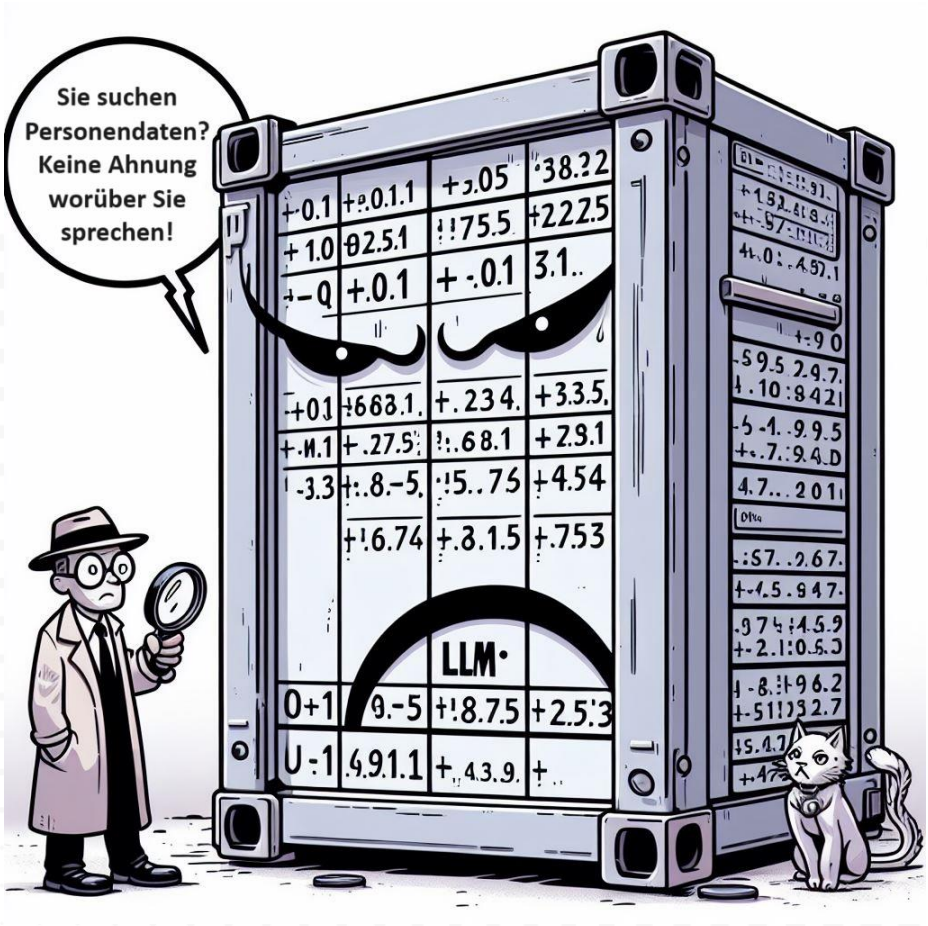
VISCHER

KI Governance.

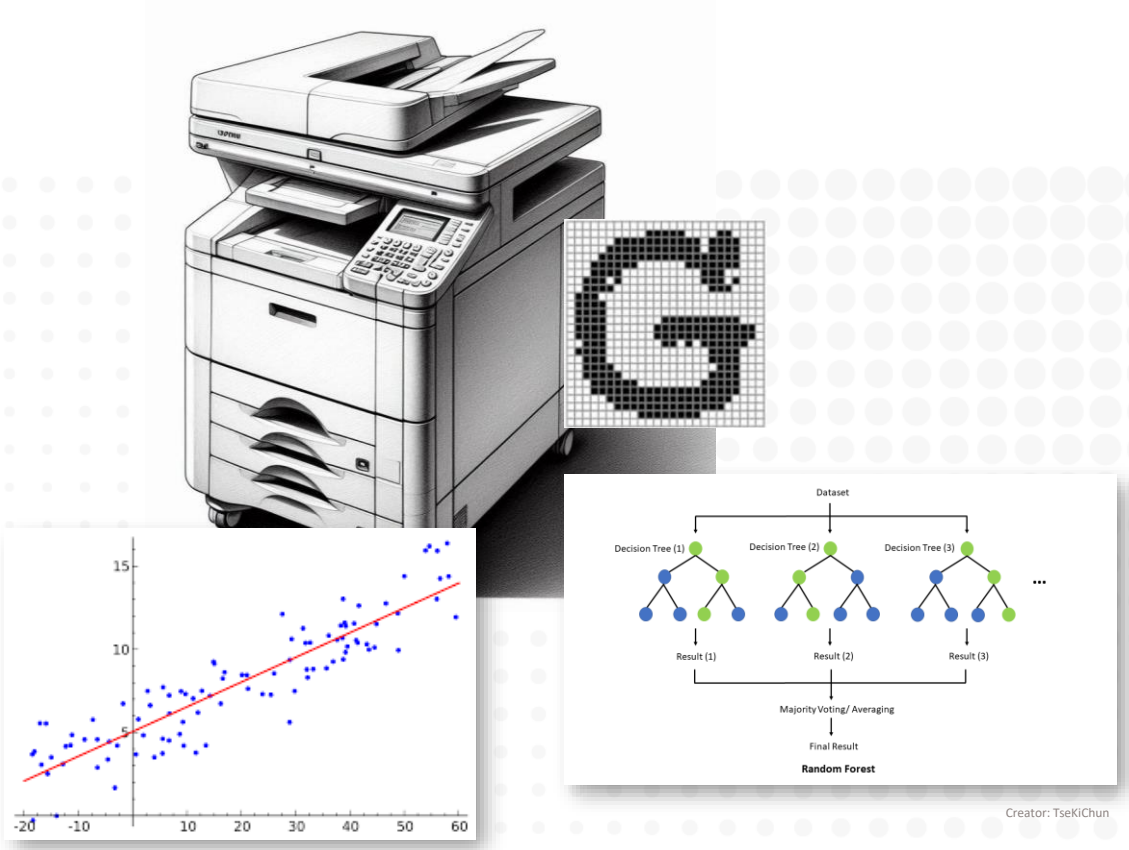
Die Fragen, die Verwaltungsräte (auch sich) stellen müssen

David Rosenthal, Partner, VISCHER AG
25. Oktober 2024

Haben Sie eine Vorstellung davon, was in einem grossen Sprachmodell steckt?



Wissen Sie eigentlich überhaupt, was KI ist?



Drei Aufgaben in Sachen KI

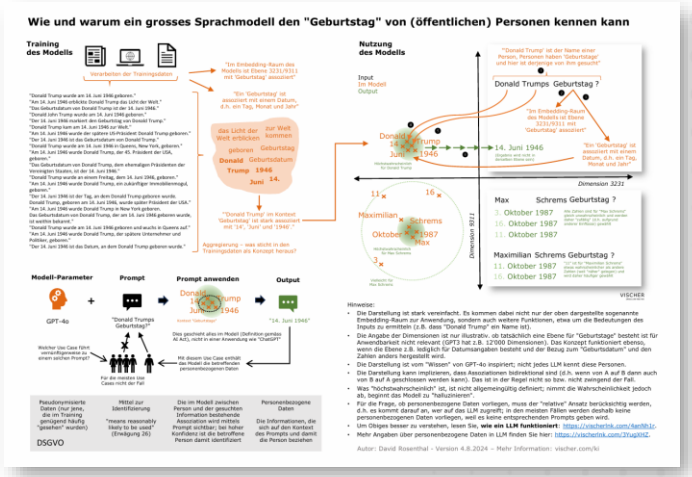
1. **Verstehen** – KI, ihre Möglichkeiten und ihre Risiken im Grundsatz kennenlernen
2. **Vorgaben** – Eine Vorstellung darüber bilden, was Ihr Unternehmen damit tun soll und was es bleiben lassen sollte
3. **Überwachen** – Ihrer Organisation kritische Fragen stellen, um sicherzustellen, dass sie die nötige Governance im Bereich KI betreibt

1. Verständnis bilden

- Sie müssen **kein Experte** sein, um über KI zu befinden
- **Aber:** Sie sollten die Wirkungsweise, Möglichkeiten und Risiken von KI im Grundsatz verstehen, namentlich …
  - den Unterschied zwischen generativer KI und analytischer KI
  - wo KI schon benutzt wird (KI: trainiert statt nur programmiert)
  - grob die Stärken und Schwächen der diversen KI-Techniken
- Persönliche Empfehlung: **Entmystifizieren** Sie das Thema!
  - Verstehen, wie Sprachmodelle und neuronale Netze funktionieren
  - Verstehen, wie KI implementiert werden kann (z.B. "RAG")
  - Verstehen, welche Machine-Learning-Techniken Ihr Betrieb nutzt
  - Verstehen, welche Schwächen und Sicherheitsrisiken diese haben

Für das Selbststudium …

Noch mehr: vischer.com/ki



vischerlnk.com/3zAoqeb

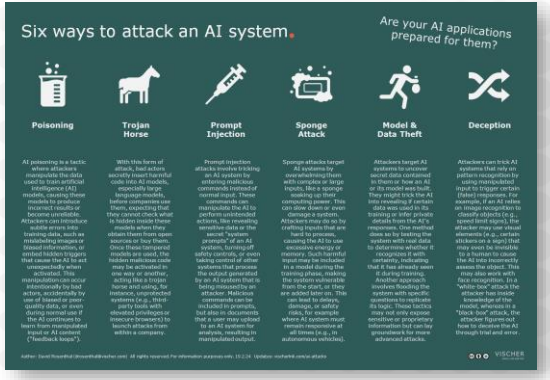
Wie funktioniert ein grosses Sprachmodell und was ist wirklich darin gespeichert?

Wie funktioniert ein neuronales Netzwerk?

vischerlnk.com/4anNh1r

Wie kann eine KI angegriffen werden?

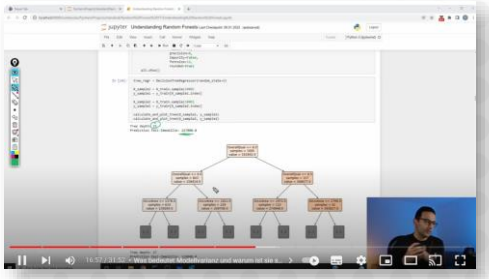
z.B. durch manipulierte Prompts, welche die Programmierung eines Chatbots überlisten



vischerlnk.com/3OPTpaA

Wer ein Beispiel zu analytischer KI sucht:

Erklärvideo "Random Forest"



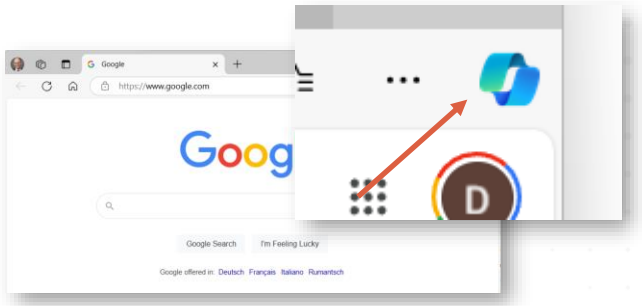
www.youtube.com/watch?v=4f3iE8Tp1cg

Für das Selbststudium

Noch mehr: vischer.com/ki

Fragen Sie ChatGPT oder "Copilot":

Erkläre mir (1) in sehr einfachen Worten für einen Nicht-Techniker, aber trotzdem sehr konkret, was "RAG" im Bereich KI bedeutet, (2) gib mir drei ganz praktische und weitverbreitete Beispiele und wie sie eingesetzt werden und (3) erkläre mir, was der besondere Vorteil von RAG ist.



(1) Was bedeutet "RAG" im Bereich KI?

"RAG" steht für **Retrieval-Augmented Generation**. Es ist ein Ansatz in der Künstlichen Intelligenz (KI), bei dem eine KI nicht nur auf das trainierte Wissen zurückgreift, sondern in Echtzeit Informationen aus externen Datenquellen abruft. Stell dir vor, die KI hat eine Art digitales Gedächtnis (ihr Modell), aber sie kann zusätzlich noch Bücher oder das Internet "lesen", wenn sie eine Frage beantworten soll. Sie holt sich also die aktuellsten Informationen von außen und kombiniert diese mit ihrem eigenen Wissen, um genauere und relevantere Antworten zu geben.

(2) Drei praktische und weitverbreitete Beispiele für RAG:

1. **Suchmaschinen (wie Google)**: RAG-Technologien kommen in modernen Suchmaschinen zum Einsatz, wo die KI auf eine Frage nicht nur allgemeine Antworten gibt, sondern auch passende Informationen aus dem Internet abruft, um genauer und relevanter zu sein. Beispielsweise bei Suchanfragen zu aktuellen Nachrichten kann die KI externe Artikel finden und in die Antwort einfließen lassen.
2. **Chatbots im Kundenservice**: Viele Unternehmen setzen KI-gestützte Chatbots ein, die mit RAG arbeiten. Wenn du eine Frage zu einem Produkt hast, kann der Chatbot nicht nur mit den Informationen aus dem allgemeinen Kundenhandbuch

2. Was das Unternehmen damit soll

- Den (bewussten) Einsatz von KI **einschränken** oder **fördern**?
- KI als blosses **Hilfsmittel** für mehr Effizienz im Alltag oder KI als **Enabler** für unternehmensspezifische Vorhaben?
  - Nicht nur an generative KI denken!
- Use Cases, die nicht nur cool sind, sondern sich **auch lohnen**?
  - Es gibt noch viel heisse Luft, ganz besonders bei Drittanbietern
  - Kosten rechtlicher und qualitativer Anforderungen beachten
- Welche **Risiken** ist das Unternehmen bereit einzugehen?
  - Risikobereitschaft nimmt zu – womit fühlen Sie sich wohl?
  - Operationelle, finanzielle, reputative und rechtliche Risiken
  - Compliance kann auch "ethische" Vorgaben umfassen


Air Canada

Chatbot verspricht Fluggast irrtümlich Rückerstattung – Airline muss zahlen

Er gewährte einen Rabatt, den es gar nicht gab: Weil ein Chatbot gegen die eigenen Richtlinien verstieß, wollte Air Canada einen Kunden auf dessen Kosten sitzen lassen. Nun entschied ein Gericht gegen das Unternehmen.

Quelle: Spiegel.de

Es ging um 500 Franken …

Die wichtigsten Compliance-Fragen

Checkliste: 18 KI-Compliance-Schlüsselfragen.

Unter vischer.com/ki finden Sie kostenlose Ressourcen zu diesen Themen sowie zu KI-Governance und Risikomanagement (keine Registrierung erforderlich)

KI = System, das Ergebnisse auf Basis eines Trainings und nicht nur einer Programmierung erzeugt

Datenschutz

- Haben wir einen angemessenen Vertrag mit den von uns genutzten Providern (z.B. einen ADV, EU SCC, Verbot der Eigennutzung unserer Daten)?
- Haben wir die Leute über die Zwecke informiert, zu denen wir Daten von ihnen bearbeiten oder erzeugen?
- Haben wir es im Griff, wenn die KI falsche oder anderweitig unzulässige Daten über sie produziert?
- Wenn eine KI wichtige Entscheidungen über sie trifft, können sie diese von einem Menschen prüfen lassen?
- Ist unsere KI vor Missbrauch und Angriffen geschützt und auch sonst sicher, insbesondere, wo wir Dritten die Nutzung erlauben (z.B. Chatbot)?
- Können wir Auskunfts- und Berichtigungsbegehren wie erforderlich umsetzen?
- Haben wir eine Risikobeurteilung für unser Vorhaben (inklusive einer DSFA) durchgeführt?

Vertragspflichten, Geheimhaltung

- Kommen wir unseren Geheimhaltungspflichten nach (z.B. beim Einsatz von Providern, Verhinderung der unerwünschten Preisgabe von Daten)?
- Untersagen unsere Verträge die von uns ins Auge gefasste Anwendung (z.B. NDA, welches die Nutzung von Daten für unsere Zwecke einschränkt)?

Schutz von Inhalten Dritter

- Füttern wir KI-Systeme nur dann mit Inhalten Dritter, soweit unsere Lizenzen oder die gesetzlichen Schranken des Urheberrechts dies zulassen?
- Vermeiden wir die Erstellung von Inhalten, die bereits bestehenden Inhalten Dritter entsprechen?

EU AI Act (noch nicht in Kraft)

- Ist klar, dass wir entweder nicht unter den EU AI Act fallen oder unser Vorhaben keine verbotene Praktik ist und möglichst auch kein "Hoch-Risiko"-KI-System (und gehen wir ansonsten richtig damit um)?
- Wenn eine KI "Deep Fakes" erstellt oder mit Menschen interagiert oder sie beobachtet, werden sie dann darauf hingewiesen gemacht?

Andere (auch ethische) Aspekte

- Vermeiden wir Diskriminierung beim Einsatz von KI?
- Behält der Mensch (wirklich) die Kontrolle über die KI?
- Können wir unsere KI-Ergebnisse rechtfertigen/erklären?
- Sagen wir es den Leuten, wie wir KI einsetzen, wenn es für sie unerwartet sein könnte, und erlauben wir ihnen gar, sich für oder gegen deren Einsatz zu entscheiden?
- Haben wir ein angemessenes KI-Testing, angemessene Überwachung und ein angemessenes Risk-Management?

VISCHER

Blog-Beitrag und weitere Infos:

https://bit.ly/3WNgxeO
https://vischer.com/ki

Hier muss jedes Unternehmen seinen Weg finden

vischerlnk.com/ki-compliance-kurz

Verantwortungsvolle KI: Die typischen Vorgaben

- Wir sorgen für **Verantwortlichkeit**
- Wir sorgen für die nötige **Transparenz**
- Wir bleiben **fair** und **schaden nicht**
- Wir sorgen für **Zuverlässigkeit**
- Wir sorgen für **Informationssicherheit**
- Wir achten auf **Verhältnismässigkeit** und **Selbstbestimmung**
- Wir respektieren fremdes und eigenes **geistiges Eigentum**
- Wir wahren die **Rechte der Betroffenen**
- Wir sorgen für **Erklärbarkeit** und **menschliche Aufsicht**
- Wir verstehen und kontrollieren die **Risiken**
- Wir verhindern **Missbräuche** unserer KI-Anwendungen

Diese Vorgaben gehen oft über das gesetzliche Minimum hinaus

Basierend auf DSG, DSGVO, UWG, URG, StGB, KI-K, AI Act

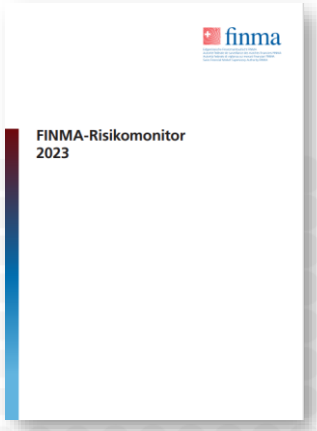
3. Kritische Fragen zur Governance

1. Wissen wir, **wo** wir (relevante) KI im Unternehmen einsetzen?
2. Haben wir Aufgaben, Kompetenzen und Verantwortlichkeiten ("**AKV**") in Sachen KI *lege artis* geregelt, inkl. Ownership?
3. Haben wir die nötige **Fachkompetenz** zum Einsatz von KI?
4. Haben wir unsere Mitarbeitenden im Umgang mit KI **geschult**?
5. Haben wir zu uns passende **Vorgaben** zum Umgang mit KI?
6. Wird KI **kontrolliert** eingeführt und der Einsatz **überwacht**?
7. Betreiben wir ein **Risiko-Management** im Bereich KI, das operationelle, finanzielle, rechtliche und reputative Risiken abdeckt? Wer? Wo sind die grössten Risiken? Massnahmen?
8. Ist der **Verwaltungsrat** in riskantere Vorhaben eingebunden?

Verfolgen Sie einen risikobasierten Ansatz

- Nicht jede Anwendung von KI hat **relevante** oder gar hohe **Risiken**
  - Fokussieren, Experimente zulassen
- Lassen Sie die Risiken **strukturiert prüfen** und dokumentieren
- Augenmerk auf **Risiko-Trigger**, z.B.
  - Biometrie, Emotionen, KI betreffend Personal, sensitive Personendaten, Berufs- und Geschäftsgeheimnisse, urheberrechtlich geschützte Inhalte, Einsatz von Service-Providern, KI-Angebote für Dritte, Automatisierung wichtiger Entscheide, Bias, Training

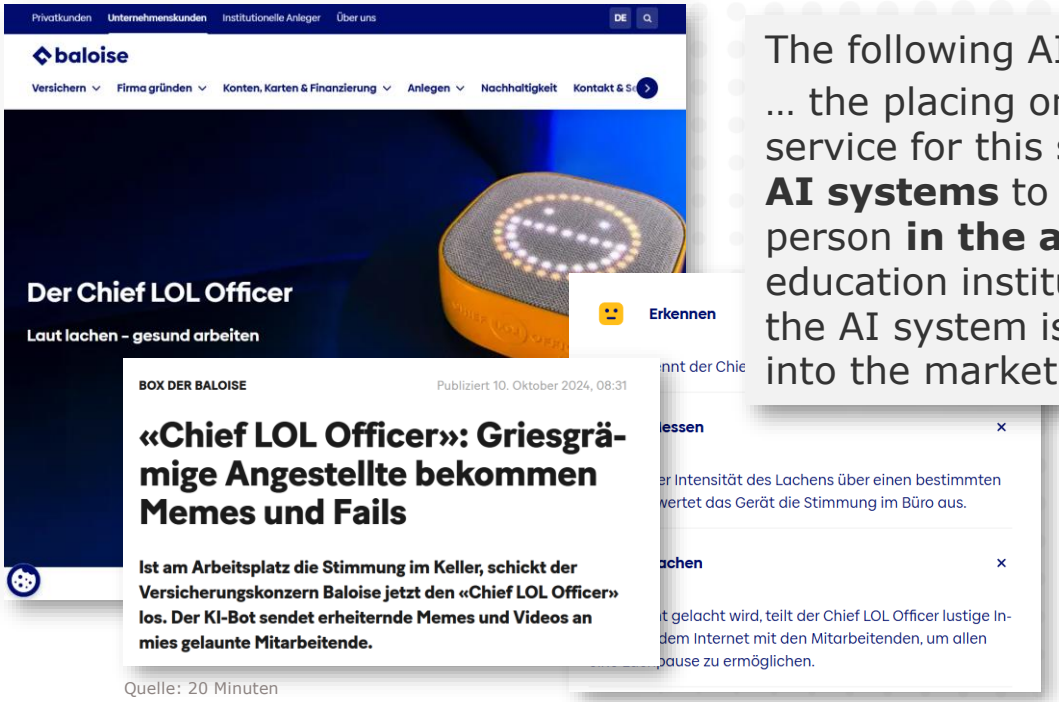
Gutes Beispiel: Erwartungen der FINMA



1. Es müssen klare Rollen und Verantwortlichkeiten sowie Risikomanagementprozesse definiert und implementiert werden. Die Verantwortung für Entscheidungen kann nicht an KI oder Drittparteien delegiert werden. Alle Beteiligten müssen über genügend Know-how im Bereich KI verfügen.
2. Bei der Entwicklung, der Anpassung und in der Anwendung von KI ist sicherzustellen, dass die Ergebnisse hinreichend genau, robust und zuverlässig sind. Dabei sind sowohl die Daten als auch die Modelle und die Resultate kritisch zu hinterfragen.
3. Die Erklärbarkeit der Resultate einer Anwendung sowie die Transparenz über deren Einsatz sind je nach Empfänger, Relevanz und Prozessintegration sicherzustellen.
4. Nicht begründbare Ungleichbehandlung ist zu vermeiden.

Blog-Beitrag:
vischerlnk.com/3MRHurk

Schlechtes Beispiel: Chief LOL Officer



Quelle: 20 Minuten

The following AI practices shall be **prohibited**:

… the placing on the market, the putting into service for this specific purpose, or the **use of AI systems** to **infer emotions** of a natural person **in the areas of workplace** and education institutions, except where the use of the AI system is intended to be put in place or into the market for medical or safety reasons;

Die für die rechtliche Prüfung zuständigen Stellen wussten offenbar nichts davon …

Wie geht es weiter?

- Die Materie mag heute für viele von uns **unheimlich und ungewohnt** sein
  - Das war das Internet auch mal
- Dies führt zum **Ruf nach mehr Regulierung** und "Ethik"
  - Themen wie Transparenz, Diskriminierung, Erklärbarkeit, Human-in-the-Loop
  - "EU AI Act" als typische Reaktion darauf
  - Bundesrat will in Kürze Schweizer Regulierungsbedarf erklären
- Aber: Wir werden uns **daran gewöhnen**
  - Wir werden unsere bestehenden Gesetze und Prozesse darauf anwenden; es wird so normal werden wie das Internet

25. Juli 1994 (time.com, Titelseite: James Porto)

VISCHER

Danke für Ihre Aufmerksamkeit!

Fragen: david.rosenthal@vischer.com

Mehr Unterlagen:
www.vischer.com/ki
www.rosenthal.ch

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