



**SDAIA**  
الهيئة السعودية للبيانات  
والذكاء الاصطناعي  
Saudi Data & AI Authority



أكاديمية طويق  
TUWAIQ ACADEMY



## AI & Data Science bootcamp T5

# OUR VOICE

صوْنَا

# SSLR

(Saudi Sign Language Recognition)

# CONTENT

- PROBLEM AND SOLUTION
- IMPACT ON SAUDI 2030 VISION
- DATA
- MODEL AND PROTOTYPES
- CHALLENGES
- APPLICATION IN REAL LIFE
- FUTURE WORK
- DEMO
- REFERENCES

# PROBLEM AND SOLUTION

**Problem:** People with Hearing Disabilities (PWHD) face significant communication challenges in healthcare due to limitations in spoken and written methods.

**Solution:** Developing an AI model that translates sign language into words aims to bridge the communication gap between People with Hearing Disabilities (PWHD) and medical staff.



# IMPACT ON SAUDI 2030 VISION

- The Ministry of Health has initiated the "We Are With You" campaign to support people with hearing disabilities.
- Our AI model empowers those with hearing disabilities, contributing to this campaign.

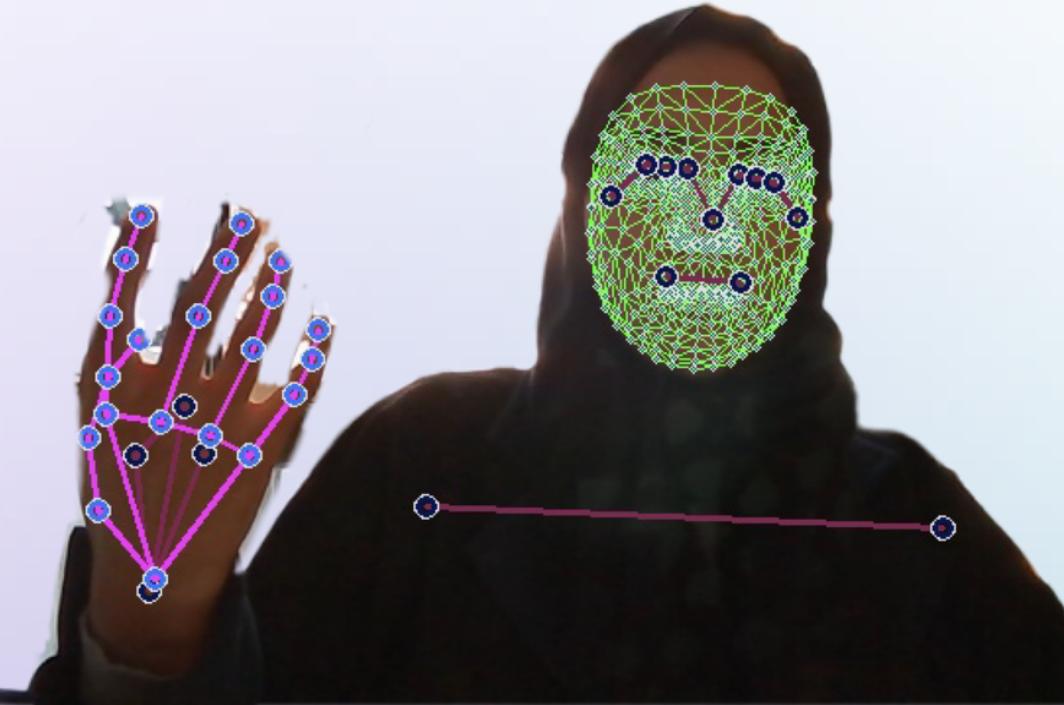


# DATA

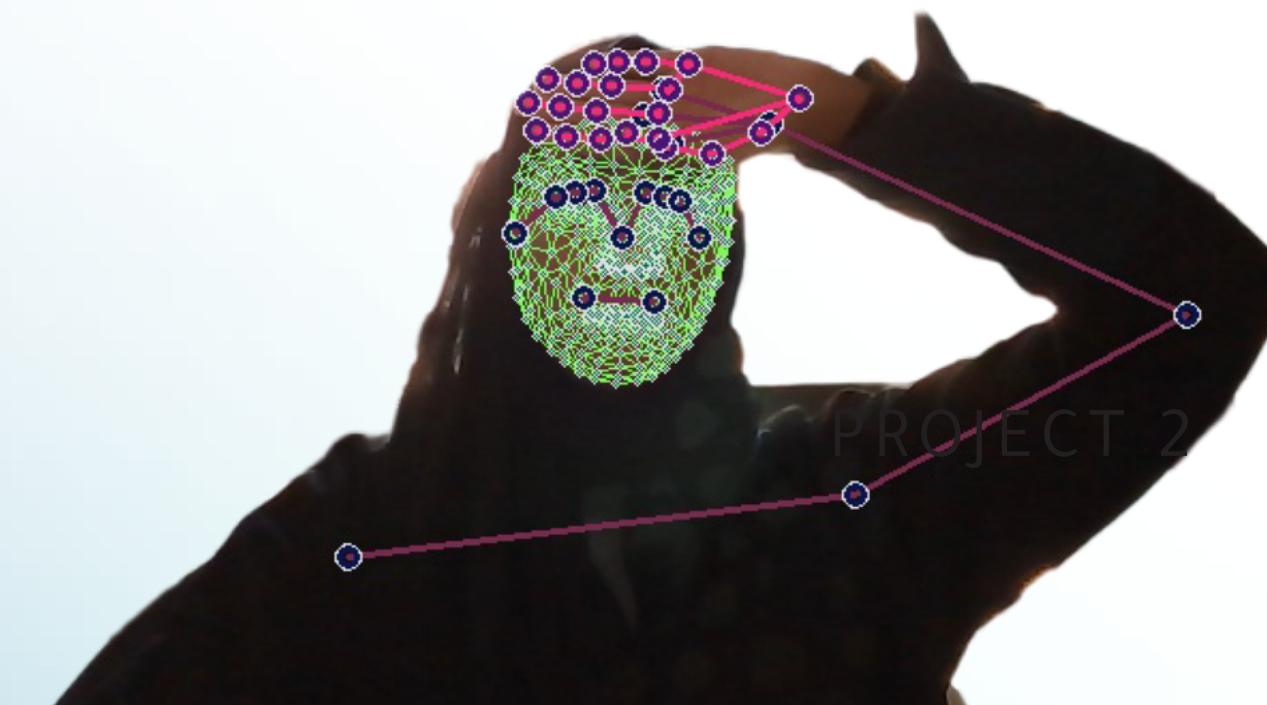
- **Data collection :**

- We recorded the dataset by the team members.
- Each word is recorded 30 times to capture variations.
- Every Video of the recorded words is split into 30 frames.

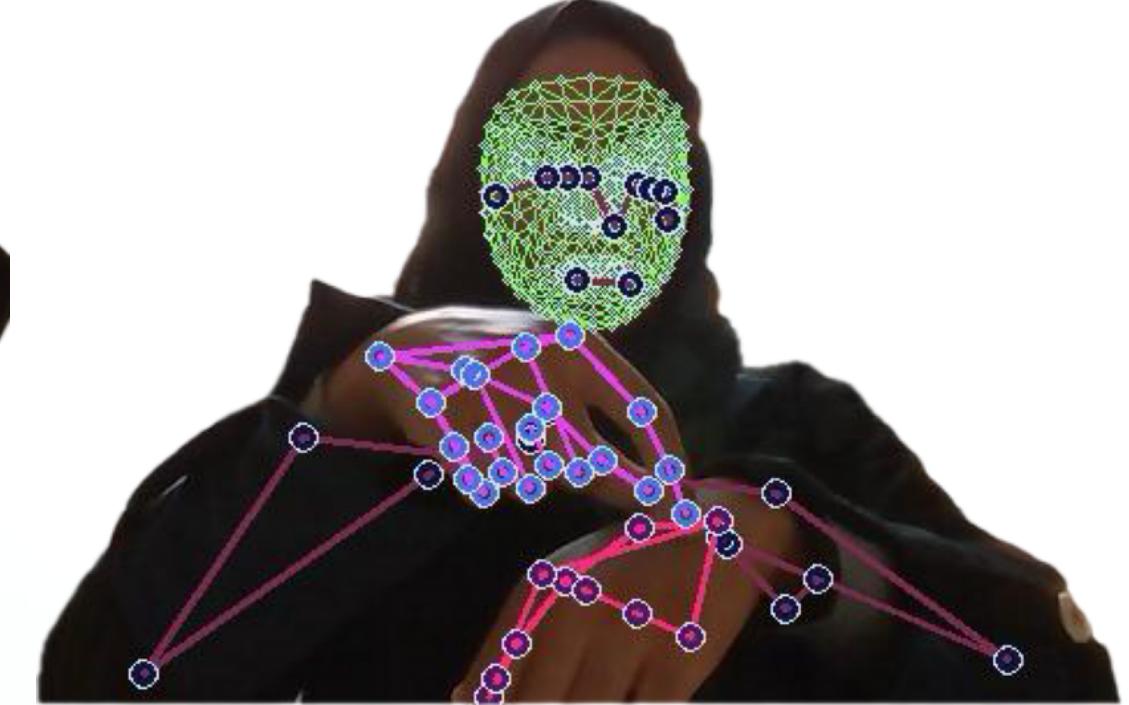
**SIGN:PAIN**



**SIGN:FEAVER**



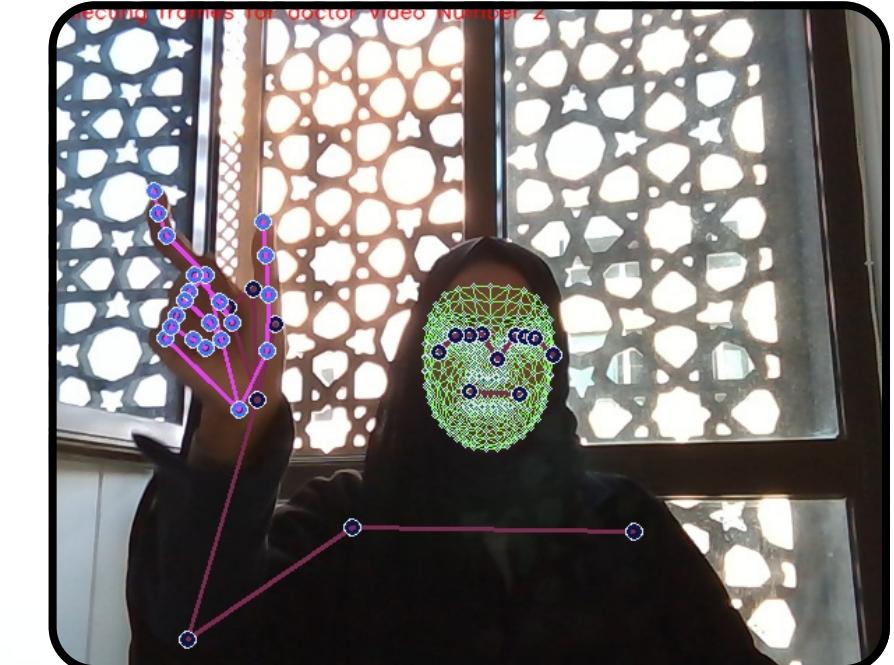
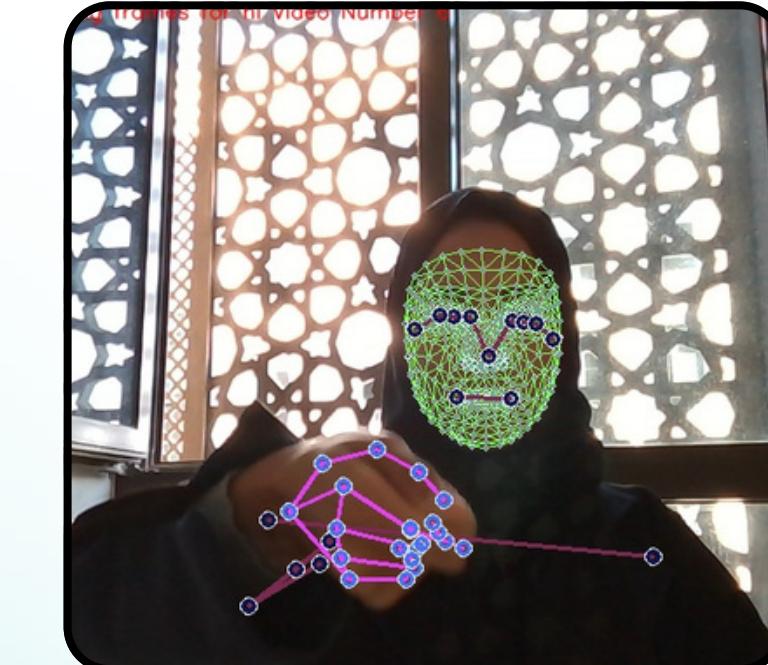
**SIGN:DOCTOR**



# DATA

- **Data Preprocessing and Features Extracting:**

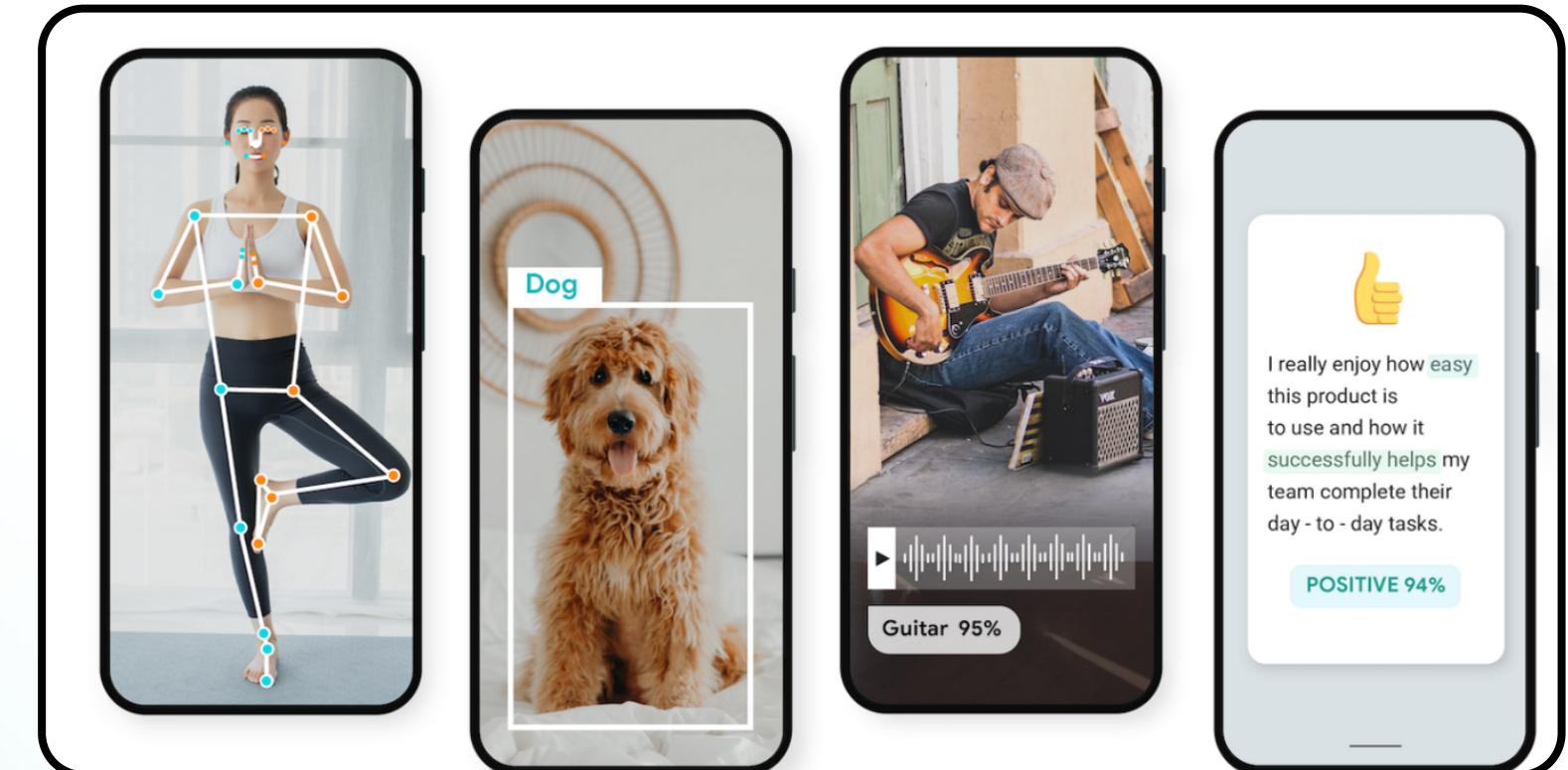
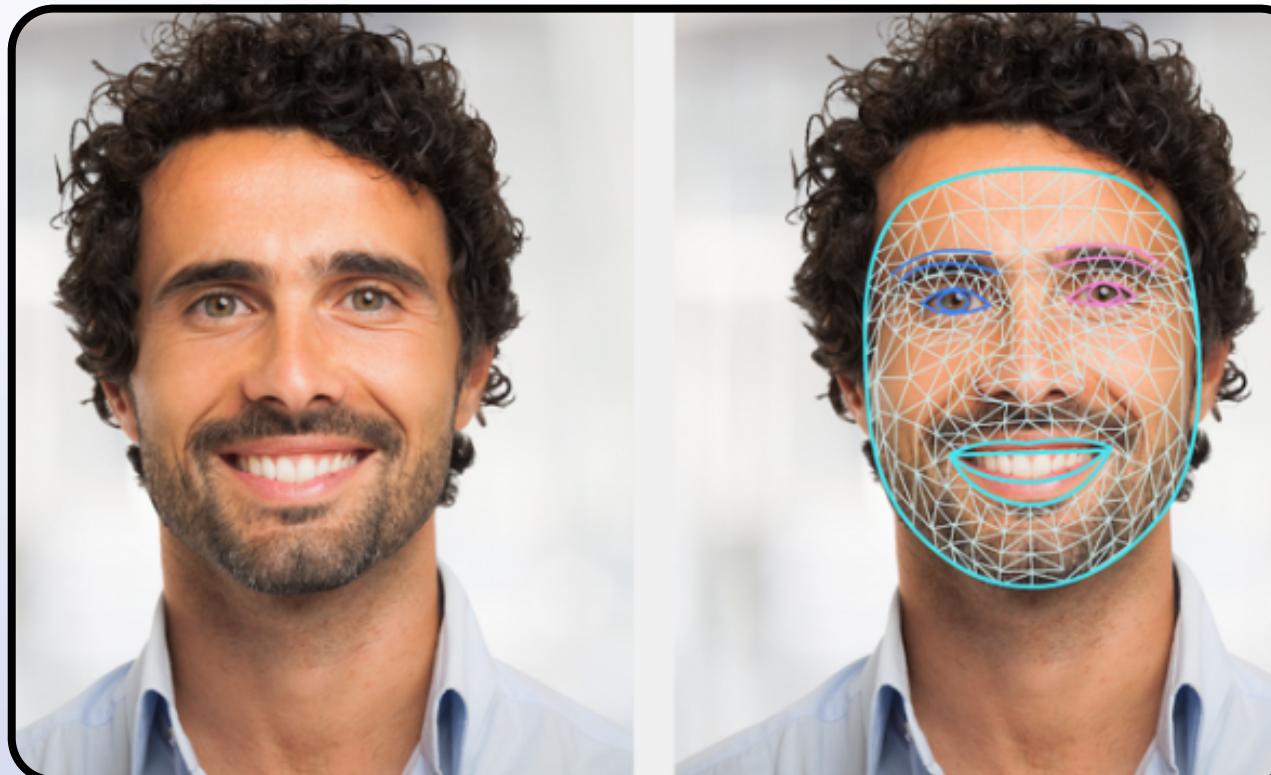
- Used The Mediapipe model to extract features from frames for both left and right hands, pose, and face.
- Each frame is saved as a numpy array, a data structure used for scientific computing.
- Numpy arrays allow efficient storage and manipulation of multi-dimensional data.
- By saving frames as numpy arrays, it becomes easier to perform further analysis and processing on the extracted features.





# MEDIAPIPE

- Versatile framework for computer vision and machine learning tasks.
- Contains pre-built models and algorithms for various multimedia analysis tasks.
- Some tasks include Pose landmark detection, image segmentation, face detection, and object detection.



# INITIAL PROTOTYPES

- **First Prototype:**

- **Dataset:** Videos recorded by team members.
- **Preprocessing:** Videos split into frames.
- **Challenge:** Varying number of frames in each video.
- **Accuracy:** Weak performance (0.05%) due to varying frame lengths.
- **Model architecture:** RNN ( Model
  - Input layer for frame features (shape: (MAX\_SEQ\_LENGTH, NUM\_FEATURES))
  - Input layer for mask (shape: (MAX\_SEQ\_LENGTH,))
  - GRU layer (16 units, return\_sequences=True) with mask input
  - GRU layer (8 units) without mask input
  - Dropout layer (dropout rate: 0.4)
  - Dense layer (8 units, activation='relu')
  - Dense layer (output units=len(class\_vocab), activation='softmax')

# INITIAL PROTOTYPES

- **Second Prototype:**

- **Dataset:** Videos recorded by team members.
- **Preprocessing:** Videos split into frames and extracted features using Mediapipe model.
- **Training Model:** GRU-RNN architecture with fixed input shape.
- **Accuracy:** Weak performance (0.53%).
- **Model architecture:** RNN-GRU(Model
  - GRU layer (64 units, return\_sequences=True, activation='relu', input\_shape=(30, 1662))
  - GRU layer (128 units, return\_sequences=True, activation='relu')
  - GRU layer (64 units, return\_sequences=False, activation='relu')
  - Dense layer (64 units, activation='relu')
  - Dense layer (32 units, activation='relu')
  - Dense layer (output units=actions.shape[0], activation='softmax'))

# INITIAL PROTOTYPES

- **Third Prototype:**

- **Dataset:** Videos recorded by team members.
- **Preprocessing:** Videos split into frames and extracted features.
- **Training Model:** Video Swin Transformer.
- **Pretrained on:** Kinetics-400 dataset for human action recognition.
- **Accuracy:** (0.73%).

# FINAL MODEL: RNN(LSTM) MODEL

- **Dataset:** Videos recorded by team members.
- **Preprocessing:** Videos split into frames and extracted features using the Mediapipe model.
- **Training Model:** LSTM-RNN architecture with fixed input shape.
- **Accuracy:** Training accuracy: 0.81 and testing accuracy: 0.73
- **Model architecture:**
  - LSTM layer (64 units, return\_sequences=True, activation='relu', input\_shape=(30, 1662))
  - LSTM layer (128 units, return\_sequences=True, activation='relu')
  - LSTM layer (64 units, return\_sequences=False, activation='relu')
  - Dense layer (64 units, activation='relu')
  - Dense layer (32 units, activation='relu')
  - Dense layer (output units=actions.shape[0], activation='softmax')

# CHALLENGES

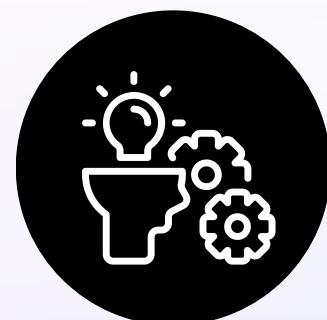
## DATA AVAILABILITY



The only available datasets for sign language were in image format.

Creating a custom dataset and refining the optimal data format for model training was time-consuming and required significant effort.

## DATA LIMITATION



Employing video classification yielded suboptimal model accuracy due to the data limitation video classification model provided poor accuracy.

## TIME LIMITATION



Despite time constraints, we acknowledge the insufficient dataset. We're actively exploring solutions to enhance data collection for future projects

# REAL LIFE APPLICATION



- In education, Saudi Sign Language Recognition (SSLR) can support PWHD students by providing real-time translation of sign language into written or spoken language.



- During emergency situations, SSLR can assist first responders in communicating with individuals who are PWHD



- SSLR enhances customer services, businesses by providing real-time translation for PWHD individuals.

# FUTURE WORK

- Improving accuracy through trying more complex models .
- Enhancing real-time performance.
- Collecting more data including all sign language vocabularies.
- Adding Voice-to-Text feature.
- Discussing potential integrations with emerging technologies or user feedback mechanisms for continuous improvement.(i.g. Tawakkalna, My Tawakkalna Health).





# TOOLS

Data collection and Pre-Processing	 <b>MediaPipe</b>	 python™	 <b>NumPy</b>	 <b>OpenCV</b>
Build model	 scikit <i>learn</i>	 <b>TensorFlow</b>	 <b>Keras</b>	
Workspace		 <b>ANACONDA.</b>	 <b>jupyter</b>	



# DEMO



# REFERENCES

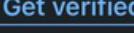
- AVAILABLE AT: [HTTPS://SSHI.SA/ . .](https://sshi.sa/) (NO DATE)  
(ACCESSED: 30 NOVEMBER 2023)
- MEDIPIPE | GOOGLE FOR DEVELOPERS (NO DATE) GOOGLE.  
AVAILABLE AT: [HTTPS://DEVELOPERS.GOOGLE.COM/MEDIPIPE](https://developers.google.com/mediapipe)  
(ACCESSED: 26 NOVEMBER 2023).
- VIDEO CLASSIFICATION WITH A CNN-RNN ARCHITECTURE: HUMAN ACTIVITY RECOGNITION (2022) YOUTUBE. AVAILABLE AT:  
[HTTPS://YOUTU.BE/EZJNYSXQDTO?SI=WFPR9MMFETKTRXJI](https://youtu.be/EZJNYSXQDTO?si=WFPR9MMFETKTRXJI) (ACCESSED: 28 NOVEMBER 2023).

# OUR TEAM

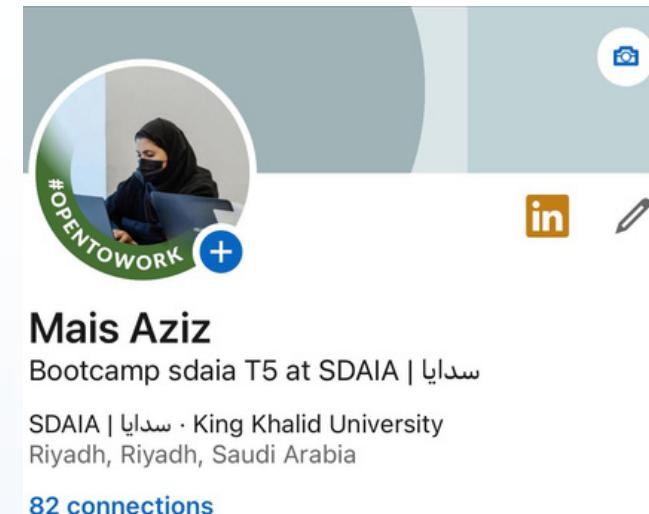
scan QR code to view Linkedin and GitHub

Dania Fallata



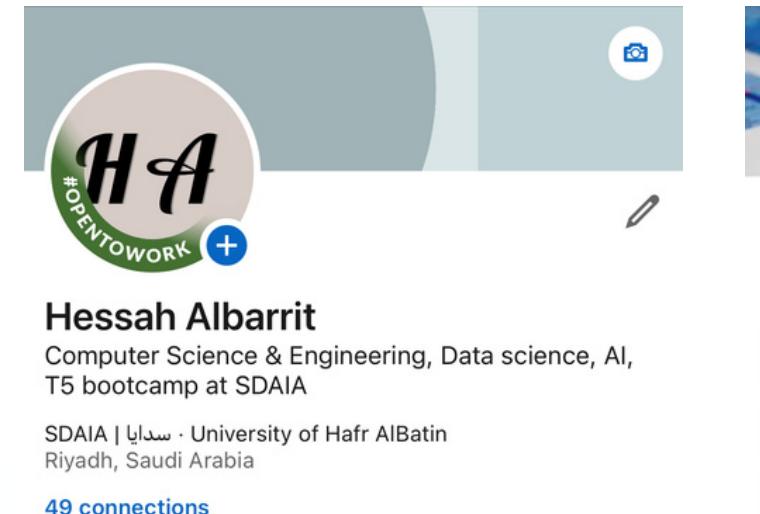
Dania Fallata  Data scientist  
SDAIA | سدايا · King AbdulAziz University Jiddah, Makkah, Saudi Arabia  
106 connections

Mais Alshahri



Mais Aziz  Bootcamp sdaia T5 at SDAIA | سدايا · King Khalid University Riyadh, Riyadh, Saudi Arabia  
82 connections

Hessah Albarrit



Hessah Albarrit  Computer Science & Engineering, Data science, AI, T5 bootcamp at SDAIA  
SDAIA | سدايا · University of Hafr AlBatin Riyadh, Saudi Arabia  
49 connections

Atheer Alzhrani



Atheer Alzhrani  Data Scientists | SQL | Excel | Python  
Al Baha University  
Riyadh, Saudi Arabia  
219 connections

Shahad Wadani



Shahad Wadani  Senior MIS | Data Scientist | AI | SDAIA T5 Bootcamp  
SDAIA | سدايا · Jazan University  
Riyadh, Saudi Arabia



**THANK YOU**