

AI for Innovation

die richtige AI-Datengrundlage schaffen

24. Juni 2026

Fraunhofer Austria als Teil der Fraunhofer-Gesellschaft

Die führende Organisation für anwendungsorientierte Forschung

Fraunhofer-Gesellschaft

DE

Hauptsitz

3,1 Mrd.

EUR Forschungsvolumen

75

Institute & Einrichtungen

32.000

Beschäftigte



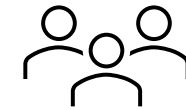
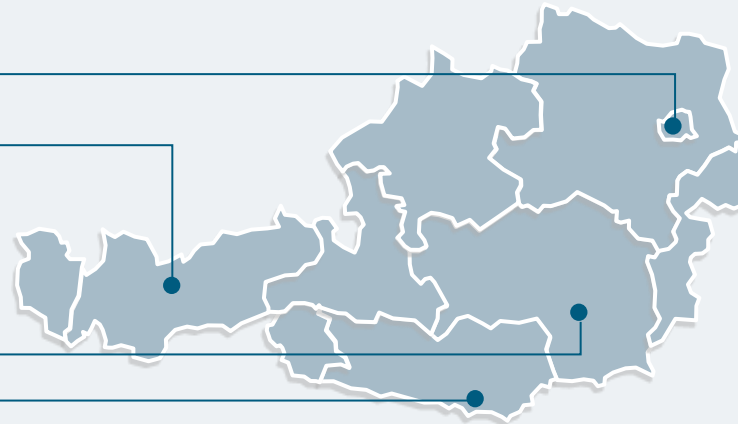
Fraunhofer Austria

Center für Nachhaltige Produktion und Logistik

- Arbeitsgestaltung und Digitalisierung
- Fabrikplanung und Produktionsmanagement
- Logistik und Supply Chain Management

Center für Data Driven Design

- Digitalisierung und Künstliche Intelligenz
- Visual Computing



135

Mitarbeiterinnen
und Mitarbeiter

4

Standorte

5

Geschäftsbereiche



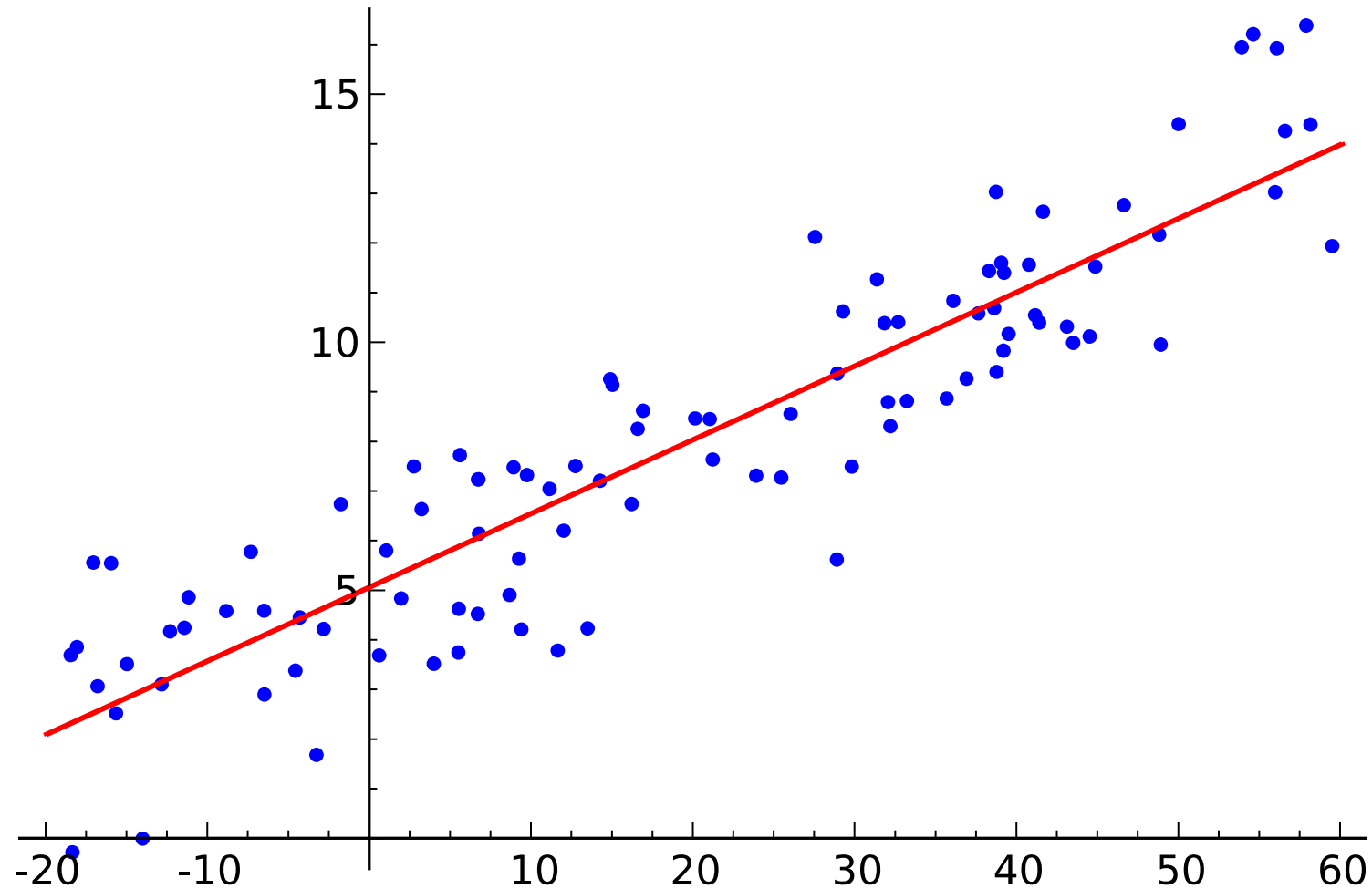
10 Mio.

EUR
Forschungsvolumen

-
- **Welche Daten brauche ich für erfolgreiche KI-Anwendungen in Forschung und Entwicklung?**
 - Wie gehe ich mit Datenqualität, Datenintegration und Datenschutz um (DSGVO, IP-Schutz)?
 - Welche Infrastruktur passt zu meinem Projekt (Cloud vs. On-Primse)?

Datenumfang

Welche Daten brauche ich für erfolgreiche KI-Anwendungen in Forschung und Entwicklung?



Datenumfang

Welche Daten brauche ich für erfolgreiche KI-Anwendungen in Forschung und Entwicklung?

- **1 byte**
8 bit – 256 Möglichkeiten
- **1 kilobyte (10^3 bytes)**
1000 Zeichen Text
- **1 megabyte (10^6 bytes)**
eine Diskette
- **1 gigabyte (10^9 bytes)**
ein USB-Stick
- **1 terabyte (10^{12} bytes)**
eine Festplatte
- **1 petabyte (10^{15} bytes)**
alle Texte aller US-Universitätsbibliotheken
- **1 exabyte (10^{18} bytes)**
jedes jemals gesprochene Wort aller Menschen
- **1 zetabyte (10^{21} bytes)**
die in 2020 generierte Datenmenge umfasst 64,2 Zetabytes

Datenumfang

Welche Daten brauche ich für erfolgreiche KI-Anwendungen in Forschung und Entwicklung?

Wetter Singapur ☆

Singapur, 1.29°N 103.85°O, 23m ü.NHN

 27 °C

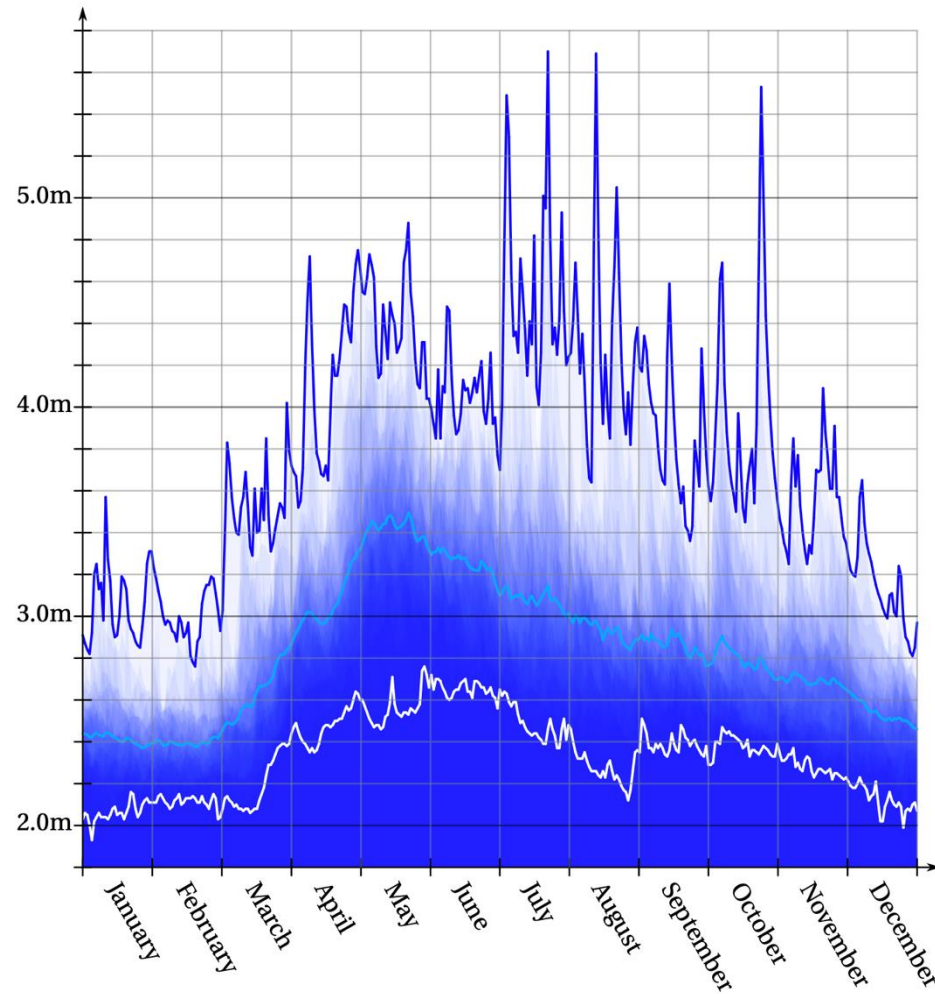
3 km/h | 16:45

MO Heute	DI Morgen	MI 24.6.	DO 25.6.	FR 26.6.	SA 27.6.	SO 28.6.
						
						
27 °C	30 °C	29 °C	30 °C	31 °C	31 °C	30 °C
24 °C	25 °C	26 °C	26 °C	27 °C	26 °C	26 °C
↗ 5 km/h	↗ 4 km/h	↗ 5 km/h	↑ 6 km/h	↑ 6 km/h	↑ 5 km/h	↑ 6 km/h
💧 10-20 mm	💧 0-10 mm	💧 10-20 mm	💧 0-10 mm	💧 0-5 mm	💧 0-5 mm	💧 0-2 mm
☀ 2 h	☀ 5 h	☀ 5 h	☀ 6 h	☀ 6 h	☀ 7 h	☀ 7 h
🌀🌀🌀🌀	🌀🌀🌀🌀	🌀🌀🌀🌀	🌀🌀🌀🌀	🌀🌀🌀🌀	🌀🌀🌀🌀	🌀🌀🌀🌀

Quelle: https://www.meteoblue.com/de/wetter/woche/singapur_singapur_1880252 (aufgerufen am 22.06.2026, 11:00)

Datenumfang

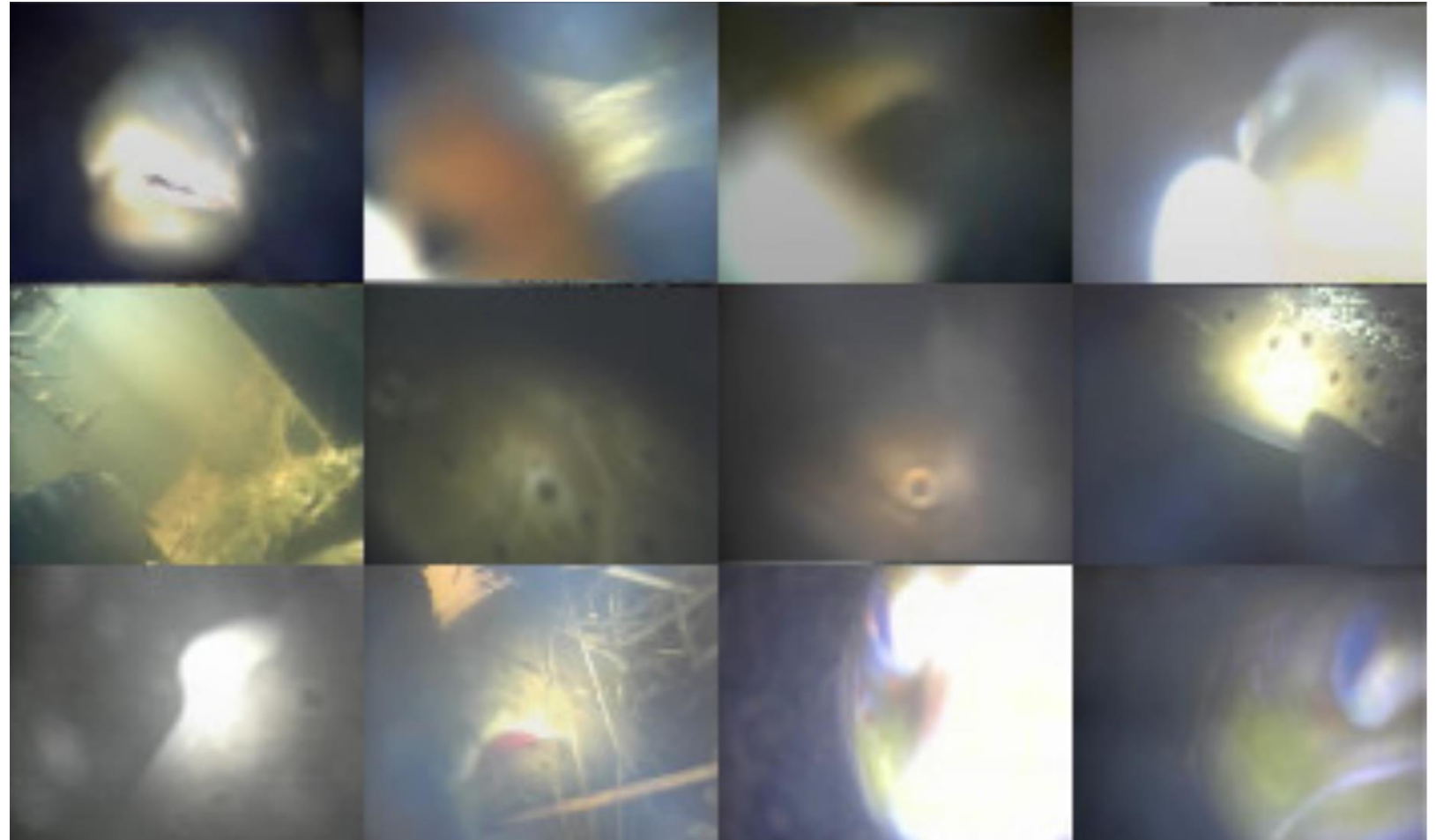
Welche Daten brauche ich für erfolgreiche KI-Anwendungen in Forschung und Entwicklung?



year	minimum	first quartile	median	third quartile	maximum	average	standard deviation
1976	2,26	2,60	2,77	3,04	3,88	2,83	±0,34
1977	2,12	2,48	2,88	3,24	4,14	2,90	±0,46
1978	2,15	2,41	2,69	3,18	4,47	2,82	±0,49
1979	2,11	2,63	2,95	3,40	4,75	3,02	±0,55
1980	2,35	2,56	2,82	3,28	3,97	2,93	±0,42
1981	2,26	2,66	2,93	3,16	4,26	2,94	±0,37
1982	2,39	2,73	2,95	3,24	4,69	3,01	±0,44
1983	2,18	2,61	2,85	3,16	3,88	2,88	±0,39
...	...	...	...	...	...	...	...
2010	2,17	2,41	2,54	2,78	3,72	2,61	±0,31
2011	2,18	2,35	2,48	2,62	3,62	2,51	±0,23
2012	2,28	2,63	2,80	3,02	5,70	2,87	±0,45
2013	2,39	2,55	2,67	3,00	4,68	2,85	±0,45
2014	2,48	2,70	2,86	3,08	4,59	2,94	±0,34
2015	2,35	2,53	2,66	2,93	3,74	2,75	±0,28
2016	2,29	2,58	2,81	3,03	4,41	2,84	±0,33
'76-'16	1,93	2,49	2,75	3,11	5,70	2,84	±0,46

Datenumfang

Welche Daten brauche ich für erfolgreiche KI-Anwendungen in Forschung und Entwicklung?

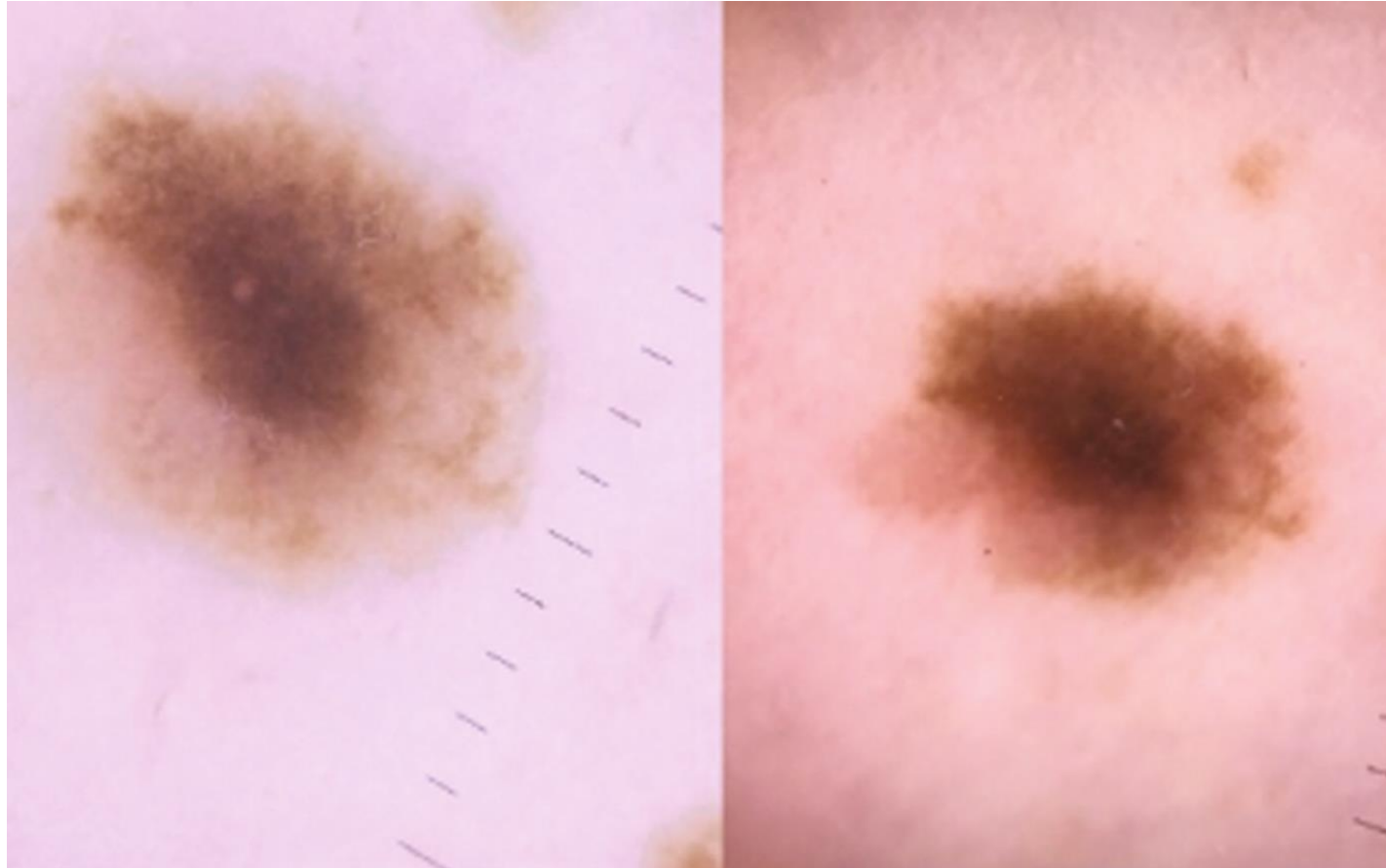


Quelle: H. Tödtmann, M. Vahl, U. Freiherr von Lukas, T. Ullrich (2020), **“Time-Unfolding Object Existence Detection in Low-Quality Underwater Videos using Convolutional Neural Networks”**, VISIGRAPP, 15:370-377.

-
- Welche Daten brauche ich für erfolgreiche KI-Anwendungen in Forschung und Entwicklung?
 - **Wie gehe ich mit Datenqualität, Datenintegration und Datenschutz um (DSGVO, IP-Schutz)?**
 - Welche Infrastruktur passt zu meinem Projekt (Cloud vs. On-Primse)?

Datenqualität

Wie gehe ich mit Datenqualität, Datenintegration und Datenschutz um (DSGVO, IP-Schutz)?



Quelle: J. K. Winkler, C. Fink, et al. (2019), “Association Between Surgical Skin Markings in Dermoscopic Images and Diagnostic Performance of a Deep Learning Convolutional Neural Network for Melanoma Recognition”, JAMA Dermatology, 155:1135-1141.

Wie gehe ich mit Datenqualität, Datenintegration und Datenschutz um (DSGVO, IP-Schutz)?



Datenqualität

Data Maturity Model

This work has been funded by the Austrian Research Promotion Agency (FFG) within the research project "PREDictions for Science, Engineering N' Technology (PRESENT)"; GA No. F09989554.



Category: Data Collection

#1: Is all the data collected?
☐ yes ☐ no ☐ not relevant

For an oral surgeon's internal statistics, the number of patients over the age of 30 who still have wisdom teeth needs to be determined. However, according to the current database, data on wisdom teeth is not available for a quarter of the target group. In order to make a valid statement, this data needs to be collected.

#2: Did you have a clear understanding of what data needed to be collected?
☐ yes ☐ no ☐ not relevant

Depending on the aim of the data analysis, it is necessary to collect the data that can be used to answer the question underlying the data analysis.

#3: Is the data coming from the right target group?
☐ yes ☐ no ☐ not relevant

When students are asked about their financial situation during their studies, no data may be collected from graduates as this would distort the results of the analysis.

#4: Was the objective of the subsequent data analysis clear when the data was collected?
☐ yes ☐ no ☐ not relevant

If the objective of the data analysis is not clearly stated, it will be difficult to collect the relevant data. For example, the objective to analyse the housing market would be an unclear objective because it does not describe what factors should be used to analyse the housing market.

#5: Did you primarily collect the data?
☐ yes ☐ no ☐ not relevant

Primary data is data collected for the purpose of data analysis. For example, in a company's market research, primary data is collected when the company's customers are asked directly about the topic. However, if the data comes from another source or from previous data collections, it is referred to as secondary data and the data was not primarily collected.

#6: Is the data publicly accessible?
☐ yes ☐ no ☐ not relevant

To analyse the performance of a professional football team, the number of wins is published by the football league. This data is publicly available. However, the mood within a football team is not publicly available. This data can only be obtained through surveys.

#7: Is the source of the data comprehensible and verifiable?
☐ yes ☐ no ☐ not relevant

The analysis of posts on a social media platform should examine the political influence of a party. When collecting data, only posts that were actually posted on the social media platform and still exist at the time of data collection may be used.

#8: Have the paradata been recorded and documented?
☐ yes ☐ no ☐ not relevant

Paradata describe the process of data collection. Questions about the acquisition process, such as "Who collected the data?", "What instrument and method were used?", "What were the environmental conditions?", or similar, need to be answered by the paradata. Paradata should be available together with the actual data.

#9: Have the metadata been captured and documented?
☐ yes ☐ no ☐ not relevant

Metadata is structured data about a resource. In the case of a book, metadata is high-level information such as the publisher, year of publication or author. Metadata is used to describe the information resource with additional data so that it can be processed automatically. Metadata should be available with the data.

#10: Has the data collection process been designed with a view to processing the data?
☐ yes ☐ no ☐ not relevant

When collecting personal information for a course, participants are asked for their gender. For further processing it is preferable to use checkboxes for gender rather than providing space for text answers that need further processing.

#11: In the case of time-dependent data, has the synchronisation of the data been considered and are the mechanisms used for synchronisation documented?
☐ yes ☐ no ☐ not relevant

Ensure that all time-related data is synchronised. Poor synchronisation can be detected when the picture and sound of a video are out of sync. Consider synchronisation for high-resolution measurement systems with thousands of measurements per second. Document the synchronisation mechanisms in the paradata and metadata.

#12: Did the way the data was collected or the type of questions asked influence the data collected?
☐ no ☐ yes ☐ not relevant

An interview focuses on specific topics in order to obtain relevant data. Could a questionnaire get the same data? Even in a questionnaire, non-objective wording can influence responses. For example, "Do you like the new canteen?" is more positive than "Are you satisfied with the food in the canteen?"

#13: Was the correct scaling (nominal, ordinal, metric) used for the data during data collection?
☐ yes ☐ no ☐ not relevant

A company asks its employees whether they use public transport to get to work, with the possible answers being daily, several times a week, once every two weeks, monthly and never. As these are ordinal data, they must also be asked in the correct order during data collection.

#14: Would the same data be recorded if data were collected several times at the same time?
☐ yes ☐ no ☐ not relevant

When measuring electrical quantities, the measurement results must be identical for several measurements if the same measurement setup, the same measurement environment and the same measurement method are used. In contrast, in a street survey of voting behaviour for a political election, it is expected that different data will be obtained depending on the person, location and time of day surveyed.

#15: Is the data collection process reproducible?
☐ yes ☐ no ☐ not relevant

A data collection is reproducible if all the processes used to obtain the data are sufficiently documented in the paradata so that it is possible to repeat the data collection. For example, when measuring physical quantities, the instrument used, the measurement setup and the measurement method must be described.

#16: Is the data collection process comparable?
☐ yes ☐ no ☐ not relevant

A data collection is comparable if the data are calibrated. For example, when measuring physical quantities, additional information about the measuring instrument, such as the last calibration, the accuracy of the instrument, and the tolerances associated with the instrument's measurements, should be included in the metadata. Without this additional information, it is not possible to make meaningful comparisons.

#17: Has the representativeness of the data been checked?
☐ yes ☐ no ☐ not relevant

Representativeness is key to making general statements from a sample. If a data collection is not representative, the results are limited. To illustrate, a statement about all students' opinions on campus facilities requires verification of representativeness and survey adherence.

#18: Is the representativeness of the data documented?
☐ yes ☐ no ☐ not relevant

A survey is only representative if its representativeness is verified and documented in the metadata.

#19: Are you authorised to use the data collected?
☐ yes ☐ no ☐ not relevant

Data sets that contain personal data, or that can be used to infer personal data, are subject to the General Data Protection Regulation (GDPR). The processing of the data and the publication of the results of any subsequent analysis are only permitted if the GDPR is complied with. Therefore, the appropriate GDPR consents must be obtained and documented alongside the data.

Category: Data Labelling

#20: Is the data available in digital form?
☐ yes ☐ no ☐ not relevant

Data is stored in a file on a computer or server.

#21: Is there still analogue data that needs to be digitised?
☐ no ☐ yes ☐ not relevant

If questionnaires are completed by hand, all questionnaires need to be digitised. Similarly, if an interview is audio recorded, it needs to be transcribed for processing.

#22: Has attention been paid to spelling in the data entry process?
☐ yes ☐ no ☐ not relevant

Spelling errors lead to different entries in the record for the same information. For example, the entries "female" and "femiale" both contain the same information about gender.

#23: Are there any abbreviations in the data that are not explained and need additional explanation?
☐ no ☐ yes ☐ not relevant

In the survey, respondents were asked to indicate their academic degree or title. For the sake of clarity, only the abbreviations of the titles are included. As academic titles vary, a detailed explanation of the abbreviations used must be provided in the metadata.

#24: Has the data been affected by the data labelling?
☐ no ☐ yes ☐ not relevant

When measuring a distance, the length is measured in kilometres due to the capabilities of the available measuring instrument used. During the data labelling process, measurements are converted to metres to improve the usability of the data. Incorrect conversions would affect the data.

#25: Is the format/syntax the same for all data within a category?
☐ yes ☐ no ☐ not relevant

The term format or syntax refers to the structure of a data entry within a particular category. For example, all dates of birth must be in the format day/month/year to facilitate consistent age analysis. Data entries in the format month/day/year would distort the results.

#26: Are all formats or syntaxes used described or documented?
☐ yes ☐ no ☐ not relevant

If a record in a category requires a special format, the format should be described in the metadata. Examples are specifications for time data (day/month/year) and gender.

#27: If one or more classifications have been used, has care been taken to avoid potential falsifications?
☐ yes ☐ no ☐ not relevant

Classification groups data into predefined classes. A survey may ask about housing and the question may be open-ended. The responses are then classified into categories: flat, owner-occupied house and living with parents. A problem is the distinction between rented and owner-occupied accommodation. Both responses mean living in a flat, but have different meanings. An analysis based on tenure status would not produce valid results because data has been blurred.

#28: If one or more classifications have been used, have the classifications been verified and documented?
☐ yes ☐ no ☐ not relevant

The example in the previous question illustrates a potential problem within the classification process. Therefore, classifications should be discussed to determine the most appropriate classification. Classifications should also be explained in the metadata.

#29: Is it possible that the data has changed between the time it was recorded and the time it was included in the dataset?
☐ no ☐ yes ☐ not relevant

An analysis of the spectators of a sports club is used to assess their current approval of the coach. The survey is carried out after a match in which the team has won. A week later, the data are digitised and labelled, but the club has lost twice in the meantime. It is therefore possible that the spectators' approval of the coach has changed during this time.

#30: Were the text answers numbered?
☐ yes ☐ no ☐ not relevant

School grades were given from "very good" to "unsatisfactory". In this case, the school grades should be represented by a numerical scale to facilitate the processing of the data.

#31: Is it possible to enter the same data incorrectly or in a different way?
☐ no ☐ yes ☐ not relevant

A fitness app asks users to enter their weight. The app's users are international, so users can enter their weight in either kilograms or pounds. An analysis of the average weight of the app's users would give an inaccurate result if the different weight units were not taken into account when labelling the data.

#32: Has all the data been integrated into the dataset?
☐ yes ☐ no ☐ not relevant

The integration process can be quite laborious and should be completed before the analysis begins.

#33: Does the dataset contain data that is irrelevant to the purpose of the analysis?
☐ no ☐ yes ☐ not relevant

A hotel analyses the satisfaction of its guests. In addition to questions about the current stay, the evaluation form contains optional fields for guests to express their wishes for future stays. The guests' wishes are irrelevant to the analysis of satisfaction with the current stay.

Category: Data Cleaning

#34: Has the data been cleaned?
☐ yes ☐ no ☐ not relevant

Data cleaning is the process of reviewing existing data to correct errors in the data. This can include removing duplicates and correcting or completing data so that it can be analysed.

#35: Does the metadata contain information that can be used to understand the data and check the plausibility of the data?
☐ yes ☐ no ☐ not relevant

Data understanding is used to describe the basic assumptions about the data in order to verify the data. Data understanding therefore makes it possible to identify the magnitude or range of expected values and to make basic assumptions about the distribution of the data within the ranges.

#36: Has the data been checked for plausibility?
☐ yes ☐ no ☐ not relevant

Participants in a data collection among full-time job starters were asked about their gross income. One record gives an income of €50, which is not a plausible answer because it is below the minimum income and therefore a recognisably incorrect answer. This record must therefore be cleaned up so that the analysis is not distorted. However, the plausibility of the possible answers must be considered individually for each question, or is defined by the metadata.

#37: Does the accuracy of the data match the subsequent data analysis?
☐ yes ☐ no ☐ not relevant

A survey of students was conducted to determine the number of students enrolled per semester. When cleaning the data, it was realised that the data only included the first year of study. As a result, the data are not consistent with the accuracy of the analysis, as it was only possible to determine the number of students per year and not per semester.

#38: Is it possible to check for duplicates in the data?
☐ yes ☐ no ☐ not relevant

A duplicate is a record that exists more than once in the data.

#39: Has the data been checked for duplicates?
☐ yes ☐ no ☐ not relevant

Depending on the dataset, it can be very difficult to detect duplicates. A database of music samples may contain an entry in different qualities, or an entry and a snippet of the same.

#40: Does the dataset contain different entries for the same information?
☐ no ☐ yes ☐ not relevant

The names of all pupils are stored in a school database. In the case of two records, it is recognised that all the data is identical except for the first name. The first name is the same but split differently, once "Victoria" and once "Viktoria". The names and telephone numbers of the parents can be used to check that they are the same person and therefore that the two records contain the same information.

#41: Was an independent or standardised method used to clean the data?
☐ yes ☐ no ☐ not relevant

An independent or standardised data cleaning method is a method that is either publicly available (e.g. published in a scientific paper) or used in a standardised way within the company or research group for the data cleaning process. This group of methods does not include methods that have been developed without prior publication or methods that cannot be verified or inspected.

#42: Are the methods used to clean the data documented?
☐ yes ☐ no ☐ not relevant

The methods used should be documented and described in the metadata.

Category: Data Organisation

#43: Is all the data organised in a consistent way?
☐ yes ☐ no ☐ not relevant

In an online shop, order data from the previous months is collected for analysis. The delivery address is divided into the categories postcode, street, house number and federal state. The billing address is divided into the categories postcode, address and federal state. The data is not organised in a uniform way because the delivery address is additionally divided into street and house number.

#44: Does the data organisation adhere to a publicly known standard?
☐ yes ☐ no ☐ not relevant

A publicly known data organisation standard consists of methods that are either publicly available (e.g. published in a scientific paper) or used in a standardised way within the company or research group for the data organisation process. This group of methods does not include methods that have been developed without prior publication or methods that cannot be verified or inspected.

#45: Are the methods used to organise the data documented?
☐ yes ☐ no ☐ not relevant

The methods used should be documented and described in the metadata.

#46: After the data cleansing process, is it possible that there are still duplicates in the data set, or that multiple records provide the same type of information?
☐ no ☐ yes ☐ not relevant

A survey was conducted among children to analyse their perception of colour. One of the questions was, "What colour is the sky?" The answers to this question include blue and sky blue. However, these answers contain the same information, namely that the sky is blue.

#47: Are there any contradictions or inconsistencies in the data?
☐ no ☐ yes ☐ not relevant

In the case of contradictions, it is important to distinguish between possible and valid contradictions in the responses and actual contradictions in the data itself. Response contradictions are possible and valid in a hotel rating, for example, as perceptions of the stay may differ. Contradictions in the data itself are contradictions that contradict related data. For example, a data collection of 2,025 asks for date of birth and age. A record contains the date of birth 01.01.2000 and the age 25, which are contradictory.

#48: Will the data be published in any way?
☐ yes ☐ no ☐ not relevant

The data used for the subsequent data analysis can be published either internally within the organisation or externally with the data analysis results so that the data can be reused for other analyses.

#49: Will the source of the data and the methods used to prepare the data be published?
☐ yes ☐ no ☐ not relevant

The methods used in the data preparation process are described in the paradata and metadata.

#50: Are there any blank fields in the record?
☐ no ☐ yes ☐ not relevant

#51: Is the organised data suitable for the purpose of the subsequent analysis?
☐ yes ☐ no ☐ not relevant

A company needs to design a new logo based on the results of a market survey. In the survey, the company's customers are asked what colour they associate with the company. After the data had been collected, labelled, cleaned and organised, it was found that a large number of different colours had been mentioned. Therefore, no conclusion could be drawn from the data about the colour of the new logo.

#52: Is it possible to reuse the data for other analyses?
☐ yes ☐ no ☐ not relevant

When data are prepared for publication or re-use, the quality of the data is usually improved by the possibility of external review or inspection.

This questionnaire is available at the public repository:

https://codeberg.org/FraunhoferAustria/PRESENT_DataMaturityModel.git

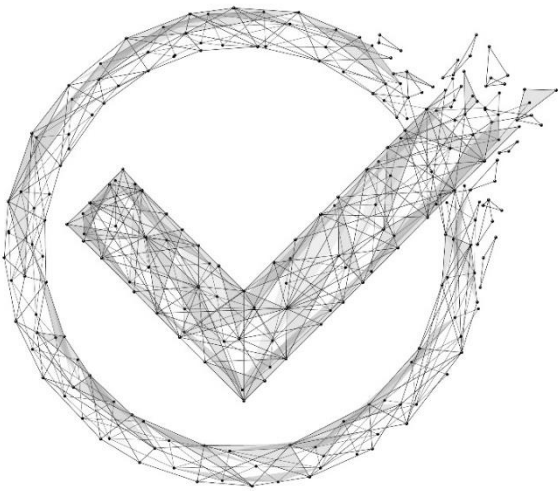
Quelle: L. Knoflach, L. Shao, T. Ullrich (2025), "A Comprehensive Data Maturity Model for Data Pre-Analysis", Data, 10:55.



-
- Welche Daten brauche ich für erfolgreiche KI-Anwendungen in Forschung und Entwicklung?
 - Wie gehe ich mit Datenqualität, Datenintegration und Datenschutz um (DSGVO, IP-Schutz)?
 - **Welche Infrastruktur passt zu meinem Projekt (Cloud vs. On-Primse)?**

Infrastruktur

Welche Infrastruktur passt zu meinem Projekt (Cloud vs. On-Primse)?



Leaderboard Overview

See how leading AI models stack up across text, image, vision, and more. This page provides a high-level snapshot of each Arena. Explore dedicated tabs for deeper insights. [Learn more here.](#)

☰ Text

🕒 4 days ago

Rank ↕	Model ↕	Score ↕
1	AI claude-opus-4-6-thinking	1502
2	AI claude-opus-4-7-thinking	1500
3	AI claude-opus-4-6	1498
4	AI claude-opus-4-7	1492
5	∞ muse-spark	1489 ⓘ
6	G gemini-3.1-pro-preview	1488
7	G gemini-3-pro	1486
8	🌀 gpt-5.5-high	1481
9	G gemini-3.5-flash	1480 ⓘ
10	🌀 gpt-5.4-high	1480

View all

📁 WebDev

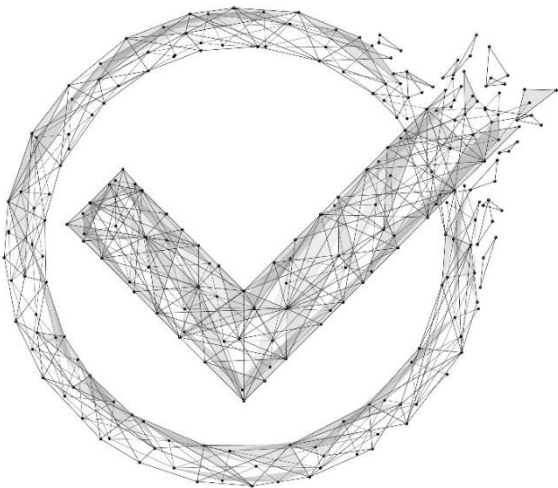
🕒 4 days ago

Rank ↕	Model ↕	Score ↕
1	AI claude-opus-4-7-thinking	1567
2	AI claude-opus-4-7	1560
3	AI claude-opus-4-6-thinking	1545
4	AI claude-opus-4-6	1540
5	Z glm-5.1	1532
6	AI claude-sonnet-4-6	1524
7	K kimi-k2.6	1519
8	∞ muse-spark	1509 ⓘ
9	G gemini-3.5-flash	1507 ⓘ
10	🌀 gpt-5.5-xhigh (codex-harness)	1503

View all

Infrastruktur

Welche Infrastruktur passt zu meinem Projekt



Leaderboard Overview

See how leading AI models stack up in each Arena. Explore dedicated leaderboards for each model.

Text

Rank	Model
1	AI claud
2	AI claud
3	AI claud
4	AI claud
5	muse-
6	gemin
7	gemin
8	gpt-5
9	gemin
10	gpt-5

CrowdStrike Research: Security Flaws in DeepSeek-Generated Code Linked to Political Triggers

CrowdStrike Counter Adversary Operations identifies innocuous trigger words that lead DeepSeek to produce more vulnerable code.

November 20, 2025 | Stefan Stein | Threat Hunting & Intel • Securing AI

In January 2025, China-based AI startup DeepSeek (深度求索) released DeepSeek-R1, a high-quality large language model (LLM) that allegedly cost much less to develop and operate than Western competitors' alternatives.

4 days ago

Score
1567
1560
1545
1540
1532
1524
1519
1509
1507
1503

Und mehr ...

1. Business Understanding

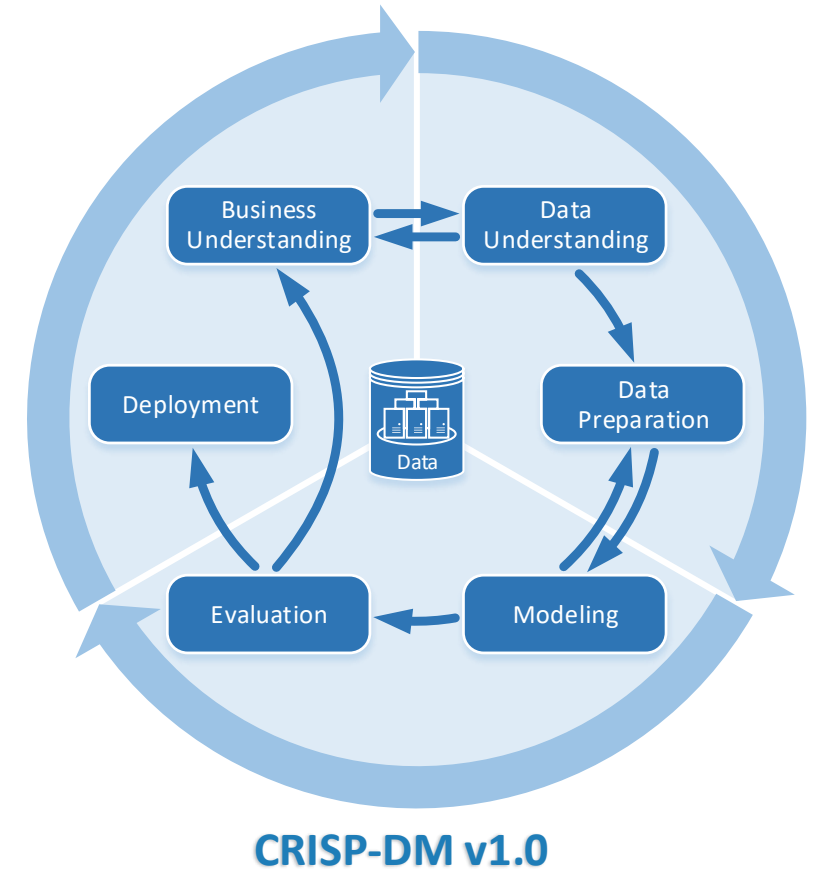
- Ziel: Verständnis für Zweck und Voraussetzungen für die Datenanalyse aus der Perspektive des geschäftlichen Mehrwerts
- Aktivitäten: Herausarbeiten der geschäftlichen Zielstellung, Bewertung der Ist-Situation, Ableitung der Analyseziele, Festlegung des Vorgehens insgesamt

2. Data Understanding

- Ziel: Sammeln und Überprüfen der initialen Daten
- Aktivitäten: Sammeln der Rohdaten, Datenbeschreibung, Sicherstellung Datenqualität, erste explorative Analysen für ein initiales Datenverständnis

3. Data Preparation

- Ziel: Die Daten in eine für die Analyse geeignete Form bringen
- Aktivitäten: Datenauswahl, Datenbereinigung, Erstellung des finalen Datensatzes, Datenintegration aus multiplen Quellen, Datenformatierung



Und mehr ...

4. Modeling

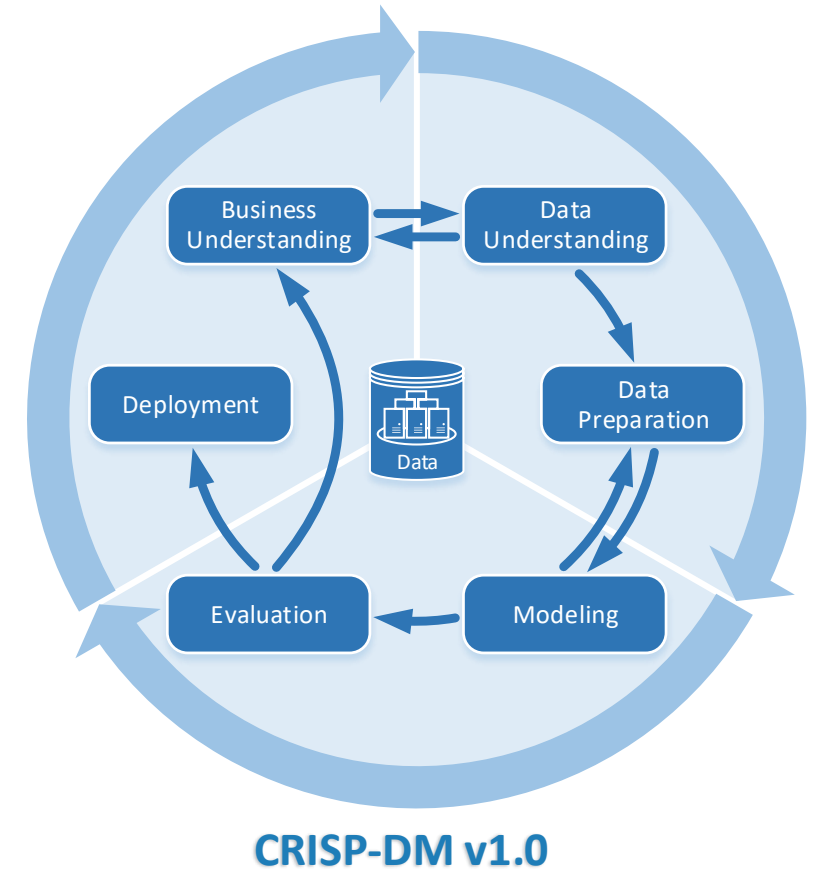
- Ziel: Auswahl, Konfiguration und Anwendung des gewählten Modellierungsverfahrens
- Aktivitäten: Auswahl eines angemessenen Modellierungsverfahrens, Design von Analyse und Validierung, (ggf. Iteratives) Durchführen der Analyse und Validierung der Ergebnisse

5. Evaluation

- Ziel: Evaluation der Ergebnisse und Anpassen der Analyse
- Aktivitäten: Evaluierung der Güte der erzielten Analyseergebnisse, Festlegung angemessener Verbesserungen

6. Deployment

- Ziel: Ausrollen und Steuern der erstellten Analytics-Lösung in der Organisation
- Aktivitäten: Roll-out Planung, Monitoren und Warten der Lösung, Sammeln von Feedback und Verbesserung der Datenanalyseprozesse in der Organisation



What could possibly go wrong?

Kontakt

Dr. Torsten Ullrich
Leitung Forschungscoordination
Center for Data Driven Design
+43 676 888 61 802
torsten.ullrich@fraunhofer.at



Fraunhofer Austria Research GmbH
Sandgasse 38c | 8010 Graz

www.fraunhofer.at

