EOSC-Future & RDA Artificial Intelligence/Data Visitation Workgroup

Deliverable 4

Guidance for Ethics Committees Reviewing AI and DV

Draft Version: AUGUST 12, 2024

Subject matter exerts from across the globe developed these recommendations to provide a European and global platform for engaging ethics, law, and social considerations in AI, data sharing, DV, and Open Science policy development. These recommendations are consensus-driven and actionable.

Prepared by *Challace Pahlevan-Ibrekic (USA), Valery Sokolchik (Belarus), Rita Sitorus (Indonesia), Perihan Elif Ekmekci (Turkey), Aliaksei Razuvanau (Belarus), Ülkücan Kaplan (Turkey), Pukovisa Prawiroharjo (Indonesia)*

Contents

List of Abbreviations

AI - Artificial intelligence

DV - Data Visitation

EMA - European Medicines Agency

FDA - Food and Drug Administration

GDPR - General Data Protection Regulation

HC - Health Canada

ML - Machine Learning

NGO - Non-Governmental Organization

REC/IRB/IEC - Research Ethics Committee/Institutional Review Board / International Ethics Committee

[In this guide, RECs that review human researches refer to IRBs as well as IECs)]

List of Figures and Boxes

List of Appendices

Draft Recommendations Summary

1. Build multi-disciplinary review committees: Identify and appoint diverse experts (ethicists, regulatory professionals, technical developers, informaticists, health system operations, etc.) to serve as - research ethics committees/ institutional review boards (REC/IRB) members for the review of research including artificial intelligence (AI) and data visitation (DV).

2. Identify potential risks: AI risks can be eliminated or mitigated once the risks presented by the specific AI are identified. While all AI includes algorithms, statistical techniques will vary, thus not all AI (and their risks) are created equal. RECs/IRBs must consider the aim of the technology and its proposed use to identify potential risks such as risk of bias and risk of physical or other harms, to name a few; involves ethical and social risk assessment (sustainability, responsibility & impact) and model risk assessment (reliability & predictability due to accuracy and data quality). Moreover, dual use capacity of the proposed AI and potential harms associated with dual use should be considered. This includes defining reliability criteria and parameters for AI (acceptable for AI use in health care), understanding scalability options and usability testing of the AI under review, and reaching consensus on potential risks.

3. Develop materials to assist with regulatory review of research including AI and DV: Checklists, guidance and policies can help REC/IRB reviewers document their assessment of the risks/benefits of using AI and DV in research. They will also guide the REC/IRB members to qualify their decisions in a way that does not exclude any significant evaluation criteria.

4. Develop targeted training programs to help identify risks of AI and DV in research: Once regulatory review material is developed, targeted trainings on how to use the material to identify risks of AI and DV is necessary, including trainings on how to differentiate between ethical and model/technical risks. These training materials should be updated regularly to catch up with the latest developments in AI technology so that reviewers have a comprehensive understanding of the capacity of the state of art technology.

5. Detect ancillary committees to assist with the review of AI and DV in research: The risks of AI and DV extend beyond research to health care operations. To overcome this obstacle, REC/IRBs may either seek collaboration with ancillary committees or in the absence of such committees, may invite specialists from different backgrounds to evaluate the risks extending to various domains.

6. Evaluate the ethical issues presented/ avoided by the use of AI/DV in research proposal. Also consider open science principles, responsible conduct of research principles, data and research integrity, and the ethical principles of the particular research area in preparing the ethical review by the REC/IRB.

Overview

AI promises valuable insights to all sectors and domains through analysis and manipulation of massive datasets, often created as a byproduct of being human in today's digital age.

This data can be used to offer technological solutions that range from the practical, such developing more efficient workflows, to the innovative and enigmatic large language or unsupervised predictive models.

Indeed, the disruptive capacity of AI in the recent reports of the rapid technological advancement and accelerated use of generative AI (e.g. Chat GPT, Murf, Fireflies, DALL- E 2) have captured the attention of government regulators, the scientific and academic society and the general public across the globe.

In the healthcare sector the disruptive nature of AI is even more significant than other sectors. The innovative statistical models and data solutions with machine learning (ML) and deep learning (DL) capacity have the potential to revolutionize healthcare delivery systems. However, ML, DL and, other models require large volumes of data to train, validate and monitor the accuracy and efficacy of the AI solutions. Data of this nature are generated daily within healthcare systems through the routine provision of standard healthcare services, rather than for research purposes specifically. Access to such data is understandably restricted by privacy regulations in order to safeguard and uphold the privacy and confidentiality of the individuals from whom it is collected.

The General Data Protection Regulation (GDPR) is a comprehensive data protection law in the European Union (EU) that aims to protect the privacy and personal data of individuals within the EU citizens. It regulates how personal data is collected, processed, stored, and shared by organizations, and grants individuals greater control over their personal data. The ethical grounds of the GDPR is sound, intending to improve transparent and transnational use and access to personally identifiable data while protecting rights and dignity of data providers, has resulted in even more siloed and disparate databases which significantly hinders AI-driven scientific discovery. Furthermore, the growing development of big data has amplified concerns surrounding the protection of privacy, confidentiality, and individuals' rights to control the use and reuse of their personal data.

To overcome this problem federated models aiming to access to diverse data without breaching the GDPR and ethical norms to protect privacy and confidentiality has been offered. Data Visitation (DV), is one of these solutions that provide effective access to diverse data without threatening the privacy and confidentiality of the individuals from whom the data are derived from. Federated models contain potential that could be beneficial in this regard considering the sophisticated data security infrastructure and complicates the chain of responsibility they require– a tenuous component of AI in general.

There has been some effort through national regulations such as the Food and Drug Administration (FDA), the Health Canada (HC), the European Medicines Agency (EMA), to establish frameworks that promote the responsible use of data in the development of AI, as well as to mitigate the unforeseen risks of its various uses.

However, use of AI in clinical research, and methods of access to the data required for such solutions, presents unique and paradoxical challenges that require close examination by REC/IRBs to identify the ethical, legal and social issues that use of the technology presents to humans.

REC/IRBs are charged with the responsibility to protect the rights, dignity and welfare of research volunteers, including use of their data that may have been collected for other, non-research purposes. Generally speaking, REC/IRBs review research involving humans according to a principled approach and usually refer to a specific set of regulatory criteria to document their assessment of the potential risks and benefits of the proposed research. With the advancement of AI and its growing use in research, REC/IRBs are confronted with new challenges when reviewing the downstream potential negative risks of AI on research participants a human society at large, the first of which is a dearth of concrete guidance for review of AI in research.

The recommendations set forth in this guidance outline basic ethical principles that will assist research ethics committees to understand the questions, methods, and procedures for reviewing AI, including use of DV in research with humans.

Target Audience for this document is REC/IRB members, researchers, data scientists, AI developers as well as other stakeholders who may be involved in development, use and, deployment of AI and data.

Multi-Disciplinary Ethics Committees

AI involves numerous fields, including theoretical computer science, mathematics, and statistics, to name a few. Likewise, ethical assessment of AI technology uses in research necessitates evaluation from legal, societal, ethical and historical perspectives. Given the diverse frames of reference, it necessary to identify and appoint equally diverse experts such as ethicists, regulatory professionals, technical developers, informaticists, sociologists, health system operations, etc. to serve as ethics committee members for the review of research including AI and DV.

Subject matter experts pertinent to AI and DV include:

1. AI designers or developers (engineers/ methodologists) who understand and use the technical framework of models used in the respective research
2. Ethicists and other trained individuals who have the capacity to understand ethical risks of AI and DV that may extend beyond the purview of a REC/IRB, and who can help provide objective ethical arguments for consideration;
3. Regulatory personnel including, but not limited to, privacy officers and attorneys, to provide a legal framework for assessment of AI and DV risks; and
4. Clinicians, informaticists and healthcare administrators who understand the clinical problem the AI or DV is attempting to solve;
5. Social scientists, representers of patient's organizations and non-governmental organizations (NGOs) and lay persons to help inform the underlying social determinants of health that impact the data (or lack thereof) used to train algorithms and the potential ethical and social implications of the proposed AI system.

REC/IRBs that include experts in the domains noted above help to improve the systematic identification of the ethical, legal and social risks of AI in research through diverse backgrounds, experience and different perspectives.

Ethical Principles for use of Artificial Intelligence in Human Research

1. Research with humans must be reviewed and approved by REC/IRBs before it can be initiated. To be approved, the proposed research must meet specific criteria that are, generally speaking, based on the principles of the Declaration of Helsinki.[1]

The use of AI and DV in human research is closely connected with implementation of principles of biomedical ethics and responsible conduct of research that are listed as; respect for individual autonomy, justice, beneficence and non-maleficence. These principles are addressed to protect human subjects from any risks that may emerge from the clinical research. Their area of use and impact is enhanced to guide the ethical review process of the research that involves AI and data use, particularly federated systems such as DV. By referring to the four principles of biomedical ethics, issues such as data integrity in research involving AI and DV, protection of the privacy and confidentiality of personal data, safeguarding the rights of data providers, avoiding biases or prejudices embedded in AI systems, enabling fair access to AI, responsible and reliable use of AI can be addressed. These considerations together with embodiment of principles of open science in the ethics review process help to build trust in both the process and results of research and science in general.

However, research involving AI, either as the focus of the investigation or as a tool for analysis, requires the same foundational considerations as other research conducted on human subjects, but requires an additional calculus that seemingly tests the flexibility of our current regulatory frameworks.

Under most of the world's current regulatory landscape (e.g. the "common rule" at 45 CFR 46, UU PDP no 27/2022; UU no 17/2023 Kes, ADD OTHER RELEVANT REGS from our respective countries), review of research that proposes secondary use of data generated or collected for non-research purposes (e.g. social media interactions, medical record information, etc.) is relatively straightforward, and often does not require close monitoring by REC/IRBs, or in some instances is completely exempt from review and oversight. This regulatory approach is permissible according to our current framework, but there are valid concerns that argue the inadequacy of the current systems to protect the rights, dignity and welfare of the individuals whose data are used to develop AI systems.[2] Furthermore, the current practice is not designed to require consideration or assessment of the full spectrum of an AI's impact on society. For example, under 45 CFR 46.11(a)(2), "the IRB should not consider possible long-range effects of applying knowledge gained in the research (e.g., the possible effects of the research on public policy) as among those research risks that fall within the purview of its responsibility." In other words, REC/IRBs in some countries are constrained to only evaluating the privacy and

---

[1] https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/

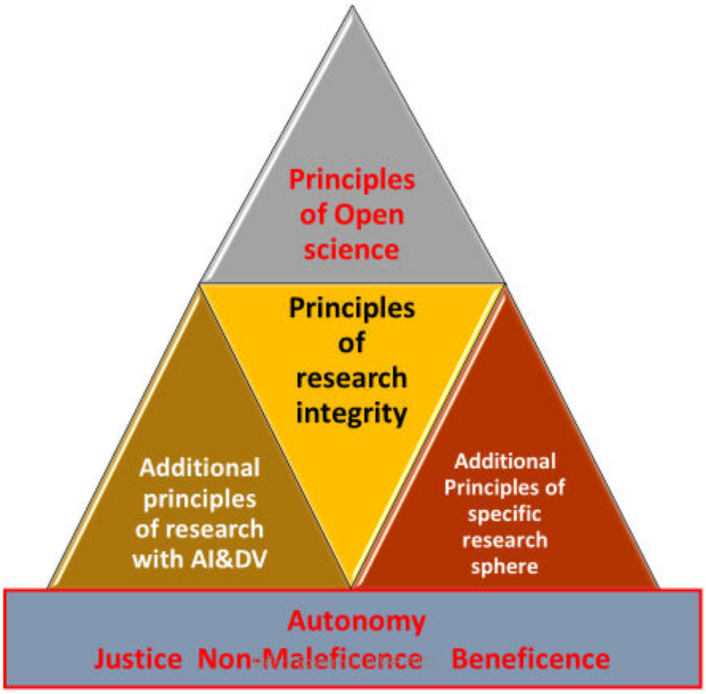
[2] SACHRP: https://www.hhs.gov/ohrp/sachrp-committee/recommendations/attachment-e-july-25-2022-letter/index.html

confidentiality risks linked to data usage for AI advancement, and they are not always required to assess the broader and more worrisome potential and unforeseen future harms that AI may pose in general.[3]

The current approach is rather inadequate, considering that clinical decision support systems are increasingly being designed with predictive capabilities that can profoundly influence the trajectory of an individual's clinical care, for better or for worse. The REC/IRB’s assessment of the risks and potential benefits of validating the AI prior to research to assess if use of the AI in the research project presents greater than minimal risk to participants enrolled in the research and if sufficient risk mitigating strategies are defined to avoid unnecessary, irreversible and/ or significant harm. Moreover, this assessment may help protecting patients and society in general before the widespread clinical deployment of the tired process from future negative impacts to society, including equitable access to care.

Drawing from these discussions, it is reasonable to assert that the ethical assessment of AI in research is informed by the fundamental principles of biomedical ethics. However, it should also be guided by both objective principles concerning the utilization of AI and subjective principles concerning the interaction between humans and AI, as depicted in the illustrations below. Moreover domain specific ethical concerns should be embedded in this framework, suggesting particular sensitivities in research in the domain of military, health or environment. Note that dual use of the research outcomes should also be taken into consideration within the ethical review process.

(Fig.1, Fig.2):



---

[3] ibid SACHRP

Figure 1. The correlation of the principles of research ethics

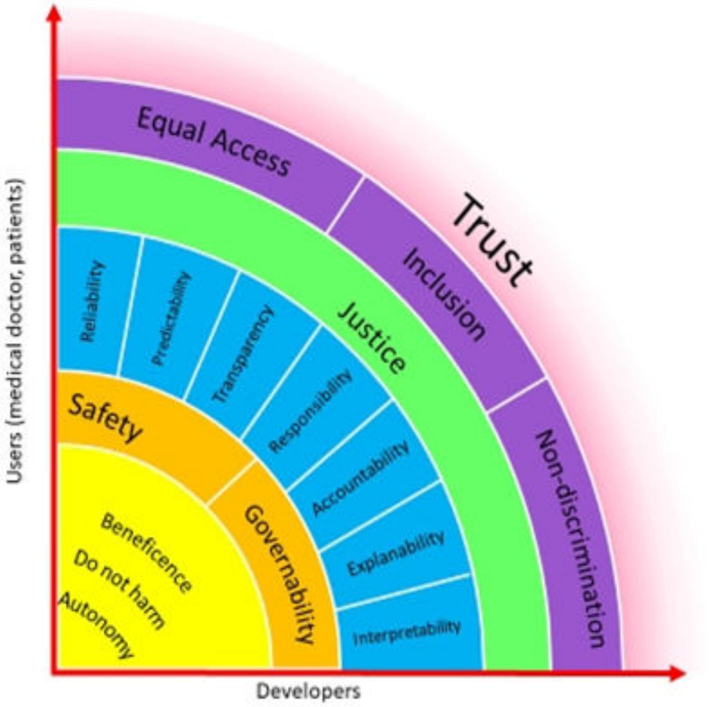


Figure 2: The main ethical principles of using AI in research

Principles for use of Artificial Intelligence in Human Research

Objective ethical principles of research with AI can be summarized as safety and robustness, responsibility and governance, accountability, explainability, transparency, effectiveness. These principles can be articulated among the core principles which determine the main ethical requirements for AI /DV research, and define the minimum criteria for the ethical compliance of the research. In addition to these core principles, the REC/IRBs should assess the research proposal in terms of data ethics, including but not limited to: data integrity, data safety, data impartiality, and the presence of controlling and preventing measures to avoid accuracy errors of the AI that is subject to the research. These requirements are governed to some extent by international and/or national laws, pertinent recommendations, and guidelines. It is paramount to take these principles into account during the ethical review of any research involving AI/DV. Failure to adhere to these principles in the research process will result in the inability to obtain approval from the REC/IRB to proceed with the research.

The above principles of ethical compliance in research with AI/DV are complemented by the core human principles of justice and trust, which are culturally influenced . The implementation of these two principles in research with AI/DV is quite difficult to evaluate sometimes due to their correlation to sociocultural attitude varying among societies and individuals and no possibility to rely on expert opinion because of the with subjective connotations of these principles. On the other hand, they hold a key position for ethical assessment of the AI/DV research to avoid biases, unfairness in terms of access to AI as well as contradictions inherent in AI. (Sokolchik, Razuvanau, 2023).

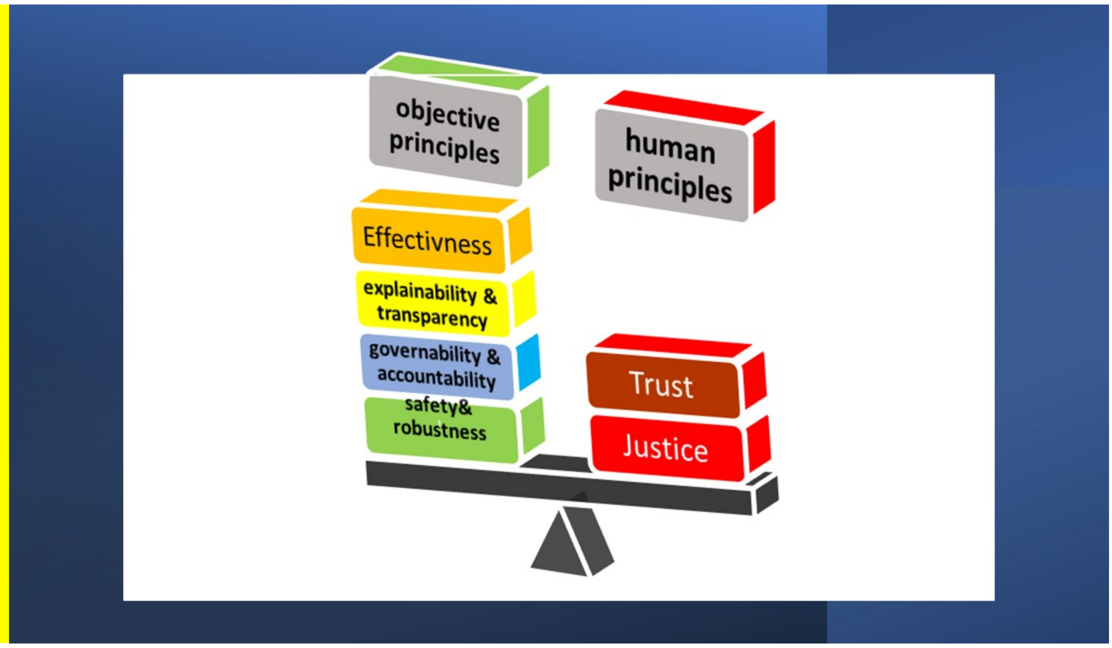


Figure 3. Objective and human principles of using AI in research

The principles that RCE/IRBs should consider when reviewing research involving AI/DV

Safety is one of the most important ethical principles that guides review of research using AI/DV. As our use of AI becomes more integrated and ubiquitous, it may be difficult to predict whether this new and rapidly developing technology will have a limit in terms of dissemination in various sectors and result in unforeseen/ unexpected/ unwanted consequences. Therefore, the importance of how REC/IRBs evaluate and determine AI research to be safe for use cannot be overemphasized. To deem an AI technology safe does not imply it must be entirely risk-free, as achieving such a state may not be realistic and may not align with prevailing medical standards. Instead, in the context of AI, to be safe implies that the risks of its use are reasonably known or at a minimum and acceptable given the known or anticipated benefits. It also requires an awareness of the population(s) most likely to incur benefits or face the risks from its use. In essence, ensuring the safety of AI necessitates the implementation of adequate security compliance measures throughout the development and deployment phases. This is crucial for preserving data integrity, verifying the accuracy and validity of model output, and establishing robust risk mitigation strategies to prevent any unexpected, disproportionately significant harms. Practically speaking, the principle of safety requires that REC/IRBs assess following:

1. The privacy and confidentiality risks introduced or deepened by the technology. This includes assessment of the potential risks incurred when data are combined with external sources, especially if disclosure could place individuals at risk of criminal or civil liability and limits to privacy (e.g. does the technology offer one hundred percent privacy? What limits to privacy may be reached in the future? How much does this breach to privacy restrict an individual's future choices?).
2. The real or perceived physical and/or psychological harms that could result from use of the AI, in general and in certain groups/populations (e.g. When the entire ecosystem is evaluated as the domain of artificial intelligence applications, how will the vulnerable groups be effected? Is there be an imbalance of risk of harm among vulnerable populations and the rest of the community?);
3. Potential dual uses or misapplication of the developed technology that could pose a significant threat to the public once deployed (e.g. Can this research have unpredictable future damages? In such a situation, are members of the ethics committee able to foresee this? Will the ethics committee have a say before the outcomes of the study are disseminated in various sectors that may pose significant risk ?);
4. The quality of the data used to develop and validate the technology. This includes assessing the potential for under- or over-fitting of the model and inaccuracies with output that in turn could lead to increased risk of harm;
5. When or if reassessment of the technology is needed to account for model drift and confirm continued efficacy;

6. The types of validation required to confirm performance of the model and plans for clinical integration of the technology, including training and feedback on usability and efficacy; and
7. the level of information and considerations for informed consent?

REC/IRBs should require researchers at the time of the initial submission of their research to communicate how they plan to ensure safety according to the above. Any research protocol that does not meet the safety criteria of the research ethics committee should be revised. Where reasonable safety cannot be achieved, or the risks outweigh possible benefits, REC/IRBs have a duty to consider if the technology should be disallowed/cancel the proposed development or use of AI.

REC/IRBs should also require plans for future data sharing, as applicable. In this context, the following questions on data security and data sharing should be answered by the AI research team, both during the research phase and when the outcomes of the research are deployed (box 1.)

Box 1. Questions for the research team (on data security and data sharing)

1. Have you set a policy to ensure data security?
2. How will the properties of the database where the data will be stored be determined? (Responsible persons, security, data access conditions, legal basis, ethical accountability, etc.)
3. What are the third parties' access conditions to this data?
4. Who will be responsible for the security and monitoring of data belonging to insufficient or vulnerable groups?

The responsibilities of the researchers in the matters mentioned here should be clearly stated by the REC/IRB and the relevant safety precautions should be documented.

Accountability is another fundamental principle that should govern review of AI in research is accountability that is defined as the ability to control the operation of AI systems and the presence of a clearly defined and hierarchical system of responsibility for actions and results produced through AI.

Accountability for AI research refers to the obligation of researchers involved in developing AI to take responsibility for the ethical, legal, and social implications of their work. This includes being transparent about their research methods, ensuring that AI systems are designed and deployed in a manner that respects ethical principles, and being prepared to address negative consequences or risks that may arise from the use of AI technology. In this respect accountability presupposes:

- the initial determination of the required parameters for using the system in a specific area and compliance with these parameters. For example, in medicine, a prerequisite for the use of AI systems is compliance with such parameters as the physical, mental and data safety of the patient, the focus of any actions on improving longevity and/ or the quality of her life, the requirement to obtain the informed consent the patient, the application of existing ethical and legal standards, ensuring fairness and avoiding discrimination;

- it also requires the presence of legal support for AI activities both on a national, regional and global scale, and strict compliance with established standards, including the presence of mechanisms for assessing the potential harm caused by AI systems with subsequent compensation/ mitigation for it.

Ensuring the accountability of scientific studies in which the said technology is used and/or developed is one of the most effective control methods that can be applied at the first stage of model development. Successfully managing this period, characterized by applications to REC/IRBs can help to ensure that AI technologies with such a strong network of influence are reasonably maintained and limited.

The principal of accountability requires REC/IRBs to consider the following when creating specialized REC/IRBs to review AI research, or for existing REC/IRBs who may be asked to take on review of AI research (box 2)

Box.2 Accountability requirements from RECs

Ethical rules as well as legal regulations should be the main basis in the evaluation processes of research by ethics committees. Therefore, ethical accountability should be the main motivation for ethics committee decisions, but national or international legislations should be taken into consideration as well.

The right to benefit from scientific progress and its application is an important part of article 15 of the International Covenant on Economic, Social and Cultural Rights. To use this right effectively, societies must be informed at a level that promotes sufficient understanding of the scope of scientific discoveries to interpret the benefit-risk balance posed by the outcomes of the research. This situation imposes responsibility on REC/IRBs for disclosing sufficient and accurate information to community and ensuring community engagement. Within this context, REC/IRBs may be expected to provide some insights to of the concerns of the society regarding research with AI and having access to using the right to benefit from scientific developments. While various methods may be employed, the significance of REC/IRBs' perspectives cannot be overlooked in developing the capacity to effectively assess these inquiries. Research that receives ethical approval by a committee and will also positively impact people's attitudes toward this technological advancement.

These discussions imply that the duties and responsibilities of ethics committees are increasing. Rather than being a group of experts isolated in an ivory tower, solely evaluating individual research projects, IRB/RECs moving towards becoming one of the key elements shaping the community's approach to technology research and its outcomes, even integrating closely with society to determine their attitudes toward such research.

Therefore, it becomes increasingly crucial for members of REC/IRBs to undergo effective and continuous education that is free from discrimination, bias, ignorance, and indifference when reviewing research involving AI. Prioritizing transparency in their decisions becomes more important than ever as well as clear definition of their position in establishing the balance between the right of society to benefit from science and the safety of people (box 3).

Box 3. Questions to the Research Team (right to science)

1. Is the “right to science” balanced with the compromise form some rights (such as being safe, protecting privacy, being treated equally)?

2. How should the right of individuals to approach scientific developments cautiously in the face of society's right to benefit from scientific developments?

3. When studies are evaluated by ethics committees, how can the boundaries of the evaluation between the society's right of access to science and the risks that society may be exposed to be determined?

Explainability, another key principle of ethical review process, is the existence of a fundamental possibility of understanding the actions of the system, the transparency of its algorithms, the accuracy of explaining the essence and results of the process of generating results (Reddy, 2022). The principle of explainability is implemented through criteria for the transparency of AI actions and is based on the principles of safety and controllability.

If a researcher does not understand the general algorithm for obtaining results using AI, she ceases to control the process, notice errors and, accordingly, the results obtained become exclusively quantitative and difficult to interpret meaningfully.

The principle of explainability presupposes the availability of appropriate qualifications of the researcher (and members of the research team), the ability to communicate with the developer, and the ability to control and promptly eliminate errors, in particular those arising from the actions of AI.

REC/IRBs must not only evaluate the presence of errors and inaccuracies in the research process, but also the potential possibility of their elimination/correction. Also, REC/IRBs, when examining research using AI/DV, pays attention to the opportunity to seek consultation from the developer for the researcher, the opportunity to obtain consultations for research subjects, the opportunity for training for research subjects (if necessary), the possibility of a rational explanation of the results obtained by the researcher.

Nonetheless, these efforts might prove somewhat futile if the REC/IRB is tasked with evaluating generative AI, which is inherently more intricate and lacks explainability. Assessing the ethics of generative AI proposals becomes more challenging and may necessitate additional expertise, particularly regarding the principle of explainability. The essential focus might be to grasp the algorithm in broad strokes and comprehend its implications rather than fully understanding its inner workings.

Efficiency principle dictates the accuracy and precision of the data collected and processed by AI systems. An important criterion for effectiveness is compliance with such a parameter as the "knowledge limit," which caps the operation of AI systems to the conditions and purposes for which the system was developed. Unauthorized changes in system operating parameters by users such as changes in temperature conditions, operating rules or scope of use, may cause additional risks emerging from unpredictable errors, distortions and, as a result, data inconsistency.

The effectiveness of AI systems also depends on the ability of users to work with the system or data. Sometimes incompetence leads to the fact that the unambiguous results of AI analysis and the predictive capabilities of the system are overly exaggerated, which only creates the appearance of objectivity, justified by the "infallibility" of the analysis carried out by AI.

The effectiveness of AI systems is determined not only by protection from errors and distortions, but also by the validity of the analyzed data. Validity requires the collection of critical, important data, which is determined by researchers and subject matter experts (Saracci, 2018). Analysis of compliance with this principle is thus associated not only with consideration of the ethical parameters of the study, but also with the verified research methodology and competency of the users.

Justice (and inclusivity) principle is also important aspect of ethical review of AI/DV research. The principle of justice and inclusivity in the ethical review of AI involves ensuring that the design, development, and deployment of AI systems consider and incorporate the diverse perspectives, needs, and experiences of all stakeholders, including individuals from different demographic groups, communities, and cultures. This principle emphasizes the importance of addressing potential biases, discrimination, and exclusionary practices in AI technologies to promote fairness, equity, and accessibility for all individuals and groups affected by AI systems. Inclusivity also entails actively involving relevant stakeholders in the decision-making process and fostering transparency and accountability in AI development and implementation processes.

Avoiding bias is also one of the main ethical concerns with AI/DV review process. AI algorithms are only as good as the data they are trained on, and if the data contains biases, the resulting algorithms will also be biased. Researchers' awareness on this issue should be taken into account and necessary precautions should be taken for possible bias.

Bias in AI algorithms can result in discriminatory outcomes, reinforcing and exacerbating existing inequalities and marginalizing certain individuals or groups based on factors such as race, gender, ethnicity, or socioeconomic status. Furthermore, considering the potential societal impact of biased AI systems, adhering to the principle of avoiding bias aligns with broader ethical imperatives to promote social justice, human rights, and the public good. It reflects a commitment to responsible innovation and underscores the ethical responsibility of researchers and developers to prioritize the well-being and dignity of individuals and communities affected by AI technologies.

Box 4. To prevent bias and discrimination

1. Validate the data set used for machine learning. This means checking that there is no intentional discrimination in the data and that it accurately represents the population being studied.

2. Apply methodologies and software solutions that detect and prevent discrimination based on factors such as race, national origin, gender, political opinion, religious belief, age, social and economic status or privacy. These tools can help ensure that AI algorithms are fair and unbiased.

3. Adapt algorithms at the regional/ national level to account for differences in socio-economic conditions and the state of national systems. An algorithm with diversity in mind ensures that everyone has access to high-quality services. .

Trust is an ethical principle that implies compliance with all of the above principles, and on the basis of these ethical principles, it forms AI users (both the researcher and research participants) a positive emotional attitude when interacting with AI, confidence in the results and actions of AI, a personal sense of reliability and security.

Building public trust is indispensable for research ethics that the limits, known risks and unpredictable/ unforeseen negative results of artificial intelligence research are shared transparently by researchers. Likewise, the society should be informed about potential risks as well as the benefits of this technology and even have a say in the health policies to be shaped for the future. The way to achieve this is to gain the trust of the society. To ensure this;

1. It is crucial to disclose the use of AI in an application. Individuals need to be informed if AI is utilized in the service process and the extent of its involvement. REC/IRBs should ensure that researchers address this disclosure requirement.

2. The benefits of AI research in every sector that it can disseminate to should be shared with the public in an understandable and transparent way.

Building community trust is also important to ensure the continuity of volunteers participating in AI research. As a result of all these efforts, people can criticize AI technologies and research according to their own intuition, without relying on hearsay. This will increase public awareness and make it easier for the public to make better decisions while having more say.

It is seen that the effects of the successful management of REC/IRBs will not be limited to ethical committees only, but will create a ripple effect from the individual to the society. The fact that this effect is positive depends on the successful maintenance of the manageability of the ethics committees monitoring and evaluating AI research.

Box 5. Artificial intelligence technologies in health policies

The inclusion of artificial intelligence technologies in both national and international health policies is closely related to a successful process management. Bringing artificial intelligence technologies to the agenda by field professionals is important in terms of explaining technology to the society. Thus, the gap between technology opponents and those who bless technology can be reduced to a more appropriate level. It can be ensured that people question their feelings of unfounded fear or unconditional trust and thus gain a more objective perspective. Another advantage is that governments can be aware of the limitations and demands of boards evaluating AI technologies. Achieving all these is closely related to the correct communication of the process to policy makers. Ultimately, this is a reflection of the successful manageability of ethics committees.

It may not always be possible to answer all these questions about artificial intelligence research, but we cannot deny our responsibility to seek these answers. The first step may be to accept that scientific research that produces AI applications, a very new and widely applied technology, will become more and more common in our lives. Instead of resisting this technology, it seems like the most logical approach for now is to be cautious and follow all its steps in a controlled way. Ethical committees will be at the center of this control mechanism. While both governments, ethics committees and researchers provide the necessary safe and free environment for the society to benefit from technological developments; It is also responsible for emphasizing the responsibilities of these developments to human dignity, to other living things and therefore to the ecosystem.

Guidance and Review Material

Checklists, guidance and policies can help REC/IRB reviewers document their assessment of the risks/benefits of using AI/DV in research. We suggest REC/IRBs to use the following basic questions to evaluate ethical compliance while assessing plausibility of the research design, methodology, aims, goals and involvement of data and human subjects.

Key-issues for REC/IRB about ethical and legal compliance of the research proposal:

1. Check if the researcher adheres to principles of open science.
2. Determine if research participants will directly interact with AI or if AI will solely process/receive data. Provide recommendations/consultations/support information for participants if direct interaction is involved.
3. Verify if members of the research team possess competencies in working with AI (e.g., special training, certificates).
4. Confirm if special training/instruction on using AI programs is provided for:
   a. Research team members
   b. Research participants
5. Assess the availability of advice on using AI from the developer or a member of the research team for:
   a. Research team members
   b. Research participants
6. Check if consultations on the planned research were conducted with stakeholders (e.g., public organizations, patient communities).
7. Determine if the research team has analyzed the suitability/opportunities of the AI system in specific socio-cultural conditions, taking into account characteristics such as religion, culture, profession, age, gender, etc., of participants/users.
8. Ensure that the form of informed consent provides information about the use of AI (scale of use), including the use of special devices, testing, questioning, etc.

Key-issues for REC about using of datasets, and participant's data

1. Will existing open databases be utilized, and has the DV algorithm been considered?
2. Have necessary permissions from authorities to use databases been obtained, if applicable?
3. Are measures in place to protect the information/data of research participants?
4. If intending to use data from participants sourced via Social Networks or other open sources:
   a. Have participants provided informed consent for this usage?
   b. Can participants access and review this data?
5. Have levels of access to research participant data within the research team been defined?
6. Is the use of AI and DV methodology for research purposes or healthcare operations/quality improvement?

7. What is the source and general characteristics of the data to be used?
8. Can end-users review factors relevant to the output being offered?
9. Is there an alternative option for using AI through human communication upon request from the research participant?

Key-issues for REC/IRBs about risks of participants and healthcare provision

1. Are there any risks introduced at this stage, if so, what are they?
2. Is the model considered a medical device subject to regulatory oversight by the FDA, HC, or EMA?
3. Is the medical device classified as a significant risk device according to established FDA, HC, or EMA criteria?
4. Does the research involve validating a previously developed model? If yes, does the validation impact clinical care?
5. If the validation impacts clinical care, can decisions be made without the use of the technology, or does the technology run parallel to clinical care and inform potential courses of action that a physician could take (driving decision-making)?

Reviewing the research in the process of its implementation (at the final stage) REC/IRBs need to investigate the research documents, additional materials of research, conversation with researcher, participants, developers (if applicable), anonymous survey for participants (if applicable).

Additional key-issues are:

1. Are research participants informed about the frequency/algorithm used by the researcher to analyze potential AI errors?
2. Have there been any instances of deceiving research participants and what actions were taken? How is the deception justified?
3. Are alternatives for the use of AI provided for research participants, if applicable?
4. Do participants know where to seek assistance with questions regarding the use of AI?

Training

Once regulatory review material is developed, targeted training on how to use the material to identify risks of AI and DV is necessary, including training on how to differentiate between ethical and model/technical risks.

AI technologies are a relatively new field of study, not only for society but also for much of the scientific community. For this reason, many scientists from different backgrounds who serve on ethics committees may have limited knowledge about artificial intelligence technologies. This is an expected situation for today rather than a deficiency. On the other hand, educating the members of the ethics committee about AI technologies and the possible ethical problems of this technology can make significant contributions to the manageability of the ethics committee processes. Maintaining the said trainings and updating their contents in line with the needs are also valuable in terms of laying the groundwork for the development of optimum training curricula and guides over time. These trainings can be enriched with the experience of experts in the field of AI and ethical evaluation of case studies. Thus, the ties between the field of ethics and artificial intelligence technologies can be made tighter.

Ancillary Reviewers

Depending on the use case, risks of AI extend beyond the purview of a REC/IRB. Projects that propose to deploy AI technology within a healthcare system or facility must include review by departments responsible for the oversight of system heath safety such as health information technology committees. Plans should include how model drift will be identified, how potential risk(s) will be mitigated and who has the authority to discontinue use of the AI.

Appendix 1. Checklist for RECs/IRBs/IECs reviewing research with AI/DV before the starting of the research

№	Ethical principle	Evaluation indicator	Evaluation category		
			Compliance in full	Partial match	Lack of conformity / not usable
1	2	3	4	5	6
1	**Safety (security) & robustness**	The planned design of the study includes informing study participants about working with AI elements (information is included in the informed consent, memos, etc.)			
		Periodic control of the use of AI by the developer is provided. There are also ways of emergency communication of the researcher or research team (hereinafter referred to as the Researcher) with the developer in the event of emergency situations.			
		Periodic control of the use of AI by the Researcher is provided.			
		Methods of emergency communication of the study participants with the Researcher in case of emergency situations are provided.			
		Protects the privacy of study participants' data			

1	2	3	4	5	6
		Algorithms for encrypting / depersonalizing the data of study participants are provided			
		An algorithm has been developed for the Researcher to analyze the risks of the study participants (physical, mental, informational, financial, legal, ethical, etc.)			
		An analysis was made of the use of the proposed AI systems in terms of their safety for the study participants (security parameters are determined in accordance with the direction of the study)			
OUTCOME ON COMPLIANCE WITH THE PRINCIPLE Safety (security) & robustness: POINT TO NOTE			Yes	No	Partially
2	**Governance & accountability** **Transparency & explainability** **Effectiveness (epistemic reliability)**[4]	The responsibility of the Researcher at each stage of the design of the work is delimited, incl. a member of the research team is identified who is responsible for coordinating the use of AI in the study			
		The researcher has a sufficient (level of education is confirmed by relevant documents; practical skills are confirmed by work experience) level of knowledge to work with AI and databases			

---

[4] When researching with Database/ Data Visitation

1	2	3	4	5	6
		The algorithm of actions of the Researcher for the constant (current) analysis of the process of applying AI / DV has been determined			
		The algorithm of actions of the Researcher in case of incorrect operation of AI systems (errors, failures, AI bias, detected contradictions in the analysis of data, etc.) is planned.			
		The researcher has developed understandable mechanisms for terminating research (refusal of a research participant to continue their participation) in case of failures/incorrect operation of AI systems. These mechanisms are known to the study participants.			
		The researcher provides for compliance with all regulatory documents for working with AI (state laws, regulations, codes, etc.)			
		The researcher provides for compliance with advisory documents on working with AI (the research project involves links to relevant documents) and DV, including ethics and guidelines			
		The researcher provides a mechanism for reasonable restriction of access /			

1	2	3	4	5	6
		expansion of access to data during the study and at its end			
		The Researcher provides and explains the system of responsibility (developer, Researcher) for AI actions that harm the research participants, society and nature, and the Researcher himself			
		Investigator provides consent to reuse data and/or publish study data			
		The investigator provides formal consent to conduct research / use their databases of third-party organizations participating in the study			
		The researcher instructed the study participants on working with the AI elements used in the study (if necessary)			
		Conducted training / briefing for members of the research team on working with databases (DV) (if necessary)			
		Participants of the study (in informed consent and orally) were clearly explained what data will be included in the databases (biobanks), what are the conditions for storing and using data (if applicable)			
		A mechanism has been developed for consultations of research participants with			

1	2	3	4	5	6
		the Researcher (on the work of AI systems in the research process)			
		A mechanism is provided for explaining the results of the study using AI (both intermediate and final) for the researcher participants and the research team.			
		The researcher ensures the validity of the data included in the databases (the methodology for obtaining data is standardized, a constant analysis of possible sources of distortion is carried out)			
		The researcher provided for the possibility of data search, their reuse and the possibility of merging data sets			
		The researcher provided for ethically and technically verified handling of any data, especially rare data, control and ethical support of data			
		The researcher provided for the responsibility of the creator and owner of databases for their correctness and compliance with ethical and legal standards, user orientation, and data stability			
		The researcher analyzed the work of the AI systems used in the study for the			

1	2	3	4	5	6
		presence of biases embedded in the programs (discrimination based on racial, national, gender, political, religious, age, cultural, linguistic principles, as well as social and economic status, information about private life, etc. .)			
TOTAL BY COMPLIANCE PRINCIPLE Governance & accountability Transparency & explainability Effectiveness: NEEDED MARK			Yes	No	Partially
3	**Justice/ fairness**	There is an alternative to using AI for participants (if applicable)			
		The researcher identified among the study participants the categories of vulnerable persons who, for various reasons, cannot use / apply AI systems, the rules for replacement and exceptions are determined.			
		An algorithm is provided by the Researcher to respond to AI errors/prejudices in relation to research participants			
		The Researcher provided for informing the participants about the available ways of contacting the Researcher (contact numbers, addresses) in case of complaints and suggestions.			

1	2	3	4	5	6
		The researcher clearly defined and justified the criteria for including / excluding participants in working with AI systems / databases			
OUTCOME ON COMPLIANCE WITH THE PRINCIPAL FAIRNESS: THINGS TO MARK			Yes	No	Partially
RESEARCH OVERALL			Approve	Do not approve	Approve with comments
indicate the comments and the way to control the elimination of comments:					

Appendix 2. Cheklist for RECs/IRBs/IECs reviewing research with AI/DV (examining the completed research)

№	Ethical principle	Evaluation indicator	Evaluation category		
			Compliance in full	Partial match	Lack of conformity / not usable
1	2	3	4	5	6
1	**Safety (security) & robustness**	The AI used was controlled by the developer. Incl. methods of emergency communication of the researcher or research team (hereinafter referred to as the Researcher) with the developer in case of emergency situations are provided.			
		Performed control of the use of AI by the Researcher. Methods of emergency communication of the study participants with the Researcher in case of emergency situations are provided.			
		Conditions have been created to protect the confidentiality of data of study participants			
		Study participant data is encrypted / depersonalized for placement in databases (according to study design and informed consent of participants)			

1	2	3	4	5	6
		The researcher identified and analyzed the risks for the study participants: physical, psychological / mental, informational, financial, legal, ethical, etc. (Underline whatever applicable)			
OUTCOME ON COMPLIANCE WITH THE PRINCIPLE Safety (security) & robustness NECESSARY MARK			Yes	No	Partially
2	**Governance & accountability** **Transparency & explainability** **Effectiveness (epistemic reliability)**[5]	The researcher conducted a periodic analysis of the processes of applying AI / DV			
		The researcher conducted a continuous analysis of the incorrect operation of AI systems (errors, failures, AI bias, detected inconsistencies in data analysis, etc.)			
		The researcher complied with all regulatory documents for working with AI / DV (state laws, regulations, codes, etc.)			
		The researcher complied with the requirements of advisory documents on working with AI (the research project requires links to relevant documents) and DV, including ethics and guidelines			
		A mechanism has been developed to reasonably restrict access / expand access to data obtained during the study.			

---

[5] When researching with Database/ Data Visitation

1	2	3	4	5	6
		Developed and implemented (if necessary) a system of responsibility of the developer, the Researcher for AI actions that harm the research participants, society and nature, and the Researcher himself			
		Investigator obtained permission (informed consent) from study participants to reuse data and/or publish study data			
		The investigator has received formal consent to conduct research / use their databases of third-party organizations participating in the study			
		Authorship is determined when preparing the final result (publications, patents, etc.). All members of the research team are mentioned (for publication). The absence of a conflict of interest is justified.			
		Consultations were held between the research participants and the Researcher (on the work of AI systems in the research process)			
		Explaining research results using AI			
		The validity of the data included in the databases has been ensured (a methodology for obtaining data has been developed, a constant analysis of possible distortions is being carried out)			

1	2	3	4	5	6
		The researcher implemented a mechanism for data search and reuse, as well as the possibility of fundamental compatibility with other data (databases)			
		The researcher implemented ethically and technically correct handling of any data, especially rare ones.			
		(If applicable) When creating databases, the Researcher provides for the responsibility of the creator and owner of databases for the correctness of the data, their compliance with ethical and legal standards, as well as data stability			
TOTAL BY COMPLIANCE PRINCIPLES Governance & accountability Transparency & explainability Effectiveness: IMPORTANT NOTE			Yes	No	Partially
3	**Justice/fairness**	There is an alternative to using AI for participants (if applicable)			
		The researcher identified among the study participants the categories of vulnerable persons who, for various reasons, cannot use / apply AI systems, the rules for replacement and exceptions are determined.			
		The AI algorithms used do not contain prejudice and discrimination based on			

1	2	3	4	5	6
		racial, national, gender, political, religious, age, cultural, linguistic, social, economic reasons, and other reasons			
		An algorithm for the Researcher's response to AI errors/prejudices in relation to research participants has been defined			
		Criteria for inclusion / exclusion of participants when working with AI systems are provided and implemented			
		There is clear information on how and when a participant can withdraw from the study without prejudice to their rights.			
		The Researcher provided for informing the participants about the available ways of contacting the Researcher (contact numbers, addresses) in case of complaints and suggestions.			
OUTCOME ON COMPLIANCE WITH THE PRINCIPLE Justice / fairness: IMPORTANT NOTE			Yes	No	Partially
4	**Trust (trustworthy)**	The rights of study participants are protected (informed consent and timely notification, data confidentiality, the right to prohibit the further use of any data and / or their storage in databases, the right to refuse to participate in the study, the right to refuse the use of AI in the study, etc.)			

<table>
<tr><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th></tr>
<tr><td></td><td></td><td>The researcher conducted a survey / survey of participants to determine the level of their confidence in the use of AI elements in research, as well as on the creation, storage, use of databases</td><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td>Timely clarification (if necessary) and a wide discussion of the results of the study among interested parties are carried out</td><td></td><td></td><td></td></tr>
<tr><td colspan="3">OUTCOME ON COMPLIANCE WITH THE PRINCIPLE
Trust (trustworthy): IMPORTANT NOTE</td><td>Yes</td><td>No</td><td>Partially</td></tr>
<tr><td colspan="3">RESEARCH OVERALL</td><td>Approve</td><td>Do not approve</td><td>Approve with comments</td></tr>
<tr><td colspan="6">Specify the comments (if any) and the way to control the elimination of comments:</td></tr>
</table>

Appendix 3/ Anonymous survey for research participants:

1. *Do you think people's right to benefit from scientific developments can override people's rights to security and privacy?*
2. *Do you think artificial intelligence technologies are adequately represented in health policies?*
3. *Is there a need for training in this field for members of research ethics committees that evaluate AI technologies research?*
4. *Do you think that research ethics committees evaluating artificial intelligence technologies should have similar governance systems in order to have a universal common approach?*
5. *Do you think it is possible to predict the place that artificial intelligence studies will take in human life?*
6. *Do you think it is possible to fully ensure the safety of the data you obtain in your research?*
7. *Who do you think is primarily responsible for ensuring the privacy of the participants in your research?*
8. *What do you think are the primarily responsible persons and/or institutions for the long-term storage of data collected in research using artificial intelligence?*
9. *What do you think about the effectiveness of storing the data obtained in artificial intelligence studies in databases?*
10. *Do you think there is a need to design an informed consent form specific to AI research?*
11. *Did you have difficulties using AI systems (including questionnaires, using various devices, programs, etc.)?*
12. *Did you have the possibility of alternative options instead of using AI (in case of difficulties and corresponding requests/wishes), incl. opportunity to opt out?*

13. *Have you been provided with clarification/advice on the use of AI systems during the research?*
14. *Have the interface / algorithms of using / final results been obtained clear enough (after explanation)?*
15. *Have you been trained for the using of AI systems during the research?*