A portable home device offering virtual companionship and mental comfort through empathetic conversation, health reminders, emotional engagement, mood upliftment, positive energy boost, personal preferences, and emergency alerts to caregivers
for elderly and lonely people.

Introduction

Problem Statement

Loneliness, Lack of companionship and **feeling left out** lead to various types of mental issues and health risks like depression, dementia and even self-harm.

More than 42 million Americans are **socially isolated** and over 30% of children are **home alone**.

Recent survey of adults aged 50 to 80 found that "one in four people said they feel **isolated from other people** at least some of the time, and one in three say they don't have regular companionship."

Lack of timely mental support has generated a strong demand for virtual companion which can emphatically engage with people.

Inability to pay unwavering **attention** to lonely patients also necessitates the availability of a voice-assisted companion device.

Moreover, there is a **scarcity of voice-assisted devices** which can emotionally engage with people using cutting edge AI.

Strong Need for Mental Companionship !







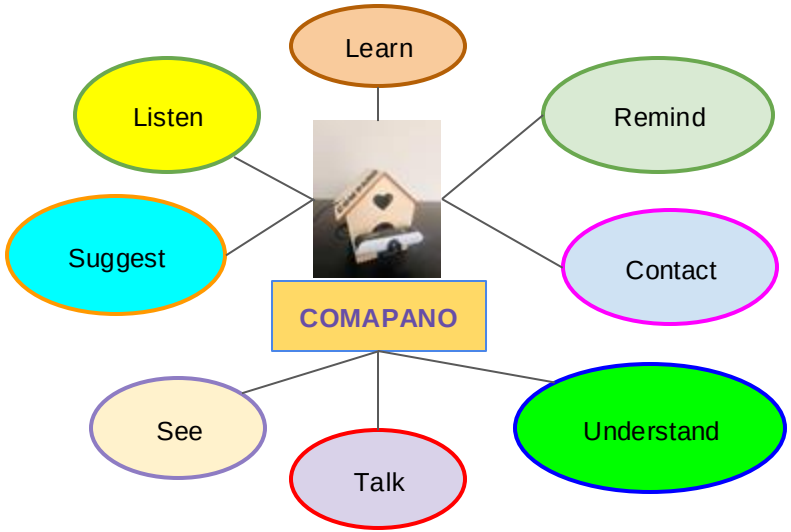
Introduction

Product Goals

- To address the pressing issues, an affordable device, was planned which can emphatically engage with a person and reduce risks of mental distress and chances of self-harm.
- It was intended to develop a software using image recognition, emotion detection and natural language processing algorithms running in a portable and user-friendly device fitted with video camera, speaker and microphone.

Product Capabilities

- Captures user facial expressions and speech
- Continues Natural Conversation
- Assists users with real-time reminders
- Help users find nearby facilities and medical centers
- Send an automated timely report to caregivers.
- Makes emergency calls to caregiver as per request by user.
- Monitors users in a friendly way and determines necessary help needed.
- Ability to start, pause and stop.
- Plays music, cracks jokes and shares uplifting quotes
- Suggests healthy lifestyle.
- Remind Medicines
- Analyze Postures



Methods

Product Engineering Design

Hardware Components

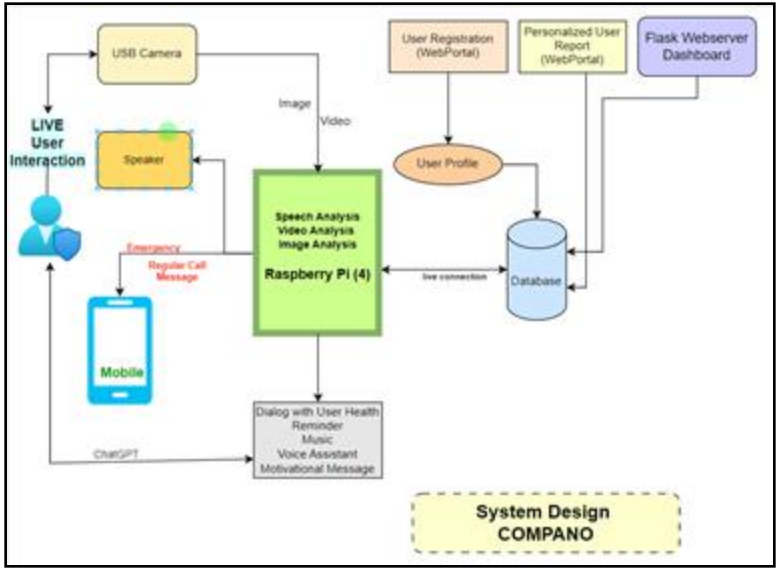
- 5V power supply
- Raspberry Pi 4
- Video Camera, Speaker, and Microphone
- Bluetooth connection on the Raspberry Pi

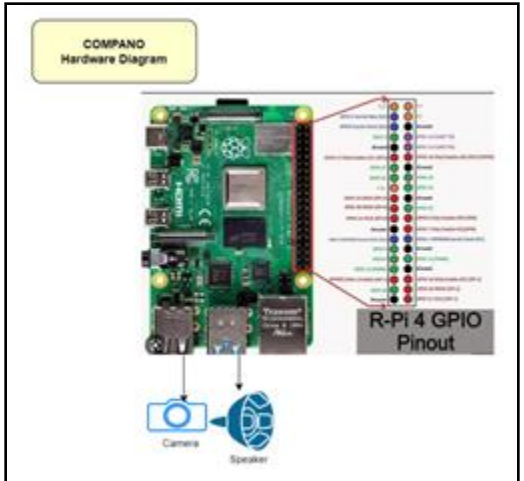
Software libraries

- Yolov5 and Yolov7 (image detection)
- Vgg (emotion detection)
- GPT-3 (speech analysis)
- Google Speech Recognition (speech to text)
- Twilio (send alerts)
- PyMongo (data management)

Integration

- Setup Raspberry Pi which is the main processing unit.
- Enable camera, speaker, and microphone for capturing video and audio.
- Yolov5, a state-of-the-art object detection and classification algorithm is used to detect faces in real-time video feeds.
- RepVGG is used to classify the emotions expressed by the detected faces.
- The AI libraries are integrated with hardware to have intelligent contextual conversation.
- GPT-3 is used to generate speech based on the detected emotion of the user.
- GTTS converts the text to speech and plays it through the speaker.



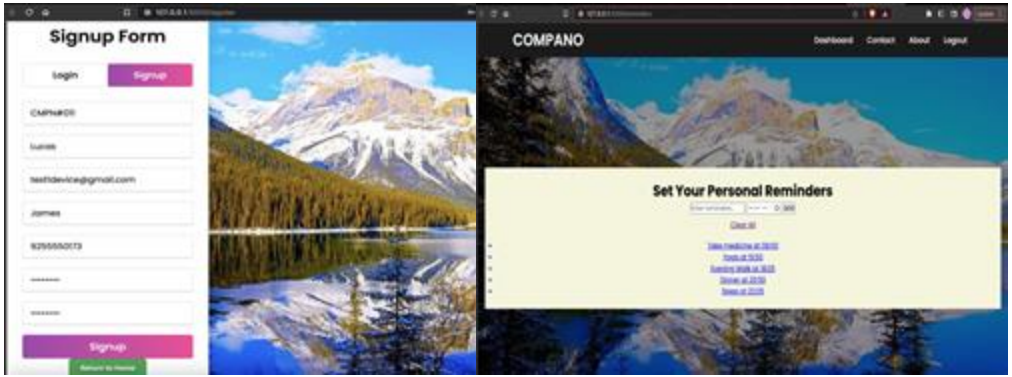


Methods

Input Data Collection & Processing

- Initially the user details are registered in database through the website
- The user details were attached to a specific device.
- The emergency contact details were captured in the user entry
- The Caregiver / User / Guardian can create and update the daily reminders for the user on the website
- The Raspberry Pi used the microphone to capture speech
- It used the video camera to capture the live facial expressions
- The device fetched the user name (associated with the Device) from the User-Device dataset.
- The Raspberry Pi used a combination of machine learning algorithms to analyze the data. It will use OpenCV, OpenPose, Pytorch, Text To Speech, Speech to Text, Yolov5 Deep Learning Framework, and NLP algorithm to analyze the data.
- GTTS, a Python library for text-to-speech conversion is used to converse with Users.
- The device uses GPT-3 model to generate speech based on the detected emotion of the user.

User Registration **Set Reminders**



Capture Images and detect emotions

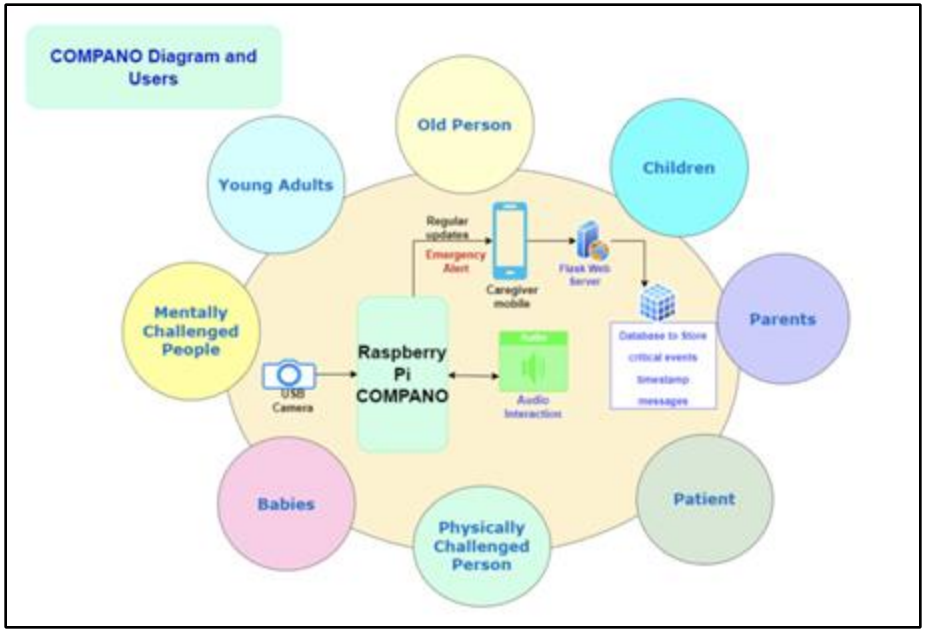
Generate unique response

Hello Lucas! Looks like something is bothering you.
You are an amazing person. Let's cheer up.
Would you like to listen to calm music?

Methods

Output Data Processing

- If the user provides permission, then the event will be logged.
- The code within the Raspberry Pi 4 detected speech and uses the speaker for talking to the user.
- The text spoken was uniquely generated by the GPT module.
- It also sent certain output messages in the form of an SMS as well as alerts to phone.
- The events were analyzed, and metrics were generated.
- Sample Output Events
    - REMINDER, PLAY_MUSIC,
    - TELL_JOKE,
    - UPLIFITING_QUOTES,
    - CALL_CONTACT,
    - SHOW EMPATHY,
    - START, STOP
- Device can intelligently process the user responses and take further actions through a continuous feedback loop.



Output Processing Example

Handle Events

Emotion

Play Music

Health Reminder

Call / Alert

Chat / laugh

Store & Analyze Events

Methods

Algorithm Testing

- Rigorous testing and research was done to select the right image detection Model and NLP Model.
- The Testing concluded that Yolov5 along with VGG Model (custom trained) work perfectly fine with Raspberry Pi microcomputer.

Functionality Testing

- Ensure User can register and login.
- Ensure Device can welcome the User correctly
- Ensure Device can detect the emotions and facial expressions correctly and greet accordingly
- Ensure Device can detect the speech correctly
- Ensure Device can continue the conversation empathetically and offer different ways to uplift the mood
- Ensure Device can send SMS message and call the Caregiver or Emergency Contact person
- Ensure Device can remind the important tasks
- Ensure Device can play calm music
- Ensure Device can capture the various events for behavioral analysis (only if user gives permission)
- Ensure Device operations can be stopped and resumed smoothly

Test Example 1: It welcomes the user and greets based on facial expression and time of the day.
USER: Appears in front of the device in a happy mood
Device:
Good Evening Lucas. Hope you're having a wonderful time.
I am glad to see you are doing well. Let me know if you need anything.

Test Example 2: It Detects an angry mood
USER: Appears in front of the device in angry mood:
Hello Lucas! It looks like something is bothering you.
You are an amazing person. Let's cheer up.
Would you like to listen to calm music?

Device: Together let's hear a few humorous jokes.
<mental uplifting quotes>

Results

- The Object Detection Algorithm accuracy was analyzed by running the Image processing Model (e.g. OpenCV, MobileSSDNet, Yolov5,) over multiple epochs in multiple batches using the given datasets
- The analysis was performed by capturing and verifying the metrics like latency, object coverage, and No. of epochs
- The images were analyzed by verifying the bounding box and predicted emotions.
- RepVGG and GPT-3 Model Performance was verified by published benchmarks.
- For example, The Emotion Detection Algorithm (RepVGG) accuracy was verified by AffectNet analysis results

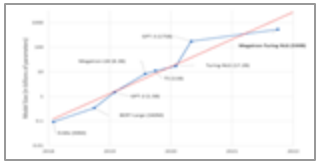
The Results confirmed that the prototype was able to meet the engineering goals of selecting and running an ensemble of image detection, emotion detection and NLP algorithms running in a portable and user-friendly device fitted with a video camera, speaker, and microphone.

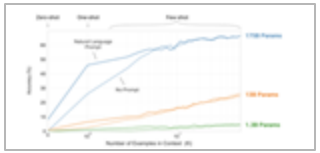
RepVGG Algorithm Test Result by AffectNet

Neutral	75374
Happy	134915
Sad	25959
Surprise	14590
Fear	6878
Disgust	4303
Anger	25382
Contempt	4250
None	33588
Uncertain	12145
Non-Face	82915
Total	420299

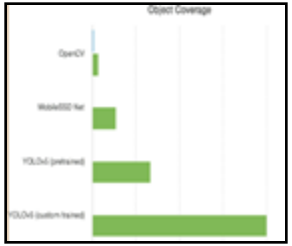
	Valence	Arousal
RMSE	0.37	0.41
CORR	0.66	0.54
SAGR	0.74	0.65
CCC	0.60	0.34

Performance of GPT-3 Model

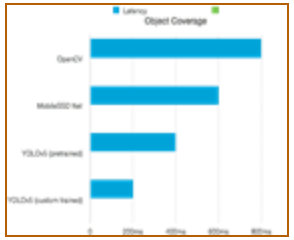




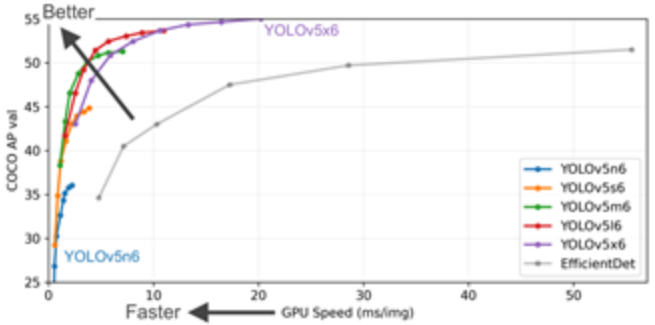
YoloV5 Object detection Results



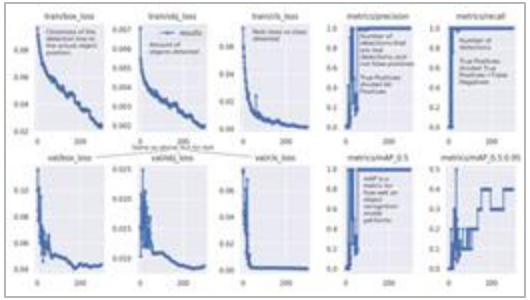
YoloV5 in terms of object detection compared to other models



time taken to detect an object by various algorithms.



Effectiveness of the specific customized model of Yolov5



Yolov5 Testing Results

Results

- As a result of the using the tuned emotion detector, the prediction accuracy improved to 95% - 98% from 90% within 200 ms latency.
- All the different interactions between the device and the person were analyzed by capturing various event details (device_id, user_id, event_name, timestamp) which can be used for behavioral analysis.

DeviceId	UserId	Event	Description	Timestamp
CMPN#123	Lucas	SAD	Device_Detected_Emotion	02-01-2023 0810 PST
CMPN#123	Lucas	HAPPY	Device_Detected_Emotion	02-01-2023 1110 PST
CMPN#123	Lucas	HAPPY	Device_Detected_Emotion	02-01-2023 1110 PST
CMPN#123	Lucas	REMINDER	Device_Reminded_Task	02-01-2023 0910 PST
CMPN#123	Lucas	PLAY_MUSIC	Device_Played_Music	02-02-2023 0910 PST
CMPN#123	Lucas	TELL_JOKE	Device_Told_Joke	02-03-2023 1210 PST
CMPN#123	Lucas	UPLIFITING_QUOTES	Device_Shared_Quotes	02-03-2023 0910 PST
CMPN#123	Lucas	NOT_FEELING_WELL	User_Emergency	02-04-2023 1410 PST
CMPN#123	Lucas	CALL_CONTACT	Device_Notified_Contact	02-04-2023 1510 PST
CMPN#123	Lucas	EMPATHY	Device_Empathizes	02-04-2023 1610 PST
CMPN#123	Lucas	HEALTH	Device_Healthy_Habits	02-04-2023 1710 PST
CMPN#123	Lucas	FEEDBACK	Device_Captures_Feedback	02-04-2023 1810 PST

Discussions

Key Challenges

Building a composite device which can simultaneously perform AI-driven emotion detection and natural conversations needs to address various types of challenges like

- selecting best model version which is compatible with hardware devices
- making multiple algorithms work together with low latency and high accuracy
- creating a feedback loop so that the Devices can continue the conversation without interruption.
- recorded video feeds also helped troubleshoot and solving key problems.

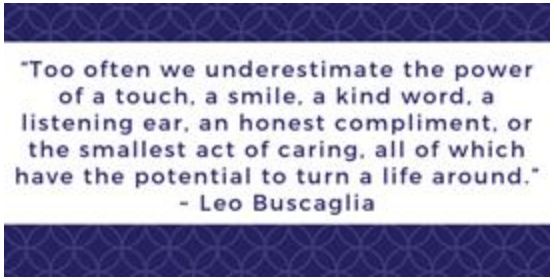
Key Observations

- The emotion and postures vary for different personas at different time of day. The initial observations gathered from the analysis of results and live interactions with Device is really encouraging as it really helped uplift the mood of the users as testing was conducted on different personas.
- Overall, the results reflect that Device achieved its engineering goal of successfully detecting emotions of people in various states and it generating unique responses for each scenario.
- It also opens up the possibilities of taking proactive life-saving measures by sharing
    - the anomalies observed (multiple times of angry moods, emergency calls) with Care Givers and
    - the qualitative data on various events with doctors for behavioural analysis and create advanced algorithmic models.

Conclusions

Key Benefits

- Device is best suited for anyone who spends most of their time at home alone and are in need of companionship and reducing mental distress.
- It helps older adults to fight social isolation and eliminate psychological insecurity.
- It is a Cost effective care companion for older adults and a great digital roommate.
- It is most suitable for people who stay Independent but still want to be connected to Family.
- A Perfect gift for Grandparents.
- Assistive technology for accessibility
- Proactive, voice-operated care designed to empower independence and support one taking control of your social, mental and physical wellbeing.



Future Work

- Device will provide additional capabilities like detecting sign languages, better speech analysis (e.g. detect coughing sounds) , assist personal healthcare (e.g. periodic reminder on latest prescribed) and posture detection for disabled patients.
- Device will provide more specific functionalities for certain Personas (like Patients, Autistic and Intellectually disadvantaged) and will be tuned to suit different types of Living Spaces (like Senior Living and Special Care)
- Device will provide a service for doctors to easily access behavioral and psychological information about the patient based on HIPAA compliance and patient's consent.
- Device will securely integrate with more devices and localized healthcare services
- Device will provide behavioral analysis insights with Doctors
- Device will continue to provide better psychological and mental assistance by interpreting different spoken languages, providing security by remembering user face.