



# KI zur Analyse der Kundenkommunikation -

wie kleine, maßgeschneiderte  
Sprachmodelle den Kundenservice  
verbessern und die Vertriebsperformance  
steigern können

**Peter Kesch**  
(SentiSquare)

# SentiSquare = Erstellung kleiner spezialisierter KI-Modelle zur Analyse der Kommunikation



## ✓ **Steigerung des Umsatzes**

Steigern Sie Ihre Konversionsrate durch eine bessere Kommunikation.

## ✓ **Höhere Kundenzufriedenheit**

Verwandeln Sie die gewonnenen Erkenntnisse in hochwertige Dienstleistungen und zufriedenere Kunden.

## ✓ **Volle Kontrolle**

Eine detaillierte Übersicht über jede Interaktion, in der Ihnen nichts entgeht.

## ✓ **100% Compliance**

Stellen Sie sicher, dass interne und gesetzliche Vorschriften befolgt werden.





## **Das Problem: Unmengen unstrukturierter Textkommunikation**

B2C-Unternehmen haben oft keinen objektiven Überblick und kein Verständnis über die Kundenkommunikation über verschiedene Kanäle hinweg.  
(E-Mails, Anrufe, Chats, SMS, soziale Medien usw.)

# Warum sollte man die Kommunikation analysieren?

- Erlangen einer Makroperspektive
- Quantifizierung
- Trendbeobachtung
- Suche nach Zusammenhängen
- Kontrolle
- Optimierung

## Einsatzbereiche:

- Laufender Betrieb
- CX
- Kommunikationsqualität
- Vertrieb
- Compliance

# Sprachmodelle

## Aktuelle Trends



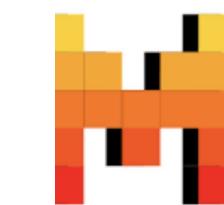
## Sprachmodelle: LM



Gemini



Claude



MISTRAL  
AI\_



SentiSquare\_



## LLMs sind überall

- Der Fortschritt in der Entwicklung der Neuronalen Netzwerke und GPUs haben die Entwicklung von LLMs ermöglicht
- LLMs sind inzwischen der Mainstream
- LLMs sind großartig



## Kleine spezialisierte KI-Modelle (SLM - Small Language Models)

### “Tech Trends For 2025: The Rise Of Specialized AI”

- “Spezialisierte KI-Modelle, die das Paradigma „Größer ist besser“ auf den Kopf stellen.”
- Quelle: Forbes, 21. November 2024
- <https://www.forbes.com/councils/forbestechcouncil/2024/11/21/tech-trends-for-2025-the-rise-of-specialized-ai-and-trust-based-innovation/>

Forbes

### “The 5 AI Trends In 2025”

- “Im Bereich der KI geht es nicht mehr nur um große Sprachmodelle (LLM) – es geht um intelligenter, kostengünstiger, spezialisierte und lokale KI-Lösungen.”
- Quelle: Forbes, 28. Februar 2025
- <https://www.forbes.com/sites/solrashidi/2025/02/28/the-5-ai-trends-in-2025-agents-open-source-and-multi-model/>

Forbes



**“Bis 2027 werden Unternehmen dreimal mehr kleine, aufgabenspezifische KI-Modelle als universelle LLM verwenden”**

<https://www.gartner.com/en/newsroom/press-releases/2025-04-09-gartner-predicts-by-2027-organizations-will-use-small-task-specific-ai-models-three-times-more-than-general-purpose-large-language-models>

**Gartner®**

Gartner, 9. April 2025

**„Kleine, aufgabenbezogene Modelle bieten schnellere Reaktionszeiten und verbrauchen weniger Rechenleistung, was die Betriebs- und Wartungskosten senkt.“**

*– Sumit Agarwal, Vice President Analyst, Gartner*

**“Die Zukunft liegt nicht in größeren Modellen, sondern in besser abgestimmten.”**

*– Andrew Ng, weltweit anerkannter Experte für KI*



# **SLM vs LLM**

## Geschwindigkeit



# 1.000.000 Kundentelefonate: Geschwindigkeit

Analyse von 5 Aspekten pro Kommunikation = 5.000.000 Analysen



Text classification speed		LLM + prompting					SentiSquare No-Code NLP Plattform
		ChatGPT 3.5 / 4	Mistralai 7B	Orca 7B	OpenChat 7B	Llama 7B	
Model size	LLM	LLM	LLM	LLM	LLM	SLM	
HW	Super strong	16GB VRAM	16GB VRAM	16GB VRAM	16GB VRAM	16GB VRAM	1 CPU 1GB RAM
time for 1 token	0.02s	0.11s	0.11s	0.11s	0.11s	/	
avg. num. tokens	80	88	28	10	106	/	
time for 1 document	1.3s	9.8s	3.1s	1.1s	11.8s	0.02s	
documents per hour	2 813	369	1 165	3 328	306	180 000	
documents per day	67 500	8 845	27 972	79 866	7 342	4 320 000	
documents per day (common project = 5 models)	13 500	1 769	5 594	15 973	1 468	864 000	
1 million documents (common project)	74 days	565 days	179 days	63 days	681 days	1.2 day	

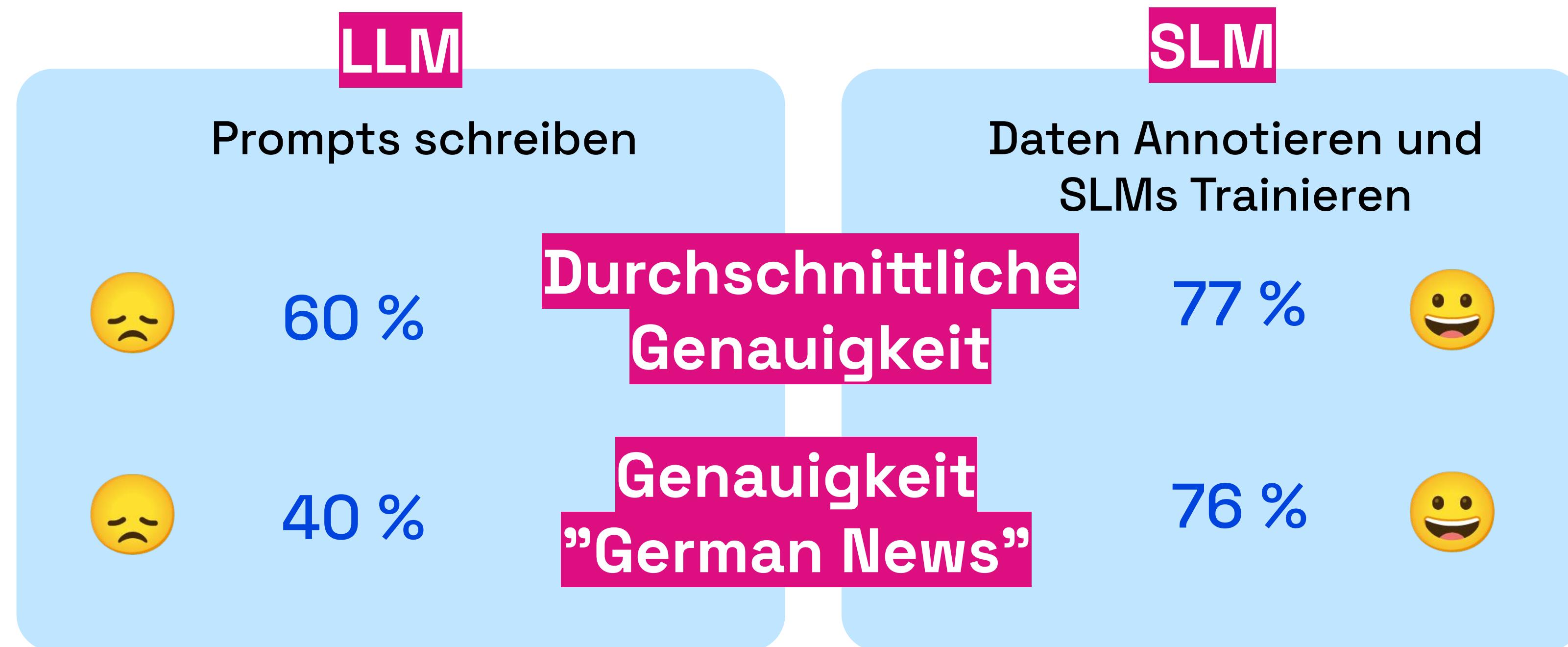
# **SLM vs LLM**

## Genauigkeit



# 1.000.000 Kundentelefonate: Genauigkeit

Effektive Prompts für LLMs schreiben oder SLMs trainieren?



Text classification accuracies	LLM + prompting							SentiSquare No-Code NLP Plattform
	ChatGPT 4o	ChatGPT 3.5	Gemini 1.0 pro	MistralAI 7B	Orca 7B	Open Chat 7B	Llama 7B	
Model size	LLM	LLM	LLM	LLM	LLM	LLM	LLM	LLM
20 News Groups (EN)	76%	69%	60%	50%	48%	53%	17%	84%
Affects in tweets (EN)	60%	57%	58%	51%	56%	60%	44%	66%
ČTK (CZ)	61%	47%	36%	31%	30%	29%	13%	79%
German news (DE)	40%	43%	36%	33%	38%	37%	34%	76%
Stack Overflow (EN)	52%	46%	41%	32%	42%	42%	34%	71%
Questions (EN)	71%	57%	46%	35%	40%	50%	14%	86%

# **SLM & LLM**

## das Dream-Team



# Maßgeschneiderte kleine Sprachmodelle

**SLM**

**Ideal für Szenarien mit sensiblen Daten, hohem Interaktionsvolumen und Anforderungen an geringe Latenzzeiten.**



Höchste Präzision bei konkreten Aufgaben



Schnelle Reaktion



Keine spezielle Hardware erforderlich



Wiederkehrende Aufgaben

**LLM**

**Dank ihres universellen Wissens ideal für Ad-hoc-Analysen und sich schnell verändernde Umgebungen.**

Gute Genauigkeit bei allgemeinen Aufgaben



Langsame Reaktion



Erfordern leistungsstarke GPUs



Detailliertes Verständnis



**Genauigkeit**

**Latenz**

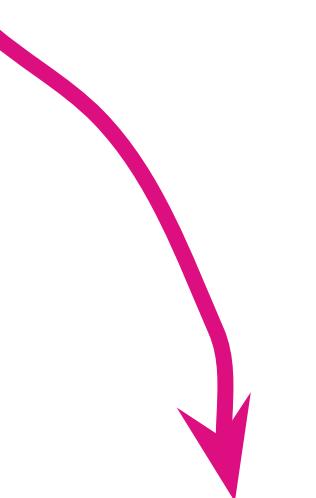
**Hardware**

**Fokus**



## SentiSquare, unsere Vision:

- KI-Modelle sind/werden zur Standardsoftware
- Die Zukunft liegt in:
  - Effiziente Erstellung (Training) und Betrieb von KI-Modellen
  - No-Code-/Low-Code-Plattformen für spezialisierte Modelle
- Hybridansatz:
  - SLM + LLM
  - Nichtgenerative KI + Generative KI

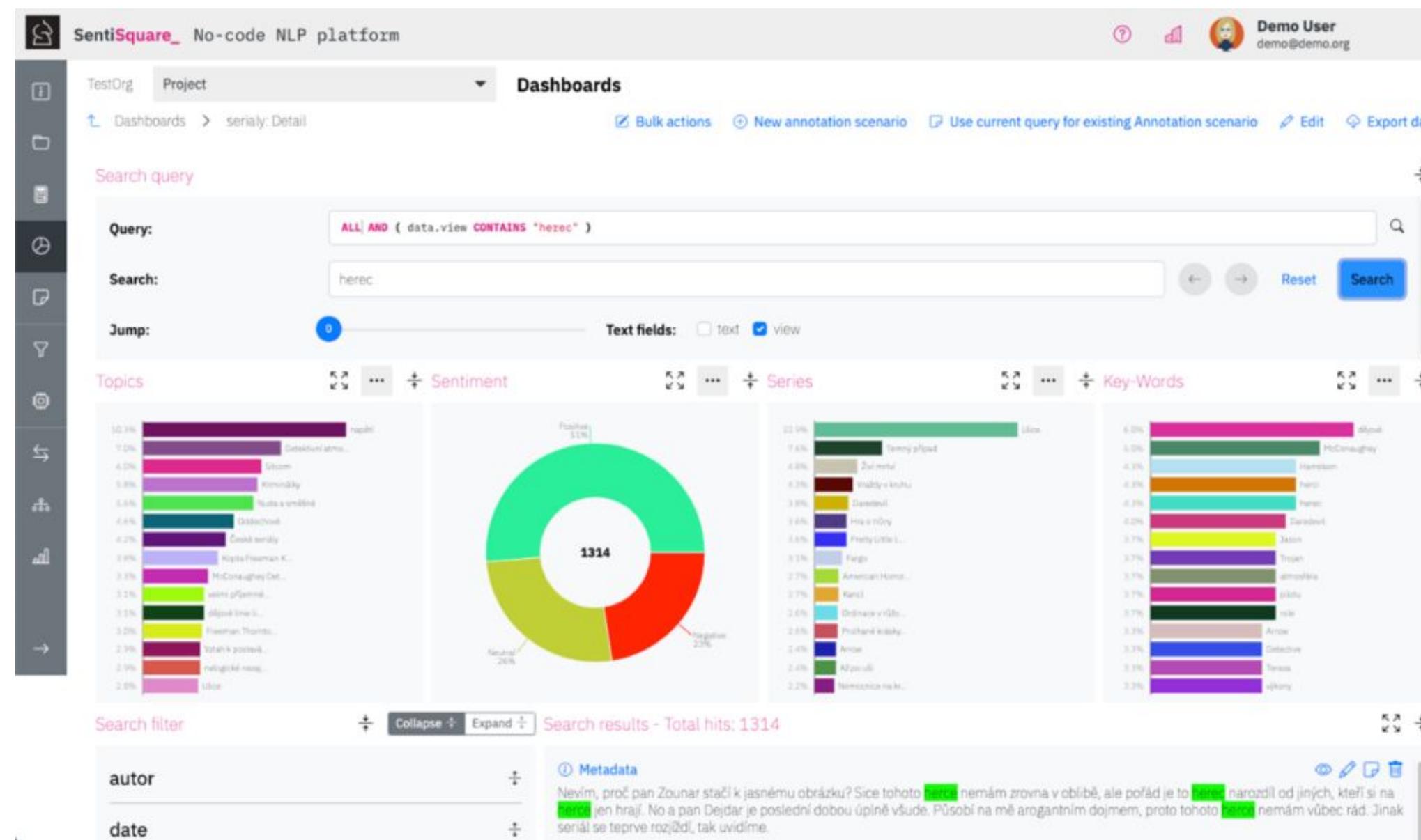


SentiSquare No-Code NLP Plattform



# SentiSquare\_

# Eigene Modelle leicht gemacht: Die SentiSquare No-Code NLP-Plattform



## → Omnichannel-Analytik

# → Gesamter SLM-Lebenszyklus

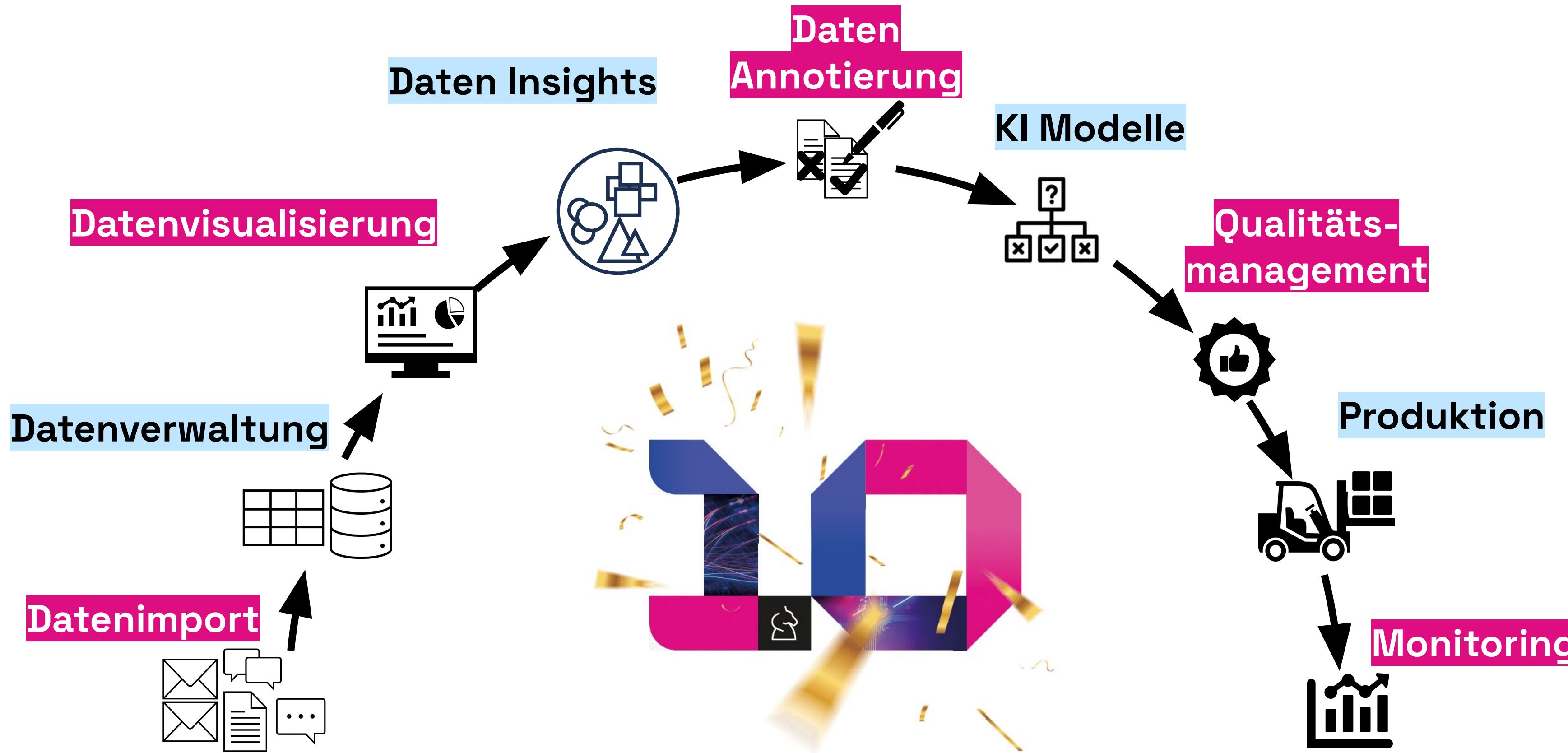
## → Sprachunabhängig

→ Autark (auch vor Ort möglich)

# → Schnell

## → Integriertes LLM

# Der gesamte Lebenszyklus der KI





**SentiSquare No-Code NLP-Plattform:**

**Erstellung kleiner spezialisierter KI-Modelle zur**

**Auswertung der Kundenkommunikation.**

# Kundendienstmitarbeiter

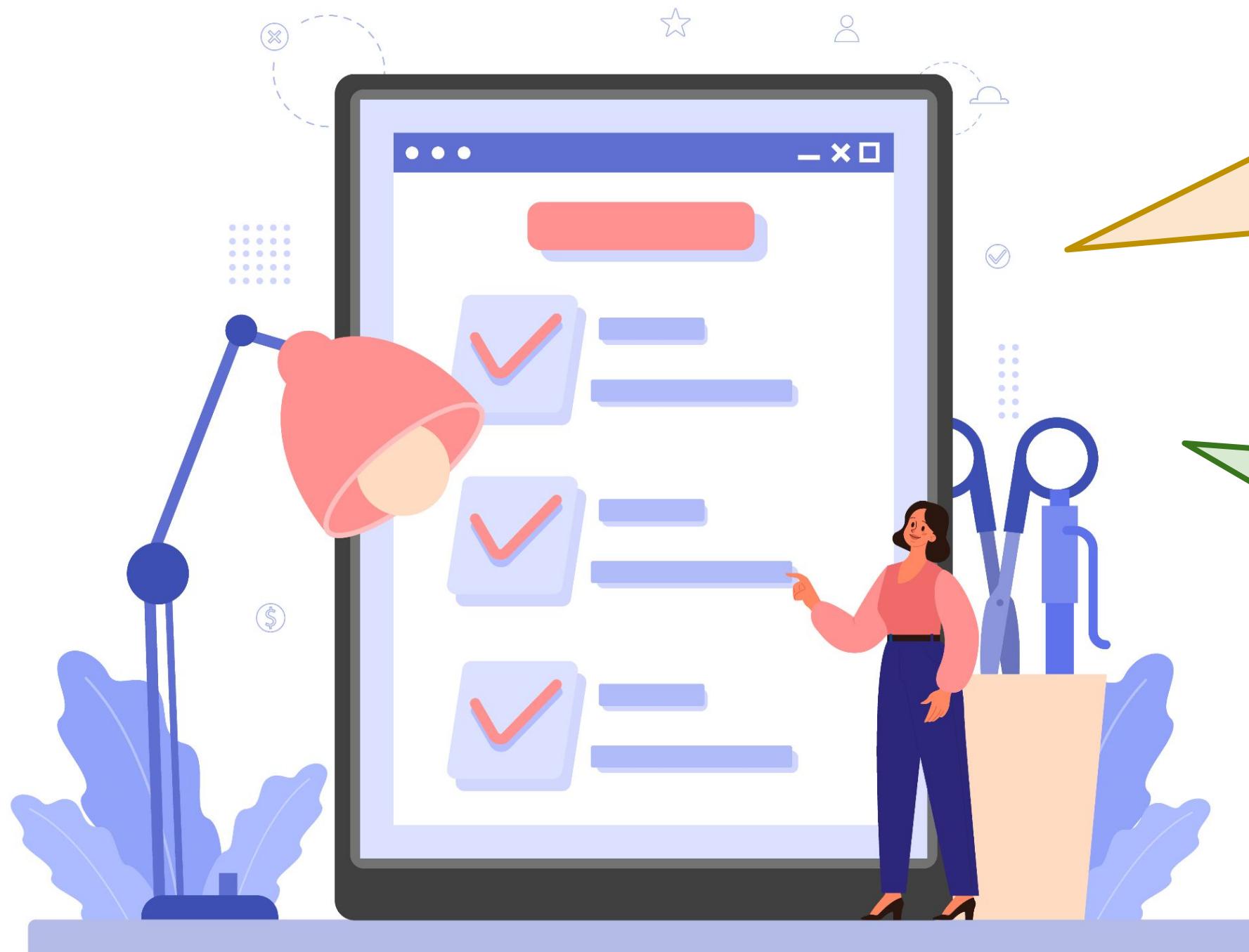


## Analyse der Kommunikation

- Was ist der Kontaktgrund?
- Ist es dringend?
- Wurde der Kunde verstanden?
- Wurde eine Problemlösung angeboten?
- Hat der Kunde die Lösung akzeptiert?
- Ist der Kunde ein Abwanderungskandidat?
- First Contact Resolution (FCR)?



# Qualitätskontrolle



## Qualitätskontrolle der Kommunikation

- Wurde das Anrufskript befolgt?
- Wurde das Kundenbedürfnis verstanden?
- Sind die Soft Skills der Sachbearbeiter gut?
- War die Stimmung der Kunden negativ?

## Verkaufsgesprächsanalyse

- Wurde versucht einen Abschluss zu erzielen
- Wird zu schnell abgebrochen?
- Warum wurde das Angebot abgelehnt?
- Wurde Cross-Selling versucht?
- Wurde das Produkt gut erklärt?



# Compliance-Abteilung



## Einhaltung der Vorschriften

- Wurde die Gesprächsaufzeichnung angekündigt?
- Wurden die internen Regeln eingehalten?
- Erfüllt es die regulatorischen Anforderungen (z.B. Vorschriften der BaFin)?

## Investitionsgespräche

- Ist dieser Anruf ein Investment Anruf?
- Wurden Empfehlungen ausgesprochen?
- Wurden Gewinne versprochen?
- Wurden Gebühren verschwiegen?
- Wurden private Kontakte weiterleiten?

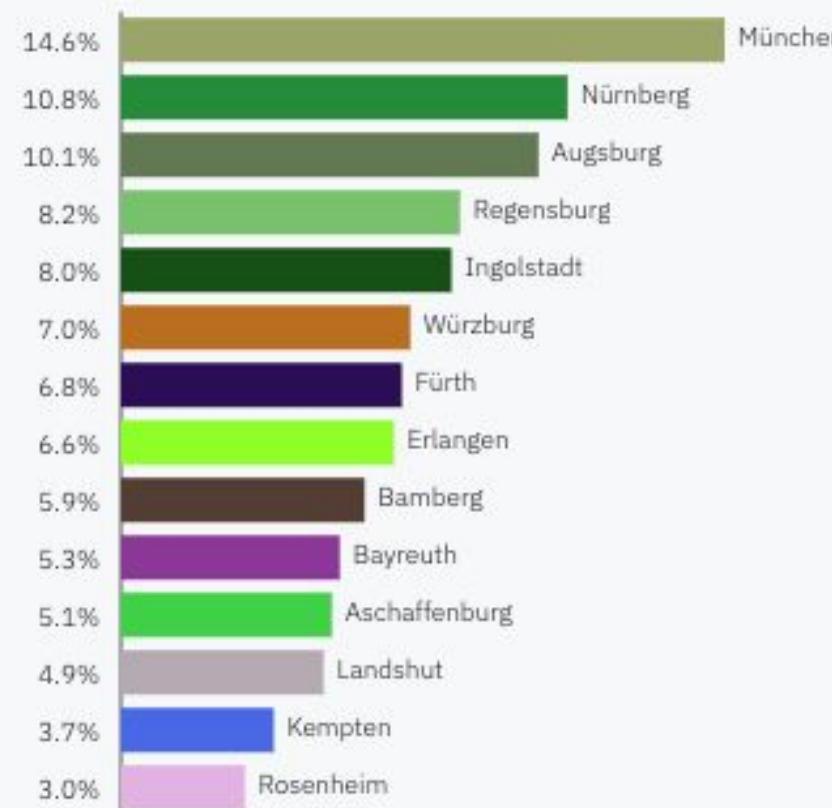


# CX und Marketingabteilung

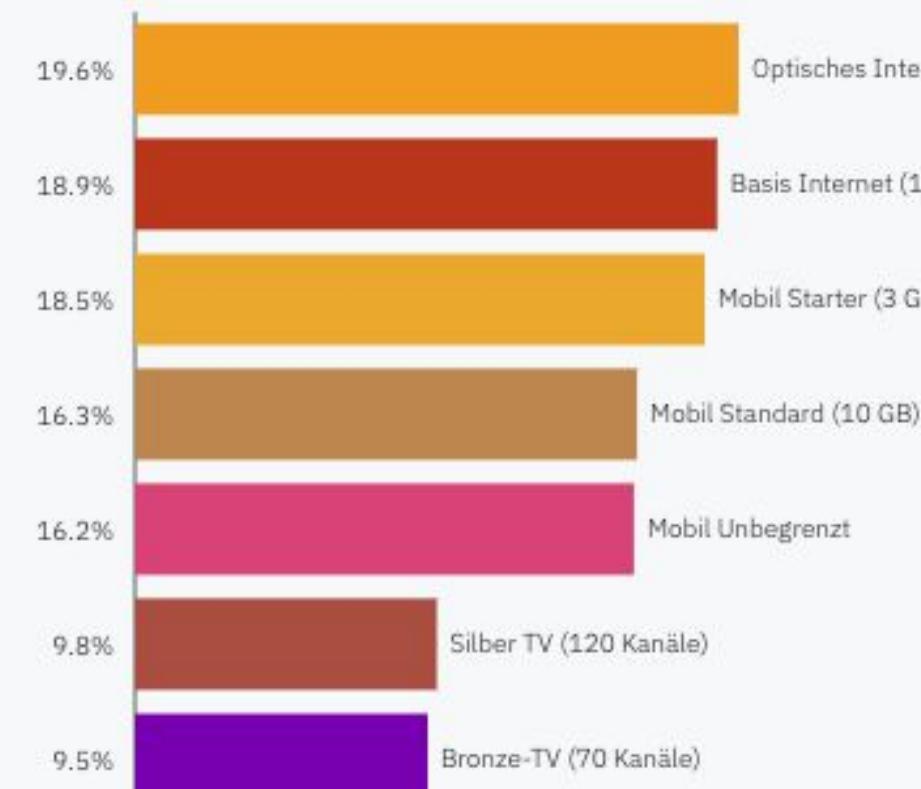


## Region

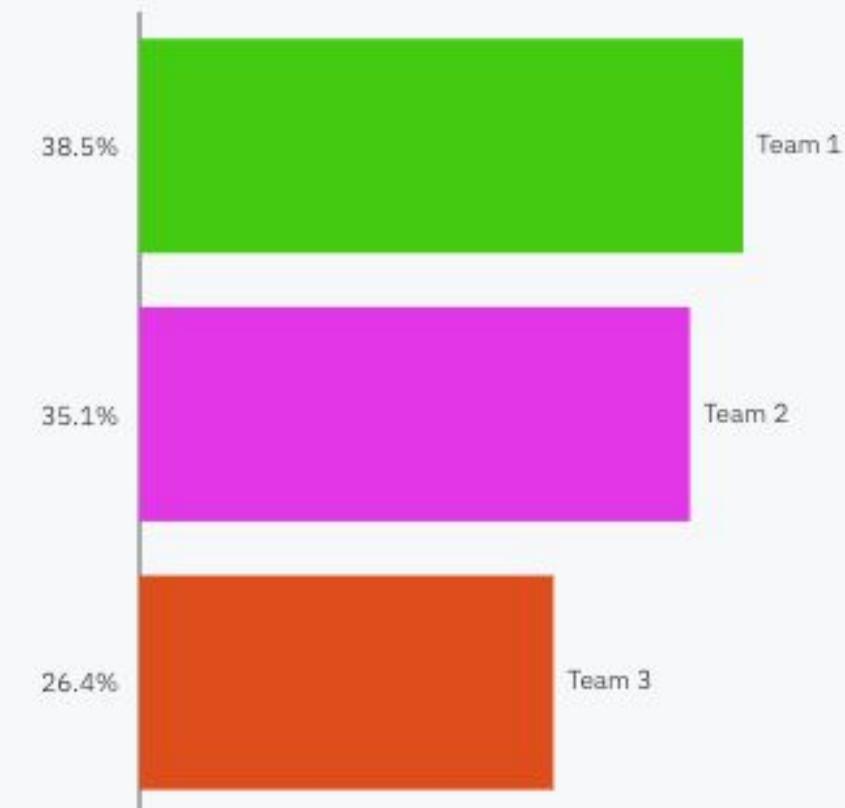
## Produkte



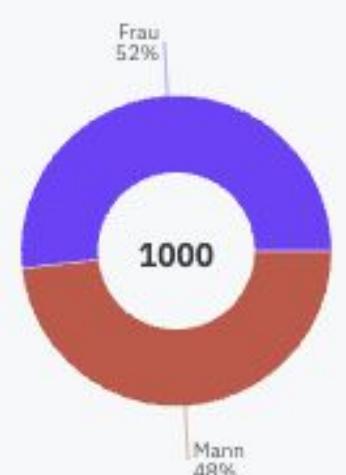
## Kampagne



## Verkaufs - Teams

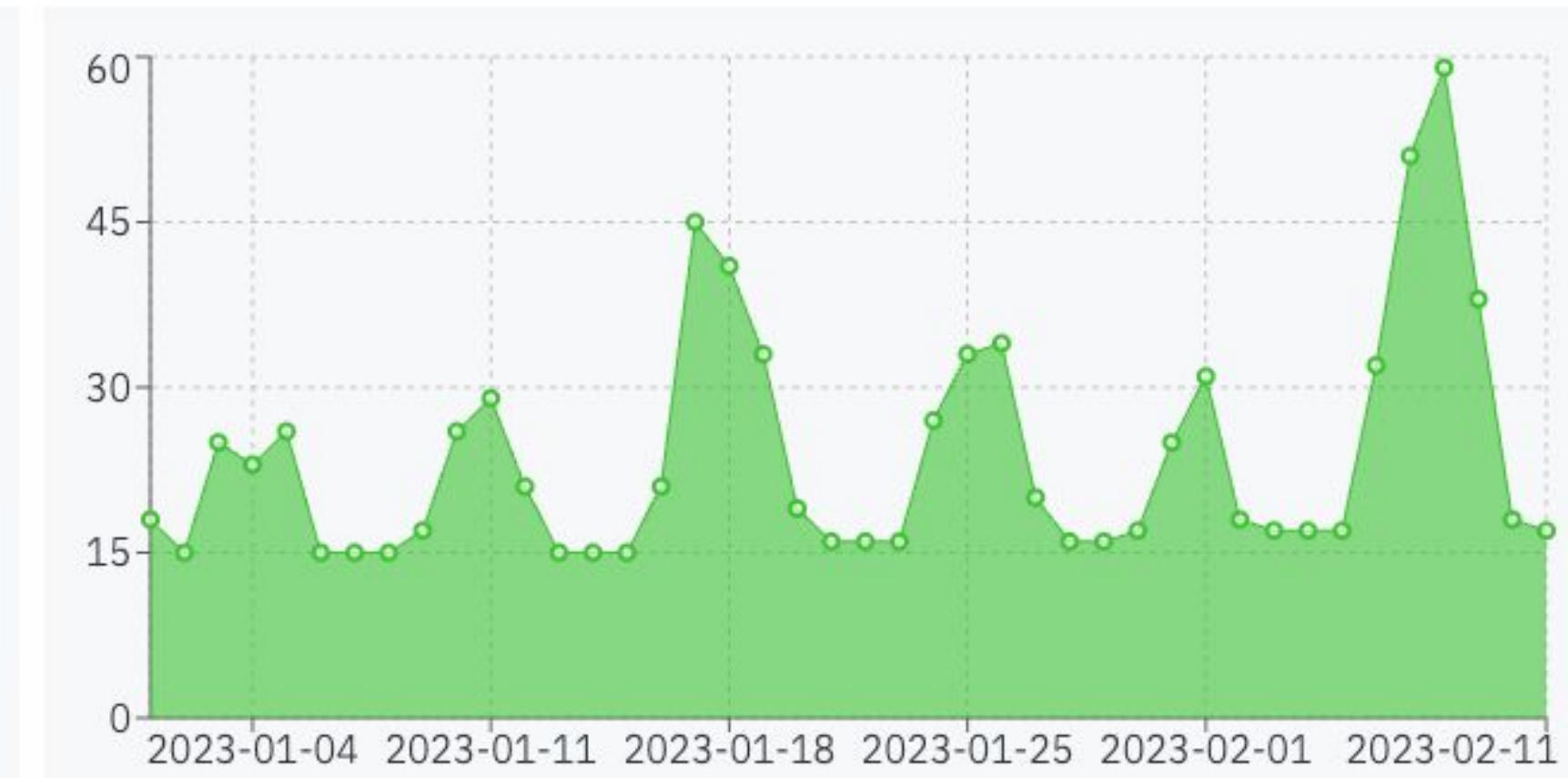
Altersgruppe... ... Geschlecht

## Agent x Geschlecht Kunde



	Frau	Mann
Matthe...	48.84 %	51.16 %
Jessica...	54.88 %	45.12 %
Benjami...	40.98 %	59.02 %
Sophia Davis	53.45 %	46.55 %
Isabella...	55.17 %	44.83 %
Alexand...	63.16 %	36.84 %
Olivia A...	62.96 %	37.04 %
William...	56.60 %	43.40 %
Amelia...	54.17 %	45.83 %
Ava Martin	48.94 %	51.06 %
James T...	52.17 %	47.83 %
Harper...	52.17 %	47.83 %
Ethan Harris	47.83 %	52.17 %
Charlotte Lee	39.13 %	60.87 %
Daniel Brown	42.22 %	57.78 %

## Zeitachse Anrufe



## Typische Ergebnisse unserer Kunden



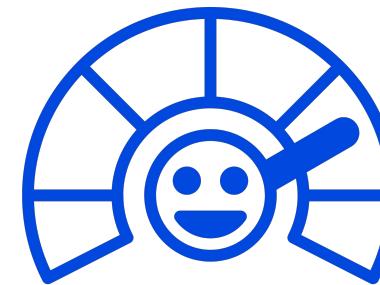
**Millionen**

Jährlich  
analysierte  
Interaktionen



**100 %**

Kommunikation  
unter Kontrolle



**+10**

NPS Punkte



**+5-10 %**

Vertriebs-  
performance

## Wie wir die Leistung von O2

### Telesales verbessert haben

#### Herausforderung:

- Mehr als 5 Millionen Anrufe/Chats pro Jahr

#### SentiSquare No-Code NLP Plattform:

- Themen und Gründe für die Kontaktaufnahme
- Kundenstimmung und -zufriedenheit
- Einhaltung von Vorschriften und Versprechen
- Qualität der Lösungen (Betreuung + Verkauf)

#### Resultat:

- 100 % Kommunikation unter Kontrolle
- NPS +9 Punkte
- Gesteigerte Vertriebsperformance
- Mehr als 30 SLMs im Einsatz

“SentiSquare No-Code NLP is the only platform that allows us to create our own AI models for text analysis without needing NLP knowledge or developer skills, while providing extensive flexibility in model creation and management of various exceptions.”

– Jana Šmatláková, Customer Interaction Analytics Product Owner, O2 CZ



O<sub>2</sub>  vodafone  

**e-on**  

 **Kooperativa**  
VIENNA INSURANCE GROUP









# SentiSquare\_

Besuchen Sie unseren Stand

Peter Kesch

kesch@sentisquare.com