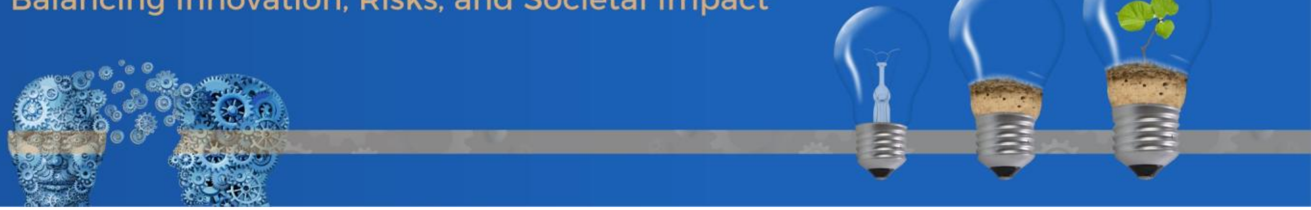
SID 2025
Sibiu Innovation Days
06 - 07 November, Sibiu - RO

EMERGING DISRUPTIVE TECHNOLOGIES:
Balancing Innovation, Risks, and Societal Impact



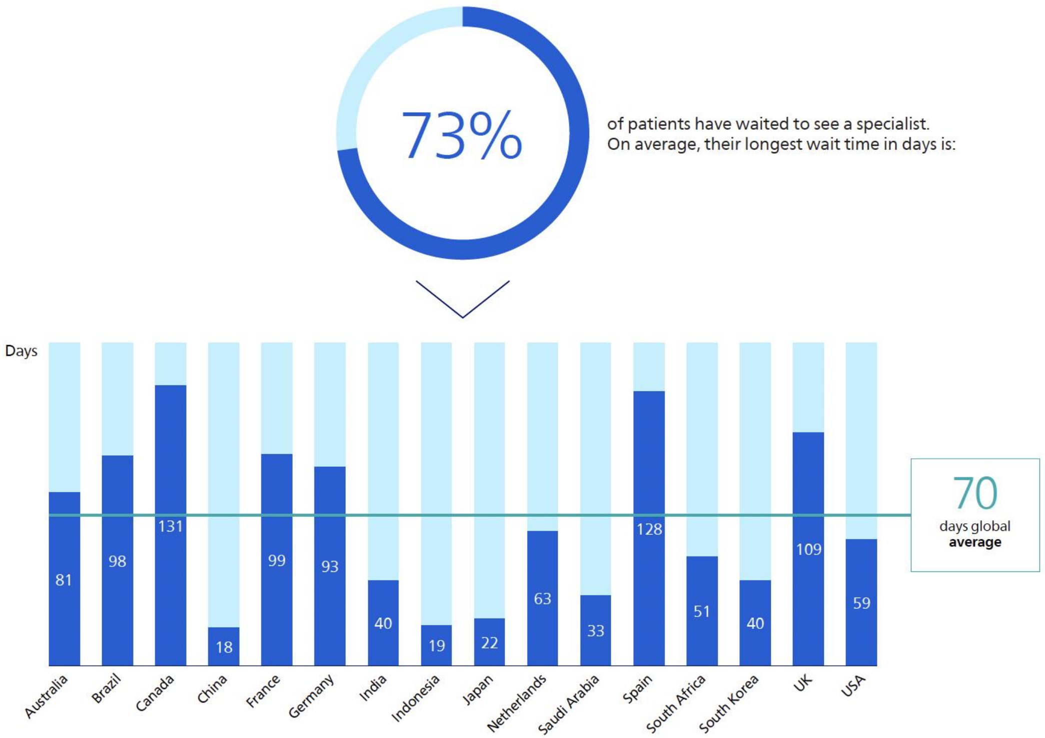
Ethical AI in Healthcare Systems: Where Fairness Meets Clinical Reality

Elena-Anca Paraschiv

Scientific Researcher, National Institute for Research and Development in Informatics - ICI Bucharest, Romania

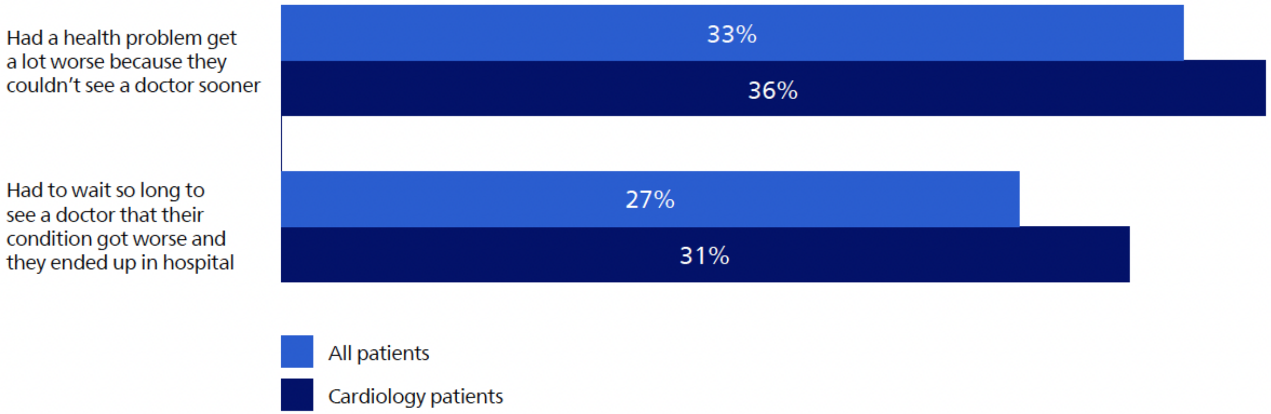
What is the clinical reality?

The clock is ticking: how care delays threaten patient health



of patients have waited to see a specialist. On average, their longest wait time in days is:

Wait times lead to worsening health, especially for cardiology patients



Philips, **Global report - Building trust in healthcare AI:** Perspectives from patients and professionals, 2025
https://www.philips.com/a-w/about/news/future-health-index/reports/2025/building-trust-in-healthcare-ai.html

What is the clinical reality?

Lost hours, lost care: the burden on healthcare professionals

77%

of healthcare professionals have lost clinical time due to issues with incomplete or inaccessible patient data

34%

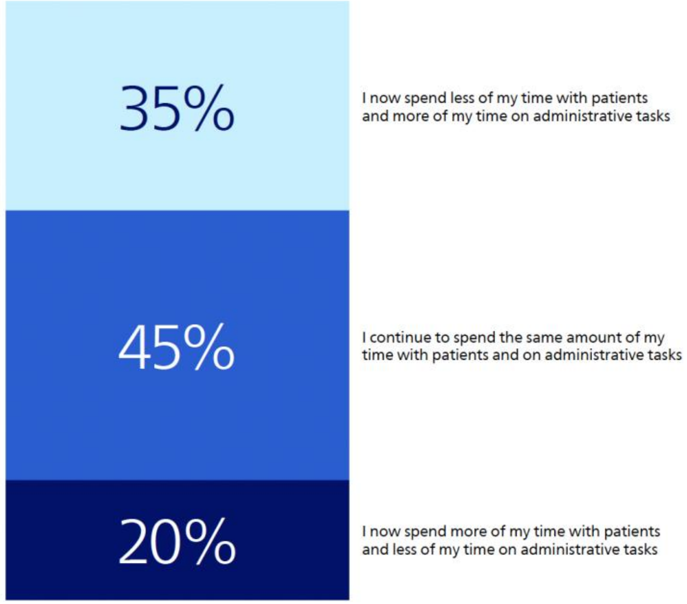
of these healthcare professionals are losing 45+ minutes of clinical time per shift

This equates to:

4+ working weeks

lost in a year per healthcare professional

Healthcare professionals losing patient time to admin



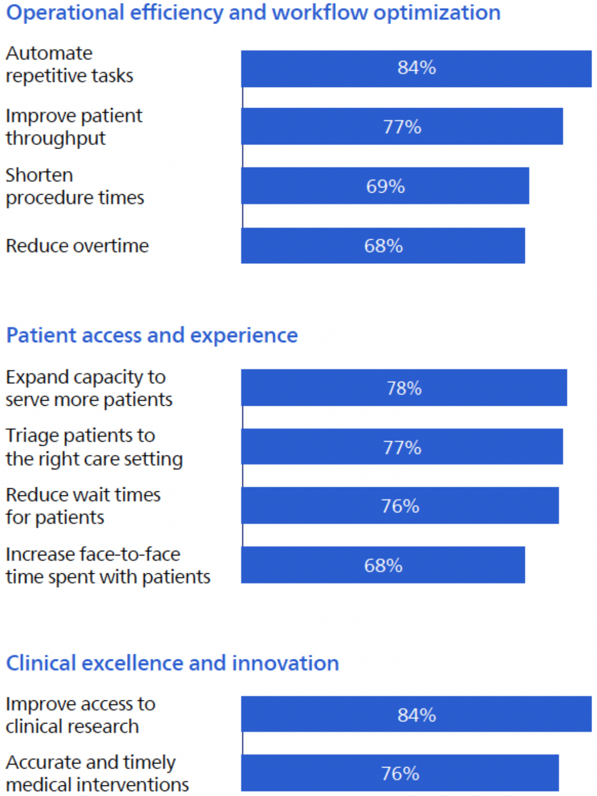
82%

of healthcare professionals say AI and predictive analytics could save lives by enabling early interventions

75%

of healthcare professionals say digital health technologies – including AI and predictive analytics – will reduce hospital admissions in the future

How healthcare professionals say AI can positively impact their department



Philips, Global report - **Building trust in healthcare AI:** Perspectives from patients and professionals, 2025
https://www.philips.com/a-w/about/news/future-health-index/reports/2025/building-trust-in-healthcare-ai.html

AI's roles in healthcare

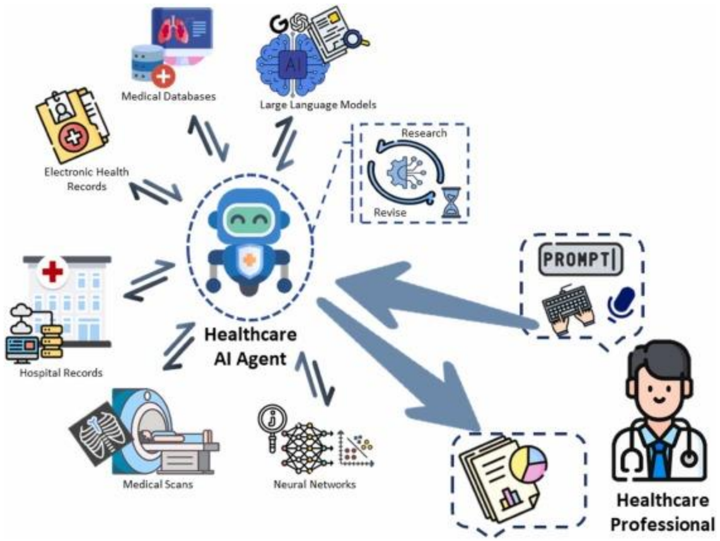
- AI is revolutionizing healthcare by enhancing various aspects such as drug development, disease diagnosis, treatment, patient monitoring, and administrative tasks.

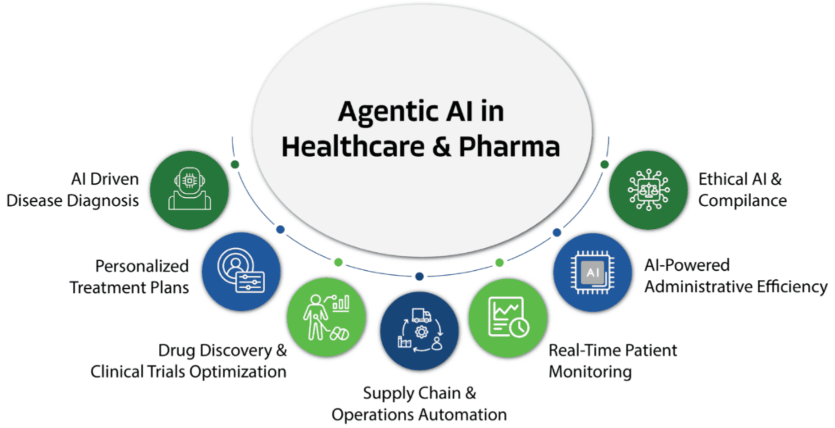
- Notable examples include Google's Med-PaLM, Stanford's CheXNet, and NVIDIA's partnership with Hippocratic AI.

- In addition to the advancements by the private sector, the World Health Organization (WHO) launched S.A.R.A.H. (Smart AI Resource Assistant for Health) in April 2024. This digital health promoter prototype, powered by generative AI, features enhanced empathetic responses in eight languages.

Looking ahead, we can expect a growing trend of collaboration among healthcare companies, technology firms, and research institutions. This synergy will drive further innovations and improvements in healthcare delivery and patient outcomes.

How AI can improve the clinical practice?




Agentic AI in Healthcare & Pharma
AI Driven Disease Diagnosis
Personalized Treatment Plans
Drug Discovery & Clinical Trials Optimization
Supply Chain & Operations Automation
Real-Time Patient Monitoring
AI-Powered Administrative Efficiency
Ethical AI & Compilance

EMORPHIS HEALTH

Transforming Healthcare with Agentic AI

EMORPHIS HEALTH

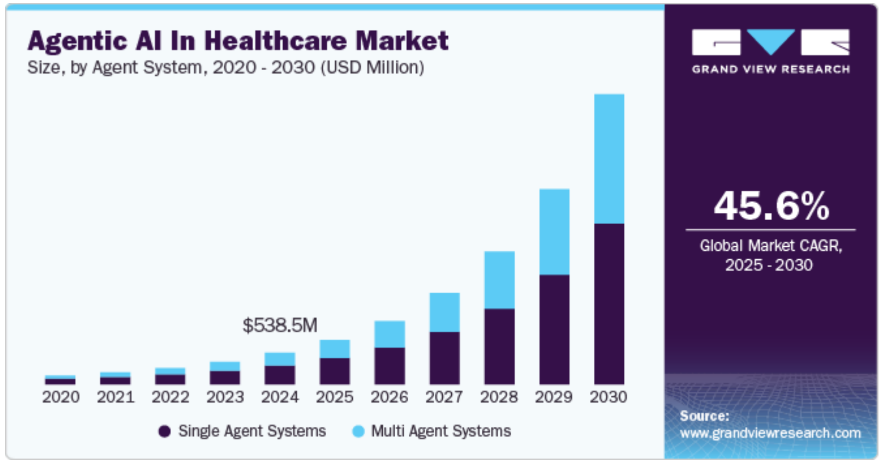
Proactive Care
Anticipates health issues before they escalate, facilitating preventive interventions.

Faster Decision-Making
Provides real-time insights from large datasets, enabling quick and informed decisions.

Personalization at Scale
Tailors healthcare plans to individual needs through continuous learning.

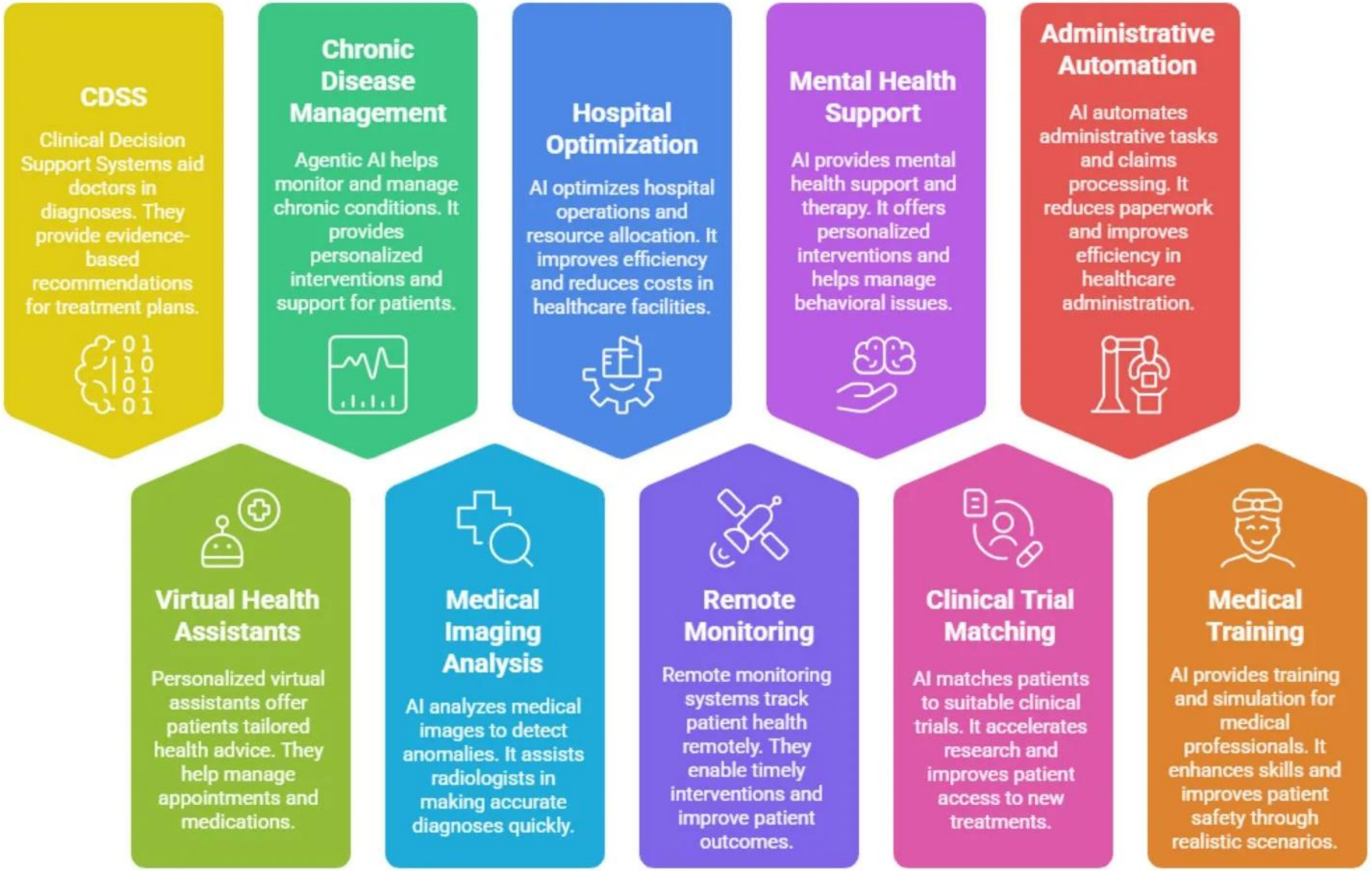
Increased Efficiency
Automates routine tasks, reducing clinician burnout and improving workflow.

24/7 Availability
Ensures constant monitoring and support, as AI agents never sleep.


Agentic AI In Healthcare Market
Size, by Agent System, 2020 - 2030 (USD Million)

Source: www.grandviewresearch.com

How AI can improve the clinical practice?



CDSS

Clinical Decision Support Systems aid doctors in diagnoses. They provide evidence-based recommendations for treatment plans.

Chronic Disease Management

Agentic AI helps monitor and manage chronic conditions. It provides personalized interventions and support for patients.

Hospital Optimization

AI optimizes hospital operations and resource allocation. It improves efficiency and reduces costs in healthcare facilities.

Mental Health Support

AI provides mental health support and therapy. It offers personalized interventions and helps manage behavioral issues.

Administrative Automation

AI automates administrative tasks and claims processing. It reduces paperwork and improves efficiency in healthcare administration.

Virtual Health Assistants

Personalized virtual assistants offer patients tailored health advice. They help manage appointments and medications.

Medical Imaging Analysis

AI analyzes medical images to detect anomalies. It assists radiologists in making accurate diagnoses quickly.

Remote Monitoring

Remote monitoring systems track patient health remotely. They enable timely interventions and improve patient outcomes.

Clinical Trial Matching

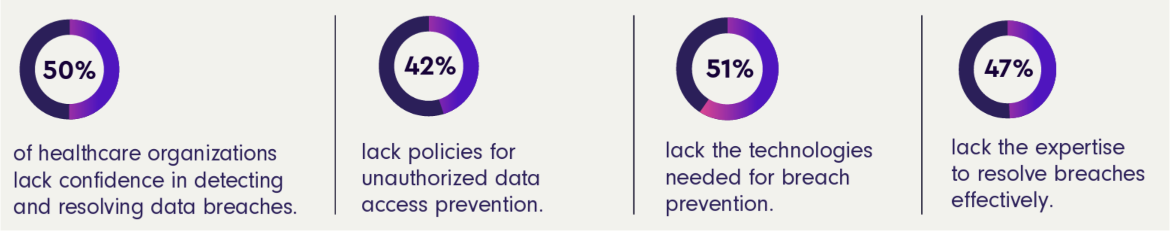
AI matches patients to suitable clinical trials. It accelerates research and improves patient access to new treatments.

Medical Training

AI provides training and simulation for medical professionals. It enhances skills and improves patient safety through realistic scenarios.

Agentic AI in Healthcare, Apps, Benefits, Challenges and Future Trends (2025)
https://emorphis.health/blogs/agentic-ai-in-healthcare/

The state of cybersecurity in healthcare



50% of healthcare organizations lack confidence in detecting and resolving data breaches.

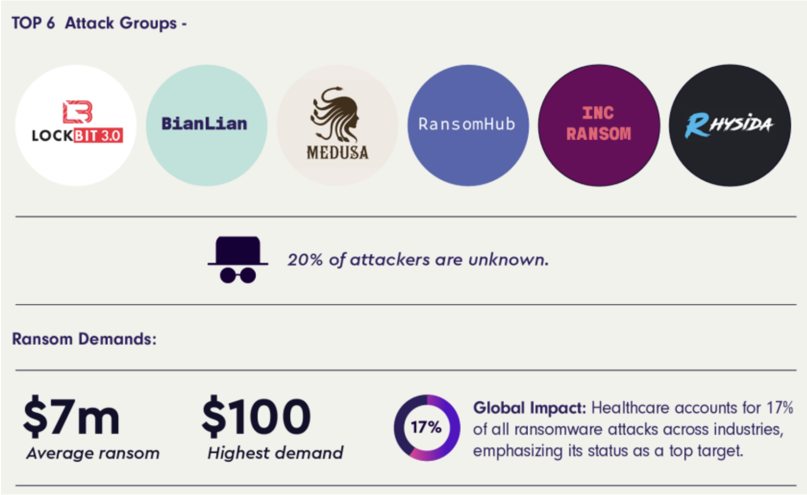
42% lack policies for unauthorized data access prevention.

51% lack the technologies needed for breach prevention.

47% lack the expertise to resolve breaches effectively.

$3.5m - The average cost of a data breach for healthcare organizations

$398 - The average cost per exposed record



TOP 6 Attack Groups -

LOCKBIT 3.0
BianLian
MEDUSA
RansomHub
INC RANSOM
RHYSIDA

20% of attackers are unknown.

Ransom Demands:

$7m *Average ransom*

$100 *Highest demand*

17% **Global Impact:** Healthcare accounts for 17% of all ransomware attacks across industries, emphasizing its status as a top target.

The state of cybersecurity in healthcare 2025. A Veriti Research Report, veriti.ai

Cybersecurity Risks in Agentic and Multi-Agent AI Systems

Adversarial attacks: AI agents can be tricked by carefully crafted inputs that are designed to mislead them.

This poses critical risks such as:

- Delivering incorrect treatment recommendations;
- Suppressing vital alerts (e.g., early signs of sepsis or cardiac arrest);
- Misleading clinicians with fabricated or irrelevant data.

Data poisoning: If attackers gain access to the training or retraining pipeline, they can inject false or biased data that causes an AI agent to learn incorrect patterns over time.

This can result in:

- Systemic bias in clinical decisions
- Agents that slowly degrade in performance while appearing normal
- Faulty risk scoring or triage escalation paths

Agent impersonation or hijacking: If security is weak, malicious actors can take control of or impersonate an agent within the network.

Such attacks can lead to:

- Cancellation or rescheduling of critical procedures
- Authorization of medications with known contraindications
- Silent failure of alerts or risk escalation triggers

Insecure communication between agents: Multi-agent systems rely heavily on constant communication—data is exchanged between agents through APIs, message queues, or local networks. If these are not encrypted or authenticated properly, they become prime targets for interception or manipulation.

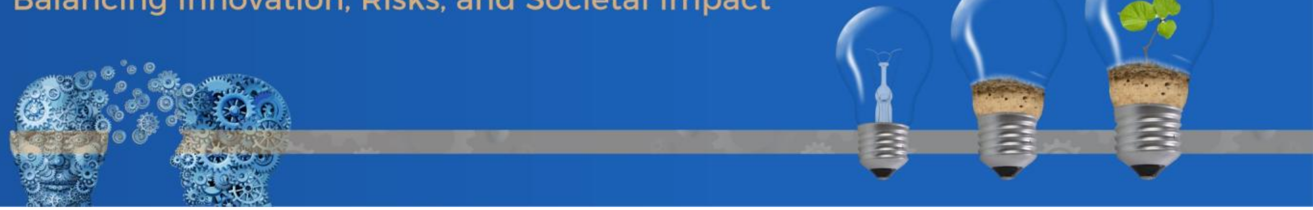
Consequences include:

- Leaking of sensitive patient data (violating HIPAA/GDPR)
- Injection of false data during agent handoffs
- Loss of integrity in collaborative clinical workflows

What Responsible AI looks like?

The question of *“what is responsible AI?”* is one being asked about every application of artificial intelligence. But when it comes to responsible AI in healthcare, at its core it must be:

- **Fair**: Designed to reduce, not reinforce, health disparities.
- **Accountable**: Containing clear lines of responsibility for decisions influenced by AI.
- **Transparent**: Systems must be explainable so clinicians and patients can understand how conclusions are reached.
- **Safe**: Built with rigorous testing and continuous monitoring to minimize harm.
- **Alignment with Human Values:** Ensuring AI systems operate in ways that enhance human well-being.



What Responsible AI looks like?

Examples of Responsible AI in Practice

Some institutions are already developing frameworks for responsible AI in healthcare.

- The World Health Organization (WHO) offers guidance emphasizing human oversight, inclusivity, and data privacy.
- STANDING Together is an initiative funded by the U.K.'s NHS AI Lab that developed recommendations for transparency of AI datasets.
- And to help uncover errors in clinical trials, the SPIRIT-AI and CONSORT-AI extensions are reporting guidelines for protocols with an AI component.

But there is still a gap between AI's use in hospitals and its oversight. A study published in 2025 found that 65% of U.S. hospitals used predictive models, but only 44% reported evaluating for bias.

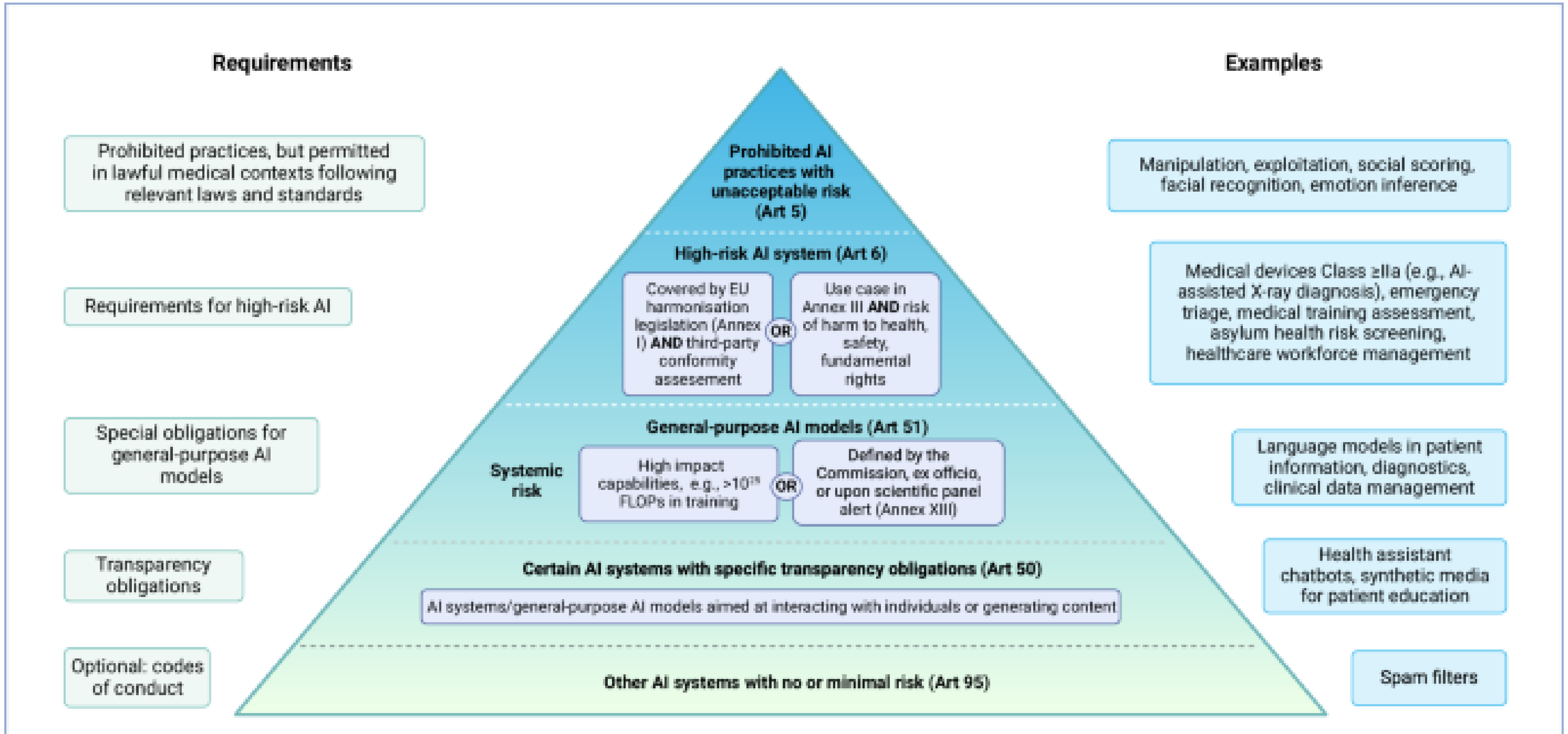
Legal frameworks governing AI in healthcare

- **The US Food and Drug Administration (FDA)** has recently issued several discussion papers on AI drug development and manufacturing medical devices and guidance on decentralized clinical trials.[3] FDA generally supports the use of AI in healthcare development and has already **reviewed and authorized over 1200 AI/Machine Learning (ML)-enabled medical devices**.[4]

- **The EU AI Act** is recognized as **the world's first comprehensive AI law**. Although most of its requirements will only come into effect from August 1, 2026, and pure research and development AI is excluded from much of its scope, the Act imposes regulatory requirements on AI systems based on four risk categories: **(1) prohibited AI, (2) high risk AI, (3) AI triggering transparency requirements, and (4) general-purpose AI**. In the context of healthcare, the middle two categories—"high risk AI" and "AI triggering transparency requirements"—are likely to be the most relevant. These categories will impose specific regulatory obligations to ensure the safe and ethical use of AI in healthcare applications.

- **General purpose AI models (GPAIM)** is also used for many hundreds of different use cases, across R&D and corporate functions. This is typically by way of customizing large language models using proprietary data. As such, the industry has been calling out for clarification regarding the extent to which such bespoke deployment of GPAIM will engage the specific EU AI Act obligations (applying from August 2, 2025).

EU AI Act in healthcare



Requirements

Prohibited practices, but permitted in lawful medical contexts following relevant laws and standards

Requirements for high-risk AI

Special obligations for general-purpose AI models

Transparency obligations

Optional: codes of conduct

Prohibited AI practices with unacceptable risk (Art 5)

High-risk AI system (Art 6)

Covered by EU harmonisation legislation (Annex I) **AND** third-party conformity assesement **OR** Use case in Annex III **AND** risk of harm to health, safety, fundamental rights

General-purpose AI models (Art 51)

Systemic risk: High impact capabilities, e.g., $>10^{25}$ FLOPs in training **OR** Defined by the Commission, ex officio, or upon scientific panel alert (Annex XIII)

Certain AI systems with specific transparency obligations (Art 50)

AI systems/general-purpose AI models aimed at interacting with individuals or generating content

Other AI systems with no or minimal risk (Art 95)

Examples

Manipulation, exploitation, social scoring, facial recognition, emotion inference

Medical devices Class ≥IIa (e.g., AI-assisted X-ray diagnosis), emergency triage, medical training assessment, asylum health risk screening, healthcare workforce management

Language models in patient information, diagnostics, clinical data management

Health assistant chatbots, synthetic media for patient education

Spam filters

https://www.nature.com/articles/s41746-024-01213-6

Ethical considerations

Data Privacy and Security

Patient Confidentiality: Ensuring that AI systems protect patient data and adhere to privacy laws like Health Insurance Portability and Accountability Act (HIPAA) in United States and General Data Protection Regulation (GDPR) in EU.
Data Security Measures: Implementing robust cybersecurity protocols to prevent data breaches and unauthorized access.
Anonymization: Techniques to anonymize patient data while maintaining its utility for AI analysis.

Bias and Fairness

Algorithmic Bias: Addressing biases in AI algorithms that can lead to unequal treatment outcomes across different patient demographics.
Inclusive Data Sets: Ensuring that training data for AI models is diverse and representative of the entire patient population.
Fairness Audits: Regularly conducting audits to assess and mitigate biases in AI systems.

Transparency and Accountability

Explainability: Making AI decisions understandable to healthcare providers and patients to ensure trust and acceptance.
Accountability Frameworks: Defining clear accountability for AI-driven decisions and outcomes, including legal and professional responsibilities.
Informed Consent: Ensuring patients are fully informed about the use of AI in their care and obtaining their consent.

Ethical considerations – EU AI Act

● High-Risk Classification:

- AI systems used in healthcare are explicitly classified as “high-risk” under the EU AI Act.
- This means AI tools used for:
  - ➢ Monitoring vital signs
  - ➢ Predicting deterioration (e.g., sepsis, AKI, cardiac arrest)
  - ➢ Recommending treatments or interventions

  are subject to strict safety, quality, and transparency requirements.

● Risk Management and Compliance:

- Providers must perform risk assessments and develop mitigation strategies before using AI in critical care.
- The system must comply with cybersecurity, bias prevention, and patient safety standards.

● Transparency and Explainability:

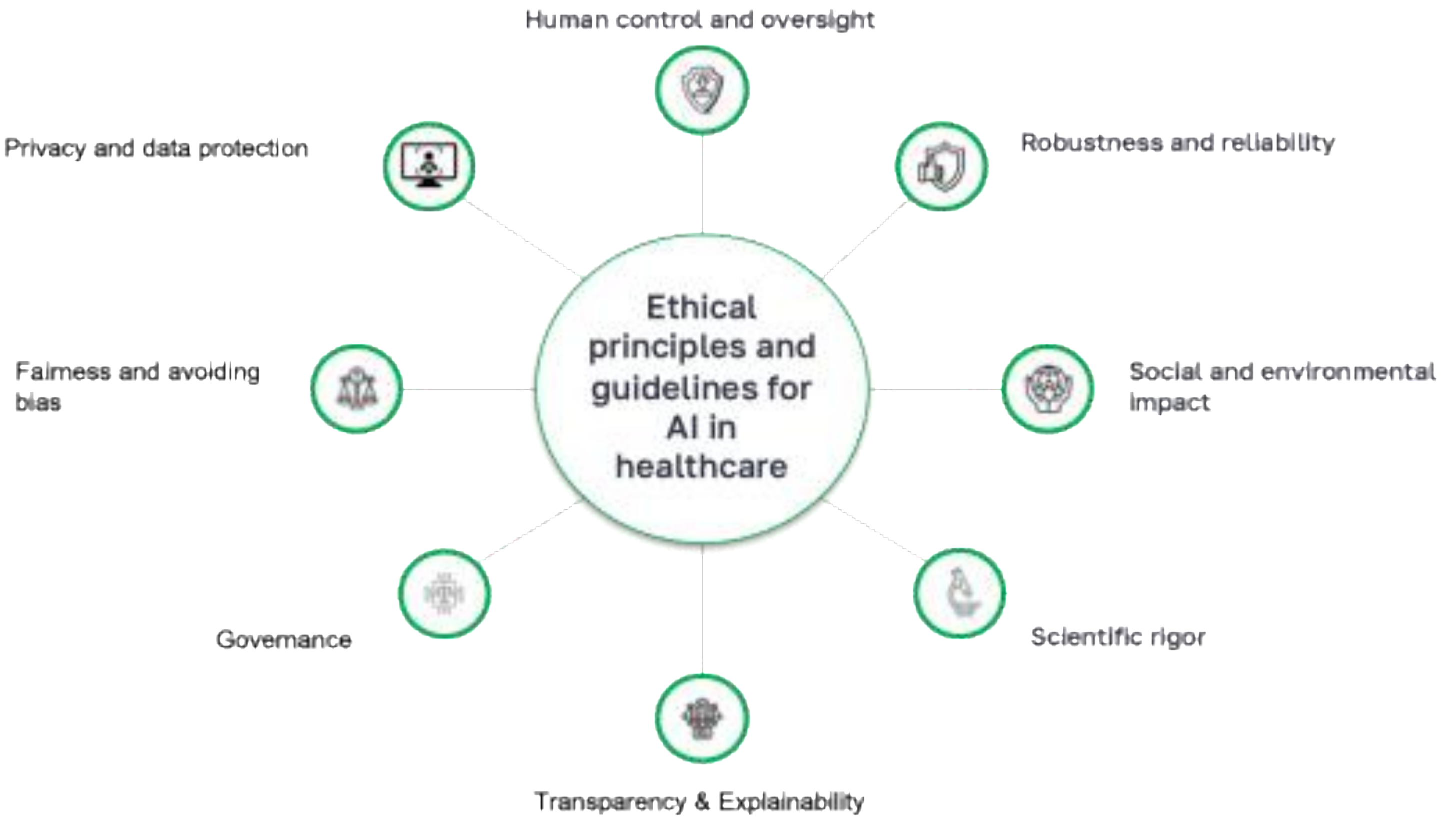
- AI systems must be interpretable to clinical staff.
- The Act mandates that healthcare professionals using AI must:
  - ➢ Be informed about how the system works
  - ➢ Be able to understand and challenge its recommendations

● Human Oversight:

- The Act emphasizes "meaningful human oversight" — clinicians must remain in control.
- The AI system cannot make autonomous decisions that bypass or replace clinician judgment.

AI must act as a support tool, not a replacement.

Building secure and ethical AI in Healthcare



Ethical principles and guidelines for AI in healthcare

Human control and oversight

Robustness and reliability

Social and environmental impact

Scientific rigor

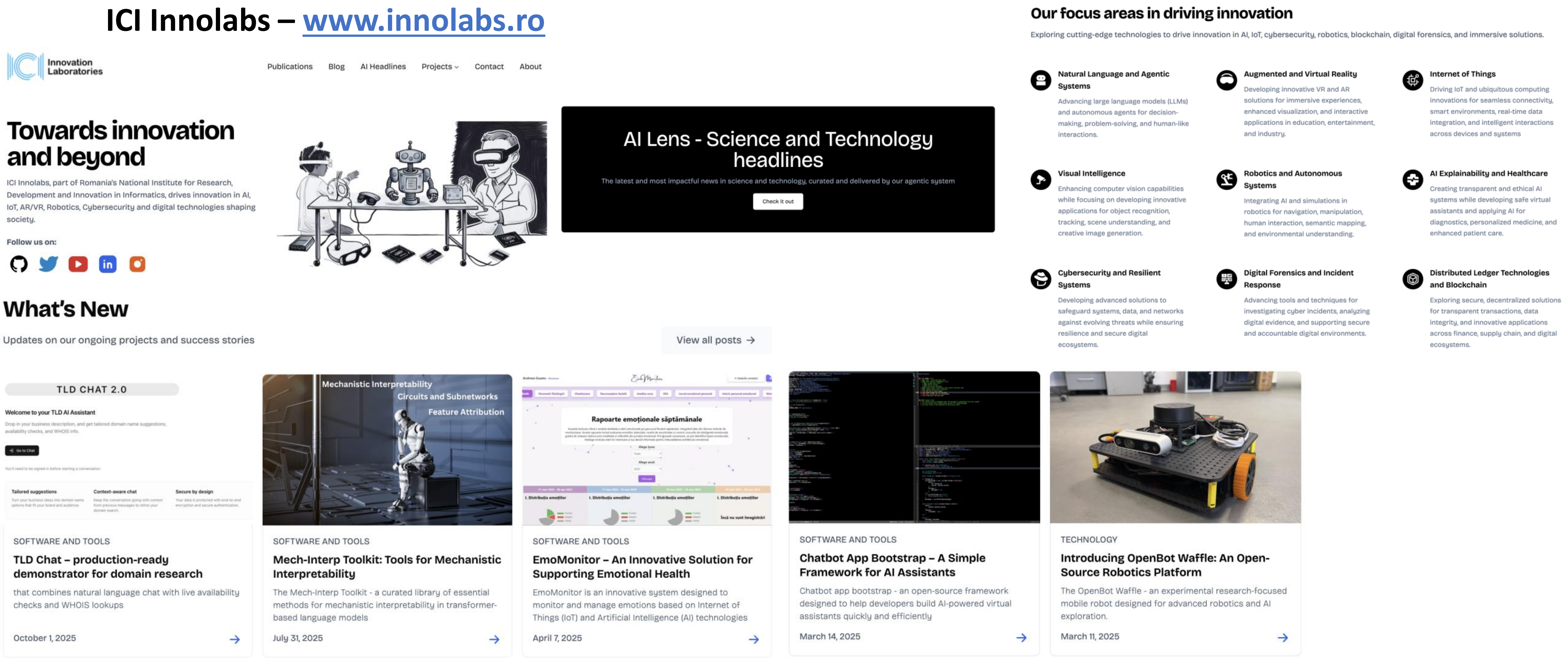
Transparency & Explainability

Governance

Fairness and avoiding bias

Privacy and data protection

ICI Innolabs – www.innolabs.ro



Innovation Laboratories

Publications · Blog · AI Headlines · Projects ⌄ · Contact · About

Towards innovation and beyond

ICI Innolabs, part of Romania's National Institute for Research, Development and Innovation in Informatics, drives innovation in AI, IoT, AR/VR, Robotics, Cybersecurity and digital technologies shaping society.

Follow us on:

AI Lens - Science and Technology headlines

The latest and most impactful news in science and technology, curated and delivered by our agentic system

Check it out

Our focus areas in driving innovation

Exploring cutting-edge technologies to drive innovation in AI, IoT, cybersecurity, robotics, blockchain, digital forensics, and immersive solutions.

Natural Language and Agentic Systems
Advancing large language models (LLMs) and autonomous agents for decision-making, problem-solving, and human-like interactions.

Augmented and Virtual Reality
Developing innovative VR and AR solutions for immersive experiences, enhanced visualization, and interactive applications in education, entertainment, and industry.

Internet of Things
Driving IoT and ubiquitous computing innovations for seamless connectivity, smart environments, real-time data integration, and intelligent interactions across devices and systems

Visual Intelligence
Enhancing computer vision capabilities while focusing on developing innovative applications for object recognition, tracking, scene understanding, and creative image generation.

Robotics and Autonomous Systems
Integrating AI and simulations in robotics for navigation, manipulation, human interaction, semantic mapping, and environmental understanding.

AI Explainability and Healthcare
Creating transparent and ethical AI systems while developing safe virtual assistants and applying AI for diagnostics, personalized medicine, and enhanced patient care.

Cybersecurity and Resilient Systems
Developing advanced solutions to safeguard systems, data, and networks against evolving threats while ensuring resilience and secure digital ecosystems.

Digital Forensics and Incident Response
Advancing tools and techniques for investigating cyber incidents, analyzing digital evidence, and supporting secure and accountable digital environments.

Distributed Ledger Technologies and Blockchain
Exploring secure, decentralized solutions for transparent transactions, data integrity, and innovative applications across finance, supply chain, and digital ecosystems.

What's New

Updates on our ongoing projects and success stories

View all posts →

TLD CHAT 2.0

Welcome to your TLD AI Assistant

Drop in your business description, and get tailored domain name suggestions, availability checks, and WHOIS info.

Go to Chat

You'll need to be signed in before starting a conversation.

Tailored suggestions Turn your business ideas into domain name options that fit your brand and audience.

Context-aware chat Keep the conversation going with context from previous messages to refine your domain search.

Secure by design Your data is protected with end-to-end encryption and secure authentication.

SOFTWARE AND TOOLS

TLD Chat – production-ready demonstrator for domain research

that combines natural language chat with live availability checks and WHOIS lookups

October 1, 2025 →

SOFTWARE AND TOOLS

Mech-Interp Toolkit: Tools for Mechanistic Interpretability

The Mech-Interp Toolkit - a curated library of essential methods for mechanistic interpretability in transformer-based language models

July 31, 2025 →

SOFTWARE AND TOOLS

EmoMonitor – An Innovative Solution for Supporting Emotional Health

EmoMonitor is an innovative system designed to monitor and manage emotions based on Internet of Things (IoT) and Artificial Intelligence (AI) technologies

April 7, 2025 →

SOFTWARE AND TOOLS

Chatbot App Bootstrap – A Simple Framework for AI Assistants

Chatbot app bootstrap - an open-source framework designed to help developers build AI-powered virtual assistants quickly and efficiently

March 14, 2025 →

TECHNOLOGY

Introducing OpenBot Waffle: An Open-Source Robotics Platform

The OpenBot Waffle - an experimental research-focused mobile robot designed for advanced robotics and AI exploration.

March 11, 2025 →

Conclusions

- **AI in healthcare must do more than perform; it must *behave* responsibly.**
  The true challenge is not technical, but ethical: ensuring that innovation amplifies fairness, not inequity.

- **Trust is the foundation of adoption.**
  Every algorithm that enters the clinic must respect privacy, accountability, and transparency — not as compliance checkboxes, but as moral commitments to patients.

- **The future of healthcare is hybrid = human and intelligent.**
  When fairness meets clinical reality, technology becomes an extension of care itself.

Thank you for your attention!

Any questions?

elena.paraschiv@ici.ro