Ciclul „INTELIGENȚA ARTIFICIALĂ ÎN MEDICINĂ”

Oct. – Dec. 2025

Coordonator ciclu: Prof. dr. Daniel LIGHEZAN

Coordonator științific: Prof. dr. Gheorghe Ioan MIHALAȘ

Tema 7

AI: Limite, Riscuri, Reglementări

Prof. Dr. Gheorghe I MIHALAS

Universitatea de Medicină și Farmacie Victor Babeș Timișoara

Centrul de Modelare a Sistemelor Biologice și Analiza Datelor

Laboratorul de Inteligență Artificială

Și

Academia de Științe Medicale

Comisia de Informatică Medicală, Inteligență Artificială și Protecția Datelor

Planul prelegerii

- **AI – beneficii și provocări**
- **Limite**
- **Aspecte etice**
- **Riscuri**
  - Actuale / Pe termen lung
  - Generale / Specifice domeniului medical
- **Cum diminuăm riscurile**
  - Reglementări
  - Educație
- **Concluzii / reflecții / discuții**

Planul tematic al ciclului de A.I. - 2025

Data	Tema prelegerii
7 oct	**Introducere în AI**. Definiții, istoric. Abordări cognitive și conexioniste. Sisteme discriminative și generative. Agenți AI. Aplicații medicale – privire generală, clasificări. Integrarea AI în activitatea medicală.
14 oct	**„Machine learning”,** învățare supervizată și nesupervizată. Analiza datelor medicale: predicții, clasificări. Date sintetice, simulări. Modelarea proceselor, proiectarea studiilor clinice, evaluare.
28 oct	**AI pentru decizia medicală**. Reprezentarea cunoștințelor, ontologii medicale, codificări. Sisteme logice, sisteme baysiene. Logica fuzzy. Sisteme expert medicale, decizie asistată. Monitorizarea pacienților.
4 nov	**Rețele neuronale**: principii, clasificări. Perceptronul, „deep learning”. Aplicații în imagistica medicală, bioinformatică, medicina de precizie. Performanță vs. transparență, criterii de validare.
11 nov	**Sisteme generative conversaționale**. Prelucrarea limbajului natural, „LLM – large language models”. Construcția și funcționarea unui Chat-bot. Aplicații medicale, chatboți specializați. Prompt engineering. Limite și capcane.
18 nov	**AI în învățământul medical și în cercetare**. Generare întrebări, cazuri clinice interactive. Utilizarea LLM în redactarea științifică. Construcția și antrenarea de chatboți specializați. Etica utilizării AI în educație și cercetare.
25 nov	**Limitele actuale, aspecte etice și juridice**. Limite: confidențialitate, bias. Transparență, sisteme explicabile XAI. Tehnologia empatică, riscuri și abuzuri. Competențe AI, responsabilități. Reglementări.

AI în medicină:
Beneficii și Provocări

- **Beneficii în:**
  - practica medicală
    - performanțe ridicate
    - management
    - relația medic-pacient, pacientul informat
  - cercetare
  - învățământ medical
- **Provocări**
  - limite – tehnice, epistemologice, organizaționale
  - aspecte etice
  - riscuri
- **Măsuri** de creștere a beneficiilor și diminuare a riscurilor

Limitele actuale ale Inteligenței Artificiale în Medicină

„When you invent the ship, you also invent the shipwreck.”

P. Virilio

Originea limitelor

- AI nu „înțelege” medicina, operează pe pattern-uri
- Modelele sunt dependente de date → bias, lipsă de generalizare
- Opacitatea modelelor mari („black box”)
- Responsabilitatea rămâne la profesionistul uman
- **Limită ≠ risc**, dar generează scenarii cu risc

Limite tehnice

- 1. Halucinații
  - – Răspunsuri fabricate, dar coerente
- 2. Inconsistență
  - – Răspunsuri diferite la aceleași întrebări *(sunt construite ad-hoc)*
- 3. Bias-ul datelor
  - - Subreprezentarea unor populații / rezultatelor negative
  - - Pericolul „infectării” cu date eronate ***(AI nu uită!)***
- 4. Lipsa generalizării
  - – Modelele „consumer” nu sunt optimizate pentru clinici
- 5. Actualizare incompletă
  - – Delay între cunoașterea medicală și versiunea modelului

Limite epistemice

- 1. Opacitate (XAI încă insuficient)
  - Justificările sunt aproximative, post-hoc
- 2. Nu înțelege cauzalitatea
  - Confuzie între corelație și mecanism clinic
- 3. Lipsa contextului uman
  - Nu percepe ton, nonverbal, preferințe pacient
- 4. Fragilitate la formulare
  - Promptul determină radical calitatea răspunsului
- 5. Fără responsabilitate
  - Nu poate evalua consecințe etice / legale

Limite practice și organizaționale

„Underuse, overuse and misuse" a AI în medicină

- Calitatea inputului (garbage-in, garbage-out)
- Prompts ambigue
- Limitările modelelor gratuite / consumer AI
- Protecția datelor în conversații
- Absența ghidurilor de bune practici pentru AIM

Aspecte etice

Principiile etice în AI medicală

Principiile etice biomedicale „adaptate” la AI

- **Beneficitate** — AI trebuie să aducă valoare reală pacientului
- **Non-maleficitate** — prevenirea erorilor, halucinațiilor, bias-urilor
- **Autonomie** — pacientul are dreptul la explicație și la consimțământ informat
- **Justiție** — evitarea discriminărilor introduse de date și algoritmi
- **Responsabilitate profesională** — medicului îi revine decizia finală

Dileme etice ale AI în medicină

Cine răspunde pentru o decizie greșită generată de AI?

- **Acces inegal la AI** → risc de „medicină cu două viteze”
- **Dependența excesivă de AI** → eroziunea competențelor clinice
- **„Pseudo-empatie”** → riscuri psihologice pentru pacientul vulnerabil
- **Autonomie afectată**: pacientul nu știe dacă sfatul e uman sau generat de AI
- ***Recunoaștere**: AI nu este agent moral; medicul este.*

Principii de utilizare responsabilă

„AI Safe Use in Medicine”

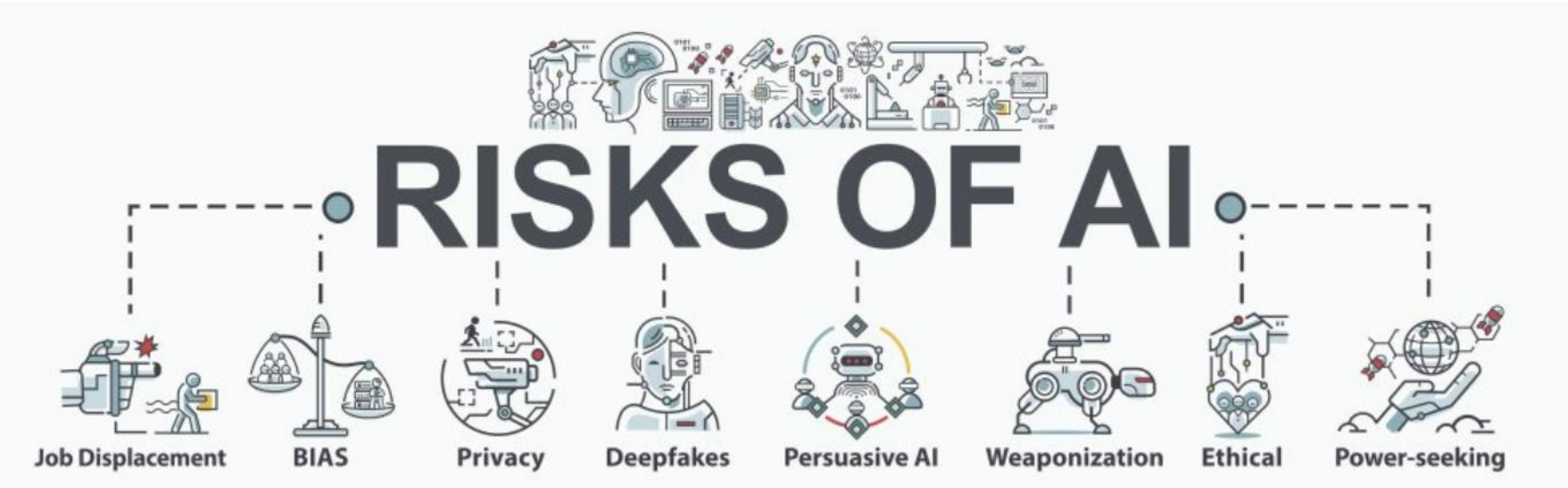
- **AI asistențială**, nu AI *decisională*
- **Decizii validate** → „human in the loop”
- **Explicabilitate minimă obligatorie** pentru cazurile critice
- **Limitarea scopului**: AI trebuie folosită doar pentru rolul declarat
- **Evitarea antropomorfizării** (în special la pacienți)
- **Documentarea deciziilor** influențate de AI

Riscurile utilizării AI în medicină

„During use / Underuse / Overuse / Misuse / Maluse”

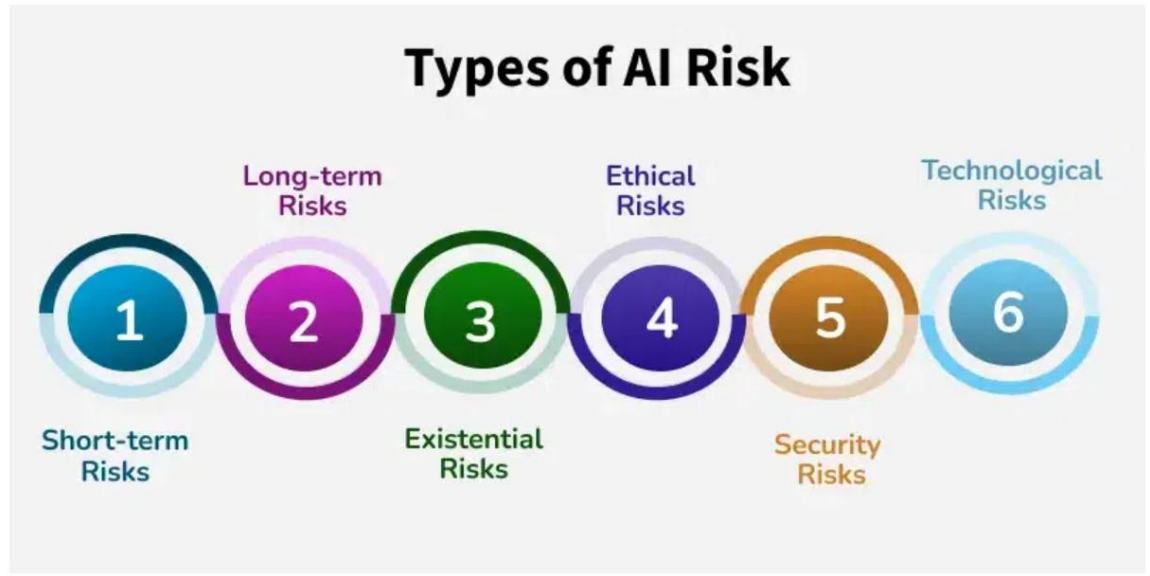
Istoria științei arată că:

- Progresul nu este liniar
- Fiecare revoluție tehnologică a adus noi provocări
- Inteligența Artificială nu face excepție



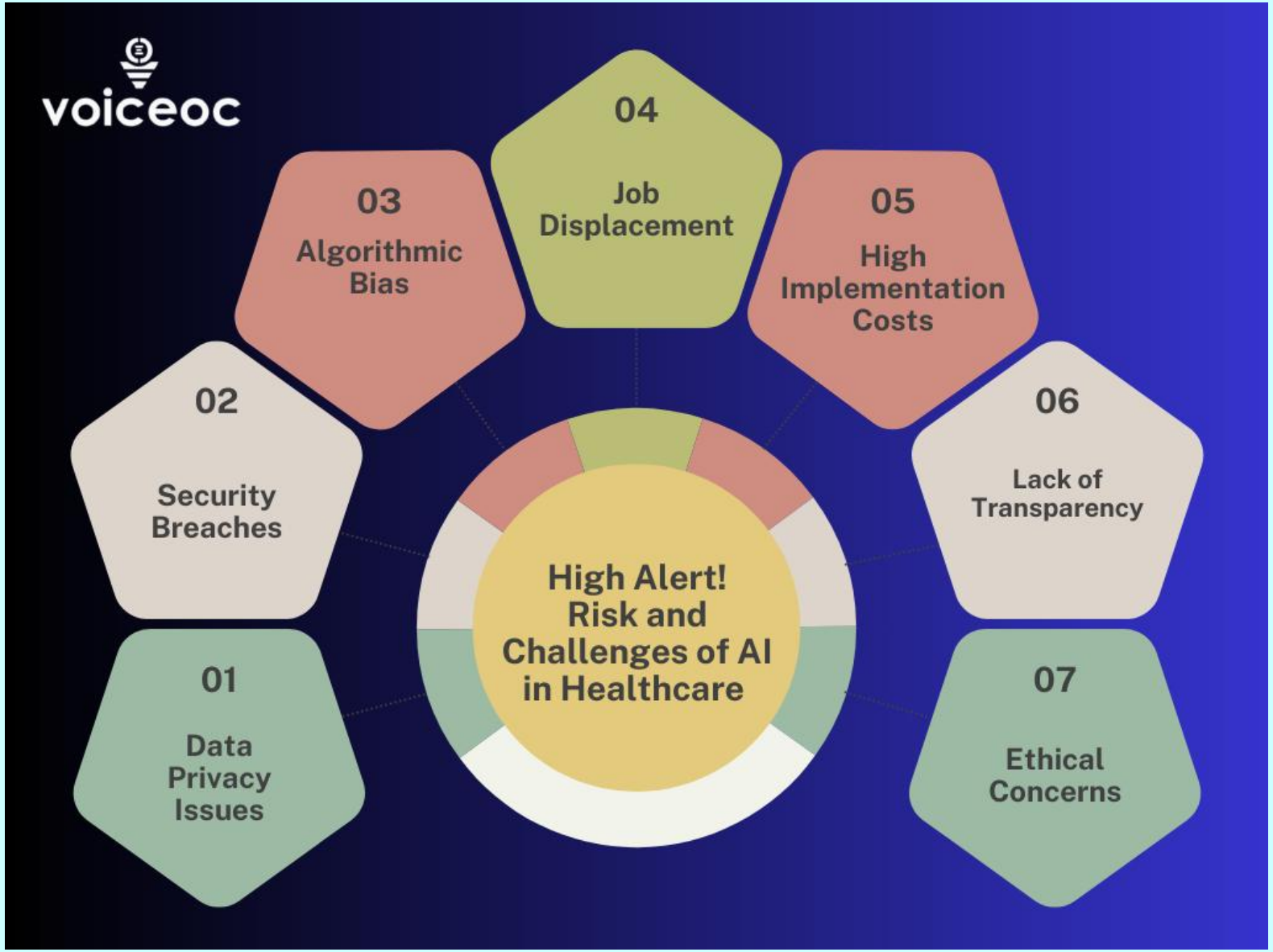
Provocări majore ale AI

- Limitele tehnice și științifice ale modelelor actuale
- Aspecte etice
- Dileme epistemice
- Aspecte legale și juridice
- Implicații organizaționale și profesionale
- Impact educațional, psihologic și cultural
- Efecte pe plan social și geopolitic
- Provocări existențiale



Riscuri generate de limite tehnice și științifice ale modelelor actuale

- **Calitatea și completitudinea datelor**
  - modelele sunt la fel de bune ca datele care le antrenează;
  - lacunele sau datele incorecte duc la erori sistematice.
- **Bias algoritmic**
  - părtinirile din date devin părtiniri ale modelului, chiar dacă nu sunt intenționate.
- **Halucinațiile / fabulațiile modelelor generative**
  - produc informații false, dar coerente, în lipsa unei verificări factuale.
- **Lipsa explicabilității (black box)**
  - modelele complexe oferă rezultate corecte, dar greu de justificat.
- **Generalizarea limitată**
  - performanță bună pe date similare cu cele de antrenament, scăzută în contexte noi.
- **Dependenta de resurse**
  - necesită volume uriașe de date, energie și putere de calcul.


voiceoc
04
Job Displacement
03
Algorithmic Bias
05
High Implementation Costs
02
Security Breaches
06
Lack of Transparency
High Alert! Risk and Challenges of AI in Healthcare
01
Data Privacy Issues
07
Ethical Concerns

Aspecte etice și juridice

- **Responsabilitatea deciziei**
  - cine răspunde când AI greșește: medicul, dezvoltatorul sau instituția?
- **Consimțământul informat**
  - pacientul sau utilizatorul trebuie să știe dacă și cum este implicată AI în procesul decizional.
- **Protecția datelor personale**
  - colectarea și procesarea masivă a informațiilor sensibile ridică riscuri de confidențialitate.
- **Transparența algoritmică**
  - modelele trebuie să poată fi auditate, interpretate și contestate.
- **Echitatea și nediscriminarea**
  - evitarea părtinirilor care pot afecta grupuri vulnerabile.
- **Utilizarea conformă scopului**
  - evitarea deturnării tehnologiilor AI către aplicații neetice sau non-medicale.

Dileme epistemice

- **Predicție fără înțelegere**
  - modelele AI operează probabilistic, fără a avea noțiunea de cauzalitate.
- **Acurateță vs. adevăr**
  - un rezultat corect statistic (sau chiar predictiv) nu este neapărat o explicație validă.
- **Halucinații și pseudo-cunoaștere**
  - AI poate genera informații plauzibile, dar complet false.
- **Eroziunea gândirii explicative**
  - se pierde treptat rolul raționamentului cauzal și clinic uman.
- **Criza sursei**
  - informația nu mai este verificabilă prin autoritate sau origine, ci doar prin coerență internă.
- **Confuzia între încredere și adevăr**
  - tonul sigur al AI poate masca incertitudinea epistemică.

Implicații organizaționale și profesionale

- **Redefinirea rolurilor profesionale**
  - AI devine partener decizional, nu simplu instrument.
- **Rezistența la schimbare**
  - teama de înlocuire sau pierderea autonomiei profesionale.
- **Necesitatea reconfigurării fluxurilor de lucru**
  - integrarea AI în procese clinice și administrative.
- **Supraîncrederea în algoritmi**
  - riscul de „automatizare a deciziei” fără judecată critică.
- **Lipsa pregătirii manageriale**
  - multe instituții adoptă tehnologia fără o strategie clară.
- **Nevoia colaborării interdisciplinare**
  - medic, informatician, jurist, bioetician – ***toți trebuie implicați!***

Impact educațional, psihologic și cultural

- **Alfabetizarea digitală insuficientă**
  - mulți utilizatori folosesc AI fără a-i înțelege principiile sau limitele.
- **Dependenta cognitivă**
  - tendința de a apela la AI înainte de a gândi critic.
- **Deplasarea motivației**
  - accent pe rapiditate și rezultat, nu pe învățare și reflecție.
- **Efecte psihologice**
  - supraîncredere, anxietate sau atașament afectiv față de sisteme „empatice”.
- **Transformări culturale**
  - schimbarea modului în care producem, transmitem și valorizăm cunoașterea.
- **Riscul pierderii autenticității**
  - comunicarea devine uniformizată, filtrată prin modele generative.

Efecte pe plan social și geopolitic

- **Inegalități de acces**
  - țările și regiunile cu resurse reduse riscă să rămână în afara revoluției AI.
- **Concentrarea puterii tehnologice**
  - dependență de câteva platforme globale și monopoluri de date.
- **Suveranitatea datelor**
  - informațiile medicale și personale pot fi stocate și procesate în afara controlului național.
- **Manipularea informațională**
  - riscul utilizării AI în propagandă, dezinformare sau control social.
- **Schimbări în piața muncii**
  - redistribuirea competențelor și accentuarea diferențelor socio-profesionale.
- **Responsabilitatea colectivă**
  - necesitatea cooperării internaționale pentru reglementare și etică globală

Provocări existențiale

- **Autonomia sistemelor**
  - pericolul ca decizia să scape treptat de sub controlul uman.
- **Dilatarea graniței dintre om și mașină**
  - AI care simulează vocea, emoția, empatia, intenția.
- **Eroziunea valorilor umane**
  - eficiența începe să înlocuiască sensul, iar performanța – compasiunea.
- **Dependența cognitivă și emoțională**
  - omul se sprijină tot mai mult pe sisteme care „gândesc” în locul său.
- **Riscul pierderii autonomiei morale**
  - delegarea discernământului către algoritmi.

De la provocări la responsabilitate

- Fiecare progres aduce nu doar beneficii, ci și **noi forme de responsabilitate**.
- Provocările identificate nu sunt obstacole, ci repere pentru acțiune.
- Soluțiile cer **echilibru între inovație și reglementare**, între libertate și protecție.
- Răspunsul nu poate fi doar tehnologic –
  trebuie să fie **instituțional, etic și educațional**.

Ce putem face?

„The future depends on what we do today.”

Mahatma Gandhi

De la provocări la responsabilitate

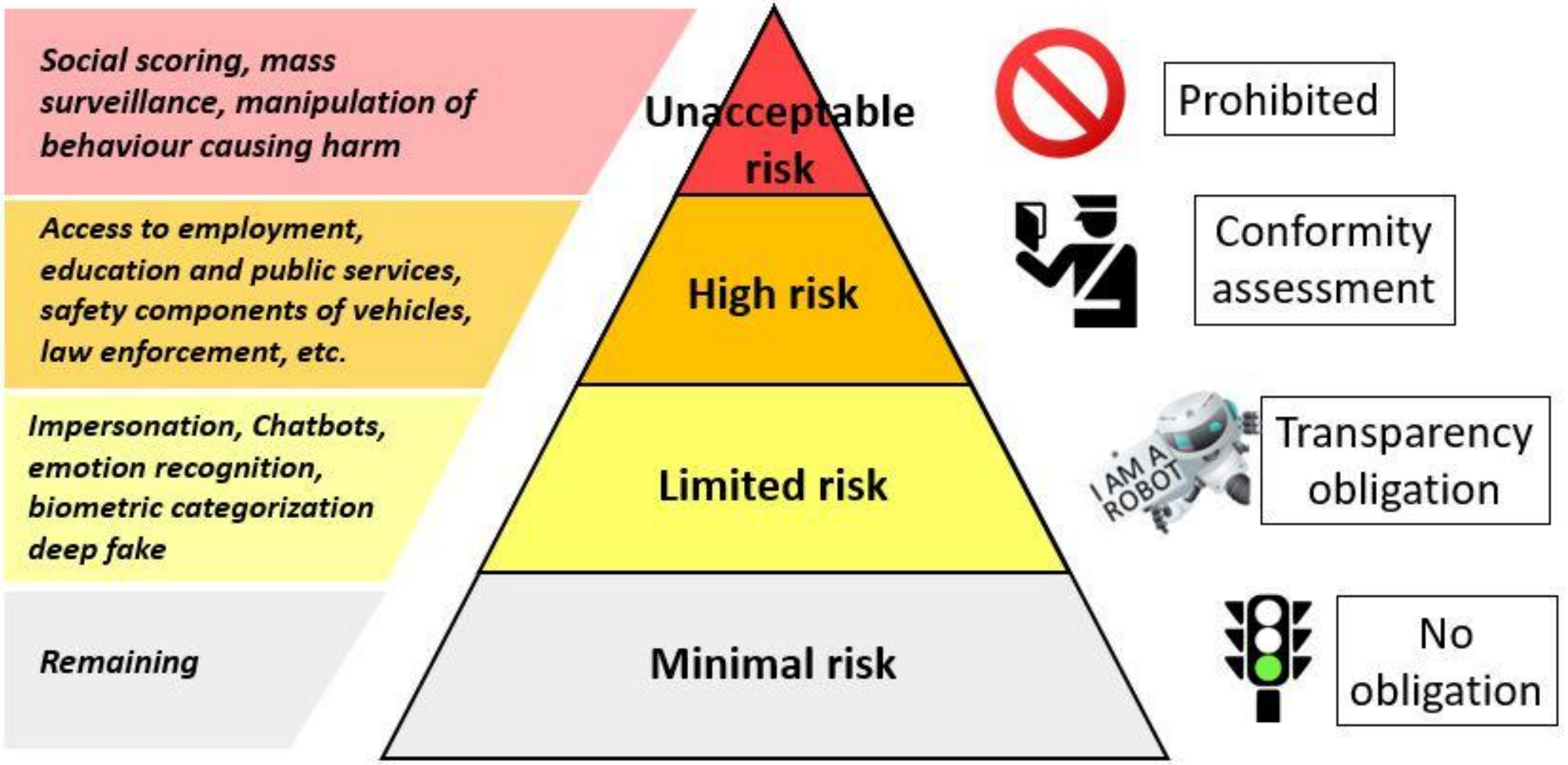
- Fiecare progres aduce nu doar beneficii, ci și **noi forme de responsabilitate**.
- Provocările identificate nu sunt obstacole, ci repere pentru acțiune.
- Soluțiile cer **echilibru între inovație și reglementare**, între libertate și protecție.
- Răspunsul nu poate fi doar tehnologic –
  trebuie să fie **instituțional, etic și educațional**.

- **DIRECȚII PRINCIPALE:**
  - Reglementări
  - Educație
  - Măsuri organizatorice și infrastructură

Reglementări

- **Ce s-a făcut deja**
  - EU AI Act: clasificare pe riscuri, cerințe stricte pentru AI medicală (risc înalt)
  - GDPR: protecția datelor sensibile, limitări privind trainingul AI cu date reale
  - FDA / EMA: cadre pentru SaMD și ML-based CDSS
  - WHO Guidance (2023–2024): transparență, responsabilitate, fairness
  - Ghiduri profesionale (radiologie, cardiologie, psihiatrie) privind utilizarea AI
  - UNESCO – 2025: Raportul *„Neurotehnologiile și drepturile omului. Un cadru global pentru inovație responsabilă”*

EU Artificial Intelligence Act: Risk levels



- **Rezultate observate până acum**
  - Mai mare atenție la documentație tehnică, audit și validare locală
  - Creșterea rolului „responsible AI teams” în spitale / clinici
  - Îmbunătățirea proceselor de anonimizare și securizare a datelor
  - Evitarea adoptării pripite a sistemelor „opace”

- **Ce se prevede și ce rezultate sunt așteptate**
  - Certificarea obligatorie a AI medicale de risc înalt
  - Trasabilitate și audit continuu pentru modele actualizabile
  - Interoperabilitate și standardizare (FHIR + AI)
  - Responsabilități clar definite pentru furnizori și utilizatori
  - Creșterea încrederii publicului și reducerea erorilor sistemice

- **Ce s-ar mai putea face**
  - Reglementarea clară a generative AI în medicină (încă insuficient abordată)
  - Audit extern obligatoriu pentru modelele mari folosite în sănătate
  - Rețele europene de „AI safety monitoring”
  - Ghiduri personalizate pe specialități (cardio, onco, psihiatrie etc.)

Educația și competențele profesionale – „AI literacy” pentru medici

- **Ce s-a făcut deja**
  - Introducerea AI în curricula medicală în unele universități (UK, Olanda, SUA, România – inițiative locale / module)
  - Cursuri pentru specialiști: radiologie, dermatologie, cardiologie
  - Resurse profesionale: ESMRMB, ESR, AMA – cursuri de AI pentru clinicieni
  - În România: „cursuri” elective în creștere

- **Rezultate până acum**
  - Creșterea conștientizării riscurilor de halucinație / bias
  - Utilizare mai responsabilă a generative AI
  - Mai multă transparență în prezentarea rezultatelor AI
  - Apariția „medicului supervizor AI” (AI-informed doctor)

- **Ce se prevede**
  - AI Literacy devine competență esențială
    (similar cu competențele digitale acum 15 ani)
  - Module obligatorii în rezidențiat
  - Fellowship-uri și certificări în AI medicală
  - Integrarea în training: prompt engineering, interpretarea algoritmilor, XAI
  - Combine clinician + data science în unele specialități
    (radiologie/oncologie)

- **Ce s-ar mai putea face**
  - Elaborarea unei curricula naționale de AI în medicina românească
  - Training pentru profesorii de medicină *(train-the-trainer)*
  - Ghiduri de *„safe AI use for clinicians”*
  - Includerea studiilor de caz reale cu erori AI pentru învățare
  - Educație Medicală Continuă *(obligatorie)* în AI pentru recertificare

Măsuri organizaționale și infrastructură

- **Ce s-a făcut deja**
  - Comisii de AI în unele spitale universitare
  - Sisteme de management al datelor (data governance)
  - Evaluări locale ale algoritmilor înainte de implementare
  - Proiecte pilot pentru AI în imagistică / triaj / administrativ

- **Rezultate observate**
  - Implementare controlată și treptată a AI
  - Evitarea utilizării nevalidate în situații critice
  - Creșterea calității datelor clinice prin standardizare

- **Ce se prevede**
  - Implementarea „AI Safety Offices” în spitale
  - Validare continuă și revalidare la fiecare update de model
  - Infrastructuri sigure pentru date: *data lake-uri* medicale
  - Sisteme hibride: decizii AI + verificare umană

- **Ce s-ar mai putea face**
  - Audit clinic al performanței AI în timp real
  - Dashboard-uri de monitorizare a erorilor AI
  - Definirea rolului de „AI clinical champion” în fiecare secție
  - Simulări și „AI drills” *(similar cu protocoalele de urgență)*

Comparație între direcții de diminuare a riscurilor AI în medicină

Direcție	Avantaje	Limite	Impact
Reglementări	Protecție sistemică; standardizare; responsabilități clare	Lente; birocratice; greu de actualizat; uneori prea generale	Protejează populația; reduce erorile sistemice
Educație	Flexibilă; se adaptează rapid; costuri reduse	Variabilă; depinde de interesul profesioniștilor; nu acoperă integral nevoile	Reduce erorile individuale; îmbunătățește utilizarea responsabilă a AI
Organizații / Infrastructură	Permite implementare responsabilă în spitale; proceduri clare	Necesită resurse, timp și competențe tehnice	Asigură siguranța utilizării AI la nivel de instituție

„În medicină, siguranța nu vine doar din reglementări,
ci din profesioniști bine formați.

Feedbackul dvs:

(1) vă rugăm completați chestionarul
(2) email: **mihalas@umft.ro**