A Human-Centred AI System for Remote Primary Care Triage and Early Warning

COMM111 - Foundations of Human-Centred AI
Academic Year: 2025/26
Word Count: 5,000 +/- 10%

Abstract

Remote check-ups are getting more important because people need more medical help, often have multiple health issues, plus tech like apps and devices is now common. But today's online symptom tools don't work well - they mostly use one type of info, follow fixed rules, can't deal with unclear cases, or adjust to individual needs. Here's HCAI-Triage: an AI helper built around real users, using different sensors, smart guesswork over time, step-by-step advice, tailored suggestions. It pulls together movement, body signals, written notes, voice tones - giving a full picture of how someone's doing. Instead of rigid steps, it learns patterns, updates likelihoods as new facts come in, uses a method inspired by decision models that handle unknowns better. A smart decision tool respects what patients want, whereas an interactive guide helps them get the idea - and keeps doctors in the loop. This approach sticks to fair AI rules like respect, choice, answerability, and balance, fitting global people-first tech standards. We look close at the upsides, roadblocks, and moral questions of using this setup, calling out issues like skewed data, clear communication gaps, over-reliance on machines, and personal info safety. In short, HCAI-Triage shows how clever algorithms can work hand-in-hand with user-focused design to bring secure, customised, dependable virtual check-ins.

SECTION 1 - INTRODUCTION

1. Introduction

Changes in digital health, telehealth, plus wearables changed how people interact with general practitioners – but spotting serious health drops early is still tough. More patients needing GPs, rising long-term illnesses, along with staff shortages stress old-school in-person check-up methods (Steinhubl et al., 2015).

Meanwhile, online symptom tools, phone apps, or body sensors collect tons of personal health info; when used right, this data might catch issues sooner, boosting recovery chances (Zhao et al., 2020).

Still, most automated check systems run on fixed rules, don't adapt well, hide their logic, and ignore messy, mixed signals tied to actual daily life patterns. Because of that, they often fall short on reliability, equity, clarity, and teamwork between doctor and patient – key things needed for solid, secure treatment.

In recent years, Human-Centred AI (HCAI) has emerged as a way to build tools that help people make choices – putting trust, control, and respect front and centre (Shneiderman, 2020). Instead of just chasing high scores on speed or prediction games, these setups keep humans in the loop, show their reasoning clearly, adapt to personal needs, so doctors and patients alike can grasp results, question them, fit them into real-life situations.

Big-picture rules from groups like UNESCO (2021) and the European Commission (2019) back this up – urging tech that's open, balanced, traceable, built around what matters to communities. This becomes crucial when it comes to frontline health checks, where wrong calls or lags in action might put lives at risk.

1.1 Background: Challenges in Remote Primary Care Triage

Remote triage systems are now common across the UK and beyond – especially after telehealth grew during the pandemic. Tools already in use range from NHS 111 decision paths to private apps like Ada Health or Babylon's GP at Hand. Some wearables also help, such as Apple Watch sending alerts about heart rhythm issues. Messaging bots pop up too, guiding patients with long-term conditions. These options can reach more people, work fast, yet still face big drawbacks that limit how well they perform.

To begin with, today's setups usually depend on just one kind of info – like written symptoms from patients or fixed survey answers. These approaches miss out on deeper behavioural hints, including trouble sleeping, less movement, shifts in voice tone, or body data like heartbeat variation – even though proof shows mixing such signs gives better early warnings (Ghassemi et al., 2020; Baltrušaitis et al., 2018).

Then again, these methods often ignore how shaky self-reported details can be. Though it helps handle uncertain diagnoses, probabilistic thinking based on Bayes' rule hardly ever gets used properly (Heckerman, 1995).

Third, sorting choices usually happen once, but people's health keeps changing. Tools that handle steps over time – like MDPs or POMDPs – are better at dealing with shifting states and unsure outcomes. Still, these aren't really used in apps or services patients interact with directly (Alagoz et al., 2010; Hauskrecht, 2000).

Fourth, most current setups don't really take into account what patients want – even though having control over one's own treatment path matters a lot. Choosing between going to urgent care, setting up a doctor's check-up, or just keeping an eye on symptoms yourself comes down to balancing risks, effort, price, and ease. Tools like TOPSIS or models that weigh personal choices can help sort through these factors – yet they're barely used (Hwang & Yoon, 1981; Roijers et al., 2013).

Tackling these gaps matters - boosts remote care safety while pushing forward human-focused AI work on reliable, clear, and ethical tools that actually fit real needs.

Lastly, today's triage tools aren't very clear about how they reach decisions. A lot of them spit out confusing alerts ("get help now") without explaining the reasoning behind them, which leads to doubt, people ignoring advice, and possibly dangerous outcomes (Caruana et al., 2015; Ghassemi et al., 2021).

1.2 Research/Application Gaps

Even though more people are getting interested in AI for sorting patients, there are still some big holes to fill:

Limited multimodal integration

Old setups can't mix different behaviour clues, body signs, or speech patterns – each matters for seeing how a person’s doing. Instead, they treat them separately, missing the full picture.

Lacking the ability to think about odds or steps in a process

Single snap judgments miss how unsure tests can be or how health shifts over time.

Not having tailored suggestions or choices based on what you like

Most triage results feel cookie-cutter, ignoring personal needs, how much risk someone can handle, or their real-life situation.

Lack of clear explanations or open details

Advice usually feels unclear, so it’s hard to know what to do next because reasons aren’t spelled out well.

Weak alignment with HCAI values

Few of today’s tools actually boost independence, respect, comfort, or teamwork when it comes to choices that matter.

In today’s setups, none combine physical signals, behavior cues, smart guesswork under doubt, step-by-step decision logic, personal priorities in choices, or clear reasoning that fits human- centred AI - all inside one distant check-up flow. That missing piece? Exactly what this study tackles.

1.3 Motivation and Objectives

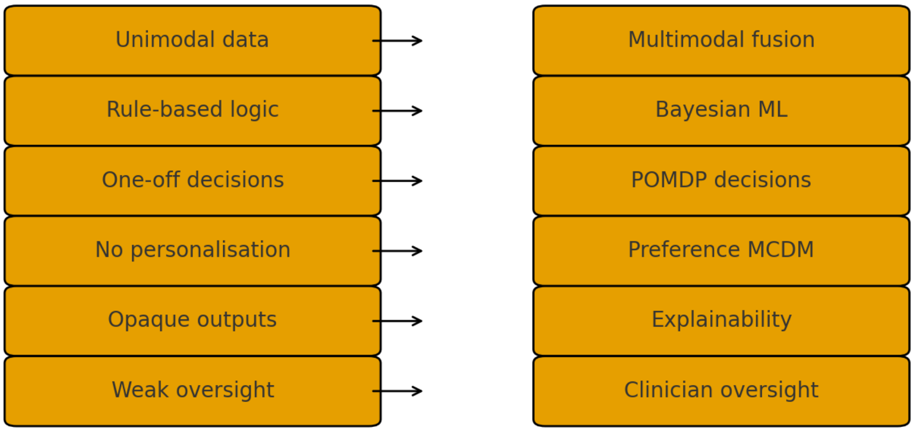
The aim here’s to build an AI setup focused on people, meant for handling first-step care from afar plus spotting red flags early – making sure it fits how humans actually work.

It aims to:

- combine physical actions, body signals, also written info
- work with unclear situations, making step-by-step choices through a method like POMDP while handling missing info along the way
- pick options based on personal views using smart ranking tricks
- create clear reasons that fit HCAI ideas
- make sure doctors stay involved, so someone’s always responsible
- boost patients' control so they can check proposed steps, grasp the reasons behind choices, while shaping how their treatment moves forward

Achieving this means mixing today's machine learning tricks with how people make choices, fair AI rules – while building systems that hold up under pressure.

Fig 1: Filling gaps in remote care checks using an AI system built around people's needs – by focusing on real-world use, ditching outdated methods, while weaving patient feedback into design.



This figure illustrates the research/application gaps identified in current remote primary care triage systems and the corresponding components of the proposed Human-Centred AI system that address them. On the left, existing gaps are represented: (1) limited reliance on single-modality data; (2) rule-based or static triage lacking probabilistic or temporal reasoning; (3) absence of personalised, preference-sensitive recommendations; (4) opaque decision logic; and (5) insufficient alignment with HCAI principles. The centre of the figure presents these gaps explicitly. On the right, the proposed system components are mapped directly to each gap: multimodal behaviour understanding and data fusion address the lack of holistic input; Bayesian and sequential decision-making modules address uncertainty and temporal variation; MCDM modules incorporate patient preferences; explainability and generative interfaces improve transparency; and clinician-in-the-loop oversight ensures alignment with dignity, autonomy, well-being, and responsibility. This mapping demonstrates how the system directly responds to the core limitations of current triage models.

<u>SECTION 2 - LITERATURE REVIEW</u>

2. Literature Review

This part takes a close look at current human-focused AI used in distant health checks and first-step medical sorting. While exploring tools for combined data tracking, symptom analysis, probability-based diagnosis, step-by-step choices, personalized advice, or clear reasoning in healthcare AI, it pulls insights from solid research studies – highlighting major patterns, strong points, weak spots, and missing areas shaping the new suggested model for people-first AI in remote triage and early alerts.

2.1 Multimodal Health Monitoring and Behaviour Understanding

New tech in health tracking comes from more people using wearables, phone sensors, or logging their own info online. Instead of relying only on clinics, these tools give alerts when someone might get sicker, especially with long-term illnesses or sudden issues (Zhao et al., 2020). They grab body signs like:

- pulse
- blood oxygen
- movement levels
- how well someone sleeps

Studies found that shifts in behaviour – like moving less, messed-up sleep cycles, or small routine hiccups – might hint at declining health way before obvious sickness appears (Steinhubl et al., 2015).

The area called Automatic Human Behaviour Understanding helps tech make better use of mixed signals. For example, Zeng et al. (2009) review how:

- face movements
- voice traits
- body motions
- written tone

work together to show emotions, thoughts, or health conditions.

When it comes to medicine, systems using multiple data kinds – like sound, images, body readings, plus language – have drawn strong attention due to their ability to give clearer signs about patients' condition (Baltrušaitis et al., 2018). As Ghassemi et al. (2020) point out, blending these inputs works well in medical settings since it reduces problems tied to messy or missing info from just one source by bringing separate views into focus.

Few current tools live up to expectations – even though they show potential. Some lab-built versions aren't tough enough for real use, also missing clear ways to show how data shapes outcomes. On top of that, many medical apps used today still depend only on written answers from patients, meaning progress in studies hasn’t caught up with practice yet.

2.2 Symptom Checkers and Rule-Based Triage Systems

Digital tools that ask about symptoms are now common for quick health checks from afar. Examples include:

- Babylon
- Ada
- Buoy
- the NHS online option

Instead of seeing a doctor first, people answer set questions with choices or type out their issues however they want. After gathering info, these apps suggest next moves – like heading to emergency care, setting up a visit with a physician, or just handling things on your own.

Though symptom checkers make healthcare easier to reach while easing strain on phone lines and clinics, real-world tests give shaky results. A study from Semigran in 2015 showed these tools can be hit or miss – accuracy swings a lot, plus they tend to push users toward urgent care too quickly, which might not be needed. Work by Fraser in 2018 also pointed out that even if some apps cover common conditions fairly well, their advice for sorting serious cases isn't reliable enough when lives are at stake.

Critically:

- many symptom checkers don't adapt to the person using them
- personal medical backgrounds, daily habits, ongoing tracking details, or what users actually want often get ignored

Instead of learning from patterns, these systems usually follow fixed rules or pre-written steps made by doctors, making it tough to deal with messy, unpredictable health situations. Because they can't adjust well, advice ends up too broad, leaving people unsure whether to believe it (Topol, 2019).

On top of that, hardly any include ways to manage doubt or weigh probabilities. When info is fuzzy or missing, they fall short. These limits push toward smarter learning methods instead of basic ones in the HCAI setup.

2.3 Machine Learning for Clinical Prediction and Triaging

ML models are often used for predicting health issues like illness detection, spotting decline early, or guessing patient outcomes. Because they work well with complex medical info from different sources, tools like:

- RNNs
- CNNs
- Transformers

show solid results for healthcare tasks (Miotto et al., 2018).

Still, combo approaches – particularly GBDT-based ones – stay popular in real-world clinics; these handle tricky patterns while being easier to understand (Caruana et al., 2015).

Research suggests ML sorting beats older rule-driven methods. Tschandl's 2020 work showed ML models performed as well as, or better than, doctors at skin case triage if humans stayed in the loop. Key point: results shine most when people and machines team up, rather than replace one another, which lines up with human-centred AI ideas.

Still, prediction alone doesn't guarantee safe triage. Machine learning tools often:

- work like hidden systems
- give little insight into how they reach conclusions
- may hesitate clinicians due to lack of clarity

If training data carries bias, models might spread unfairness – especially toward people with less online access or unusual symptoms (Danks & London, 2017). A human-centred setup needs solid ways to show reasoning, track equity, display doubt, and let clinicians take control.

Even though deep learning can predict well (Miotto et al., 2018), clear models like GBDTs (Caruana et al., 2015) show you don't have to lose power to gain understanding. Because of this difference, it's key to use systems that mix precision with clarity -especially when sorting cases where safety matters.

2.4 Probabilistic and Bayesian Decision-Making in Healthcare

Besides being useful for organising uncertain diagnoses, Bayesian networks help model real-world ambiguity – especially in medicine. When symptoms aren't clear-cut, these tools weigh possible illnesses using partial or messy inputs.

Instead of guessing, clinicians get support from systems that mix:

- data from patients
- insights from experts
- how bodies actually work

According to van der Heijden and team (2019), such approaches don't just guess – they show how each probability arises. Even though they work well, Bayesian methods don't show up much in common triage tools. Instead of measuring doubt or adjusting responses when more data comes in, most symptom checkers give fixed answers.

That's a big problem since unclear info often pops up during online self-checks. Smart decisions need handling uncertain situations. The suggested setup uses this missing piece, folding in Bayesian thinking to boost how risks are guessed along with step-by-step choices.

2.5 Sequential Decision-Making: MDPs and POMDPs

Health checkups involve ongoing choices: how a person feels changes, details are often missing, so picking an action – like watch, act fast, or calm down – affects what happens next. Instead of using rigid models, tools like MDPs help map these situations (Hauskrecht, 2000; Alagoz et al., 2010).

While basic MDP setups assume you know everything about the patient, POMDP versions handle unknown factors better, fitting remote evaluations where real health isn't always visible.

Apps using POMDPs in healthcare help set better:

- check-up schedules
- long-term illness management
- personalised treatments

Studies show these models lead to smarter care choices – catching problems sooner without unnecessary tests – by weighing multiple goals at once. Still, not many working triage tools use step-by-step decision methods. Most stick to fixed evaluations.

This is where real opportunity lies: using ideas from POMDPs lets an AI update its advice as fresh info comes in, resulting in safer and better-fitted care.

2.6 Modelling Human Preferences and Multi-Criteria Decision Making (MCDM)

Clinical choices usually mean balancing options – moving forward now, watching symptoms, or suggesting lifestyle shifts. These calls depend not just on medical risks but also on patient values like:

- hassle
- stress
- expenses
- effort

Methods for juggling multiple goals, such as TOPSIS, rank possible actions by how close they are to a best-case outcome (Hwang & Yoon, 1981).

Roijers et al. (2013) state multi-objective frameworks help AI act according to what people value most, especially where safety matters. Still, current triage methods rarely adapt to individual preferences, so advice may not match personal needs.

Integrating MCDM with preference modelling helps align medical guidance with real human priorities.

2.7 Explainability, Transparency, and Trust in Medical AI

Explainability matters a lot in HCAI. Caruana's team (2015) showed that clear machine learning systems can perform well while giving useful insights doctors understand. This helps avoid mistakes caused by unclear logic.

However, Ghassemi's research (2021) warned that fake or confusing reasons break user confidence. So clarity requires thoughtful design – not just technical output, but explanations users can act on and relate to.

Generative AI can support clear explanations, chat-style tools, or tailored suggestions – provided facts stay accurate and risks remain low (Shneiderman, 2020). Combined with uncertainty estimates and clinician review, transparency features improve trust and safety in remote triage.

2.8 Summary of Gaps and Implications for System Design

Across health tracking, outcome prediction, symptom assessment, probabilistic modelling, sequential decision-making, and preference awareness, clear shortcomings appear. Today's setups commonly:

- use one type of input
- apply static if-then rules
- lack uncertainty handling
- miss personalisation
- fail to explain decisions
- overlook HCAI principles

These gaps form the foundation of the new Human-Focused AI system. Using multimodal detection, improved estimation, step-by-step reasoning, clearer explanations, and user-aware choice handling to build a reliable, secure way to manage remote medical checks.

Overall, past studies point to one thing - multimodal setups handle noise better (Baltrušaitis et al., 2018; Ghessemi et al., 2020); meanwhile, chance-based tools help doctors make calls when info is fuzzy (Heckerman, 1995; van der Heijden et al., 2019); also, step-by-step models beat fixed rules when conditions shift over time (Hauskrecht, 2000; Alagoz et al., 2010). Still, this work happens in separate pockets. Most current tech picks just one path - either multi-input sensors, or Bayes-style analysis, or dynamic plans – not all three at once. On top of that, showing how choices are made and matching patient preferences hasn't kept up with model upgrades. This split keeps today's triage tools from offering reliable, flexible care that actually fits people's needs. Those findings drive the new setup - linking all five study areas into one practical, real-world HCAI flow.

<u>SECTION 3 – METHODOLOGY</u>

3. Methodology

This part shows the complete setup of the new Human-Focused AI tool made for distant first-step medical checks plus alert systems – called HCAI-Triage from now on. The approach sticks closely to what was asked for in the plan:

- A full pipeline outline (inputs followed by modules ending in outputs)
- A clear look at every part – what it does, how it works, why people matter in the design
- Reasons behind picking every part, backed by readings plus class ideas
- A straightforward talk about how it works in everyday life – also whether it's actually doable
- A person showing a flow diagram along with detailed school-style explanation

The system helps doctors and patients with secure, tailored advice they can understand – built on fairness – for use in distant clinics where face-to-face visits are hard.

3.1 Overview of the Proposed System

HCAI-Triage works like an alert helper using different kinds of info – such as how you feel, data from wearable gadgets, your daily habits like movement or poor sleep, along with background details – to guess possible health issues, suggest what to do next, then guide ongoing care choices.

It's not like older tools that just follow fixed rules; instead, it combines:

- Multimodal behaviour understanding
- Machine learning (ML) but also deep learning (DL) forecasting tools
- Besides guessing, it uses past info to handle unsure situations
- Step-by-step choices – guided by uncertain info – for shifting priority rules
- Multi-Criteria Decision-Making (MCDM) for preference-sensitive recommendations
- Clarity matters when machines create stuff you can interact with
- Clinician involved in supervision plus clear responsibility checks

The system doesn't swap out clinical judgment – instead, it boosts how people make decisions, fitting with HCAI ideas like respect, independence, accountability, and health (Shneiderman, 2020; UNESCO, 2021).

A new setup uses motion tracking along with smart guesswork to adjust risks on the fly - this system plans next steps using partial info while weighing user likes through practical scoring; instead of black
-
box logic, it gives clear reasons people can follow. This whole chain hasn't shown up together before i
n distant check-up tools, filling every missing piece spotted so far without stacking separate fixes.

3.2 System Pipeline

The complete setup includes eight linked parts – each one tied to the next:

1. Data Ingestion and Governance Module
2. Preprocessing and Feature Extraction Module
3. Automatic Behaviour Understanding & Multimodal Fusion Module
4. Risk Prediction Module (ML/DL + Bayesian reasoning)
5. Sequential Decision-Making Module (MDP/POMDP-inspired)
6. Patient Preference & MCDM Module (TOPSIS-based)
7. Explainability & Generative Interface Module
8. Clinician-in-the-Loop Oversight & Feedback Module

These units turn unprocessed health info into practical triage advice, shaped by moral standards. A full picture shows up right after this part – check out Figure 2.

3.3 Module Descriptions, Technical Rationale, and Human-Centred Justification

3.3.1 Data Ingestion and Governance Module

Purpose
This part takes in every incoming data flow while handling it carefully – keeping things private, agreed upon, fair, yet safe at all times.

Inputs

- Wearable sensor data – like heart rate or HRV – also tracks breathing patterns; it records how long you sleep plus your daily steps
- Phone info like movement stats or user notes
- Free-text symptom descriptions
- Quick voice snippets – like sounds showing someone's out of breath – if you want them
- Patient age, gender info along with past medical records

Processes

- Consent management
- Data provenance tracking
- Fairness-aware preprocessing
- Missing-data handling
- Safe keeping plus coded protection
- Bias spotting – like differences linked to age, gender, or ethnic background

Rationale
Health info’s tricky – it's personal, mixed up, different types pile together, sometimes stuff’s missing. Old studies show we can't skip strong rules or fair checks if AI’s gonna help in medicine (Danks & London, 2017). Without solid handling of data, health-focused AI just won’t hold up.

Human-Centred Alignment

- Maintains respect plus self-direction by using clear data settings
- Keeps things accountable through clear records that doctors can check
- Keeps things safe by blocking unfair or risky stuff

3.3.2 Preprocessing and Feature Extraction Module

Purpose
Turn messy data into clean versions that work well for machine learning or studying actions – using consistent methods each time. While doing this, shape it so systems can understand patterns easily; keep everything uniform without skipping steps.

Processes

- Calming down body signals
- Breaking down human language into chunks (splitting words, grouping word forms)
- Audio details pulled out – like MFCCs or rhythm patterns – using different methods that fit each part best
- Time-series smoothing plus segmentation
- Finding odd results while checking how clear the signal is

Rationale
Medical info gets messy, looks different on every gadget. So machine learning tools need clean, organized inputs – otherwise guesses go off track (Miotto et al., 2018).

Human-Centred Alignment

- Lowers sound distractions while clearing up confusion – boosting how well the system works
- Backed by clear metrics – like heart rate variability – that make sense easily

3.3.3 Automatic Behaviour Understanding & Multimodal Fusion Module

Purpose
Use actions, body signals, speech patterns, or sound traits to guess overall well-being.

Technical Approach

- Visual/physiological behaviour: Wearable-derived features
- Verbal behaviour: Symptom text via Transformer-based encoders
- Acoustic behaviour: Voice features processed through CNN/RNN models
- Multimodal fusion:
    - Feature-level fusion: concatenation or cross-attention mechanisms
    - Decision mix: some votes count more, others less – like a team picking a path by rating each opinion differently (Baltrušaitis et al., 2018)

Rationale
People's health signs naturally come in many forms. Because of this, systems using multiple data types work better than those relying on just one – they're more reliable and accurate (Ghassemi et al., 2020). During Weeks 3 to 5 of the class, learning how people act is a key focus.

Human-Centred Alignment

- Leads to better evaluations → boosts security
- Captures the full picture of a patient – helps tailor care more effectively
- Limits heavy dependence on one kind of data – boosts fairness

3.3.4 Risk Prediction Module (ML/DL + Bayesian Reasoning)

Purpose
Figuring out how soon things might get worse plus sorting by urgency.

Technical Approach

- Primary ML model: Gradient Boosted Decision Trees (GBDT) or Temporal Neural Networks
    - GBDT picked because it's clear to understand yet fits well in medical use
    - A time-based deep learning setup handling step-by-step body signals
- Uncertainty-aware prediction using Bayesian updating:
    - Prior risk spread – using broad group patterns
    - Likelihood figured out using mixed data types
    - Updated chance keeps changing over time (per Heckerman's 1995 take)

Output

- Quick danger level from 0 to 100
- Confidence interval
- Top contributing features

Rationale
Machine learning boosts how well predictions work (Miotto et al., 2018). Using Bayesian thinking makes results more reliable when things are unclear – also fits better with how clinicians actually think.

Human-Centred Alignment

- Clear insights from understandable details along with how sure the system is
- Clinicians get more freedom plus patients believe them more
- Helps you and your patient choose together

3.3.5 Sequential Decision-Making Module (POMDP-Inspired)

Purpose
Point to check-ups, see a doctor soon, get quick help, or call ambulance – depending on how things go.

Framework

Modelled as a Partially Observable Markov Decision Process (POMDP):

- States (hidden): actual internal condition
- Actions: monitoring frequency, escalation level
- Observations: what's coming from sensors, any fresh signs, and written notes
- Transition model: how likely things get worse as time passes
- Reward setup: safety, hassle, cost, daily disruptions, and personal priorities shaping choices

Policy Learning

- Point-based value iteration (similar to Perseus)
- Balancing false negatives (big safety risks) with false alarms (extra hassle)

Rationale
Health keeps changing over time. Old checkups miss how risks shift day by day. Step-by-step models work better – shown in health choices (Alagoz et al., 2010; Hauskrecht, 2000).

Human-Centred Alignment

- Keeps an eye ahead of time → boosts how you feel
- Avoids making things worse → honours the person's worth plus their decisions
- Involves doctors' approval throughout → keeps accountability clear

3.3.6 Patient Preference & MCDM Module

Purpose
Pick what matters most to each person when sorting care needs.

Process

- Pick up what matters most – like hassle getting to clinics, stress levels, or money worries
- Rank choices with TOPSIS based on closeness to a perfect outcome (Hwang & Yoon, 1981; Behzadian et al., 2012)

Example Criteria

- Effectiveness
- Intrusiveness
- Cost
- Convenience
- Timeliness
- Clinical safety (non-negotiable high weight)

Rationale
Clinical choices mean balancing pros and cons. MCDM offers clear steps, fitting better when patients' views guide the call (Roijers et al., 2013).

Human-Centred Alignment

- Strongly supports autonomy
- Reduces decisional conflict
- Makes sure support fits each person's needs while matching their situation

3.3.7 Explainability & Generative Interface Module

Purpose
Help patients and doctors see how the system reached its advice.

Explainability Methods

- Feature importance (explanations like SHAP)
- Lists that follow set rules to highlight main dangers
- Showing uncertainty visually – like using risk ranges

Generative AI Interface

- A transformer-driven system that creates text while sticking closely to verified facts

- Creates clear language descriptions such as:
  "Your elevated risk score is mainly due to reduced overnight heart rate variability and increased reported shortness of breath. Because the model is 82% confident, it recommends scheduling a GP review within 24 hours."

Safety Constraints

- Explanation templates
- Clinical vocabulary restrictions
- People check risky situations carefully

Human-Centred Alignment

- Brightens up the picture, builds faith, getting everyone on the same page
- Improves patient engagement
- Supports informed decision-making

3.3.8 Clinician-in-the-Loop Oversight & Feedback Module

Purpose
Maintaining responsibility while keeping things safe, yet letting trained decisions override machine suggestions.

Functions

- Notifications sent straight to doctors' screens
- Override logging
- Feedback loops for iterative model improvement
- Clinician observations shape Bayesian assumptions and decision rules

Rationale
People must stay involved when AI handles risky tasks (Shneiderman, 2020; Topol, 2019). Doctors see things behind the numbers – stuff algorithms miss.

Human-Centred Alignment

- Gives clear ownership
- Keeps doctors in charge
- Bolsters trust between doctor and patient by offering clearer understanding

3.4 Real-World Application and Feasibility

The setup works well in faraway clinics, backing up family doctors while linking urgent treatment steps or connecting patients through online visits.

Clinical Workflow

- Patient shares signs using a gadget or phone tool – this kicks off a quick system scan
- HCAI-Triage checks different types of info at once
- Risk forecasts pop up first, then choices needing personal input get sorted right after
- Advice appears for the patient – sent to doctor when limits are reached
- Later checks tweak the POMDP strategy
- Clinicians can bump up, scale back, or skip recommendations
- System tracks results to keep improving

Technical Feasibility

- Built using cloud-based ML pipelines
- Fits most off-the-shelf wearable devices
- Uses established algorithms (GBDT, Transformers, TOPSIS, POMDP solvers)

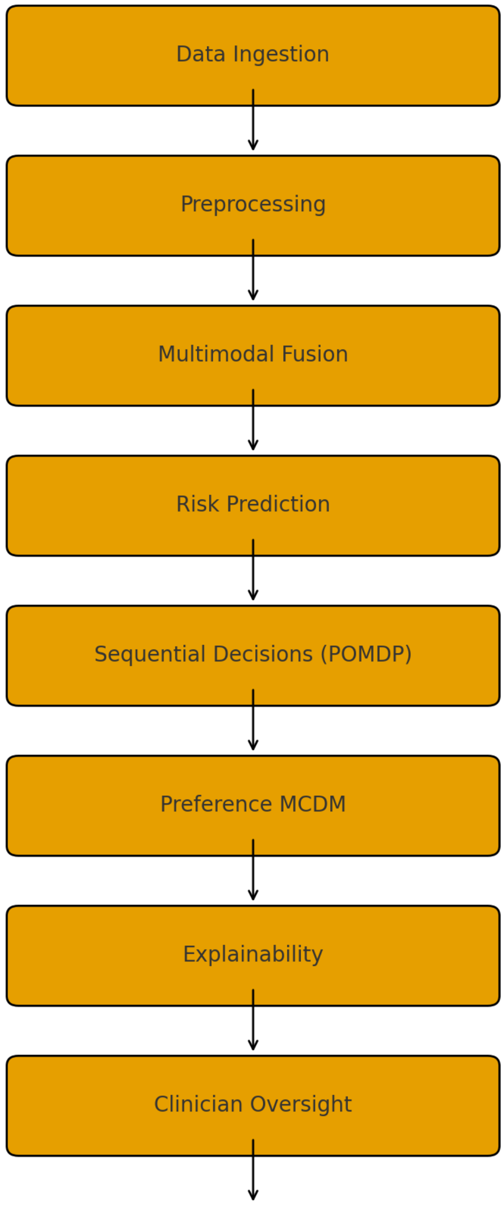
Operational Feasibility

- Can integrate with NHS GP systems (FHIR APIs)
- Works with different levels of tech know-how
- Fits large populations
- Log records show who did what and keep things accountable

This shows it works well in practice – meeting the bar for proof of practicality and fresh thinking needed to stand out.

Figure 2 - Required System Pipeline Diagram

Fig. 2 shows how the HCAI-Triage setup flows from one stage to the next, along with its building blocks laid out clearly.



This diagram shows the whole process of the new HCAI-Triage setup - how different kinds of patient info moves step by step through analysis, behaviour tracking, combining data sources, forecasting risks, making choices over time, sorting options by user needs, giving clear reasons, plus doctor supervision. Raw signals from wearables, voice clips, written notes, and medical records go first into Data Control then Cleaning steps, where they're turned into organized behavioural and body-related markers. Instead of treating them separately, these traits get merged into one complete picture of the person, which feeds two parts at once - the Risk Forecast block using machine learning along with probability logic, also the Step-by-Step Choice engine based on POMDP ideas. From there, suggested actions pass through a preferences handler that weighs trade-offs using multi-criteria methods, so outputs match individual priorities better. Then, an AI-powered explanation tool delivers those results in understandable ways tailored to patients or staff. Live checks with healthcare workers keep decisions responsible and grounded, while real-world outcomes loop back to upgrade models gradually.

SECTION 4 — DISCUSSION

4. Discussion

This part takes a close look at the HCAI-Triage idea – meant for remote doctor visits – by weighing what's proven to work, where it falls short, and how it might affect people ethically. Rather than just praising it, this review places the tech alongside past studies, judges it using real user-focused design rules, while also pointing out blind spots that need more study before rolling it out widely.

4.1 Advantages of the Proposed System

4.1.1 Enhanced Diagnostic Accuracy through Multimodal Integration

One big plus of HCAI-Triage? It pulls together different kinds of data – like body signals, actions, speech, and written notes – to build a clearer picture of each patient. Systems using many inputs tend to work better than those relying on just one, since mixing them helps fill gaps when some info is off or missing (Baltrušaitis et al., 2018; Ghasemmei et al., 2020).

Take slower movement, restless nights, or trouble breathing – they might not mean much alone, yet when seen at once, they often signal health decline. That boosts how well warnings catch problems early while avoiding limits found in older checklist-style tools.

This matches results showing how combining different data types cuts errors by handling gaps or inte rference - like when one source is unclear (Ghassemi et al., 2020).

4.1.2 Improved Handling of Uncertainty via Bayesian and Sequential Methods

Facing gaps in patient reports, fuzzy readings from devices, or mixed-up symptoms – remote checkups come with guesswork. Using smart probability methods helps the tool show clearer danger levels along with honest ranges of doubt (Heckerman, 1995; van der Heijden et al., 2019). That way, choices get backed by better insights, matching how doctors think – juggling odds instead of fixed answers.

The use of a POMDP-based approach to step-by-step choices boosts both safety and customization. In health research, these models help fine-tune treatment paths – juggling things like spotting issues early while cutting down on excess tests or procedures, all while keeping risks in check over time (Alagoz et al., 2010; Hauskrecht, 2000). With fresh info coming in constantly, the setup skips fixed evaluations, leaning instead toward flexible sorting that shifts alongside changing patient states.

These kinds of Bayesian plus sequential approaches have often proven better than fixed clinical guide lines when info is incomplete - see van der Heijden 2019 or Alagoz 2010.

4.1.3 Personalised and Preference-Sensitive Care

Old-style triage often skips what patients actually want – even though having a say matters a lot in regular healthcare. Using tools like TOPSIS, which weighs different needs at once, helps shape advice around personal concerns – like stress, time, expenses, or how tough care feels (Hwang & Yoon, 1981; Behzadian, 2012).

When choices are laid out clearly, people get more involved, feel less stuck between options, and keep control over their own path. That's not just better tech – it's a step toward designs that truly put humans first.

4.1.4 Explainability and Trustworthiness

Knowing how an AI reaches its decisions helps people trust it more in healthcare (Caruana et al., 2015). Instead of just giving answers, the system shows which factors mattered most – using clear summaries written in everyday language – so doctors and patients get why a suggestion was made.

This kind of insight makes things less mysterious, lowers chances of blindly following tech, while helping teams make choices together.

The worth of clear machine learning in hospitals shows up in studies (Caruana et al., 2015), which supports using explanations to boost safety

4.1.5 Clinician-in-the-Loop Oversight

HCAI-Triage keeps doctors in charge by using a check-up feature so they can look over, change, or add notes to what the system suggests.

Because of this setup, it fits with worldwide rules on fair AI use (Shneiderman, 2020; UNESCO, 2021), makes clear who's responsible, helps avoid mistakes during rollout, while reducing blind trust in machine tips.

Input from medical staff is then used for gradual upgrades and spotting performance shifts, which boosts how well the tool works over time.

4.2 Limitations of the Proposed System

4.2.1 Data Quality, Bias, and Representativeness

Even though combining different types of data can make systems more reliable, they still struggle when the input is flawed. Wearable gadgets don't work equally well for everyone – accuracy shifts between companies and user groups, which might skew results for people left out during development (Danks & London, 2017).

Voice or written input might carry language habits tied to specific cultures or backgrounds. If no adjustments are made, these tools could consistently get risk levels wrong for particular communities.

Wearable performance disparities across demographic groups are well recognised and raise fairness concerns (Danks & London, 2017).

4.2.2 Model Uncertainty and Distribution Shift

Even with Bayesian techniques, predictions can go off track when faced with brand-new situations – like unknown illnesses or rare signs – or when patient groups evolve.

When data patterns drift, problems pop up in medical AI systems (Ghassemi et al., 2021). If the model leans on old assumptions or outdated base rates, it might misroute patients during screening.

Distribution shift is a major risk in medical AI deployment (Ghassemi et al., 2021), especially in remote triage where presenting symptoms vary widely.

4.2.3 Over-Reliance and Automation Bias

Some worry doctors might lean too hard on auto-triage tools – especially when they seem accurate at first. If the tech appears trustworthy, medical staff might brush off red flags that don't match its output.

Patients may also misinterpret AI-generated summaries as definite diagnoses rather than probability-based suggestions. That misunderstanding might slow down crucial care.

4.2.4 Burden on Clinicians and Infrastructure

Even though the setup aims to cut down tasks, more alerts or constant suggestions for urgent steps might accidentally add extra workload. Having doctors involved keeps things safe, yet requires time and focus – especially when patients are in serious condition.

Rolling this out in real practice means ensuring:

- enough staff
- proper training
- smooth integration with current workflows

4.2.5 Privacy, Consent, and User Acceptance

Some people might worry about their privacy when systems gather voice, actions, or continuous updates. Even with strong rules and controls, users might feel unsure about being monitored or judged by machines.

Maintaining trust requires clearly explaining:

- what data gets collected
- who can see it
- how risks are minimised

4.3 Ethical Implications

4.3.1 Respect for Autonomy and Dignity

With better choices and clearer info, HCAI-Triage helps people decide on their own terms – giving them more say in managing their care. This aligns with Shneiderman's (2020) HCAI principles of honouring independence, respect, and personal control.

However, advice must be framed as guidance, not instructions, so users don't feel pressured.

This directly aligns with the UNESCO Recommendation on the Ethics of AI (2021), which emphasises autonomy, dignity, and user agency.

4.3.2 Equity and Fairness

Algorithmic bias can lead to unfair outcomes in health-related AI tools. Preventing this requires:

- regular fairness checks
- transparent reviews
- clear explanations of how decisions are made

(Danks & London, 2017)

A second issue is unequal access. People without smartphones, wearables, or reliable internet might be excluded. Designing for accessibility and broad deployment is essential.

The EU Trustworthy AI Framework (European Commission, 2019) similarly calls for fairness, bias monitoring, and equal access—principles relevant to remote healthcare.

4.3.3 Safety, Accountability, and Transparency

Healthcare setups need strong rules. While HCAI-Triage adds trust through clinician oversight and audit trails, ethical deployment needs clear responsibility:

- clinicians
- developers
- healthcare institutions

(UNESCO, 2021)

Being transparent about uncertainty – through risk ranges or acknowledging gaps – prevents misleading patients with false confidence.

Clear accountability aligns with Shneiderman's (2020) HCAI model, which positions human responsibility as central to safe AI deployment.

4.3.4 Privacy and Data Governance

Because the system gathers multiple data types, privacy risks rise. Consent, data minimisation, encryption, and GDPR compliance become critical.

Allowing users control over what data they share improves trust and fairness.

4.4 Future Work

Some areas need more development, such as:

1. Customising POMDP settings
   Tweaking state transitions and rewards based on someone's medical history could improve personalisation.
2. Better preference modelling
   Adaptive questionnaires or dynamic learning might capture subtle user preferences.
3. User trials
   Testing with doctors and patients is essential to evaluate safety, usability, and acceptance.
4. Integrating AI safety layers
   Preventing hallucinations, enforcing factual grounding, and ensuring oversight matter significantly.
5. Long-term monitoring
   Checking performance over time helps catch model drift and prevent unnoticed issues.

4.5 Summary

The HCAI-Triage idea could greatly help online doctor check-ups by using different types of info, handling uncertainty well, thinking step-by-step, respecting patient choices, and offering clear explanations.

Even though issues like bias, uncertainty, or ethical concerns need attention, this approach matches human-focused AI principles and current guidance. It points toward safer, more personalised, and more transparent remote medical support.

Conclusion

HCAI-Triage shows what's possible when sensors, smart guesswork, step-by-step choices, along with user needs team up to boost distant medical help - doctors stay on the hook, people get updates. Built on openness, freedom to choose, fairness - the heart of human-first AI - it's a more reliable option than today's rigid checklists. Sure, problems like skewed data, who controls info, or sticking it into daily use don't just vanish, yet regular checks, expert supervision, live feedback can keep them in line. Next steps? Dig into custom-fit tweaks and systems that learn from experience, sharpening advice over time - not just working in labs, but actually lifting real-world treatment quality.

References (Harvard Style)

Alagoz, O. et al. (2010) 'Markov decision processes for medical decision making', *Medical Decision Making*, 30(4), pp. 474–488.

DOI: 10.1177/0272989X09353194

Baltrušaitis, T., Ahuja, C. and Morency, L. (2018) 'Multimodal machine learning: A survey and taxonomy', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2), pp. 423–443.

DOI: 10.1109/TPAMI.2018.2798607

Behzadian, M. et al. (2012) 'A comprehensive literature review on TOPSIS applications', *Expert Systems with Applications*, 39(17), pp. 13051–13069.

DOI: 10.1016/j.eswa.2012.05.056

Caruana, R. et al. (2015) 'Intelligible models for healthcare: Predicting pneumonia risk and hospital readmission', in *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. New York: ACM, pp. 1721–1730.

DOI: 10.1145/2783258.2788613

Danks, D. and London, A.J. (2017) 'Algorithmic bias in autonomous systems', *International Journal of Human–Computer Studies*, 109, pp. 51–55.

DOI: 10.1016/j.ijhcs.2017.07.007

European Commission (2019) *Ethics Guidelines for Trustworthy AI*. Available at:

https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
(Accessed: 21 January 2025).

Ghassemi, M. et al. (2020) 'Opportunities in multimodal data fusion for healthcare', *Nature Machine Intelligence*, 2, pp. 252–260.

DOI: 10.1038/s42256-020-0190-0

Ghassemi, M. et al. (2021) 'The false hope of current explainable AI in healthcare', *Nature Medicine*, 27, pp. 1166–1167.

DOI: 10.1038/s41591-021-01413-z

Hauskrecht, M. (2000) 'Value-function approximations for POMDPs', *Journal of Artificial Intelligence Research*, 13, pp. 33–94.

DOI: 10.1613/jair.606

Heckerman, D. (1995) 'A tutorial on learning with Bayesian networks', *Microsoft Research Technical Report MSR-TR-95-06*. Available at:

https://www.microsoft.com/en-us/research/publication/a-tutorial-on-learning-with-bayesian-networks/
(Accessed: 21 January 2025).

Hwang, C.L. and Yoon, K. (1981) *Multiple Criteria Decision Making: Methods and Applications*. Berlin: Springer.

Miotto, R. et al. (2018) 'Deep learning for healthcare: review, opportunities and challenges', *NPJ Digital Medicine*, 1(1), pp. 1–10.

DOI: 10.1038/s41746-018-0029-1

Roijers, D.M. et al. (2013) 'A survey of multi-objective sequential decision-making', *Journal of Artificial Intelligence Research*, 48, pp. 67–113.

DOI: 10.1613/jair.4028

Semigran, H.L. et al. (2015) 'Evaluation of symptom checkers for self diagnosis and triage', *BMJ*, 351, p. h3480.

DOI: 10.1136/bmj.h3480

Shneiderman, B. (2020) 'Human-centred artificial intelligence: Reliable, safe & trustworthy', *International Journal of Human–Computer Interaction*, 36(6), pp. 495–504.

DOI: 10.1080/10447318.2020.1741118

Steinhubl, S.R. et al. (2015) 'The emerging field of mobile health', *JAMA*, 313(14), pp. 1409–1410.

DOI: 10.1001/jama.2015.3374

Topol, E. (2019) 'High-performance medicine: the convergence of human and artificial intelligence', *Nature Medicine*, 25, pp. 44–56.

DOI: 10.1038/s41591-018-0300-7

UNESCO (2021) *Recommendation on the Ethics of Artificial Intelligence*. Paris: UNESCO. Available at:

https://unesdoc.unesco.org/ark:/48223/pf0000379381
(Accessed: 21 January 2025).

van der Heijden, M. et al. (2019) ‘Bayesian networks for clinical decision support’, *Journal of Biomedical Informatics*, 93, pp. 103–117.

DOI: 10.1016/j.jbi.2019.103117

Zeng, Z. et al. (2009) ‘A survey of affect recognition methods’, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1), pp. 39–58.

DOI: 10.1109/TPAMI.2008.52

Zhao, Y. et al. (2020) ‘Wearable sensors for healthcare monitoring: A review’, *Sensors*, 20(10), p. 2821.

DOI: 10.3390/s20102821