#### Key Risk Factors of Al Applications ReMeP Annual Conference 2024 Peter Kieseberg



/fh///

st.pölte



/fh/// st.pölten

# Major impact





#### Now and the Future





### New security challenges



New attacks & risks



Stability & validity



Privacy



Control over data / models



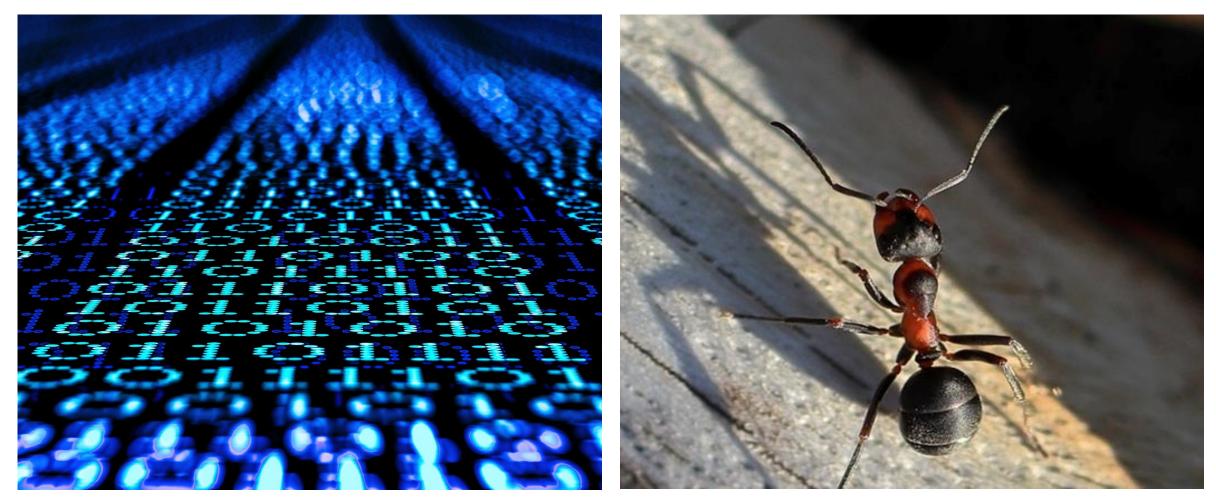
#### Bias & Data preparation



Procurement

# Data Defined Software & Explainability





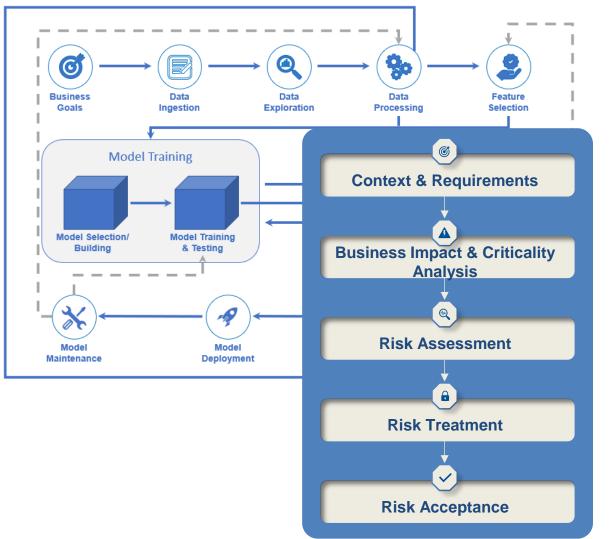
#### /fh/// st.pölten

# Ubiquitous AI $\rightarrow$ What do we actually do?



- What does the system do?
- Am I currently using AI?
- What risks arise?
- How do I deal with them?
- How do I make sensible decisions?
- Attribution problem

# Al Risk Management



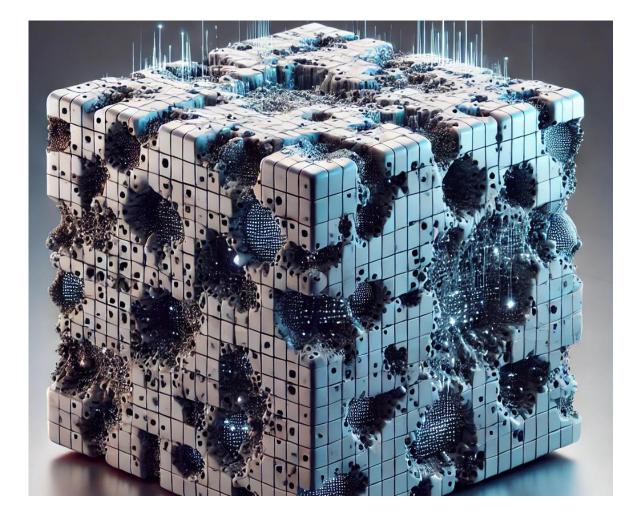
- Not ever AI is the same!
- Black box vs. understanding AI
- Risk & technology assessment
- Threat Intelligence
- Procurement vs. in-house development vs. customisation

/fh//

st. pölte

#### **Data & Data Preparation**





- (Non-) intentional bias, quality
- Data Cleansing
- Data pre-processing
- Reliable sources
- Protecting valuable data
- Dominant data pools / monopolies

# Models & transparency





- Transparency
  - Who trains? What/what is trained on?
  - ChatGPT wrapper?
- Patching
  - Data and models
- Reinforcement learning
- Re- & Post-Processing
- Logging and monitoring

Nutzungszeiten, Anzahl Datensätze ....)?

Bei Virtualisierung:

- a. Zuordnung des Containers zu einer physischen Maschine oder einer Geo-Location möglich?
- b. Abschottungsmaßnahmen zwischen den Containern?

Source Code				
Wofür?	Funktionalitätsgarantie (macht die Komponente was sie soll oder ungewollt mehr?) bzw. Weiterentwicklungsmöglichkeiten			
Indikatoren	Aufgaben	Nicht-Auf	gaben	Erweiterbarkeit
1. Wird der Source-Code der Anwendung zur Verfügung gestellt (auch bei Cloud und SaaS)?				
	↓ Ja ↓			↓ Nein ↓
Kann die Anwen	rden? Wer g	Wer garantiert die Korrektheit? (v.a. on-premise)		
Ja ↓	Ne	ein		<b>_</b>
Sind dafür spez notwendig?		Wie wird garantiert, dass Source Code und Compilat einander entsprechen?		
Ja	Ne	<b>i</b> n		
Auflistung notwendig. Kann in diese Einblick				

- How much in-house modification possible?
- A slight case of overselling ...
  - LLMs everywhere
- What do we buy?
  - Full system, model, AI as a service ...
- Who is responsible?

## The future - Explorative scenario analysis





### Chances

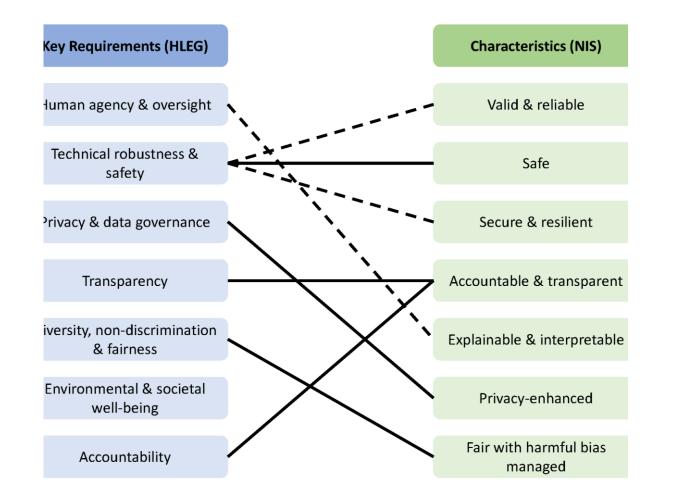




- CISO can benefit
- Time for risk management
- Documentation → YSRYFUM
- Thinking before deployment
- Traceability for data Questioning data
- Creation of own AI components & AI security industry

# Trustworthy vs. Controllable AI









# Question of jurisdiction





# Thank you very much!



