

VISCHER

Generative AI.

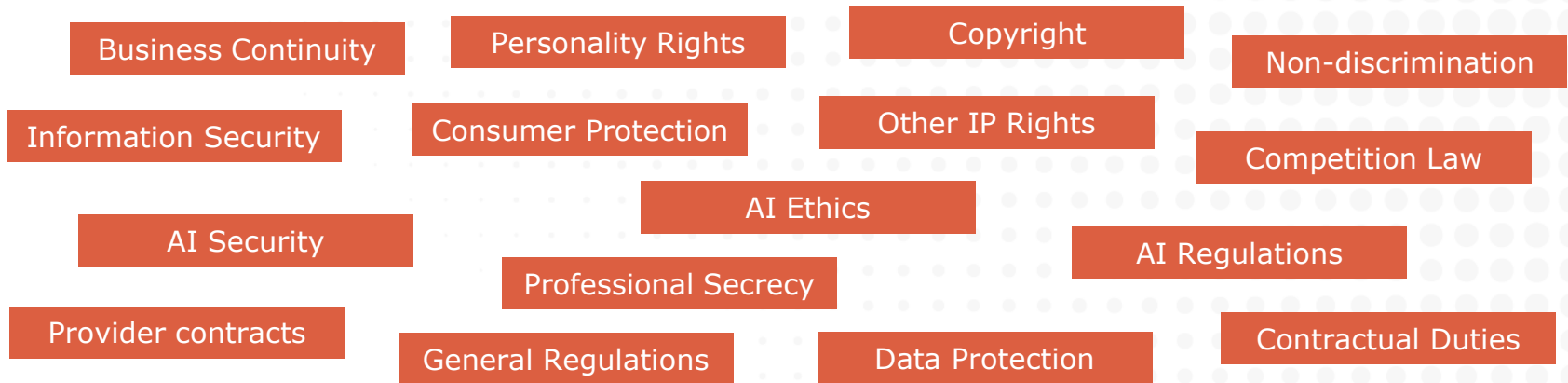
Handling risk management in practice & setting
up an appropriate AI Governance Structure

Rolf Auf der Maur & David Rosenthal, VISCHER AG
AI House Davos
January 16, 2024

Risks? There are many ...

- Operational risks
- Legal & compliance risks
- Reputational risks

} Understand, control and accept them



Some observations

- Many **existing rules** of law (e.g., data protection) **work well** also for GenAI; new regulations seem to target providers of AI
- Do not confuse legal requirements with **ethical principles**
- Expect a "legal" **demystification**; we will realize that with proper contracts and provider setups, feeding personal data and third-party content into GenAI systems is often less problematic than anticipated; the main concern is their output and its use
- Most GenAI projects are also **cloud projects**, which may result in additional requirements and issues for risk management
- A lack of **transparency** and **quality standards** concerning the models, their training, and AI offerings in general will continue to exist and make compliance and risk assessments difficult

How to manage AI risks

1. Identify and discuss the applicable **compliance baseline**
2. Issue an appropriate **policy** for governing and using AI
3. Establish a proper **governance** (1st line, 2nd line, reporting)
4. Establish an **overview** of what is going on ("ROAIA")
5. Enable and **train** employees in the proper use of GenAI
6. Establish a process for a compliance and **risk assessment** of every AI service or other use of AI
7. Have the **2nd line assist** the 1st line in doing the assessment
8. Install **monitoring**, re-assessment and incident **reporting** processes and act upon findings and reports

Doing AI risk assessments

- Many AI projects are **not very problematic**
 - E.g., using a chatbot for day-to-day tasks
- Is your AI project a **high-risk** for the company?
 - Training your own model
 - Taking important automated decisions
 - Interaction with many people on sensitive topics
 - We would consider legal if this were about us
 - A high "shitstorm" factor
 - Prohibited activity or high-risk system as per EU AI Act
 - Offering AI applications to third parties
 - Large investment or project of strategic importance

GAIRA Light

Get it at vischerlnk.com/gaira

GAIRA
Comprehensive

If so, then do a
comprehensive risk
assessment

GAIRA Light

Doing a risk assessment with GAIRA Light.

To assess your AI application, follow the nine steps. Expect 60 minutes for a comprehensive assessment to answer the questions of your provider(s). For 12

Compliance housekeeping questions

1	2	3	4	5	6
3.08 Do we have a proper data processing agreement with our data processing provider(s)? ⁽¹⁾	Yes	OK			
3.09 Do we properly address the transfer of personal data to "unsafe" countries with them? ⁽⁵⁾	Yes	Yes, we have the EU SCC in place with a TIA.	OK		
3.10 Are our provider(s) approved for use with official/professional secrecy data (where necessary)? ⁽⁶⁾	Yes		OK		
3.11 Do our provider(s) contracts protect the confidential or legally protected data we may be using? ⁽⁴⁾	Yes		OK		
3.12 Have we performed a DPIA insofar necessary (with no high risks for data subjects)? ⁽⁷⁾	No		Warning!	OK	DPO
3.13 Have we amended the privacy notice (where necessary)?	Yes	No amendment is necessary.	OK		Accepted
3.14 Have we amended the records of processing activities (where necessary)? ⁽⁸⁾	Yes		OK		

Risk questions (25)

Answer	Reason incl. measures taken/planned	Assessment	2nd Line Comment	By whom?	Risk Handling

Step 4: Conclusion

Instruction: Once you have come to step 4, you are almost done. The first line below will give you a recommendation based on the above responses you have made. Let the 2nd line functions and other stakeholders (e.g., other business functions with whom you have discussed your risk assessment) add any comments they wish and consider them. At the end, you are asked to assess the overall risk level of your application (low, medium, high, very high) in terms of the risk for the organization. The value can then be used for the Records of AI Activities (ROAIA). Finally, record your decision as how you wish to proceed and provide any comments you deem fit for the record (don't forget to update the status and date in the top section of the form).

Recommendation: 7 Your application does not appear to have any unaccepted risks. You can proceed, subject to any other objections or changes to your answers above.

Comments or objections by 2nd Line: No comments.

Comments or objections by others:

Overall risk level: Low (to be assessed by the application owner; please include this in the ROAIA)

Decision on how to proceed: 8 Implementation (to be taken by the application owner)

Comments by application owner:

Get the GAIRA excel for the "Light" version showing the "Light" assessment details may be necessary, too. Discuss this with your DPO or other legal experts.

Author: David Rosenthal (drosenthal@vischer.com) All rights reserved. No legal advice. 10.1.24 Updated: vischer.com/updates VISCHER

Ten questions to start with when using LLMs

- **Where** is the input ("prompts") sent to and processed?
- Is there a **data processing contract** with the provider?
- Is **data security** sufficiently ensured?
- Is the input used for **provider training** of the model?
- **What** input are employees allowed to make?
- Is the output ("completions") **monitored** by the provider?
- How do we deal with **inaccurate/unwanted output**?
- Must and can **"data leakage"** be avoided?
- Do we have to **point out** that we use AI and how?
- How do we handle **data subject requests** that we may get?

Check out your GenAI project for risks. Based on "GAIRA LLM"

Do you plan to use an application based on generative AI in a company project? If so, then check out the following points.

1. Is your application likely to involve high risks for the company?

If you cannot confirm every one of the following points, then your project likely involves high risks and you should do a comprehensive risk assessment (e.g., GAIRA Comprehensive). Otherwise proceed to step 2.

- We do not create or further train the AI model we use for our application (excluding "RAI")
- We do not let our application take decisions on other people that they will find important
- Our application will not interact with a large number of people concerning sensitive topics
- We'll would not feel uncomfortable if the application were used by others with our/my data
- The application does not have the potential to cause negative headlines in the media
- The application does not qualify as a prohibited or high-risk system as per the EU AI Act
- We will be using our AI application only for our purposes and not offer it to third parties
- The application does not require a large investment, and it is not of strategic importance

2. How are you dealing with the typical risks of generative AI?

If you cannot confirm every one of the following points in view of the existing or planned measures to control the risks, then you should discuss the legal & reputational risks of your project internally.

- Our use of AI is fully contained or protected data to be fine with our contractual obligations
- Our provider contracts are in line with data protection law and our other legal obligations
- Our input/output will not be monitored by our provider(s) or we are fine with it
- Our input/output will not be used by our provider(s) for their own business or we are fine with it
- We have ensured that our AI solution will not leak confidential data or personal data to others
- Our application neither systematically processes sensitive personal data nor data any profiling
- We have a sufficient legal basis for processing personal data (where legally required)
- We are using personal data only for one or several purposes (for one that data to be expected)
- We limit the personal data collected/used to the minimum necessary for the purpose
- We keep personal data in connection with the solution only for as long as needed
- We are able to comply with data subject requests (e.g., access, correction, deletion)
- We have adequate data security and business continuity measures in place
- We inform people about our use of AI where this is relevant for their interaction with us
- The public and those affected by our use of AI will generally not find it unfair or objectionable
- We seek measures to avoid or deal with errors, inaccuracies, or biases in the output (e.g., bias)
- Our application will be tested extensively prior to its use, including against adversarial use
- Our use of AI does not cause unintended repercussions for others (e.g., damage, discrimination)
- Our use of AI cannot be considered as being based on reporting vulnerabilities of individuals
- We have human oversight where our AI solution could take or influence key decisions
- We will be using a widely recognized, quality AI model with a behavior we understand
- Our use of AI will neither mislead nor deceive anyone
- We have measures in place to detect and remediate behavior of our AI, to log it and to react to it
- Our use of AI respects the dignity and individual autonomy of those affected by it
- The users of our AI solution will be instructed, trained and monitored in the proper use
- We do not foresee any other uncontrollable issue in connection with our use of AI

If you are processing personal data, do not forget to also do a Data Protection Impact Assessment, amend your privacy notice and update your records of processing activities (ROPA), each where necessary. Also comply with your other company rules and procedures on AI, including ethics guidelines that may exist.

Do you want to properly document your above risk assessment? Get the GAIRA Excel for Risk at <https://vischerlnk.com/gaira>. It includes both a worksheet for the "light" version with the points above and one for a comprehensive risk assessment and documentation.

Author: David Rosenthal, drosenthal@vischer.com, All rights reserved. No legal advice. © 2024 Vischer - solutions.com

vischerlnk.com/ai-riskcheck

GAIRA Comprehensive / 1

Generative AI Risk Assessment (GAIRA)

Draft for public comment (31.10.2023) - with Data Protection Impact Assessment (DPIA) included

Overview: GAIRA is intended to allow an organization to make a holistic risk assessment of its AI applications. It works best with medium and large sized and higher risk projects. GAIRA is to be completed under the lead and the responsibility of the "Business" or the application owner, whereas IT and information security experts and internal law functions such as legal, compliance and the data protection officer should help (comparable to a data protection impact assessment, which is actually included in GAIRA). On typical approach is to have the application owner or project manager prepare the form and then have a workshop with all stakeholders for going through the assessments, discussing and completing them. The ultimate decision is with the "Business" or the application owner.

Company: Bank ABC
Department: Wealth Management
Application owner: Peter Parker
Status and date of risk assessment: In progress, 2023-12-01

Name of application: Project Alpha

Scope of assessment includes: The implementation and use of the system.

Step 1: Description of App

Instruction: Complete application unless it covers the risks of the app. This will give you a better understanding of the app.

Area in which the app is used: What is planned?

Interests pursued:

General technical measures:

Procedures involved:

Description of input:

Description of output:

Training or fine-tuning:

Categories of data:

Categories of person:

Project plan:

The positive effects:

Category of AI application:

Four phases of generative AI:

Risk Radar

Assessment period: 3 years

Assessors: Hans Binner (P/Pr), Peter Parker (AI), Dennis Doe (P/Pr), Lukas Langemann (Legal/DPO)

Risk area:	What is the application (if not covered by the application)?	What measures or other controls are in place to prevent or reduce the risk?	Probability of occurrence	Financial harm	Reputational harm	Regulatory harm	Criminal actions	Other negative effects	Harm to the bank	Comments for e.g., assumptions, No	Risk Handling	Risk Owner
4.01	Our use of AI at such a level is permitted by the public, as long as we do not use it for unlawful or otherwise illegal purposes in the context of our business.	- We inform clients and they agree to our use of AI at such a level. - We do not use it for unlawful or otherwise illegal purposes in the context of our business.	Low	Low	Medium	Low	N/A	N/A	Low	Yes, we covered the overall risk and the app is not used for unlawful or otherwise illegal purposes in the context of our business.	Accepted	AD
4.02	Our use of AI at such a level is permitted by the public, as long as we do not use it for unlawful or otherwise illegal purposes in the context of our business.	- We inform clients and they agree to our use of AI at such a level. - We do not use it for unlawful or otherwise illegal purposes in the context of our business.	Low	Low	Medium	Low	N/A	N/A	Low	Yes, we covered the overall risk and the app is not used for unlawful or otherwise illegal purposes in the context of our business.	Accepted	AD
4.03	Our use of AI at such a level is permitted by the public, as long as we do not use it for unlawful or otherwise illegal purposes in the context of our business.	- We inform clients and they agree to our use of AI at such a level. - We do not use it for unlawful or otherwise illegal purposes in the context of our business.	Low	Low	Medium	Low	N/A	N/A	Low	Yes, we covered the overall risk and the app is not used for unlawful or otherwise illegal purposes in the context of our business.	Accepted	AD
4.04	Our use of AI at such a level is permitted by the public, as long as we do not use it for unlawful or otherwise illegal purposes in the context of our business.	- We inform clients and they agree to our use of AI at such a level. - We do not use it for unlawful or otherwise illegal purposes in the context of our business.	Low	Low	Medium	Low	N/A	N/A	Low	Yes, we covered the overall risk and the app is not used for unlawful or otherwise illegal purposes in the context of our business.	Accepted	AD

- Full AI risk assessment including a data protection impact assessment (DPIA)
- Requires a list of technical and organizational measures
- Requires more time (several hours for the business to complete)
- Covers also reputational risks

Available for free at
vischerlnk.com/gaira



GAIRA Comprehensive / 2

- Use the **same** proven **approach** as for doing a **DPIA**
 - Have the application owner describe the application
 - Have the application owner list all measures intended to prevent "problems" and to comply with law and internal policies
 - Go through the list of risk scenarios, and have the application owner and others assess the relevant risks; typically, additional measures will pop-up – add them to the list of measures
 - Ensure that someone is responsible for each measure
- **Top five DP risks** are usually accuracy, secrecy, data leakage, provider contracts and data subject rights
 - Ethics and transparency are usually not (yet) an issue
- Don't forget: The application **owner**/business has to **decide**

VISCHER
SWISS LAW AND TRADING

Download it at
vischerlnk.com/qaira

What works well for data protection (the "Records of Processing Activities" or ROPA), also works well for AI applications ("Records of AI Activities")

AI helping out on DP risk assessments

Data Protection Impact Assessment (DPIA)
Version 17.12.2023 - Private Sector CH-FADP/GDPR

Note: Instructions for completing this DPIA and for the AI-assisted filling aid (optional, only in the Excel with macros) can be found at the end of this page, on the right and in the one pager and vic [FAQ](#)

Organization (Controller):

Department:

Internally res:

Status of the

Project Name:

	Possible unintended negative consequences	How we mitigate the risk	How we assess the residual risk	Possible consequences for the person	Probability of occurrence (all in all)	Risk (1-16)
4.01	Personal data used in the project is leaked due to an error or intentionally to unauthorized third parties . They misuse it to the detriment of the affected individuals.	1. Implement robust encryption techniques for voice print and other personal data both in transit and at rest to prevent unauthorized access. 2. Conduct regular security audits and vulnerability assessments on the system to detect and fix any potential loopholes. 3. Establish strict access control measures, ensuring only authorized personnel can access the stored voice prints and personal data.	In the event that personal data used in the project is leaked to unauthorized third parties, the affected person may face substantial harm. For example, if an unauthorized party gains access to the voice print and personal data, they could potentially use this information to impersonate the caller and commit fraudulent activities. This could lead to financial loss, reputational damage, or even legal consequences for the affected person. Although the probability of this occurring is very low due to the implemented countermeasures, the potential severity of the consequences emphasizes the importance of maintaining a robust data protection framework.	Substantial	Very low	Low (3)
4.02	Personal data used in the project is leaked due to an error or intentionally to an unauthorized internal person . This person misuses it to the detriment of the affected	- Training: This measure is chosen because it ensures that all users are familiar with the correct procedures for handling personal data, reducing the risk of unintentional leaks or	The specific residual risk in this scenario is that personal data used in the project is leaked to an unauthorized internal person who then misuses			

2. Necessity / Proportionality

2.01 Why the data processing is necessary or why, from the perspective of the data subjects, there is no less

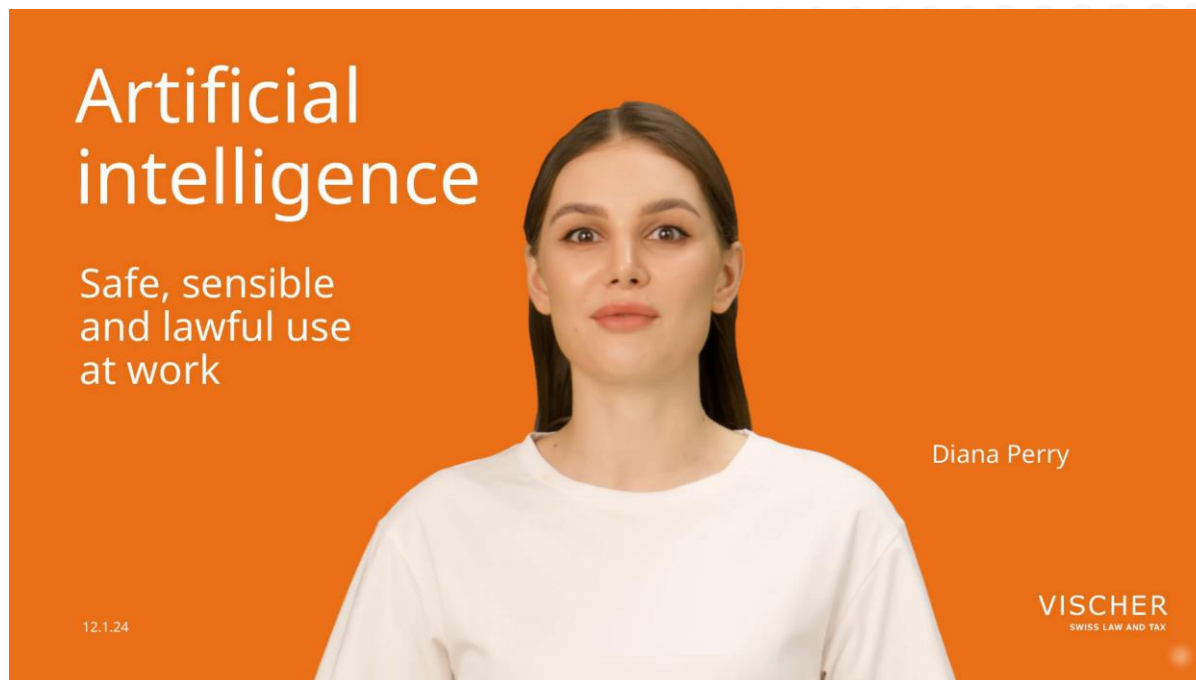
[vud.ch/dpia](#)

Data Protection Impact Assessment
[vud.ch/dpia](#)

Data protection impact assessments are a useful tool for dealing with data privacy risks. vud.ch provides a template for conducting data protection impact assessments.

With AI-assisted assistance

An AI that warns and charms about her kind



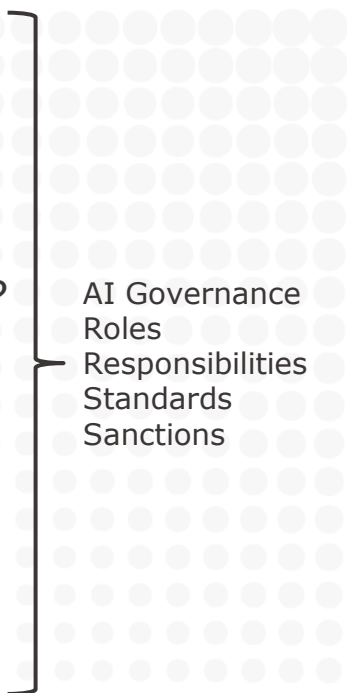
Feel free to share the video in your office with anyone you think should learn the three key points Diana presents



vischerlnk.com/ai-intro

[http://www.youtube.com/@VISCHERLegalInnovationLab](https://www.youtube.com/@VISCHERLegalInnovationLab)

AI Governance in Practice

- Does your organization need an AI Compliance Officer?
 - Which department is the most appropriate for AI Compliance?
 - What are the rights and obligations of the AI Compliance Officer?
 - Should AI Principles be adopted and implemented (law and ethics)?
 - Does every AI Application have a business owner (accountability)?
 - Which risk assessment methodology should be applied?
 - Should a standard be adopted (voluntarily)?
 - ISO/IEC 42001 Standard on AI management system (AIMS), adapted on 18 December 2023
 - Artificial Intelligence Risk Management Framework (AI RMF 1.0) by the U.S. National Institute for Standards and Technology
 - Sanctions for non compliance?
- 
- AI Governance
Roles
Responsibilities
Standards
Sanctions

TBD: Corporate AI Principles

- 11 Principles for the responsible use of AI:
 - We ensure accountability
 - We provide the necessary transparency
 - We remain fair and do no harm
 - We ensure reliability
 - We ensure information security
 - We pay attention to proportionality and self-determination
 - We respect others' and our own intellectual property
 - We protect the rights of those affected
 - We ensure explainability and human oversight
 - We understand and control the risks
 - We prevent misuse of our AI applications

AI Principles
Legal
Ethics
Corporate
Culture

VISCHER

Thank you for your attention!

Questions: drosenthal@vischer.com, ram@vischer.com

Zürich

Schützengasse 1
Postfach
8021 Zürich, Schweiz
T +41 58 211 34 00

www.vischer.com

Basel

Aeschenvorstadt 4
Postfach
4010 Basel, Schweiz
T +41 58 211 33 00

Genf

Rue du Cloître 2-4
Postfach
1211 Genf 3, Schweiz
T +41 58 211 35 00



AIgen

Visit vischer.com/ai and the
VISCHER Legal Innovation Lab
vischer.com/en/lil