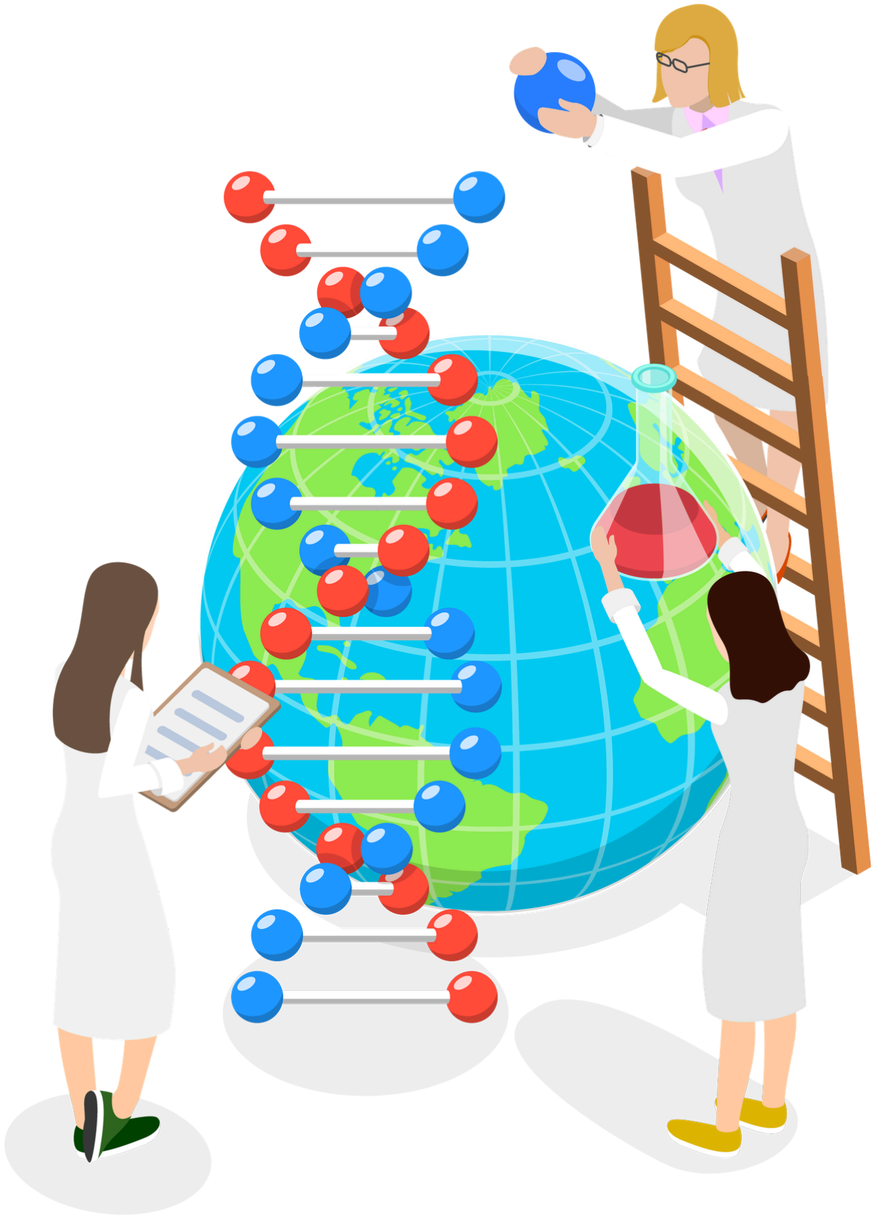
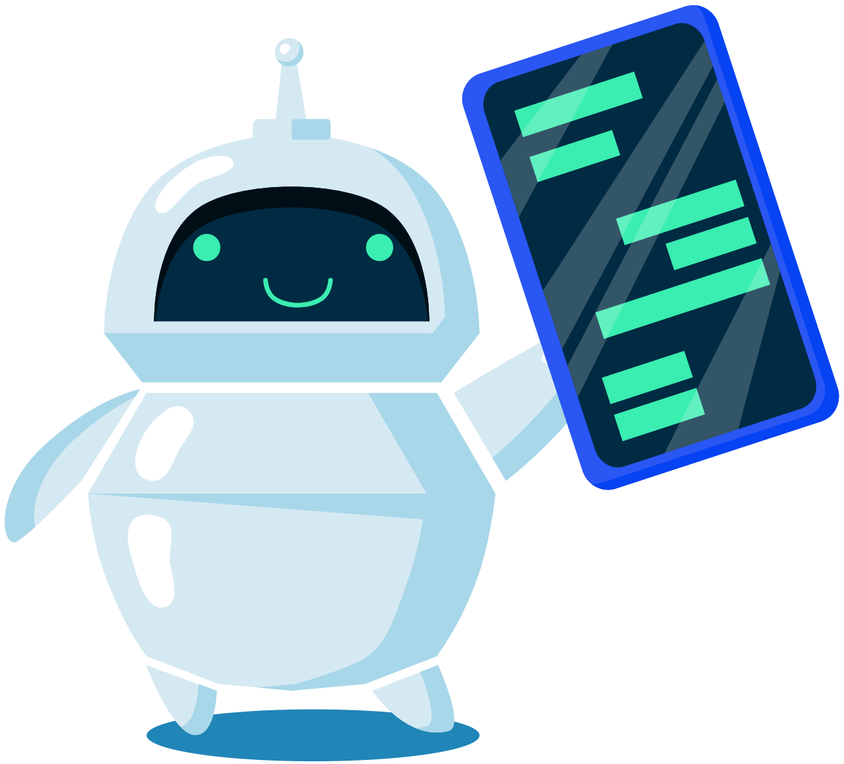
Applied AI in Healthcare

Understanding AI systems in clinical and real-world contexts



Institutul Național de Cercetare-Dezvoltare în Informatică ICI București



Administrative AI Applications for Operational Efficiency

Institutul Național de Cercetare-Dezvoltare în Informatică ICI București

Statistics. Understanding the complexity and necessity of administrative applications.

Case Study: University of California Medical Centers

Table MDA.1: Operating Statistics

	DAVIS	IRVINE	LOS ANGELES	SAN DIEGO	SAN FRANCISCO	TOTAL
Licensed beds						
2023	646	459	801	799	1,199	3,904
2022	646	459	801	799	1,250	3,955
2021	646	418	801	799	1,290	3,954
Admissions						
2023	33,123	22,609	38,436	36,156	44,309	174,633
2022	31,953	22,147	37,742	35,701	42,776	170,319
2021	29,953	21,885	35,691	34,311	40,895	162,735
Average daily census						
2023	620	411	787	671	877	3,366
2022	591	380	745	647	827	3,190
2021	560	364	698	594	774	2,990
Discharges						
2023	33,193	22,573	38,458	36,130	44,342	174,696
2022	31,888	22,136	37,689	35,704	42,852	170,269
2021	29,916	21,885	35,617	34,103	40,761	162,282
Average length of stay (days)						
2023	7.0	6.6	7.5	6.8	7.2	7.0
2022	6.8	6.3	7.2	6.6	7.0	6.8
2021	7.0	6.1	7.2	6.4	6.9	6.7

https://www.ucop.edu/uc-controller/financial-reports/systemwide-reports/medical-center-reports/22-23/medical-center-report-2023.pdf

Resource Planning and Scheduling

Problem complexity: An average-sized hospital has hundreds of employees, dozens of rooms and equipment units, and thousands of weekly appointments. Each variable is dependent on all the others.

If a surgery takes longer than expected:

- The next surgery scheduled in that operating room is delayed
- Subsequent surgeries involving the same medical team are delayed
- Patients remain longer in hospital wards
- Fewer beds become available
- Other emergency admissions are postponed

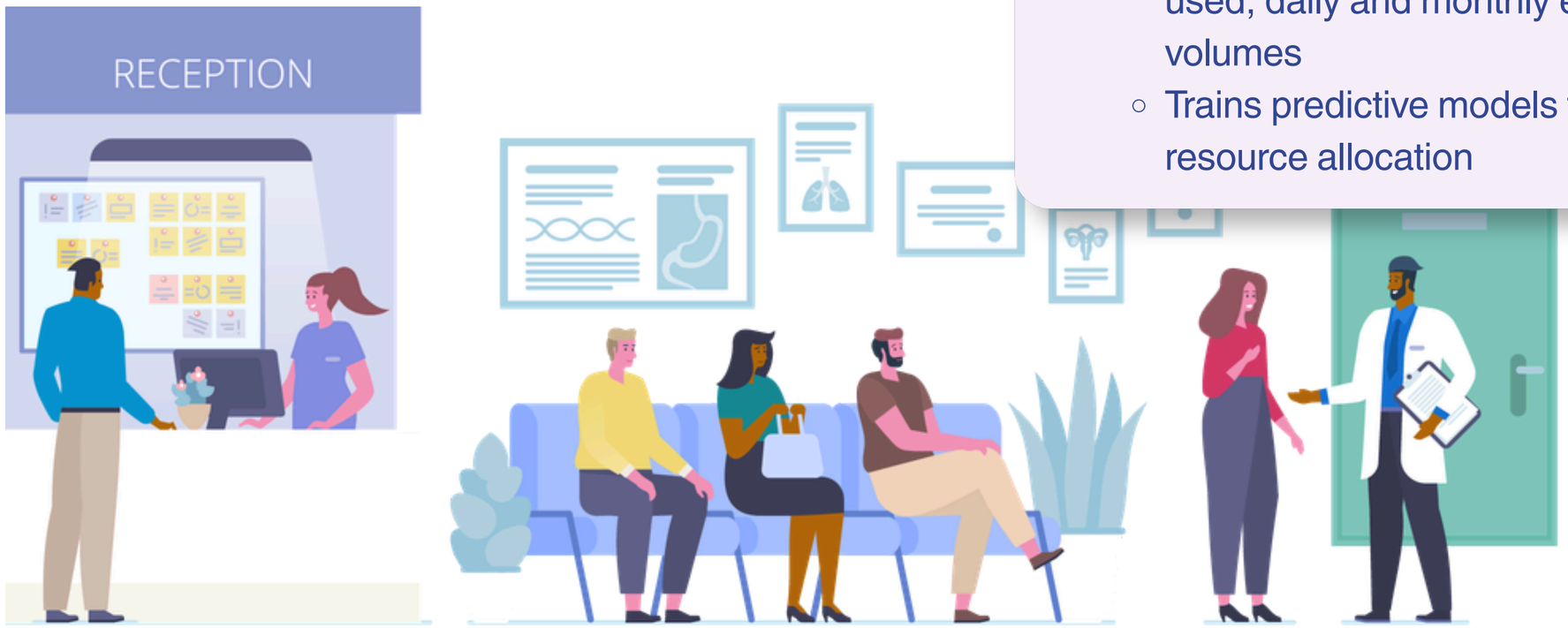
This system has numerous cascading interdependencies.

Resource Planning and Scheduling

How to:

- Track staff, rooms, and equipment
- Ensure availability
- Optimize everything efficiently in real time?

- Traditional planning: Manual processes or Excel-based systems
- AI approach:
  - Uses historical data: Number of surgeries over the past 3 years, duration of each procedure, resources used, daily and monthly emergency department volumes
  - Trains predictive models to optimize scheduling and resource allocation



How the Prediction System Works

INPUT DATA

Surgery history

Procedure durations

Resources used

Emergency Department (ED) patient flow

External factors (weather, traffic, events)

↓

AI MODEL

Learns from historical data

Identifies patterns

Generates predictions

↓

OUTPUTS

Number of ED patients (next day)

Estimated surgery durations

Resource demand

Delay risk

The model can say:
"Under conditions similar to next Tuesday, the hospital will likely receive X emergency patients, and Y surgeries will exceed their estimated duration."

Limitation: Models are only as good as the data they are trained on. Accurate predictions require complete and consistently recorded data.

Resource Planning and Scheduling

Medical staff

Anticipates patient volume
Automatically adjusts shifts
Reduces both overload and idle time
Enables crowding prediction

Patient flow prediction

ML models trained on historical data can estimate patient volume days in advance, allowing proactive resource adjustments.

Resource Planning and Scheduling

Medical instruments

Aligns surgical instrument availability with surgery schedules
Prevents last-minute cancellations

Rooms and equipmen

Optimizes operating room and equipment usage
Minimizes scheduling conflicts

Medical Documentation

Documentation and reporting

- Automatic generation of discharge letters
- Assisted completion of patient records (EHR)
- Periodic reports and hospital statistics
- AI-suggested diagnosis coding (ICD-10)
- Voice-to-text transcription for consultations

Doctors in Europe spend 25% to 50% of their time on documentation:

- *Discharge letters*
- *Patient records*
- *Reports*
- *Procedure coding for insurance reimbursement*

This is a major contributor to burnout.

- **Auto-completion.** EHR systems suggest: Diagnoses, ICD-10 codes, frequently prescribed medications based on past entries and statistical patterns.
- **Voice-to-text transcription.** The doctor speaks during the consultation. The system transcribes and structures the content into the correct fields.
- **Automated document generation.** The system listens to the doctor–patient interaction. Generates a draft of the discharge letter and consultation note. The doctor reviews and approves it.

Reporting and Standards

Standards define the format of data used by systems.
What happens if hospital systems cannot communicate?

- Duplicate data
- Medical errors
- Delayed decisions

Standards and Interoperability

- HL7 FHIR - standardized exchange of clinical data (modern REST API-based standard)
- ICD-10 / ICD-11 - international disease classification
- DICOM - standard for digital medical imaging
- GDPR & HIPAA - legal frameworks for data protection
- CE / FDA - certification of AI-based medical devices

AI makes compliance with these standards easier.

Reporting and Standards

🔗 HL7 FHIR (Clinical Data)

- With HL7 FHIR: Lab results are automatically integrated into the patient's record
- Without it: The doctor manually enters them or searches in another system

🏷️ ICD-10 / ICD-11 (Diagnoses)

- Example: „Type 2 Diabetes" → E11
- Importance: Billing, statistics, national reporting

🖼️ DICOM (Medical Imaging)

Enables system compatibility.
An MRI performed in one hospital can be viewed in another.

🔐 GDPR / HIPAA & ✅ CE / FDA

GDPR / HIPAA → Protect patient data
CE / FDA → Validate AI systems before clinical use

Scenario

An AI system recommends a treatment for a patient but does not have access to the full medical history (previous admissions, allergies, prior treatments). As a result, the model's recommendation is incorrect.

The doctor notices this and **does not administer the suggested medication**, choosing instead a treatment aligned with the patient's history.

What was the main issue?

A. The AI model wasn't trained properly

B. The infrastructure does not allow access to relevant data

C. AI should not be used for this type of decision

Submit

Scenario

An AI system recommends a treatment for a patient but does not have access to the full medical history (previous admissions, allergies, prior treatments). As a result, the model's recommendation is incorrect.

The doctor notices this and **does not administer the suggested medication**, choosing instead a treatment aligned with the patient's history.

What was the main issue?

A. The AI model wasn't trained properly

B. The infrastructure does not allow access to relevant data

C. AI should not be used for this type of decision

Submit

What would have solved the problem most effectively?

What would have solved the problem most effectively?

A. More data for model training

B. Additional training for the doctors using the AI systems

C. Integrations of systems through standards (e.g., FHIR)

D. Prohibiting AI in clinical decisions

Submit

What would have solved the problem most effectively?

What would have solved the problem most effectively?

A. More data for model training

B. Additional training for the doctors using the AI systems

C. Integrations of systems through standards (e.g., FHIR)

D. Prohibiting AI in clinical decisions

Submit

AI for Healthcare Financial Management

Impact (based on studies from USA 2022–2023):

- ~30% reduction in billing errors in hospitals using AI
- 2-5× faster procedure coding compared to manual processes
- ↓ 40% time spent on financial reconciliations

Automated Billing

Processing and validation of invoices, DRG code verification, detection of discrepancies before submission to insurers. Reduces claim rejections and improves reimbursement rates

Cost Estimation

Predictive models estimate the cost of a care episode based on historical similar cases

Fraud Detection

Identifies unusual billing patterns: incorrectly recorded procedures, duplicate entries, statistically improbable code combinations

Predictive Budgeting

Forecasts departmental expenses based on: historical data, seasonality, expected patient volume

A hospital losing €10 million annually due to incorrect billing can recover up to €3 million simply by automating code validation.
How billing errors occur: Incorrect DRG code selected for a procedure, invalid combinations of codes, missing required documentation. AI systems trained on insurer rules can detect these errors before submission.

Understanding Fraud Detection: AI systems can flag suspicious statistical patterns, such as a physician reporting significantly more complex procedures than their peers, or a department with an unusually high complication rate requiring reintervention.
An unusual pattern does not automatically mean fraud. It may reflect more complex patient cases, or a physician handling high-risk patients. AI only raises a flag. Humans always make the final decision.

Reducing Administrative Workload

Appointments and cancellations

Patients can schedule, reschedule, or cancel without phone calls: available 24/7

Triage

Collects symptoms and relevant history before consultation

Medication and treatment information

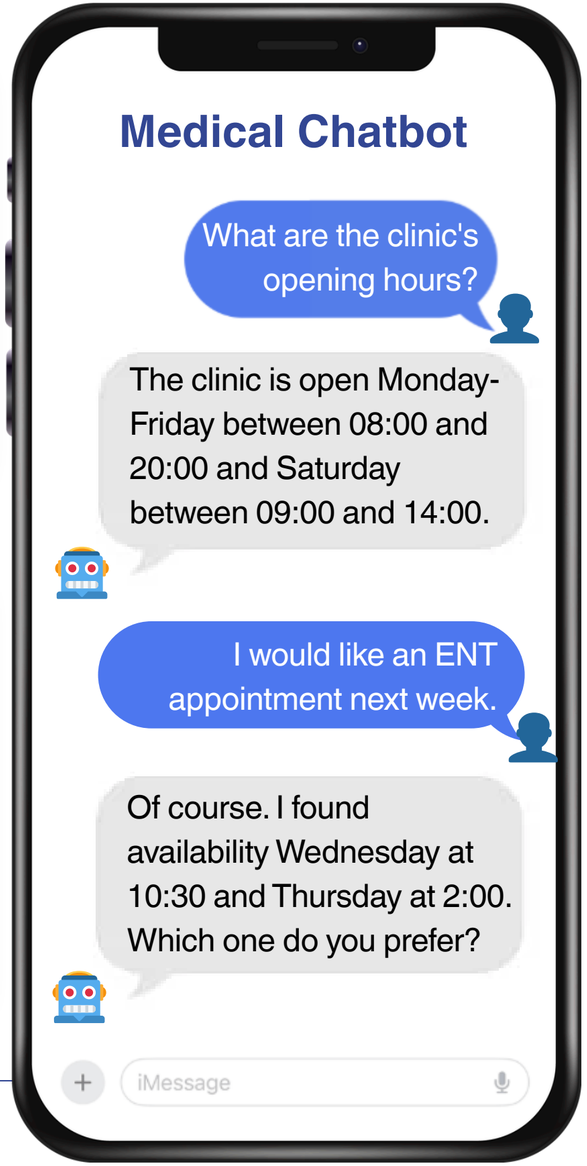
Instant answers to common questions, reduces staff workload

Automated reminders & follow-up

Pre-visit notifications, post-procedure instructions, treatment adherence monitoring

Access to documents and results

Patients can directly access: lab results, prescriptions, medical letters



Chatbots

the most visible point of contact between patients and AI today.

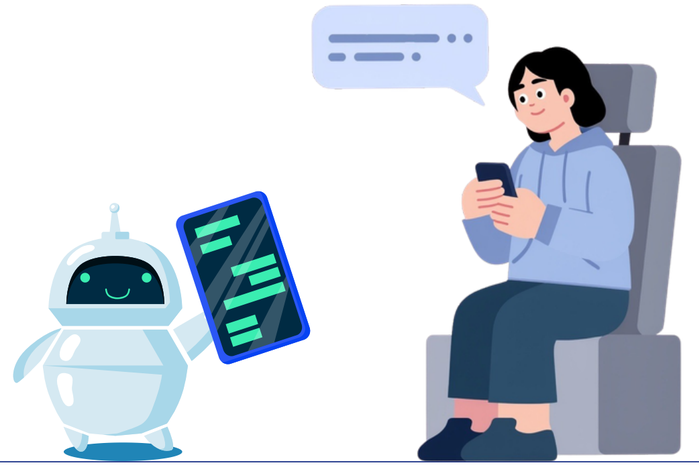
- **Administrative chatbots:** handle bookings, provide schedule information, send reminders. They do not make clinical recommendations. They do not interpret symptoms.

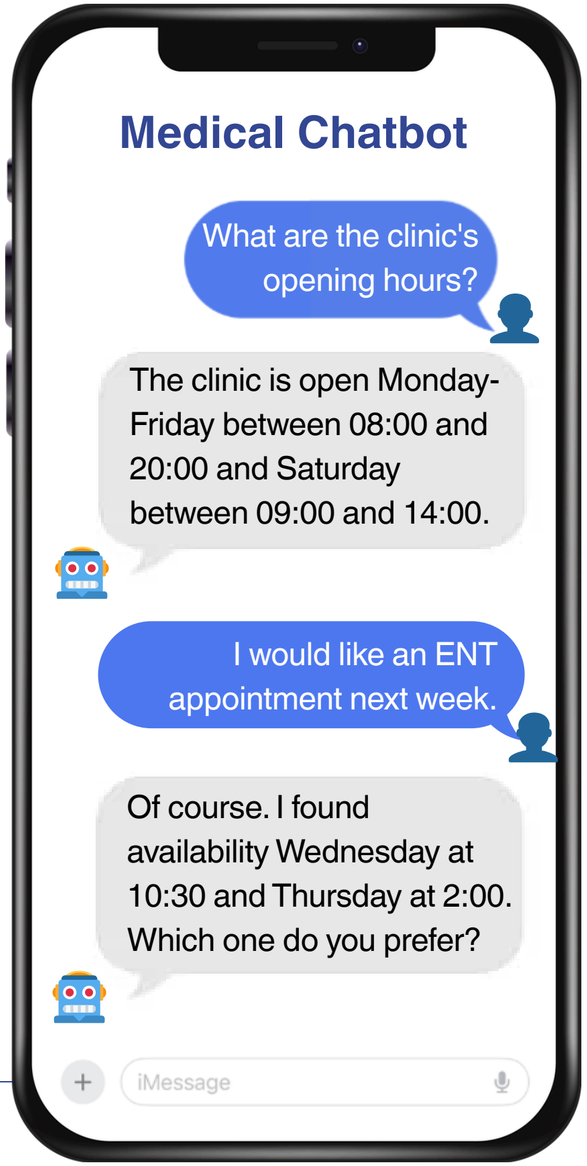
A medical center receives hundreds of calls per day, half of them for appointments. If a large portion of these calls can be handled by a chatbot, front-desk staff can focus on cases that truly require human interaction.

- **Clinical triage chatbots:** ask patients to describe their symptoms and attempt to guide them toward the appropriate level of care: emergency, primary care, or self-care.

Pre-consultation function: recording symptoms before the consultation. It does not directly make a clinical decision, but the collected data will influence the clinical decision.

Who is responsible if the information collected by the chatbot is incomplete or incorrectly structured and the physician makes a decision based on it?

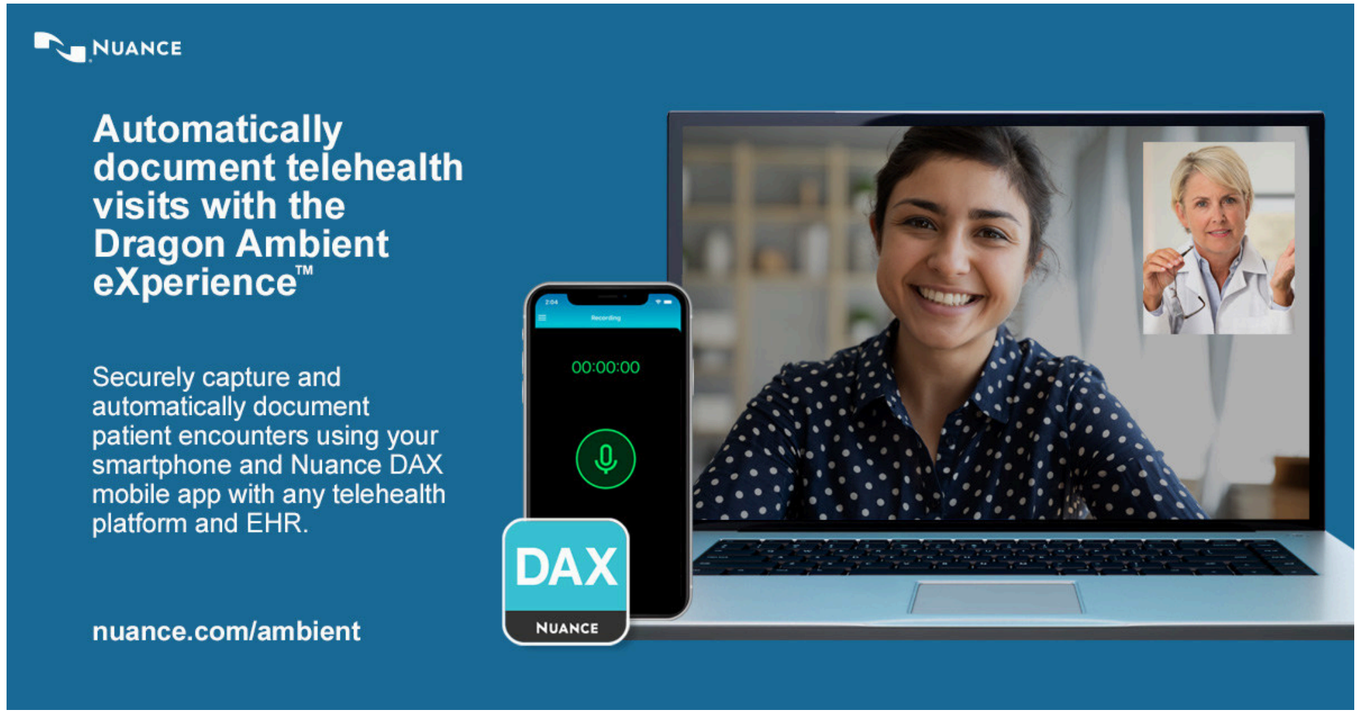




AI tools currently used in healthcare organizations

Domain	Tool
Clinical documentation & EHR	Nuance DAX: voice-to-clinical-notes Suki AI: voice assistant for physicians for EHR Epic NoteWriter: EHR summaries
Administrative (Operations & planning)	Qventus: patient flow prediction, operating room scheduling LeanTaaS: bed optimization for surgical units Olive AI: back-office automation (sold most of its assets in 2023) GE Healthcare Comand Center
Patient interaction	Ada Health (symptom checker) Babylon Health (clinical chatbot, telemedicine services, health monitoring) (bankrupt in 2023, acquired by eMed Healthcare UK) Sensely (virtual assistant)

Nuance DAX: voice-to-clinical notes



- Clinical documentation system
- Records the doctor-patient conversation during the consultation and automatically generates a structured clinical report in the EHR, which the physician then reviews and approves.

Consultation

↓

Mobile device (or microphone)

↓

ASR - Automatic Speech Recognition (with specialized clinical vocabulary)

↓

Speaker diarization (doctor/patient)

↓

NLP - extraction of relevant clinical information

↓

LLM - generation of structured notes (SOAP format)

↓

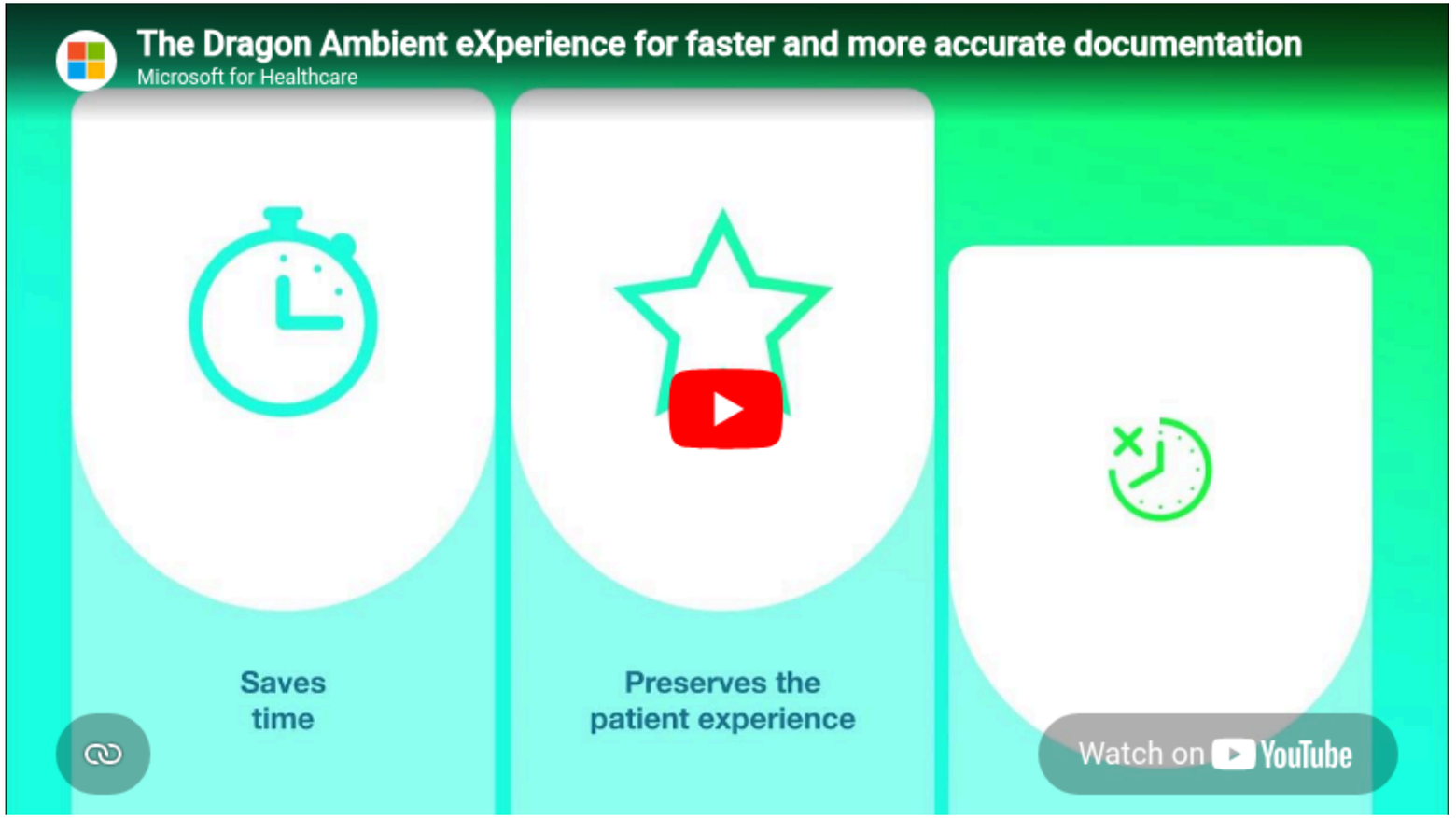
EHR Draft → reviewed and approved by the physician

SOAP

Subjective (patient-reported symptoms),
Objective (measurable observations),
Assessment (clinical diagnosis/interpretation),
Plan (treatment, follow-up, and goals).

Nuance DAX: voice-to-clinical notes

Used by over 550,000 physicians in the United States. Integrated with Epic, Oracle Health, and other EHR systems.



https://videos.microsoft.com/healthcare-industry-solutions-webinar-series/watch/oWP7riSLq6srYDeqjC4gbK

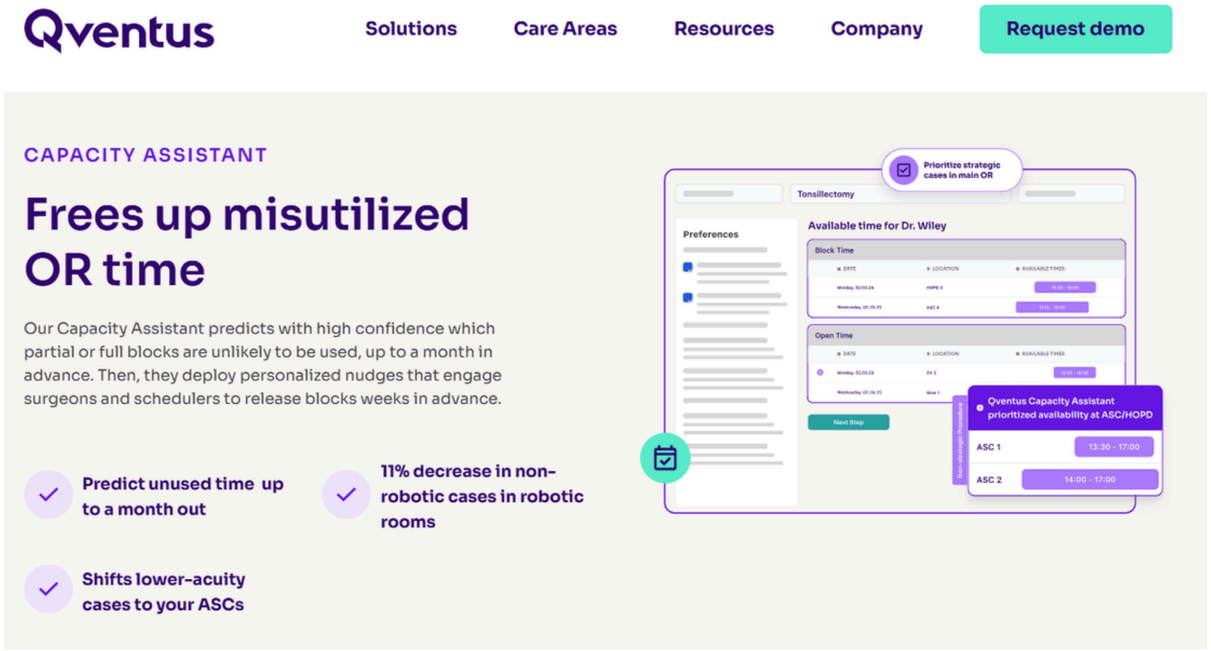
Nuance DAX: voice-to-clinical notes

Time spent on documentation	50-70% reduction in time required to write clinical reports (Nuance/Microsoft studies, 2022-2023)
Physician satisfaction	70% of physicians report reduced documentation-related burnout
Report quality	AI-generated reports are considered equivalent or superior to manual ones

Rubber-stamping	Overloaded physicians may quickly approve drafts without carefully reviewing them
Documented hallucinations	Reported cases where diagnoses appear in reports but were never discussed during the consultation
Language dependency	Performance drops significantly for consultations conducted in languages other than English

Major limitation: Most published data comes from Nuance/Microsoft (limited independent peer review)

Qventus - Optimizing hospital workflows using ML



- Surgical Scheduling - optimization of operating room schedules
- Discharge Planning - prediction and acceleration of patient discharges
- Capacity Management - real-time management of available beds

Historical EHR data (surgery durations, complications)

+

OR scheduling data (surgeon preferences, equipment, rooms)

+

Staff data (shifts, availability)

+

Supply chain data (availability of instruments, blood)

↓

Predictive models (e.g., Gradient Boosting, Ensemble Methods)

↓

Actionable recommendations for schedulers

Example: Dr. Jones's surgery on Thursday has a 73% chance of exceeding its scheduled duration by more than 30 minutes, based on the last 47 similar cases. Suggestion: extend the time slot or reschedule the next case.

Qventus - Optimizing hospital workflows using ML

Inpatient Capacity Solution Overview | Qventus

Qventus - Optimizing hospital workflows using ML

What implementations show
15-25% reduction in operating room idle time in large US hospitals
Discharges accelerated by 1-2 hours on average in pilot units
Positive ROI reported within 12-18 months after implementation
Reduction in staff overtime costs

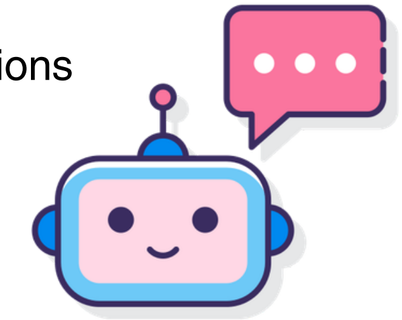
⚠️

Qventus optimizes for operational efficiency - costs and time, but cannot optimize for quality of care or patient satisfaction.

Conditions required for the system to work
Integrated data infrastructure: works only if EHR, OR scheduling, and staffing systems communicate in real time
Sufficient historical data volume: models require at least 2-3 years of high-quality data for reliable predictions; a small hospital with 3 operating rooms does not generate enough data
Adoption by staff: if surgeons ignore recommendations or schedulers do not trust the system, it cannot be effective
Local retraining: a model trained on US hospital data cannot be directly applied to a Romanian hospital without retraining on local data

Administrative chatbots in healthcare

Example based on real-world implementations (Infermedica, Gyant, Ada Health):
“MedBot”



What it can do:

- Appointments, rescheduling, cancellations, 24/7
- Answers to 150+ frequently asked questions (hours, location, test preparation)
- Sends reminders and pre-consultation instructions
- Collects demographic and administrative data for pre-registration
- Automatic handoff to a human operator when it cannot respond

User input (text / voice)

↓

Natural Language Understanding (identifies intent + extracts relevant information)

↓

Dialog Manager (manages conversation flow and session state)

↓

Integration with external systems (backend) (calendar, EHR, FAQ databases)

↓

Natural Language Generation (forms a natural response)

↓

Output to the user

Chatbots - Risks across the boundary between administrative and clinical.

Spectrum of healthcare chatbots

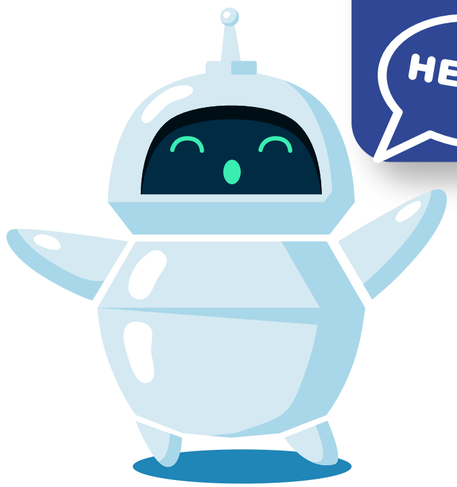
Type	Examples of tasks	Risk	Regulation
Purely administrative	Appointments, opening hours, location, documents	Low	Minimal
Pre-consultation	Symptom collection, history, allergies	Moderate	Requires validation
Triage	„Should I go to the emergency room?"	High	Medical device
Clinical recommendation	Suggests diagnosis, treatment	Critical	FDA/CE required

According to the EU AI Act (2024): a chatbot that provides clinical triage recommendations = high-risk medical device. Certification, transparency, and human oversight are mandatory.

Chatbots - Risks across the boundary between administrative and clinical.

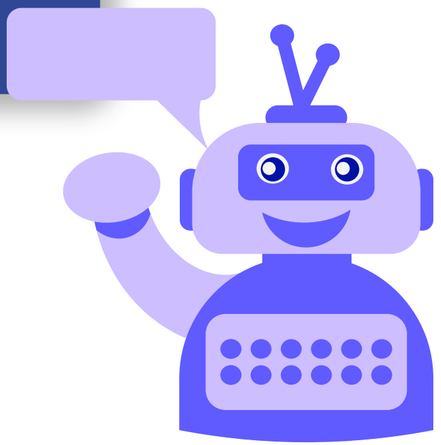
PURELY ADMINISTRATIVE
LOW RISK

"Your appointment is on Monday at 10:30 with Dr. Jones, 2nd floor, room 14."
→ No clinical decision. No medical risk.



CLINICAL TRIAGE
HIGH RISK

"Your symptoms are not urgent, you can stay at home." → This is a clinical triage decision.
If it is wrong, the consequence may be that a serious medical condition is treated too late.



Classify the tool

Read the 6 scenarios below. For each, choose the correct category.

LOW RISK

purely administrative
no clinical decision

MODERATE RISK

administrative with clinical implications

HIGH RISK

clinical

Scenario	Answer
An AI system reorders a radiologist's worklist, prioritizing studies with a high probability of pathology	
An AI system predicts which patients are ready for discharge within the next 24 hours and alerts the discharge team	
A chatbot confirms a patient's appointment and sends colonoscopy preparation instructions	

Classify the tool

Read the 6 scenarios below. For each, choose the correct category.

LOW RISK

purely administrative
no clinical decision

MODERATE RISK

administrative with clinical implications

HIGH RISK

clinical

Scenario	Answer
An AI system reorders a radiologist's worklist, prioritizing studies with a high probability of pathology	🔴 High risk - Clinical decision
An AI system predicts which patients are ready for discharge within the next 24 hours and alerts the discharge team	🟡 Moderate risk - Administrative with clinical implications
A chatbot confirms a patient's appointment and sends colonoscopy preparation instructions	🟢 Low risk - Administrative

Classify the tool

Read the 6 scenarios below. For each, choose the correct category.

LOW RISK

purely administrative
no clinical decision

MODERATE RISK

administrative with clinical implications

HIGH RISK

clinical

Scenario	Answer
An AI system automatically extracts information from medical letters and enters it into the patient record	
An AI system recommends the optimal treatment for a patient based on medical history and lab results	
An AI system optimizes the operating room schedule based on each surgeon's historical procedure durations	

Classify the tool

Read the 6 scenarios below. For each, choose the correct category.

LOW RISK

purely administrative
no clinical decision

MODERATE RISK

administrative with clinical implications

HIGH RISK

clinical

Scenario	Answer
An AI system automatically extracts information from medical letters and enters it into the patient record	🟡 Moderate risk - Administrative with clinical implications
An AI system recommends the optimal treatment for a patient based on medical history and lab results	🔴 High risk - Clinical decision
An AI system optimizes the operating room schedule based on each surgeon's historical procedure durations	🟢 Low risk - Administrative

Propose an AI Administrative System

You are an AI consultant for a regional hospital in Romania with 400 beds, 3 operating rooms, a functional EHR but no API integration, and an implementation budget of €150,000. The hospital has three main problems:

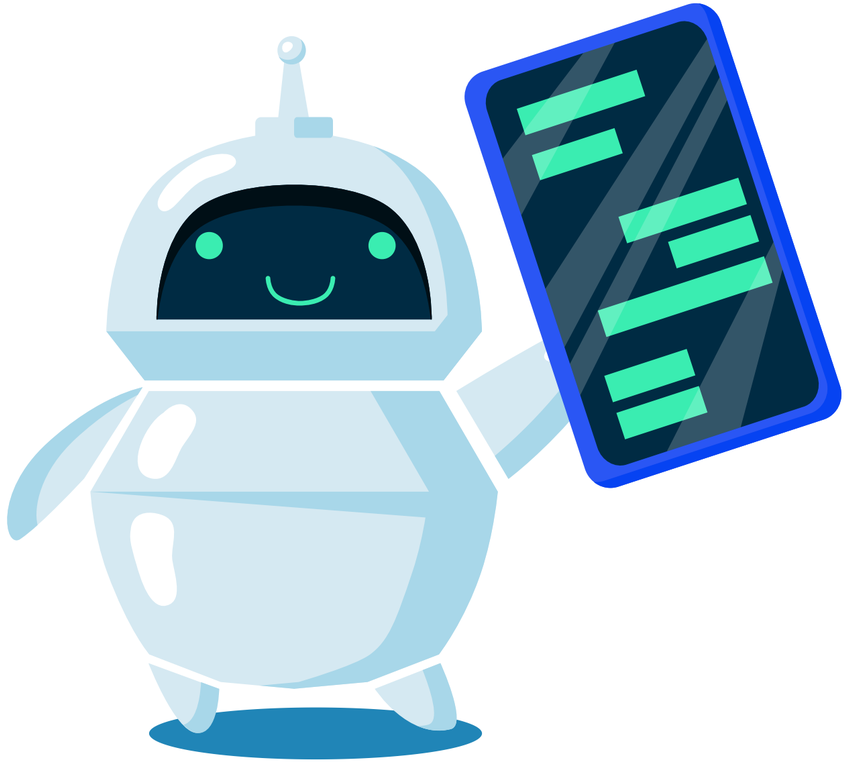
Operating room cancellation rate of 22% due to poor scheduling

Physicians spend on average 2.5 hours/day on documentation

Call center receives 800 calls/day, 60% for appointments or simple questions

Propose 2 administrative AI solutions. For each, mention:

- *What problem does it solve?*
- *What tool/approach is used?*
- *What data is needed?*
- *What infrastructure is missing?*
- *Biggest risk?*
- *How do you measure success?*



Ethical Considerations: Risk, Bias & Responsibility

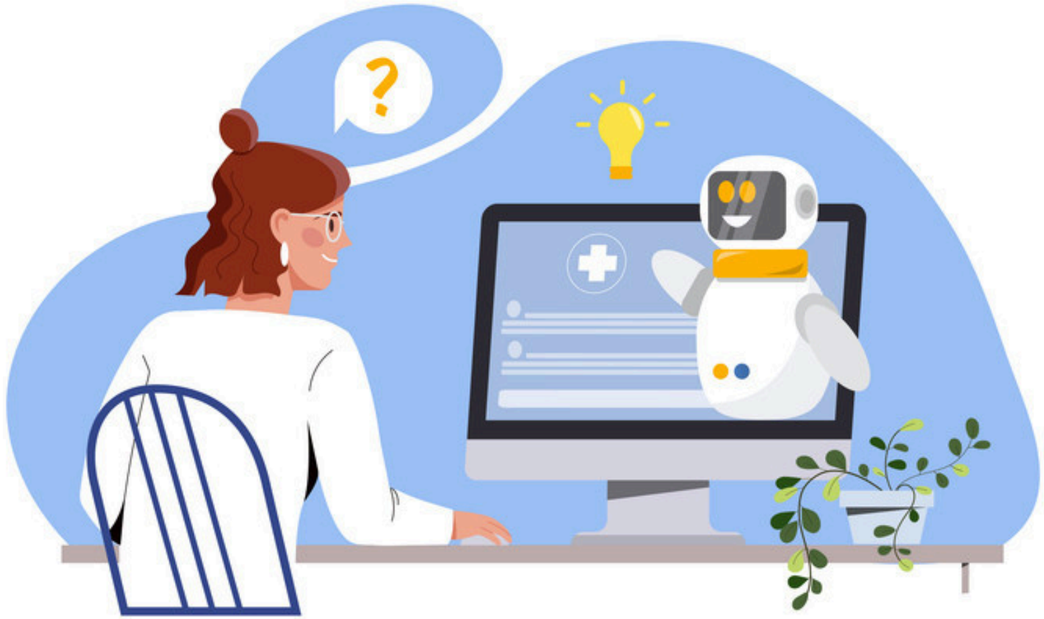
Responsible Use of AI in Healthcare

Just because an AI system functions correctly from a technical standpoint, does that mean it should be used? Is it truly safe? Are its outputs fair and equitable? And if the model produces an incorrect result, who is ultimately responsible?

Bias in AI

What is bias?

Bias in an AI system refers to a systematic error in its predictions or recommendations that disproportionately affects certain groups of people.



A widely cited example in the medical literature:

An algorithm used in the United States to prioritize patients for intensive chronic disease management was trained on historical healthcare cost data.

The underlying assumption seemed reasonable: patients who generate higher healthcare costs likely require more care.

However, there was a critical flaw. African American patients, due to longstanding economic and social barriers, had less access to healthcare services. As a result, they incurred lower healthcare costs. The algorithm therefore inferred that these patients were healthier, even though, on average, they had a higher burden of untreated medical conditions.

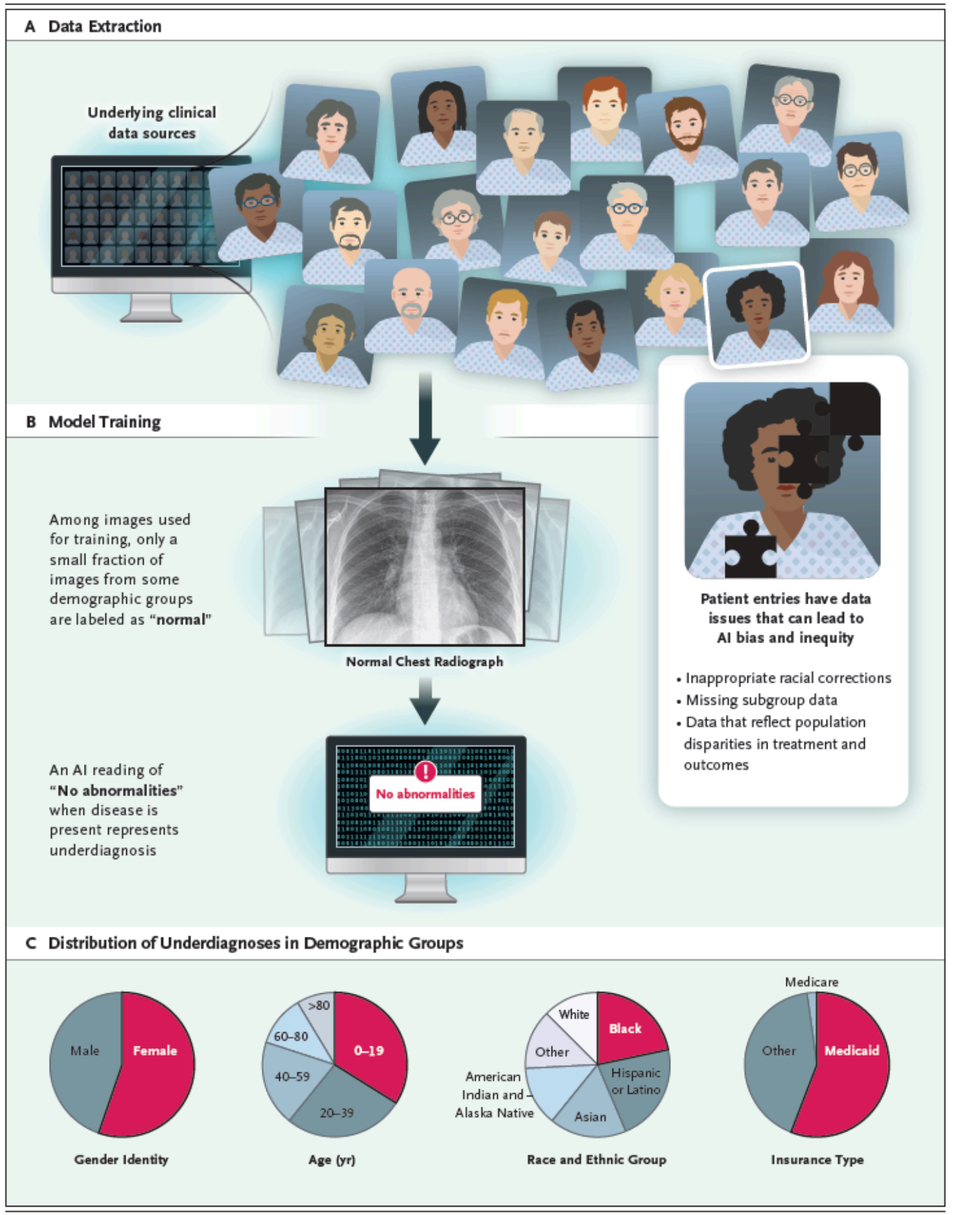
In this case, “cost” was used as a proxy for “health needs”, but it was an imperfect proxy, and one that was correlated with race.

https://apnews.com/racial-bias-in-health-care-software-aids-whites-over-blacks-b5fb4e1ae6e944fd979568df1fdb6df4

Bias in AI

Sources of Bias in AI Systems:

1. **Unrepresentative training data.** If a model is trained predominantly on data from European men aged 40-60, its performance will likely be weaker for women, older adults, or individuals from other ethnic backgrounds.
2. **Labeling bias.** If the clinical reports or diagnoses used as training data reflect physicians' biases, the model will learn and reproduce those same patterns.
3. **Proxy variables.** Models may rely on seemingly neutral variables (e.g., ZIP code, insurance type) as proxies for sensitive attributes such as ethnicity or income, leading to indirect discrimination.
4. **Feedback loops.** When AI systems influence clinical decisions, and those decisions become part of future training data, existing biases can become amplified over time.



Ferryman K, Mackintosh M, Ghassemi M. Considering Biased Data as Informative Artifacts in AI-Assisted Health Care. N Engl J Med. 2023 Aug 31;389(9):833-838. doi: 10.1056/NEJMra2214964. PMID: 37646680.

Bias in AI: Impact and Detection Challenges

Who is affected?

Women. Cardiovascular risk algorithms are often trained predominantly on male populations. Atypical heart attack symptoms (more common in women) are underrepresented.

Ethnic minorities. Particularly affected in fields such as dermatology (where models are often trained on lighter skin tones) and genomics.

Patients from rural or low-income backgrounds. Reduced access to healthcare leads to less available data, which in turn lowers model performance for these populations.

Why is bias difficult to detect?

Aggregate performance metrics. A model with 92% overall accuracy may perform at 85% for one subgroup and 96% for another. The global metric masks these disparities.

Lack of interpretability. When a system outputs a risk score, clinicians see a single value, not how performance varies across subgroups.

Biased test data. If the test dataset is as unrepresentative as the training data, validation will fail to detect underlying bias.

Delayed manifestation. Some biases only become apparent months or years after deployment, once sufficient real-world data accumulates.

Automation Bias

Definition: The tendency to follow recommendations from an automated system, even when they contradict one's own judgment or when there are indications that the system may be wrong.

This phenomenon was first extensively studied in aviation. Pilots who rely heavily on autopilot systems may react more slowly (or incorrectly) in emergency situations that require manual control. A similar pattern is observed in healthcare.

Effects
Radiologists using AI detection tools may pay less attention to areas not flagged by the system.
Clinicians tend to place greater trust in AI-generated risk scores than in their own judgment, even when they possess contextual information unavailable to the model.

Over-reliance and Deskilling

Definition: The gradual erosion of clinical or administrative skills due to consistent and excessive delegation to automated systems.

Deskilling is a slow, cumulative process. It is difficult to study because it unfolds over time and is highly context-dependent. Performance may appear adequate as long as systems function correctly, with problems only becoming visible during system failures or in the absence of technical support.

Effects
Reduced ability of physicians to interpret radiological images without AI assistance after prolonged use
Staff losing the ability to manage operating room workflows when automated systems fail
Decline in manual documentation efficiency
Erosion of critical thinking skills when they are not regularly exercised

Ethical Principles

FAIRNESS

A system is fair if it treats all patients equitably, regardless of gender, age, ethnicity, or geographic location. An AI system should deliver comparable performance and non-discriminatory outcomes across different patient groups, without amplifying existing healthcare inequalities.

1. What population was the model trained on?
2. Has performance been evaluated across demographic subgroups?
3. Are there significant performance differences between groups?
4. If so, are these differences ethically and clinically acceptable?

Ethical Principles

ACCOUNTABILITY

For any AI-assisted decision, there must be a clearly identifiable human actor who is responsible, able to justify the outcome, and to be held accountable for it.
AI systems cannot bear legal or moral responsibility. A human must always be held accountable.

1. Who signs off on the final decision?
2. Is there an audit trail of AI interventions?
3. What happens when the AI makes an error?
4. Does the vendor contract clearly define responsibility?

Ethical Principles

TRANSPARENCY

Transparency means that the functioning of an AI system is understandable, explainable, and verifiable at an appropriate level for different stakeholders: professionals (clinicians, auditors), institutions (hospitals, regulators), patients (in an accessible form).

1. Can the model's reasoning be explained? (Explainable AI)
2. Are the training datasets documented and accessible?
3. Is performance reported across subgroups?
4. Are patients aware they are interacting with an AI system?

Legal Framework: GDPR & EU AI Act

Core Principles:

- Health data is classified as a special category of personal data (Article 9) and requires enhanced protection.
- Processing is only permitted under strict conditions (e.g., public health interest or medical purposes), and consent must be explicit and informed, especially for automated processing.

Patient rights:

- Right not to be subject to solely automated decisions (Article 22)
- Right to request human intervention and to contest decisions
- Right of access, erasure (“right to be forgotten”), and data portability

Organizational obligations:

- Implementation of key principles:
  - data minimization
  - privacy by design & by default
- Conducting a Data Protection Impact Assessment (DPIA) for high-risk systems
- Appointing a Data Protection Officer (DPO) when processing sensitive data at scale
- Ensuring robust data security (technical and organizational measures)

Legal Framework: GDPR & EU AI Act

Core Principle:

AI systems are regulated based on their level of risk: the greater the potential impact on patients, the stricter the requirements.

Risk classification in healthcare:

- High risk - AI used for diagnosis, treatment, or surgical assistance → strict obligations
- Limited risk - chatbots, virtual assistants → transparency requirements
- Minimal risk - low-impact applications → no specific obligations

Requirements for high-risk systems (most clinical applications and some administrative ones):

- Pre-deployment conformity assessment
- Mandatory human oversight (human-in-the-loop)
- Logging and audit trails for decision traceability
- Continuous post-deployment monitoring
- Transparency toward users
- High data quality standards (relevant, representative, low bias)
- Risk management throughout the system lifecycle
- Comprehensive technical documentation

Liability and Responsibility

Who is responsible when AI makes a mistake?

Responsibility Ecosystem

Actor	Responsibility
Clinician	Final clinical decision, regardless of AI recommendations
Hospital / Healthcare provider	Proper implementation, staff training, and ongoing monitoring of system performance
AI vendor	Ensuring transparency, performance in line with specifications, and minimizing defects
Regulator	Market oversight and withdrawal of unsafe systems

'Responsibility gap': responsibility is shared, but no single actor holds it entirely.

AI as a Medical Device

Software as a Medical Device (SaMD):

Software intended by the manufacturer to be used for one or more medical purposes, without being part of a physical medical device.

If a digital system influences a clinical decision (diagnosis, treatment, monitoring, or prevention), it is considered a medical device, regardless of whether it runs in the cloud, on a smartphone, or in a browser.

Software	**Medical Device?**
Excel used for administrative statistics	❌ No
App showing clinicians lab results	✅Yes
Chatbot answering scheduling questions	❌ No
Chatbot advising patients on whether to go to the emergency department	✅ Yes
ML algorithm detecting atrial fibrillation on ECG	✅ Yes
Dashboard displaying vital signs without interpretation	⚠️ Depends

Risk Classification

EU Classification Framework

In the European Union, medical software is classified based on the risk it poses to patients.
The greater the impact on clinical decision-making, the higher the risk class and the stricter the regulation.

Class	Risk Level	Exemples	Conformity Assessment
Class I	Low	Scheduling or management software with minor clinical functions	Self-declared by manufacturer
Class IIa	Low-moderate	Decision support for non-critical diagnosis, general monitoring	Notified Body
Class IIb	Moderate-high	AI-assisted imaging diagnostics, vital sign monitoring with alerts	Notified Body
Class III	High	Software influencing treatment decisions for serious conditions, surgical assistance	Notified Body + rigorous clinical evaluation

A Notified Body is an independent, third-party organization designated by an EU Member State to assess whether high-risk products (like medical devices, machinery, or toys) meet strict safety regulations before being placed on the European market

Risk Classification

Applied Example

Scenario: A web platform that collects data from medical devices (ECG, pulse oximeter, glucometer), processes it using an ML model, displays interpretations to clinicians, and sends alerts to patients.

Risk classification evolves with functionality:

Raw data collection only (no interpretation) → Class I

+

Threshold-based display and automated alerts → At least Class IIa

+

ML model classifying arrhythmias → Class IIb or III

+

Automated treatment recommendations → Class III

Certification Process

SaMD Certification in the EU: From Prototype to CE Marking

Step 1. Classification

Determine the risk class according to MDR Annex VIII (Class I / IIa / IIb / III)

↓

Step 2. Technical Documentation

Device description and intended purpose

Risk analysis (ISO 14971)

Clinical performance data

User information

Post-market surveillance plan

↓

Step 3. Clinical Evaluation

Review of published clinical data

Own clinical validation data

Demonstration of equivalence or dedicated clinical studies (depending on class)

Certification Process

SaMD Certification in the EU: From Prototype to CE Marking

Step 4. Conformity Assessment

Class I: Self-declaration + registration in EUDAMED

Class IIa / IIb / III: Audit by a Notified Body

↓

Step 5. CE Certification

Registration in EUDAMED

↓

Step 6. Post-Market Surveillance (PMS)

Continuous monitoring

Incident reporting

Documentation updates for significant changes

Certification Process

SaMD Certification in Romania (ANMDMR)

In Romania, the National Agency for Medicines and Medical Devices (ANMDMR) is the competent authority that:

- *Registers medical devices on the national market*
- *Monitors reported incidents*
- *Can suspend or withdraw unsafe devices*
- *Coordinates with the European Commission and other Member States*

Class	Estimated Duration
I	3-6 months
IIa	12-18 months
IIb	18-24 months
III	24-36 months

Classify the Medical Software

For each of the software systems described below, determine and justify:

a) Is it Software as a Medical Device (SaMD)? (Yes / No / Depends)

b) If yes, what is the risk class? (I / IIa / IIb / III)

1. A web platform that receives data from a connected glucometer, automatically calculates glucose trends, and sends an alert to the physician if values exceed a predefined threshold.

2. An AI system that analyzes retinal fundus images and returns a probability score for diabetic retinopathy, displayed to the specialist alongside the original image.

3. A hospital's drug inventory management software that alerts the pharmacist when a drug is close to the expiration date.

4. A telemedicine platform that enables video consultations between physician and patient, without any automated processing of medical data.

5. A system that continuously monitors a patient's vital signs in the ICU (heart rate, SpO_2, blood pressure) and automatically generates deterioration alerts based on an Early Warning Score derived from combined parameters.

Classify the Medical Software

For each of the software systems described below, determine and justify:

a) Is it Software as a Medical Device (SaMD)? (Yes / No / Depends)

b) If yes, what is the risk class? (I / IIa / IIb / III)

Scenario	SaMD?	Class	Rationale
1	Yes	IIa	Automated alerts based on clinical data; software influences clinical decision-making
2	Yes	IIb/III	Automatically classifies a pathology; AI-assisted diagnosis with direct impact
3	No	-	Logistical function, no connection to clinical decisions
4	No	-	Communication infrastructure only
5	Yes	IIb/III	Real-time critical monitoring with automatic alerts, vulnerable patients

Data Governance: Principles and Obligations

Data governance is the foundation for the responsible use of AI in healthcare, enabling full control over how data is collected, processed, and used.

These principles function as an integrated system. If one is missing, significant risks may arise.

CONFIDENTIALITY

Medical data is accessed only by authorized personnel.

In practice: Role-based access control, encryption (in transit and at rest), audited access logs.

INTEGRITY

Data remains accurate, complete, and unaltered unless there is a justified reason.

In practice: Input validation, error detection, documented correction mechanisms, data versioning.

AVAILABILITY

Authorized users can access data when needed.

In practice: Backups, redundancy, disaster recovery plans, clearly defined uptime SLAs.

TRACEABILITY

Any action performed on data is recorded and attributable to a specific actor.

In practice: Comprehensive audit trails: who accessed what, when, and why; change logging.

TRANSPARENCY

Patients are informed about what data is collected, how it is used, and who can access it.

In practice: Clear notifications, accessible privacy policies, explanations of the role of AI in data processing.

FAIRNESS

Datasets are representative to prevent bias and inequities.

In practice: Assessment of representativeness during training, and post-deployment monitoring across demographic subgroups.

A system may be technically robust, but if it lacks transparency or fairness, it can become ethically and legally problematic.

Patient Rights in AI-Assisted Healthcare

1. Right to Consent and Control

Patients must provide explicit consent for the collection, processing, and secondary use of their health data.

- *Consent must be specific, informed, and revocable at any time*
- *Withdrawal of consent must not affect the care received*

Implication for AI: If a patient's data is used to train a model, this constitutes a separate purpose and requires separate consent.

2. Right to Information and Access

Patients have the right to know what data is collected, how it is processed, and to access their medical records, including outputs generated by AI systems that concern them.

- *Information must be clear and accessible, not purely technical or legal*

Implication for AI: If an AI system generates a risk score or recommendation that influences clinical decisions, the patient has the right to request an explanation.

Patient Rights in AI-Assisted Healthcare

3. Right to Correction and Erasure

Patients can request the correction of inaccurate data and, under certain conditions, its deletion.

Implication for AI: If a model was trained on incorrect data that is later corrected, what happens to the model? This raises the need for re-evaluation and potential retraining.

4. Right to Data Portability

Patients can request their data in a structured, machine-readable format to transfer it to another provider.

Implication for AI: This underpins the need for interoperability standards. Without standardized formats (e.g., HL7 FHIR), true portability becomes practically impossible.

5. Right to Object to Automated Decision-Making

Patients have the right not to be subject to decisions made solely by automated systems that significantly affect them.

Implication for AI: No AI system in healthcare should make final clinical decisions without human involvement. This is both an ethical requirement and a legal right.

The Interoperability Ecosystem in Healthcare

AI-generated image (ChatGPT)

Interoperability as an Ethical Requirement

Definition:

Interoperability is the ability of IT systems, medical devices, and AI systems to exchange, interpret, and use data reliably and securely, regardless of vendor or platform.

Beyond a technical concept, interoperability should be understood as an ethical requirement in modern digital healthcare systems.

The Four Levels of Interoperability

1. Technical - Systems can physically communicate (networks, communication protocols)

↓

2. Syntactic - Data follows a common format (e.g., HL7 FHIR, DICOM, standardized CSV)

↓

3. Semantic - Data has the same meaning across systems (e.g., SNOMED CT, LOINC, ICD)

↓

4. Organizational - Institutions have agreements and processes enabling data exchange

Interoperability as an Ethical Requirement

Lack of interoperability is not just a technical limitation. It has direct ethical implications for patient care, fairness, and accountability.

Patient Safety

If clinicians lack access to a patient's complete medical history due to non-communicating systems, clinical decisions may be incomplete or even incorrect.

Equity

If AI systems function only in well-equipped hospitals, patients in smaller or rural settings are excluded. System fragmentation reproduces and amplifies existing inequalities.

Traceability and Accountability

If data cannot be consistently tracked throughout the care pathway, auditing AI-driven decisions becomes impossible. Without traceability, accountability breaks down.

Right to Data Portability

Patient data portability is practically impossible without semantic interoperability.

Standards in the Context of AI in Healthcare

To address these challenges, a wide range of international standards have been developed to structure safety, interoperability, and quality in digital health systems. Together, these standards create a coherent ecosystem that ensures systems operate safely, efficiently, and interoperably.

Standard	Domain	Purpose	Who Should Know It
ISO 14971	Risk Management	Structured methodology for identifying and mitigating risks in medical devices	Engineers, SaMD manufacturers
ISO 13485	Quality - Medical Devices	Quality management system for organizations developing medical devices	Manufacturers, QA teams
IEC 62304	Medical Software Lifecycle	Processes for design, testing, maintenance, and modification of medical software	Software engineers, system architects
ISO/IEC 27001	Information Security	Information security management system	IT teams, CISOs, data teams
ISO 27799	Health Data Security	ISO 27001 applied specifically to healthcare data	Healthcare IT, DPOs
HL7 FHIR	Clinical Interoperability	Standardized format for exchanging clinical data via REST APIs	Engineers, system integrators
DICOM	Medical Imaging	Format and protocol for digital medical imaging	Biomedical engineers, imaging IT
SNOMED CT / LOINC / ICD	Medical Terminology	Standard vocabularies for coding diagnoses, procedures, and lab tests	Medical informaticians, clinicians

Standards in the Context of AI in Healthcare

Digital health systems are subject to complex requirements related to safety, security, and interoperability, defined through complementary international standards.
These standards cover the entire lifecycle: from design and development to certification, deployment, and post-market monitoring.

Design → ISO 14971 (risk management) + IEC 62304 (software lifecycle)

↓

Data Security → ISO 27001 + ISO 27799

↓

Interoperability → HL7 FHIR + DICOM + SNOMED CT

↓

CE Certification → ISO 13485 + ISO 14971 (explicitly referenced in MDR)

↓

Post-market → ISO 14971 (risk updates) + IEC 62304 (change management)

Interactive Case Study: Who Is Responsible?

A hospital in Romania implements an AI system for discharge prioritization (similar to Qventus). The system predicts that a 71-year-old patient admitted with pneumonia can be safely discharged the next day. An overworked physician follows the recommendation without thoroughly reviewing the medical record. The patient is discharged, readmitted 6 hours later with sepsis, and passes away three days later. An internal investigation reveals that the AI system did not have access to the most recent laboratory results from that morning, due to lack of integration within the hospital's IT infrastructure.

Team A

ATTENDING PHYSICIAN

What responsibility do you bear?
What could you have done differently?
What are the mitigating circumstances?

Team B

HOSPITAL MANAGEMENT

What is the institution's responsibility?
What processes were missing?
What should have been in place before deployment?

Team C

AI VENDOR

What responsibility does the vendor have?
What should have been clearly communicated?
What are the limitations of the product?

Team D

REGULATOR

What is the regulator's role?
What could have prevented this situation?
What should be improved in regulation or practice?

Responsible Adoption: How Do You Evaluate an AI System?

The responsible adoption of AI in healthcare begins with a structured, critical evaluation, both of the product itself and of how it will be integrated into clinical practice.

Questions About the Product

- What data was the model trained on? (population, country, time period, size)
- Are there independent, peer-reviewed validation studies?
- Has it been validated on a population similar to ours (demographically, clinically, infrastructurally)?
- What performance metrics are reported? (Accuracy alone, or also sensitivity, specificity, AUC, especially across subgroups?)
- How recent is the model? (A model trained in 2019 on pre-COVID data may be less relevant than one trained in 2025 on current data)

Questions About Implementation

- What IT infrastructure is required? Do we have it?
- How does it integrate with our existing systems (EHR, PACS, etc.)?
- What happens when the system makes a mistake? Is there a feedback and correction mechanism?
- Who is responsible for support, maintenance, and updates?
- Does the contract clearly define liability in case of an incident?
- What training do users receive, and how long does adoption take?

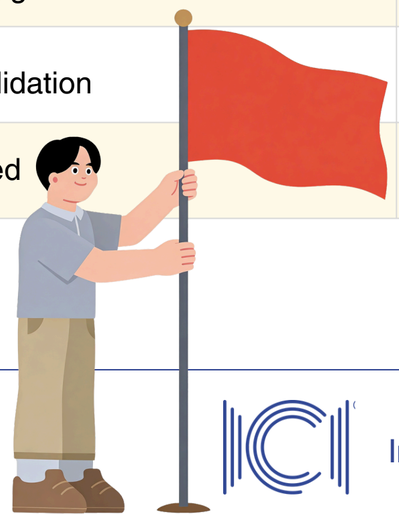
Spot the Red Flags

Read the following statements (from AI vendor marketing materials).
For each, decide: Red Flag 🚩 or OK ✅, and why.

Nr.	Statement
1	“Our AI system has 97.3% accuracy in detecting pneumonia on chest X-rays.”
2	“Validated on 180,000 images from 23 hospitals in 8 countries, study published in Radiology (2023).”
3	“Our AI completely eliminates the need for a radiologist in routine cases.”
4	“CE-certified Class IIa decision-support system. Not an autonomous diagnostic tool.”
5	“Doctors save 2 hours per day using our system - satisfaction guaranteed or your money back.”
6	“Implementation in 2 weeks, positive ROI guaranteed within 6 months.”
7	“Our model is updated annually with new data and revalidated before each release.”
8	“The system cannot be subject to external audit due to intellectual property constraints.”

Spot the Red Flags

Nr.	Statement	Verdict	Rationale
1	97.3% accuracy (no context)	🚩	Overall accuracy without dataset details, class balance, or subgroup performance is potentially misleading
2	Large, multi-center, peer-reviewed study	✅	Strong dataset, geographic diversity, independent validation
3	Replaces radiologists	🚩	No CE/FDA-approved system claims full replacement of clinicians. Clinically and legally unfounded
4	CE Class IIa, decision support	✅	Correct classification, clear communication of limitations
5	“Money-back guarantee”	🚩	Marketing-style claims inappropriate for medical devices; suggests lack of robust evidence
6	2-week implementation, guaranteed ROI	🚩	Unrealistic timeline; ROI guarantee is not credible in clinical environments
7	Regular updates + revalidation	✅	Good practice and transparency
8	No external audit allowed	🚩	Unacceptable for high-risk systems; auditability is a regulatory requirement



What Cannot Be Delegated to AI

Even as AI systems become more advanced, there are clear limits, both technical and ethical.
AI can support medical decisions, but it cannot replace them in areas involving clinical judgment, responsibility, and the human dimension of care.

Current Technical Limits	Permanent Ethical Limits
Causal reasoning: AI identifies correlations, not true causation	Final clinical decision: legal and moral responsibility remains with the physician
Out-of-distribution generalization: poor performance in unfamiliar scenarios (e.g., new diseases, new populations)	Communicating serious diagnoses: empathy and human connection cannot be replaced
Implicit context integration: AI cannot understand nuanced context (patient anxiety, family pressure, situational factors)	Informed consent: patients have the right to discuss decisions with a human
Real-time adaptation: models do not automatically update with new clinical evidence	High-stakes ethical decisions: e.g., resource allocation, end-of-life care

Responsibility for errors: AI cannot be held accountable. A human must always be responsible.

Evaluate the System

You are part of the committee of a regional hospital in Romania.
You receive a proposal for an AI-based discharge management system.

Vendor Claims:

- Discharge prediction accuracy: 94.1%
- Training data: 2.3 million care episodes (USA, 2018-2022)
- Integration: compatible with Epic, Cerner, Oracle Health
- Implementation: 3-4 weeks
- Certification: No CE marking (currently in progress)
- Published studies: 2 internal (non-peer-reviewed)
- Post-implementation support: 12 months included
- Price: €180,000/year
- Contract clause:
- “The vendor is not liable for clinical decisions made based on system recommendations.”

Team Task

1. Identify 3 red flags, with justification
2. List 2 critical missing pieces of information you would require before making a decision
3. Provide a final recommendation:

- ✅ Approve
- ⏸️ Approve conditionally (state conditions clearly)
- ❌ Reject (with justification)