MISSION KI

Nationale Initiative für
Künstliche Intelligenz
und Datenökonomie

MISSION KI

DFKI-side Kick-off for MISSION KI / TrustifAI

14.10.2024

MISSION KI Team

Gefördert durch:

Bundesministerium für Digitales und Verkehr

aufgrund eines Beschlusses des Deutschen Bundestages

dfki ai Deutsches Forschungszentrum für Künstliche Intelligenz
German Research Center for Artificial Intelligence

acatech
DEUTSCHE AKADEMIE DER TECHNIKWISSENSCHAFTEN

MISSION KI

- 10:00 - Welcome & Short Introduction Round
- 10:15 - Introduction to MISSION KI
- 10:45 - Test Platform
- 11:30 - Quality Platform
- 12:15 - Lunch Break
- 13:00 - Overview of Use Cases
- 13:10 - Use Case 1 - Skin Lesion Analysis
- 13:30 - Use Case 2 - Decision-Support for Tumorboards
- 13:50 - Use Case 3 - Ambulance Capacity Planning
- 14:10 - Use Case 4 - Treatment Plans for Psychological Care
- 14:30 - Use Case 5 - Intubation Planning from Endoscopy Videos
- 14:50 - Coffee Break
- 15:15 - WP8: Community Building
- 15:45 - Administrative & Project Interfaces
- 16:15 - Wrap Up

MISSION KI

Introduction Round

Introduction to MISSION KI

MISSION KI

Project Structure

- since 06/2024
- only pillar 2

- since 01/2023?
- all pillars

- Subcontractor

- “Angebotspartnerschaft“

1

PILLAR – 1

Connecting
data spaces
across sectors

2

PILLAR – 2

Creating transparent
AI Quality & Test standards

3

PILLAR – 3

Supporting the growth
of AI Innvoations

MISSION KI

Confidence in use, planning security for the development of AI

PILLAR – 2

Creating transparent AI quality standards

1 **quality and testing standards** to strengthen competitiveness

2 Innovation & Quality centres as **contact points & experience spaces**

pwc provides a strong **project partnership** for **pillar 2**

AIQ TÜV AI.LAB C.AI Fraunhofer IAIS VDE

Premises of the quality standard

- Voluntary
- Compatible
- Resilient
- Comparable
- Accessible
- Efficient

MISSION KI

From Use Cases to an industrial standard

MISSION KI
Advantages

- **Compatibility** with existing regulations such as the **EU AI Act, NIST, G7 CoC***
- **Voluntary** standard that focuses on **low-risk AI systems**
- **Efficiency** and **accessibility** for **economic benefit,** including for SMEs and startups

WP1

Preparation of the **use case landscape** and a **risk framework** for the development of test approaches

WP2

Draft and development of the **criteria catalog** and **testing procedure** for the voluntary AI quality standard

WP3

Testing, validation and **adaptation** of the criteria catalog and testing approach as well as development of a comprehensive **testing tool landscape**

WP4

Consolidation, **publication** of the AI quality standard and **establishment** for broad acceptance in the AI market

MISSION KI

Our Use Case Landscape

The collection consists of 1331 Use Cases from different sources

Use Cases From PLS
(Plattform Lernende Systeme)

Use Cases from AppliedAI‘s collection

Use Cases from an Open Call

- Selection of **5 representative use cases** to **test, validate and develop** MISSION KI's testing approaches - cooperation of the companies is a prerequisite
- Systematic analysis and grouping of use cases using the dimensions of **industry sector, technology field, main risk and risk class according to the EU AI Act**

MISSION KI

MISSION KI is based on existing and established Frameworks

System for a practicable and application-oriented quality standard

Testing system and overall assessment according to the **VCIO* method** of the VDE SPEC

Extension through **risk-based approach** and **application context** (+R) and **concrete technical tests** (+T) of the Fraunhofer test catalog

Selection of **values, criteria, indicators and observables** based on **relevant catalogs** such as the AIC4 and **customer requirements**

- ✓ Based on **existing frameworks**
- ✓ VCIO-approach incorporates a **systematic, accessible method**
- ✓ Risk analysis provides adaptation of testing to use case context and modularity of testing
- ✓ Incorporation of **concrete test procedures** and **evidences**

Values of the AI Quality standard for testing

MISSION KI

Values for manufacturer-related testing of the AI application

Vision of the test landscape

MISSION KI

Infobox

1. The **Infoportal** is the first point of contact for the examinee who would like to have the object to be tested tested. It provides information for exploration and orientation and gives access to the individual **test centers**, which are also directly accessible to the examinee.
2. An **AI test environment** provides the resources required for testing depending on the use case.
3. **Test procedures and methods** provide the test center with the procedure for the test process and the tests to be carried out.
4. The **AI test interface** guides the test specimen through the test, provides secure access to data and models, and provides test results and other feedback.
5. The **AI test platform** executes various test workflows and test tools and is always in communication with the test environments as well as compliant with the test methods.

acatech MISSION KI Nationale Initiative für Künstliche Intelligenz und Datenökonomie

MISSION KI

Project Structure

Bundesministerium für Digitales und Verkehr

- since 06/2024
- only pillar 2

dfki ai Deutsches Forschungszentrum für Künstliche Intelligenz *German Research Center for Artificial Intelligence*

- since 01/2023?
- all pillars

acatech DEUTSCHE AKADEMIE DER TECHNIKWISSENSCHAFTEN

SDS

- Test Platform (WP1)
- Use Case 1 (WP3)
- Use Case 2 (WP4) JG|U
- Use Case 4 (WP6) UNIVERSITÄT TRIER
- Community building (WP8)

DSA

- Quality Platform (WP2)
- Use Case 3 (WP5) Deutsches Rotes Kreuz

AV

- Use Case 5 (WP7) UKS Universitätsklinikum des Saarlandes

pwc

- Subcontractor

AIQ TÜV AI.LAB C.AI Fraunhofer IAIS VDE

- “Angebotspartnerschaft“

Mission KI

AP8 Mission KI IQZ & Community Building
Mission KI Plattform
AI quality & test standards
AP2 Quality Platform
AP1 (medical) Test Platform
Application
Feedback
Application
Feedback
Development
Improvement
Testing
Deployment
AP3 – Use Case 1 – Skin Lesion
AP4 – Use Case 2 – KITTU
AP5 – Use Case 3 – DRK
AP6 – Use Case 4 – Uni Trier
AP7 – Use Case 5 – Uni Saarland

MISSION KI

Timeline for AP1&2

		2024			2025			
		Q2	Q3	Q4	Q1	Q2	Q3	Q4
AP1	**Test platform**							
1.1	Analysis of the requirements							
1.2	Analysis of the relevant test platforms							
1.3	Concept and prototype of a test platform for trustworthy AI systems in medicine							
1.4	Analysis, categorization and evaluation of existing testing tools							
1.5	Adaptation and, if necessary, development of new testing tools for use cases 1 and 2							
1.6	Preparation and implementation of prototype tests of the use cases							
1.7	Evaluation and gap analysis for the medical testing platform							
1.8	Development of recommendations for action for the medical testing platform and its tools							
AP2	**Quality platform**							
2.1	Language of the AI quality requirements							
2.2	Knowledge base (knowledgebase)							
2.3	Automated reasoning							
2.4	Debugger							

MISSION KI

Medical Test Platform

Adriano Lucieri

What We Want – (SDS-side) Project Goals

1. Clear Definition of AI Quality & Trustworthiness

   What exactly is meant by *Transparency* & co.?

   What does it take an AI system to be considered *Transparent*

   Definitions

   Requirements

2. Bridge theoretical definitions to (as many) operationalizable aspects (as possible)

   What are desiderata for the different aspects of trustworthiness?

   What are the available test tools and which desiderata do they quantify?

   Subrequirements

   Categorization Scheme

3. Propose new ways of (AI-supported) quantitative testing of AI systems

   Thinking outside the box → How can AI be used to support the exploration of Vulnerabilities and Biases?

   How can this be leveraged to extend the quantitative testing?

   Example 1:
   New unsupervised Bias detection method
   → Quantitative testing with existing Fairness methods

   Example 2:
   AI-based methods for API exploitation to automatically test for system vulnerabilities

4. Propose a collection of AI test tools (including datasets & synthesis methods)

   Software toolbox / library?

   Framework?

   Toolbox / Library

   Categorization Scheme

5. Evaluate the test tool collection & framework on use cases

What We Want

Regulatory & Normative Requirements
AI Quality & Trustworthiness
Aspects of TAI
Human agency & Oversight
Technical Robustness & Safety
Privacy and Data Governance
Transparency
Diversity, Non-Discrimination and Fairness
Societal and Environmental wellbeing
Accountability
Requirements
Subrequirements
Observables
Requirement 1
Subreq. 1,1
...
Subreq. 1,s
Observable 1 - m
Subreq. K,1
Subreq. K,s
Observable 1 - m
Requirement n
Observable 1 - m

AI Quality & Trustworthiness
Aspects of TAI
Human agency & Oversight
Technical Robustness & Safety
Privacy and Data Governance
Transparency
Diversity, Non-Discrimination and Fairness
Societal and Environmental wellbeing
Accountability
Horizontally & vertically extendable
Requirements
Subrequirements
Observables
Requirement 1
Subreq. 1,1
...
Subreq. 1,s
Observable 1 - m
Subreq. K,1
Subreq. K,s
Observable 1 - m
Some requirements might have no, or no technical observables
Requirement n
Observable 1 - m
Qualitative Subset
Quantitative Subset
• Documentation & risk Analysis
• User & expert studies
AI could support?
• Technical (model & data related)
• Measurable or approximatable
Main Focus

Combined Top Down & Bottom Up Approach

What do we want?

Aspect	Transparency
Requirement	Explainability
Subrequirement	Correctness

„Property of an AI system to **express important factors influencing the AI** system results **in a way that humans can understand**“

„An explanation should not just give any rationale, but justify **exactly whats happening inside the model.**“

What do we have?

Observable	Interpretation of Infidelity
Test Tool	Captum → Infidelity
Test Method	Infidelity Method

Implemented in: https://captum.ai/api/metrics.html#infidelity

Introduced in Yeh et al. 2019 (https://arxiv.org/abs/1901.09392)

Q: How does the predictor function itself change in response to significant perturbation?

INFIDELITY = (x - (f( ) - f( )))^2

Binary Noise | Original Explanation | Original prediction | Perturbed prediction

$$\mathrm{INFD}(\Phi, \mathbf{f}, \mathbf{x}) = \mathbb{E}_{\mathbf{I} \sim \mu_{\mathbf{I}}} \left[\left(\underbrace{\mathbf{I}^T \Phi(\mathbf{f}, \mathbf{x})} - \underbrace{(\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x} - \mathbf{I}))} \right)^2 \right]$$

Significant Perturbation (I) multiplied with the explanation of the unperturbed image.

Prediction of the perturbed input subtracted from the original input‘s prediction.

For a „good“ infidelity value:

If the relevance of perturbed features is big / small, the difference in prediction should be big / small

Combined Top Down & Bottom Up Approach

Definitions are important!

What do we want?

Aspect	Transparency	
Requirement	Explainability	„Property of an AI system to **express important facto**[illegible]**ncing the AI** system results **in a way that humans can understand**“
Subrequirement	Correctness	„An explanation should not just give any rationale, but justify **exactly whats happening inside the model.**“

Categorization scheme required to map observables to requirements!

- What does it approximate / quanitfy?
- What are the limitations of the approximation?
- & more things like compatibility (input, output modality, etc.).

What do we have?

Observable	Interpretation of Infidelity	„The change in prediction should be aligned with the model‘s explanation when the input data is significantly perturbed.
Test Tool	Captum → Infidelity	Implemented in: https://captum.ai/api/metrics.html#infidelity
Test Method	Infidelity Method	Introduced in Yeh et al. 2019 (https://arxiv.org/abs/1901.09392)

TrustifAI-SDS: In- and Outputs

MISSION KI / TrustifAI (SDS)

Abstract Test Framework (Definitions & Hierarchy)

Combination of PwC & own.

Categorization Scheme (Applicable Method)

Classifying test tools.

Test Tool Collection & Categorization

Including Test Datasets & Supporting Tools

Implementation?

e.g., python library

Utilization

Abstract Test Framework (Definitions & Hierarchy)

Combination of PwC & own.

Categorization Scheme (Applicable Method)

Classifying test tools.

Test Tool Collection & Categorization

Including Test Datasets & Supporting Tools

Implementation?

e.g., python library

Progress so far

- Identification & familiarization with **regulations** (EU AI Act, MDR, IVDR), **standards** (ISO, IEEE, etc.), and **guidelines** (NIST, HLEG, etc.).
- (Ongoing) extraction of requirements for testing.
  - → (Preliminary result: Requirements in literature are very vague, as expected!)
- (Ongoing) survey on test methods for
  - Fairness (Ko Watanabe)
  - Explainability (David Dembinsky)
  - Privacy (Deepak Pathak)
- (Starting) survey on synthetisation and usage of test dataset (Stanislav Frolov)

MISSION KI

Quality Platform

Sergey Redyuk - DSA

MISSION KI

Requirements, Policies, Guidelines, EU AI Act

Training and Experience

Assisted, Continuous Improvement of AI models

Testing, Certification, Audit

Quality Platform

Goals

- WP2.1 A prototype of a standard terminology ("ubiquitous language") for the specification of AI requirements
- WP2.2 Curating an expandable knowledge base of AI quality requirements and measures
- WP2.3 Expansion of the knowledge base to include logical inference units for the automatic inference of requirements validations
- WP2.4 AI "debugger" to automatically analyze the AI development process and to recommend preventive measures during the AI development process in different application areas
- A mapping of AI quality requirements and corresponding tests of the AI test platform (WP1.4 and WP2.1)

"…

This document ***avoids specific guidance*** on management processes. The organization can combine generally accepted frameworks, other International Standards and its own experience to implement crucial processes such as risk management, life cycle management and data quality management which are appropriate for the specific AI use cases, products or services within the scope."

ISO/IEC 42001 IT — AI — Management system

Quality Platform

Modes of operation

AI Debugger (WP2.4)

- Identify vulnerabilities and suggest mitigation actions
- Human-in-the-loop assistant
- End-to-end development process, model as white-box

Integration with the Test Platform (e.g.)

- Choosing relevant tests
- Manage test dependencies and conditions
- Using test scores to identify vulnerabilities

AI Debugger. Parsing Example

```python
data = pd.read_csv('compas-scores-two-years.csv')

# Selecting features and target variable
# Features to be used for classification (adjust as necessary)
features = [
    'age', 'sex', 'race', 'juv_fel_count', 'juv_misd_count', 'juv_other_count',
    'priors_count', 'days_b_screening_arrest', 'c_charge_degree'
]

# Target variable (binary classification: recidivism within two years)
target = 'two_year_recid'

# Splitting the data into features and target
X = data[features]
y = data[target]

# Preprocessing
# Define the categorical and numerical columns
categorical_features = ['sex', 'race', 'c_charge_degree']
numerical_features = [
    'age', 'juv_fel_count', 'juv_misd_count', 'juv_other_count', 'priors_count', 'days_b_screening_arrest'
]

# Preprocessing for numerical data
numerical_transformer = StandardScaler()

# Preprocessing for categorical data
categorical_transformer = OneHotEncoder(drop='first')

# Combine preprocessing steps
preprocessor = ColumnTransformer(
    transformers=[
        ('num', numerical_transformer, numerical_features),
        ('cat', categorical_transformer, categorical_features)
    ])

# Create a pipeline that combines preprocessing with the classifier
pipeline = Pipeline(steps=[
    ('preprocessor', preprocessor),
    ('classifier', LogisticRegression(random_state=42))
])

# Split the data into training and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# Train the model
pipeline.fit(X_train, y_train)

# Predict on the test set
y_pred = pipeline.predict(X_test)

# Evaluate the classifier
accuracy = accuracy_score(y_test, y_pred)
conf_matrix = confusion_matrix(y_test, y_pred)
```

AI Debugger

Data Science Ontology

Requirements for Trustworthy AI

Threats & Vulnerabilities

Mitigation Actions

```
"""
        PREFIX rdf: <{RDF}>
        PREFIX rdfs: <{RDFS}>
        PREFIX owl: <{OWL}>
        SELECT ?sibling
        WHERE {{
            <{class_uri}> rdfs:subClassOf ?parent .
            ?sibling rdfs:subClassOf ?parent .
            FILTER (?sibling != <{class_uri}>)
        }}
"""
```

Tests

- Testing procedure
- Metric
- Scoring functions
- Conditions
- see WP1 :D

Auxiliary Entities

LUNCH BREAK

12:15 PM – 01:00 PM

MISSION KI

Use-Case Overview

Adriano Lucieri

MISSION KI

Goals of Use Case APs

Goals:

Evaluation of Quality & Test Platforms

Initial Development & Improvement of Use Cases

- Different
  - data modalities
  - problem statements
  - development phases
  - cooperation partners

MISSION KI

Use Case Overview & Timeline

		2024 Q2	Q3	Q4	2025 Q1	Q2	Q3	Q4
AP3	**Use case 1 - Image-based diagnosis of skin cancer**							
3.1	Preparation of the AI system for prototype testing							
3.2	Adaptation and development of new testing tools for the application							
3.3	Support of PoC testing							
3.4	Evaluation and gap analysis of the AI system							
3.5	Reworking the AI system to fulfill the test requirements and retesting							
3.6	Analysis of the test procedure from the use case perspective and recommendations for action for the medical testing platform							
AP4	**Use case 2 - Tumor board decisions based on tabular data**							
4.1	Preparation of the AI system for prototype testing							
4.2	Adaptation and development of new testing tools for the application							
4.3	Support of PoC testing							
4.4	Evaluation and gap analysis of the AI system							
4.5	Reworking the AI system to fulfill the test requirements and retesting							
4.6	Analysis of the test procedure from the use case perspective and recommendations for action for the medical testing platform							
AP5	**Use case 3 - Spatio-temporal data in rescue services**							
5.1	Detailed analysis of the use case with regard to testing and risk							
5.2	Demand forecast							
5.3	Forecast of remaining resource utilization							
5.4	Decision support, identification of bottlenecks and improved area coverage. Recommendations for action for the test platform.							
AP6	**Use case 4 - Text and survey data in psychiatric treatment and training**							
6.1	Integration of the quality and testing platform and feedback to it from the use case							
6.2	Involvement of users (psychotherapists) as a key aspect for the development of trustworthy AI systems							
6.3	Analysis of the transcripts							
6.4	Analysis of the rating data of therapeutic behavior							
6.5	Development of an explainable AI model for the improvement and evaluation of clinical interventions							
6.6	User study and feedback							
AP7	**Use case 5 - Endoscopy videos for safer intubation**							
7.1	Integration of the quality and testing platform and feedback to it from the use case							
7.2	3D reconstruction and feature extraction from endoscopy videos							
7.3	Risk prediction for intubation							
7.4	Tests and requirements for implementation							
7.5	Future planning							

Owner	Scope
SDS Adriano Lucieri	Testing & improvement
SDS Dominique Mercier	Testing & improvement
DSA Sergey Redyuk	Development & testing
SDS Christoph Maerz, Johannes Ruf, David Riebschlaeger	Development & testing
AV René Schuster	Development & testing

Mission KI

AP8 Mission KI IQZ & Community Building
Mission KI Plattform
AI quality & test standards
AP2 Quality Platform
AP1 (medical) Test Platform
Application
Feedback
Application
Feedback
Development
Improvement
Testing
Deployment
AP3 – Use Case 1 – Skin Lesion
AP4 – Use Case 2 – KITTU
AP5 – Use Case 3 – DRK
AP6 – Use Case 4 – Uni Trier
AP7 – Use Case 5 – Uni Saarland

MISSION KI

Use-Case 1:
Skin Lesion Analysis

Adriano Lucieri

Caddis

Cannot Express Himself!

Abilities:

- Detects skin cancer
- High detection rates (Sensitivity / Specificity)

Limitations:

- Mute
- Paralysed
- No common concepts

Patients hesitate to be treated by Caddis

Accountability

Trust

Explainability

Caddis

Cannot Express Himself!

Who is responsible in case of wrong diagnosis?

„I don‘t want to be treated by a machine!“

Why did it give this diagnosis?

Computer **A**ided **D**iagnosis System based on Deep Neural Networks (DNNs)

Making CADdis use common concepts

Testing with Concept Activation Vectors (TCAV)

- Train DNN classifier
- Extract concept dataset activations
- Train linear concept classifier
- Computation of TCAV score

A. Lucieri, M. N. Bajwa, S. Alexander Braun, M. I. Malik, A. Dengel and S. Ahmed, "On Interpretability of Deep Learning based Skin Lesion Classifiers using Concept Activation Vectors," *2020 International Joint Conference on Neural Networks (IJCNN)*, Glasgow, United Kingdom, 2020, pp. 1-10, doi: 10.1109/IJCNN48605.2020.9206946.

Testing with Concept Activation Vectors (TCAV)

- Train DNN classifier
- Extract concept dataset activations
- Train linear concept classifier
- Computation of TCAV score

A. Lucieri, M. N. Bajwa, S. Alexander Braun, M. I. Malik, A. Dengel and S. Ahmed, "On Interpretability of Deep Learning based Skin Lesion Classifiers using Concept Activation Vectors," *2020 International Joint Conference on Neural Networks (IJCNN)*, Glasgow, United Kingdom, 2020, pp. 1-10, doi: 10.1109/IJCNN48605.2020.9206946.

Testing with Concept Activation Vectors (TCAV)

- Train DNN classifier
- Extract concept dataset activations
- Train linear concept classifier
- Computation of TCAV score

A. Lucieri, M. N. Bajwa, S. Alexander Braun, M. I. Malik, A. Dengel and S. Ahmed, "On Interpretability of Deep Learning based Skin Lesion Classifiers using Concept Activation Vectors," *2020 International Joint Conference on Neural Networks (IJCNN)*, Glasgow, United Kingdom, 2020, pp. 1-10, doi: 10.1109/IJCNN48605.2020.9206946.

exAID

Select Dataset

Current: /Datasets/Patient_1/
Face_01

Class Predictions

Concept Classification

exAID

Select Dataset

Current: /Datasets/Patient_1/Face_01

Class Predictions

Absence Presence

Nevus

Melanoma

Concept Classification

Concept Classification

Absence 0 Presence

Typical Pigment Network

Atypical Pigment Network

Streaks

Regular Dots & Globules

Irregular Dots & Globules

Blue Whitish Veil

Making CADdis show us the concepts

Concept Localization Maps (CLMs)

- Sequentially occlude image patches.
- Feed occluded images to DNN with concept head.
- Generate heatmap of prediction differences.

Lucieri A., Bajwa M.N., Dengel A., Ahmed S. (2020) Explaining AI-Based Decision Support Systems Using Concept Localization Maps. In: Yang H., Pasupa K., Leung A.CS., Kwok J.T., Chan J.H., King I. (eds) Neural Information Processing. ICONIP 2020. Communications in Computer and Information Science, vol 1332. Springer, Cham. https://doi.org/10.1007/978-3-030-63820-7_21

Selective Concept Localization

Hexagon Pentagon Tripod

Ellipse Hexagon Triangle

Lucieri A., Bajwa M.N., Dengel A., Ahmed S. (2020) Explaining AI-Based Decision Support Systems Using Concept Localization Maps. In: Yang H., Pasupa K., Leung A.CS., Kwok J.T., Chan J.H., King I. (eds) Neural Information Processing. ICONIP 2020. Communications in Computer and Information Science, vol 1332. Springer, Cham. https://doi.org/10.1007/978-3-030-63820-7_21

exAID

Select Dataset

Current: /Datasets/Patient_1/
Face_01

Class Predictions

Absence Presence

Nevus

Melanoma

Concept Classification

Absence 0 Presence

Typical Pigment Network

Atypical Pigment Network

Streaks

Regular Dots & Globules

Irregular Dots & Globules

Blue Whitish Veil

Transparency:
Regular Dots & Globules

Absence Presence

Making CADdis formulate a full explanation

Concept-based Verbal Explanations

Lucieri, A., Bajwa, M. N., Braun, S. A., Malik, M. I., Dengel, A., & Ahmed, S. (2022). ExAID: A multimodal explanation framework for computer-aided diagnosis of skin lesions. *Computer Methods and Programs in Biomedicine*, *215*, 106620.

Concept-based Verbal Explanations

Melanoma

The image has been identified as Melanoma due to moderate evidence of Irregular Dots & Globules, and strong presence of Blue-Whitish Veil, and absence of Pigment Networks and Regression Structures.

Nevus

The image has been identified as Nevus due to moderate evidence of Streaks, and absence of Atypical Pigment Networks, Regression Structures and Blue-Whitish Veil.

Melanoma

The image has been identified as Nevus due to absence of Atypical Pigment Network and Regression Structures, despite strong evidence of Streaks, and strong evidence of Irregular Dots & Globules, and strong evidence of Blue-Whitish Veil.

Nevus

The image has been identified as Nevus due to moderate evidence of Typical Pigment Networks, and absence of Atypical Pigment Network and Blue-Whitish Veil, despite moderate evidence of Regression Structures, and moderate evidence of Irregular Dots & Globules.

Nevus

The image has been identified as Nevus due to moderate evidence of Streaks, despite moderate evidence of Dots & Globules.

Nevus

The image has been identified as Melanoma due to moderate evidence of Blue-Whitish Veil, despite moderate evidence of Regular Dots & Globules.

Lucieri, A., Bajwa, M. N., Braun, S. A., Malik, M. I., Dengel, A., & Ahmed, S. (2022). ExAID: A multimodal explanation framework for computer-aided diagnosis of skin lesions. *Computer Methods and Programs in Biomedicine*, *215*, 106620.

exAID

Select Dataset

Current: /Datasets/Patient_1/
Face_01

Class Predictions

Concept Classification

exAID

Select Dataset

Current: /Datasets/Patient_1/
Face_01

The image has been identified as Melanoma due to mild presence of Atypical Pigment Network and mild presence of Regular Dots & Globules.

Class Predictions

Absence Presence

Nevus

Melanoma

Age	64
Ethnicity	Caucasian
Gender	Male
Lesion Loc	Face

Concept Classification

Absence 0 Presence

Typical Pigment Network

Atypical Pigment Network

Regular Dots & Globules

Diagnosis

The image has been identified as Melanoma due to mild presence of Atypical Pigment Network and mild presence of Regular Dots & Globules.

Concepts

Diagnostic Mode

exAID

Select Dataset

Current: /Datasets/Patient_Dataset_1/

Overlay: Select overlay

Filter: Select filter

Highlight by: -

Sort by: -

Latent View

Concept Influences

Target Class: Nevus

Data Distribution

conv2_3 conv3_4 conv4_6

Classification Metric

Metric: Precision

Confusion Matrix

Educational Mode

MISSION KI

Use-Case 2: Decision-Support for Tumorboards

Dr. Dominique Mercier

Improving treatment recommendations

- Every fourth new cancer diagnosis belongs to the urinary and genital organs.
- Recommending the best therapy is complex (individuality of each patient).
- It requires an extensive knowledge that can hardly be surveyed by individual experts.

Support tumor recommendations and enhance them using external information

Our Vision

- Develop an AI-system in urology showing physicians and patients all the options available.
- Relieving the burden on those involved in the process and to optimizing therapy decisions.
- Integration in the process of digitalization of the healthcare system.
- Increase evidence-based treatment to improve the quality of treatment and long-term survival of patients.

Goals

AI-generated support in the interdisciplinary urologic-oncologic tumor board recommendation incl. explainability by means of study data

- Reduction human errors and risks
- Improvement of the evidence-based therapy recommendation
- Better life and patient satisfaction
- Improvement of the survival rates

Support tumor recommendations and enhance them using external information

Implementation

- **Preprocessing:** Pattern Matching, Regular Expressions, Normalization, Standardization
- **High- and Low-Level Classification:** Machine Learning (Boosting, Deep Neural Networks)
- **Matching and Study Evaluation:** Large Language Models (LLMs)

KITTU - Workflow

1. Align data to data types (improve representation, resolve complex data types)
2. Normalize data
3. Merge values for better data distribution (remove / edit underrepresented data)

KITTU - Workflow

```
1. Do a high-level classification
2. Do a low-level classification
```

High-Level	Low-Level
➢ Surgery ➢ Medication ➢ Aftercare ➢ Best supportive Care	➢ Radicale zystectomies ➢ Re-TUR-B ➢ Nephrouretrectomies ➢ Gemcitabine ➢ Vinflunine ➢ Nivolumab

KITTU - Workflow

Preprocessing → Classification → **Explanation** → Related Studies

1. Do Patient-specific explanation
2. Do Class-specific explanation

KITTU - Workflow

Preprocessing → Classification → **Explanation** → Related Studies

1. Do Patient-specific explanation
2. Do Class-specific explanation

KITTU - Workflow

Preprocessing → Classification → Explanation → **Related Studies**

```
Kriterium:  Empfohlenes Medikament muss jeden der Werte Methotrexat UND Vinblastin UND Adriamycin UND Cisplatin UND G-CSF haben.
Extraktion: Arzneimitteltumortherapien: Chemotherapie 3 hat den Wert NONE. Empfohlene Medizin hat den Wert Gemcitabin/Cisplatin.
Auswertung: NEIN
========================================
Kriterium:  TNM Klassifkation (pathologisch): Einteilung nach Lokalisation muss einen der Werte 1 ODER 2 haben.
Extraktion: Klassifizierung nach Standort hat den Wert Ureter (C66; D09.1 HL) (2).
Auswertung: JA
========================================
Kriterium:  Aktuelle Situation: aktuelles UICC muss einen der Werte 5 ODER 6 ODER 7 ODER 8 ODER 9 ODER 10 haben.
Extraktion: Aktuelle Situation: aktuelle UICC hat den Wert III (5) Antwort: Aktuelle Situation: aktuelle UICC hat den Wert III (5)
Auswertung: JA
```

1. Show studies that apply to the patient

Currently LLM approaches

MISSION KI

Use-Case 3: Ambulance Capacity Planning

Sergey Redyuk

MISSION KI

WP5 Use case 3 - Ambulance Capacity Planning

In emergencies, every minute counts. In the event of a heart attack, one minute saved increases the survival rate by **24%**.

400+
ambulances

700k+
emergencies

<15 minutes
response time

WP5 Use case 3 - Spatio-temporal data in rescue services

Goals

- WP5.1 Detailed analysis of the use case with regards to testing and risk
- WP5.2 Demand forecasting
- WP5.3 Prediction of the remaining resource utilization
- WP5.4 Decision support, identification of bottlenecks and improved area coverage
- Recommendations for action for the test platform

DRK – the Ambulance Service Process

Improve QoS and decision-making policies

WP5.1 Use case analysis

Testing, risks, and compliance with the EU AI Act

- Revision of the ambulance service process
- Summary of ethical, legal, and other regulatory requirements (both explicit and implicit aka best practices)
- Description of emergency calls (ultimately, with references to the existing data) that depict bottlenecks, worse-case scenarios, and high optimization potential
- Parts of the ambulance service process supported by information systems and how they operate
- Interfaces and means for human oversight that are already in place, even if the current application does not involve AI tools

WP5.2 Demand forecasting

How many ambulances would region A require in a given timeframe Y?

- Efficient dispatching
- Reducing / managing the workload of rescue stations and ambulance crews
- Point processes (spatio-temporal)

WP5.3 Prediction of the remaining resource utilization

Time-to-availability of a given ambulance

Factors

- type of operation
- the clinical picture
- location and accessibility of the operation
- whether other forces such as an emergency doctor need to be called in
- location of the target hospital
- capacity utilization or handover time of the target hospital
- recovery time of the ambulance vehicle (e.g. disinfection in the case of infectious diseases)
- distances to be covered and traffic

WP5.4 Decision support, identification of bottlenecks and improved area coverage

Support dispatcher' decision-making
and identify possible future bottlenecks

- Design of services (where the majority of existing laws happens to focus)
- Logistics
- Service analytics
- Can we advise on the necessity to acquire more resources on a county level in a data-driven manner?

Discussion is ongoing

WP5 Challenges

Emergency response is a complex, multi-faceted, dynamic optimization problem

- Specifics of the dispatching policies
  - Ambulance with the fastest estimated response time
  - Possible reassignment
  - Triage and priority rules
- Optimization objectives
  - Fair coverage, population and surface coverage
  - Coverage computed as equality or equity
  - Crew fatigue and minimization of excessive relocation
  - Uncovered demand (demand that could not be met)
  - Preparedness (regional level as a whole; coverage probability - for given fleet location, fraction of the region where the 95%-15-minute rule likely won't work; probability that ambulances are occupied at the time of a new call)
- Volatile signals
- Event-based architecture

MISSION KI

Use-Case 4:
Treatment Plans for Psychological Care

Johannes Ruf & David Riebschlaeger

TrustifyAI- AP6: Enhancing and Evaluating Clinical Interventions

Group FLLM – Christoph Maerz, Johannes Ruf, Tahseen Rizvi, Philipp Schneider

dfki ai

MISSION KI Nationale Initiative für Künstliche Intelligenz und Datenökonomie

Enhancing and Evaluating Clinical Interventions

Goals of AP6

Developing a trustworthy AI system for enhancing and evaluating clinical interventions

Combination of unstructured text data and quantitative indicators

Complement feedback from experienced colleagues

Focus on quality and trustworthiness

Explainability and transparency

Ease of use and end-user adoption

Data protection and privacy

Handling of highly sensitive patient data

Development of methods for joint system development while preserving privacy

Testing and evaluation: Testing of the quality platform from WP2

Use of the AI system to test the platform from WP1

Evaluation of the applicability of the project results

Work Packages

- WP 6.1: Integration of the quality and testing platform as well as feedback to it from the use case
- WP 6.2: Involvement of users (psychotherapists) as a key aspect for the development of trustworthy AI systems
- WP 6.3: Analysis of Transcripts
- WP 6.4: Analysis of Rating Data on Therapeutic Behavior
- WP 6.5: Development of an explainable AI model for the improvement and evaluation of clinical interventions
- WP 6.6: User Study and Feedback

Data

- Video (short introduction of the case, very short video of a patient)
- Situation description of the video
- Transcript of patient statements
- Transcript of therapist's responses or spoken words
- Self-assessment of the therapist
- Assessment by two raters
- Ratingmanual
- Preprint about how the rating schema was developed

Simplified Clinical Intervention **Pipeline**

* Advancing with research in advanced reasoning with critics and multiple agents including multiple rag requests.

Your thoughts and ideas?

MISSION KI

Use-Case 5:
Intubation Planning from Endoscopy Videos

René Schuster

TrustifAI

AP 7 – Use Case 5 – Safe Intubation

Status Quo

Everyday life @ UKS

- Worldwide >230 M anesthesia per year
- ~ 25 K at UKS per year
- Anesthesia
  → lung paralyzed
  → intubation

Risks

- 8 % of intubations are difficult or impossible
- Within seconds: Threat to life
- 90 % of hard cases are unexpected!
- Last resort: Tracheotomy
- High stress on medical staff

Chances

- At UKS, anesthesia and ENT doctors work closely together

- Endoscopic video of every air tube before narcosis
  → Data
  → Prediction of difficulty

Goal

- Prediction of difficulty
- Suggestion of visual/medical indicators

Methods

Part 1

- Training of a (video) classifier
- Continuous validation
  - Generalization across patients?
- Analysis (Biases, Robustness, Explainability)

Part 2

- Explainable AI
- Which parts of the air tube are relevant for the decision
- Hints for the doctors

Part 3

- Monocular non-rigid structure-from-motion
- Challenges
    - Deformations
    - Distortions
    - Reflectivity (wet surfaces)
    - Low textures, homogeneous areas

Kairanda et al. “φ-SfT: Shape-from-Template with a Physics-Based Deformation Model”, CVPR 2022.
Golyanik et al. “Intrinsic Dynamic Shape Prior for Fast, Sequential and Dense Non-Rigid Structure from Motion with Detection of Temporally-Disjoint Rigidity”, 3DV 2020.
Su et al. “A Shape Completion Component for Monocular Non-Rigid SLAM”, ISMAR 2019.

Data

First Data Package(s)

- Initially, very imbalanced class distribution
- Now: 79 videos (balanced)
- 4 Cormack scores
- Binary classification: I/II vs III

Score	# Patients	# Frames
1	40	30349
2	8	3706
3	31	42965
4	0	0

Score	# train seqs	# train frames	# val seqs	# val frames
I/II	38	27233	10	6822
III	25	34384	6	8581

Dataset Overview

- Central frame for each sequence

train
I-II

train, III

val
I-II

val, III

First Results

Cormack-Classification with ResNet101

Accuracy: 92.31 % (val set)

- Single image
- The two categories are balanced
- The model generalizes
- Some C3 samples are misclassified

Suspicion: Color/intensity bias

Color Bias

- Average grayscale intensity per class, training data

AP 7 ←→ AP 1/2

Thank you for your attention!

DFKI GmbH
Department Augmented Vision
Trippstadterstr. 122
D-67663 Kaiserslautern

René Schuster
rene.schuster@dfki.de
https://av.dfki.de/

Coffee break

2:50 PM – 3:15 PM

MISSION KI

AP8 – Community Bulding

Adriano Lucieri

MISSION KI

Goals of AP8

Timeline:

		2024			2025			
		Q2	Q3	Q4	Q1	Q2	Q3	Q4
AP8	Mission AI contact point and community							
8.1	Collecting and analyzing challenges from the industry							
8.2	Thematic workshops on trustworthy principles							
8.3	Curation of showroom exhibits							
8.4	Monthly TrustifAI consultation							
8.5	Final event to present the project results							

Goals:

Strengthening the MISSION KI community

External presentation & promotion of MISSION KI

- Innovation & Quality Center (**IQZ**) **Kaiserslautern** as the **primary point of contact** with competence on **AI quality & trustworthiness**
- Building a **strong local AI community** that is **well connected**
- Targeting **founders**, **startups**, **SMEs**, and **other interested stakeholders**
- **Knowledge transfer** through
  - **Workshop series** on trustworthy AI
  - **Consultation** service
  - Showroom **exhibits**

Workshop Series

- Series of workshops on topics around Trustworthy AI
- First workshop last week (10.10.2024) on the EU AI Act
- Next workshops:
  - 29.11.2024 (AI & Leadership)
  - January 2025 (AI Transparency)
  - Many more to come!

If you have ideas for topics & potential speakers, let us know ☺!

MISSION KI

Save the Date! - AI & Leadership Workshop

Agenda

01 AI101: A brief introduction to current AI technologies

02 Leadership redefined: AI meets leadership

- Use of AI systems at management level
- Leadership in the context of AI
- Human-machine collaboration
- AI & Ethics

03 Trustworthy AI: the key to safe innovation

- Pitfalls of modern AI systems
- The cornerstone of trustworthiness
- Trustworthiness in practice

When?
Friday, 29.11.2024

Where?
IQZ Kaiserslautern

IQZ Kaiserslautern – Consulting Services

- Individual consultation on topics around AI quality and trustworthy AI such as:
  - AI test tool selection
  - Achieving AI Quality by Design
  - Navigating the EU AI Act
  - Best practices for trustworthy AI solutions
  - Networking between community members

©DFKI 2024

MISSION KI

Exhibits at Deutsches Technikmuseum Berlin & Project 42

App

Guided Lesion Management

Guided image capturing and anamnesis. Easy management of new and past cases.

Immediate xAI Feedback

Online analysis using latest state-of-the-art explainable Deep Learning algorithms.

Direct Doctor's Consultation

Direct, teledermatological communication interface connecting patients to expert dermatologists.

- SkinDoc demonstrators in 42 and DTM Berlin from 1.Q 2025.
- More demonstrators welcome!

MISSION KI

MISSION KI advertising channels

IQZ Website with:

- Newsletter
- Appointment booking
- Latest News

Online survey to determine the needs of stakeholders

MISSION KI LinkedIn page

MISSION KI

Administrative & Project Interfaces

Ludger van Elst

MISSION KI

Wrap Up

Ludger van Elst