VISCHER

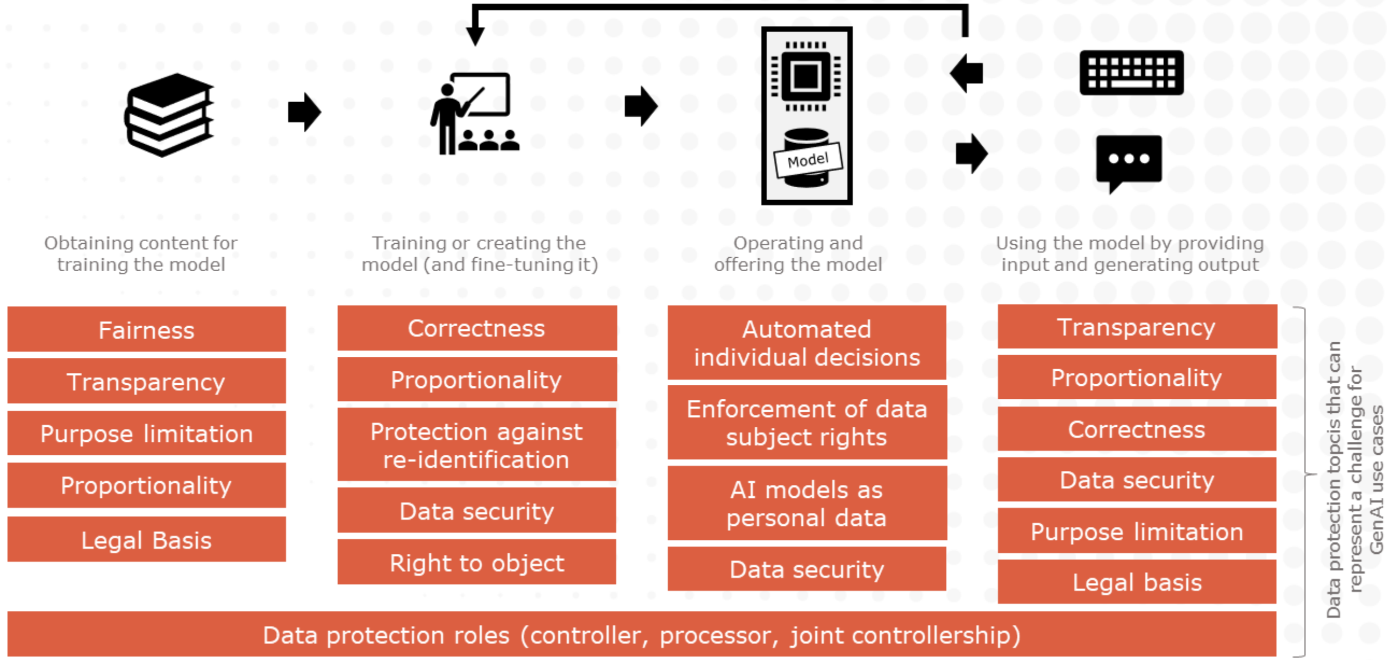
Generative AI.

Privacy & Risk Management

David Rosenthal, Partner, VISCHER AG
October 31, 2023

Plenty of data protection topics

Four phases of GenAI:



Obtaining content for training the model	Training or creating the model (and fine-tuning it)	Operating and offering the model	Using the model by providing input and generating output
Fairness	Correctness	Automated individual decisions	Transparency
Transparency	Proportionality	Enforcement of data subject rights	Proportionality
Purpose limitation	Protection against re-identification	AI models as personal data	Correctness
Proportionality	Data security	Data security	Data security
Legal Basis	Right to object		Purpose limitation
			Legal basis

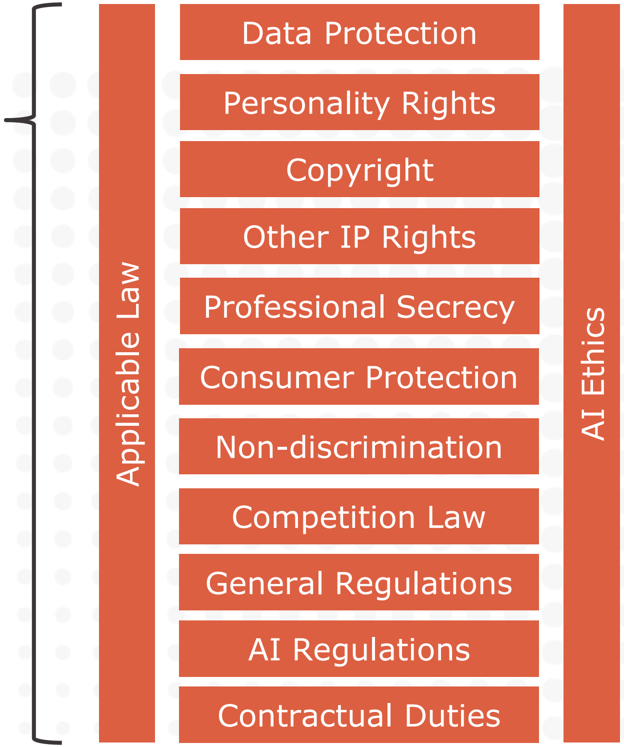
Data protection roles (controller, processor, joint controllership)

Data protection topcis that can represent a challenge for GenAI use cases

Download chart at: https://www.rosenthal.ch/downloads/Rosenthal-generative-ai-data-protection-issues.pdf

And there is more …

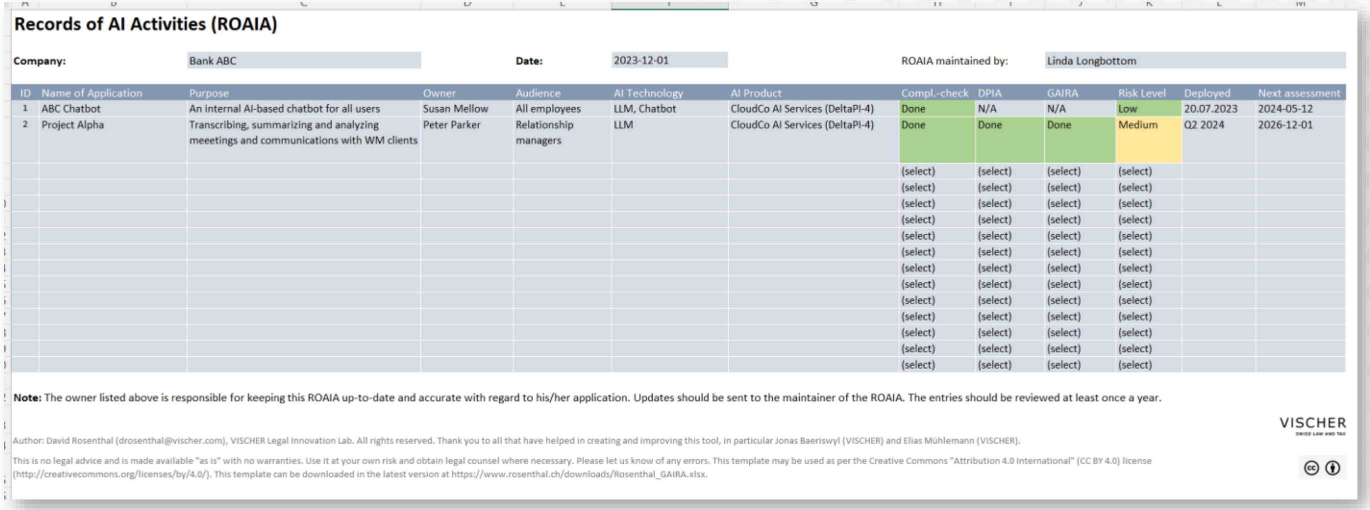
- **Other areas of law** also need to be considered when using generative AI
  - This is true even before the AI Act and similar regulations will apply
- Private and public sector organizations generally **feel unsure** about the legal conditions for using generative AI
  - They issue general guidelines, mainly advising not to use personal data
- At the same time, IT & business are pushing to implement such systems, and end users are **simply using them**



How to manage the risks

1. Gain ownership of the (legal) topic, or at least part of it
2. Establish an overview of what is going on, create the ROAIA

Create the "ROAIA"



Records of AI Activities (ROAIA)

Company: Bank ABC

Date: 2023-12-01

ROAIA maintained by: Linda Longbottom

ID	Name of Application	Purpose	Owner	Audience	AI Technology	AI Product	Compl.-check	DPIA	GAIRA	Risk Level	Deployed	Next assessment
1	ABC Chatbot	An internal AI-based chatbot for all users	Susan Mellow	All employees	LLM, Chatbot	CloudCo AI Services (DeltaPI-4)	Done	N/A	N/A	Low	20.07.2023	2024-05-12
2	Project Alpha	Transcribing, summarizing and analyzing meeetings and communications with WM clients	Peter Parker	Relationship managers	LLM	CloudCo AI Services (DeltaPI-4)	Done	Done	Done	Medium	Q2 2024	2026-12-01
							(select)	(select)	(select)	(select)		
							(select)	(select)	(select)	(select)		
							(select)	(select)	(select)	(select)		
							(select)	(select)	(select)	(select)		
							(select)	(select)	(select)	(select)		
							(select)	(select)	(select)	(select)		
							(select)	(select)	(select)	(select)		
							(select)	(select)	(select)	(select)		
							(select)	(select)	(select)	(select)		
							(select)	(select)	(select)	(select)		
							(select)	(select)	(select)	(select)		
							(select)	(select)	(select)	(select)		
							(select)	(select)	(select)	(select)		

Note: The owner listed above is responsible for keeping this ROAIA up-to-date and accurate with regard to his/her application. Updates should be sent to the maintainer of the ROAIA. The entries should be reviewed at least once a year.

VISCHER
SWISS LAW AND TAX

Author: David Rosenthal (drosenthal@vischer.com), VISCHER Legal Innovation Lab. All rights reserved. Thank you to all that have helped in creating and improving this tool, in particular Jonas Baeriswyl (VISCHER) and Elias Mühlemann (VISCHER).

This is no legal advice and is made available "as is" with no warranties. Use it at your own risk and obtain legal counsel where necessary. Please let us know of any errors. This template may be used as per the Creative Commons "Attribution 4.0 International" (CC BY 4.0) license (http://creativecommons.org/licenses/by/4.0/). This template can be downloaded in the latest version at https://www.rosenthal.ch/downloads/Rosenthal_GAIRA.xlsx.

Download it at https://www.rosenthal.ch/downloads/Rosenthal_GAIRA.xlsx

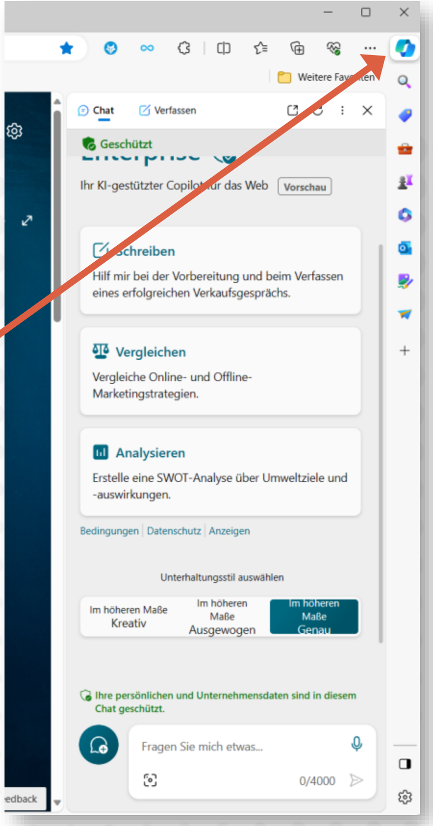
What works well for data protection (the "Records of Processing Activities" or ROPA), also works well for AI applications ("Records of AI Activities")

How to manage the risks

1. Gain ownership of the (legal) topic, or at least part of it
2. Establish an overview of what is going on, create the ROAIA
3. Understand where existing tools permit the use of GenAI

Understand the tools ...

Users can log into Bing with their corporate accounts, but Microsoft does not act as a processor when processing their data and is, thus, not bound by existing agreements



Ihr KI-gestützter Copilot für das Web

Schreiben
Hilf mir bei der Vorbereitung und beim Verfassen eines erfolgreichen Verkaufsgesprächs.

Vergleichen
Vergleiche Online- und Offline-Marketingstrategien.

Analysieren
Erstelle eine SWOT-Analyse über Umweltziele und -auswirkungen.

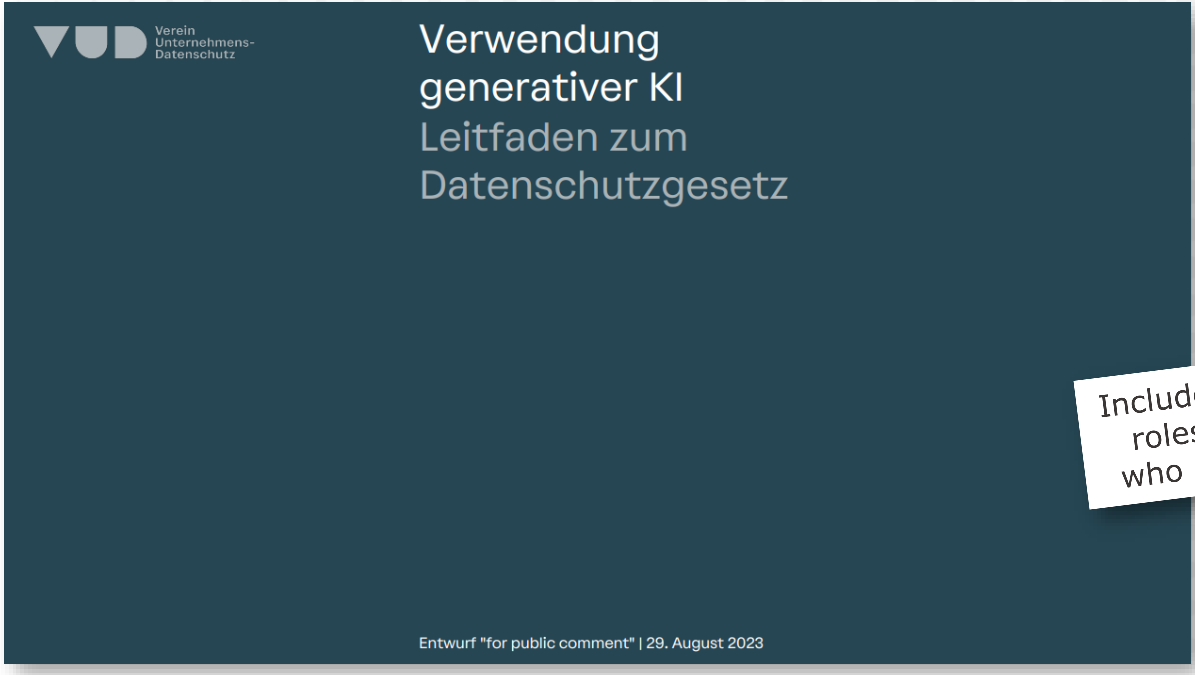
Unterhaltungsstil auswählen

Ihre persönlichen und Unternehmensdaten sind in diesem Chat geschützt.

How to manage the risks

1. Gain ownership of the (legal) topic, or at least part of it
2. Establish an overview of what is going on, create the ROAIA
3. Understand where existing tools permit the use of GenAI
4. Instruct and train employees in the proper use of GenAI
5. Establish that any new AI service used will go by you first, so you make sure that there are proper contracts and the basic data protection compliance requirements are met

More information



Verein Unternehmens-Datenschutz

Verwendung generativer KI

Leitfaden zum Datenschutzgesetz

Entwurf "for public comment" | 29. August 2023

Includes a discussion about the roles and responsibility, i.e. who is the controller of what?

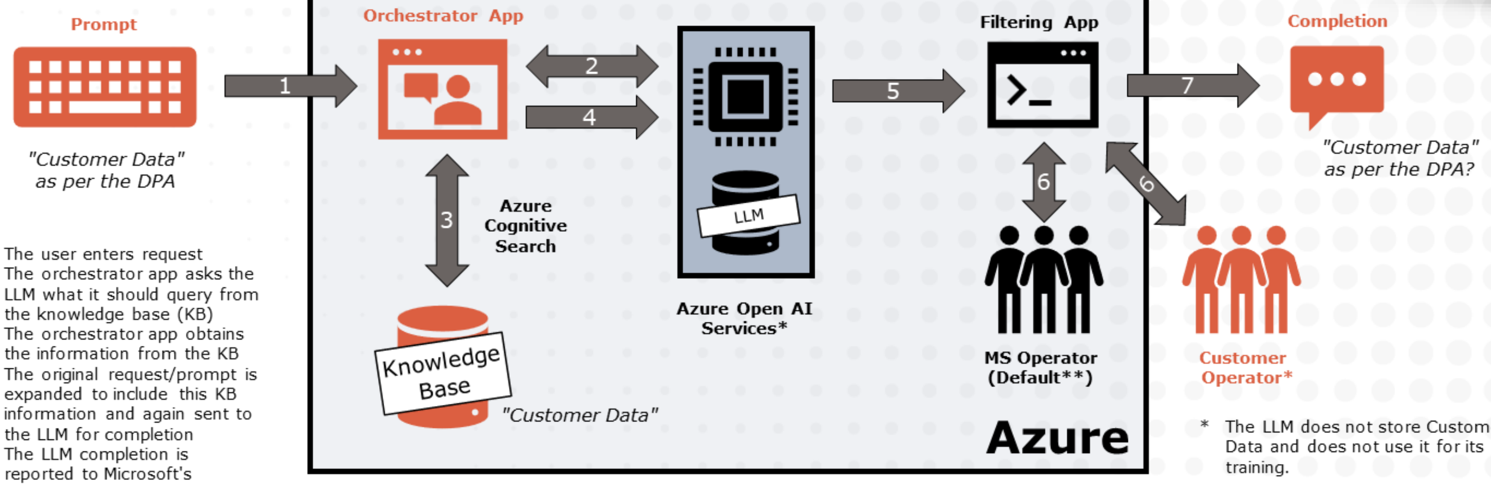
www.vud.ch

Ten basic questions for practice

- **Where** is the input ("prompts") sent to and processed?
- Is there a **data processing contract** with the provider?
- Is **data security** sufficiently ensured?
- **What** input are employees allowed to make?
- Is the input used for **provider training** of the model?
- 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?

Retrieval Augmented Generation

Example Microsoft with OpenAI



1. The user enters request
2. The orchestrator app asks the LLM what it should query from the knowledge base (KB)
3. The orchestrator app obtains the information from the KB
4. The original request/prompt is expanded to include this KB information and again sent to the LLM for completion
5. The LLM completion is reported to Microsoft's filtering app for ethics & compliance checking
6. If necessary, the completion is checked manually
7. The completion is played out

🟥 From/for Customer

⬛ From Microsoft

→ Use of Azure within the framework of the existing contract conditions

Info: https://learn.microsoft.com/en-us/legal/cognitive-services/openai/data-privacy

* The LLM does not store Customer Data and does not use it for its training.

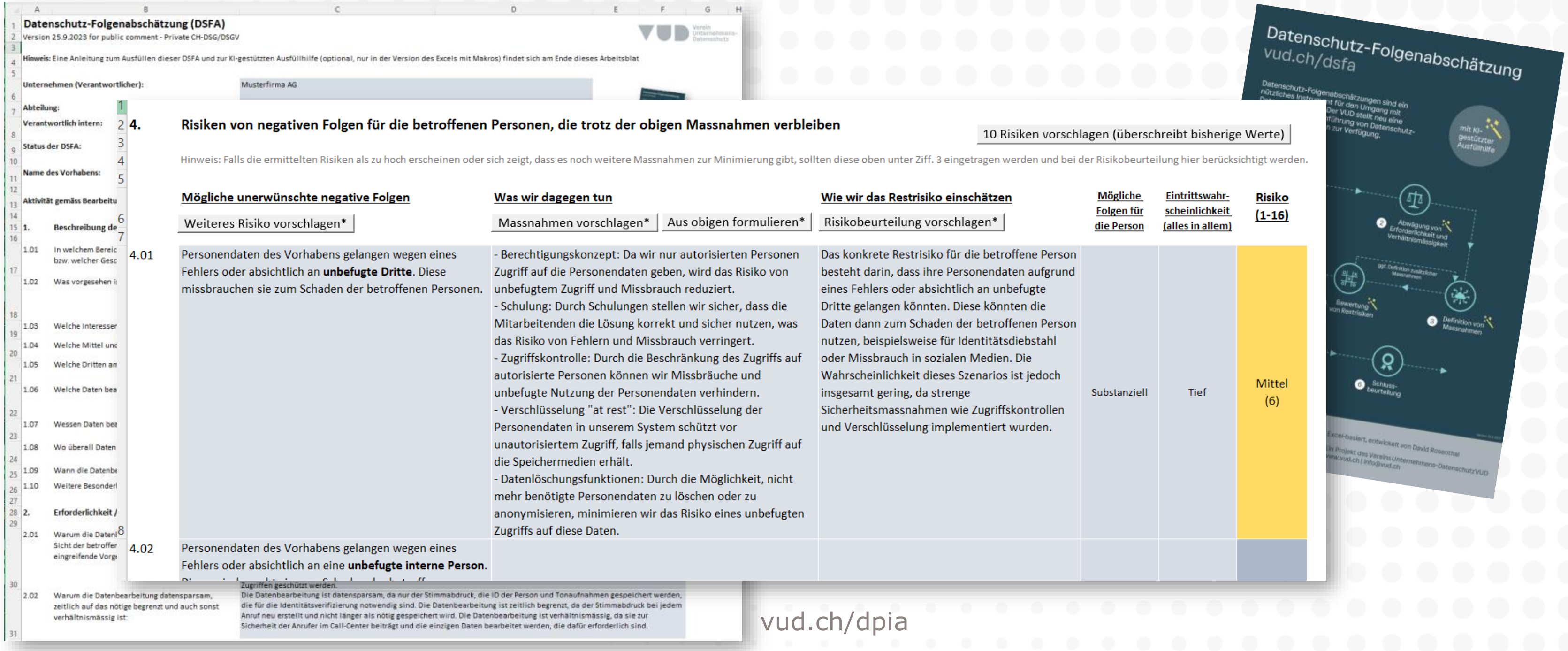
** Customer may be able to opt out of Microsoft's ethis & compliance verification of the completion (and do it itself using Microsoft's monitoring tools, if necessary).

How to manage the risks

1. Gain ownership of the (legal) topic, or at least part of it
2. Establish an overview of what is going on, create the ROAIA
3. Understand where existing tools permit the use of GenAI
4. Instruct and train employees in the proper use of GenAI
5. Establish that any new AI service used will go by you first, so you make sure that there are proper contracts and the basic compliance requirements are met
6. Have the application owner perform a DPIA (as applicable) and a structured and documented assessment of GenAI risks

VISCHER

Tool for Data Protection Impact Assessments



Datenschutz-Folgenabschätzung (DSFA)

Version 25.9.2023 for public comment - Private CH-DSG/DSGV

Hinweis: Eine Anleitung zum Ausfüllen dieser DSFA und zur KI-gestützten Ausfüllhilfe (optional, nur in der Version des Excels mit Makros) findet sich am Ende dieses Arbeitsblat

Unternehmen (Verantwortlicher):	Musterfirma AG
Abteilung:	
Verantwortlich intern:	
Status der DSFA:	
Name des Vorhabens:	
Aktivität gemäss Bearbeitu	

1. Beschreibung de

1.01 In welchem Bereic bzw. welcher Gesc

1.02 Was vorgesehen i

1.03 Welche Interesser

1.04 Welche Mittel unc

1.05 Welche Dritten an

1.06 Welche Daten bea

1.07 Wessen Daten bea

1.08 Wo überall Daten

1.09 Wann die Datenbe

1.10 Weitere Besonderl

2. Erforderlichkeit /

2.01 Warum die Datenl Sicht der betroffer eingreifende Vorg

2.02 Warum die Datenbearbeitung datensparsam, zeitlich auf das nötige begrenzt und auch sonst verhältnismässig ist: Zugriffen geschützt werden. Die Datenbearbeitung ist datensparsam, da nur der Stimmabdruck, die ID der Person und Tonaufnahmen gespeichert werden, die für die Identitätsverifizierung notwendig sind. Die Datenbearbeitung ist zeitlich begrenzt, da der Stimmabdruck bei jedem Anruf neu erstellt und nicht länger als nötig gespeichert wird. Die Datenbearbeitung ist verhältnismässig, da sie zur Sicherheit der Anrufer im Call-Center beiträgt und die einzigen Daten bearbeitet werden, die dafür erforderlich sind.

4. Risiken von negativen Folgen für die betroffenen Personen, die trotz der obigen Massnahmen verbleiben

10 Risiken vorschlagen (überschreibt bisherige Werte)

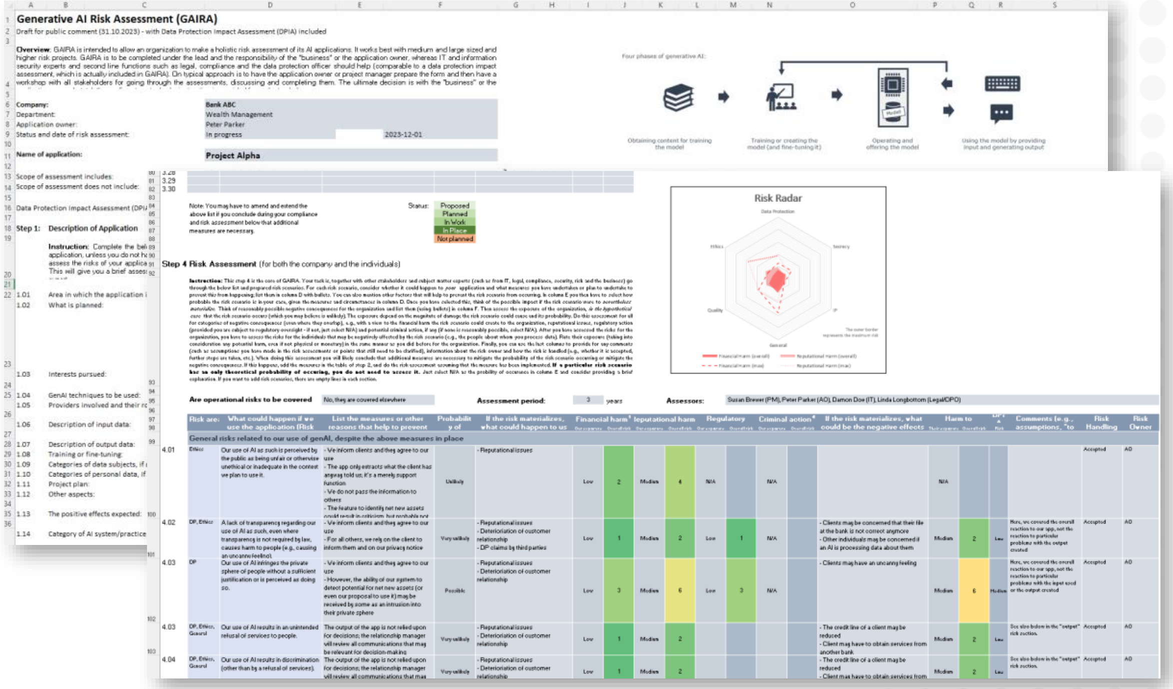
Hinweis: Falls die ermittelten Risiken als zu hoch erscheinen oder sich zeigt, dass es noch weitere Massnahmen zur Minimierung gibt, sollten diese oben unter Ziff. 3 eingetragen werden und bei der Risikobeurteilung hier berücksichtigt werden.

	Mögliche unerwünschte negative Folgen	Was wir dagegen tun	Wie wir das Restrisiko einschätzen	Mögliche Folgen für die Person	Eintrittswahrscheinlichkeit (alles in allem)	Risiko (1-16)
	Weiteres Risiko vorschlagen*	Massnahmen vorschlagen* Aus obigen formulieren*	Risikobeurteilung vorschlagen*			
4.01	Personendaten des Vorhabens gelangen wegen eines Fehlers oder absichtlich an **unbefugte Dritte**. Diese missbrauchen sie zum Schaden der betroffenen Personen.	- Berechtigungskonzept: Da wir nur autorisierten Personen Zugriff auf die Personendaten geben, wird das Risiko von unbefugtem Zugriff und Missbrauch reduziert. - Schulung: Durch Schulungen stellen wir sicher, dass die Mitarbeitenden die Lösung korrekt und sicher nutzen, was das Risiko von Fehlern und Missbrauch verringert. - Zugriffskontrolle: Durch die Beschränkung des Zugriffs auf autorisierte Personen können wir Missbräuche und unbefugte Nutzung der Personendaten verhindern. - Verschlüsselung "at rest": Die Verschlüsselung der Personendaten in unserem System schützt vor unautorisiertem Zugriff, falls jemand physischen Zugriff auf die Speichermedien erhält. - Datenlöschungsfunktionen: Durch die Möglichkeit, nicht mehr benötigte Personendaten zu löschen oder zu anonymisieren, minimieren wir das Risiko eines unbefugten Zugriffs auf diese Daten.	Das konkrete Restrisiko für die betroffene Person besteht darin, dass ihre Personendaten aufgrund eines Fehlers oder absichtlich an unbefugte Dritte gelangen könnten. Diese könnten die Daten dann zum Schaden der betroffenen Person nutzen, beispielsweise für Identitätsdiebstahl oder Missbrauch in sozialen Medien. Die Wahrscheinlichkeit dieses Szenarios ist jedoch insgesamt gering, da strenge Sicherheitsmassnahmen wie Zugriffskontrollen und Verschlüsselung implementiert wurden.	Substanziell	Tief	Mittel (6)
4.02	Personendaten des Vorhabens gelangen wegen eines Fehlers oder absichtlich an eine **unbefugte interne Person**.					

Datenschutz-Folgenabschätzung
vud.ch/dsfa
Datenschutz-Folgenabschätzungen sind ein nützliches Instrument für den Umgang mit … Der VUD stellt neu eine …führung von Datenschutz- … zur Verfügung.
mit KI-gestützter Ausfüllhilfe
2 Abwägung von Erforderlichkeit und Verhältnismässigkeit
ggf. Definition zusätzlicher Massnahmen
Bewertung von Restrisiken
3 Definition von Massnahmen
6 Schluss-beurteilung
Excel-basiert, entwickelt von David Rosenthal
Ein Projekt des Vereins Unternehmens-Datenschutz VUD
www.vud.ch | info@vud.ch

vud.ch/dpia

Tool for Generative AI Risk Assessment (GAIRA)



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 second line 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

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:

Scope of assessment does not include:

Data Protection Impact Assessment (DPIA

Step 1: Description of Application

Instruction: Complete the bel application, unless you do not ha assess the risks of your applica This will give you a brief asses

1.01 Area in which the application i

1.02 What is planned:

1.03 Interests pursued:

1.04 GenAI techniques to be used:

1.05 Providers involved and their ro

1.06 Description of input data:

1.07 Description of output data:

1.08 Training or fine-tuning:

1.09 Categories of data subjects, if

1.10 Categories of personal data, if

1.11 Project plan:

1.12 Other aspects:

1.13 The positive effects expected:

1.14 Category of AI system/practice

Four phases of generative AI:

Obtaining content for training the model
Training or creating the model (and fine-tuning it)
Model
Operating and offering the model
Using the model by providing input and generating output

3.28

3.29

3.30

Note: You may have to amend and extend the above list if you conclude during your compliance and risk assessment below that additional measures are necessary.

Status: Proposed
Planned
In Work
In Place
Not planned

Step 4 Risk Assessment (for both the company and the individuals)

Instruction: This step 4 is the core of GAIRA. Your task is, together with other stakeholders and subject matter experts (such as from IT, legal, compliance, security, risk and the business) go through the below list and prepared risk scenarios. For each risk scenario, consider whether it could happen to *your* application and what measures you have undertaken or plan to undertake to prevent this from happening; list them in column D with bullets. You can also mention other factors that will help to prevent the risk scenario from occurring. In column E you then have to select how probable the risk scenario is in your case, given the measures and circumstances in column D. Once you have selected this, think of the possible impact if the risk scenario were to *nevertheless materialize*. Think of reasonably possible negative consequences for the organization and list them (using bullets) in column F. Then assess the exposure of the organization, *in the hypothetical case* that the risk scenario occurs (which you may believe is unlikely). The exposure depend on the magnitude of damage the risk scenario could cause and its probability. Do this assessment for all for categories of negative consequences (even where they overlap), e.g., with a view to the financial harm the risk scenario could create to the organization, reputational issues, regulatory action (provided you are subject to regulatory oversight - if not, just select N/A) and potential criminal action, if any (if none is reasonably possible, select N/A). After you have assessed the risks for the organization, you have to assess the risks for the individuals that may be negatively affected by the risk scenario (e.g., the people about whom you process data). Rate their exposure (taking into consideration any potential harm, even if not physical or monetary) in the same manner as you did before for the organization. Finally, you can use the last columns to provide for any comments (such as assumptions you have made in the risk assessments or points that still need to be clarified), information about the risk owner and how the risk is handled (e.g., whether it is accepted, further steps are taken, etc.). When doing this assessment you will likely conclude that additional measures are necessary to mitigate the probability of the risk scenario occurring or mitigate the negative consequences. If this happens, add the measures in the table of step 2, and do the risk assessment assuming that the measure has been implemented. **If a particular risk scenario has an only theoretical probability of occuring, you do not need to assess it.** Just select N/A as the probility of occurrence in column E and consider providing a brief explanation. If you want to add risk scenarios, there are empty lines in each section.

Are operational risks to be covered No, they are covered elsewhere

Assessment period: 3 years

Assessors: Susan Brewer (PM), Peter Parker (AO), Damon Doe (IT), Linda Longbottom (Legal/DPO)

	Risk are:	What could happen if we use the application (Risk	List the measures or other reasons that help to prevent	Probability of	If the risk materializes, what could happen to us	Financial harm[1]		Reputational harm		Regulatory		Criminal action[4]		If the risk materializes, what could be the negative effects	Harm to		DPIA Risk	Comments (e.g., assumptions, "to	Risk Handling	Risk Owner
						Our exposure	Overall risk	Our exposure	Overall risk	Our exposure	Overall risk	Our exposure	Overall risk		Their exposure	Overall risk				
General risks related to our use of genAI, despite the above measures in place																				
4.01	Ethics	Our use of AI as such is perceived by the public as being unfair or otherwise unethical or inadequate in the context we plan to use it.	- We inform clients and they agree to our use - The app only extracts what the client has anyway told us; it's a merely support function - We do not pass the information to others - The feature to identify net new assets could result in criticism, but probable not	Unlikely	- Reputational issues	Low	2	Medium	4	N/A		N/A			N/A				Accepted	AO
4.02	DP, Ethics	A lack of transparency regarding our use of AI as such, even where transparency is not required by law, causes harm to people (e.g., causing an uncanny feeling).	- We inform clients and they agree to our use - For all others, we rely on the client to inform them and on our privacy notice	Very unlikely	- Reputational issues - Deterioration of customer relationship - DP claims by third parties	Low	1	Medium	2	Low	1	N/A		- Clients may be concerned that their file at the bank is not correct anymore - Other individuals may be concerned if an AI is processing data about them	Medium	2	Low	Here, we covered the overall reaction to our app, not the reaction to particular problems with the output created	Accepted	AO
4.03	DP	Our use of AI infringes the private sphere of people without a sufficient justification or is perceived as doing so.	- We inform clients and they agree to our use - However, the ability of our system to detect potential for net new assets (or even our proposal to use it) may be received by some as an intrusion into their private sphere	Possible	- Reputational issues - Deterioration of customer relationship	Low	3	Medium	6	Low	3	N/A		- Clients may have an uncanny feeling	Medium	6	Medium	Here, we covered the overall reaction to our app, not the reaction to particular problems with the input used or the output created	Accepted	AO
4.03	DP, Ethics, General	Our use of AI results in an unintended refusal of services to people.	The output of the app is not relied upon for decisions; the relationship manager will review all communications that may be relevant for decision-making	Very unlikely	- Reputational issues - Deterioration of customer relationship	Low	1	Medium	2					- The credit line of a client may be reduced - Client may have to obtain services from another bank	Medium	2	Low	See also below in the "output" risk section.	Accepted	AO
4.04	DP, Ethics, General	Our use of AI results in discrimination (other than by a refusal of services).	The output of the app is not relied upon for decisions; the relationship manager will review all communications that may	Very unlikely	- Reputational issues - Deterioration of customer relationship	Low	1	Medium	2					- The credit line of a client may be reduced - Client may have to obtain services from	Medium	2	Low	See also below in the "output" risk section.	Accepted	AO

- Risk assessment includes a basic DPIA
- Separate compliance check
- ROAIA template

Download it at https://www.rosenthal.ch/downloads/Rosenthal_GAIRA.xlsx

GenAI risk assessments

- Use the **same approach** as for **DPIAs**:
  - 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**

How to manage the risks

1. Gain ownership of the (legal) topic, or at least part of it
2. Establish an overview of what is going on, create the ROAIA
3. Understand where existing tools permit the use of GenAI
4. Instruct and train employees in the proper use of GenAI
5. Establish that any new AI service used will go by you first, so you make sure that there are proper contracts and the basic compliance requirements are met
6. Have the application owner perform a DPIA (as applicable) and a structured and documented assessment of GenAI risks
7. Install monitoring, re-assessment and incident reporting processes and act upon findings and reports

Final remarks

- The data protection rules **we already know** and are used to generally work well also for GenAI applications
- Do not confuse data protection and **data ethics**
- There will be a "legal" **demystification**; people will realize that with proper contracts and processor setups, feeding a GenAI system with personal or secret data is not necessarily a big risk or problem – the main concern is the 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

VISCHER

Thank you for your attention!

Questions: drosenthal@vischer.com

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

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

www.vischer.com