

Data Protection Impact Assessment for Artificial Intelligence - Practical Considerations

1. Demonstrate accountability through the DPIA for AI

Impact on individual's fundamental rights must be at the core of discussions about AI. It is crucial that organizations properly consider fundamental rights as part of any DPIA relating to AI systems. Data protection impact assessments are ideal opportunities for organizations to demonstrate accountability for decisions made in the process of design, or procurement, of AI systems.

While much of the focus and guidance regarding DPIAs has been on data protection issues, the broader fundamental rights aspect of DPIAs has been less talked about, despite a DPIAs potential to provide effective roadmaps for organizations to identify and control risks posed by AI. A DPIA for AI should consider the inherent challenges of bias, scale, and complexity of this technology and provide coverage for all stages of AI lifecycle. An understanding of the source of the risk, such as the data, people, process, and technique, is needed to effectively mitigate the risks in AI.

With the sudden emergence of AI into the workplace, it is essential that business leaders are equipped with practical approaches to managing the complex world of artificial intelligence, whilst unlocking the considerable value that can be created by it. This whitepaper highlights key practical considerations for conducting data protection impact assessments (DPIAs) for AI.

Here are 4 building blocks for a DPIA for AI:

The outcome of a DPIA should enable organizations to minimize the risks of processing by putting in place effective policies, procedures, and measures.

To trust organizations they interact with, regulators, business partners and individuals need to have the comfort that AI risks are managed.

AI risk management can enhance organizations reputation on the market and give them a competitive edge, helping businesses to thrive and grow.

AI can involve several processing operations that are themselves likely to result in a high risk for the right and freedoms of individuals, such as use of data matching, invisible processing, and tracking of location or behaviour, evaluation or scoring, systematic monitoring, large-scale processing.

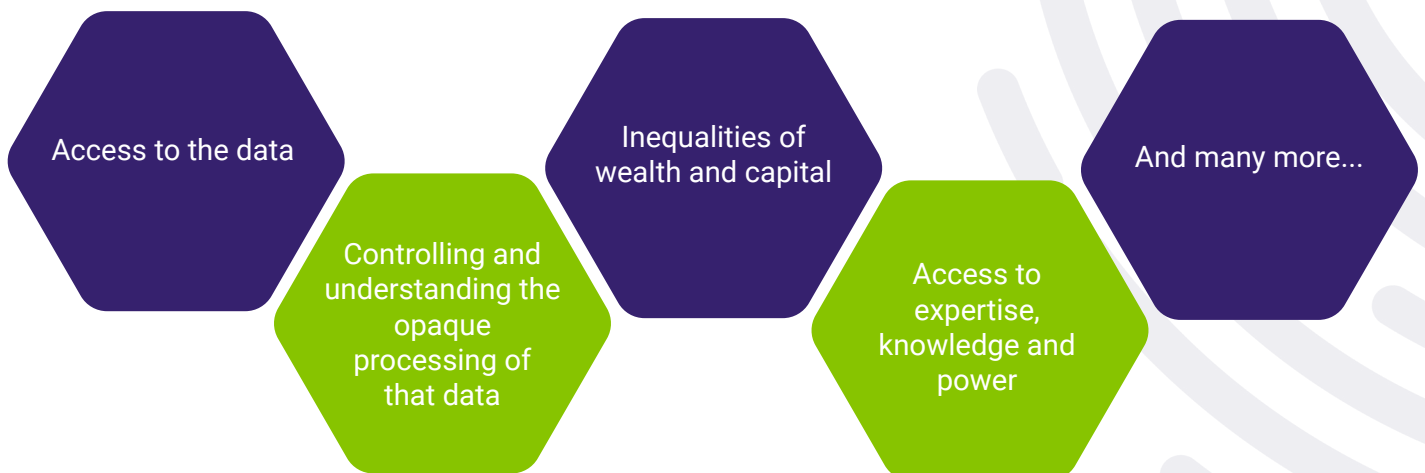
2. Fundamental rights as part of the DPIA for AI

Fundamental rights are typically described in broad terms and therefore entail general formulations, making it difficult for organizations to determine whether their AI system creates risks to one of those rights. Besides that, the concept of fundamental rights is often perceived as abstract, making it potentially challenging for organizations to concretely ascertain how their operations might pose risks to these rights, and how they may apply to, and potentially impact, the design and deployment of AI system and processes.

For easier reference academics Heleen Janssen, Michelle Seng Ah Lee and Jatinder Singh propose to group fundamental rights into clusters:

- Privacy rights cluster, comprising the right to private life, personal autonomy, the sanctity of the home, physical privacy, communication secrecy, data protection, the development of one's identity, including the right to hold a conviction or a belief, or not to hold any belief.
- Non-discrimination rights cluster such as the right to equal treatment and the prohibition of discrimination.
- Freedom rights cluster comprising freedom of expression, freedom to receive and to impart information, including the expression of convictions and beliefs or religious freedom, the right to assembly or voting rights.
- Procedural rights cluster, which includes the right to access to a court, to an effective legal remedy and the right to a fair trial and
- Social, economic, and cultural rights cluster such as the right of access to healthcare, affordable housing, education, or social benefits.

The technological context is characterized by systemic power asymmetries inherent to the current data processing practices of digital organizations. The asymmetries commonly originate from an unequal distribution between those organizations and individuals, in terms of;



It has been therefore, recognized that private organization activities may harm consumer, employee, or patient fundamental rights.

According to Subject Matter Experts, when considering the AI impact on individuals it is important to look at:

Allocative harms - where an individual is made worse off in terms of the resources available to them such as a lower salary for the same work or denied the opportunity for a job interview or credit based on, for example, gender. Currently, within the job market, algorithms are widely deployed in support of the hiring process with a tendency to centre on the employer perspective.



Amazon recruitment algorithm (Dastin 2018)

In 2014, Amazon started developing an internal AI system to streamline their recruitment process. Using the CVs of past applicants as training data, the system would analyse incoming CVs and rate the candidates for further evaluation. Very quickly though, the system was found to rate candidates for technical jobs in a gender-biased way. The system was found to penalize any CVs which indicated the applicant to be a woman. This included mentions of attending things like a women's chess club, or an all-women college. Amazon reportedly attempted to debias the system but ended up scrapping the whole project instead. The system was never used in actual recruiting

Representational harms such as denigration, stereotyping, misrecognizing, denigrating, meaning, leading to undermining human dignity.

Google Image search (2015)

Back in 2015, software engineer Jacky Alciné pointed out that the image recognition algorithms in Google Photos were classifying his black friends as "gorillas." Google said it was "appalled" at the mistake, apologized to Alciné. Google then blocked its image recognition algorithms from identifying gorillas preferring to limit the service rather than risk another miscategorization.

3. Types of assessments to inform the DPIA for AI

In addition to conducting a DPIA for AI, organisations may undertake other types of impact assessments either on a mandatory or voluntarily basis. Such assessments could be for instance:



Voluntarily Algorithmic Impact Assessments (AIA)

To determine the algorithmic harm (i.e., biased hiring algorithm, the unexplained credit denial, and the unsafe medical AI) of an automated decision-system. An AIA-based would require the creator of an algorithmic system to assess its potential socially harmful impacts before implementation and create documentation that can be used later for accountability and future policy development.

Datasheet

Which describes how machine learning models may perform under different conditions, and the context behind the datasets they are trained on, which may help inform an impact assessment.

Mandatory Equality Impact Assessments (EqIA)

Aiming to prevent discrimination against individuals who are members of a protected category basis on personal characteristics such as race, religion or belief, disability, sex, gender reassignment, sexual orientation, age, marriage or civil partnership, pregnancy, and maternity.

Model cards

For model reporting as a framework for machine learning performance characteristics that report details of the datasets used to train and test machine learning models. Model cards are accompanying trained machine learning models that provide benchmarked evaluation in a variety of conditions, such as across different cultural, demographic, or phenotypic groups (i.e., race, geographic location, sex, and intersectional groups (i.e., age and race, or sex and skin type) that are relevant to the intended application domains.

4. DPIA for AI steps

By applying the existing guidance and what we have learned from the GDPR compliance mechanisms we can integrate Fundamental Rights Impact Assessments (FRIAs) with DPIAs in four steps, as follows:

Step 1 - Document the processing activities of the AI systems

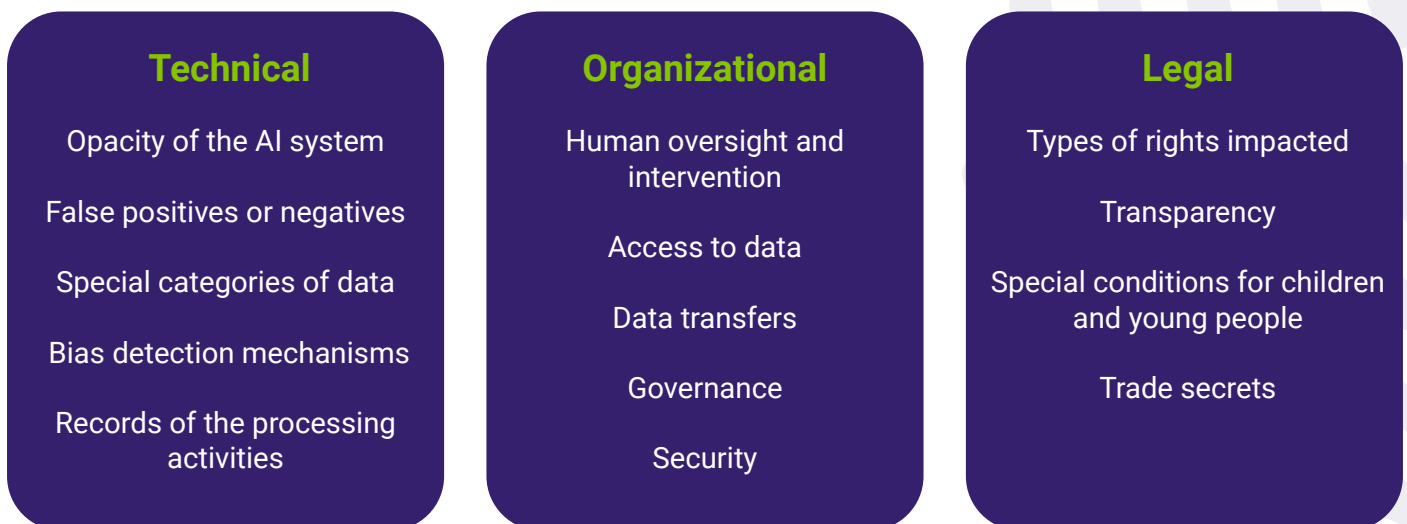
It can be difficult to describe the processing activity of AI systems, particularly when they involve complex models and data sources. However, such a description is necessary as part of a DPIA for AI. In some cases, although it is not a legal requirement, it may be good practice to maintain two versions of an assessment, with:

- The first presenting a thorough technical description for specialist audiences; and
- The second containing a more high-level description of the processing and explaining the logic of how the personal data inputs relate to the outputs affecting individuals (this may also support in fulfilling obligations to explain AI decisions to individuals).

Step 2- Identify and assess risks to fundamental rights of individuals

Organizations must assess risks to rights of individuals throughout all stages of the system's lifecycle, from commissioning, design, operation, and investigation. Continuous examining and testing of all technical and organizational processes of the AI system will help organizations verify if the system satisfies the organization's objectives and legitimate interests and crucially, if such objectives are balanced against the rights and interests of other stakeholders, including data subjects. The standard formula for assessment is risk = impact or likelihood of occurrence. An accurate assessment of risk assists better places the organization to adjust the system to account for risks, to help ensure that mitigation measures are appropriate.

Example of factors indicating a high risk to fundamental rights could include:



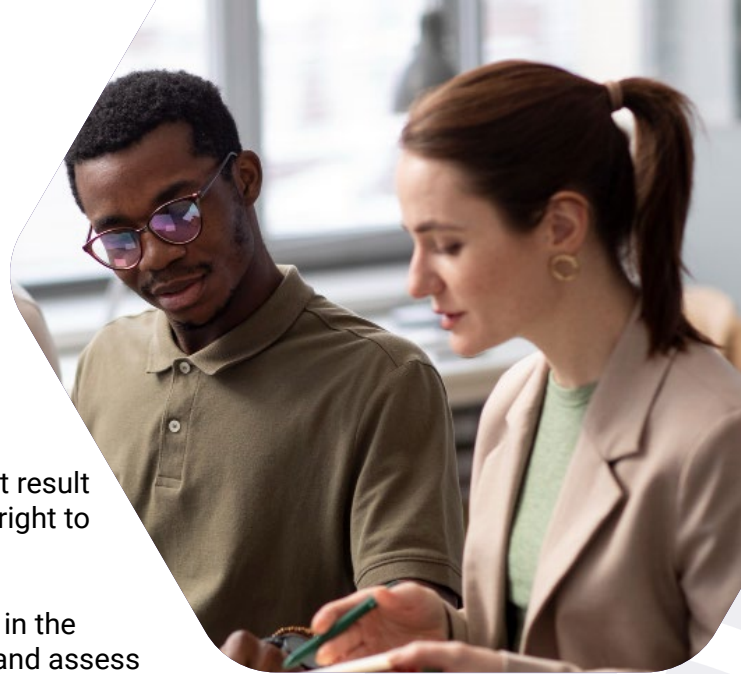
The risks to rights might not be identified at the time of analyzing, examining, and testing the system, but rather through the system's operation, its future impacts, and the way it is interrogated, and managed. This is the reason why organizations should conduct DPIAs for AI not only at the design stages, but also regularly.

Considering the likelihood of occurrence of these risks, an aggregated risk score must be developed. This score must result in a determination of 'high', 'medium' or 'low' risk for each right to be infringed by the system.

The risks to rights might not be identified at the time of analyzing, examining, and testing the system, but rather through the system's operation, its future impacts, and the way it is interrogated, and managed. This is the reason why organizations should conduct DPIAs for AI not only at the design stages, but also regularly.

Considering the likelihood of occurrence of these risks, an aggregated risk score must be developed. This score must result in a determination of 'high', 'medium' or 'low' risk for each right to be infringed by the system.

Organizations should also consider stakeholders involved in the development or deployment of AI system's supply chain, and assess whether any activities of any suppliers and contractors could contribute to potential risks to fundamental rights.



Step 3 Balance organizational interests with individual's fundamental rights (necessity and proportionality)

In this step organizations must analyze and balance incompatibilities between organizational benefits and interests, and potential harms and ultimately identify the extent to which trade-offs are necessary, and proportional. Such an approach requires organizations to consider all elements in context. Where a proper balance cannot be found, the use of an AI system may create unacceptable risks and result in rights violations, although worth noting that these risks might potentially be addressed by employing a range of mitigating measures.

Necessity

Organizations (especially private) are often deploying AI systems to achieve legitimate business interests (commercial or economic interests, business-processing efficiencies, time, and cost-savings, etc.). In this step of the assessment organizations must explain why deploying the system is suitable for the legitimate purposes they pursue. They must answer questions about the necessity of a system in a way that must be purely evidence based. Where uncertainty exists over a system's suitability for achieving a particular purpose, that system becomes problematic.

Additionally, organizations must document if the use of their AI system is necessary to achieve the purposes, and whether other, less intrusive means exist to accomplish the purposes. In some situations, an organization might have to find an appropriate trade-off where their purpose can be accomplished through approaches that might be less effective or efficient, but that are also less rights invasive.

Systems are generally regarded as not reaching the necessity threshold whenever they are inaccurate (with e.g., high-rate false positives), or where they do not help an organization to achieve their purpose, for instance, if no cost-saving will be achieved.

Proportionality

It is important to note that typical commercial purposes (i.e., to increase profit, to reduce costs, to develop strategies to improve the organization's market position, to research into customers behavior to identify new markets, to improve service delivery, achieve faster, cheaper and more efficient internal processing, etc.) will often NOT by themselves be sufficient to justify the use of AI systems, specifically where risks to fundamental rights were identified in Step 2. If high risks have been identified, it is strongly recommended that alternative methods to achieve purposes are sought.

Step 4 Mitigate risks

Example 1 The Automotive Industry: Self-Driving Cars

"x,y,z, organisations poor safety culture, faulty technology, poor attention to the interface between its system and its drivers, and careless treatment of its employees predictably led to a high risk of car crashes."

Example 2 AI and Recruitment: Hiring Employees

In 2014, a famous brand infamously built a recruitment program to automate its search for talent. It trained the program on resumes submitted over a 10-year period. In 2015, the team noticed that the AI system was biased against women. It downgraded resumes that included the word "women's" and penalized graduates of all-women's colleges. The AI system had replicated the bias in the data set, in which most successful job applicants over the past ten years had been male. This is a classic illustration of the problem of "garbage in, garbage out," in which the quality of machine learning models is "only as good as the quality of [training] data."

Example 3 AI and Public Health: Access to Care and Prescriptions

A risk-prediction algorithm widely used in the healthcare sector erroneously overlooked high-risk black patients, reducing the number of black patients identified for the high-risk care management programs by more than half. They found that the algorithm was not looking directly at health needs but at health costs. Since at a given number of chronic illnesses black patients generated lower costs than white patients, the result of both healthcare inequities and social factors, they were being incorrectly labelled by the algorithm as having lower health risks.

The degree of mitigatory intervention depends on circumstances. The outcome must be a diminishing of concerns while strengthening individuals and society trust in an organization's use of the AI systems. On top, effective mitigating measures can help organizations comply with legal and accountability obligations.

Example of mitigating measures are:

- Prevent continuous data capture
- Minimise terms for data retention
- Consider on-device processing rather than centralized in-company processing
- Give individuals meaningful tools to manage processing
- Avoid third-party data sharing for commercial purposes
- Use AI systems that can be reviewed
- Treat all personal data as special categories of data
- Implement meaningful human intervention
- Build in mechanisms for GDPR data subject requests
- Regularly evaluate the entire AI system to identify non-compliant areas

AI systems can fail in multiple ways. There are different examples on how a risk can be framed.



Conclusion

The importance of examining the impacts of AI systems on fundamental rights impact assessment will become part of the EU policy. To assist organizations with the development of fundamental rights compliant AI, we suggest that organizations should put processes in place to assess in detail the need for a data protection impact assessment, including an assessment of the necessity and proportionality of the processing operations in relation to their purpose.

At Wrangu we have years of experience working on complex issues in privacy and data protection, compliance, risk, and security. We are leveraging this experience to prepare for the next wave of technological revolution. [Contact us](#) today or [visit our website](#) today to find out more.


Author



Petruta Pirvan is a Senior Principal Privacy Consultant at Wrangu specializing in data protection implementation, management and educational programmes. She is a member of the European AI Alliance and she holds the IAPP CIPP/E, CIPP/US, CIPM and FIP certifications.

Contact Us


Wrangu BV


 De Entree 99-197
1101 HE Amsterdam
Netherlands

 +31 (0)20 399 9842

 hello@wrangu.com

Wrangu UK Limited

 Level 1, 6 Bevis Marks
London EC3A 7HL
United Kingdom

 +44 (0)20 4583 1177

 hello@wrangu.com