



Prüfplattform für KI in der Medizin

BMC Online Seminar

Elora Schörverth

Fraunhofer HHI

Standards & Prüfkriterien



Datenqualität

Ethische und rechtliche Einschränkungen.



Sicherheit

Kein unberechtigter Zugriff auf die Daten.



Bias

Klassenrepräsentation, Aufteilung des Datensets.



Risikoanalyse

Gefährdungsfaktor für Patient*innen.



Metriken

Klare Definitionen für Performanz, Bias, Robustheit, Unsicherheit, Erklärbarkeit.



Expertise

Teilhabe von Mediziner*innen an Entwicklung.

Über FG-AI4H

ITU/WHO Fokusgruppe, deren Ziel es ist eine standardisierte Assessment Plattform für KI-basierte Methoden in der Medizin zu entwickeln.



AI4H - Assessment Plattform

01

Vielfalt

Anpassbare Evaluation und diverse Modellarten.

02

Fairness

Datensätze unserer Testfälle werden überprüft.

03

Fragebogen

Befragung zu Anwendungsfall, Datensätze und Entwicklung.

04

Sicherheit

Geschlossene Umgebung für das Testen mit unbekanntem Datensätzen.

Zwei Methoden



Text-basiert

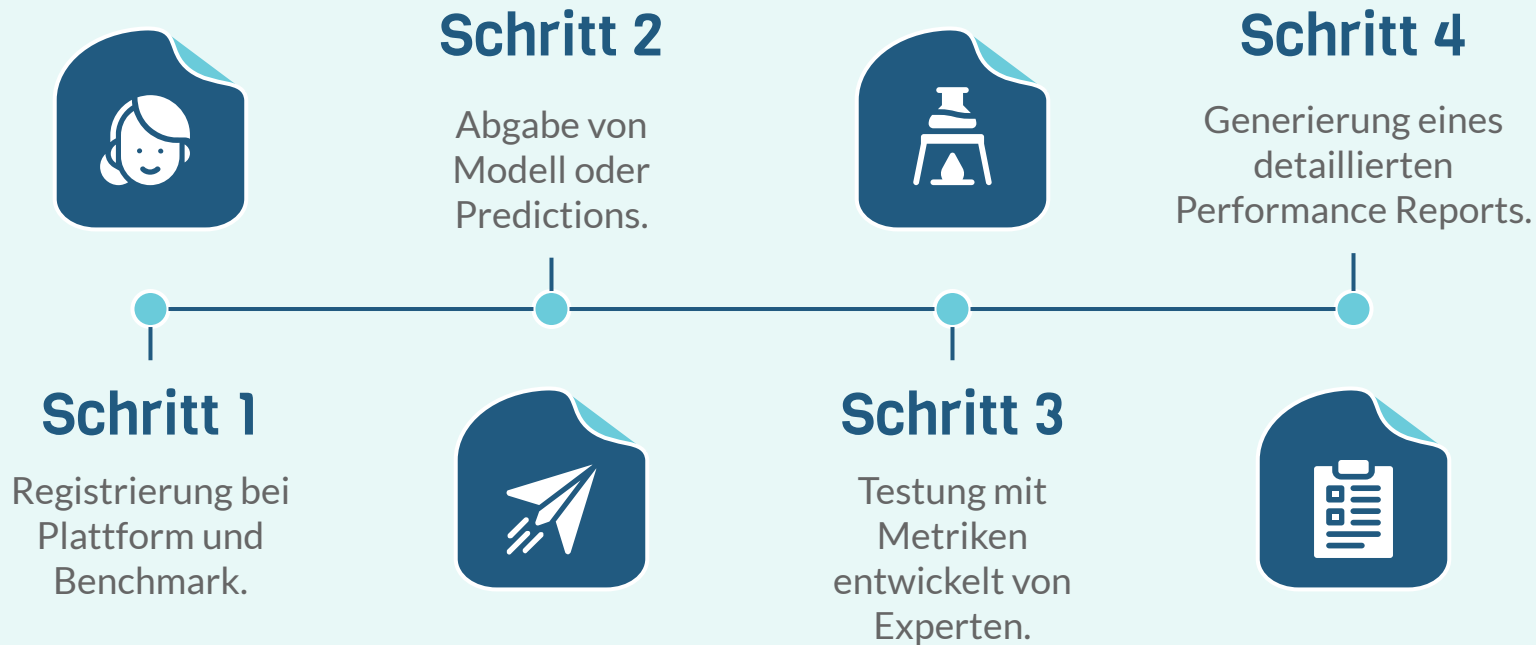
Teilnehmer laden die Ergebnisse des Models hoch, welche mit den Daten verglichen werden.



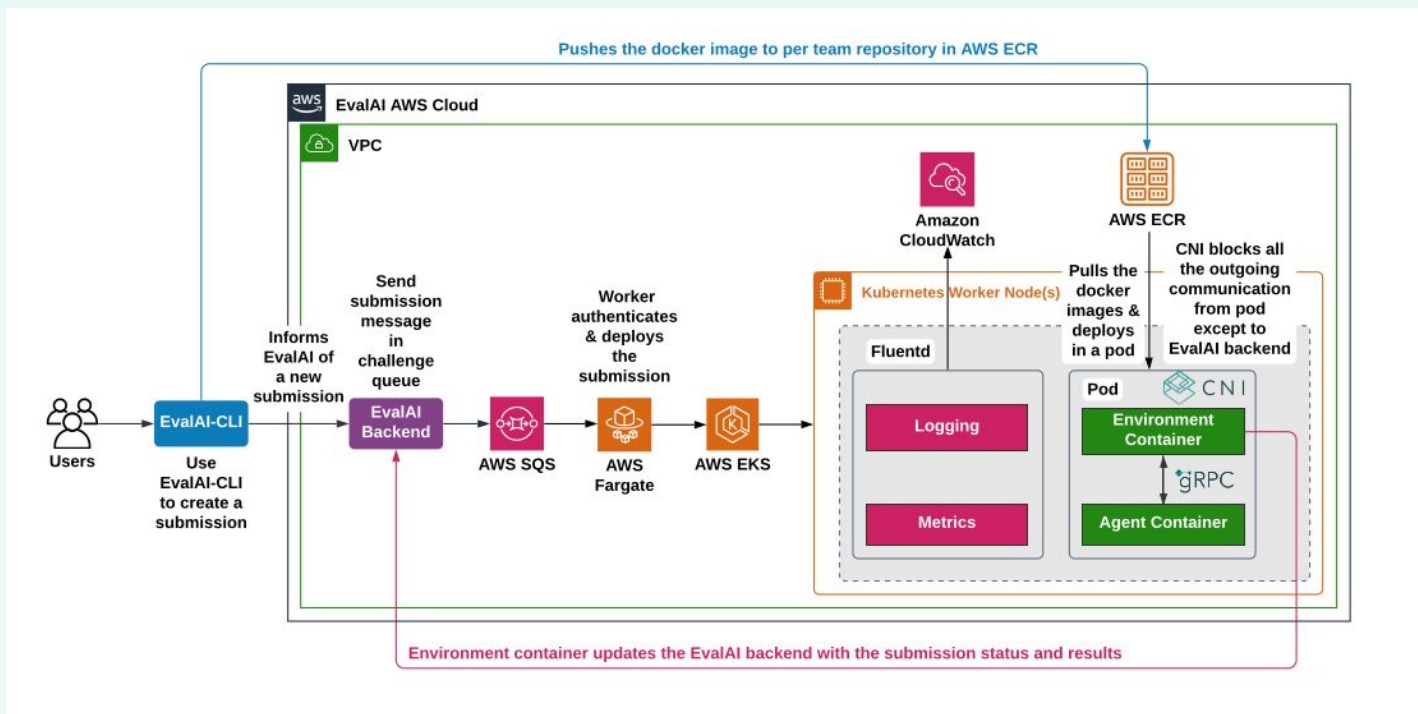
Docker-basiert

Teilnehmer geben ihre KI Modelle als Docker Container ab. Die Modelle werden in einer sicheren Umgebung gegen die Daten getestet.

Evaluation - Ablauf



Evaluation - Software



Jain (2021) p.22

Unsere Anwendungsfälle

| Team Name | Topic Group | ML Domain | Clinical Domain | Regulatory Domain | Ethical Domain | Audit Developers | Meta Advisory Team |
|-----------|---|---|--|---|--|--|---|
| Team - A | TG- Symptoms (Driver: Henry Hoffman) | Frank Klawonn | OPEN | Carolin Prabhu | Marta Lemanczyk | Bastiaan Quast | OPEN |
| Team - B | TG- Ophthalmology (Driver: Arun Stroff) | Muhammad Abdullah (Luca Gilli) | Aaron Lee Xiao Liu | Sheena.macpherson | Manuel Bierwirth | Danny Xie Li Elora-Dana Schörverth | OPEN |
| Team - C | TG- Malaria (Driver: Rose Nakasi) | OPEN | Giovanna Gutierrez | Martin Meyer | Andrea Romaoli | Kaushik Manjunatha Gustavo Rodriguez-Nunez | OPEN |
| Team - D | TG- Psychiatry (Driver: Nicolas Langer) | Stefan Haufe Markus Wenzel | OPEN | Shobha Iyer | Lukas Picek | Nikita Agarwal | Anuja Negi Mahta Mousavi Johannes Stadinger |
| Team - E | TG-Snake (Driver: Rafael Ruiz de Castaneda) | Lukas Picek | Isabelle Bolon Lukas Picek | Dominik Schneider | Amel Ghoulia | Luca Gilli | OPEN |
| Team - F | TG- Neuro, (Driver: Marc Lecoultrre) | Alberto Merola Erick Alvarado Ferath Kherif Sebastian Niehaus | Ferath Kherif Sebastian Niehaus Eva Weicken | Alberto Merola Catherine Lowe | Yi Zeng | Erick Alvarado | OPEN |
| Team - G | TG- Outbreaks (Driver: Auss Abbood and Stéphane Ghazzi) | OPEN | Matthew Arentz | OPEN | Daniela Paolotti | Steffen Vogler | OPEN |
| Team - H | TG- Radiology (Driver: Darlington Ahiale Akogo) | Ian Shadforth Saul Calderon Ramirez Jana Fehr Vasanthakumar Venugopal William Wasswa Shada Alsalamah | Andrew Murchison Jana Fehr Vasanthakumar Venugopal William Wasswa Shada Alsalamah Giovanna Gutierrez | Regina Geierhofer Saul Calderon Ramirez Shada Alsalamah | Vasanthakumar Venugopal William Wasswa | Md. Golam Rasul | OPEN |
| Team - I | TG- Dental (Driver: Joachim Krois) | OPEN | OPEN | OPEN | OPEN | OPEN | OPEN |
| Team - J | AI-Vengens Collab (Driver: Luis Oala) | Jerry Fadugba Steffen Vogler | Andrew Murchison Vasanthakumar Venugopal | Catherine Lowe Pat Baird | Amel Ghoulia | Jerry Fadugba | OPEN |

● Phase 1, 12.21

Planung der Anwendungsfälle und erste Tests.

● Phase 2, 01.22

Migration der Tests auf die Plattform.

● Phase , 02.22

Resultate und Veröffentlichung der Reporte.

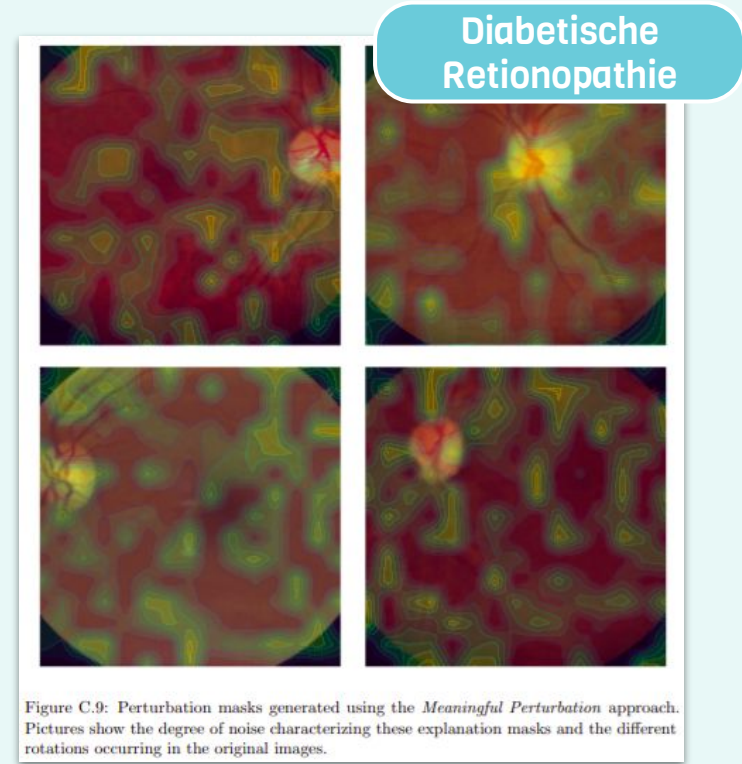
Unsere Anwendungsfälle

The screenshot displays the 'aiaudit.org' web application interface. At the top, there is a navigation bar with the 'AI:+' logo, 'Home', 'Dashboard', and 'Documentation' links on the left, and 'Hi admin' and a 'Logout' button on the right. A left sidebar contains icons for 'Dashboard', 'All Auditing Tasks', 'Hosted Auditing', and 'Participating Teams'. The main content area is titled 'Ongoing Auditing Tasks' and features a grid of eight task cards. Each card includes a title, a thumbnail image, the organizing team, start and end dates, and a 'View Details' button.

| Task Title | Thumbnail Description | Organized by | Starts on | Ends on |
|----------------------------------|--|--------------------|----------------------------|----------------------------|
| Retinopathy Text-Based Challenge | Fraunhofer Heinrich-Hertz-Institut logo | admin_team | Jan 1, 2019 1:00:00 AM CET | Jun 1, 2099 1:59:59 AM CET |
| TG Outbreaks | Red virus particles | TG - Outbreaks | Jan 1, 2019 1:00:00 AM CET | Jun 1, 2099 1:59:59 AM CET |
| TG Radiology | Chest X-ray | TG Radiology | Jan 1, 2019 1:00:00 AM CET | Jun 1, 2099 1:59:59 AM CET |
| TG Dental | Dental X-rays | TG Dental | Jan 1, 2019 1:00:00 AM CET | Jun 1, 2099 1:59:59 AM CET |
| TG - Psychiatry | Word cloud with terms like 'mental health', 'psychiatry', 'mind', 'psyche' | TG - Psychiatry | Jan 1, 2019 1:00:00 AM CET | Jun 1, 2099 1:59:59 AM CET |
| TG - Snake | Snake | TG - Snake | Jan 1, 2019 1:00:00 AM CET | Jun 1, 2099 1:59:59 AM CET |
| TG - Symptoms | Infographic with people icons | TG - Symptoms | Jan 1, 2019 1:00:00 AM CET | Jun 1, 2099 1:59:59 AM CET |
| TG - Ophthalmology | Person's face with eye being examined | TG - Ophthalmology | Jan 1, 2019 1:00:00 AM CET | Jun 1, 2099 1:59:59 AM CET |

Erste Tests

- Beeinträchtigung der Sicht, verursacht durch Diabetes.
- Modell ist ResNet-101 CNN, trainiert auf 82.000 Bildern
- Evaluierung auf Testset mit 848 Bildern, 50% davon mit DR.
- Resultate
 - Unterrepräsentierung von Individuen mit Diabetes +30 Jahre (N=6)
 - Performanzdisparität für Individuen mit Alter >70 Jahre



Erste Tests

```
def evaluate(test_annotation_file, user_submission_file, phase_codename, **kwargs):
    print("Starting Evaluation....")

    print(kwargs["submission_metadata"])
    gt = pd.read_json(test_annotation_file)
    gt = gt.replace(to_replace="retinopathy", value="1")
    gt = gt.replace(to_replace="normal", value="0")
    pred = pd.read_json(user_submission_file)
    pred = pred.replace(to_replace="retinopathy", value="1")
    pred = pred.replace(to_replace="normal", value="0")

    gt_vector = gt['Diagnosis'].values.astype(int)
    #for ground truth id find matching prediction:
    p_vector = []
    for id in gt['Id']:
        val = pred[pred['Id']==id]['Prediction'].values
        p_vector.append(int(val[0]))
    p_vector = np.array(p_vector)

    acc = accuracy_score(gt_vector, p_vector)
    f1 = f1_score(gt_vector, p_vector)
    prec = precision_score(gt_vector, p_vector)
    recall = recall_score(gt_vector, p_vector)

    output = {}
    if phase_codename == "dev":
        print("Evaluating for Dev Phase")
        output["result"] = {}
        {
            "train split": {
                "Accuracy": acc,
                "F1": f1,
                "Precision": prec,
                "Recall": recall,
            }
        }
    ]
    # To display the results in the result file
    output["submission_result"] = output["result"][0]["train_split"]
    print("Completed evaluation for Dev Phase")
```

```
{
  "Id": 1,
  "Prediction": "retinopathy"
},
{
  "Id": 2,
  "Prediction": "normal"
},
{
  "Id": 3,
  "Prediction": "normal"
},
{
  "Id": 4,
  "Prediction": "normal"
},
{
  "Id": 5,
  "Prediction": "retinopathy"
},
{
  "Id": 6,
  "Prediction": "normal"
},
{
  "Id": 7,
  "Prediction": "retinopathy"
},
{
  "Id": 8,
  "Prediction": "normal"
},
{
  "Id": 9,
  "Prediction": "normal"
},
{
  "Id": 10,
  "Prediction": "normal"
},
{
  "Id": 211,
  "Prediction": "retinopathy"
},
{
```

Erste Tests

AI: + Home Dashboard Documentation Hi admin Logout

Dashboard
All Auditing Tasks
Hosted Auditing
Participating Teams

Fraunhofer Heinrich-Heinz-Institut Retinopathy Text-Based Challenge

Organized by: admin_team
Published
Starts on: Jan 1, 2019 1:00:00 AM
Ends on: Jun 1, 2099 1:59:59 AM

★ 0
Toggle Participation

[Overview](#) [Evaluation](#) [Phases](#) [Submit](#) [My Submissions](#) [All Submissions](#) [Leaderboard](#) [Manage](#)

[My Report](#)

Leaderboard

All submissions ranked by performance

Phase: Dev Phase, Split: Train Split Private

B - Baseline submission * - Private submission Sort by latest

| Rank | Participant team | Accuracy | F1 | Precision | Recall | Last submission at |
|------|------------------|---------------------|------|-----------|--------|--------------------|
| 1 | Host_64300_Team | B 0.89 | 0.89 | 0.90 | 0.89 | 23 hours ago |

Referenzen

- FG-AI4H (2021) *Updated call for proposals: Use cases, benchmarking, and data*. Available at https://www.itu.int/en/ITU-T/focusgroups/ai4h/Documents/FGAI4H-CfP_UC_Benchm_Data.pdf
- Jain, Rishabh (2021) *Code-Upload AI Challenges on EvalAI*. Georgia Institute of Technology. <http://hdl.handle.net/1853/64704>
- Oala et al. (2020) *ML4H Auditing: From Paper to Practice*. Available from <https://proceedings.mlr.press/v136/oala20a.html>

Danke!

Haben Sie noch Fragen?

elora.schoerverth@hhi.fraunhofer.de
luis.oala@hhi.fraunhofer.de

hhi.fraunhofer.de/abteilungen/ai.html
itu.int/go/fgai4h/
aiaudit.org/

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon and infographics & images by Freepik