for situations when it is used by law enforcement agencies to investigate crimes);
- Establishing transparency rules for systems designed to interact with people, emotion recognition systems, and systems used to create or process images, audio or video content;
- Creation of a single European AI market, which involves the application of regular EU rules on free movement of goods and technologies and, at the same time, the applicability of single and unified rules for working with AI technologies on the EU territory;
- Establishment of a supervisory body at the EU level – the European Artificial Intelligence Board (EAIB) and the parallel creation of national regulators (or the authorization of already existing bodies to supervise the AI market). In addition, there is a requirement to notify supervisory authorities of incidents in high-risk systems (such as malfunctions, cases of discrimination, data leaks, etc.), and a general list of such systems is planned to be kept in the form of the EU Database. However, as part of the provisional agreement, a consensus was reached on the creation of a new Office for Artificial Intelligence, which will develop more detailed standards in this area. It will make regulation more centralized and somewhat move it from national levels to the EU level regarding groundful issues. The exact scope of powers that will stay with national regulators will depend on the implementation of the Act, though.
- Implementation of the system of fines. Before the political agreement, the fine scale provided for fines ranging up to EUR 30 million or 6% of the company's global annual turnover (in the case of prohibited practices such as real-time biometric tracking or data protection violations), up to EUR 20 million or 4% of the global annual turnover of the company's turnover (for other violations), up to EUR 10 million or 2% of the global annual turnover (for

reporting incorrect, incomplete or false information to regulators). After the provisional agreement, the scale of fines varies between EUR 35-7.5 million and 7-1.5% of the global annual turnover of the company. Such changes are quite a wake-up call for the industry since they indicate an increase in fines, including for procedural violations (such as reporting to regulators).

- Establishing a regulatory sandbox (an initiative described in more detail latter in this study).

Draft Directive on AI Liability (EU). Unlike the AI Act, the Directive is not a document of direct effect – that is, the states must take additional measures for its implementation and develop appropriate mechanisms at the national level. This document was developed based on the Report on Safety and Liability Implications of Artificial Intelligence, Robotics, and the Internet of Things, which outlined quite a few challenges associated with automated decision-making. The document is complementary to the general rules on civil liability and to the AI Act itself. The latter, in contrast to the fairly common opinion among the non-experts, in no way indicates who should be responsible for the errors of AI systems in individual cases – starting from self-driving cars and ending with something like ChatGPT. Therefore, there was a proposal at the EU level to develop separate rules of the game:

- Access to information about a potential violation, which can serve as an important prerequisite for further prosecution – notably, the ability to claim damages;
- Identification of a person responsible for a violation, including determining the stage at which an error in the operation of the AI system occurred – for example, against whom civil proceedings for compensation should be initiated: developer, tester, user, etc.;
- The right to demand the disclosure of necessary evidence based on a court order in cases where it is impossible to otherwise assess whether there was an error in the relevant AI system;
- Reforming the burden of proof system. In particular, a plaintiff whose rights have been violated by an AI system can justify their claims by the fact that the company did not follow the rules regarding due diligence in the development and application of systems, which caused damage. For that, it is proposed to introduce a presumption of the existence of such a causal relationship between carelessness and damage.

International Guiding Principles for Organizations Developing Advanced AI Systems and Code of Conduct for Organizations Developing Advanced AI Systems (EU). The Guiding Principles and the Code of Conduct were developed

within the framework of the so-called Hiroshima process. Both documents are a fairly short summary of the main principles and approaches for regulating the AI. At the same time, their goal is the unification of the industry and the creation of a platform for dialogue rather than the development of standards. As follows from the content of the documents, they are planned to be 'living' tools – that is, to be regularly updated and supplemented with relevant information (for example, as the generative AI develops, it will definitely be worth preparing separate provisions and a vision for its development, testing and application). Here is the summary of documents:

- **Guiding Principles**: conducting risk assessment during the entire lifecycle of the systems, monitoring possible ways of malicious use of systems and developing mechanisms to reduce such risks, adhering to the principle of transparency (including on systems' vulnerabilities), disseminating information about incidents related to AI systems (for example, data leaks), developing crisis and security policies, security mechanisms (including physical security, such as on storing information on servers in countries with a high index of human rights violations), developing a system for marking AI-generated content, prioritizing effective means of risk mitigation (in fact, choosing a proactive rather than a reactive approach), as well as prioritizing the development of systems that bring public good (educational, healthcare, climate change) in compliance with international technical standards and requirements for data protection and IP requirements;
- **Code of Conduct**: addresses the Guiding Principles detailing how they should be implemented in practice. In particular, the document states, for example, what risks may occur in the systems (such as their use in the field of manufacturing chemical, biological or nuclear weapons – every year this risk seems more and more realistic), the content of public reports to maintain transparency (regarding the capabilities of the systems, detailed analysis of potential damage to personal data or threats of discrimination, etc.). That is, the document is rather an 'action plan' for the developers of AI systems, which tells exactly how to implement the standards and requirements provided by the Guiding Principles.

Draft Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law (Council of Europe). Within the framework of the special Committee of the Council of Europe on Artificial Intelligence, a document aimed at the codification of key principles and standards on AI regulation has been under development for the past five years. Unlike the AI Act, the Convention is more general and is rather an attempt to unite as many states as possible with universal obligations (in addition to the countries of the Council of Europe, the US, Canada, Australia and many others are actively involved in its development). Notably, its main provisions (in version of December 2023) provide for:

- Introducing the concept of AI, which focuses on the possibility of automated systems to help a person make decisions or do it for them, based on statistical and mathematical methods;
- Respect for human rights, democratic processes and the rule of law as a focus of the Convention and any regulatory document developed on its basis at the national level;
- Principles of AI development and application: transparency and oversight, accountability, equality and non-discrimination, respect for privacy and data protection, reliability and security, safe innovation, as well as ensuring inclusivity in standards (regarding persons with disabilities and children) and interpreting the Convention in a wider context – that is, in the light of other human rights instruments;
- Development of a mechanism for assessment of impact of the systems on human rights, democracy and the rule of law and applicability of such a mechanism throughout the lifecycle of systems;
- Provision of legal remedies (although the Convention lacks details on what means should be provided and what will happen if they are lacking), procedural safeguards, implementation of risk assessment and interpretation of the methodology for such assessment (so that those who apply it to verify the systems created AI understand the principles and approaches of risk reduction).

At the same time, at the beginning of 2024, fierce debates are raging around the Draft Convention – notably, regarding the scope of its future provisions. Currently, the word goes among the experts involved in the development of the Convention that the proposed rules cannot apply to the private sector and to developments in the defence sector (or, more precisely, national security issues). As a result, the scope of the document is potentially only narrowed to technologies used by the government in the civil sector, giving rise to an extremely high number of risks for users. In response to this, representatives of civil society drafted an open letter to the countries participating in the discussion with a call not to narrow the scope of the Convention down.

Guidelines for Convention 108+ on AI (Council of Europe). This fairly short document supplements (or rather modernizes the interpretation of) the Convention 108+ and allows its application in current context. Compared to other regulatory acts and drats, these Guidelines are quite generic (as is the entire Convention 108+), but important from the perspective of the framing the AI systems in matters of personal data protection. What is this document about?

- General recommendations. For example, conducting a risk assessment and applying the general principles provided by Convention 108+ to the field of AI;

- Recommendations for developers, manufacturers and service providers. For example, these actors are asked to develop a system based on a value approach. This implies a focus on the protection of human rights and observance of due diligence in the development and application of AI systems, and the requirement of a risk assessment. In addition, industry representatives are recommended to create expert and advisory bodies that will allow receiving feedback on the compliance of products with the human rights standards. In addition, the key focus is development of guarantees and mechanisms of liability for violations;
- Recommendations for legislators and regulators. In particular, they are advised to pay attention to the development of a system of responsibility, certification mechanisms and codes of conduct, where relevant, to extend the requirements in the field of public procurement to the field of AI, to keep the decision-making system dependent on humans and not on automated systems, to increase the level of expertise of national regulators, and invest resources in the development of digital literacy.

European Ethical Charter on the Use of AI in the Judicial Systems and their Environment (Council of Europe). The Charter is a soft law document that essentially codifies five basic principles of AI application in the field of justice, and also highlights typical examples of systems use and (in)security generated by such application. What principles are we talking about?

- The principle of respect for fundamental rights: ensuring the compatibility of development and implementation of AI tools and services with fundamental rights in the field of justice;
- The principle of non-discrimination: prevention of the development or strengthening of any discrimination between individuals or groups of individuals;
- The principle of quality and security: regarding the processing of judicial decisions and data, using certified sources and intangible data with models developed on an interdisciplinary basis, in a secure technological environment;
- The principle of transparency, impartiality and fairness: making data processing methods are accessible and understandable, allowing external audit;
- The principle of user control of systems: ensures that users are informed participants and in control of their choices;
- Four categories are also proposed: the use of AI in the field of justice, which is encouraged by the authors of the Charter, use that require reservations and additional scientific research, and use that should be based on the strictest limitations.

Guidelines on the Responsible Implementation of AI in Journalism (Council of Europe). After long negotiations, where Ukraine's National Council for Television and Radio Broadcasting also took part, the Steering Committee on Media and Information Society of the Council of Europe agreed on the final text of the advisory document on the use of AI in the work of the media. Overall, the Guidelines outline a framework and provide media outlets w ith an algorithm for decision-making: do newsrooms really need to use AI systems in their work? Key questions that need answers before using such systems include:

- the decision to use AI must correspond to the mission of the media, be an editorial decision supported by a regular risk assessment and provide for a balance of the interests of all those affected by such a decision;
- distinguishing between systems developed by the media independently and systems that the media uses as part of publicly available technologies (in particular, chatbots such as ChatGPT, Bard/Gemini and others). So, the media must ensure the quality of data in AI systems, obtain data lawfully and comply with other requirements stipulated by the law regarding personal data, ensure editorial supervision of the quality of such systems operation (which should be a practical procedure rather than a 'tick in a box');
- it is critical to emphasize that traditional journalistic standards – such as balance, truthfulness, respect for private life and impartiality, as well as other ethical requirements – are also applicable to AI systems that the media has decided to use;
- when the AI systems are used, especially for the purpose of content generation, the media must label such artificially generated information in a format and manner understandable to the audience.

The document also establishes obligations for platforms and Internet intermediaries that provide space for the distribution of media materials. In addition, certain obligations – notably, regarding the development of legal norms and recommendations – rest on the states.

Declaration on the Manipulative Capabilities of Algorithmic Processes and Recommendation on the Human Rights Impacts of Algorithmic Systems (Council of Europe). The Declaration is a short document adopted in 2019 that highlights an array of problems associated with the use of algorithmic systems. Its main focus is raising awareness of risks, particularly in the field of public administration. For example, it mentions the problems such as the impact of algorithmic systems on communication between authorities and citizens; data protection in the context of the development of such systems; the impact on the autonomy and self-determination of a person' as well as the ability of technologies to influence the freedom of decision-making in economic,

social and other areas. The Declaration does not offer detailed proposals, using a general recommendation to take proportionate measures and monitor the risks that arise all the time.

The Recommendation is aimed primarily at the governments and their regulatory bodies, which should take effective legislative and practical measures to protect human rights from the negative impact of algorithmic systems. In fact, the document is a follow-up to the Declaration and a kind of proposal of an algorithm for solving the problems and challenges outlined therein. The Recommendation is quite generic and, being adopted in 2020, does not cover some of the current challenges that came up relatively recently. However, the document sets out key principles applicable to the AI and the obligation to respond in good faith to challenges in this field:

- Involvement of all stakeholders in the discussion of regulatory initiatives (in particular, the government should invite industry, civil society, representatives of the academia, etc.);
- Promoting the development of expert potential in public and private institutions that use algorithmic systems;
- Considering the impact of algorithmic systems on the media, the environment and other vulnerable areas;
- A statement of core principles, which include democratic participation, awareness, institutional frameworks, secure data management and system testing, as well as transparency, accountability and guarantees of legal remedies. Moreover, the security measures include the duty of constant supervision of such systems.

Recommendations of the Council on Artificial Intelligence (OECD). The Recommendations are one of the first documents to comprehensively address the topic of developing AI systems in accordance with ethical requirements and human rights standards, although they are focused on reformatting the ecosystem in society. Overall, the document is fairly short and structured, and is based on OECD research on the impact of AI on society. The Recommendations are divided into two parts – general principles, which aim at developers of AI systems, and recommendations for those who make legislative decisions and impact policy making:

- Five general principles include: sustainable development and well-being, justice and human-centered approach (notably, the rule of law and respect for human rights), transparency and well-reasoned decisions (understandable for users), reliability and security, and accountability;

- Five policy and regulatory recommendations include: investing in AI research and development, creating a digital ecosystem, creating an enabling legal environment for the development of AI, preparing for the transformation of the labor market and creating human capital capable of interacting with AI, international cooperation on creating a reliable AI.

The OECD recently updated its definition of the AI to make it harmonized with the provisions of the upcoming AI Act. The main change was about adapting the concept to generative AI – previously, the definition required a certain level of human involvement or contribution to the system's decision-making process. Currently, the OECD is communicating with the relevant EU bodies regarding amendments to the Artificial Intelligence Act to ensure that the new regulation is viable and compliant with standards in the field of human rights protection.

Recommendation on the Ethics of Artificial Intelligence (UNESCO). Aiming primarily at the public sector, the Recommendation outlines key principles and values that the governments should implement in national practice when developing approaches to regulation. It also includes recommendations for the private sector and for civil society (the latter are mentioned as an integral element of the policy development and hands-on application to AI technologies, especially those that perform functions in the public interest). The document is quite human-centric and focused on the protection of human rights and democratic values. Notably, key emphases are:

- Compliance with the principles of proportionality, do-no-harm, justice and non-discrimination, sustainable development, safety and security, data protection and human supervision of AI, transparency and validity of decisions, responsibility, as well as digital literacy and involvement of the broadest range of stakeholders;
- Ensuring an ethical assessment of the impact of AI systems and taking into account the results when making management decisions, developing a secure digital ecosystem, creating a legal framework and policies on data protection (taking into account AI systems and their impact on privacy), developing a methodology for evaluating the efficiency of AI policies;
- Ensuring international cooperation on the regulation of AI and development of ethical standards for the development and use of such systems;
- Monitoring, assessment and reduction of negative impact in thematic issues (gender, environment, culture, education and research, communication and information environment, economy and work, health and well-being).

Regulatory considerations on artificial intelligence for health (WHO). Quite a few organizations with a narrower profile have also begun to contribute to the

development of the AI standards. The documents they developed are mostly analytical in nature with small portions of recommendations for key stakeholders (developers, legislators and those who apply AI). Notably, the WHO developed the Regulatory considerations on AI for health, published in October 2023. The document contributes to general trend of developing both ethical and legal frameworks, focusing in detail on six key issues:

- Building trust: it is important to ensure transparency and documentation of the entire product lifecycle and track development processes;
- Risk management, in particular, the use of technologies as intended, human intervention, continuous learning (in terms of updates for meeting societal realities and needs), learning models and cyber security threats, should be considered comprehensively and with the aim of simplifying the models as much as possible;
- External data validation and a clear view of the intended use of AI helps ensure security and ease regulation;
- Ensuring data quality: by carefully evaluating previous versions of systems (learning from positive and negative experiences, evaluating causes of error, etc.), it is vital that systems do not reinforce biases and errors;
- Challenges related to important and complex regulations, such as GDPR in Europe and Health Insurance Portability and Accountability Act (HIPAA) in the US, should be assessed with an understanding of jurisdiction and consent requirements for privacy and data protection issues;
- Fostering collaboration between regulators, patients, healthcare professionals, industry, and government partners can help ensure that the AI systems meet appropriate standards throughout their lifecycle.

The Bletchley Declaration (AI Safety Summit). The document was developed and signed by the AI Safety Summit member states and focuses mainly on ensuring sustainable development, economic growth and innovation. 29 states (including Ukraine) agreed that the world is currently facing many challenges, risks and opportunities, especially from generative AI. At the same time, some criticized the content of the document for avoiding mentions of open AI systems, lack of focus on the development of strict regulation and too generic approach. What does the Declaration say?

- An international approach: The Declaration clearly highlights that the nature of AI systems means that their risks are "*best addressed through international cooperation*" – recognizing the cross-border nature of the technology and little benefit to countries taking a fully independent approach;

- Key streams: While being fairly generic, the Declaration highlights specific concerns about advanced applications affecting cybersecurity, biotechnology and disinformation, reflecting broader concerns about how the AI can be used by 'bad actors' to harm citizens or undermine democratic processes;
- Space for disagreement: in the framework of international cooperation, it is openly recognized that the "*approaches may differ depending on national circumstances and the applicable legal framework.*" Surely, it reflects many countries being nervous, because they rely too much on the EU regulatory trends through the upcoming Artificial Intelligence Act;
- Dependence on AI developers: the focus of the Declaration is a call for cooperation, transparency and accountability from private companies;
- Common standards: the Declaration calls for developing generally accepted principles of transparency, standards and testing, as well as for using capabilities of the public sector to manage and monitor the AI systems.

Paris Charter on AI (RSF & Others). The international NGO Reporters Without Borders, in partnership with representatives of civil society, journalists, media and experts on AI, drafted a list of principles for the use of AI technologies in media and journalism. The Charter contains ten key recommendations. Reporters Without Borders consider that four of them are key:

- The media should be guided by ethical principles when choosing technologies, including the application or development of AI systems;
- Human decisions must remain central to editorial decisions;
- The media should help society clearly distinguish between authentic and artificially generated content;
- The media must participate in global AI regulatory processes and defend the viability of journalism in negotiations with technology companies.

Equally important are the issues of responsibility that rests on the media, regardless of who created a certain material – a system or a real person. Another fundamental principle is to ensure pluralism of opinions when using content prioritization systems (notably, this is about search engines embedded in media websites).

II. Foreign regulation: to regulate or not to regulate – that is the question

The situation around national initiatives is somewhat different. First, most of them attempt to assess risks and opportunities for the market of their developers, and therefore mostly use soft rules and promote ethical standards. Second, national markets differ greatly – therefore, regulatory needs and requests are not uniform across national approaches. So, what is the position of the AI regulation champions – the governments that host most of the advanced representatives of the industry?

White Paper and Guidelines for Secure AI System Development (the United Kingdom). In 2020, the Information Commissioner's Office (ICO) developed guidance on the use of AI, having published the Guidance on the AI Auditing Framework, which outlines the risks associated with the impact of AI on rights and freedoms and suggests strategies to avoid or mitigate them. OCI later issued a Guide explaining decisions made with AI, as well as a Guide on AI and data protection. In addition, the government has also published the Standards of Algorithmic Transparency, a document which name is self-explanatory. However, these initiatives are generic and aimed at adapting current standards to the field of AI. The first fundamental documents were the White Paper and the Guidelines for Secure AI System Development – the soft law documents that defined key milestones for the future regulation and development of the industry.

The flexible and adaptive approach in use in the UK shows that the government is aware of the risks associated with AI, and at the same time the recognizes the lack of information for the development of comprehensive regulation. Since heated discussions currently take place around generative AI as part of the AI Act development process, it is quite possible that a wait-and-see approach will allow for the development of a regulatory framework that will be most conducive to the safe development of the technology. Meanwhile, research organizations like Ada Lovelace Institute emphasize that the lack of regulation is quite harmful in the medium term, because it makes the regulatory space unpredictable for the industry, creating risks for human rights. Well, what is the **White Paper** about?

- The five principles of AI are: safety, security and robustness; appropriate transparency and explainability; fairness; accountability and governance; contestability and redress, avoidance of the use of AI (objection against automation of processes);
- Empowering already existing bodies and expanding the scope of effective acts so that they cover the field of AI; support of already existing initiatives,

mainly in the form of financial programs from the government, but without strict regulation of the areas and types of proposed projects – that is, open tenders without creating a clear government request;

- Prioritizing the development of the industry and the voluntary implementation of the principles in development and application of AI, contextual application of these standards and adaptation to market conditions;
- Priority areas of activity: monitoring and evaluation of the effectiveness of the legal framework and implementation of principles, focusing on innovation support; assessment and monitoring of AI-related risks in the economy; horizon scanning and analysis of gaps and challenges; supporting test initiatives to help the developer bring new technologies to the market; quality education; promoting compatibility with international regulatory frameworks;
- Priority of participation in international AI development programs, forums for discussing challenges and priorities in the field of innovation, participation in development of international standards and coordination of the UK regulatory model with the international framework to open the market for foreign developers.

The Guidelines for Secure AI System Development are a slightly more technical document. It sets standards for secure system design, development, implementation, and maintenance. The document specifically targets AI developers. In a nutshell, the document is an explanation of how general standards (transparency, comprehensibility, technical security of physical infrastructure, etc.) work. It is a very practical guide on how to build a supply chain at all stages of the life cycle of an AI system (notably, in the context of data access and processing, its origin and quality, etc.). An important element is also the management of risks and incidents during the stage of implementation and support of the system.

Draft AI Bill of Rights and Biden's AI Executive Order (the US). First of all, when analyzing the US legal framework on innovation regulation, it should be noted that it varies significantly from state to state. Specialized regulations mostly have no federal effect. Electronic Privacy Information Center concluded that as of the beginning of August 2023, there were about 50 regulatory initiatives at the state level (including 30 documents that were registered as draft laws in 2023). Some of them are sectoral: they regulate AI and employment, protection of personal data, use of AI in the public sector, etc. At the same time, there are quite a few proposals regarding the regulation of generative AI, prevention of general harm from AI, and other conceptual matters. However, all local regulations will have to comply with federal legislation in one way or another. Here are two main initiatives that propose to regulate AI:

- **Draft Bill on AI Rights**. The document proposes to lay the foundation for the regulation of AI in the US, highlighting five key principles that should be followed by industry representatives: safety and reliability of systems, protection against discrimination by algorithms, protection of privacy, notification and explanation (regarding interaction with AI and its nature), human alternatives and the right to refuse. At the same time, the Draft Bill emphasizes the need to further adopt a number of laws for implementing these principles in practice in various areas.
- **Biden's AI Executive Order**. The huge document contains ideas for regulation of many areas, having a sectoral approach to the development of a regulatory framework. Notably, it covers new standards of safety and reliability (the requirement for developers of critical systems to share the results of security checks of the US government, protection against cyber fraud with the help of AI, etc.), privacy protection (prioritization and state support of AI creation technologies that are safe for privacy, development of recommendations to assess efficiency of data protection policies and practices by state authorities), anti-discrimination (focus on algorithmic bias and eliminating it, ensuring fairness in the justice system), protection of the rights of consumers, patients, students (ensuring responsible use of AI and developing the education sector), protection of labor rights (monitoring the impact of AI on the labor market), ensuring competition in the innovation market (openness and honesty of processes, promoting research in different parts of the state at the local level), responsible and effective use of AI by the government (hiring experts in the field of AI, providing training to staff, mapping the needs of various government structures). Overall, the document is a very detailed guide to the new developments that are to be expected in the US regulation.

Another critical initiative is the draft AI Disclosure Act, which proposes to label all content created with generative AI and basically calls for transparency about the technical makeup of the system, so users can understand how it works. It is noteworthy that this is not the first proposal of a kind – similar ideas have already been presented in the Senate.

Measures for Management of Generative AI Services, **Provision on the Administration of Deep Synthesis Information Services**, **Regulations on Management of Algorithmic Recommendations in Information Services**, **and Conclusions on Strengthening Ethics and Management in Science and Technology** **(China)**. In general, China has advanced a lot not only in the development of AI systems, but also in the regulation of this field. The most recent act, which was published just in midAugust 2023, is a law on generative AI – and it is the world's first hard law regulatory act in this field. However, in addition

to it, China still has a lot of developments in regulation of AI technologies. Let's find out how the legislation currently looks and how it differs from the European approach.

- **Measures for generative AI**. The document applies to the use of all generative AI technologies for the provision of services to individual clients in the People's Republic of China (the scope of law applies specifically to services on the territory of the state). The law is designed in response to the surge in popularity of AI-powered chatbots such as ChatGPT, and has an emphasis on generating text and training data. It requires providers to ensure that training data and content produced is '*true and accurate.*' It means that the act regulates the information sector, with some additional regulation in the field of data protection. It is important that China became the first country among those actively developing generative AI to regulate the chatbots. The law proposes to make the developers of AI systems responsible for the results of such technologies, and it is prohibited to provide such services anonymously (for example, it will be illegal to create and distribute an application with elements of generative AI without revealing the identity of the developer).
- **Provision on Internet services of deep synthesis**. This act provides regulation of so-called deepfakes – notably, the technologies that create any kind of synthetic media (including for entertainment purposes). The key novelty is the ban on the use of such technology to generate false news, as well as the requirement to label any artificially generated content (that is, an analogue of the labelling requirement, which now appears in every draft regulatory act). In addition, the act also contains requirements for monitoring the operation of systems, and also establishes the responsibility of developers in case of violations.
- **Regulations on algorithmic recommendations**. The main motivation behind the development of this regulation was the fear of strong impact of algorithms on content filtering and news prioritization. Key obligations of Internet service providers include the obligations not to use algorithms to violate the rights of others, as well as to create a body responsible for the safety of algorithms. In addition, the act essentially introduces a requirement for regular evaluations of the impact of algorithms on human rights, as well as protection against the use of algorithms to spread false news or manipulation. In addition, the act provides for enhanced protection of children's rights, labor rights and other industry norms.
- **Conclusions on ethics and management**. The document focuses on internal ethics and governance mechanisms to be used by researchers and technology developers, with AI identified as one of three areas of particular concern, alongside developments that directly affect life and

healthcare. Unlike other acts, this regulation is more technical. However, it is also an important indication that China is trying to regulate ethical issues at the legislative level (which is a kind of oxymoron). This is a wake-up call, just like many other human rights issues in China), because ethics and ethical standards are something that should mostly be left to self-regulatory bodies and industry rather than government regulation.

In addition to these instruments, there is a large array of other legal acts – mostly very thematic and narrow-profile in their nature. Key difference from the European approach is segmental regulation and the solution of specific problems rather than development of general rules of the game for all developers of AI systems.

However, even thematic acts in China are growing in number, and the number of areas under regulation is constantly growing, too. In addition to the acts on algorithmic recommendations in Internet services and deepfakes, there are many recommendations, for example, regarding the work of Internet intermediaries. In its study, Carnegie Endowment for International Peace also mentions a number of advisory documents that serve as a kind of strategy for AI developments, prioritization of areas and regulatory acts. Some of the most noteworthy include the regulation on the protection of personal information, which takes into account the existence of algorithmic systems and specific risks for personal data. In addition to regulation, the Chinese government also creates financial opportunities for startups and actively supports developers of technologies that are considered socially beneficial.

Industry self-regulation. There are quite a few initiatives to develop self-regulatory acts (mostly collective), or individual public statements by AI development companies regarding the standards applicable to their products. Surely, such claims often crash against the granite walls of reality since the procedure to prosecute violations is lacking, and it is often impossible to verify compliance with technical promises in practice. However, there are also positive examples – mainly from large corporations that have financial resources and a team big enough to invest in additional methods of protection, assessment of the impact of systems on human rights, etc. A good example of individual initiatives is the recent announcement by Microsoft regarding the launch of a system for labelling artificially generated content. Notably, the company demonstrates good faith in implementing AI systems and takes possible precautions to avoid abuse. Similarly, Google has published seven AI principles that it adheres to in the development and application of technology.

One of the most powerful collaborations aimed at the responsible use of AI is a non-commercial initiative called Partnership on AI. Within the framework of this coalition, research on the impact of AI systems on ethical principles and human

rights is conducted and recommendations are developed for the safe use of technology. Notably, it concerns the development of documents that codify the values promoted by the industry within the processes of creating AI technologies. Similarly, Google, Microsoft, OpenAI and Anthropic announced the establishment of a supervisory body that will check for violations of standards in the field of AI.

Also, the AI Act at the EU level leaves many issues specifically for self-regulation, including assessment of risks and threats and independent development of methodologies, determination of the risk status of the system, compliance with ethical principles, etc. In this vein, it will be interesting to see its final text and evaluate the confidence index in the industry against the backdrop of the latest political agreements between the European Parliament and the European Commission.

Other initiatives. Other ideas regarding regulation and development of standards (as well as attempts to implement them) appear regularly. Most of such moves showcase that key actors are not certain about the success of the proposed Council of Europe Convention or unable to strike an agreement on the AI Act (or uncertain about the applicability of its standards outside the EU). For example, not so long ago, 18 countries signed an international agreement on safe AI, which once again emphasizes the need to create safe and human rights-oriented AI systems. It does not offer specific requirements and obligations – however, even this step is yet another declaration of intentions to finally regulate AI and develop at least ethical requirements in this area. Another interesting initiative is the attempt to develop a system for labelling AI-generated content. So far, research centers of the European Parliament have certain progress in this field.

Last year, the UN created an AI Advisory Body, which consists of experts on AI, on human rights and on development of standards for new technologies. Its work is based on the principles of inclusion, respect for public interest and international law, involvement of all interested parties and special attention to data management (an integral part of AI management). The Advisory Body recently issued its first Governing AI for Humanity Report. Like most other fundamental recommendations in this field, the Report contains two key components – a list of principles in the field of AI governance and institutional functions (of developers, users and other actors related to this field).

As a brief overview of key developments in regulation and standardization of AI proves, most of the advisory documents at the international level are similar in terms of approaches and focus on transparency and ethics of AI systems. At the same time, national strategies, managerial decisions and declarations are focused more on creating a favorable environment for innovation – in particular, through a powerful industry lobby. As a result, most documents are either

framework instruments or sets of principles of a generic nature, or have not yet turned into a legally binding regulation, and when it will happen so is a highly relevant question.

III. Ukrainian standards: in-between ethics and law

What regulations are currently effective in Ukraine? In December 2020, the Concept of the Development of Artificial Intelligence in Ukraine entered into force (with minor additions regarding the prioritization of the fields of legislative regulation in 2021). It is a strategic document that prioritizes the directions of AI development and identifies priorities for the Ukrainian industry. It does not provide any standards for the safe creation and use of AI, as Ukrainian NGOs have repeatedly stated. This trend was causing concern back in 2020, when active development of technologies was stimulated without any red lines. The Action Plan did not change the situation. It was adopted for the implementation of the Concept and it sets tasks for several specialized state bodies – the Ministry of Health, the Ministry of Education, the Ministry of Culture, other ministries, including the MDT. The list of tasks mainly refers to possible priorities and areas for the implementation of AI technologies, but not to the development of a regulatory framework, the introduction of standards or security guarantees, including in the field of human rights protection.

The situation changed in 2023, when the MDT presented **the Roadmap for the AI Regulation in Ukraine**. This document was developed with participation of the Expert Committee on the Development of Artificial Intelligence in Ukraine – a platform for cooperation between the Ministry, the industry, academia and representatives of civil society. In general, the document is based on bottom-up approach – development of the standards starts at the industry level, with the gradual development of legislative regulation. This will allow national developers to better adapt to emerging regulation, and the government will be better able to understand how to implement obligations under the AI Act (which can create quite a few challenges for Ukraine due to the lack of access to most EU institutions). Unlike the aforementioned Concept, the Roadmap proposed specific mechanisms for the development of the industry, including the so-called 'crash test' regarding the protection of human rights. What ideas are proposed for implementation in the near future?

- **White book**. By the end of 2024, it is planned to prepare, discuss and publish a White Paper (following the British model), which will help to fit AI technologies into the existing framework of effective Ukrainian legislation at least until the development of comprehensive regulation. In particular,

the document aims to outline key approaches and make future regulation predictable for the industry. The White Paper will also draw heavily on the approaches of the AI Act, which will later be incorporated into national law as part of European integration commitments. In essence, the White Paper will serve as a more detailed guide and a mapping of regulatory perspectives. The first draft was presented during The AI State: Government Tech with Startups event, noting that the document is currently planned as an expanded version of the Roadmap with detailed measures that will be implemented by the government.

- **General and sectoral recommendations**. In 2023, it was planned to start working on general recommendations for the development and use of AI, which would outline a system of values and key principles for the creation of such systems, regardless of their scope of application or purpose. A little later, in 2024, they are going to develop sectoral recommendations as well – they will be more detailed and focused on specific social needs, challenges of certain industries (such as healthcare or law enforcement) and prospects for standards that will directly address such challenges. Also, the recommendations will provide briefly guidance on the creation of self-regulation tools and mechanisms. In practice, the sectoral Recommendations on the responsible use of AI in the media were the first to appear. They outline key approaches, ethical principles and values that the industry must adhere to (including the distinction between generated and authentic content, labelling of generated materials, transparency regarding the use of AI, etc.).
- **Regulatory sandbox**. As early in as March 2023, the MDT announced the launch of a 'regulatory sandbox' that will enable industry product tests in a protected closed environment on closed data sets, effectiveness check of new systems and compliance with standards in the field of human rights, etc. This will be done before the products enter the market. Inception phase for the project is planned for 2024, and the launch, for 2025. Work is currently underway on the technical component of the 'regulatory sandbox.' For this purpose, the MDT conducts a survey regarding the format and possibilities of a project, the needs of the industry, and the demand for this tool. Notably, due to limited capabilities, it is initially planned to only extend tests to the projects in the field of AI having significant social value.
- **AI labelling procedure**. Labelling includes opening up the processes and methods involved in data annotation, creation of software architecture, use of third-party components, and types of output data. It ensures compliance with the principle of transparency, which gives users more predictability in the quality of results, and standardization enables planned use of approaches to improve AI systems. Use of open data criteria for labelling

results is mandatory and enables establishing a balance between the need for transparency and confidentiality (preserving commercial secrets).

- **AI content labelling**. Content labelling helps establish the authenticity of content produced, helping to identify and mitigate potential misuse of material produced by systems with generative AI. Development of a system for marking the content produced by AI is important to ensure transparency of operation of these systems and their use (useful initiatives in this regard include Content Authenticity Initiative, C2PA, etc.).
- **Human rights impact assessments**. The road map proposes mechanisms for assessing the impact of AI systems on human rights, democracy and the rule of law. Among such tools, the key is the HUDERIA project – a methodology that helps identify risks in systems, assess their impact and develop a mechanism for mitigating negative consequences. The process involves engagement of various stakeholders for comprehensive identification of gaps and potential dangers. Currently, participation of Ukrainian representatives in the HUDERIA pilot project has already been agreed. Additionally, there are a number of other similar initiatives, including AIIA, Canvas, NIST AI Risk and many others that can be used to make such an assessment.
- **Development of codes of conduct**. Such documents are normally produced when the industry representatives unite into a self-regulatory body (or sectoral bodies) that serve as a platform for communication, discussion of values and principles of operation of the AI sector in Ukraine. Such voluntary commitments, participation in their development and their recognition will help identify *bona fide* players on the market. Moreover, it will become a platform for collaboration between companies ready to adhere to ethical standards (notably, public-private partnerships).
- **Development of comprehensive legislative regulation (until 2027)**. Considering the fact that the AI Act has not yet been adopted at the EU level, and the prospects remain rather vague, at the moment it is too early to discuss the development of national regulation of AI in Ukraine – notably, because regulatory acts will then have to be updated or completely changed if they differ from EU standards. Even after the adoption of the AI Act, many issues will still be there to prevent Ukraine from transposing it into national regulation. Like any document of a similar nature, the AI Act has quite a lot of references to EU institutions, which Ukraine does not yet have access to. In addition, similar to the Digital Services Act, there may be problems in the AI sector in terms of implementation of the Regulation even at the EU level, let alone the candidate countries. Therefore, there is no need to hurry with the development of national regulations, at least in order to avoid numerous amendments to already adopted laws. The

Roadmap envisages the development and adoption of comprehensive regulation in 4 years, when the system will be tested on the European market.

- **Other instruments** that will be developed at the national or international level (including bottom-up ones – that is, from industry to the state).

Innovation Development Strategy. In December 2023, the MDT presented a document that is going to outline the demands and priorities in the field of innovation until 2030. While Ukrainian innovations may seem overly futuristic, in real life this is not the case, as the Strategy provides a detailed analysis of foreign practices. Among them, Japan's Society 5.0 (large-scale digitization of public services), the AI startup ecosystem in Taiwan, the long-term innovation strategy in the UK (until 2050), and many other practices. The Diia Information Center also mentions a number of other services that_the MDT considers to be exemplary in the field of AI development: Patenttranslate (a patent translation service in 32 languages), Serenata.ai (a service for civic control of public expenditures), and Kaggle (a platform for competitions on AI model creation).

In the Ukraine's Strategy, AI was prioritized in many fields, including defense, education, health care and public administration. The Strategy goals include the need to develop regulation in the field of AI, to develop ethical and normative rules, as well as to encourage industry to work in priority areas. The regulatory sandbox, Government BI and GovTech AI Center of Excellence (the opening of which was announced at the beginning of January 2024) are named the three main projects in the field of AI. To implement the Strategy, it is proposed to restructure the administration in the following way

- the MDT will be responsible for overall coordination;
- the Innovation Development Council is empowered with managing discussions, preparation and approval of the Strategy;
- Deputy Ministers for Digital Transformation (CDTO) in ministries, for the development of innovative policies in various fields;
- CDTOs in the regions, for policy implementation in the regions;
- The State Agency for the Development of Innovations, for the implementation of the Strategy and investment raising.

The developed model (at the levels of both the Roadmap and the Strategy) is detailed, multi-layered and quite ambitious, especially in the short-term perspective. And it is important to note in this regard that the proposed measures are aimed at the private sector and the establishment of public-private partnership, with the development of ethical standards or voluntary testing of systems (crash

test for compliance with human rights) being an eventual outcome. However, many AI technologies are already actively used, and not all of them are low-risk – for example, entertainment-related. On the contrary, most technologies developed for the public sector have a significant impact on human rights. Therefore, we have to find out what are the current initiatives for the use of AI by the government and whether the sword of Damocles sometimes hangs above human rights due to the promotion of digitalization.

IV. Public administration: Diia in action

In July 2023, a survey by the Razumkov Center showed that only about 15% of Ukrainians actively use AI technologies. However, digitization is unrolling quickly and at many levels. One of its drivers is the digital reform of public administration sector. State bodies regularly announce the implementation of automated systems in decision-making processes, data analysis or provision of public services. Even Ukrainian diplomats began to use AI in their work, as recently stated by the Minister of Foreign Affairs Dmytro Kuleba during EquAllity Hackathon. The methods of applying AI in diplomacy are quite safe though: analysis of large data sets, development of communication strategies, analytics and search queries.

However, there are more complex initiatives in the field of public administration, where AI has more 'autonomy' and ability to affect the organisation of public services, mass data processing for the purposes of statistics, and the adoption and verification of other systems. In addition, in the avalanche of announcements from the MDT, one can find ideas of monitoring social networks and other public platforms, which may be of concern. What initiatives are we talking about and are they safe from the human rights perspective?

State-in-smartphone application Diia (*Action*). Although initially Diia had no AI elements – it did not make any independent decisions or a significant contribution to decisions made by civil servants – currenttly there is an idea to expand the capabilities of the application with the help of an additional **AI tool called Nadiia (*Hope*)**. According to the Minister of Digital Transformation Fedorov, AI in Diia will take the form of a virtual assistant that will advise users of the platform, explain technical aspects or, for example, advise a person where to find the nearest administrative service centre. The plan is to implement this project jointly with OpenAI and Microsoft. Negotiations with them about the integration of AI into Diia are currently underway.

That is, it is planned to develop a chatbot that will give correct answers to any questions related to public services, without errors and in easy and understandable

language. So far, according to the developers, it has not been possible to achieve this outcome – therefore, the system is being tested by a special team and is not available to the general public. At the same time, it is about the suboptimal efficiency of the technology as of now. Notably, Fedorov emphasized that from a security perspective, Nadiia will be very reliable, because the architecture of the system will not entail the use of personal data. The team working on the development of Nadiia notes that there are plans to develop an assistant that will predict a person's needs in certain types of services (in particular, public services) and proactively offer them. This resembles a kind of ads targeting of public services, and the key question is about the procedure to implement this idea, specifically – is it possible to do profiling and mapping of needs without analyzing the user's personal data (their location).

Another initiative within the Diia application is the **Diia Office project**, designed to digitize the work of ministries, make it transparent and ensure civic control over the performance of authorities. According to Fedorov, regular citizens will not have access to all the functions of the system available to civil servants. However, they will be able to monitor the distribution of competences and the responsible persons for the implementation of specific objectives, the terms of their performance, the allocation of funds, etc. At the same time, civil servants will be able to conduct surveys or propose ideas within their ministry/department, schedule meetings or communicate among themselves.

Another goal of this innovation is to introduce gamification into the work of officials. Notably, civil servants will be able to evaluate performance of their colleagues, receive an award in Diia Office, and monitor progress in meeting the goals of other departments. Currently, the project is at the stage of beta testing, and the legal regulation for this type of application is also not ready. And this can become a real issue, given how much effort legal changes in the field of civil service require.

State Statistics Service. In the Action Plan for the Implementation of the Concept for the Development of AI, a significant role is assigned to collaboration between the MDT and State Statistics Service for improving information processing and public administration evaluation. Notably, the MDT has launched a pilot project on digitization and capacity building of public authorities – Government BI (GBI). Currently, the project is under beta testing. After is successful completion, it is planned to scale it up to other authorities. The State Statistics Service was engaged in the development of the project – it provides data that is used to create the AI system.

In general, the initiative provides for the launch of analytical tools – including AI – to improve and speed up management decisions. The developers emphasize

that the GBI system will not replace human decisions, but will only help perform organizational tasks. The last stages of the project provide for the development of BigData structures that will ensure the development of the open data system with the participation of the State Statistics Service (the digitization of which goes on since 2021). This is an extremely important initiative, because analysts have long been requesting for the opening of statistical data for forecasting the economy (which is also supported by the Ministry of Finance).

In the second half of 2023, a new portal of the State Statistics Service was launched, including reorganization of the administrative structure, creation of an internal IT system and transfer of data for the past thirty years into digital form. It will also speed up data processing and strengthen the involvement of the State Statistics Service in public administration processes.

Voice of Citizens. Among other digitization and digitalization projects, an interesting initiative by the MDT is worth mentioning. It is called Voice of Citizens, and its ambitious goal is to scan everything that happens in social networks in real time. The idea is to collect information about discussions on forums, on Facebook, Instagram, Telegram and others, in order to identify social problems and needs: where services are poorly provided, what complaints citizens have, etc. Subsequently, the information is planned to be transferred to contact centers and communicated to relevant authorities.

When presenting the project, the MDT emphasized that Ukraine is currently "*probably the best country in the world where various technologies can be used*," and Voice of Citizens has the opportunity to bridge the gap between bureaucratic processes and reality. According to the announcements of the project in the media, the initiative will be integrated with Nadiia virtual assistant – this way, the citizens will be able to record voice messages with complaints and suggestions, and civil servants, to receive feedback in real time.

However, it is important to be aware of the difference between purposeful reporting problems to government agencies with the help of a virtual assistant (voice or text complaints) and indiscriminate monitoring of social networks (which are legally a non-state space). In the context of this project, monitoring, firstly, may involve the processing of personal data (since online complaints can often be very personalized), secondly, the implementation of such technologies requires separate legal regulation, because the impact this initiative on human rights is very significant. And although now it is too early to talk about technical risks until more information is known about the architecture of the project, conceptually, this system goes against the government's duty not to interfere in the freedom of expression.

V. The information sphere: on top of the news mountain

The use of AI in the information space is always high on the agenda: from the already mentioned ChatGPT which does not distinguish adequate information from Russian propaganda to the creation of deepfakes, with the videos of Zelenskyi, Zaluzhny and the scandal with mimicking Klitschko being prominent example. And it is prevalence of disinformation that prompts us to look for mechanisms to counteract manipulative, propagandistic and – at the same time – artificially generated content. Currently, public authorities are quite cautious about disinformation, and relevant ministries are trying to promote media and digital literacy (for example, the Ministry of Culture has a Filter project).

During the conference "Artificial Intelligence and Disinformation: Exposure of Digital Propaganda," Deputy Minister of Culture and Information Policy Taras Shevchenko emphasized that the National Information Security Strategy provides for a number of goals and tools, including the use of technologies to counter disinformation. At the same time, the Ministry currently does not have specific systems, applications or partnerships with tech companies established, at least in terms of the application of AI technologies or their development for countering Russian propaganda.

The Center for Strategic Communications and Information Security uses AI to monitor the media space and analyze online publications for disinformation. They use automated tools such as the SemanticForce and Attack Index platforms. The former analyzes the information space, including social networks, and identifies information trends. The latter applies machine learning, cluster analysis, and computational linguistics to detect malicious narratives, predict future information attacks, detect automated disinformation systems or coordinated campaigns, and categorize dangerous content. However, the Strategic Communications Center also does not have its own developments or projects, at least not yet.

At the same time, at the end of October 2023, the Ministry of Culture announced cooperation with Google in the field of combating disinformation and preserving cultural heritage. One of the planned areas of cooperation is to develop the AI technologies. Therefore, the lack of national AI developments in the information space is not a sentence, and on the contrary, collaboration with tech giants can provide more effective and safer measures for the protection of human rights. No updates on this topic since October 2023 have been observed, though. Therefore, the hope remains that at the operational level, this matter does not get stuck in bureaucratic shackles.

Also, the MDT reported that Ukrainian developers will create the Mantis Analytics platform to monitor and analyze manipulations in the information space. This includes analysing media, social networks and information platforms, and plotting the data on an interactive map, which is available publicly. Similar initiative exists within Brave1 project – a defense tech cluster which focuses on providing technological solutions to bring victory closer. It is critical that in contrast to Voice of Citizens, Brave1 is announced as a platform related to the field of defense. Therefore, there is a possibility that the monitoring of social networks will stop after the Ukrainian victory.

However, at the centralized/national level, there are still quite a few technologies used to counter disinformation and other information threats. For comparison, the US Department of Defense signed a contract with DeepMedia startup to use AI technologies for detecting deepfakes. What is important is that the US plan to use such opportunities for various purposes, including to detect Russian propaganda. The tools which DeepMedia plans to develop were trained on samples of 50 languages to identify if the content is real or artificially generated. These practices prove that disinformation is a serious concern, and that it is possible to engage third-party companies to develop software that can be useful and effective in countering information aggression.

VI. Educational sector: time to fulfill the Dream (Mriia)

Online education first knocked on the door with the onset of COVID-19, when instruction migrated to Zoom, Teams, and GoogleMeets. The results were not too encouraging, as the survey showcased a deteriorated education outcome, or at least the performance of students in schools. The situation became even more complicated during a full-scale invasion (especially the winter of 2022-2023), when limited access to the Internet, constant need to move between the classroom and the shelter, and the destruction of the schools caused a significant decline in the education outcomes. Moreover, it became necessary to combine instruction for students who went abroad and those who stayed on the territory of Ukraine.

In response to the challenges of the last four years, the Minister of Digital Transformation Fedorov announced the launch **of the digital assistant** called **Mriia** (***Dream***). It was presented on September 1, 2023. During the presentation, President Zelenskyi emphasized that the application is designed to help children unlock their potential and, based on recommendations, provide an algorithm which subjects should be prioritized, which educational tools to choose, which materials may be interesting. The main goal of Mriia is to analyze the place of study,

the qualifications of the teachers and the content that the child consumes for determining the trajectory of their further education. At the practical level, Mriia will allow to offer innovations, look for grants, sections for additional classes, etc. – that is, to provide necessary information on request (similar to the Nadiia project).

It was initially expected that the project will be launched in late autumn or early winter. However, in November, Fedorov presented the application interface and announced that the full version is planned for launch for 2024. It will enable access to school notebooks, teacher data, curricula, proposals on changes, etc.

It is noteworthy that a number of fakes about Mriia have been already disseminated. Ostap Stakhiv's video about the Mriia application driving children into a digital concentration camp – under the pretext that the AI determines their life priorities, behavior and future profession – has been widely reposted on the Internet. Stakhiv also noted that Mriia should allegedly establish the order of communication between parents and children. According to fact-checking by VoxCheck, no public source has ever indicated that the app will affect parent-child relationships in any way. Instead, it will provide an opportunity for parents to communicate with teachers in Mriia online. In general, especially in the circumstances of distance learning, the app can become a very convenient way of communication and improve the quality of education.

'Choose Your Dream Profession' project. The Ministry of Education and Science jointly with the Institute for the Modernization of the Content of Education and the Association of Innovative and Digital Education are implementing a pan-Ukrainian project called "Choose your dream profession." Notably, this project provides an opportunity to take an AI-based career guidance test that helps schoolchildren determine their interest in one or more professions. According to data from the end of July 2023 available on the Ministry of Education and Science website, as many as 100,000 education seekers have already taken the test. This project is one of many tools used to help choose a profession.

In addition to school education initiatives, there are plans to promote the development of AI in universities. Notably, among the ideas for the AI development, the option of creating research laboratories together with the Committee for the Development of Artificial Intelligence in Ukraine under the MDT and the Ministry of Education and Science is gaining momentum. One of these laboratories already exists at Ukrainian Catholic University – currently it is the first lab, and it operates in a test format. It is planned to develop a standard package of contracts between companies and universities to facilitate the formalization of cooperation. In addition, the MDT has developed a catalog of curricula for future students with information on training specialists in the field of AI – for example, on machine learning, data analysis, language processing, etc.

Also, the MDT develops **educational series** about AI that outline the features and challenges associated with the creation and use of automated systems. Some of the series are aimed at younger audiences, but there are also online courses aimed at representatives of the industry – for example, on human rights compliance when creating the latest technologies. Some of the most noteworthy courses also include a course on using AI for business development and personal branding. Its development was announced as another collaboration between the MDT and Google Ukraine. This is an extremely urgent matter, since the survey conducted by the Academy of Sciences of Ukraine jointly with Projector Creative & Tech Institute in December 2023 indicated that 70% students are already using AI. Therefore, the development of skills on safe and ethical use of technology is highly timely.

Another critical thing is that unfortunately, not all initiatives integrate safe and human rights-oriented AI. Lectures often cover the topics of how technology can be useful and how to use it most effectively. At the same time, it is very rarely about data protection or the larger impact on human rights. The collaboration with Google focuses on the business benefits of technology rather than system testing, crisis protocol development, or risk assessment. And this is one of challenges in developing training programs, because the issues of ethics and protection of human rights are key in the creation and application of systems. Therefore, exclusion of such topics can be very harmful in the future. At the same time, as of December 2023, at least 5% schoolchildren use AI to do homework and write papers, which does not always happen virtuously. The lack of regulations – at least at the level of universities (let alone the guidelines from the Ministry of Education and Science) – is definitely a poor contribution to ethical use of technology, especially if it is freely available.

VII. Healthcare: who keeps a finger on the pulse?

The Strategy for Development of Telemedicine in Ukraine, approved in July 2023, does not mention any AI systems or the prospects for their development in the healthcare sector. Sectoral legal framework does not provide any regulation, either. As a result, it happened often that the government's cooperation with private services or the development of own government initiatives was hampered due to the lack of standards and basic principles for the use of AI in the field of healthcare, at least in the Ukrainian legal field. For example, the government cooperated quite closely with Helsi.me, which had quite a few high-profile cases of illegal processing and use of personal data, which was in violation to the purpose of collecting such data. It goes without saying that most of the data was sensitive, because it is about health.

At the national level, elements of data processing by an automated system were used in the Central 103 project. It was about creating and maintaining an ecosystem that monitors the quality of emergency healthcare and speeds up the arrival of an ambulance. In fact, the system coordinated dispatch and other services, allowing to reduce the time for exchanging information. However, unfortunately, the website currently does not have available reports for past periods, and data on the status of network development has been reset. Most likely, this indicates a pause in the project.

Currently, there are two large projects that openly announced development and implementation of AI systems – BrainScan and System Carebits initiatives. Both projects were launched in 2023 with the support of international partners and under the coordination of the Ministry of Health. From a legal perspective, such systems do not exist, and therefore the final responsibility for technical errors still rests with the doctor. However, let us look outside the legal field to find out what practical features these systems have and what risks they can create.

BrainScan. The project was launched as a pilot in Odesa in early September 2023. Later, the program was also used in front-line Kramatorsk in Donetsk region. The Ministry of Health noted that the use of AI for the analysis of CT images of the brain in the prefrontal regions showed good results and high efficiency. This system speeds up diagnosing brain diseases or injuries, especially in cases where time is a critical factor. According to the program ToR, it provides opinion about brain activity to the doctor already 5 minutes after the start of the computer tomography analysis of the brain. This is done automatically. It is important that the final decision regarding the treatment protocol is made by the doctor, and therefore the AI does not perform any independent intervention in human health.

System Carebits. This program is an online telemedicine platform that enables remote diagnosis of fetal development in pregnant women with the help of portable devices. The device analyzes indicators and sends the results to the doctor's device (phone or computer). The program also helps doctors consult and communicate both with pregnant women and with colleagues. This initiative is especially relevant in view of the full-scale invasion when doctors often cannot be physically present during childbirth or during the period of pregnancy. In particular, there have already been successful cases of online counseling during childbirth in Mykolaiv region, when the woman in labor could not be taken to the hospital due to shelling and physical danger. Currently, about 180 health care institutions from all regions of Ukraine participate in the program, and doctors receive the training to use the system. The first devices were transferred to Ukraine as part of humanitarian aid along with an unlimited number of licenses to use the system. Similarly to BrainScan, final decisions are made by a doctor, so the AI is not able to independently intervene in human health.

To sum up, the application of AI in the field of healthcare at the national level is currently at a nascent stage, because not many projects are implemented with the government support and coordination. At the same time, the industry's efforts should not be underestimated – notably, a number of Ukrainian startups have a lot of projects aimed at facilitating the processes of healthcare, internal administration of hospitals, inter-agency coordination, etc. (Liki.24, Tabletki.ua, Doc.ua, etc.). However, appropriate legal standards for the development and application of such technologies for both public and private sectors are lacking.

VIII. Social sector: how to support a support system?

With the onset of a full-scale invasion, the need for and amount of social support increased multiple times. Surely, a lot of needs are covered with the resources from charitable organizations, international donors, foreign foundations and institutions, and even businesses. However, the burden on the government is still very high. And this is not only about financial resources and their limitations in view of Russia's active aggression. Administration of the processes of providing and distributing social support, prioritizing the most vulnerable population groups, assessing people's social status, etc. are also a challenge. Therefore, there was a strong need for digitization and automation of these processes.

'Modernization of the system of social support of the population of Ukraine' project. This comprehensive system aims to automatically identify prerequisites for violations of the law when receiving state social support. The project is implemented by the Ministry of Social Policy with the World Bank support. It was initially launched as a pilot in several regions. What does the system provide? An algorithm with AI elements analyzes data on violations when receiving social support and creates 'risk profiles' – a set of features of recipients who most often resort to fraud and abuse when receiving social support. At the initial stages, the method was tested in 10 social support departments, and the AI contribute is self-learning from its findings – that is, the algorithms are adjusted based on the available sample, statistical data and feedback on the results of its work – notably, when the users mark them as correct or incorrect.

To launch the system, it was necessary to collect an extremely large array of data from social inspectors, specialists of the Ministry of Social Policy, the outcomes of social payments, statistics of the Ministry of Finance, etc. Based on this information, it was possible to develop a profile of 'typical offenders.' Currently, little is known about the system, and the Ministry of Social Policy has not published the results of a test run. It raises questions about the effectiveness of the system and its impact on the rights of people whose requests for social support were rejected as they matched the 'risk profile.'

While during the discussions in the Ministry of Finance, the initiators of the system referred to foreign experience of applying social ranking mechanisms in the banking sector (in particular, in Romania and Moldova), it is worth noting that such technologies are on the border between high-risk and prohibited systems in accordance with the draft AI Act, especially if they are applied by the government. After all, the bank can use a person's profile exclusively for the purpose of concluding an agreement with them on opening an account, lending or buying securities, or providing other banking services. That is, the list of potential risks is comprehensive and clear. At the same time, providing the profiles of vulnerable and marginalized groups – the groups that most often rely on social support – to the government is somewhat dangerous. Eventually, with a change of government or deteriorated social problems, this may well turn into lists of persons for persecution or oppression, as is currently happening in China. Therefore, the majority of human rights defenders are actively opposing to the implementation of such systems even in the private sector, let alone government projects in this field. Similarly, sooner or later the question of compliance with the European standards will arise – notably, in the field of data protection and prohibition of discrimination against people on the basis of social status.

'National Staff Reserve' project. The press service of the State Employment Service reported that the initiative is represented by a social elevator that – with the help of AI – can help Ukrainians find a job or start-up a business. In particular, the system will be able to determine the range of professions suitable for a person depending on their abilities, qualifications and level of communication and management skills. Based on data analysis, an interactive CV will be created. It can be uploaded to the Unified job portal, used for hiring, or as an account of personal strengths and professional qualities. If a person is into entrepreneurship, the system will offer them to take online courses to improve such skills. At the same time, AI does not make decisions for a person and does not provide advisory functions, and the public description of the model indicates that there is a fairly low risk of possible biases, because the results will be generated on the basis of information provided by the person themselves.

Also, the government plans to add a case management tool to the **Unified Information System of the Social Sector**. Beta testing of the project is currently underway in four pilot areas. This initiative involves a coordinated approach to providing social support to vulnerable people and families. As part of the case management subsystem, electronic offices of recipients of social services, providers of such services and of case managers will be created. Each person who claims to receive support or already receives it will have their own profile that will enable assess situation of this person or family, as well as prevent vulnerabilities. The use of AI and algorithms has already enabled speeding up the processing

of applications from several weeks to 2-3 days. It gives hope that after the end of beta testing, the system will be even more efficient in both speed and accuracy of results.

To sum up, optimization of the provision of social support is necessary in the modern context, however, methods and proposals should be treated with extreme caution. Notably, we should not resort to technologies that can later be used for political pressure, oppression or persecution of vulnerable groups, especially if there is no supervision and effective national legal framework to regulate these technologies.

IX. Tax and customs: the fate of high-profile announcements

American developers from Salesforce noted that AI can help in building an effective tax system. They created an automated AI Economist system, which is based on the *reinforcement learning* approach. It teaches the system to rationally set the tax rate and evenly distribute taxes. It is assumed that the system remembers successful decisions and later uses them to make strategic decisions. As a result, it turned out that the AI managed to calculate the optimal tax rate under the given conditions, although from an economic perspective, the decision was considered rather extraordinary – high tax rates for employees with the highest and lowest levels of income and a low tax rate for employees with an average level of income. Are such systems viable and can they really be used in Ukraine?

In June 2023, the State Tax Service announced several priorities of digitalization at the forum "Key anti-corruption and institutional changes for the recovery of Ukraine." Generally, they covered the areas of big data, security of data, communication and services, and some administrative issues. At the same time, it was emphasized during the Forum that the collection of data on tax obligations, conducting inspections and investigations should not turn into a kind of state surveillance. That is, in this case, public authorities oppose profiling, in contrast to what happens in the social sector. However, unfortunately, no one has shared any further details regarding AI in fiscal sector, so it is quite possible that the project is currently on hold or at least its implementation is delayed.

At the same time, since 2019, the news columns have been abuzz with information that AI is being actively used by the State Tax Service to monitor agricultural land (with the help of satellite data) to check whether it is being used for its intended purpose – for example, if there is no illegal construction there. Such technologies are likely to be limited in use during times of war, but nonetheless the initiative

is an example of positive use of AI to optimize resources and prevent breaches. However, in the near future, we should expect an increase in the number of projects in the field of AI, since the news reports abroad have been extensively covering successes of systems in exposing tax fraud. For example, in France, 20,000 undeclared private swimming pools were discovered with the help of AI.

This practice is nothing new – foreign fiscal authorities have been using AI for quite some time. For example, the US Internal Revenue Service has announced that it will use AI tools to identify potential tax violations. Moreover, the technology is planned for use precisely to track the so-called 'complex' tax evasion schemes, which are mainly implemented by individuals with high incomes. In addition, the Ukraine's Bureau of Economic Security (BES) recently announced cooperation with the US Internal Revenue Service. Interestingly, as early as 2022, the BES created a working group to develop a Terms of Reference to create an AI system. Its mission would be to forecast risks and threats in the economic sector by looking for and analyzing relevant information. It is quite possible that Ukrainian-American cooperation may also relate to the development of AI technologies for financial monitoring.

Also, in April 2022, the **State Customs Service** obtained access to the AI HS Code Recommendation Platform from the World Customs Organization. This tool provides a list of recommended codes for a specific product. At the same time, the list is formed on the basis of the inputted product description and statistical data on customs clearance of goods with a similar description for previous periods. In fact, it enables automating the classification of goods and avoid situations where the lack of data in the database or suboptimal competence of employees complicates the process or causes incorrect inspection results. Therefore, training of employees of the State Customs Service is important so that they use the system correctly, because in the Ukrainian context, a critical problem is the lack of digital competences of the personnel of public authorities and services rather than no access to technologies.

Exchange of experience can be instrumental in this regard. For example, in May 2023, the conference "Innovations in public administration – working with large volumes of databases and artificial intelligence in tax and customs affairs" was held in Hungary. During the event, participants discussed the processes of digitalization of commercial document flow and implementation of AI for control functions regarding financial transactions. Among other things, the experts discussed the difficulties of using the Harmonized System for the classification of certain categories of goods and ways to solve these problems. Since the customs control issues are largely cross-border, it is important to bolster international cooperation and collective solutions to technical challenges.

X. Justice: looking for an electronic Themis

Let us start with foreign experience again. Some time ago, the COMPAS system was in active use in the US With the help of predictive analytics, it could predict the risks of committing crimes (such as the risk of recidivism) or identify the so-called 'hot spots' of crime. In practice, it turned out that the system strongly reflected all social prejudices and inequalities. For example, when racial or ethnic discrimination was widespread in a certain region and has historically manifested in court decisions, giving such court decisions to AI meant teaching it to act according to a similar algorithm. As a result, key advantages of automation – such as impartiality, accuracy and speed – were mostly nullified. In particular, the COMPAS performance indicator for predicting recidivism was about 65%, which is not much better than simple guessing. According to research by Dresel and Farid, a regular person having 20 times less information about the convict than the automated system was able to show the same level of result accuracy as COMPAS. Research by the Dnistrianskyi Center indicates that no AI has yet achieved accuracy rates of at least 90%, and even an 80% probability of a correct answer is extremely rare. What does this mean in practice? Every fifth (!) decision to reduce or not to reduce the term of imprisonment will potentially be wrong when it is made on the basis of AI data rather than on a contextual case-by-case analysis.

It is fair to admit that COMPAS is quite a radical technology that makes decisions on critical life matters. As foreign practice showcases, the justice sector is one of the most challenging and debatable in the context of the AI application. However, Ukraine currently develops extremely diverse projects and initiatives in this sector. Are they all safe and reliable?

Cassandra project – COMPAS in Ukrainian style. The Ministry of Justice has begun to use AI-based software with a quite romantic name Cassandra in test mode. In practice, they propose a system that – based on a questionnaire consisting of 97 questions – will determine an individual's propensity to repeat crimes (recidivism), which will later be integrated into a pre-trial report. Minister of Justice Denys Maliuska emphasized that in a few years of machine learning, Cassandra will have enough data to learn "*to analyze answers not only to simple questions, but also to analyze all other data that is available about the offender*." There are currently no data on the effectiveness of the system. This may indicate that the transition to the second stage of the project – when the AI goes beyond the questionnaire and analyzes the arrays of data comprehensively – has not yet taken place. One of the reasons for a prolonged test period may be a change in government priorities to those having more direct relation to the military sector.

Cassandra has already managed to receive several positive reviews from experts who believe that the automated system is capable of solving several problems facing the Ministry of Justice and the probation system as a whole. Notably, advantages of AI such as the ability to more comprehensively evaluate data sets and isolate cause-and-effect relationships, to keep records of prisoners and the needs of the government in terms of providing necessary resources to prisons, etc. were reported.

At the same time, the description of the system shows that it does not fundamentally differ from the COMPAS initiative (except for the fact that in Ukrainian society, the institutional discrimination is somewhat less prevalent than in the US). However, the system's ability to effectively determine the propensity of repeated offense, as well as prevention of other discrimination (for example, based on sex, age, place of residence or income level) still stays high on the agenda. Ukrainian activists and human rights defenders sound the alarm when it comes to Cassandra, because the risks of wrong results of the system can be very significant. In this case, the only positive factor is that the final decision is still made by an individual, not by AI – at least as long as Cassandra operates in the test version.

Court-in-Smartphone. The first announcements of the Court in a smartphone were made as early as 2021, when on the 30th anniversary of Ukraine's independence, President Zelenskyi announced the launch of this project as an attempt to combat abuses in the judicial sector. The initiative to introduce a full-fledged electronic court was supported by multiple stakeholders, including by the judges of the Supreme Court. For example, in her presentation on the digitalization of Ukrainian courts, Judge Olena Kibenko expressed her support for this idea because of numerous opportunities that online justice opens up. In general, the idea of a Court in a Smartphone is far from new, but the content of such initiatives varies depending on the state – for example, in China, blockchain is used to certify the authenticity of evidence, France has been actively talking about the introduction of AI into the justice system since 2018, and in Estonia, rumor had it that a 'robot judge' would be developed for consideration of minor cases (although the latter turned out to be false).

However, there is often a huge gap between an idea and its implementation. Notably, quick and effective launch of the electronic justice system became a serious challenge. Its quality was a subject to criticism by the State Judicial Administration (SJA). In fact, until legal framework was amended, the adoption of documents in electronic form was left to the discretion of every individual judge. In practice, neither side of the proceedings wanted to risk rejection or additional hassles. Therefore, the full-fledged Ukrainian reform regarding the Court in a Smartphone began with the adoption of amendments to the Law "On the Judiciary and the Status of Judges," which aimed to simplify the

procedure for accessing the court system, transfer many elements of the justice system to a digital format, as well as speed up document processing and dataset analysis. This was possible due the launch of the Unified Judicial Information and Communication System (UJICS). Its tasks are to:

- developing a single information space for the bodies of the justice system,
- inter-agency circulation and exchange of information,
- accelerated consideration of court cases and proceedings,
- work automation,
- transition to electronic versions of documents, digitization of court archives,
- fast access of UJICS users to information taking into account the access rights,
- confidentiality, integrity, availability of information in UJICS,
- harmonization of law enforcement by courts.

However, creation of UJICS is only the first step towards the application of AI in the judicial system. After the digitization of document circulation and of databases, the next initiative was to develop a system that can resolve minor cases without a judge. This idea emerged as part of the Action Plan for the Implementation of the Concept of Development of Artificial Intelligence in Ukraine, drafted by the MDT and approved by the High Council of Justice (HCJ). Back in 2022, there was a plan to launch a similar project in one of the courts of first instance, and it had to be about the AI considering court cases of administrative proceedings with a formal composition. According to the HCJ, in the long run it will enable identification of common mistakes, prevent corruption and harmonize case law in this area. After the test period, a gap analysis was planned to identify legislative changes that were necessary for adequate implementation of the project at the national level.

In practice, due to full-scale Russian aggression, the implementation of this project had to be postponed – the MDT, HCJ and the Supreme Court advertised the launch in the second or third quarter of 2023. However, no news regarding the progress, gaps, or the need for certain legislative changes related to the Court in a Smartphone have yet been publicly announced. Most likely, due to red tape, the trial period was prolonged once again. However, experts emphasize that the launch of this system is absolutely necessary in view of the martial law and increased danger of offline meetings, as well as limited access to courts in the frontline regions.

Discussions are still ongoing about the possibility to fully incorporate such novelties in the Ukrainian legislation. Although the European Convention on Human Rights (Article 6 – the right to a fair trial) and the European Ethical Charter on the Use of AI in Judicial Systems and Their Environment do not prohibit the use

of AI for the resolution of legal disputes, Article 127 of the Constitution of Ukraine indicates that judicial proceedings can only be performed by a judge (in certain circumstances – by jurors). And it is noteworthy that the Concept of Development of Artificial Intelligence envisages not only the development of uniform standards for accounting of court decisions and other data of proceedings, but also "*issuing court decisions in cases of minor complexity*." If this provision is interpreted literally, it probably does not correspond to the Constitution. Therefore, the Action Plan specifies that at the initial stages, only the previous practice in minor cases will be analyzed in order to develop optimal solutions and investigate trends. That is, it makes no sense to wait for the complete replacement of judges by AI systems in the near future. At least, unless amendments to the Constitution are made.

In addition, in a recent interview, Deputy Minister of Digital Transformation Oleksandr Borniakov noted that there are plans to strengthen Ukrainian courts with the help of Clearview AI, which is quite an alarm given the reputation of this company. On the other hand, Borniakov mentioned the success of several Ukrainian initiatives, including the **Court-on-the-Palm-of-Your-Hand.** This project was developed by an interest group with financial support from international donors, and now it is actively used in the public sector as well – for example, local courts recommend it for preparing for court proceedings and even distribute instructions for use. At its core, Court in the Palm of Your Hand is a combination of registers – the register of court decisions and 14 other thematically related databases that use search words and 39 other filters to find necessary information, which is available in the format of open data. The project has two versions: simplified (free of charge) and advanced (paid, with all relevant filters). At the same time, access for journalists is free of charge.

Subsequently, as part of the project, an additional module of automatic analysis was made. Its name is WINCOURT. It can evaluate the similarity of the documents that the user uploads to earlier documents that were used to make judicial decisions in similar cases, and provides a forecast regarding the success of the court proceedings. At first, the system's effectiveness was mediocre. However, the system is regularly updated and supplemented with new solutions. As a result, users are able to test their defense strategy in court. For commercial litigation, there is a service with similar functions called Verdictum PRO.

It is ironic that at a time when the AI is being actively integrated in the judiciary and the field of justice in general, the very technology is becoming the subject of lawsuits. In addition to reasonable warnings about high risks of the proposed technologies, there are also more down-to-earth challenges. Notably, experts emphasize the need to properly digitize Ukrainian courts first – for example, at least to provide all local courts with access to a stable Internet – and only then to introduce high technologies. Given the power outages, regular air strikes and

frequent inability of judges to conduct offline hearings, this request is more than timely. Addressing this issue is partially possible through the adoption of draft law #8358, which will enable remote work of judges and online proceedings. However, the draft law remains under consideration since January 2023. Without a developed digital infrastructure and legal regulation, any project – even the safest one such as Court in a Smartphone – will be no more than an idea.

XI. About the law enforcement sector, or where military technologies migrate

In the poll by Kantar Ukraine as of late 2023, 73% Ukrainians indicated that AI can make life much easier. People see its greatest potential in the areas of production optimization (54%) and the fight against corruption (51%). At the same time, share of those who support the idea of applying AI in the security sector – notably, in the work of law enforcement agencies – is also quite high and stands at 38%. And it is rather ironic that notwithstanding the survey findings, the use of AI systems to maintain public order and investigate crimes is already underway. And started much earlier than most people think it did. What is happening to AI technologies in law enforcement?

Video surveillance cameras. Since 2019, human rights defenders have sounded the alarm about the use of video surveillance cameras with face recognition. Notably, as part of the Safe City program implemented by local self-government bodies, most regional centers were equipped with such cameras. At the same time, the Law "On Local Self-Government" does not grant municipal authorities any powers in the field of monitoring of or working with biometric data to maintain public order. Some local councils even managed to adopt by-laws, such as Regulation on the use and operation of the video surveillance system of the city of Zaporizhia. However, even such by-laws do not provide legal grounds and appropriate powers to local self-government bodies. When the discourse on the use of AI by the police was just starting to top the agenda, human rights organizations were already actively criticizing this modality, insisting on the need to at least develop a system of safeguards against abuse.

However, in 2021, the Ministry of Internal Affairs (MIA) published on its website the Safe Country project, which was a scale up of the Safe City initiative. It is noteworthy that during the project presentation, no single word was said about legislative changes or the need to expand the powers of law enforcement agencies. The public speeches of the Deputy Minister Ihor Bondarenko at the beginning of 2022 also did not clarify the legal framework for implementation of this system. Similar initiatives were already proposed in 2019, at least in the form of draft laws that try

to define the powers of state bodies. At the same time, the expansion of powers provided the police and other law enforcement agencies with unlimited access to outdoor cameras, as well as other information and communication systems in general. This move faced considerable criticism from the Main Research and Expert Department of the Parliament – for the same reason that in 2019 and 2021 the initiative did not gain significant support.

A new iteration of this discussion may start at any moment. In fact, in early 2024, media headlines actively highlight the idea of a single video surveillance platform. Apart from the very fact of a complex surveillance system, the motivation that prompted the MIA to take this is also noteworthy. As it turned out from the Schemes investigation, thousands of cameras installed throughout Ukraine use Russian TRASSIR software, and before reaching the consumer's phone or computer, video from these cameras ends up on servers in Moscow owned by companies connected to FSB. These cameras were actively purchased by both private and state enterprises, and the restriction on the use of this technology was only made on February 27, 2022 – that is, after the start of full-scale invasion. The government entities and state-owned enterprises that used the TRASSIR software included Chornobyl NPP, Poltava City Council, and the Sea Ports Administration. It is important that at the time when the Schemes journalists interviewed the Deputy Mayor of Poltava, the Safe City system was effective, and nobody was sure if the software has been changed.

In addition, in 2023, the AziGuard software from the Romanian company AziTrend entered the Ukrainian market. This software consists of TRASSIR but it is sold as a Romanian product. Therefore, the danger of using technical equipment from the aggressor country has not disappeared. Moreover, if currently some state bodies still use software of Russian origin, the introduction of a single video surveillance platform can be dangerous.

In Zakarpattia, AI-supported cameras are used to identify people who enter the territory of the region. According to the Head of Zakarpattia Oblast Military Administration, "*the system monitors and updates the law enforcement officers about unjustified presence of people near infrastructure facilities, and respective measures are taken – detention or questioning of citizens.*" Such measures were actively used in May 2022. Then the persons were checked if their ID photos and car plates are available in the system, and if not, the police performed additional control of them. That is, unlike the usual situation of a database of wanted persons, in this situation, there was a general database of all residents of the region. That is, police were keeping huge arrays of sensitive data. It is not known if the system is being used now, but given that the security threats in Zakarpattia oblast have now decreased – and will further decreased after the end of martial law – use of this system would clearly be disproportionate.

As early as 2021, there was a plan to install the Vezha system in Vinnytsia, which was able to recognize faces, determine the parameters of people for further identification, as well as read license plates on cars moving even at a speed of up to 250 km/h. Since the announcement, no updates were made on this matter, but at the end of 2023, a new video surveillance system was announced to be implemented in Vinnytsia. Although it takes place within the framework of a project on a unified video surveillance platform, the initiative may be a follow-up to the Vezha project. It is important to understand if the technologies that will become part of the overall integrated system are reliable, because until February 2022, municipal authorities were not concerned about it. Also, it would be necessary to identify the scope of legal powers and practical opportunities that law enforcement agencies will be granted after the system is launched.

Clearview AI. The launch of this facial recognition system is related to the beginning of the full-scale invasion. The idea was to use it for identifying the dead Russians in order to further establish those responsible for war crimes. However, in practice, the technology migrated to the non-military sectors, too – for example, the developer provided access to this service to the MIA. In an interview in April 2023, the Minister of Digital Transformation Fedorov noted that Clearview AI services have already been implemented in the activities of 14 government agencies (Clearview AI currently mentions 18 agencies on its website). According to its long-term plans, the US company announced the idea of introducing innovations in Ukrainian customs and banking sectors. Also, it plans to open its office in Kyiv, but so far no progress has been made in this field. And that's probably for the better.

How is Clearview AI currently used Ukraine? Purposes mentioned include family reunification, refutation of false posts in social networks, increased security at checkpoints (identification of persons at checkpoints), identification of dead soldiers and detection of Russian spies. Statistics as of November 2023 indicate that more than 350,000 searches were performed in system. Media interviews mention that at least 230,000 of those searches were for fallen soldiers. That is, about 100,000 requests involved analysis of personal data of living persons. The MDT repeatedly noted the effectiveness of Clearview AI technology and speeding up processes, which enabled redistributing resources for more important military tasks. If the technology is so effective, what is the problem about it?

- Goals and accuracy. In an interview to BBC, human rights activists and security experts emphasize that when Clearview AI is used as the main and only basis for making life-and-death decisions, it is clearly a wrong move. It is critical that the law enforcement agencies that use the technology are also aware of this. Unfortunately, there are currently no legal (or even by-law) restrictions. So, in theory, detaining a person or even using a weapon

can sometimes be motivated by the outcome of the system. At the same time, the accuracy of outcomes is far from 100%. Notably, information about errors in the identification of individuals using Clearview AI was already reported in the news. Having no confidence in the system's reliability, it should definitely not be used in risky contexts, when an emotional reaction to the system's performance, fear or other factors can be a driver for fatal consequences.

- Violation of legislation. Clearview AI has been repeatedly criticized for illegal collection of personal data – creation of a database based on information collected from social networks, including private pages and images, as rea-life practice proves. Illegal practices in the field of data protection have become not only the subject of theoretical debate, but also the reason for numerous fines (at least in five countries). For example, Clearview AI was already made to pay EUR 20 million to the Italian regulator for illegal use of personal data. Similar disputes arose with the French regulator, where the company has already been fined several times. The main reason for fines is illegal data scraping – that is, the collection of large volumes of user data without their consent, and often even without notification of data processing.
- Technical security. Given that the company does not clearly disclose the principles how the database is created, it is quite likely that the information analyzed in Ukrainian context becomes part of the general database. Although today Clearview AI actively declares that it will not cooperate with any actor that resorts to violations of human rights (for example, Russia), no one can guarantee with absolute certainty that the technology will not fall into the hands of the occupying power. Moreover, there are no guarantees that after another fine for non-compliance with data protection rules, the company will not start cooperation with countries that have higher indices of human rights violations. In this case, it is not even about Russia, but about possible transfer of data to China, Iran or other authoritarian countries.
- Reputational risks for Ukraine. It is no secret that Clearview AI's reputation is far from the best. The company, having many fines and complaints in national jurisdictions, is trying to use the war in Ukraine as a way to whitewash its reputation. In particular, Clearview AI regularly emphasizes that the system is effective in the context of armed conflict, without actually responding to criticism – such as that the company is engaged in illegal data collection. Ukraine also regularly emphasizes the benefits of the system (for example, for identifying members of illegal armed groups in Crimea), ignoring critical questions about the legality of the technology.
- European integration. Given that the draft AI Act updated after the political agreement bans data scraping, Clearview AI will most likely end up on the

list of banned AI systems. Accordingly, the fact that Ukraine uses the services of this company can significantly affect European integration efforts, and sooner or later it will have to make a choice – to use effective yet illegal technologies in violation of European integration requirements, or to look for alternatives that are more in line with human rights.

Among multiple problems, key challenges are the lack of legal regulation of such systems in Ukraine and lack of any guarantees that the Clearview AI software will not migrate somewhere else tomorrow – for example, to street video surveillance cameras or DIIA application. Considering the number of risks, widespread use of illegally created technologies cannot be justified by public necessity.

Investigation. Security and law enforcement bodies began to actively develop or participate in development of AI systems in order to more effectively monitor compliance with legislation, track violations and investigate crimes. Currently, most AI systems are planned for use in the field of economic crimes, because it means an easier data analysis procedure compared to, for example, crimes against life and health of a person. However, it is quite likely that such technologies will be scaled up in the next few years, given Ukraine's active international cooperation and the exchange of experience with the countries where AI often plays a key role in investigations. What are Ukrainian law enforcement agencies using now?

- **Security Service of Ukraine (SSU)**. At iForum, the Head of the Cyber Security Department of the SSU Illia Vitiuk talked about the initiative to start using AI. Notably, this initiative includes development of SSU own neural networks, the recognition of enemy equipment (through the analysis of images and videos from surveillance cameras), as well as the work in information space (such as monitoring open resources to counter disinformation). Details regarding the scale of the use of technologies are currently not being made public. However, it is important that monitoring initiatives do not go beyond martial law and do not become a part of security services routine after Ukraine's victory.
- **National Police of Ukraine (NPU)**. The police forces around the world are extensively using AI to track down criminals, and Ukraine is no exception. Back in 2017, the Department of Patrol Police created an aerial reconnaissance unit equipped with the cutting-edge models of drones. Notably, they were used to detect illegal poppy and hemp crops, deforestation and plantations, places of illegal amber and coal mining, etc. In addition, after Russia's full-scale invasion, NPU started using AI to identify people at checkpoints – both facial recognition using systems like Clearview AI and systems for document verification, database searches, etc.

Also, AI is being used to find missing persons, including children. Notably, on the initiative of the Commissioner for Persons Missing Under Special Circumstances, Ukraine is trying to find persons who ended up on the territory of Russia as a result of hostilities or because of the war crimes – for example, forced resettlement. For this, Reunite Ukraine app is used. It was developed in cooperation with the US company Find My Parent. The app enables creating a personal profile and a profile of the person you are looking for. The system then compares the data of various completed profiles and, when matches are established, people can exchange messages. The more information the system collects, the greater the chances of finding lost relatives. Every person can also register in the app and submit information about Ukrainian children who are in Russia or in the occupied territory.

- **National Anti-Corruption Bureau of Ukraine (NABU), Specialized Anti-Corruption Prosecutors' Office (SAPO)**. As part of the National Informatization Program, the plan is to strengthen the capacity of anti-corruption bodies to process large data sets. NABU Director Semen Kryvonos reported that the Agency plans to use AI during the investigation of criminal cases to analyze data sets and track relationships – notably, based on the analysis of traffic, power of attorney and registration, facial recognition and transcribing audio into text messages. Kryvonos clarified that the system is capable of self-learning and will be developed by a private company. Although specific details were not announced, according to the description, the Clearview AI software has such capabilities, among other things. Given that Ukrainian state authorities are eager to cooperate with this 'tycoon of personal data,' it's time to recall unpleasant feelings about privacy violations. The only thing that makes us hope for the best is that the system will be used under the control of operators. But will it really reduce the risks rather than lead to targeted attacks on specific entrepreneurs, activists, media representatives or other vulnerable groups? There are currently no guarantees, even on paper.
  In addition, NABU and SAPO are currently using the eCase Management System for exchange of documents with the High Anti-Corruption Court (HACC) and internal business processes (such as setting deadlines, creating a calendar of tasks, operational data exchange, etc.). For example, by the end of 2023, 70 procedural document templates are available in the system, which removes almost 90% of the paperwork that detectives had to do. The system operates on the basis of a joint order of the NABU, the Prosecutor General's Office, the Council of Judges and the HACC, which in 2021 approved the Regulations on the Information and Telecommunication System of Pretrial Investigation. After the test period, they decided to develop a regulation that clearly regulates the procedure for using the system, conditions for access, and its capabilities.

- **National Agency for the Prevention of Corruption (NAPC)**. In December 2023, the NAPC announced the use of new automated tools for verifying declarations. It is about comparing the data in the declaration with other registers, checking data against certain formulas (for example, regarding the signs of illegal enrichment, unjustified assets, etc.). Only declarations with the lowest risk rating will be automatically verified, and declarants will be notified of this verification. In this way, it is planned to verify 30% of the total number of declarations. In general, automation is a logical step, because the number of declarations that need to be verified has increased after the adoption of amendments to the anti-corruption legislation.
  In addition, the NAPC uses AI to monitor political advertising on the Internet, which includes analysis of textual information using automated systems, among other things. However, there are currently no details on this initiative.
- **Bureau of Economic Security (BES).** As already mentioned earlier, the BES actively cooperates with the US Internal Revenue Service to detect tax crimes. In addition, the BES is currently working on the creation of a neural network based on AI, which will search and analyze information and predict risks in economic sector. According to the BES, this will help investigate economic crimes without human factor and thus reduce risks of corruption and minimize the shadow sector. A large working group was set up for project development, featuring representatives of the Prosecutor General's Office (PGO), the Ministry of Internal Affairs, the SSU, the State Tax Service, and even the Armed Forces of Ukraine. The only question is the scope of monitoring that can be performed through this system and the availability of safeguards against abuse and excessive interference in private life of individuals and business activities.

To sum up, there are quite a few initiatives to use AI in the field of law enforcement, but most of the systems proposed for use are described in very general terms. It often makes it impossible to establish how intrusive they are or may become for human rights. Lack of transparency is one of the biggest problems, including for discussing technologies used by law enforcement agencies, with representatives of academic institutions, civil society, and the public at large. In other cases, such as facial recognition cameras or Clearview AI, the problem is about high risk or outright illegality of the systems. Unfortunately, effective Ukrainian legislation does not provide sufficient safeguards against abuse or potential violations, as well as a mechanism to challenge the harm that such systems may cause. Moreover, in many cases the government has access to alternative mechanisms that are not that intrusive, but prefers mechanisms that are easier to implement.

XII. Defense sector: industry as a united front

The Head of the Security Service of the United Kingdom (MI5) Ken McCallum and his colleagues from the United States, Canada, Australia and New Zealand emphasize the threats from AI, especially in terms of the race of developments in the military sector. For example, the Israel Defense Forces actively use AI to plan airstrikes and logistical support of the army. So far, such technologies are used with human supervision and authorization of each critical action. But how long will it last, considering the dynamics of technology development in modern conflicts?

The ideas to use AI systems in Ukrainian defense sector were discussed long before the full-scale invasion. Notably, in August 2021, Ukroboronprom and the MDT signed a memorandum on digitalization of the defense sector, including with the help of AI. Among the priorities, cyber security and development of new types of weapons, such as AI drones, were discussed. These novels have become more relevant with the new round of Russian aggression in Ukraine in February 2022.

DOU recently made a large guide of companies developing military technology, finding that there are currently about 46 of them. Many of the initiatives have guided AI systems. Since the full-scale invasion, the line between private developers of military technologies and the government initiatives has blurred, because all technical innovations one way or another end up in the units of the General Staff of the Armed Forces of Ukraine, the National Guard or the Ministry of Defense of Ukraine (MoD). Therefore, in this analytical material, we will consider projects that are supported, coordinated, financed or directly implemented by the government. Among the priorities of the government, Minister Fedorov highlighted recording of the movement of equipment and enemy personnel, shooting down of missiles and effective targeting of drones.

Clearview AI. As already mentioned in the analysis of AI use by law enforcement agencies, the main purpose of using Clearview AI in Ukraine was an attempt to find solutions to military challenges. Such challenges include identification of dead Russian soldiers (for further notification of their relatives) and of Russian military personnel who committed war crimes on the territory of Ukraine (for bringing them to justice). Although both the recognition of faces of deceased persons and the search for war criminals can justify the use of such technologies, many questions about the company's reputation, reliability and further use of services remain unanswered. Moreover, there should be strong guarantees that Clearview AI will not be used to make life-and-death decisions without additional verification and involvement of human factor.

At the same time, even face recognition at checkpoints using Clearview AI will no longer be proportional in all cases. When it comes to using similar measures at the borders between, for example, Vinnytsia and Khmelnytskyi regions where subversive groups are unlikely to be expected in the third year of a full-scale war, it is highly critical to look for less intrusive and more appropriate technologies for human rights.

Even when it comes to using the company's services for the purposes of gaining an advantage in an armed conflict, a clear distinction should be made between military use of AI and its expansion to the civilian sector. The latter option, analyzed in detail in the previous section, is clearly undesirable. While the former has alternatives that could have been tested and launched in two years of full-scale Russian aggression. For example, the armed forces has already started using Primer (for gathering and streaming audio, de-noising, translating and extracting key statements that relate to information on the battlefield) and Scale AI (image analysis). Moreover, Ukrainian developers YouControl and Artellence created the TyKhto system, which enables checking passports and personal IDs of individuals and comparing them with Myrotvorets database. A combination of such services may well replace the inherently illegal services of Clearview AI.

Systems of situational awareness. Currently, non-combat automated technologies such as situational awareness systems are in active use. Armed forces use quite a few of them – notably, Vegvisir Core system was announced not so long ago, and the specialized NGO Aerorozvidka has many projects in this area. At the same time, the two largest and most extensive projects in this field are Palantir and Delta. Their more targeted analogs are Kropyva, Dzvin (in relation to which there were accusations of money laundering), Virazh Planshet, Bronia, GisArta and others. Let us find out what their features are and if they have risks similar to those of Clearview AI, and if so, are there adequate safeguards against misuse and potential harm?

- **Palantir.** At the beginning of the full-scale invasion, a leading IT company – one of the largest providers of software and cloud solutions in Ukraine – announced a partnership with the MDT. Their main product is Palantir Edge AI, a modular system that contains different data sets and can be configured for different tasks. In practice, it is often compared to Lego for process automation. Currently, the most common application of the system is terrain visualization, or transferring data from satellites and other sources to the map. One of the latest developments of the company is Skykit – a mobile intelligence center that is already actively used on the battlefield. At the same time, system developers emphasize that they are not data providers, as they only offer technical tools and analytics. In fact, the system collects information from commercial providers/satellites and

combines data from various sources. The system proved effective during the de-occupation of Kherson oblast. It is important that the system receives data legally, and the amount of received and filtered information can be adapted to the request – that is, the number of sources varies depending on the purpose and time intervals.

The system is also used to expose war crimes. This application and expansion of capabilities actually took place at the request of the PGO, which announced the need to use technology to systematize and analyze evidence of war crimes committed by the Russians.

The government plans to continue cooperation with Palantir in the field of reconstruction of Ukraine and digitalization. Notably, the MDT signed a memorandum on further cooperation with the company, and Palantir is already working to create products in the field of education, demining and other areas. For example, for demining, the platform integrated 82 datasets combining 6 million buildings, 60,000 trains and 1 million road segments. With the help of satellite images, recent activity in these territories and the level of damage is established, which helps in prioritizing the demining efforts. The company already has established an office in Ukraine, and recently announced a project on the development of AI to optimize product procurement and stock replenishment. Also, the company actively cooperates with Ukrainian developers.

- **Delta.** Ukrainian analogue of Palantir, which allows for real-time analysis of combat data, integrates information about the enemy from various sensors and sources, and can work on any device – even a mobile phone. In fact, the system transfers information to cloud storage, and in the future will also support critical servers. It is important that the system developed by the Center for Innovation and Development of Defense Technologies of the Ministry of Defense of Ukraine is built according to NATO standards, which was certified in June 2023. In addition, two chatbots of the MDT and SSU – eVorog and STOP Russian War, respectively – are integrated into Delta. It is noted that the system is ready to integrate even F-16 fighters. Some information in the system is verified manually, and access to certain layers of the system can also be limited to certain groups of people. It is also noted that at Delta, AI is constantly learning and improving the system.

  In general, Delta is considered one of the most reliable systems in Ukraine in the military sector, and unlike Palantir, it is only used for military purposes. However, the Russians managed to hack two accounts with the help of phishing – that is, the weakness was about digital skills of users rather than technical protection of the system. The situation was quickly corrected by removing access for these accounts.

That is, both systems – as well as their aforementioned sectoral analogues – are cloud-based and allow information to be operated independently of servers.

In addition, they show a fairly high level of technical protection, and in the case of Delta, also a clear military purpose. Moreover, unlike Clearview AI, the systems legally operate with data and process sensitive information.

Army of drones. The government program provides for the systematic purchase, repair and replacement of drones. As part of the program, about 2,000 drones equipped with AI have already been delivered to the frontline. Their main functions include safe reconnaissance, adjusting artillery fire and finding even well-disguised opponents. Also, the AI system enables stabilizing drones and maintaining a pre-selected target. The number of national projects in this field is constantly increasing after the Resolution of the Cabinet of Ministers of Ukraine, which stimulates the production and purchase of drones, as well as after UAH 40 billions of investments in Ukrainian production.

One example of such AI drones is SAKER SCOUT system, which is able to independently recognize and record the coordinates of enemy equipment, transmitting information to the command post in real time. The developers assure that the system can also work completely autonomously, distinguishing 64 types of targets. The project develops both scout drones and 'kamikaze' drones.

However, one of the most critical challenges is adequate supervision of the activities of AI systems, especially in the military sector. Currently, a discussion is unfolding regarding Russia's use of drones with fully autonomous guidance systems. The possibility of misidentification of targets attracts much criticism. Even Palantir CEO Alex Karp notes that there are "huge ethical issues" on the battlefield, which often involves a dilemma between properly protecting the state and trying not to lose a military advantage. Should Ukraine respond with similar technologies, aiming for a kind of status quo in the arms race? Currently, experts are concerned that even the use of anti-tank autonomous drones is illegal, as it ultimately leads to human casualties. Therefore, international organizations call not to abandon human-in-the-loop for the sake of quickly gaining a military advantage. However, finding an alternative that will be effective enough to outweigh the use of autonomous drones by the Russian side remains an issue which may turn into a real challenge.

Brave1. It is the government cluster coordinated by the MDT, the Ministry of Defense, the General Staff of the Armed Forces, the National Security and Defense Council, the Ministry of Economy and the Ministry of Strategic Industries, and unites representatives of the military industry under 'one umbrella.' Within this cluster, developers have the opportunity to receive government grants and support, as well as access to research and development centers. Any developer can join, and priority areas include AI, which has been declared a priority for 2024. Currently, Brave1 has registered 35 AI development projects, 29 were certified

for limitary purposes – meaning that they are effective in practice and meet humanitarian law standards.

One of the high-profile AI developments within the cluster is Griselda situational awareness system. It is integrated into Delta and enables intelligence collection. According to the description, it takes 28 seconds from the time the information appears in the system to its receipt. The information goes through 4 stages of verification and only after that it is transferred to the military. The system works according to the principle of diversification of data sources. The number of initiatives within the framework of Brave1 is constantly increasing, and the very existence of the cluster helps not only coordinate, but also control the type and quality of military technologies currently being developed in Ukraine.

Other technologies. As early as in June 2022, it was reported that Kyiv is defended by AI-equipped air defense systems that determine the flight trajectories of objects and allows to act in advance if necessary. In addition, the military plan to use advanced medical technologies – notably, the so-called smart tokens, and rapid development of bionic prosthetics is anticipated to address the needs of injured servants. Also, the applications for simulating combat operations and further strategic training, separate messengers for the military, technologies for searching for dead and missing military personnel, and many other sectoral initiatives are currently developed. Most of them are targeted solutions to problems in the defense sectors rather than complex solutions. The Ministry of Defense even organizes hackathons to find the best solutions and further collaborate with such developers. Almost every initiative is private, which raises the question of how cross-border trade in such technologies and export control from Ukraine will be regulated at the later stages.

After all, many technologies related to the military sector are used in other sectors, too. For example, the Ministry of Agrarian Policy and Food is applying AI to map potentially dangerous areas. It significantly speeds up the demining processes of liberated territories and enables protecting human lives when identifying dangerous zones. Demining will take place within the framework of the Strategy developed by the Ministry of Economy of Ukraine that has a strong emphasis on AI systems. As noted, Palantir has a huge role to play. Currently, there are many developments in the field of demining machines, as well as UAVs, which can quickly and effectively detect mines. After the end of the war, the Ministry of Defense plans to offer Ukraine a new strategy for the development of the armed forces – Future Force Concept that will take into account technical capabilities and effectively use them to build a domestic and global security architecture.

Recommendations

Active and widespread use of AI in Ukrainian public sector proves that digitization efforts are effective and this topic is a buzz. However, efforts to develop an adequate legal framework that can support practical novelties started relatively recently. Moreover, they largely reflect international trends that are common, yet avoiding many problems inherent in Ukrainian legal system and context. Although most initiatives are in line with the digitalization trends, some of them need both practical refinement and legislative regulation. Therefore, the Digital Security Lab Ukraine (DSLU) suggests paying attention to the following in developing AI policies:

- In the future, it is necessary **to develop a general regulatory framework** that will adequately incorporate international standards in the field of AI at the national level. At the same time, DSLU supports postponing the implementation of the AI Act for determining how Ukraine can adapt the requirements of the document without accessing to European institutions, as well as to understand how the regulation should be applied in practice. With this in mind, DSLU welcomes the introduction of 'soft' regulation, including development of general and sectoral recommendations and encouragement of self-regulatory institutions.
- In addition to general AI regulation, **thematic and targeted amendments to legislation** have to be designed and implemented to provide citizens with adequate safeguards against abuse. Notably, this is especially relevant for the sectors of justice, law enforcement, and the military.
- It is necessary to adjust the legal framework on **public procurement** to AI systems – for example, to develop by-laws that set minimum standards for developers of AI systems designed for the public sector, requirements for conducting tenders and other guarantees. Notably, DOZORRO implements a machine learning model to identify high-risk procurement, but it is necessary to regulate the procurement modalities of the AI systems themselves, especially considering that many developments involve public-private partnerships.
- It is necessary to develop an acceptable **model of the regulator in the field of AI** and adequately fit it into the functioning legal system. In particular, availability of thematic regulators such as the National Council for Television and Radio Broadcasting and the future regulator in the field of personal data protection should be considered. Development of the regulator's model also involves the design of the complaints system – the definition of the range of subjects of complaints, topics and which body will consider them.

- It is necessary to **get rid of AI systems that work on Russian, Belarusian, Chinese or other software** originating from countries with high index of human rights violations. The government must make sure that after discovering any connection of AI systems with such countries, they are not used at least in public sector. Notably, this is about at least reacting to the investigation of Schemes regarding the TRASSIR system, and ideally, about banning such systems on the territory of Ukraine even for private actors. The government must also control that such software does not enter Ukrainian market – and in particular the public sector – as a foreign product imported through third countries.
- The government needs to **develop in advance a strategy for the transition period** between martial law and peacetime after the Ukrainian victory, which will provide for legal and practical mechanisms for terminating the use of technologies exclusively intended for martial law. Notably, it is necessary for warning foreign companies that provide wartime AI services about a smooth and clear algorithm for terminating the use of some services or applying completely different standards to them, such as a more restrictive approach in assessing the proportionality of interference with human rights, etc.
- The government should **avoid using technologies that by default violate human rights**, such as international standards in the field of personal data protection, anti-discrimination, create risks of persecution, etc. For example, alternatives to Clearview AI should be explored, especially given that the AI Act bans data scraping and considering other problems related to Ukraine's European integration processes.
- The government should develop **a system of technical standards** that provides for minimum requirements for system developers and assess the prospects of creating a body responsible for the certification of AI systems, or assigning such duties to an already existing body.
- The government should **assess the risks of AI systems** used in the public sector at all stages of the lifecycle of such systems. Various methodologies may be used as a guidance, including HUDERIA – a methodology where Ukraine is already included in the test phase. Impact assessment should take into account the possibility of a chilling effect on human rights, and the government should aim to minimize it and develop safeguards against abuse.
- The government should **develop a unified system of labelling AI-generated content** to ensure greater transparency regarding functioning of the systems, as well as to avoid irresponsible dissemination of false information.

- When providing public services with the help of AI systems, the government must **ensure the ability to easily and clearly communicate with human operators**. In particular, the situations where the automated system is loopy or takes too much time to talk to the bot should be avoided. Examples of such systems in the private sector include popular banking chatbots described in Texty.org study and telephone operator systems that have very inconvenient algorithms preventing the exercise of the right not to be affected by automated systems.
- The government should provide **open source for AI systems** used in the public sector, except for technologies in the field of security and defense, to enable bug tracking by independent experts and effective public control of the type and features of the applicable systems, their impact on human rights.
- After announcing the initiative to introduce an AI system in a public sector, the government should **regularly update information** on the results of test periods, problems encountered in application of the system, the stage of project implementation, etc. to ensure transparency and allow effective public control over the implementation of projects, their reliability and safety.
- The government should **continue international cooperation** both at the level of collaborations with companies and within the framework of regulatory initiatives such as the Council of Europe's Committee on Artificial Intelligence. It will make it possible to monitor the cutting-edge trends and shape the international agenda, taking into account Ukrainian experience in the application of AI.
- The government should **ensure participation of civil society** in discussing ideas for the introduction of AI systems in the public sector, and take into account the results of such discussions when making political and legislative decisions, developing visions and strategies.
- The government should invest efforts in **strengthening digital literacy of public servants**. Notably, training projects within Diia.Education serve as good practices in this field. However, more targeted training is also encouraged, especially for public servants who regularly deal with complex high-risk AI systems.
- When **creating training programs for AI developers,** the government should include in the curriculum courses on ethics and international standards of human rights protection to ensure that AI systems that are developed are ethical and compliant with human rights and democratic principles by default.

Finally, no matter how the international regulation in the field of AI unrolls, development of national practices must comply with basic human rights principles and standards, regardless of the availability of industry regulation. Monitoring of this compliance is impossible without involving all stakeholders and creating an environment that facilitates free discussion. And it is critical to remember that ignoring challenges and dangers today will create more problems tomorrow. Therefore, it is better not to postpone the search for solutions for distant future, but to look for ways to balance the interests of stakeholders now.

Digital
Security
Lab