DORDA

AI Act – Ein Drahtseilakt zwischen Sicherheit und Innovation

Cyber Crime Forum – 12.6.2025

Axel Anderl

Managing Partner und Leiter des IT/IP und Datenschutzteams

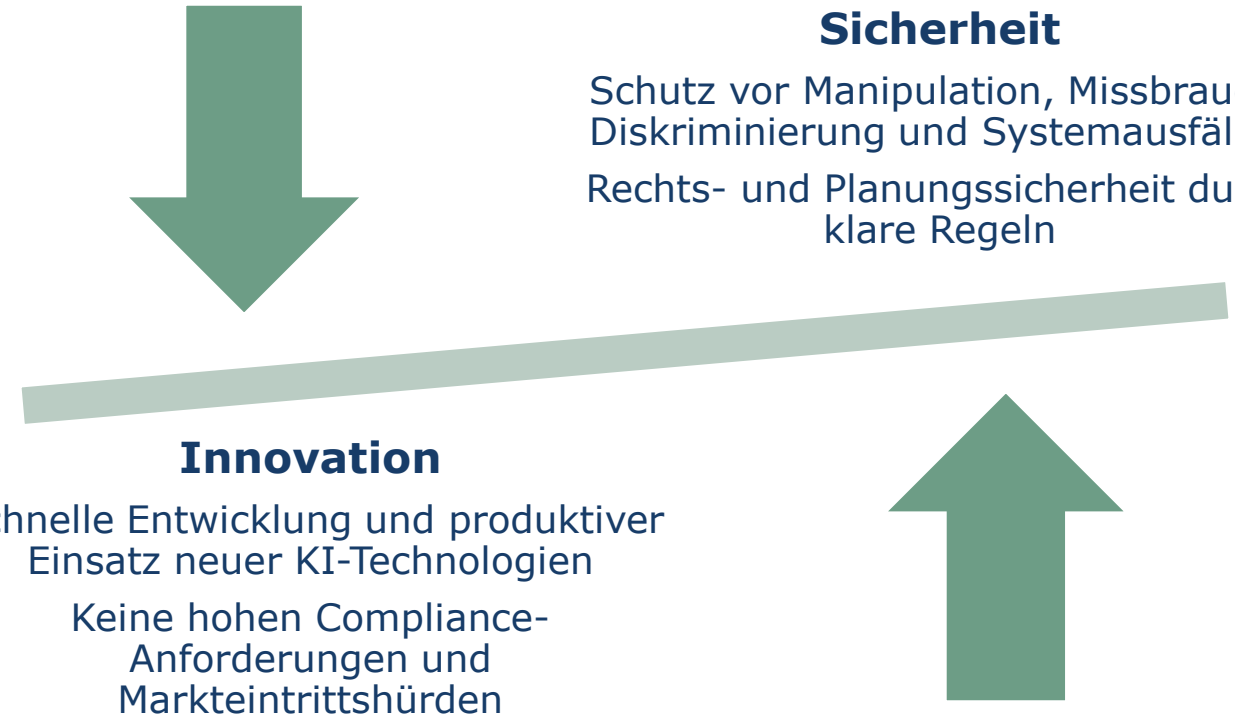
- Schwerpunkte: IT- und IP-Recht, insb Out- und Cloudsourcing, Cybersecurity, NISG Compliance, Urheberrechtsabgabe und Kunstrecht
- Empfohlen als first tier für IT, IP und Datenschutz bei Legal500 und IT und IP bei Chambers Europe
- Legal500 Hall of Fame für TMT
- Leading Individual für IP bei JUVE
- TIER 1 Media Law International 2025
- Zwölf ILO Client Choice Award für Information Technology & Internet
- Absolvent der Universität Wien (Dr iur 2005) und des Universitätslehrgangs für Informationsrecht und Rechtsinformation der Universität Wien (LL.M. 2001)
- Autor zahlreicher Fachpublikationen, ua #Cybercrime (LexisNexis), NISG Kurzkommentar (Manz), KI-VO Kurzkommentar (Manz), Handbuch UWG (Linde), "*IP in der Praxis*" (Verlag Manz), #blockchain (LexisNexis)
- Vortragender und Lektor an zahlreichen Hochschulen und Fachhochschulen sowie bei diversen Seminaranbietern
- Board-Mitglied von ITechLaw und Co-Chair Start-Up Committee

Chambers TOP RANKED Europe 2025 Axel Anderl

Axel Anderl

axel.anderl@dorda.at

Sicherheit vs Innovation



Sicherheit

Schutz vor Manipulation, Missbrauch, Diskriminierung und Systemausfällen

Rechts- und Planungssicherheit durch klare Regeln

Innovation

Schnelle Entwicklung und produktiver Einsatz neuer KI-Technologien

Keine hohen Compliance-Anforderungen und Markteintrittshürden

Ziel, Anwendungsbereich und Regelungsmechanismus

Ziele

- Menschenzentrierte, vertrauenswürdige KI
- Hohes Schutzniveau für Gesundheit, Sicherheit, Grundrechte und vor schädlichen Auswirkungen
- Innovationsschutz

Anwendungsbereich

- **Sämtliche in EU** angebotene, betriebene oder eingeführte KI
- **Ausnahmen** für militärische, rein private Zwecke sowie Forschung

Regelungsmechanismus

- Einstufung von KI-Systeme in **Risikoklassen**
- **Compliance-pflichten** nach Gefahr der KI
- **Risk based approach**

Haftung in der Lieferkette

Anbieter

Hersteller und Vertreiber unter eigenem Namen

Einführer

Importeur mit Niederlassung in der EU, der die KI aus einem Drittland in die EU einführt

Händler

Anbieter auf dem Unionsmarkt

Betreiber

Verwendung in eigener Verantwortung

Risiko-Kategorien

Verbotene KI

zB:

Social Scoring

umfassender Scan von Gesichtern zur Erstellung von Datenbanken

bestimmte Formen Emotionserkennung

---

absolutes **Verbot**

Hochrisiko KI

zB:

KI-Tools für Bewerberauswahl oder Credit-Scoring

KI als Sicherheits-bestandteil von Maschinen

Umfangreiche **Compliance-Pflichten** bei Entwicklung und Betrieb; verpflichtende Zertifizierung

KI-Modelle mit allgemeinem Verwendungs-zweck

zB:

OpenAI GPT

Modell für große Anzahl **unbestimmter Zwecken** eingesetzbar

bestimmte KI

zB:

Midjourney

diverse Chatbots

KI-System **interagiert direkt** mit natürlicher Person oder erzeugt **synthetische Outputs**

Anforderungen an Cybersicherheit

- Verpflichtung für Anbieter von Hochrisiko-KI-Systemen und KI-Modellen mit allgemeinem Verwendungszeck und systematischen Risiken

Hochrisiko-KI-Systeme

Gewährleistung Cybersicherheit während gesamten Lebenszyklus

- Manipulation der Trainingsdatensätze
- Manipulation von trainierten Komponenten
- Fehlerhafte Eingabedaten

KI-Modellen mit allgemeinem Verwendungszweck mit systematischem Risiko

Angemessenes Maß an Cybersicherheit des KI-Modells und physischer Infrastruktur

- Widerstandsfähigkeit gegenüber böswilliger Angriffe Dritter (zB Ausnutzung der KI-Prozesse, Leistung oder Sicherheit)
- Sicherheit des KI-Systems (zB vor "*data poisoning*") und der IKT-Infrastruktur

- Zusätzliche Anforderungen durch Cyber Security Act und Produktsicherheitsverordnung

Beispiele – KI in der Cybersicherheit

KI- und LLM-basierte Verteidigung

- **Identify**: Überprüfung von Softwareschwachstellen des Quellcodes
- **Protect**: Erstellung adaptiver und personalisierter Cybersecurity-Schulungen oder –Empfehlungen
- **Protect**: Dynamisch und optimierte Backup-Planung
- **Protect**: Anti-Virus bzw -Malware Lösungen
- **Detect**: Überwachung des System- und Netzwerkverkehrs, um anormale Aktivitäten und mögliche Eindringversuche zu erkennen
- **Detect**: Analyse und Identifizierung von Phishing-Nachrichten und –Websites (zB mittels Webcrawaling)
- **Respond**: Suchen von ähnlichen Vorfällen
- **Recover**: Zukünftige Reaktionspläne basierend auf historischen Berichten und Protokolle identifizieren

Angriffe auf und mit KI

- Automatisierte **Angriffe** (zB DDoS-Angriffe)
- Erstellung personalisierter **Phishing-Nachrichten** aus öffentlichen Daten
- Erstellung von **Phishing-Websites**
- Erraten von Passwörtern mit LLMs, die mit Passwortlecks trainiert wurden
- **Manipulation** von Trainingsdaten (Poisoning Attacks)
- **Daten- und Modellextraktionen** durch gezielte (automatisierte) Abfragen
- **Veränderung** der Funktionsweise GenAI-System (zB für Hassreden oder Diskriminierung)
- **Deepfakes**

Beispiel – Deepfake Millionenbetrug

- Vorfall aus 2024 in UK
- Mitarbeiter erhielt Phishing-E-Mail zu "*geheimer Transaktion*"
- Auslöser für Video-Konferenz mit Deepfakes
  - KI-generierte Personenbilder und Stimmen, inklusive CFO des Unternehmens
- Mitarbeiter veranlasste Überweisung iHv EUR 13 Millionen an diverse Konten

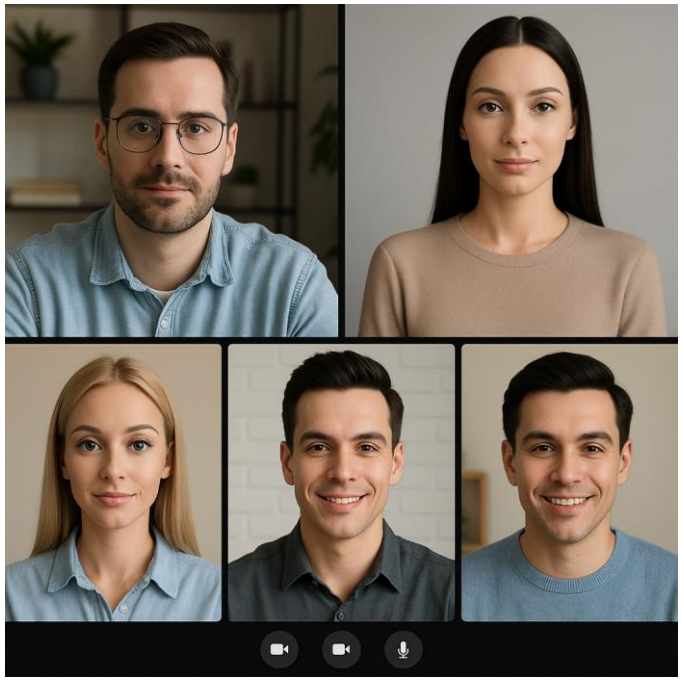
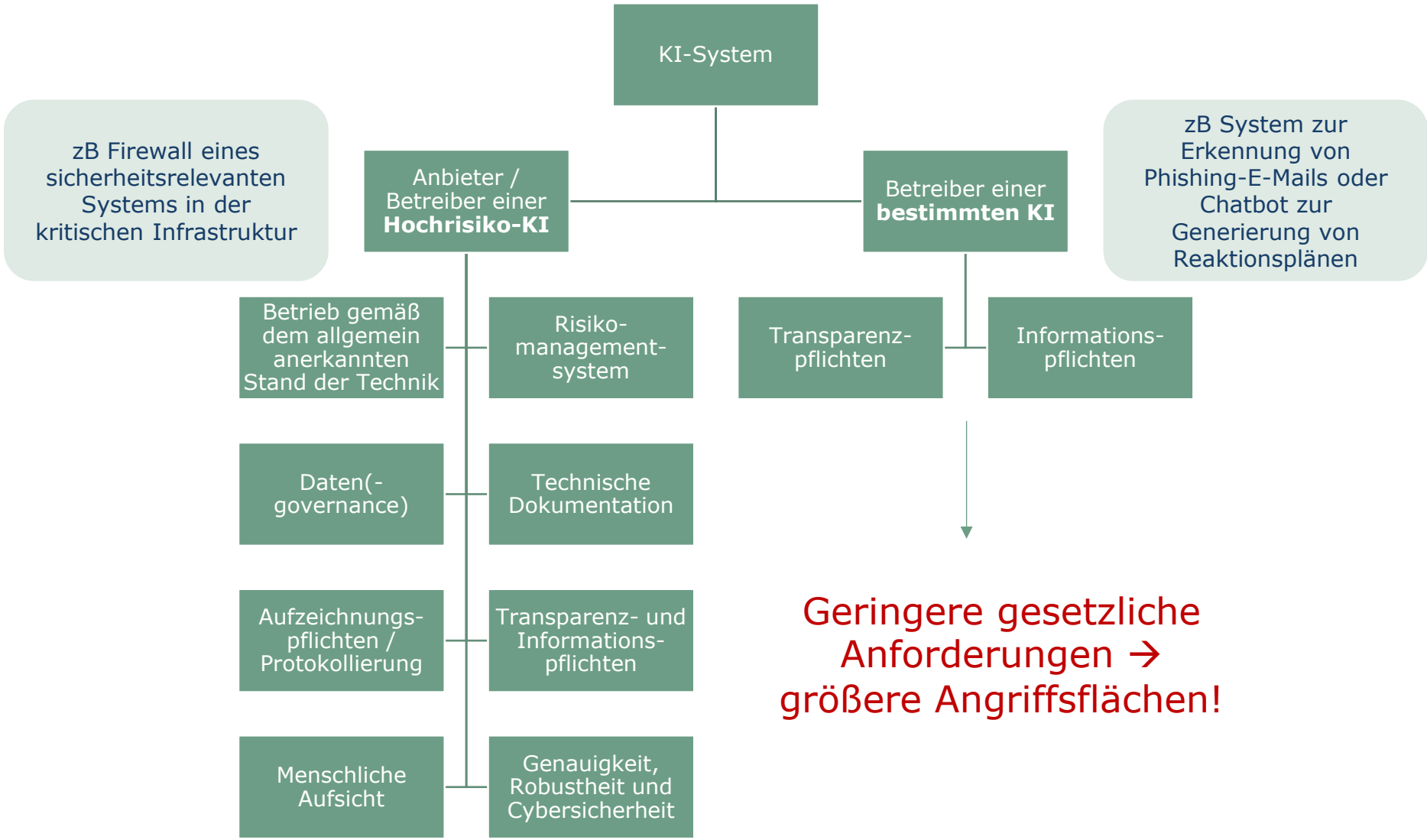


Bild generiert mit ChatGPT 4o am 10.6.2025 mit folgendem Prompt: Generiere mir bitte ein Bild. Auf dem Bild sind fünf Personen zu sehen. Eine Person ist ein Mensch, die restlichen Personen sind Deepfakes und entsprechend darzustellen. Die Situation ist wie in einer Video-Konferenz darzustellen.

Link: Deepfakes in Video-Konferenz: Geklonter Finanzvorstand ordnet Millionenbetrug an - n-tv.de

Risikoerhöhung aufgrund unterschiedlicher gesetzlicher Pflichten im AI Act



Vielen Dank für Ihre Aufmerksamkeit!

DORDA

Dr Axel Anderl, LLM

T: +43 1 533 47 95 – 23

axel.anderl@dorda.at

Client Choice Award Lexology 2011-2025: IT & Internet
Hall of Fame Legal 500: TMT
TIER 1 Legal500 2007-2025: TMT
TIER 1 Legal500 2020-2025: Data Privacy & Data Protection
TIER 1 Legal500 2021-2025: Intellectual Property
BAND 1 Chambers Europe 2008-2025: TMT:IT
BAND 1 Chambers Europe 2025: TMT:Data Protection
BAND 1 Chambers Europe 2025: IP

DORDA

CLARITY.