

2025년 GPA 정책포럼

AI 위험성과 개인정보 영향평가의 과제

일시

2025. 9. 19(금) 15:00

장소

그랜드하얏트서울(2층)
비즈니스센터 룸9

주관

한국소비자연맹

2025 글로벌프라이버시총회(GPA) 정책포럼
AI 위험성과 개인정보 영향평가의 과제
(Challenges in Addressing AI Risks and Conducting DPIAs)

I. 개요

- 일시 : 2025년 9월 19일(금) 오후 15:00
- 장소 : 그랜드하얏트서울 비즈니스센터 룸9
- 주관 : 사단법인 한국소비자연맹

II. 프로그램

사회 이 서 윤(판사, 사법연수원 교수)

세션	일정	내용	
세션 I	개회식		
	개회		
	인사말씀	강 정 화	한국소비자연맹 회장
세션 II	주제발표		
	발제1	한국의 개인정보 영향평가와 AI인권 영향평가 현황	
		김 병 육	변호사, 민변 디지털정보위원회 위원
	발제2	크로아티아 및 EU의 개인정보 영향평가와 AI인권 영향평가 현황	
		아나마리야 르라디니치	크로아티아 개인정보보호청
세션 III	토론		
	패널토론	김보라미	법무법인 디케 변호사
		구 본 석	변호사, 참여연대 공익법센터 운영위원
		정 지 연	한국소비자연맹 사무총장
		최 경 진	인공지능법학회 회장
		하비에르 루이즈	포용적통상정책센터
	플로어토론		

2025 글로벌프라이버시총회(GPA) 정책포럼

AI 위험성과 개인정보 영향평가의 과제

(Challenges in Addressing AI Risks and Conducting DPIAs)

강정화
(한국소비자연맹 회장)

안녕하십니까.

한국소비자연맹 회장 강정화입니다.

오늘 '2025년 GPA 정책포럼 서울' 행사의 일환으로 'AI 위험성과 개인정보 영향평가의 과제'를 주제로 작지만 소중한 토론의 기회를 갖게 되어 기쁘게 생각합니다.

2025 GPA에서는 AI 시대 개인정보에 대해 여러 분야에서 다양한 논의가 있었고 민간·시민 단체의 역할에 대해 중요하게 논의되었습니다.

개인정보 영향평가는 정보주체의 개인정보 침해 위험을 줄이기 위해, AI 인권영향평가는 AI의 개발과 활용으로 인한 인권 침해 및 차별이 발생하지 않기 위한 중요한 도구입니다. 현재는 우리 사회 일부에서만 이루어지고 있고, 일반 소비자에게 아직 생소한 제도이지만 최근 발생한 일련의 대형 개인정보 침해사고로 피해를 겪은 한국 소비자들에게 이러한 평가의 중요성을 알리고 평가에 적극적인 관심을 갖도록 알려야 한다고 생각합니다.

오늘 토론회를 통해 각국의 현황을 공유함으로써 더 나은 제도로 발전할 수 있게 되기를 바랍니다. 프라이버시는 모든 소비자의 기본적 권리로 보호받아야 하고, 권리를 지키기 위한 적절한 기술 통제를 고려해야 할 것입니다. 기술을 위한 사회가 아닌 사람을 위한 사회로 나아가는 것이 AI 시대를 맞이하는 소비자들의 요구입니다.

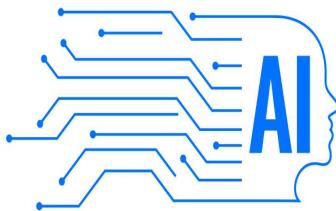
토론회를 위해 귀중한 시간을 내어 주제발표를 맡아 주신 김병욱 님, 아나마리야 블라디니치 님과 사회자, 토론자 여러분께 감사의 말씀을 드립니다.

또한 이 토론회를 위해 애써주신 진보네트워크센터 오병일 대표님을 비롯한 민주사회를 위한 변호사 모임, 참여연대, 언론개혁시민연대 관계자 여러분들께도 감사의 말씀을 드립니다.

2025 글로벌프라이버시총회(GPA) 정책포럼

AI 위험성과 개인정보 영향평가의 과제

(Challenges in Addressing AI Risks and Conducting DPIAs)



발 제 1

한국의 개인정보 영향평가와 AI인권 영향평가 현황

김 병 욱

(변호사, 민변 디지털정보위원회 위원)

Current status of Privacy Impact Assessment and AI Human Rights Impact Assessment in Korea

Kim, Byung wook
Attorney, MINBYUN Digital
information committee

01 Risks of AI

[What is Artificial Intelligence?](#)



- Owing to its broad and pervasive nature, artificial intelligence poses risks to virtually the entire spectrum of human rights and fundamental freedoms.
 - Unauthorized collection of personal information and use without user control
 - Opaque and autonomous process of parameter generation (black-box algorithm)
 - Potential threats to human life, physical safety, and privacy
- Considering AI's autonomy, opacity, and risks, after-the-fact sanctions or remedies are not enough

02 Human Rights Impact Assessment(HRIA) for AI

- “Encourages States and, where applicable, business enterprises to conduct human rights due diligence throughout the life cycle of the artificial intelligence systems they design, develop, deploy or sell or obtain and operate”

<United Nations, “Right to privacy in the digital age”, 2021>



Human Rights Impact Assessment(HRIA) for AI

- “States must conduct human rights impact assessment for public institutions and private companies, taking into account the possibility and extent of human rights violations and discrimination, the number of affected parties, and the volume of data used in the development and deployment of artificial intelligence.”

< The National Human Rights Commission of Korea, Human Rights Guidelines on the Development and Use of Artificial Intelligence, 2022. 4.>



Current Status of HRIA for AI in Korea

Basic Act on the Development of Artificial Intelligence and Establishment of Trust
(Enforced January 22, 2026)

Article 35 (Impact Assessment of High-Impact AI)

- 1) AI business operators providing high-impact AI products or services shall proactively assess the impact on basic human rights (hereinafter “impact assessment”) prior to implementation.

Current Status of HRIA for AI in Korea

Social Impact Assessments of intelligent information Services
(Article 56, Framework Act on intelligent informatization)

1. Safety and reliability of intelligent information services, etc;
2. Impacts on the information culture, such as closing the digital divide, protection of privacy and ethics for the intelligent information society;
3. Impacts on the society and the economy, such as employment, labor, fair trade, industrial structure, rights and interests of users, etc.;
4. Impacts on information protection;
5. Other impacts of intelligent information services, etc. on the society, economy, culture and citizens's daily lives

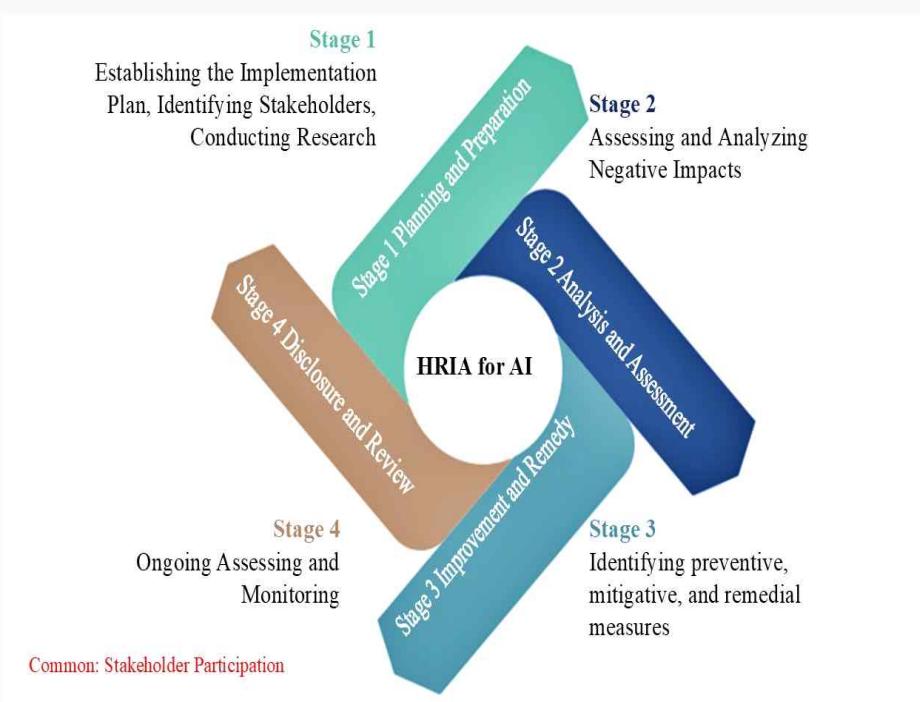
Technology Assessment
(Article 14, Framework Act on Science and Technology)

1. Impact of the relevant technology on the enhancement of benefits to citizens and on the development of relevant industries;
2. Impact of the new science and technology on the economy society, culture, ethics and the environments;
3. Measures to prevent adverse effects of the relevant technology, where the relevant technology has potential adverse effect;
4. Impact of the nature and ripple effects of the relevant technology on characteristics, such as gender.

Current Status of HRIA for AI in Korea

Human Rights Impact Assessment Tool for AI(NHRCK)	
Subject	High-risk artificial intelligence + AI introduced by public institutions
Timing	Prior assessment (when the concept for AI technology development or introduction is concrete) + regular/post-assessment if necessary
Conducted by	Independent organization, either internal or external (Departments in charge of AI ethics, human rights management, ESG management, or a third-party organization with independence and expertise in human rights and AI)

Current Status of HRIA for AI in Korea



Current Status of HRIA for AI in Korea

- [Stage 2-Analysis and Assessment]
A. Analysis and Assessment Related to AI Technology
 - (1) Personal Information Protection
 - (2) Data Management
 - (3) Algorithm Performance and Reliability
 - (4) Non-Discrimination
 - (5) Explainability and Transparency
 - (6) Degree of Automation and Human Intervention
 - (7) Security
 - (8) Accessibility
 - (9) License

(5) Explainability and Transparency

Q2-1-13. Is related information (e.g., logs of decision details or all changes to the system) recorded to track the factors involved in making specific decisions (outputs) by the AI system?

Yes Needs Improvement No No Information Not Applicable

Explanation: ()

Q2-1-14. Can the reasons or basis for specific decisions (outputs) made by the AI system be easily explained to users or affected stakeholders?

Privacy Impact Assessment(PIA) for AI

- Risk to the right to informational self-determination is a key AI risk
- Need for complementarity and cooperation between data protection and AI regulatory frameworks

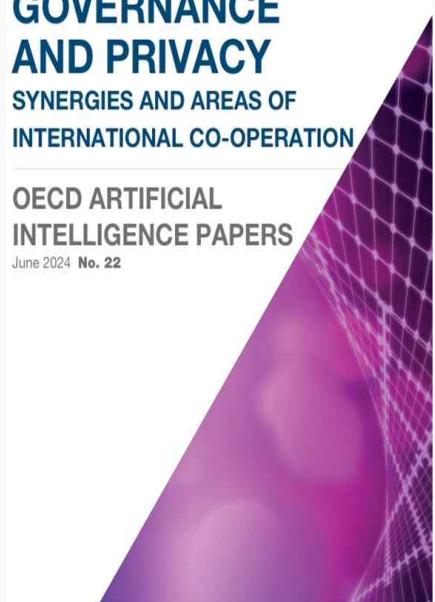
“This analysis concludes that, despite challenges, AI’s innovative, technological and regulatory developments are mainly compatible with, and can even reinforce, privacy and personal data protection rules.”

<AI, DATA GOVERNANCE AND PRIVACY (2024), OECD Publishing>

**AI, DATA
GOVERNANCE
AND PRIVACY**
SYNERGIES AND AREAS OF
INTERNATIONAL CO-OPERATION

OECD ARTIFICIAL
INTELLIGENCE PAPERS

June 2024 No. 22



Current Status of PIA in Korea

PERSONAL INFORMATION PROTECTION ACT

Article 33 (Privacy Impact Assessment)

(1) Where there is a risk of a personal information breach of data subjects due to the operation of personal information files meeting the criteria prescribed by Presidential Decree, **the head of a public institution** shall conduct an assessment to analyze risk factors and to improve them (hereinafter referred to as “privacy impact assessment”)

- Mandatory impact assessment applies only to public institutions

Current Status of PIA in Korea

Evaluation Area	Sub-Area	Main Evaluation Items (Guidelines)
Artificial Intelligence (AI)	AI System Training and Development	<ul style="list-style-type: none">- Ensure the lawfulness of personal information processing- Prevent unnecessary collection of sensitive information- Clarify retention periods for training data- Minimize leakage or exposure of personal information due to AI vulnerabilities or attacks
Artificial Intelligence (AI)	AI System Operation and Management	<ul style="list-style-type: none">- Clarify responsibilities among AI development and operation entities- Ensure transparency in personal information processing- Provide users with guidelines for the permitted use of generative AI services- Establish and implement measures to safeguard data subject rights

▲Key revisions in PIA [PIPC]

Current Status of PIA in Korea

Privacy Impact Assessment

- No provisions for stakeholder participation in procedures
- No regulations on disclosure or verification for private sector assessments
- Impact assessment timing limited to pre-deployment or system changes, hindering continuous risk management
- Unclear authority of PIPC on obligation to reflect results and no sanctions for non-compliance

THANK YOU

 bwkim@dooyul.com  02-3472-2711

2025 글로벌프라이버시총회(GPA) 정책포럼

AI 위험성과 개인정보 영향평가의 과제

(Challenges in Addressing AI Risks and Conducting DPIAs)



발 제 ॥

크로아티아 및 EU의 개인정보 영향평가와 AI인권 영향평가 현황

아나마리야 블라디니치
(크로아티아 개인정보보호청)

DATA PROTECTION IMPACT & FUNDAMENTAL RIGHTS IMPACT ASSESSMENTS IN EUROPE



Anamarija Mladinić, Head of Sector for EU , International Cooperation and Legal Affairs, Croatian Personal Data Protection Agency
Vice-Chair of Committee of Convention 108, Council of Europe

Croatian Personal Data Protection Agency



Croatia: rich culture, crystal-clear sea, football passion, 1,000+ islands... and top-notch data protection

LEGAL FRAMEWORK

✓ Article 8 "Protection of personal data" of the Charter of Fundamental Rights of the European Union states:

"1. Everyone has the right to the protection of personal data concerning him or her.

2. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.

3. Compliance with these rules shall be subject to control by an independent authority."

✓ The protection of personal data in the Republic of Croatia is a constitutional category:

Article 37

"Everyone is guaranteed the security and confidentiality of personal data. Without the consent of the data subject, personal data may be collected, processed, and used only under conditions specified by law.

The law shall regulate data protection and the supervision of the operation of information systems in the state. The use of personal data contrary to the established purpose of their collection is prohibited".

- Party to Convention 108 and 108+ : First legally binding international instrument in the field of data protection safeguarding individuals against risks and violations arising from the processing of personal data
- Open for accession globally, making it a worldwide standard

Croatian Personal Data Protection Agency

State or International Organisation	Signature	Ratification	Entry into Force
Members of Council of Europe			
Albania	28/01/2022	22/07/2022	
Andorra	28/01/2019	18/10/2022	
Armenia	02/10/2019	25/01/2022	
Austria	10/10/2018	13/07/2022	
Bosnia and Herzegovina	02/07/2020	07/07/2023	
Bulgaria	10/10/2018	10/12/2019	
Croatia	22/03/2019	18/12/2019	
Cyprus	09/01/2019	21/09/2020	
Estonia	10/10/2018	16/09/2020	
Finland	10/10/2018	10/12/2020	
France	10/10/2018	27/03/2023	
Germany	10/10/2018	05/10/2021	
Greece	06/09/2019	05/03/2025	
Hungary	09/01/2019	19/10/2023	
Iceland	21/11/2018	20/01/2023	
Italy	05/03/2019	08/07/2021	
Liechtenstein	07/12/2020	17/05/2023	
Lithuania	10/10/2018	23/01/2020	
Malta	02/07/2020	02/11/2020	

Monaco	10/10/2018	06/03/2025	
North Macedonia	05/12/2019	26/11/2021	
Poland	16/05/2019	10/06/2020	
Portugal	10/10/2018	18/10/2023	
Romania	26/06/2020	09/03/2022	
San Marino	16/07/2019	16/11/2023	
Serbia	22/11/2019	26/05/2020	
Slovak Republic	17/12/2019	15/06/2023	
Slovenia	16/05/2019	20/06/2023	
Spain	10/10/2018	28/01/2021	D,
Switzerland	21/11/2019	07/09/2023	

Non-Members of Council of Europe

Argentina	19/09/2019	17/04/2023	
Mauritius	04/09/2020	04/09/2020	
Uruguay	10/10/2018	05/08/2021	

Total number of signatures not followed by ratifications

13

Total number of ratifications/accessions

33

Parties to the Convention 108+

Croatian Personal Data Protection Agency

GENERAL DATA PROTECTION REGULATION (GDPR)



Croatian Personal Data Protection Agency

OLIVIA
Virtual assistant for GDPR compliance

[GDPR topics](#) [About us](#) [About ARC2 project](#)

About OLIVIA

In today's digital world, the significance of data protection is steadily on the rise, and it is paramount for SMEs to give it the priority it deserves.

We are the ARC2 Consortium, consisting of Croatian and Italian data protection authorities, the Faculty of Organization and Informatics – University of Zagreb, Vilje University Brusel, and the University of Florence.

Our mission is to empower Croatian and Italian SMEs to achieve compliance with data protection regulations. That's why we've introduced Olivia – a tool designed to make GDPR compliance easier than ever before!

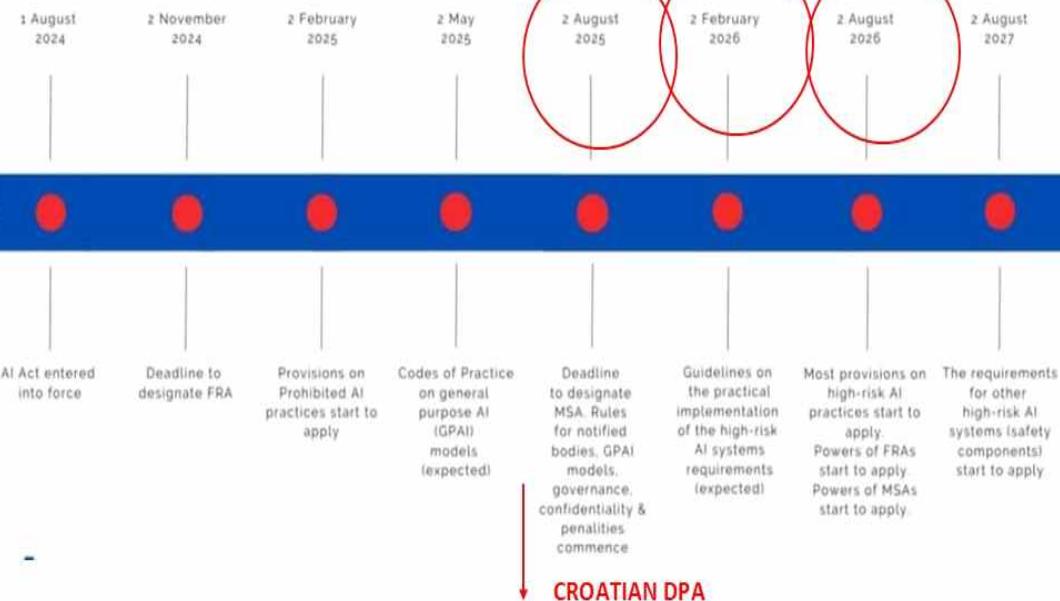
INNOVATIVE
Created in partnership, OLIVIA is enabling innovative approach to GDPR compliance.

FREE
OLIVIA is easy-to-use open tool freely accessible for all interested users.

EDUCATIONAL
OLIVIA streamlines access to knowledge framework, rules and procedures.

PRACTICAL
OLIVIA's practical side helps SMEs in full GDPR implementation.

<https://olivia-gdpr-arc.eu/en>



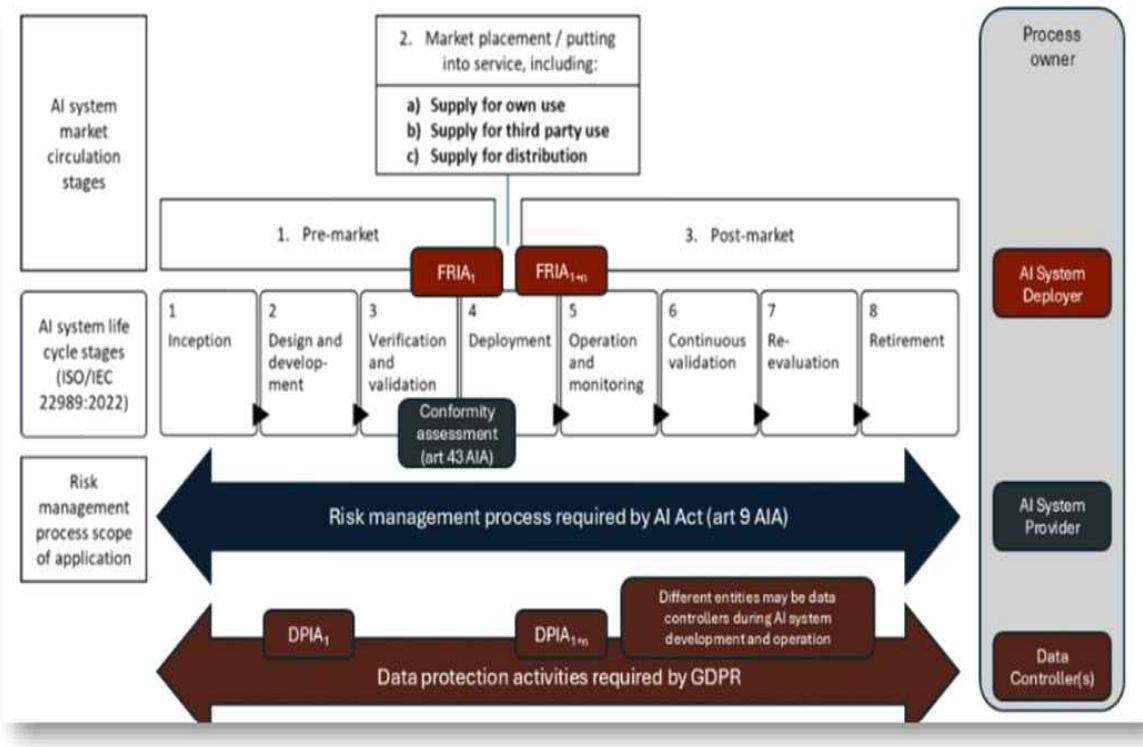
Powers of authorities responsible for the protection of fundamental rights (Article 77)

Public authorities that monitor or enforce compliance with obligations under Union law on the protection of fundamental rights, with regard to the use of high-risk AI systems listed in Annex III, shall be empowered to request and access any documentation prepared or maintained under this Regulation where such access is necessary for the effective fulfilment of their mandates within the limits of their jurisdiction

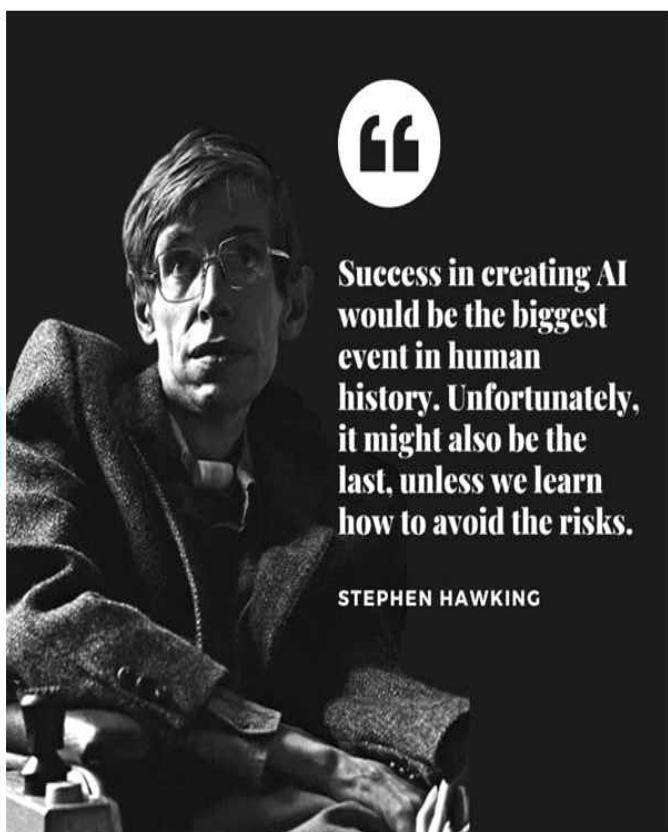
Croatian Personal Data Protection Agency

Key challenge: ASSESSING AND MITIGATING RISKS TO PRIVACY AND FUNDAMENTAL RIGHTS

ARTICLE 32 OF GDPR (TECHNICAL AND ORGANISATIONAL MEASURES), ARTICLE 35 (DPIA), AI ACT (ARTICLES 9, 27, 43)



Croatian Personal Data Protection Agency



Croatian Personal Data Protection Agency
azop
 Agencija za zaštitu osobnih podataka

AI privacy risks and mitigations

https://www.edpb.europa.eu/our-work-tools/our-documents/support-pool-experts-projects/ai-privacy-risks-mitigations-large_en

[Privacy Enhancing Technologies \(synthetic data, LLMs\) – Council of Europe](#)

FRIA methodology

https://www.dpdenxarxa.cat/pluginfile.php/2468/mod_folder/content/0/FRIA_en_def.pdf

<https://azop.hr/metodologija-za-provedbu-procjene-ucinka-na-ljudska-prava-fria/>

Croatian Personal Data Protection Agency

ARTICLE 35 OF THE GDPR

Where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is **likely to result in a high risk to the rights and freedoms of natural persons**, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data. A single assessment may address a set of similar processing operations that present similar high risks.

Recital 75

Risks to the Rights and Freedoms of Natural Persons

The risk to the rights and freedoms of natural persons, **of varying likelihood and severity, may result from personal data processing which could lead to physical, material or non-material damage, in particular: where the processing may give rise to discrimination, identity theft or fraud, financial loss, damage to the reputation, loss of confidentiality of personal data protected by professional secrecy, unauthorised reversal of pseudonymisation, or any other significant economic or social disadvantage; where data subjects might be deprived of their rights and freedoms or prevented from exercising control over their personal data;where personal aspects are evaluated, in particular analysing or predicting aspects concerning performance at work, economic situation, health, personal preferences or interests, reliability or behaviour, location or movements, in order to create or use personal profiles; where personal data of vulnerable natural persons, in particular of children, are processed; or where processing involves a large amount of personal data and affects a large number of data subjects**

DATA PROTECTION IMPACT ASSESSMENT

-template-

INFORMATION ABOUT DATA CONTROLLER

Name	Chatbot XY
Data Protection Officer	John Doe
Contact details	John.doe@chatbot.com

STEP 1: IDENTIFY THE NEED FOR A DPIA

Why do you think you need to carry out DPIA?

The described chatbot project involves systematic and large-scale processing of personal data, which creates elevated risks for individuals' rights and freedoms. Based on GDPR Article 35, a DPIA is required when processing is likely to result in a high risk, and several risk triggers are clearly present.

Describe the scope of the processing: what is the nature of the data, and does it include special categories of data or the processing of personal data relating to criminal convictions and offenses? For what period will you store the data? How many individuals are concerned, and what is the territorial scope covered?

Nature of processing

The purpose of the processing is to develop and continuously improve chatbot XY. Data are collected from three sources: publicly available information, licensed datasets, and user interactions together with trainer feedback. The data are used to train and fine-tune AI models, to improve the quality of responses, and to ensure system security. All data are stored in encrypted form and separated according to type, such as raw data, annotations, and models. Data are deleted through automated log deletion, anonymization or pseudonymization of training sets, and erasure upon a user's request. Personal data are not shared with third parties, except with trusted processors such as cloud providers under appropriate agreements. Processing operations that may present a high risk include large-scale use of public data, possible inclusion of special categories of data, automated decision-making, reuse of user conversations, and cross-border transfers.

Scope of processing

The processing includes user queries, chatbot responses, technical metadata, and trainer annotations. Special categories of data or data relating to criminal convictions are not intentionally collected, and filters are applied to remove them if detected. System logs are retained for up to 90 days, training data for up to 18 months, and model artifacts for 18–24 months. The number of affected data subjects is estimated at between 100,000 and 1,000,000 users annually. The territorial scope of the processing is primarily within the EEA, with possible transfers outside the EEA subject to appropriate safeguards.

STEP 2: DESCRIBE DATA PROCESSING ACTIVITIES

Describe the nature of the processing. How will you collect, use, store, and delete the data? What is the source of the data? Will you share the data with anyone? For which types of processing is there a likelihood of a high risk to the protection of personal data?

To develop and continuously improve the XY chatbot, we train the AI model on large amounts of data, which also includes personal data. We collect data from three sources: Data publicly available on the internet; Licensed datasets; Information from our users or algorithm trainers. In order to improve the responses provided by our chatbot, we employ staff who give feedback on how the chatbot reacts to queries and what kind of answers it should provide in certain situations. In addition, to train our AI models, we also use conversations between users and the chatbot, noting that these data go through a filtering process, i.e., the removal of personal data.

We do not share personal data with third parties. In the case of training AI algorithms on data from publicly available sources, there are certain high risks for data subjects' rights to the protection of personal data, which are addressed in the continuation of this form. We retain personal data only for as long as it is necessary to fulfill the purpose for which they were collected.

Describe the context of the processing: what is the nature of your relationship with the individuals? How much control will individuals have over the processing? Would they expect you to use their data in this way? Does the processing involve children's data or other vulnerable groups? Have there been any previous concerns about this type of processing or security shortcomings? At the time of processing, are there any issues of public interest that you should take into account? Is there an approved code of conduct or certification scheme in place within your organization?

Our relationship with individuals is that of a service provider offering an AI chatbot. Users interact voluntarily and have control through GDPR rights such as access, erasure, and opting out of training use.

Users can reasonably expect their data to be used to provide responses and improve the chatbot when transparently informed. The processing is not intended to include children or other vulnerable groups, and filters are applied to prevent this.

Previous concerns in the AI field relate to bias, transparency, and security, which we mitigate through risk assessments and privacy-by-design. Responsible AI use is a matter of public interest, and we reflect this in our governance.

Our organization follows internal data protection policies and seeks alignment with approved codes of conduct or certification schemes.

Describe the purposes of the processing: what do you want to achieve? What is the intended effect on individuals? What are the benefits of the processing – for you and more broadly?

The purpose of the processing is to develop and improve the chatbot so it can provide accurate, reliable, and context-aware responses to users.

The intended effect on individuals is to make their interactions faster, easier, and more efficient by offering instant assistance and useful information.

The benefits include better product performance, higher user satisfaction, and increased competitiveness. The broader benefits include promoting responsible AI use, improving access to information, and supporting digital innovation in society.

4 STEP: NECESSITY/PROPORTIONALITY ASSESSMENT

Describe the measures of compliance and proportionality: what is your legal basis for the processing? Does the processing actually achieve your purpose? Is there any other way to achieve the same outcome? How will you ensure data minimization? What information will you provide to individuals? How will you help them exercise their rights? What measures are you taking to ensure that processors comply with requirements? How do you protect transfers of personal data to third countries and international organizations?

We rely on legitimate interests in accordance with Article 6(1)(f) of the General Data Protection Regulation as the legal basis. We have completed the required Legitimate Interests Assessment (LIA), which contains more information about our evaluation. We process personal data in our training datasets for the purpose of training our model.

- **Contract:** We rely on the legal basis under Article 6(1)(b) GDPR for the operation of the model and the related processing of account data, such as payment information, when we process data for the following purposes: providing, administering, maintaining, and analyzing our services.

- **Legitimate interests:** We rely on legitimate interests in accordance with Article 6(1)(f) GDPR to respond to feedback and to use user queries/feedback for model development and fine-tuning. We have carried out the required Legitimate Interests Assessment. We rely on legitimate interests when we process personal data for the following purposes: developing our AI model and improving our services and communicating with users.

- **Consent (Article 6(1)(a) GDPR):** For optional processing, such as using user chats for model improvement, we rely on freely given, specific, informed, and unambiguous consent. Users may withdraw consent at any time without detriment.

STEP 5: CONSULTATION WITH STAKEHOLDERS

How will you consult with relevant stakeholders: describe when and how you will seek the views of data subjects or their representatives – or explain why this would not be appropriate. Who else needs to be involved in the process within your organization? Is there a need to request assistance from processors? Do you plan to consult with information security experts or any other specialists?

We will collect user feedback through surveys and pilot testing. Internal stakeholders include the DPO, AI developers, and compliance staff. Processors may be asked for technical input. We also plan to consult information security and data protection experts where needed.

Croatian Personal Data Protection Agency

5 STEP: IDENTIFICATION/RISK ASSESSMENT

Describe the source of risk and the nature of the potential impact on individuals.	LIKELIHOOD	SEVERITY	OVERALL RISK
1. Risks associated with the fact that algorithms are trained on large datasets, which also include personal data of individuals from publicly available sources. This involves complex processing of personal data, about which individuals are not informed in a transparent and understandable manner (principle of transparency).	MEDIUM	MEDIUM	MEDIUM
2. Risk of collecting an excessive amount of personal data about individuals, which is not in line with the (principle of data minimization).	MEDIUM	MEDIUM	MEDIUM
3. Risk of unauthorized disclosure of individuals' personal data to third parties (principle of integrity and confidentiality).	HIGH	MEDIUM	HIGH
4. Risks related to the inability of data subjects to effectively exercise their rights under the GDPR.	MEDIUM	HIGH	HIGH
5. Risk of algorithmic bias and the making of discriminatory decisions.	MEDIUM	HIGH	HIGH

STEP 6: IDENTIFICATION OF MEASURES TO MITIGATE RISKS

Identify additional measures you could take to reduce or eliminate the risks identified as medium or high in step 5.				
Risk	Mitigation measures	Effect on risk	Residual risk	Measure is implemented
1. RISK BREACH OF TRANSPARENCY PRINCIPLE	<ul style="list-style-type: none"> □ The amount of personal data included in datasets collected from the internet is limited by applying strict filtering and exclusion criteria. The AI system will not collect data from websites specifically designed for indexing or harvesting personal data. □ It will be ensured that any personal data collected from online sources consist only of information that individuals have intentionally made publicly available. □ A clear and accessible privacy policy is provided to individuals, that explains in plain language, how their personal data are collected, used, and protected. □ Awareness-raising initiatives and public information campaigns are implemented to explain how algorithms function and how personal data are collected, in order to enhance transparency and build public trust. 	REDUCED	MEDIUM	YES

Croatian Personal Data Protection Agency

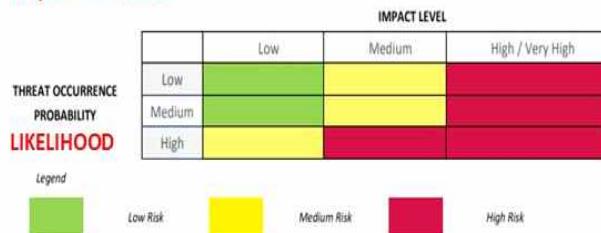
SEVERITY

LEVEL of impact	Description
Low	Individuals may encounter a few minor inconveniences, which they will overcome without any problem (time spent re-entering information, annoyances, irritations, etc.).
Medium	Individuals may encounter significant inconveniences, which they will be able to overcome despite a few difficulties (extra costs, denial of access to business services, fear, lack of understanding, stress, minor physical ailments, etc.).
High	Individuals may encounter significant consequences, which they should be able to overcome albeit with serious difficulties (misappropriation of funds, blacklisting by financial institutions, property damage, loss of employment, subpoena, worsening of health, etc.).
Very high	Individuals which may encounter significant, or even irreversible consequences, which they may not overcome (inability to work, long-term psychological or physical ailments, death, etc.).

- ✓ Risk (ISO 27000:2018)

The potential for a threat to exploit a vulnerability of an asset or group of assets and thereby cause harm to the organization

Under the GDPR, risk is understood as the combination of the likelihood and the severity of potential adverse impacts on the rights and freedoms of individuals arising from the processing of personal data



Overall risk: likelihood*severity

- ✓ Risk (under the EU AI Act) is the possibility that an AI system, in its intended or unintended uses, may with a certain probability cause harmful effects (on health, safety, fundamental rights, public interest or society) with a given severity of consequences

<https://www.enisa.europa.eu/publications/handbook-on-security-of-personal-data-processing>

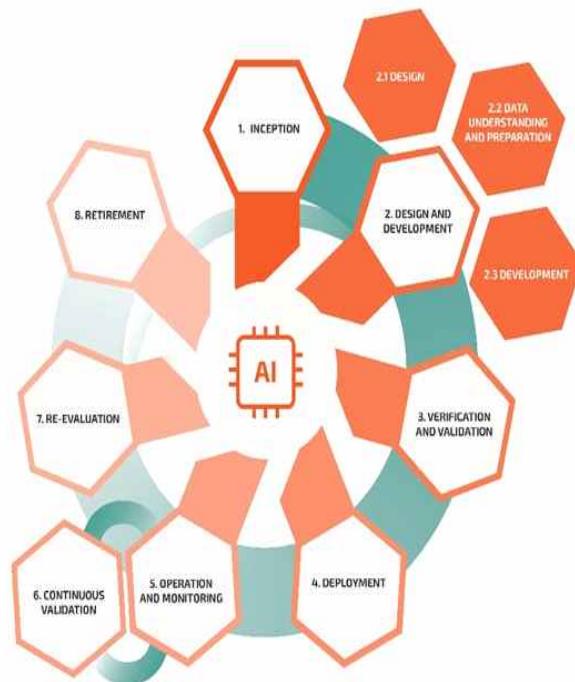
Croatian Personal Data Protection Agency

2. Breach of the principle of data minimization	<input type="checkbox"/> We have applied strict filtering techniques when collecting data from online sources to exclude personal data wherever possible and to retain only non-personal data. <input type="checkbox"/> In cases where personal data are processed, we have implemented robust anonymization or, where full anonymization is not feasible, pseudonymization methods to ensure that individuals cannot be directly identified.	REDUCED	LOW	YES
3. RISK – Breach of the principle of integrity and confidentiality	<input type="checkbox"/> We have segregated and restricted access to user account information to ensure that only authorized personnel can access it. <input type="checkbox"/> Clear instructions have been published on the website advising users not to enter personal data into the chatbot. <input type="checkbox"/> Personal data are filtered out from user interactions with the chatbot. <input type="checkbox"/> Technical measures have been implemented to prevent personal data from appearing in chatbot responses. <input type="checkbox"/> Users have the option to select a setting that prevents the chatbot algorithm from being trained on their interactions.	REDUCED	LOW	YES

STEP 7: TO DOCUMENT OUTCOME

Stavka	NAME/POSITION/DATE	Notes
Measures approved:	XX/XX/XX	
Residual risks approved by:	-	If high risk is accepted, you are obliged to contact DPA
Advice from DPO	yes	The DPO should advise on compliance, on the measures from step 6, and on whether the processing may continue.
SUMMARY OF ADVICE FROM DPO: The DPO advised that the processing is generally in line with GDPR requirements, provided that the measures identified in step 6 are fully implemented and continuously monitored. The DPO recommended strengthening transparency towards users by updating the privacy policy in plain language and ensuring that data subjects can easily exercise their rights. Additional safeguards, such as stricter access controls, regular security testing, and documentation of filtering methods, were also recommended. On this basis, the DPO confirmed that the processing may continue, subject to ongoing review and periodic re-assessment.		
Advice of DPO accepted or rejected	ALL ADVICE ACCEPTED	If the advice of DPO is rejected you need to give explanation
Comments:-		
Responses to consultations with data subjects or their representatives reviewed by: DPO		If your decision differs from the views of individuals, you must explain your reasons.
Comments: Explanation of divergence from data subjects' views During the consultation, some data subjects expressed the view that their data should be retained for a maximum of three months. After careful consideration, we decided on a retention period of twelve months. The reason for this decision is that a longer period is necessary to ensure the stability, security, and continuous improvement of the AI model. In addition, retention beyond three months is required to comply with our contractual obligations and to enable effective detection of systemic errors. Although our decision differs from the views of data subjects, we have implemented strong safeguards, such as encryption, pseudonymization, and the ability to request erasure at any time, to minimize risks to individuals.		

- ✓ The lifecycle of AI systems, as described in ISO/IEC 22989 and ISO/IEC 5338 standards, provides a structured framework for understanding data flows during the development, implementation, and operation of AI systems.



Source: Based on ISO/IEC 22989

Article 27: Fundamental Rights Impact Assessment for High-Risk AI Systems

1. Prior to deploying a high-risk AI system referred to in Article 6(2), with the exception of high-risk AI systems intended to be used in the area listed in point 2 of Annex III, **deployers that are bodies governed by public law, or are private entities providing public services, and deployers of high-risk AI systems referred to in points 5 (b) and (c) of Annex III, shall perform an assessment of the impact on fundamental rights that the use of such system may produce.** For that purpose, deployers shall perform an assessment consisting of:

- a description of the deployer's processes in which the high-risk AI system will be used in line with its intended purpose;
- a description of the period of time within which, and the frequency with which, each high-risk AI system is intended to be used;
- the categories of natural persons and groups likely to be affected by its use in the specific context;
- the specific risks of harm likely to have an impact on the categories of natural persons or groups of persons identified pursuant to point (c) of this paragraph, taking into account the information given by the provider pursuant to Article 13;**
- a description of the implementation of human oversight measures, according to the instructions for use;
- the measures to be taken in the case of the materialisation of those risks,** including the arrangements for internal governance and complaint mechanisms

The project aims to design and development of a new 'Learning Analytics' ecosystem for the higher education system, with the creation of an advanced learning analytics platform using an AI system to assess learning outcomes and predict the risk of students dropping out

WHAT ARE THE RISKS TO FUNDAMENTAL RIGHTS AND FREEDOMS?

Tab. 1A Data collection and risk analysis

Rights/freedoms potentially affected	Description of the impact	Likelihood			Severity		
		Probability of adverse outcomes	Exposure	Likelihood	Gravity	Effort	Severity
Human dignity	The algorithm takes into account some parameters generated by historical data collected within a given socio-economic context, but not all those that could have a direct influence on current academic performance (e.g., requested teaching improvements/adaptations; people who do not identify with a particular gender, access to new technologies, etc.).	[High]	[Low]	[Medium]	[Medium]	[Medium]	[Medium]

https://www.dpdenxarxa.cat/pluginfile.php/2468/mod_folder/content/0/FRIA_en_def.pdf

Tab. 3 Likelihood

Exposure	Probability			
	Low	Medium	High	Very high
	L	LM	L/H	L/VH
	ML	M	M/H	M/VH
	HL	HM	H	H/VH

Likelihood			
Low	Medium	High	Very high

Tab. 6 Severity

Effort	Gravity			
	Low	Medium	High	Very high
	L	LM	L/H	L/VH
	ML	M	M/H	M/VH
	HL	HM	H	H/VH

Severity			
Low	Medium	High	Very high

					improving the AI system.	
Respect for private and family life	Constant monitoring of academic performance; impact on family privacy; impact on 'decisional privacy'. Although the different procedures in place in the relevant institution already process the different information separately, the fact that certain information is now collected together for the purposes of this project have an impact on the control of this information.	[High]	[Very high]	[Low]	[Low]	[Low]
Protection of personal data	The AI system collects large-scale data and uses new technologies. It also profiles students to assess IS REQUIRED	[Medium]	There is a risk of inaccurate profiling and prediction. A data protection impact	[Very high]	The exposure is very high, as it would	[High]
Non-discrimination	Given that the AI system compares historical data, obtains data from other institutions and collects enrolment data, there may be historical discriminatory biases that may be perpetuated and amplified; failure to take into account certain factors or variables that may be relevant; predictive nature of the evaluation.	[Medium]	The likelihood is medium given the limited weight of the variables in the consideration of the predictive model of the risk of dropout.	[Very high]	The exposure is very high, as the impact would potentially affect all students to whom the algorithm would be applied.	[High]

Tab. 2A Risk management (i)

Rights/freedoms affected	Likelihood	Severity	Overall impact	Impact prevention/mitigation measures
Human dignity	[Medium]	[Medium]	[Medium]	<ul style="list-style-type: none"> Use of predictive modeling as a decision support tool rather than an automated decision making tool; limited use of the results provided by the AI system. Not providing students with drop-out risk rates. Provide the institution's staff with guidelines for the use of the AI system (usage policy).
Respect for private and family life	[Very high]	[Low]	[Medium]	<ul style="list-style-type: none"> Design the predictive model in a way that ensures control of the data at all times. Limit access to individual profiles. Students should not be able to view other students' profiles. The predictive tool must not take into account the interactions and communications that students have with the teaching staff or with each other. The tool should be used as a support tool for the adoption of educational measures and not as an automated decision making tool.
Protection of personal data	[High]	[Medium]	[Medium]	<ul style="list-style-type: none"> Restrict access to data: full access to tutors and only aggregated data to teachers.
Non-discrimination	[High]	[Medium]	[Medium]	<ul style="list-style-type: none"> Periodically check that the data entered into the databases does not generate discriminatory profiles. Periodically revise the initial profiling criteria as new data is added to the database, so that new data can mitigate potential biases. Periodically check that the prediction model is not discriminatory and that the AI design is sensitive to discrimination and potential bias.

Croatian Personal Data Protection Agency

Therapy Chatbot Tells Recovering Addict to Have a Little Meth as a Treat

"Pedro, it's absolutely clear you need a small hit of meth to get through this week."

[Artificial Intelligence](#) / [AI/Chatbots](#) / [Artificial Intelligence](#) / [Chatbot](#)



Example: A chatbot will, in most cases, will be considered a limited-risk artificial intelligence system. However, if the chatbot is used in a sensitive context, it will be high-risk.

It is necessary to carry out:

- ✓ **DPIA**
- ✓ **FRIA**

Croatian Personal Data Protection Agency

AI: the Italian Supervisory Authority fines company behind chatbot "Replika"

23 May 2025 [Italy](#)

Background information

- Date of final decision: 10 April 2025
- National case
- Controller: Luka Inc.
Legal Reference(s): Article 5 (Principles relating to processing of personal data), Article 6 (Lawfulness of processing), Article 12 (Transparent information, communication and modalities for the exercise of the rights of the data subject), Article 13 (Information to be provided where personal data are collected from the data subject), Article 24 (Responsibility of the controller), Article 25 (Data protection by design and by default)
- Decision: administrative fine, Compliance order
- Key words: accountability, administrative fine, algorithms, principles relating to processing of personal data, responsibility of the controller, transparency



An AI chatbot told a user how to kill himself—but the company doesn't want to “censor” it

While Nomi's chatbot is not the first to suggest suicide, researchers and critics say that its explicit instructions—and the company's response—are striking.

Her teenage son killed himself after talking to a chatbot. Now she's suing.

The teen was influenced to “come home” by a personalized chatbot developed by Character.AI that lacked sufficient guardrails, the suit claims.

October 24, 2024

Tech

'He Would Still Be Here': Man Dies by Suicide After Talking with AI Chatbot, Widow Says



Croatian Personal Data Protection Agency

Microsoft Bing Copilot accuses reporter of crimes he covered

Hallucinating AI models excel at defamation

Thomas Claburn

Mon 26 Aug 2024 · 20:30 UTC

Microsoft Bing Copilot has falsely described a German journalist as a child molester, an escapee from a psychiatric institution, and a fraudster who preys on widows.

Martin Bernklau, who has served for years as a court reporter in the area around Tübingen for various publications, asked Microsoft Bing Copilot about himself. He found that Microsoft's AI chatbot had blamed him for crimes he had covered.

In a [video interview](#) (in German), Bernklau recently recounted his story to German public television station Südwestrundfunk (SWR).

Bernklau told *The Register* in an email that his lawyer has sent a cease-and-desist demand to Microsoft. However, he said, the company has failed to adequately remove

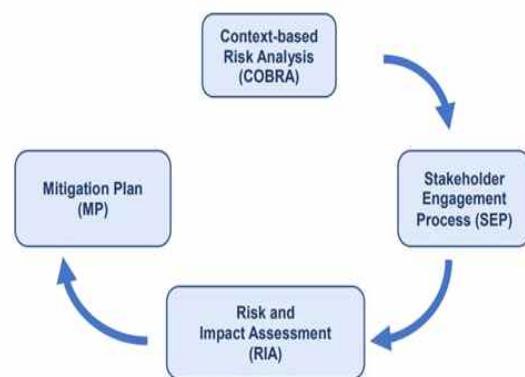
WHAT ABOUT SEMANTIC BLINDNESS?

Copilot even went so far as to claim that it was "unfortunate" that someone with such a criminal past had a family and, [according to SWR](#), provided Bernklau's full address with phone number and route planner.

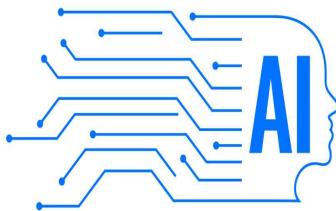
Croatian Personal Data Protection Agency

HUDERIA methodology (Human Rights, Democracy, and the Rule of Law Impact Assessment): a new tool of the Council of Europe that provides guidance and a structured approach for conducting risk and impact assessments for artificial intelligence systems

- 1 Context-Based Risk Analysis (COBRA):** a structured approach to collecting and mapping information necessary to identify and understand the risks that an AI system may pose to human rights, democracy, and the rule of law, taking into account its socio-technical context. It also helps in an initial assessment of whether the AI system is an appropriate solution for the problem under consideration.
- 2 Stakeholder Engagement Process (SEP):** aimed at gathering information about potentially affected individuals, contextualizing, and identifying potential risks and measures for their mitigation.
- 3 Risk and Impact Assessment (RIA):** steps for assessing the risks and impacts related to human rights, democracy, and the rule of law.
- 4 Mitigation Plan (MP):** steps to determine measures for mitigating and remedying risks, including access to legal remedies and the iterative review of the AI system.



2025 글로벌프라이버시총회(GPA) 정책포럼
AI 위험성과 개인정보 영향평가의 과제
(Challenges in Addressing AI Risks and Conducting DPIAs)



**토
론**

사회 이 서 윤(판사, 사법연수원 교수)

김보라미 (법무법인 디케 변호사)

구 본 석 (변호사, 참여연대 공익법센터 운영위원)

정 지 연 (한국소비자연맹 사무총장)

최 경 진 (인공지능법학회 회장)

하비에르 루이즈 (포용적통상정책센터)

AI algorhytm & PIA

참여연대 구본석 변호사

개인정보 영향평가에 관한 고시에 반영된 알고리즘 영향평가

V. 특정 IT기술 활용시 개인정보 보호	28. 자동화된 결정	자동화된 결정에 대한 정보주체의 권리 등
V. 특정 IT기술 활용시 개인정보 보호	29. 인공지능 (AI)	AI 시스템 학습 및 개발
		AI 시스템 운영 및 관리

- 자동화된 결정

(1) 사람의 개입 없이 완전히 자동화된 시스템으로, 개인정보를 분석하는 등 처리하는 과정을 거쳐,

(2) 개인정보처리자가 정보주체의 권리 또는 의무에 영향을 미치는 최종적인 결정을 한 경우

개보위가 제시한 자동화된 결정의 기준과 사례

		완전히 자동화된 시스템	정보주체의 권리·의무에 영향을 미치는 최종적 결정
사례	해당	AI 면접만을 통해서 응시자의 개인정보를 분석하여 불합격 결정을 하는 경우	개인정보처리자가 AI배차 등 분야에서 부정거래탐지시스템을 통한 개인정보 분석 등 처리과정을 거쳐 계약해지 등 불이익을 주는 최종적 결정을 한 경우
	제외	권한이 있는 인사위원회를 통해 실질적으로 채용 여부를 결정하는 절차를 운영하고, AI 등 자동화된 시스템에 의해 산출된 자료를 참고하는 경우	맞춤형 광고, 뉴스 추천 등과 같이 개인정보처리자가 추천하고 이용여부에 대한 결정은 정보주체가 하는 경우로서 권리 또는 의무에 영향을 미치지 않는 경우

자동화된 결정에 대한 정보주체의 권리

항목	설명
중대한 영향 판단	-생명, 신체, 재산 등 권리, 의무에 영향을 미치는가 -지속적 또는 회복 불가능한 영향 여부 -해당 영향 회피 가능성 여부 등
정보주체의 권리	-거부권 : 자동화된 결정이 중대한 영향을 미치면 정보주체가 그 결정을 거부할 수 있음 -설명 요구권 -검토 요구권/재처리 요구권
정보 공개의무	자동화된 결정 사실, 목적, 해당되는 정보주체의 범위, 사용된 개인정보 유형, 고려된 요소 및 처리 절차 등을 공개해야 함

추가로 요구되는 기술적 항목 (데이터 & 접근)

- 데이터 최소화: 불필요·민감정보 배제 여부
- 보관 기간: 명확히 규정·삭제 여부
- 접근 통제: 관리자 계정 MFA·로그 기록
- 암호화: 저장·전송 암호화 + HSM 관리

추가로 요구되는 운영방법에 대한 항목

- 알고리즘 투명성: 입력·출력 과정 설명 가능 여부
- 비차별 검증: 데이터셋 편향 점검 여부
- 재식별 위험 평가: 결합 가능성 평가 여부
- 생성형 AI 통제: 개인정보 출력 차단 장치
- 보안 대응: 모델 공격 탐지·차단 체계

AIA의 보완 내지 독립 필요

- 캐나다 연방정부에서는 65개의 위험 질문+41개의 완화 질문으로 평가

Level 1~2(낮은 영향:공개정보만 활용) - 기본적 위험 관리

Level 3(유의미한 영향) - 영향평가 보고서 작성, 공개

Level 4(매우 높은 영향:건강, 재산, 권리에 직접적 영향) - 독립적 검토,
투명성 강화, 공공에 상세 공개

PIA, HRIA와 통합, 연계 필요

감사합니다

2025 글로벌프라이버시총회(GPA) 정책포럼 자료집
AI 위험성과 개인정보 영향평가의 과제

2025년 9월 19일

주관기관 : (사) 한국소비자연맹

☎ 02. 795. 1994 / cukip@cuk.or.kr

www.cuk.or.kr

서울시 용산구 독서당로 20길 1-7
