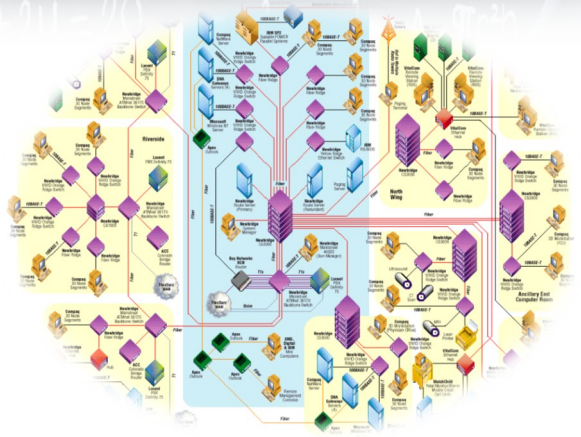




# Temi regolatori e problematiche da affrontare nell'impiego della AI in Sanità (IA, Big Data, Infrastrutture, e Regolamentazione) - COD. 21

*Ing. Mauro Grigioni*

*Commissione Sistemi Informativi Sanitari – Ordine degli Ingegneri di Roma*

*già Direttore del Centro Nazionale Tecnologie Innovative in Sanità Pubblica, ISS*

# Indice

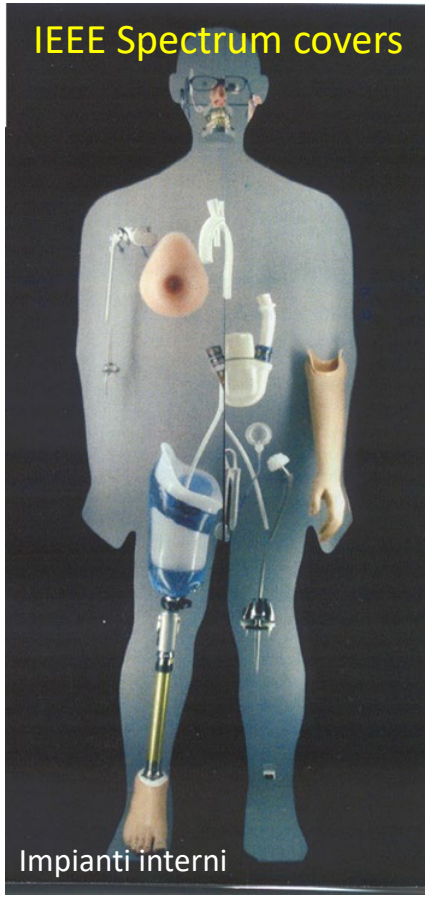
*Prefazione :*

*Una visione critica dei concetti tecnico-normativi ed etico-legislativi*


- Introduzione per il contesto attuale (Slide 3 - 8)
- Il mondo dei Dispositivi Medici, MDR 2017/745 e le normative, in particolare per i Software (Slides 9 - 38)
- ML/DL e ecosistemi digitali (Slides 39 – 78)
- Etica dell'IA - Orientamenti (82 – 89 )
- AI Act o RIA (Slides 86 – 125)

# Innovazione ubiquitaria e digitale nelle Tecnologie Sanitarie parallela a quella sociale

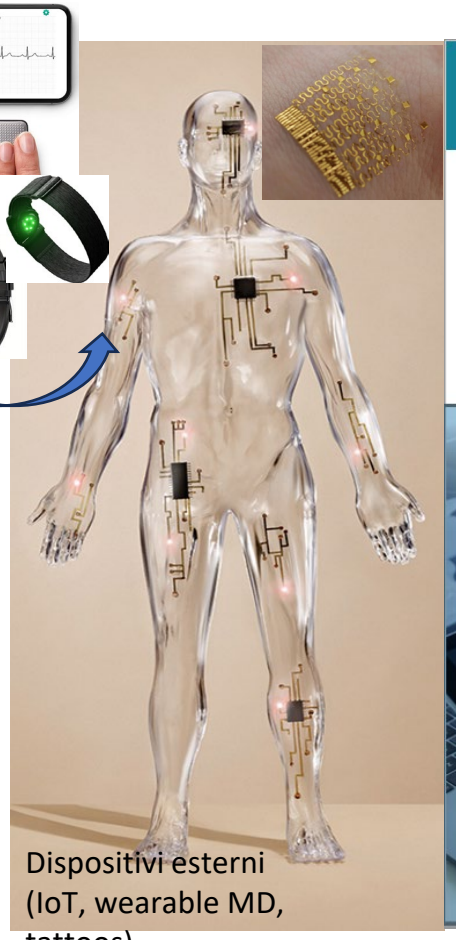
By the mid-1960s, it became clear that the provisions of the FFD&C Act were not adequate to regulate the complex medical devices of the times to ensure both patient and user safety. Thus, in 1969, the Cooper Committee was formed to examine the problems associated with medical devices and to develop concepts for new regulations.



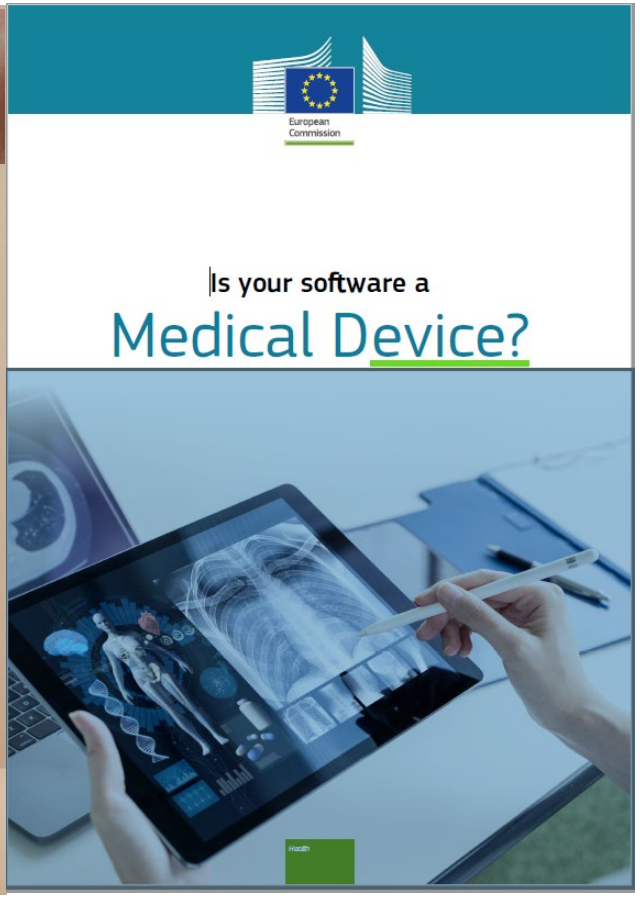

Impianti interni



Domotica e app



Dispositivi esterni (IoT, wearable MD, tattoos)




*Sviluppo tecnologico dagli anni 70 in poi*

# CONTESTO - TECNOLOGIE SANITARIE denominate  Dispositivi Medici

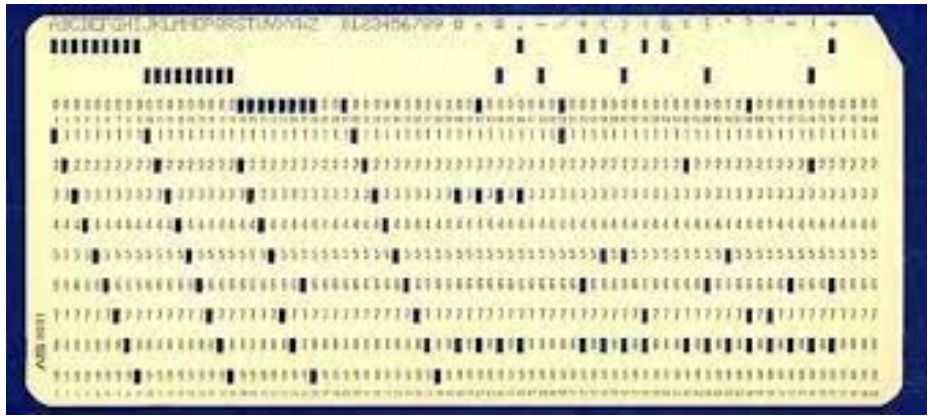
- Dispositivi Medici (sono oltre 300.000 classi di prodotti)

- Esempi : dal **cerotto** agli **elettromedicali**, dagli **impiantabili** fino alle **grandi apparecchiature**, dai **software** complessi alle **App**

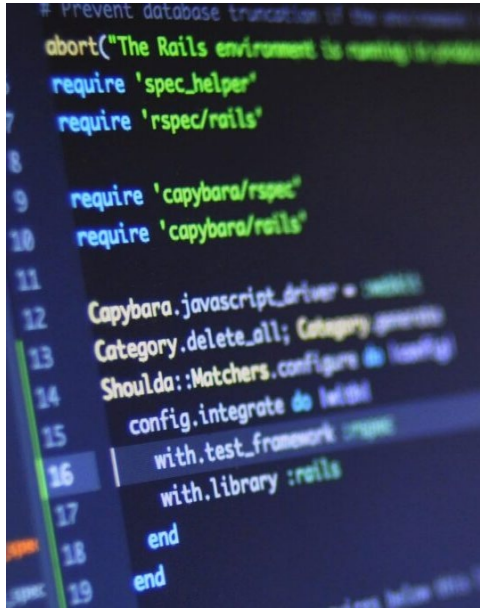
- Pervasività:
- sostituiscono la sola terapia medica ove è dimostrata la superiorità (ad es interventistica),
- completano un protocollo assieme alle terapie mediche opportunamente ristudiate, grazie a grandi studi (grandi dati = Big Data), (Stent Medicati)
- sono sistemi di autocura/automonitoraggio, per la domiciliazione
- Possiamo aggiungere tra le tecnologie in Sanità le recenti innovazioni digitali a complemento della conoscenza e del progresso delle soluzioni Diagnostico-Terapeutiche (non tutte sono DM, ad es. Big Data, BlockChain, Telemedicina (IoT etc…))

http://www.medtecheurope.org/

# Cosa sono i SW?








- Il termine sarebbe stato creato durante la seconda guerra mondiale; tecnici del Royal Army britannico erano impegnati nella **decrittazione dei codici tedeschi di Enigma**, di cui già conoscevano la meccanica interna (detta hardware, componente dura, nel senso di ferraglia) grazie ai servizi segreti polacchi. La prima versione di Enigma sfruttava tre rotori per mescolare le lettere. Dopo il 1941, ad Enigma venne aggiunto un rotore, e il team di criptoanalisti inglesi, capitanati da Alan Turing, si dovette **interessare non più alla sua struttura fisica, ma alle posizioni in cui venivano utilizzati i rotori della nuova Enigma**. Dato che queste istruzioni erano scritte su pagine solubili nell'acqua (per poter essere più facilmente distrutte, evitando in tal modo che cadessero nelle mani del nemico) furono chiamate **software** (componente tenera), in contrapposizione all'hardware.

- Il significato moderno del termine lo si deve alle istruzioni date ai computer. Dal 1950 in poi l'analogia tra l'hardware ed il corpo umano e quella tra il software e la mente umana è stabilizzata.
- Anche **Turing** ha sostenuto che il progresso tecnologico sarebbe riuscito a creare, entro il 2000, delle macchine intelligenti (in grado cioè di «pensare» autonomamente) atte alla risoluzione dei problemi.

*By Wikipedia*



# SSN Digitale

→ *Informazione*

- La ricerca biomedica e la clinica sono/saranno ad **alta intensità di dati** e l'obiettivo di raggiungere l'analisi dei dati a scala di popolazione renderà le cose ancora più difficili.

→ *Big Data*

- Il raggiungimento di questo obiettivo richiederà una **comprensione semantica comune tra l'analisi dei dati e i sistemi che contengono i dati**, molto probabilmente biobanche data-centriche profondamente integrate con il loro ente di ricerca clinica di riferimento/istituto sanitario.

→ *Interoperabilità*

- Le principali **interfacce** tra questi due sistemi riguardano lo scambio di dati con i dettagli sulla loro provenienza e le informazioni **EHR**. Entrambi hanno esigenze critiche in termini di espressività semantica e standardizzazione

→ *Normazione*

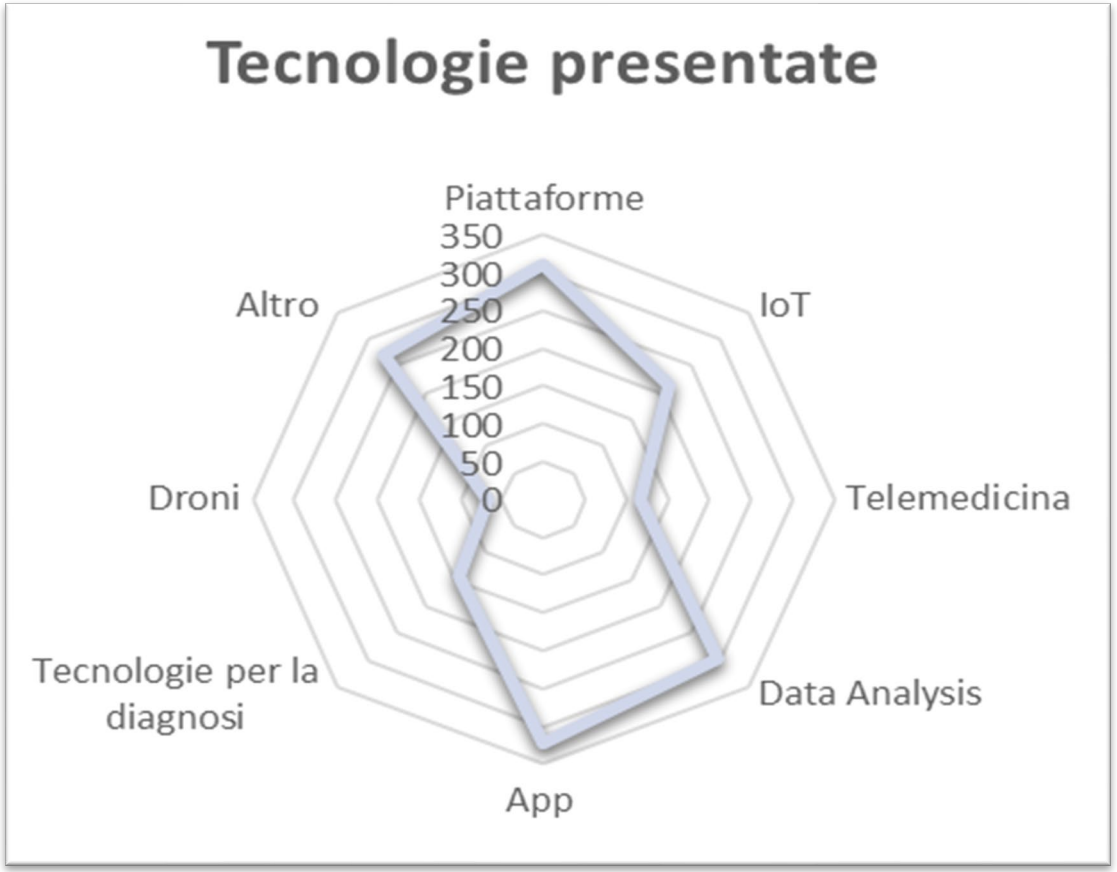
# SSN Digitale

https://digitalhealth.london/wp-content/uploads/2018/04/DigitalTherapeuticsNHS.pdf






# Il mondo Digitale al tempo del Covid:
## OFFERTA DI NUOVE TECNOLOGIE ORIENTATE AL COVID PER IL «MID»



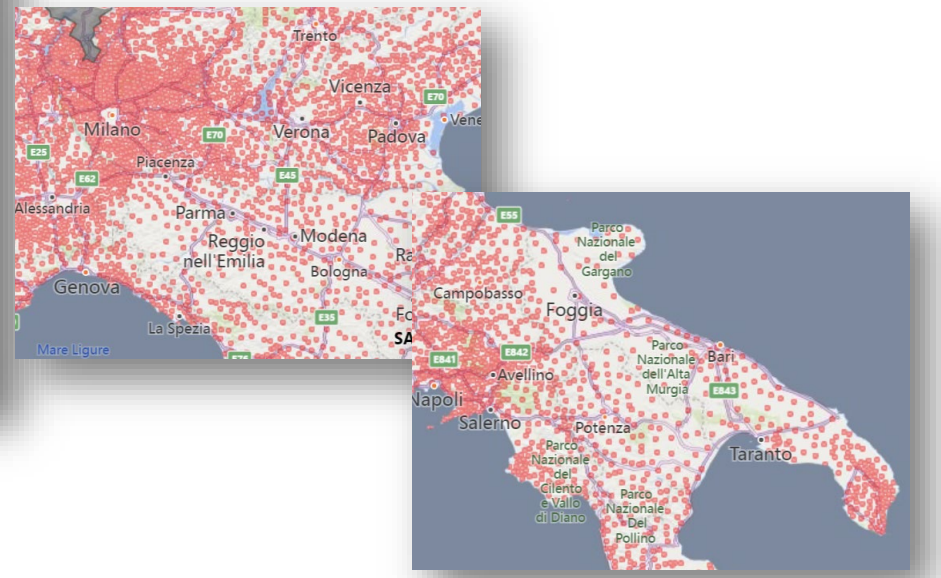

- **Elementi:**
  - Cartella Clinica Elettronica (**EHR**)
  - Sicurezza dei dati e delle informazioni
  - GDPR Compliant
  - Supporto informatico diffuso (Cloud)
  - Sistemi **GIS** (continuità assistenziale) RealTime
  - Diagnostica per immagini e di laboratorio
  - Intelligenza Artificiale (per lo più **Chat Bot**)
  - Interoperabilità dei Sistemi Informativi
  - Gestione informatizzata dei farmaci
  - Fascicolo Sanitario Elettronico (FSE)
  - Sistemi di Business Intelligence
  - Firma Digitale
  - **Digital twin**

*Decreto Presidenza ISS per il GdL : Tecnologie Sanitarie a supporto del contrasto al Covid – 19 Responsabile M. Grigioni*

# Salute e Territorio nel PNRR: la rete dell'assistenza nella Missione 6




DM 77 Allegato 1
…**farmacie convenzionate** con il SSN ubicate uniformemente sull'intero territorio nazionale, costituiscono presidi sanitari di prossimita' e rappresentano un elemento fondamentale ed integrante del Servizio sanitario nazionale




*Standard organizzativi - Da Domenico Mantoan Direttore Generale AGENAS*

# Dispositivi medici in Europa: lo sviluppo del quadro regolatorio (cogente)

- *Direttiva 90/385/CEE* (dispositivi medici impiantabili attivi)
- *Direttiva 93/42/CEE* (dispositivi medici)
- *Direttiva 98/79/CE* (IVD, diagnostici in vitro)
- *Direttiva 2001/83/CE* (medicinali per uso umano)

- *Revisione della Direttiva DM: Direttiva 2007/47/CE (inserito SOFTWARE)*

- Medical Device Regulation 2017/745 (DM) .../746 (IVDM)

- Regulation (EU) 2023/607 del 15 March 2023, (provides **medical device** manufacturers more time to certify their **medical devices)**

- AI Act - Regolamento 2024/1689 del 13 giugno 2024

# Standards

- CEI EN 62304 software per dispositivi medici - processi del ciclo di vita del software

- ISO/IEC 25012-24 sulla qualità dei dati, *(indicatori misurabili)*
- qualità del software ISO/IEC 25010-22-23
- qualità dei servizi ISO/IEC TS 25011-25
- ISO/IEC 25059:2023 - Software engineering: Systems and software Quality Requirements and Evaluation (SQuaRE); Quality model for AI systems)
- ISO/IEC 42001 Information Technology – Artificial intelligence – Management system.

*Riferimenti per l'industria e per le istituzioni.....!!!!!*

La valutazione del prodotto si situa in un punto focale delle fasi di applicazione dell'ISO/IEC 25000. A tal fine può essere di ausilio l'ISO/IEC 25040 "**Evaluation process**", tenendo conto dei modelli di qualità e delle misurazioni, divenuto norma nazionale nel 2021 **UNI CEI ISO/IEC 25040:2021**.

# EU AI Act standardization request

The EU AI Act was developed as a horizontal, product-oriented regulation, and the request for harmonized standards within the EU AI Act marks a unique approach in which standards adherence is an element of conformity with the Regulation. Although EU harmonized standards are voluntary, the development of standards aligned with mandatory legal requirements can help mitigate uncertainties surrounding definitions and interpretations. Historically, these ambiguities have required clarification through legal proceedings.

Ten harmonized standards are being developed by **CEN-CENELEC** in collaboration with the European Telecommunications Standards Institute (ETSI) as part of the EU AI Act standardization request. These include:

- Risk management for AI systems
- Governance and quality of datasets used to build AI systems
- Record keeping through logging capabilities by AI systems
- Transparency and information provisions for users of AI systems
- Human oversight of AI systems
- Accuracy specifications for AI systems
- Robustness specifications for AI systems
- Cybersecurity specifications for AI systems
- Quality management systems for providers of AI systems, including post-market monitoring processes
- Conformity assessment for AI systems

# (MDR) Articolo 105 - Compiti dell'MDCG

Ai sensi del presente regolamento, l'MDCG ha i seguenti compiti:

a) contribuire alla valutazione degli organismi di valutazione della conformità e degli organismi notificati che hanno presentato una domanda, conformemente alle disposizioni di cui al capo IV;

b) consigliare la Commissione, su sua richiesta, nelle questioni riguardanti il gruppo di coordinamento degli organismi notificati istituito a norma dell'articolo 49;

c) contribuire allo sviluppo di orientamenti volti a garantire un'attuazione efficace e uniforme del presente regolamento, in particolare per quanto riguarda la designazione e la sorveglianza degli organismi notificati, l'applicazione dei requisiti generali di sicurezza e prestazione e lo svolgimento delle valutazioni cliniche e delle indagini cliniche effettuate dai fabbricanti, della valutazione effettuata dagli organismi notificati e delle attività di vigilanza;

d) contribuire a monitorare costantemente il progresso tecnologico e a valutare se i requisiti generali di sicurezza e prestazione stabiliti dal presente regolamento e dal regolamento (UE) 2017/746 sono adeguati per garantire la sicurezza e la prestazione dei dispositivi e contribuire in tal modo a identificare se è necessario modificare l'allegato I del presente regolamento;

e) contribuire all'elaborazione di norme sui dispositivi, di SC e di orientamenti scientifici, inclusi gli orientamenti specifici per dispositivo, sulle indagini cliniche di taluni dispositivi, in particolare i dispositivi impiantabili e quelli appartenenti alla classe III;

f) assistere le autorità competenti degli Stati membri nelle loro attività di coordinamento in particolare in materia di classificazione e determinazione dello status normativo dei dispositivi, indagini cliniche, vigilanza e sorveglianza del mercato, compresi l'elaborazione e il mantenimento di un impianto per il programma europeo di sorveglianza del mercato al fine di conseguire l'efficienza e armonizzazione della sorveglianza del mercato dell'Unione, ai sensi dell'articolo 93;

g) fornire consulenza, di propria iniziativa o su richiesta della Commissione, nella valutazione di tutte le questioni relative all'applicazione del presente regolamento;

h) contribuire allo sviluppo di pratiche amministrative armonizzate nel settore dei dispositivi negli Stati membri.

# Prodotti per Digital Health (MDCG)

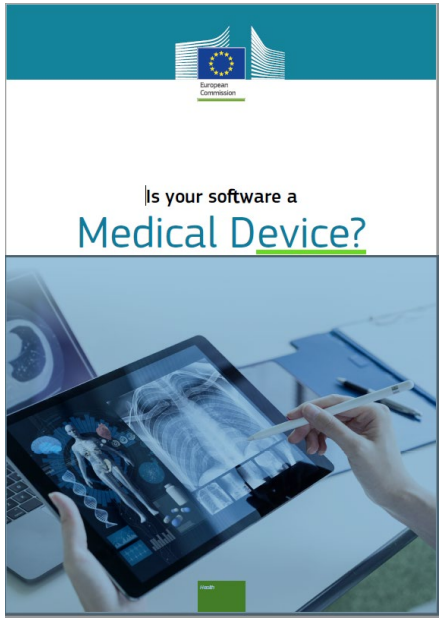
La categorizzazione dei prodotti per Digital Health può avvalersi di strumenti decisionali adottati a livello comunitario

## MDCG 2019-11

**Guidance on Qualification and Classification of Software in Regulation (EU) 2017/745 – MDR and Regulation (EU) 2017/746 – IVDR October 2019**

This document, which primarily targets medical software manufacturers, defines the criteria for the qualification of software falling within the scope of the new medical devices regulations and provides guidance on the application of classification criteria for software under Regulation (EU) 2017/745 –MDR and Regulation (EU) 2017/746 – IVDR.

[…] The criteria specified in this document shall also apply to applications (commonly referred to as apps), **may they be operating on a mobile phone, in the cloud or on other platforms.**



# Medical Device SW: Documenti di riferimento EU

- Il documento **MDGC 2019-11** "**Guidance on Qualification and Classification of Software in Regulation** (EU) 2017/745 - MDR and Regulation (EU) 2017/746" dell'Ottobre 2019 è una guida che aiuta gli stakeholders nella qualificazione e classificazione del software fornendo esempi di classificazione.

- **Il software è definito, nel citato documento, come un set di istruzioni che processa dati in ingresso e crea dati in uscita, fornendo la definizione di software dispositivo medico qualificando lo Scopo Medico**

- controlla direttamente un dispositivo medico (hardware) (ad es. un software per il trattamento radioterapico),

- fornisce informazioni decisionali mediche immediate (ad es. un software per la misurazione di parametro (glucosio nel sangue),

- fornisce supporto agli operatori sanitari (ad es. per l'interpretazione Ecg/immagini sulla base del quale il medico decide diagnosi e terapia). (*Accessorio esteso*)

- **MDCG 2020-1 Guidance on Clinical Evaluation** (MDR) / Performance Evaluation (IVDR) of Medical Device Software March 2020.

- Questa guida riguarda **il software per il quale il produttore dichiara uno scopo medico specifico.**

- **Tale software ha quindi un BENEFICIO CLINICO e richiede EVIDENZA CLINICA all'interno della propria valutazione di conformità.**

# Definizione di DM del Regolamento MDR2017/745

- «**dispositivo medico**»: qualunque **strumento**, **apparecchio, impianto**, **software**, sostanza o altro prodotto, utilizzato da solo o in combinazione, compreso il software destinato dal fabbricante ad essere impiegato specificamente con **finalità diagnostiche e/o terapeutiche** e necessario al corretto funzionamento del dispositivo, destinato dal fabbricante ad essere impiegato sull'uomo a fini di:
- — diagnosi, **prevenzione**, controllo, terapia o **attenuazione di una malattia**;
- — **diagnosi, controllo, terapia, attenuazione o compensazione di una ferita o di un handicap;**
- — studio, **sostituzione** o modifica dell'anatomia o **di un processo fisiologico**;
- — intervento sul concepimento,
- la cui azione principale voluta nel o sul corpo umano non sia conseguita con mezzi farmacologici né immunologici né mediante metabolismo, ma la cui funzione possa essere assistita da questi mezzi

I dispositivi devono soddisfare i pertinenti **requisiti essenziali di sicurezza e prestazione** prescritti (allegato I) in considerazione della loro **destinazione**.

# Percorso dell'innovazione tecnologica



# Attori principali nel processo di immissione di un DM sul mercato

- EUDAMED
- Autorità Competente
- Fabbricante (opp. Mandatario)
- Organismo Notificato
- *Enti Istituzionali di riferimento*

# Conformità e classificazione dei DM

•Le procedure di valutazione della conformità per i dispositivi delle classe I possono essere svolte, in linea di massima, sotto la sola responsabilità del fabbricante, dato lo scarso indice di vulnerabilità che possiedono questi prodotti;

•per i dispositivi della classe IIa **l'intervento obbligatorio di un organismo notificato deve riguardare la fase di fabbricazione**;

•per i dispositivi delle classi IIb e III i quali possiedono un elevato potenziale di rischio, è necessario **un controllo da parte di un Organismo Notificato sia nella fase di progettazione dei dispositivi che nella fase di fabbricazione**

# Capo V SEZIONE 1 CLASSIFICAZIONE, Art. 51

I dispositivi medici sono suddivisi nelle classi I, IIa, IIb e III.

La classificazione è in funzione della destinazione d'uso (durata di utilizzo, modalità di funzionamento, organo in cui il dispositivo esercita la sua funzione) e dei rischi che comporta.

**Allegato VIII – Regole di classificazione**

⇩

22 regole di classificazione

# Requisiti Generali di Sicurezza e Prestazione (Allegato I).

Sono Requisiti il cui soddisfacimento costituisce la condizione necessaria per l'immissione in commercio.

- Capo I — Requisiti generali rivolti alla sicurezza intrinseca del dispositivo
- Capo II — Requisiti relativi alla progettazione ed alla fabbricazione e sono suddivisi in diversi gruppi;
- Capo III — Requisiti riguardanti le informazioni fornite con il dispositivo (etichettatura)

Esiste presunzione di conformità -> si presume conforme ai Requisiti Generali di Sicurezza e Prestazione il dispositivo medico fabbricato in conformità delle norme armonizzate comunitarie e delle norme nazionali che le recepiscono.

# METODOLOGIE DI VALUTAZIONE PER I DISPOSITIVI MEDICI

- **Certificazioni** (processo cogente)(**Norme armonizzate**)
- **Analisi dei rischi (ISO 14971)**
- Conformità a norme particolari/specifiche (processo volontario)
- Valutazione In Vitro, Ex Vivo, Clinica, in Silico *(Fascicolo Tecnico per gli studi clinici, le valutazioni, le gare regionali etc…)*

- Health Technology Assessment (HTA) (→ AGENAS)
- Valutazioni di esito (→ enti e autorità competente)
- Revisioni sistematiche (→ associazioni scientifiche)

# Mitigazione dei rischi e Prove Tecniche

**Analisi dei rischi e prove tecniche** di verifica per la mitigazione dei Rischi in accordo con:

- ***norme armonizzate*** *(ISO UNI EN) (MDSW IEC ciclo di vita del SW)*

*•Le norme armonizzate europee previste dai regolamenti sui dispositivi medici sono elaborate dal CEN e dal CENELEC in qualità di enti di normazione europei sulla base della richiesta di normazione presentata dalla Commissione ai sensi del regolamento (UE) n. 1025/2012. Dopo la pubblicazione dei loro riferimenti nella* ***Gazzetta ufficiale dell'Unione europea*** *da parte della Commissione,* ***l'uso volontario di tali norme conferisce una presunzione di conformità ai requisiti delle direttive a cui intendono riferirsi****.*

- norme particolari (ASTM,…)
- esperienze tecnico scientifiche (ricerca, innovazione)

# Il MDSW nel quadro regolatorio dei DM

Come specificato nel Considerando 6 della **Direttiva 2007/47/CE**, "Occorre chiarire che un software è di per sé un dispositivo medico quando è **specificamente destinato** dal fabbricante ad essere impiegato per una o più delle **finalità mediche** stabilite nella **definizione di dispositivo medico**.

Anche se utilizzato in un contesto sanitario, il **software generico** non è un dispositivo medico."

Anche il **Regolamento (EU) 2017/745** (MDR) [1] considera espressamente l'esistenza dei MDSW (v. Articolo 2 – Definizioni).
Sia per la Direttiva MDD (All IX - Criteri di classificazione, 1.4) che per il MDR (art. 2) il MDSW è considerato un **dispositivo medico attivo**.

Al di là della definizione, invece, notevoli **differenze** sono date in fase di regole di classificazione.

# Il MDSW nel quadro regolatorio dei DM

## MDD

Non esistono regole di classificazione che citino direttamente il software come DM. Risultano applicabili però le regole 9, 10, 11, 12, in quanto riferite a **DM attivi.**
(v. MEDDEV 2.1/6, July 2016:

**"Stand alone** software that meets the definition of a medical device shall be considered as an **active medical device.**
This means that rules 9, 10, 11 and 12 of Annex IX to Directive 93/42/EEC may apply"

## MDR

**Regola 11** - Il software destinato a fornire informazioni utilizzate per prendere decisioni a fini diagnostici o terapeutici rientra nella classe IIa, a meno che tali decisioni abbiano effetti tali da poter causare:

- il decesso o un deterioramento irreversibile delle condizioni di salute di una persona, nel qual caso rientra nella classe III,
- un grave deterioramento delle condizioni di salute di una persona o un intervento chirurgico, nel qual caso rientra nella classe IIb.

Il software destinato a **monitorare i processi fisiologici** rientra nella **classe IIa,** a meno che sia destinato a monitorare i **parametri fisiologici vitali**, ove la natura delle variazioni di detti parametri sia tale da poter creare un pericolo immediato per il paziente, nel qual caso rientra nella classe IIb.

Tutti gli altri software rientrano nella classe I."

# Linea Guida MED DEV /MDCG

*Med Dev 2.1/6 – Luglio 2016 --- MDCG 2021-24 Guidance on classification of medical devices, October 2021*

**Fase decisionale 1**: se il software autonomo è un programma per computer, potrebbe trattarsi di un dispositivo medico.

Se il software non è un programma per computer, allora è un **documento digitale** e quindi non è un dispositivo medico.

Esempi di programmi per computer sono **applicazioni software, macro, script, librerie collegate dinamicamente, file batch, fogli di stile e qualsiasi documento contenente istruzioni di formattazione o filtro attive**.

Esempi di **documenti digitali** sono file di **immagini, file DICOM, registrazioni ECG digitali, risultati numerici di test e cartelle cliniche elettroniche (EHR).**

Nota: Mentre l'EHR di solito non è un programma per computer, il **sistema EHR**, cioè il software di scrittura, recupero, rappresentazione, ecc. delle informazioni contenute nell'EHR sono un programma per computer. Questo è simile a quello dei file DICOM rispetto a un PACS.

**Fase decisionale 2**: se il software è incorporato in un dispositivo medico anziché in un software autonomo, **deve essere considerato parte di tale dispositivo medico**. Se si tratta di software autonomo **Stand Alone**, procedere alla fase 3 della decisione.

**Fase decisionale 3**: se il software non esegue un'azione sui dati, o esegue un'azione limitata all'archiviazione, alla comunicazione, alla "ricerca semplice" o alla compressione senza perdita di dati (ovvero utilizzando una procedura di compressione che consenta l'esatta ricostruzione dei dati originali) **non è un dispositivo medico.**

L'alterazione della rappresentazione dei dati a scopo di abbellimento non rende il software un dispositivo medico. In altri casi, anche quando il **software altera la rappresentazione dei dati per uno scopo medico, potrebbe trattarsi di un dispositivo medico**.

La "ricerca semplice" si riferisce al recupero di record mediante l'abbinamento dei metadati dei record con i criteri di ricerca dei record, ad esempio le funzioni della biblioteca. La ricerca semplice non include i software che forniscono risultati di ricerca interpretativi, ad esempio per identificare i risultati medici nelle cartelle cliniche o sulle immagini mediche.

# Linea Guida MED DEV/MDCG

*Med Dev 2.1/6 – Luglio 2016 --- MDCG 2021-24 Guidance on classification of medical devices, October 2021*

Il software destinato a **creare o modificare informazioni mediche** potrebbe essere qualificato come dispositivo medico. Se tali modifiche vengono apportate per facilitare i compiti percettivi e/o interpretativi svolti dagli operatori sanitari durante la revisione delle informazioni mediche (ad esempio durante la ricerca nell'immagine di risultati che supportano un'ipotesi clinica sulla diagnosi o sull'evoluzione della terapia), il software potrebbe essere un dispositivo medico.

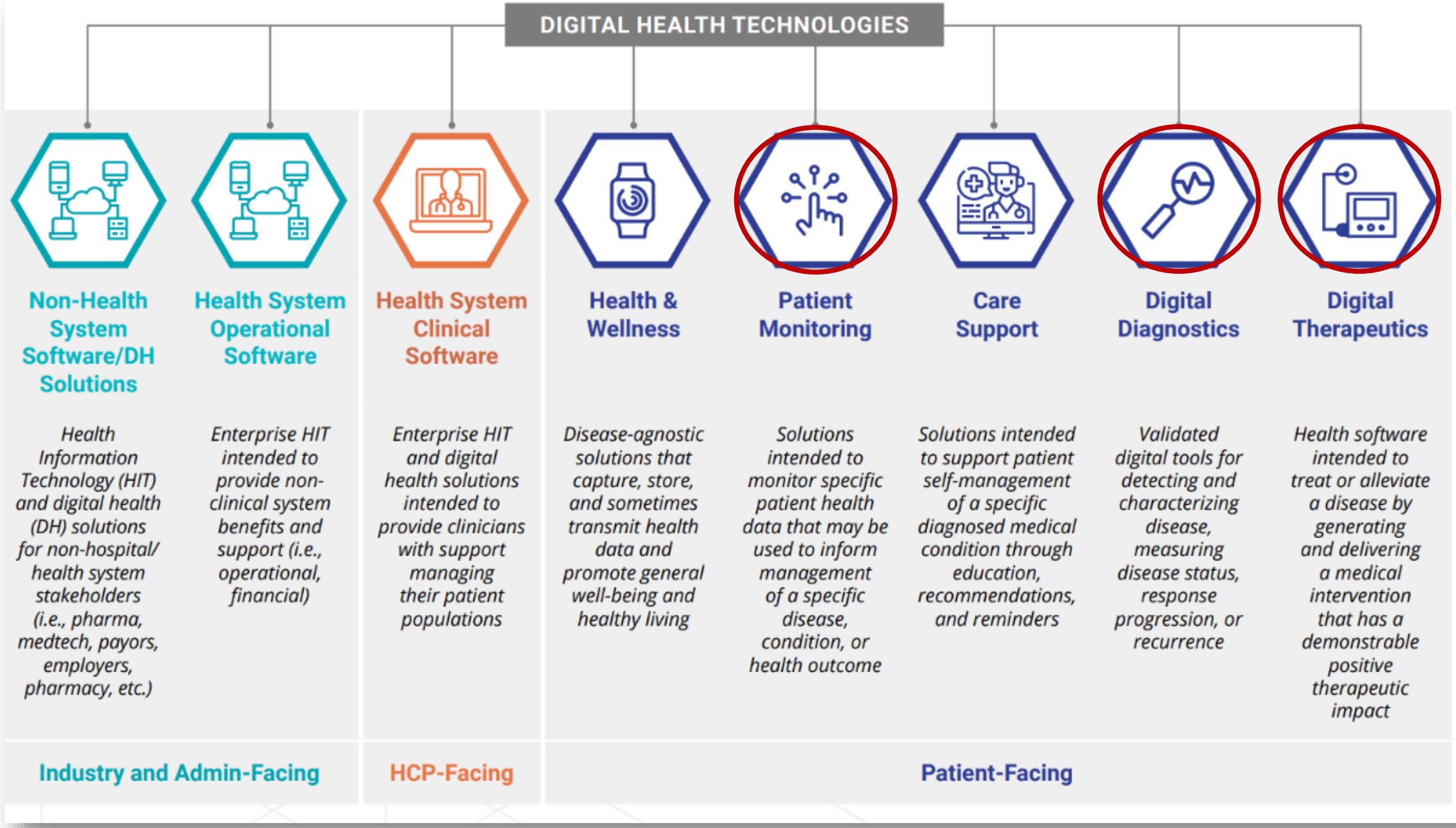
Nota: la visualizzazione delle immagini di solito comporta alterazioni alla rappresentazione perché vengono utilizzate tecniche come la modifica del contrasto, il miglioramento dei bordi, la manipolazione della scala di grigi, l'attenuazione, la nitidezza, lo zoom e il ridimensionamento.

Le alterazioni possono includere ricostruzione, compressione con perdita, filtraggio, riconoscimento di modelli, modellazione, interpolazione, trasformazione, classificazione (ad es. punteggio dei tumori rispetto a criteri specifici), segmentazione, registrazione (ad es. mappatura di un set di dati su un modello o atlante o su un altro set di dati, ad esempio registrazione di un'immagine MRI su un'immagine TC), calcoli, quantificazione, qualificazione (ad es. confronto di dati con riferimenti), rendering, visualizzazione, interpretazione, ecc.

**Fase decisionale 4**: un esempio di software a beneficio dei singoli pazienti è un software destinato ad essere utilizzato per la valutazione dei dati dei singoli pazienti per supportare o influenzare l'assistenza medica fornita a quel paziente. Esempi di software che **non sono considerati a beneficio dei singoli pazienti** sono quelli che aggregano i dati della popolazione, forniscono percorsi diagnostici o terapeutici **generici**, letteratura scientifica, atlanti medici, modelli e modelli, nonché software per **studi epidemiologici o registri**.

**Fase 5**: se il fabbricante intende specificamente utilizzare il software per uno degli scopi elencati all'articolo 1, paragrafo 2, lettera a), della direttiva 93/42/CEE, il software deve essere qualificato come **dispositivo medico**.**(Definizione di DM)**

Tuttavia, se il fabbricante intende solo uno scopo non medico, come la fatturazione o la pianificazione del personale, non è un dispositivo medico. Nota: un'attività come la posta elettronica, la messaggistica web o vocale, l'elaborazione di testi e il backup non è di per sé considerata uno scopo medico, ai sensi della direttiva 93/42/CEE.


DIGITAL HEALTH TECHNOLOGIES
Non-Health System Software/DH Solutions
Health System Operational Software
Health System Clinical Software
Health & Wellness
Patient Monitoring
Care Support
Digital Diagnostics
Digital Therapeutics
Health Information Technology (HIT) and digital health (DH) solutions for non-hospital/ health system stakeholders (i.e., pharma, medtech, payors, employers, pharmacy, etc.)
Enterprise HIT intended to provide non-clinical system benefits and support (i.e., operational, financial)
Enterprise HIT and digital health solutions intended to provide clinicians with support managing their patient populations
Disease-agnostic solutions that capture, store, and sometimes transmit health data and promote general well-being and healthy living
Solutions intended to monitor specific patient health data that may be used to inform management of a specific disease, condition, or health outcome
Solutions intended to support patient self-management of a specific diagnosed medical condition through education, recommendations, and reminders
Validated digital tools for detecting and characterizing disease, measuring disease status, response progression, or recurrence
Health software intended to treat or alleviate a disease by generating and delivering a medical intervention that has a demonstrable positive therapeutic impact
Industry and Admin-Facing
HCP-Facing
Patient-Facing

| DHT Category | Health & Wellness | Patient Monitoring | Care Support | Digital Diagnostics | Digital Therapeutics |
|---|---|---|---|---|---|
| **Overview** | • Disease-agnostic digital health solutions that primarily capture and store general health data and promote healthy living | • Digital solutions intended to monitor specific health data, which may be interpreted by physician for clinical management | • Digital solutions intended to help patients better manage their care of a specific disease or medical condition | • Validated digital tools and software that deliver a diagnosis or prognosis of a specific disease or medical condition | • Health software intended to treat or alleviate a specific disease or medical condition by generating and delivering a medical intervention |
| **Claims** | ✗ No claims to treat, improve, or diagnose a medical condition | ~ May make non-clinical claims to assess patient data | ~ May make non-clinical claims to improving health-adjacent measures (e.g., adherence) | ✓ Make a clinical claim to diagnose or assess a specific disease or medical condition | ✓ Make a clinical claim to treat or alleviate a specific disease or medical condition |
| **Intervention Delivery** | ✗ Does not deliver a medical intervention | ~ Collects health data to inform HCP decision making around medical intervention | ~ May recommend actions for patients to better manage care or inform HCPs but does not deliver medical intervention | ✓ Software drives medical intervention through a formal diagnosis or assessment | ✓ Software itself generates and delivers a medical intervention |
| **Evidence Requirements** | ✗ Not required | ~ Non-clinical claims to assess patient data, must be validated and meet a regulatory agency's quality requriements | ~ Any non-clinical claims (e.g., adherence) must be validated and meet a regulatory agency's quality requirements | ✓ Diagnostic accuracy must be validated and meet a regulatory agency's quality requirements | ✓ Efficacy claims must meet a regulatory agency's quality requirements |
| **Regulatory Implications** | ✗ No regulatory oversight | ~ May require regulatory approval and labeling | ~ May require regulatory approval and labeling | ✓ Regulated solution with label for indication, usage, evidence, warnings, etc. | ✓ Regulated solution with label for indication, usage, evidence, warnings, etc. |

# Linea Guida MDCG

*MDCG 2021-24 Guidance on classification of medical devices, October 2021*

| Class | Rule 11 | Examples |
|---|---|---|
| IIa | Software intended to provide information which is used to take decisions with diagnosis or therapeutic purposes is classified as class IIa,<br>except if such decisions have an impact that may cause: | • MDSW intended to rank therapeutic suggestions for a health care professional based on patient history, imaging test results, and patient characteristics, for example, MDSW that lists and ranks all available chemotherapy options for BRCA-positive individuals.<br>• Cognitive therapy MDSW where a specialist determines the necessary cognitive therapy based on the outcome provided by the MDSW. |
| III | — death or an irreversible deterioration of a person's state of health[1], in which case it is in class III; or | • MDSW intended to perform diagnosis by means of image analysis for making treatment decisions in patients with acute stroke. |
| IIb | — a serious deterioration of a person's state of health[1] or a surgical intervention, in which case it is classified as class IIb. | • A mobile app intended to analyse a user's heartbeat, detect abnormalities and inform a physician accordingly.<br>MDSW intended for diagnosing depression based on a score resulting from inputted data on patient symptoms (e.g. anxiety, sleep patterns, stress etc.). |
| IIa | Software intended to monitor physiological processes is classified as class IIa, | • MDSW intended to monitor physiological processes that are not considered to be vital.<br>• Devices intended to be used to obtain readings of vital physiological signals in routine check-ups including monitoring at home. |
| IIb | except if it is intended for monitoring of vital physiological parameters[3], where the nature of variations of those parameters is such that it could result in immediate danger to the patient, in which case it is classified as class IIb. | • Medical devices including MDSW intended to be used for continuous surveillance of vital physiological processes in anaesthesia, intensive care or emergency care. |
| I | All other software is classified as class I. | • MDSW app intended to support conception by calculating the user's fertility status based on a validated statistical algorithm. The user inputs health data including basal body temperature |

- intended purpose,
- intended population (*e.g.* diseases to be treated /diagnosed),
- context of use (*e.g.* intensive care, emergency care, home use)
- information provided
- decisions to be taken.

# Il MDSW nel quadro regolatorio dei DM

(MDCG 2019-11, **Guidance on Qualification and Classification of Software in Regulation (EU) 2017/745 – MDR and Regulation (EU) 2017/746 – IVDR**) → (Rule 11a)

**Decision Support Software**

In generale, si tratta di strumenti informatici che combinano database e algoritmi di informazioni mediche generali con dati specifici del paziente. Hanno lo scopo di fornire agli operatori sanitari e/o agli utenti raccomandazioni per la diagnosi, la prognosi, il monitoraggio e il trattamento dei singoli pazienti. **They are qualified as medical devices.**

I sistemi di pianificazione del **trattamento radioterapico** hanno lo scopo di calcolare il dosaggio delle radiazioni ionizzanti da applicare ad uno specifico paziente. Si ritiene che controllino, monitorino o influenzino direttamente la fonte di radiazioni ionizzanti e sono qualificati come **dispositivi medici**.

- I sistemi di **pianificazione dei farmaci** (ad esempio la **chemioterapia**) hanno lo scopo di calcolare il dosaggio del farmaco da somministrare ad uno specifico paziente e pertanto sono qualificati come dispositivi medici.
- I sistemi di rilevamento assistito da computer (**CAD**) hanno lo scopo di fornire informazioni che possono suggerire o escludere condizioni mediche, quindi qualificabili come **dispositivi medici** (MDSW). Ad esempio, tali sistemi sarebbero in grado di analizzare automaticamente le immagini a raggi X o interpretare gli ECG.

# Requisiti relativi alla progettazione e fabbricazione (ciclo di vita del SW, qualità dei dati) - MDR

Requisiti di un software legati alla **gestione delle informazioni**

**17 Sistemi elettronici programmabili — dispositivi contenenti sistemi elettronici programmabili e software che costituiscono dispositivi a sé stanti**

17.1 I dispositivi contenenti sistemi elettronici programmabili, compresi i software, o i software che costituiscono dispositivi a sé stanti, sono progettati in modo tale da garantire **la riproducibilità, l'affidabilità e le prestazioni in linea con la destinazione d'uso per essi prevista**. In caso di **condizione di primo guasto** sono previsti mezzi adeguati per eliminare o ridurre, per quanto possibile, i rischi che ne derivano o il peggioramento delle prestazioni.

17.2 Per i dispositivi contenenti un software o per i software che costituiscono dispositivi a sé stanti, il **software** è sviluppato e fabbricato conformemente allo **stato dell'arte**, tenendo conto dei ***principi del ciclo di vita dello sviluppo*, della *gestione del rischio***, compresa la **sicurezza delle informazioni**, della verifica e della convalida.

Requisiti software su **piattaforme mobili**

17.3. I software di cui al presente punto destinati a essere usati **in combinazione con piattaforme di calcolo mobili** sono progettati e fabbricati tenendo conto delle peculiarità della piattaforma mobile (ad esempio dimensioni e grado di contrasto dello schermo) e di fattori esterni connessi al loro uso (variazioni ambientali relative al livello di luce o di rumore).

# Risk Analysis 14971

**Analisi del rischio**:
Identificazione dell'uso previsto/ scopo previsto (fase 1),·
Identificazione del pericolo (fase 2)
Stima del rischio (fase 3)

Valutazione del rischio:· Decisioni sull'accettabilità del rischio (fase 4)

Controllo del rischio:· Analisi delle opzioni, · Implementazione, · Valutazione del rischio residuo,· Accettazione complessiva del rischio

Informazioni post-produzione:· Esperienza post-produzione, · Revisione dell'esperienza di gestione del rischio

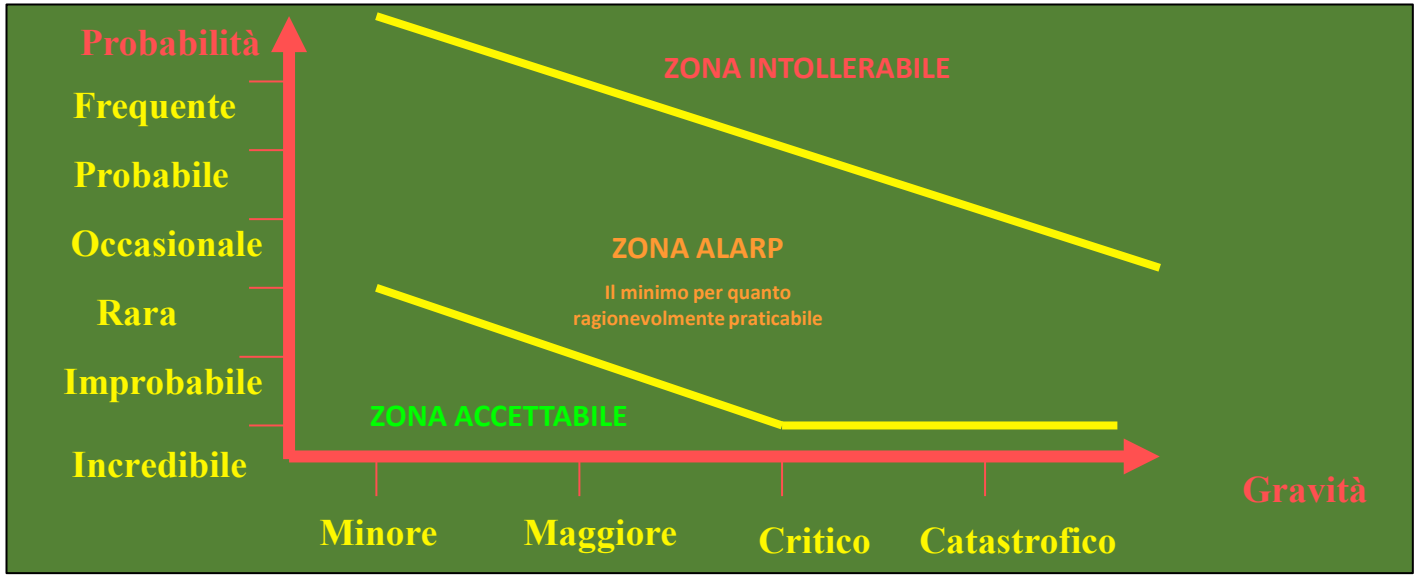
# Concetti principali dell'Analisi dei Rischi

Tabella 1. Rappresentazione schematica dei principali concetti legati all'analisi dei rischi

| | RISCHIO | PCD | DANNO |
|---|---|---|---|
| Espressione di: | *potenzialità d'incidente* | *causa scatenante* | *effetto reale* |
| | Non può essere causa d'incidente in sé, in quanto necessita della causa scatenante | Può manifestarsi a diversi livelli funzionali come: procedura, produzione (nei processi), ecc. | La gravità del danno dipende dal contesto in cui si verifica |
| Rimedio: | *prevenzione* | *identificazione* | *protezione* |

- Identificazione delle PCD (Possibili Cause di Danno). Le cause possono essere legate a diversi fattori.
- Determinazione della probabilità *P* che la singola PCD produca il danno corrispondente
- Determinazione della Gravità del Danno G.

**Le probabilità possono essere stimate elaborando le informazioni provenienti da letteratura tecnica, studi clinici, esperienza del fabbricante , informazioni da file di reclamo, risultati di calcoli e prove. Anche la gravità del danno associato ad un rischio può essere ricavata dalla letteratura o norme vigenti (eventuali quantificazioni possono essere convenzionali o reali)**



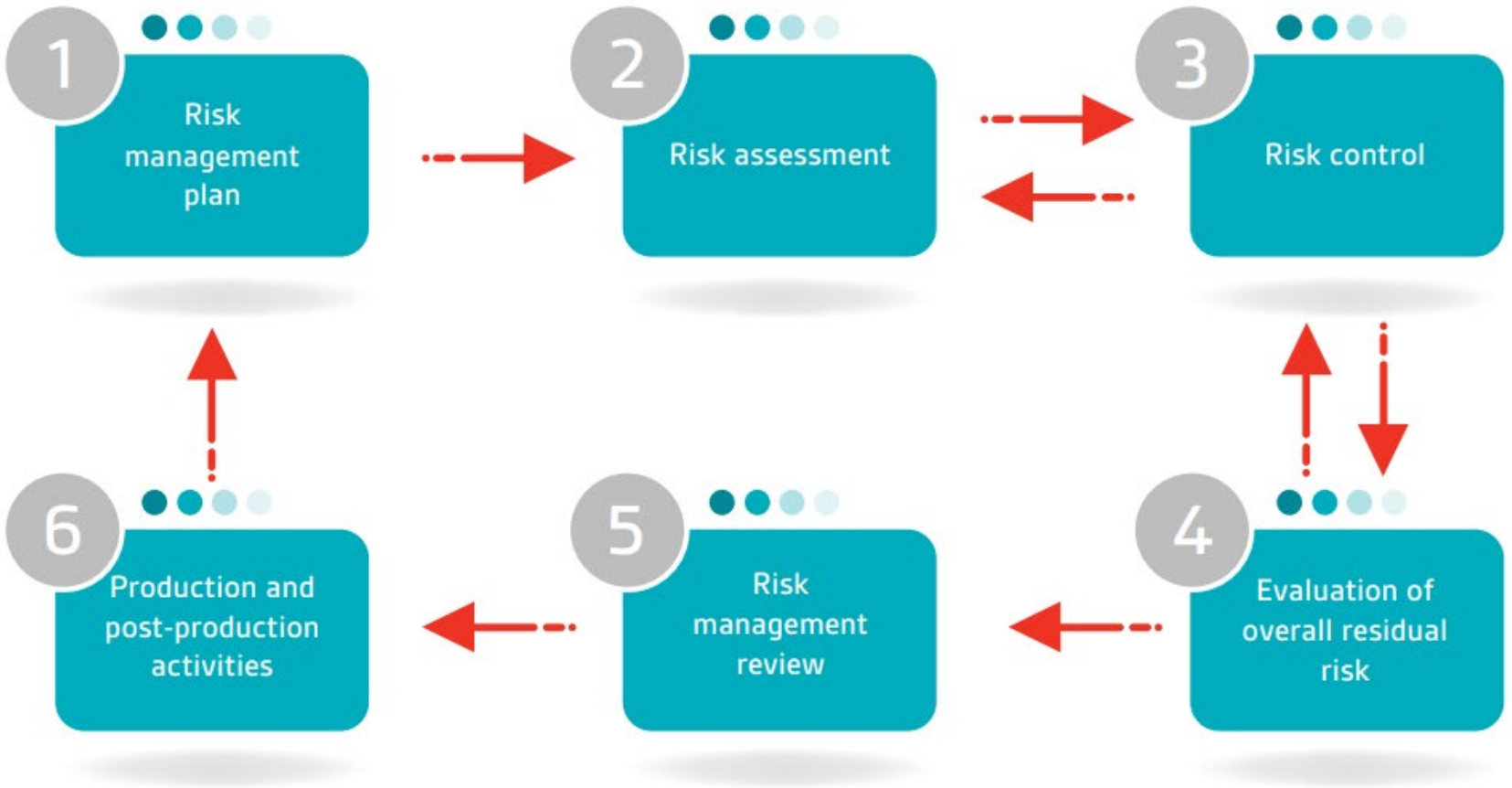
# New Risk Management 14971

Medical Device White Paper Series

## Risk management for medical devices and the new BS EN ISO 14971

**Jos van Vroonhoven**, Philips, The Netherlands,

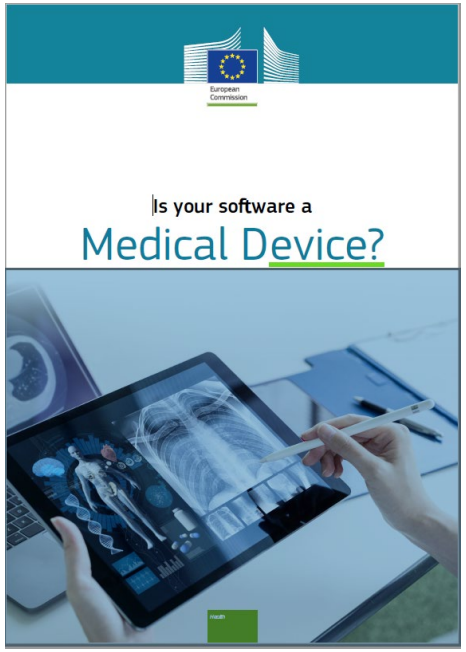
Figure 1 – The six process steps in the risk management process of BS EN ISO 14971 [1]



# *Il SaMD nel quadro regolatorio dei DM*
# *«Osservatorio SaMD dell'ISS/TISP» 2019-2022*

***Accordo di collaborazione*** tra il Ministero della Salute - D.G.D.M.S.F. e l'Istituto Superiore di Sanità

**Obiettivo**: definizione di metodologie per la classificazione e la valutazione di software stand-alone come dispositivi medici (SaMD) e dei software non DM (SnMD), ma comunque di interesse per la salute, per **favorire la sorveglianza ed una adozione regolata (qualificazione e posizionamento della tecnologia)**




APP
Digital Therapy
Digital Rehabilitation
AI
EHR
Platform
Digital Twin
…




"DISCLAIMER: This application is intended for reviewing only. It is currently neither CE nor FDA cleared. Not intended for diagnostic use!"

# EN ISO/IEC TS 82304-2 Health and wellness apps - Quality and reliability



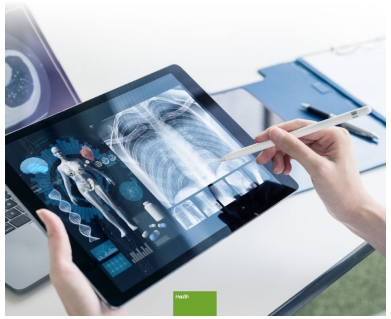

Health and Quality Label


5.1 Product information
5.1.1 Product
5.1.2 App manufacturer
5.2 Healthy and safe
5.2.1 Health requirements
5.2.2 Health risks
5.2.3 Ethics
5.2.4 Health benefit
5.2.5 Societal benefit
5.3 Easy to use
5.3.1 Accessibility
5.3.2 Usability
5.4 Secure data
5.4.1 Privacy
5.4.2 Security
5.5 Robust build
5.5.1 Technical robustness
5.5.2 Interoperability


# L2E: Label to Enable

# MDSW & AI

Is your software a
Medical Device?



La qualificazione di un SW come DM è indipendente dalla tecnologia implementata nel SW!
Un MDSW può ricadere nel RIA (AI Act), in funzione della tecnologia utilizzata

## Decision steps to assist qualification of Medical Device Software (MDSW)

**Medical Device Software (MDSW):** Software that is intended to be used, alone or in combination, for a purpose as specified in the definition of a "medical device" in the Medical Devices Regulation (MDR) or *In Vitro* Diagnostic Medical Devices Regulation (IVDR).

Is your software a Medical Device Software?

1. Is the product a 'Software' according to the definition of MDCG 2019-11?
   - X No → NOT COVERED BY MDCG 2019-11
   - Yes →
2. Is the software an 'MDR Annex XVI device', an 'Accessory' for a medical device according to Art. 2(2) of the MDR or IVDR or 'software driving or influencing the use of a (hardware) medical device'?
   - X No →
   - Yes →
3. Is the software performing an action on data different from storage, archival, communication or simple search?
   - X No →
   - Yes →
4. Is the action for the benefit of invidual patients?
   - X No →
   - Yes →
5. Is the software a Medical Device Software (MDSW) according to the definition of MDCG 2019-11?
   - X No →
   - Yes →

COVERED BY THE MEDICAL DEVICES REGULATIONS*

NOT COVERED BY THE MEDICAL DEVICES REGULATIONS*

**Medical devices Regulations*** refers to the two applicable regulations:
Regulation (EU) 2017/745 on Medical Devices (MDR) and Regulation (EU) 2017/746 on *In Vitro* Diagnostic Medical Devices (IVDR)

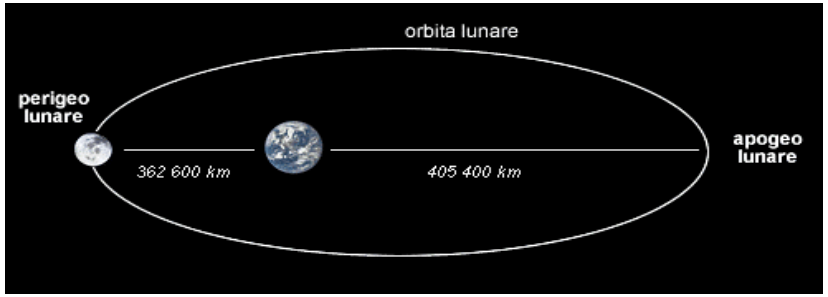
# Scenario : Ecosistemi digitali e Big Data

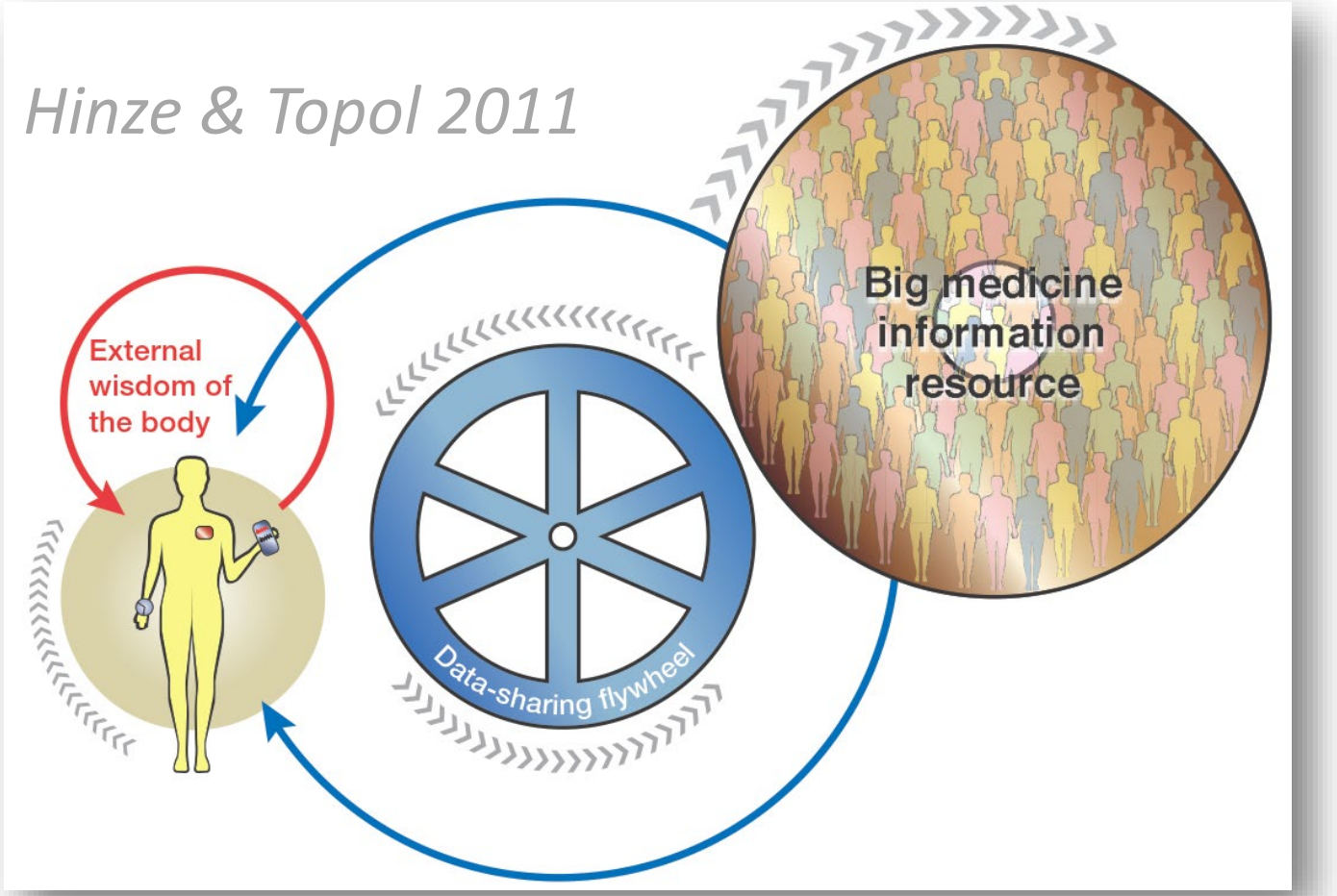
Forbes, Jun 16, 2021

**Harnessing Healthcare's Data Explosion With AI-Based Natural Language Processing**

In 2020, the amount of healthcare data created globally was an estimated 2,314 exabytes — which is an unfathomable amount when you consider a single exabyte is equivalent to one billion gigabytes.

2314 exabytes -> pila di DVD alta **633972 km**




Each individual gets its own medical data **biosensors**, **imaging**, physical examination tools and laboratory tests, comprising a new 'external' wisdom of the body. Such data are fed into the **flywheel** of the engine and eventually, when there are enough individuals amassed into a **big medicine resource**, there is a breakthrough to form a **valuable medical knowledge resource**. That, too, provides external **feedback** to the individual for optimal prevention and medical treatment.

# *Machine learning*

L'approccio di apprendimento automatico (*machine learning*) è guidato dall'emergere dei Big data e dall'accelerazione hardware (GPU), ed è in grado di estrarre autonomamente caratteristiche incorporate nei dati, con una rappresentazione di livello superiore e più astratto dalle caratteristiche dei dati grezzi di input.

*A machine learning algorithm is an algorithm that is able to learn from data.*

Arthur Samuel (1959) defined machine learning as a "field of study that gives computers the ability to learn without being explicitly programmed". That is, machine-learning programs have not been explicitly entered into a computer.

Mitchell (1997) provides a succinct definition:

"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."

Esamples of **tasks**: classification, regression, structured output, machine translation, denoising, segmentation ...

# Prospettive in Sanità EU

**Healthcare**

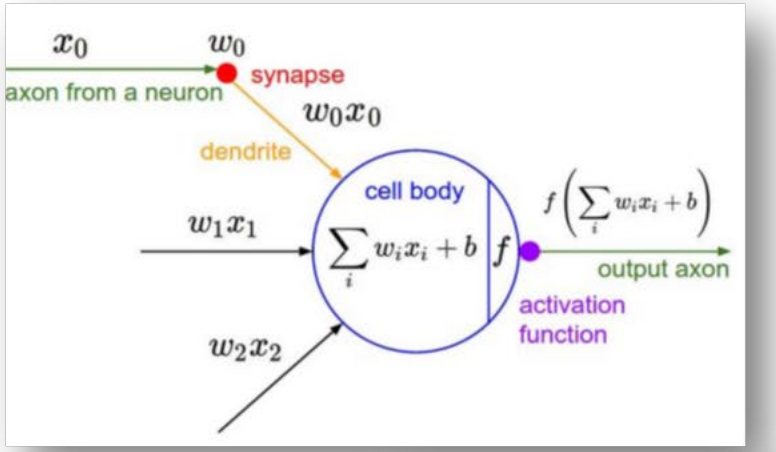
Supporting diagnosis and early identification of potential pandemics are two of the areas in healthcare with the biggest potential. This is the sector where the biggest medium-term impact (0 to 3 years) has been identified.[1] However, concerns over the privacy and protection of sensitive health data still need to be addressed.




| | | | | |
|---|---|---|---|---|
| **Health care** | ▪ **$300 billion possible savings** in the United States using machine learning tools for population health forecasting<br>▪ **£3.3 billion possible savings** in the United Kingdom using AI to provide preventive care and reduce nonelective hospital admissions | ▪ **30–50% productivity improvement** for nurses supported by AI tools<br>▪ Up to **2% GDP savings** for operational efficiencies in developed countries | ▪ **5–9% health expenditure reduction** by using machine learning to tailor treatments and keep patients engaged | ▪ **$2 trillion–$10 trillion savings** globally by tailoring drugs and treatments<br>▪ **0.2–1.3 additional years** of average life expectancy |

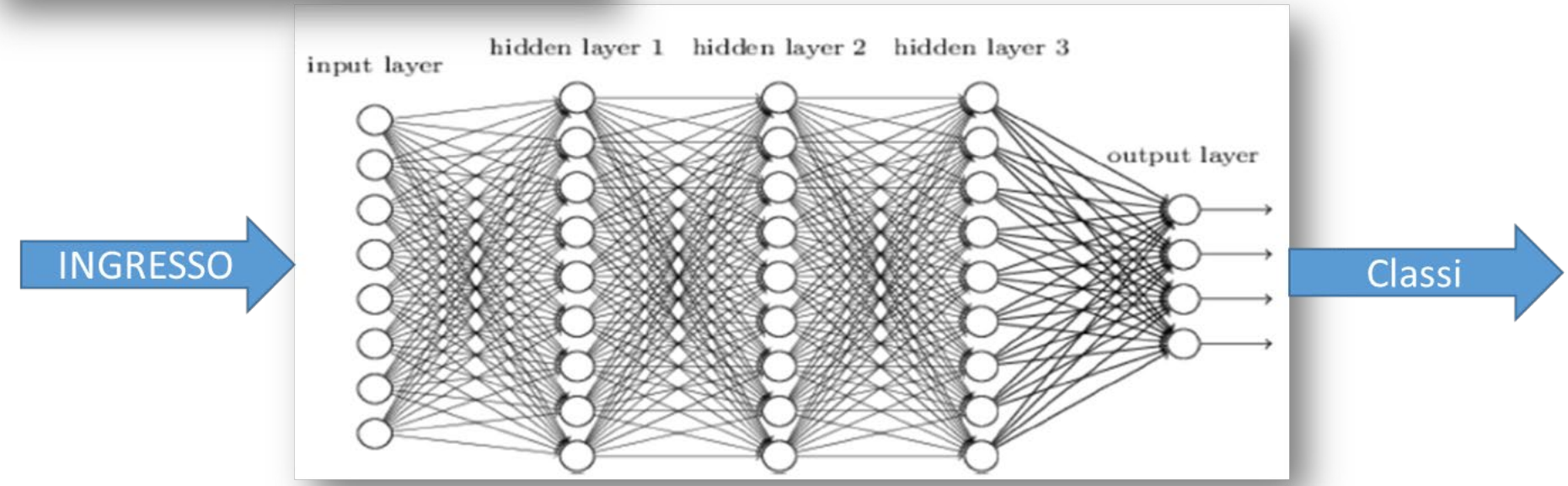
Harnessing the economic benefits of Artificial Intelligence November 2017

# Machine/Deep Learning questo sconosciuto ? → Formazione

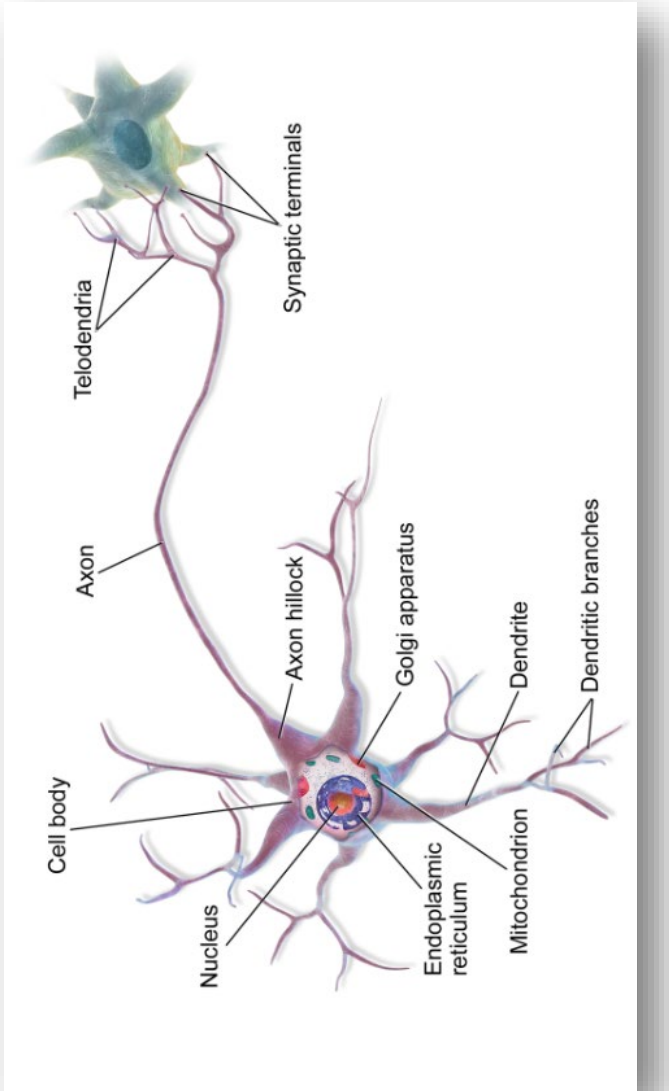


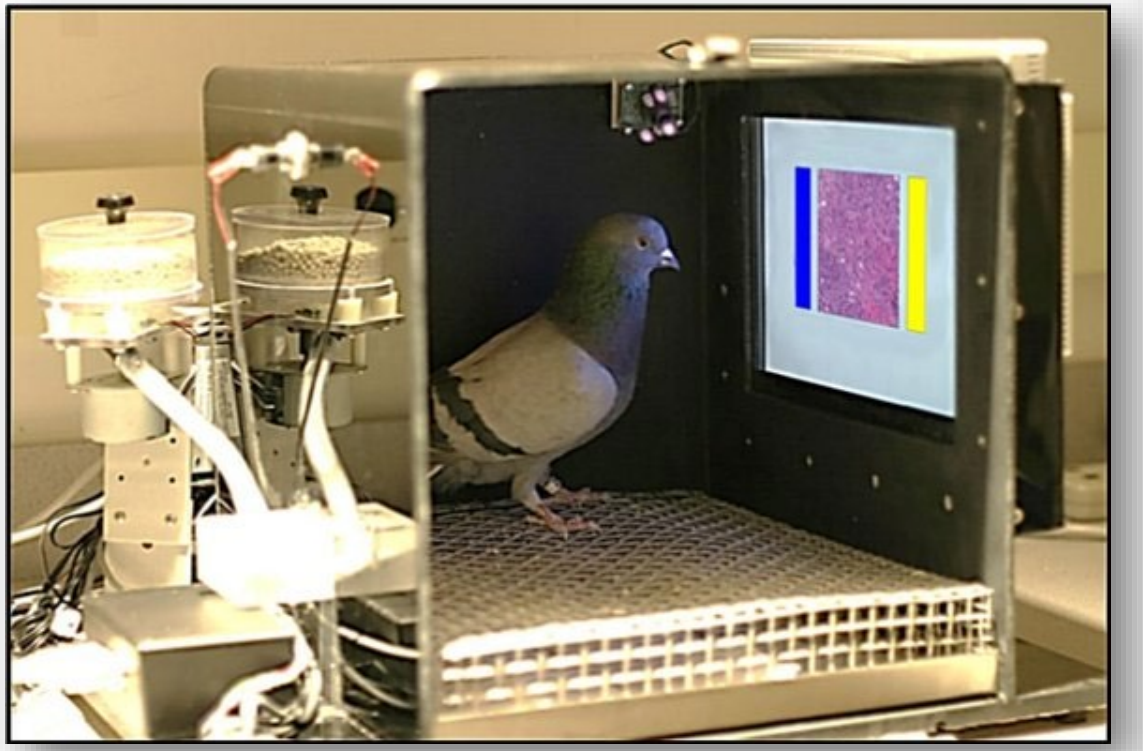

***Deep Feedforward Neural Network***

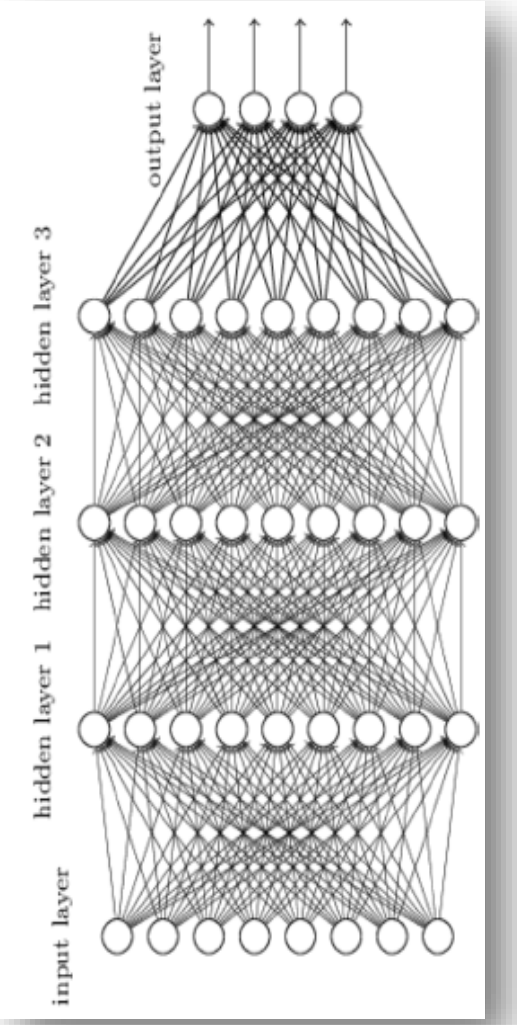
**Strati di neuroni che eseguono operazioni anche semplici (somme e moltiplicazioni)**



# https://www.cs.toronto.edu/~lczhang/360/lec/w01/pigeon.html





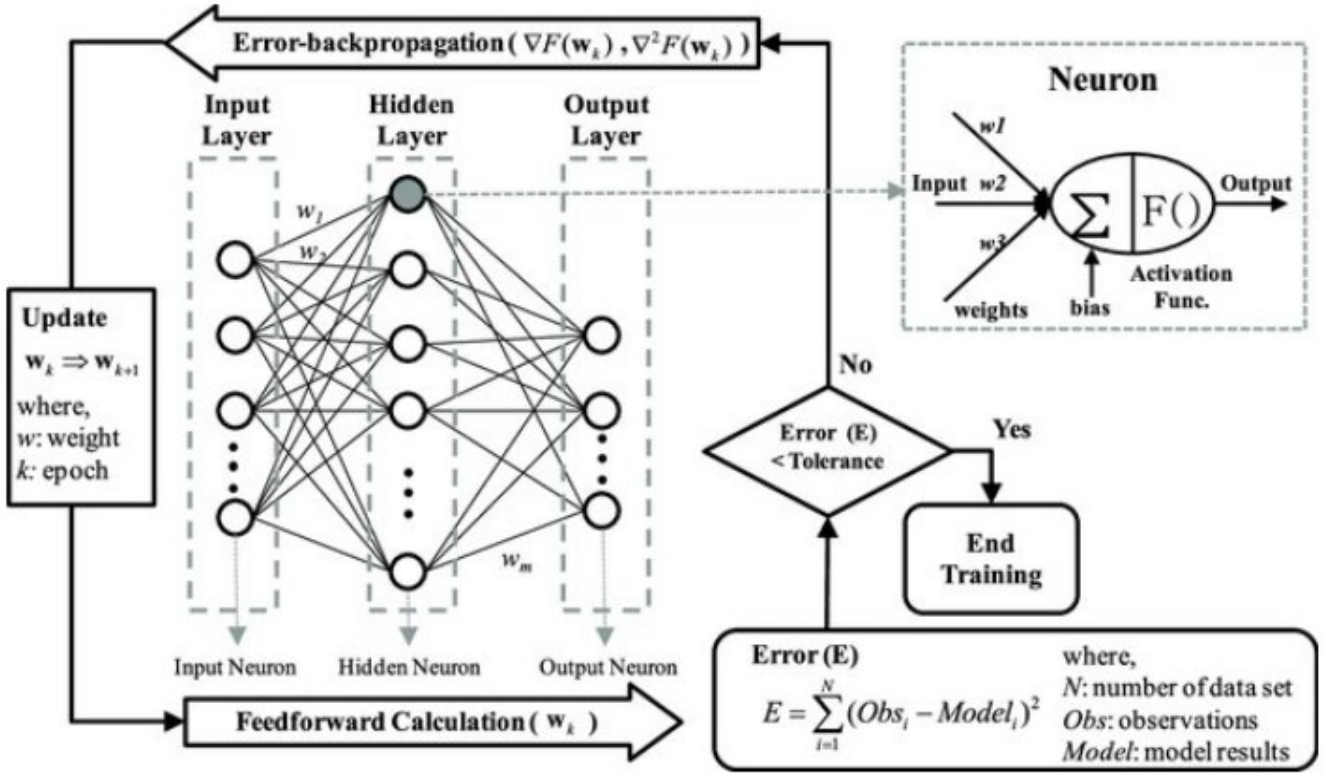
# *Learning: Che cos'è un'epoca in ML/DL ?*

Un'epoca nell'apprendimento automatico significa :

- un passaggio completo del set di dati di addestramento attraverso l'algoritmo.
- Il numero di questa epoca è un iperparametro importante per l'algoritmo. Specifica il numero di epoche o passaggi completi dell'intero set di dati di training che passano attraverso il processo di training o apprendimento dell'algoritmo.
- In un epoca il passaggio completo del set di dati può essere parzializzato in Batch o lotti del campione (token per gli LLM)

# *Learning: Backpropagation*

- L'algoritmo di **retropropagazione** è probabilmente l'elemento costitutivo più fondamentale in una rete neurale. È stato introdotto per la prima volta negli anni '60 e quasi 30 anni dopo (1989) reso popolare da Rumelhart, Hinton e Williams in un articolo intitolato "Apprendimento delle rappresentazioni mediante errori di retropropagazione".
- L'algoritmo viene utilizzato per addestrare efficacemente una rete neurale attraverso un metodo chiamato regola della catena.
- In termini semplici, dopo ogni passaggio in avanti attraverso una rete, la retropropagazione esegue un passaggio all'indietro regolando i parametri del modello (pesi).

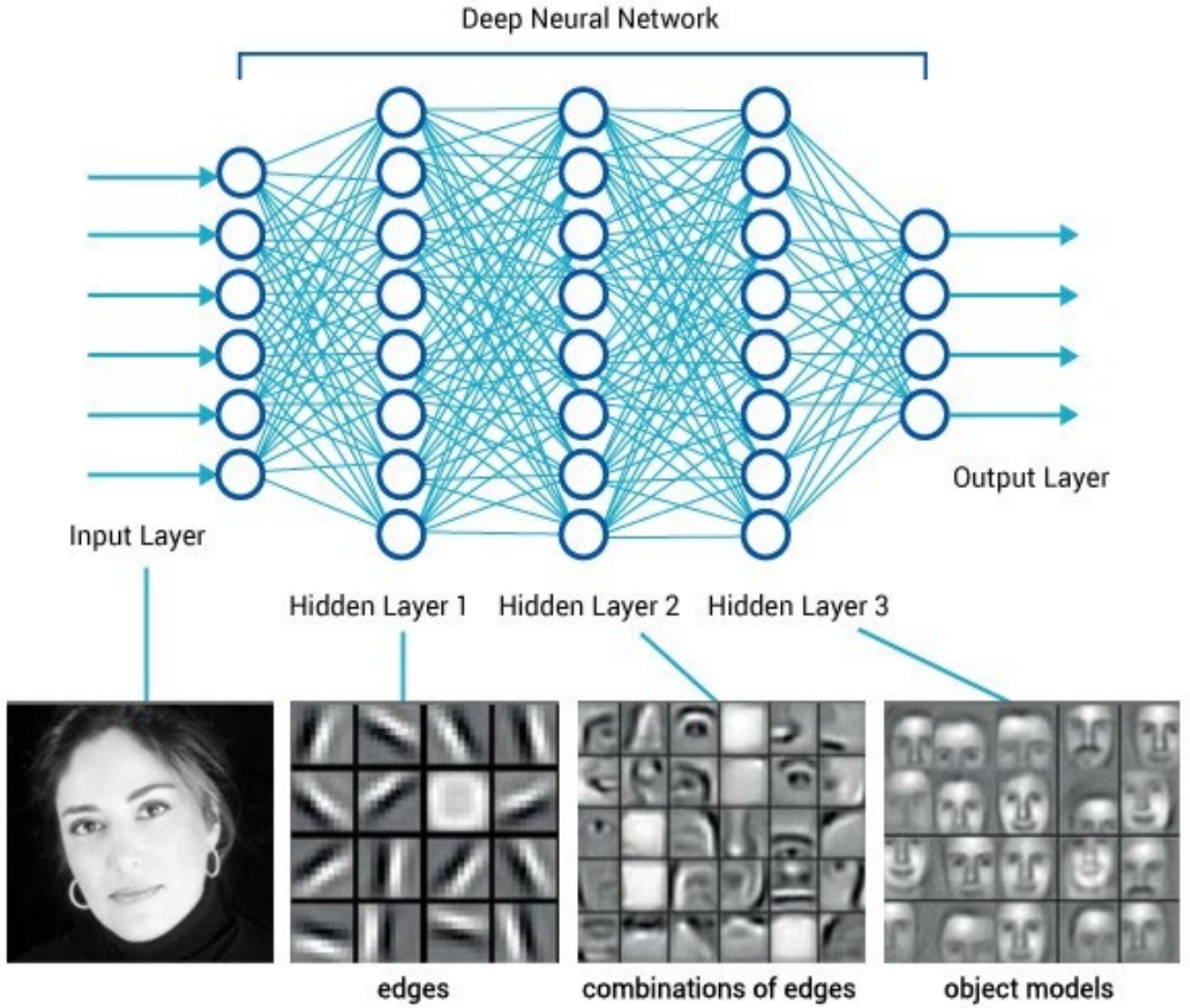



Artificial Neural Network ensemble modeling with conjunctive data clustering for water quality prediction in rivers Artificial Neural Network ensemble modeling with conjunctive data clustering for water quality prediction in rivers 2015

# AI - Deep learning

*Rappresentazione del funzionamento di un modello di deep learning. Questo sistema rappresenta il **concetto** dell'immagine del viso di una persona partendo con l'estrarre **concetti semplici**, come ad esempio angoli e contorni, che a loro volta sono definiti in termini di bordi. Più si procede nel profondo della rete, più questi semplici concetti vengono combinati tra loro per rappresentare caratteristiche complesse e arrivare infine a definire il **modello** dell'oggetto di partenza*

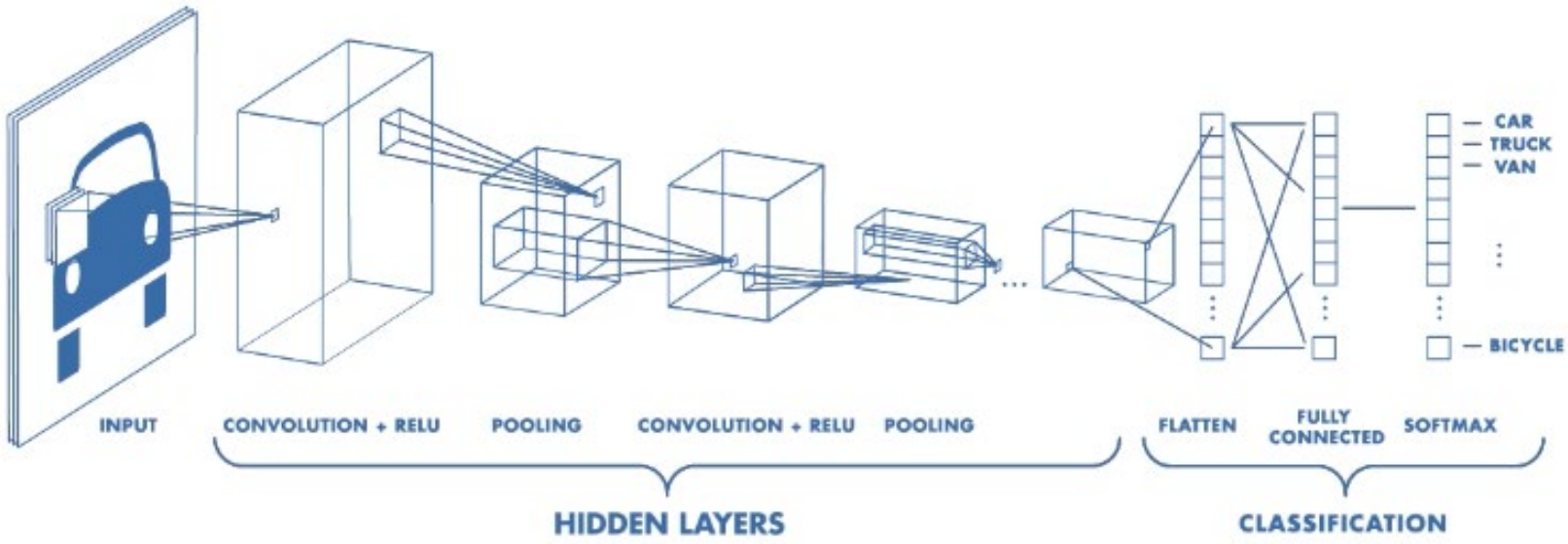
*[https://medium.com/diaryofawannapreneur/deep-learning-for-computer-vision-for-the-average-person-861661d8aa61].*



# Deep learning - Convolutional neural network

Deep learning is a subset of Machine learning.
Il deep learning è una classe di algoritmi di apprendimento automatico che utilizza più livelli per estrarre progressivamente funzionalità di livello superiore dall'input grezzo. Ad esempio, nell'elaborazione delle immagini, gli strati inferiori possono identificare i bordi, mentre gli strati superiori possono identificare i concetti rilevanti per un essere umano come cifre, lettere o volti.
In generale, DL si riferisce a reti neurali con un gran numero di livelli nascosti
Aggiungendo più livelli e più unità all'interno di un livello, una rete profonda può rappresentare funzioni di complessità crescente.

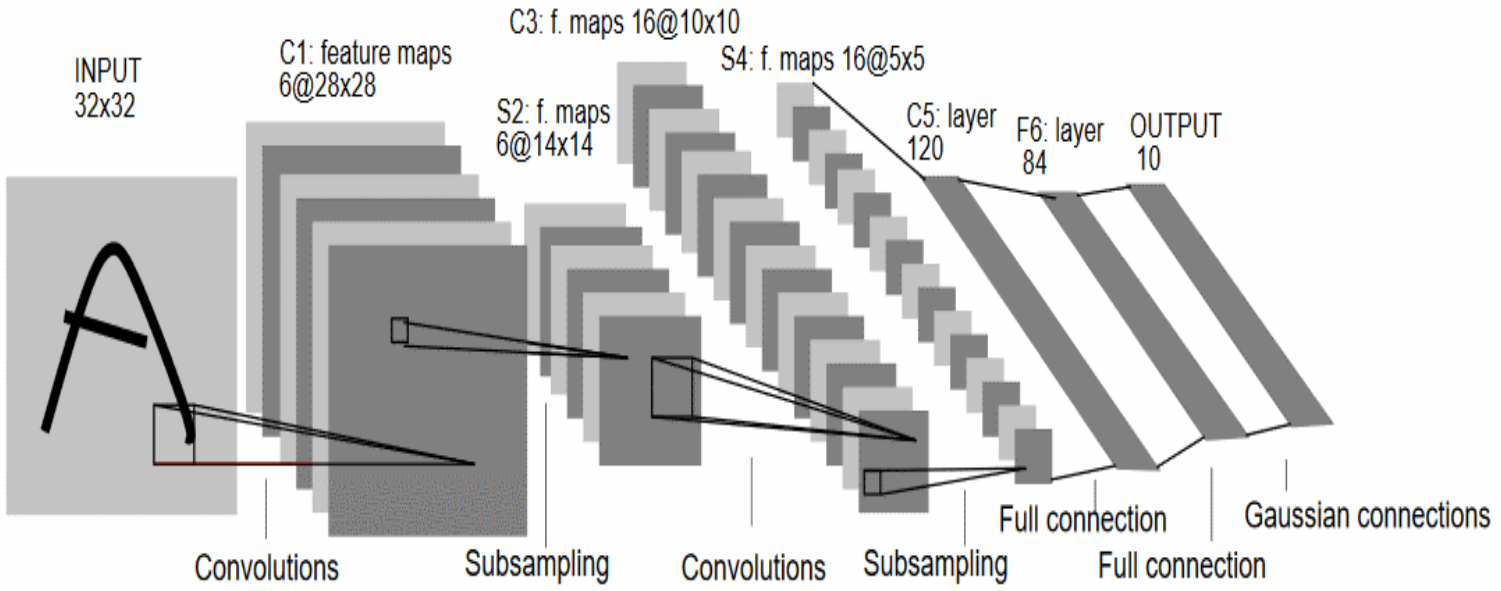
https://www.mathworks.com/videos/introduction-to-deep-learning-what-are-convolutional-neural-networks--1489512765771.html

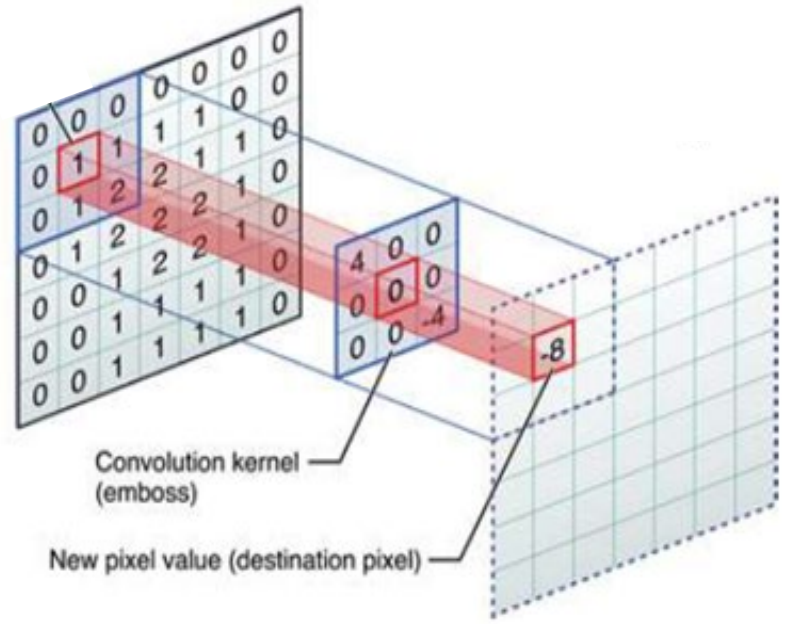



ReLU: rectified linear unit

https://www.mathworks.com/videos/introduction-to-deep-learning-what-are-convolutional-neural-networks--1489512765771.html

# DEEP LEARNING e RETI CONVOLUZIONALI




*Yann LeCun , et al. Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks. 3361 (10) 10995*

# Machine learning

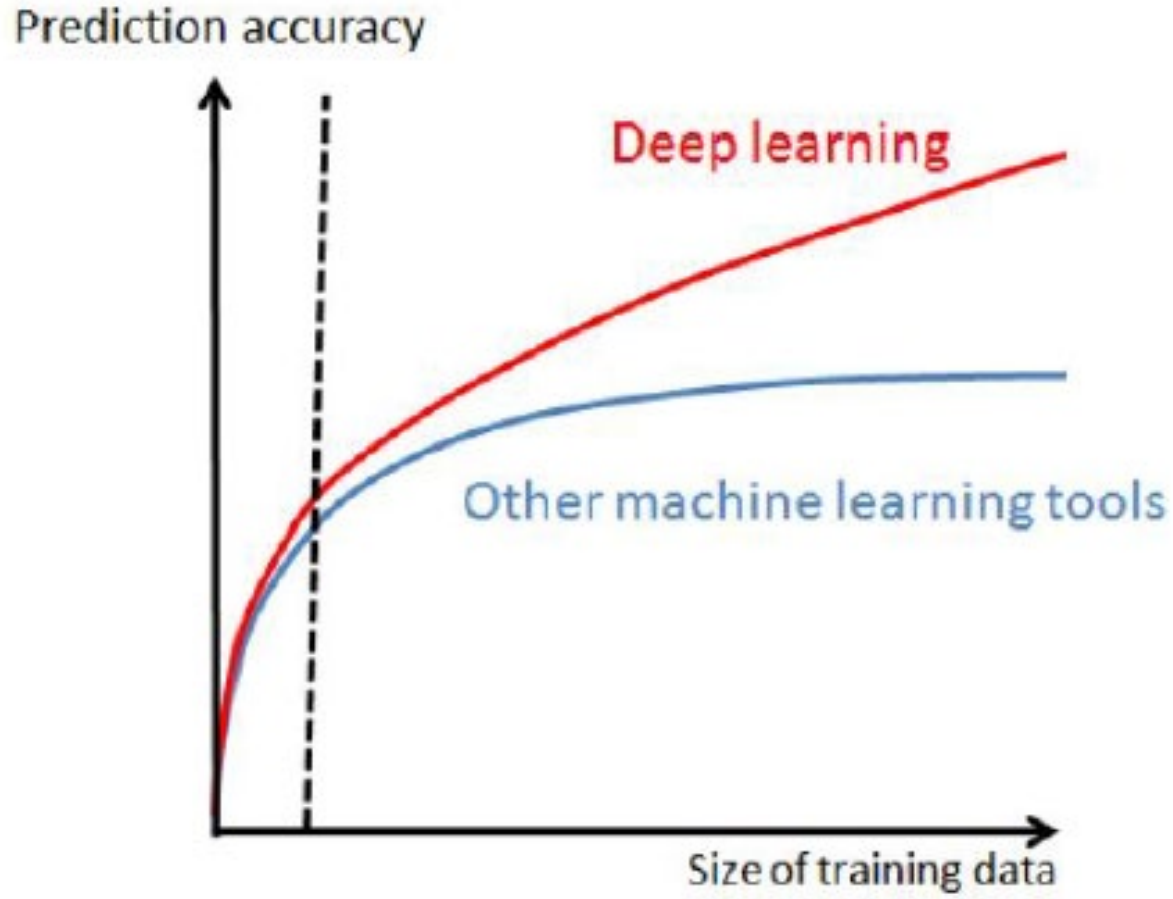
***Training phase***

- apprendimento non supervisionato (gli algoritmi sperimentano un set di dati contenente molte caratteristiche, quindi apprendono le proprietà utili della struttura di questo set di dati)
- L'apprendimento semi-supervisionato è un approccio all'apprendimento automatico che combina una piccola quantità di dati etichettati con una grande quantità di dati non etichettati durante l'addestramento. L'apprendimento semi-supervisionato si colloca tra l'apprendimento non supervisionato (senza dati di training etichettati) e l'apprendimento supervisionato (con solo dati di training etichettati).
- apprendimento supervisionato (gli algoritmi sperimentano un set di dati contenente funzionalità, ma ogni esempio è anche associato a un'etichetta o a una destinazione)

***Testing phase***

La misura delle prestazioni P viene valutata usando un set di dati di test separato dai dati usati per il training del sistema di Machine Learning.

# Deep learning




La superiorità delle tecniche di deep learning rispetto ad altri approcci si manifesta quando sono disponibili grandi quantità di dati di training. (Credits: D. Maltoni)

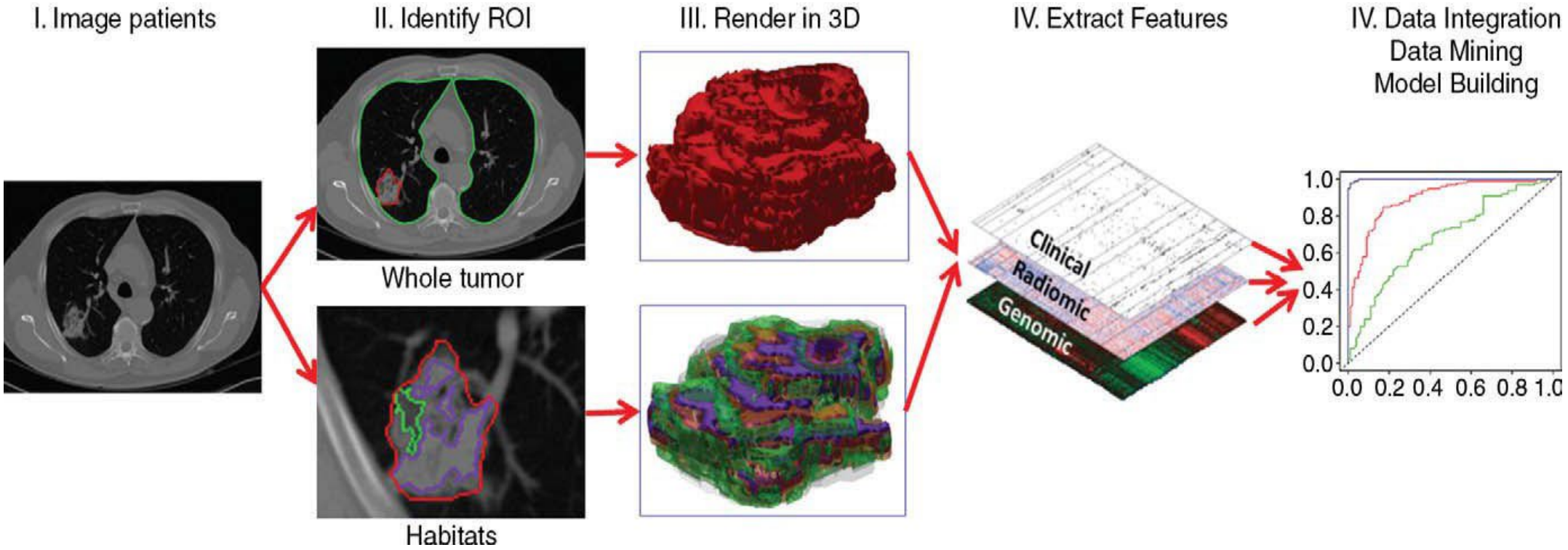
# Machine Learning questo sconosciuto

Quanto deve imparare?

- *Rule of thumb*:
- Un algoritmo di Deep Learning con Supervisione (cioè con immagini annotate con l'esito chiaro della patologia)  può raggiungere **performance accettabili** con circa **5,000 esempi annotati** per patologia, e raggiungere o eccedere la **performance umana (ad es radiologo) quando è stato addestrato con un "**dataset" contenente almeno **10 milioni di esempi annotati**.
- (*Deep Learning, Goodfellow et al. 2016*)

- *Un  bimbo ha bisogno di molta meno istruzione per riconoscere cani o gatti (ne basta anche uno) !!!!!*

# Scienze Radiologiche (processi con grandi apparecchiature: Tac, Adronterapia, Interventistica…) → Radiomica




Il diagramma di flusso mostra il processo della radiomica e l'uso della radiomica nel supporto alle decisioni. Il work-up del paziente richiede che le **informazioni provenienti da fonti disparate siano combinate in un modello coerente per descrivere dove si trova la lesione, cos'è e cosa sta facendo**. La radiomica inizia con l'acquisizione di immagini di alta qualità. Da qui è possibile identificare una regione di interesse (ROI) che contiene l'intero tumore o sottoregioni (cioè gli habitat) all'interno del tumore. Questi sono segmentati con modifiche dell'operatore e vengono infine renderizzate in tre dimensioni (3D). Le caratteristiche quantitative sono estratte da questi volumi renderizzati per generare un report, che viene inserito in un database insieme ad altri dati, come i dati clinici e genomici. Questi dati vengono usati per sviluppare modelli diagnostici, predittivi o prognostici per i risultati di interesse

# Metodi automatici

La sfida è:

**Riconoscere e distinguere**

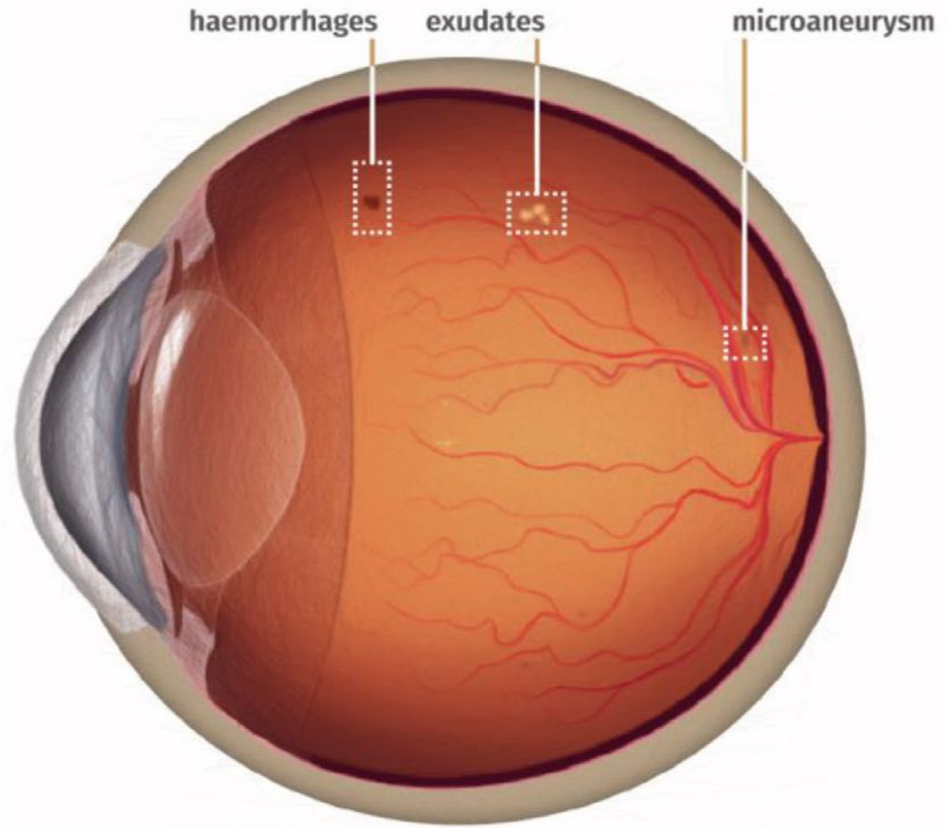
**malformazioni** (shape/colore)

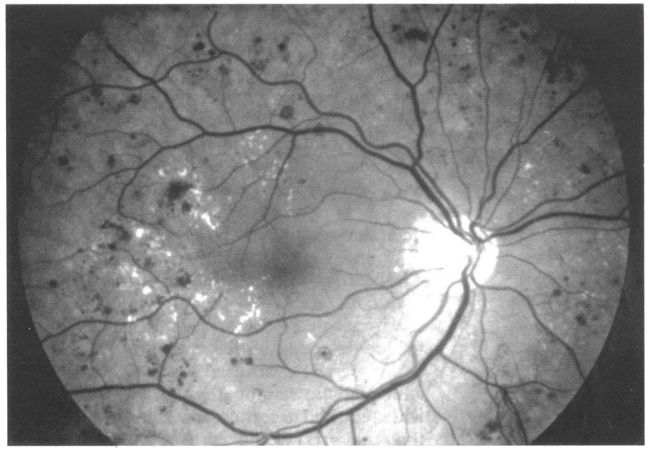
in un campo fortemente

caratterizzato dal **fondo giallo e rosso (bassa dinamica)**,

non simile alla foto in bianco e nero.

Necessità di uno o più algoritmi che possano separare le diverse patologie!!!



Information Sciences 420 (2017) 66–76

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins



# Automated segmentation of exudates, haemorrhages, microaneurysms using single convolutional neural network

Jen Hong Tan[a], Hamido Fujita[b,*], Sobha Sivaprasad[c], Sulatha V. Bhandary[d], A. Krishna Rao[d], Kuang Chua Chua[a], U. Rajendra Acharya[a,e,f]

[a] Department of Electronics and Computer Engineering, Ngee Ann Polytechnic, Singapore
[b] Iwate Prefectural University, Faculty of Software and Information Science, Iwate 020-0693, Japan
[c] NIHR Moorfields Biomedical Research Centre, London, UK
[d] Department of Ophthalmology, Kasturba Medical College, Manipal 576104 India
[e] Department of Biomedical Engineering, School of Science and Technology, SIM University, Singapore
[f] Department of Biomedical Engineering, Faculty of Engineering, University of Malaya, Malaysia

accumuli di liquido, lipidi e proteine che fuoriescono dai vasi sanguigni e si depositano all'interno dello strato retinico. Possono essere "duri" (giallastri, con margini netti) o "molli" (grigio-biancastri, con margini sfumati, spesso detti "cotton wool"). La loro presenza è spesso indicativa di patologie come retinopatia ipertensiva o diabetica

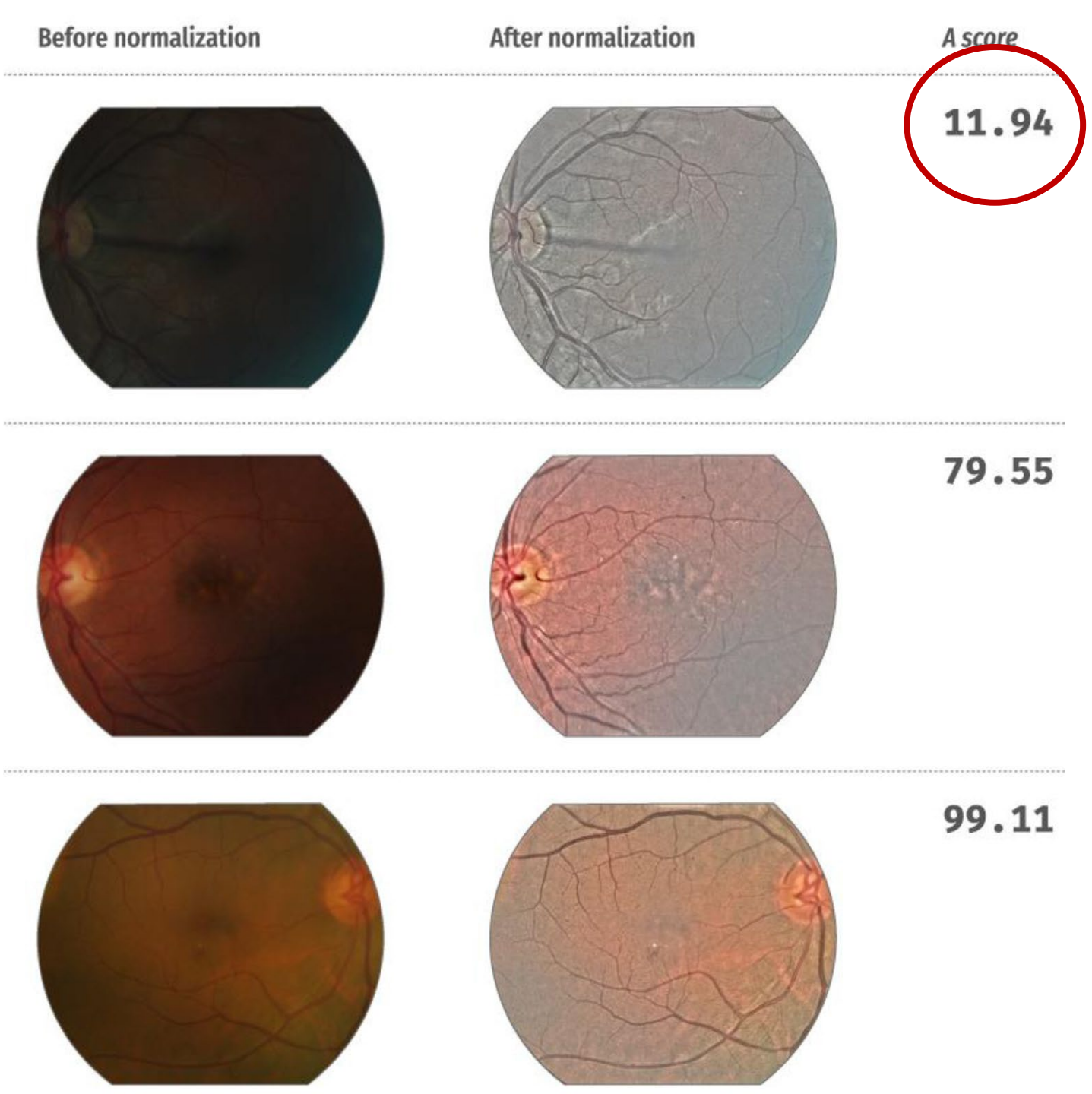
# Pre-Elaborazione
## (Qualità del dato)

Normalizzazione:

Le immagini vengono ridefinite in size, convertite da RGB in uno spazio di comodo (LUV),

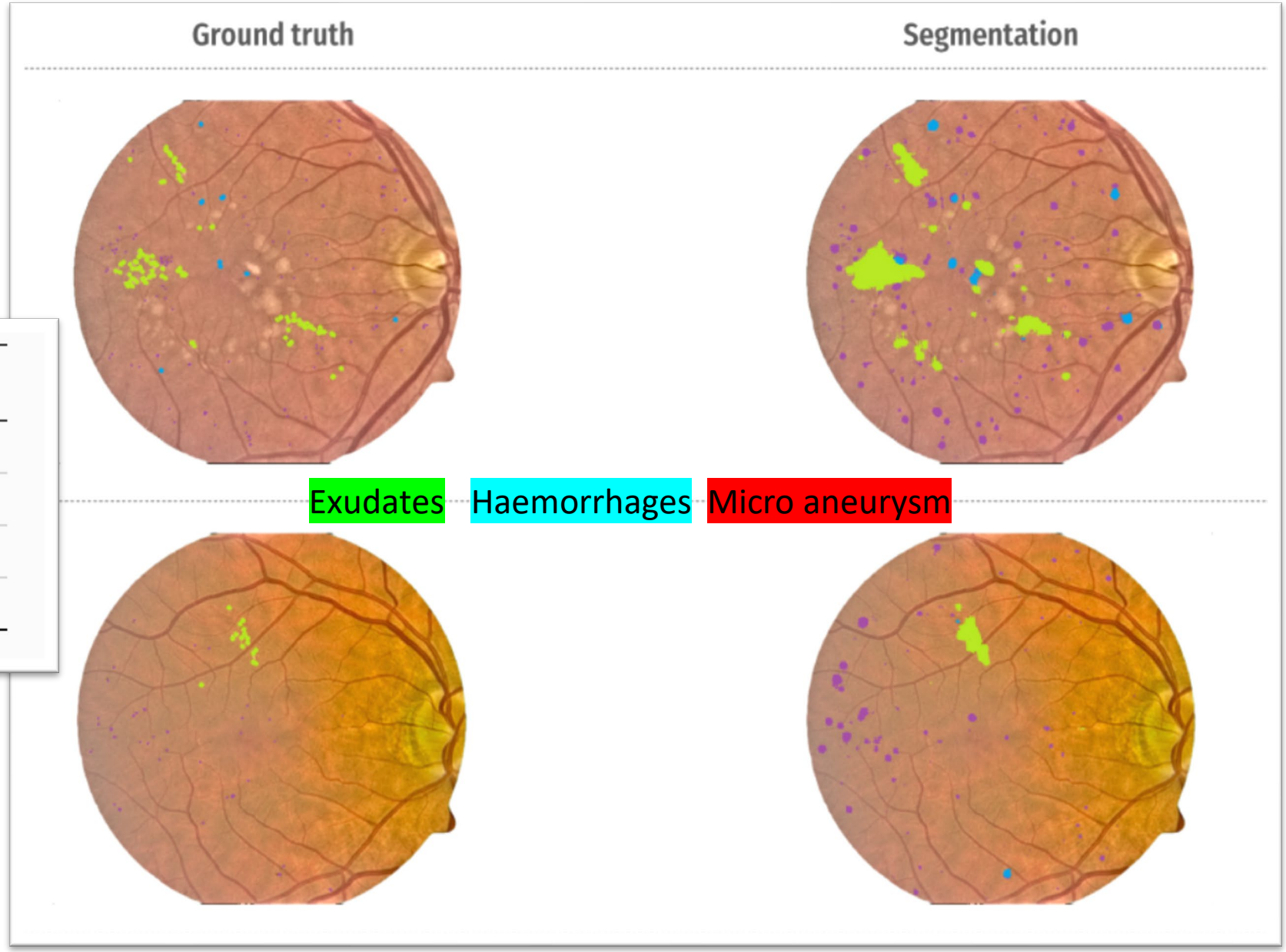
Uniformata la luminanza (i neri diventano grigi e mostrano eventuali particolari), e riconvertite in RGB, quindi normalizzata.

**A** ci dice quanto è grigia l'immagine, cioè presa in **condizioni inappropriate di luce**



# Sensibilità
# Specificità

| Class | Sensitivity | Specificity |
|---|---|---|
| **Background** | 0.9572 | 0.7875 |
| **Exudates** | 0.8758 | 0.9873 |
| **Haemorrhages** | 0.6257 | 0.9893 |
| **Micro-aneurysms** | 0.4606 | 0.9799 |



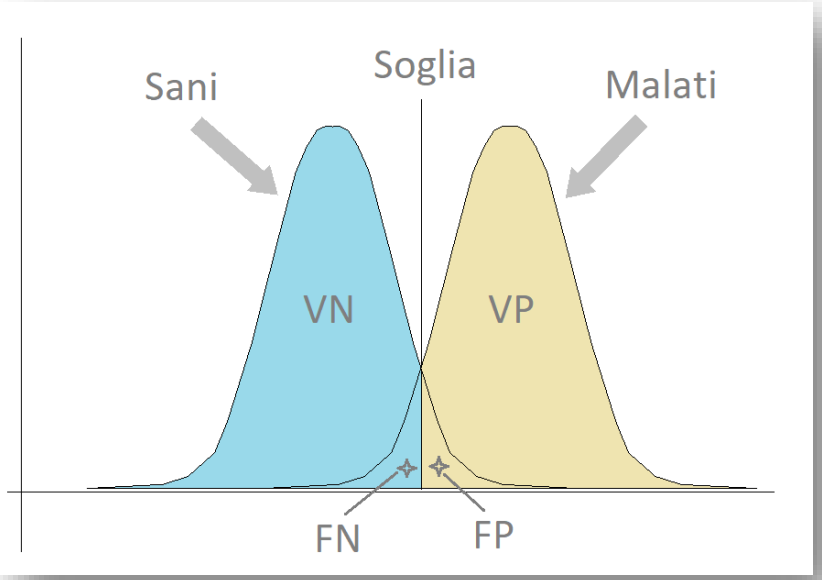

Differenti siti vengono uniti o amplificati

# Sensibilità e Specificità di una misura / test

- In statistica, la **sensibilità** di un test corrisponde alla sua capacità di dare un risultato positivo quando l'ipotesi è verificata.
- **specificità**, invece, misura la capacità di un test di dare un risultato negativo quando l'ipotesi non è verificata.

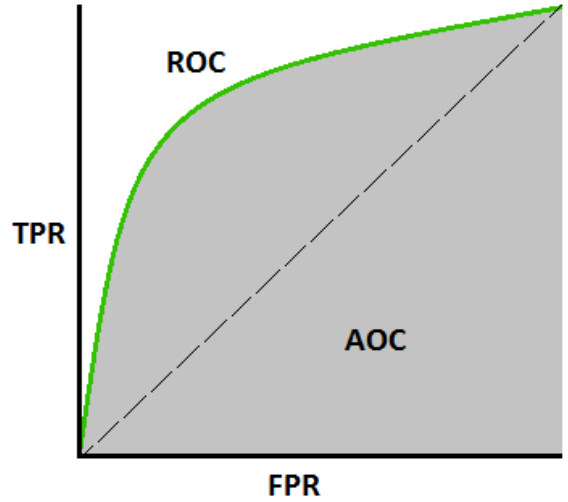
I risultati sono classificati in quattro categorie:

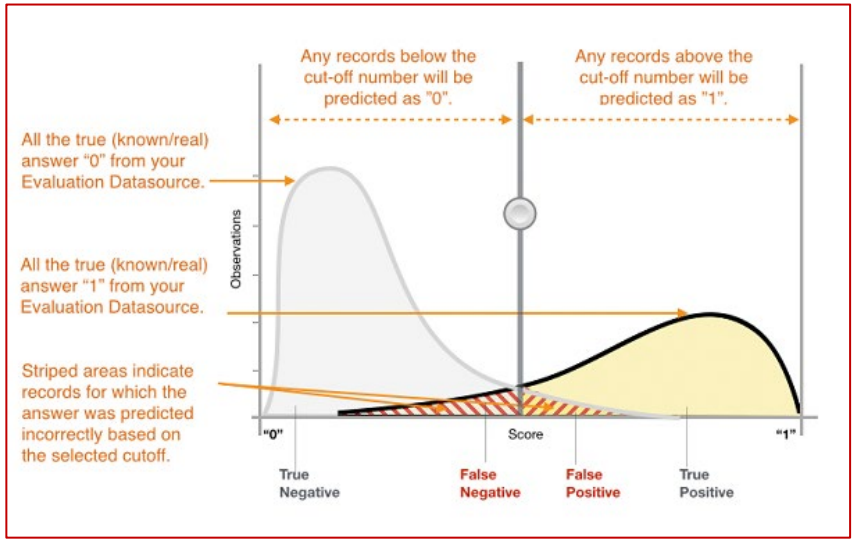
- **I veri positivi VP**: il numero di persone che risultano positive al test e sono malate (**sensibilità**)
- Falsi positivi FP: il numero di persone che risultano positive al test e non sono malate
- Falsi negativi FN: il numero di persone che risultano negative al test e sono malate
- **I veri negativi VN**: il numero di individui che sono risultati negativi al test e non erano malati (**specificità**)

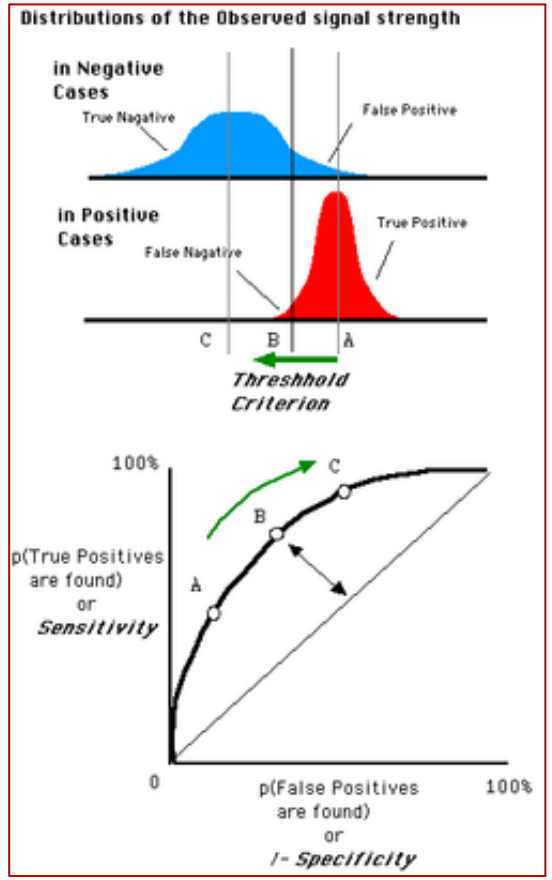


# ROC – AUC

- La curva ROC (Receiver Operating Characteristic) è un grafico utilizzato in AI per valutare le prestazioni di un modello di classificazione binaria, mostrando la relazione tra il tasso di **veri positivi** (TPR) e il tasso di **falsi positivi** (FPR) a diverse soglie. In sostanza, aiuta a visualizzare quanto bene un modello riesce a distinguere tra due classi (es: positivo/negativo, spam/non spam)
- AUC sta per Area Under The Curve e si riferisce a singolo valore numerico che riassume le prestazioni complessive del modello.
- Un AUC pari a 1 indica un modello perfetto, mentre un AUC pari a 0.5 un pessimo modello. Vedrai più avanti uno schema per consultare i valori intermedi che sono poi quelli reali.



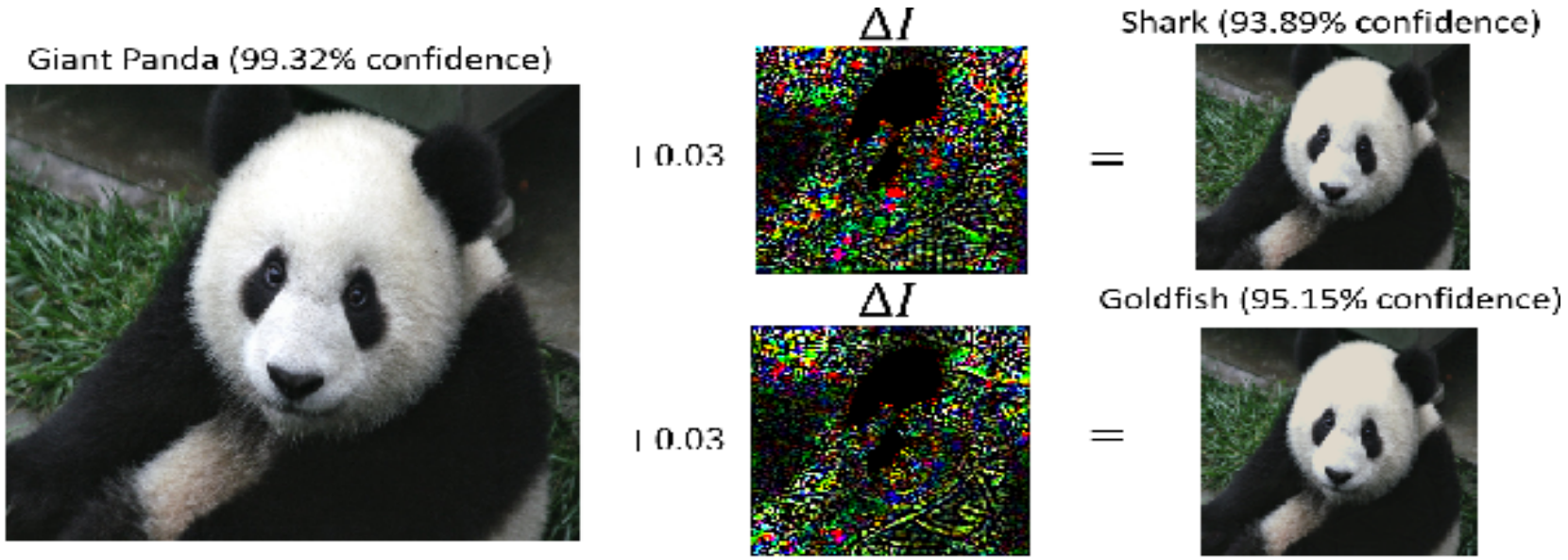



# Requisiti tecnici essenziali

Robustezza

Spiegabilità

# Adversarial examples (AE) – Criticità




From: Xin Li, Fluin Li, IEEE International Conference on Computer Vision, 2017

**Impercettibile perturbazione** per l'occhio umano, stravolge la predizione

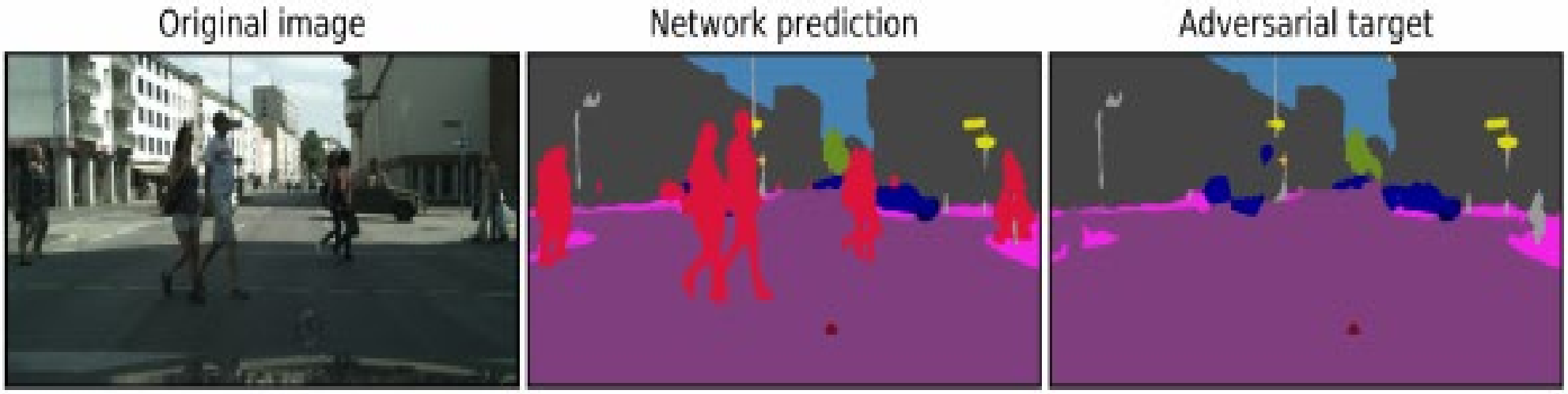
«*AE's have become an indisputable threat to the security of modern AI systems based on DNN*»
arXiv:1807.01069

La **«piccola» perturbazione** può provenire da un'azione intenzionale (minaccia) ma anche da un malfunzionamento della strumentazione, potenzialmente non rilevabile con i protocolli di controllo di qualità attualmente applicati.

# Robustness of AI tools

Le reti neurali profonde sono state recentemente trovate **vulnerabili** a campioni di input ben progettati, chiamati **esempi contraddittori**. Gli esempi contraddittori sono impercettibili per l'uomo, ma possono facilmente ingannare le reti neurali profonde nella fase di test/distribuzione.

Un avversario può costruire esempi fisici contraddittori e confondere i veicoli autonomi manipolando il segnale di stop in un sistema di riconoscimento dei segnali stradali o rimuovendo la segmentazione dei pedoni in un sistema di riconoscimento degli oggetti.

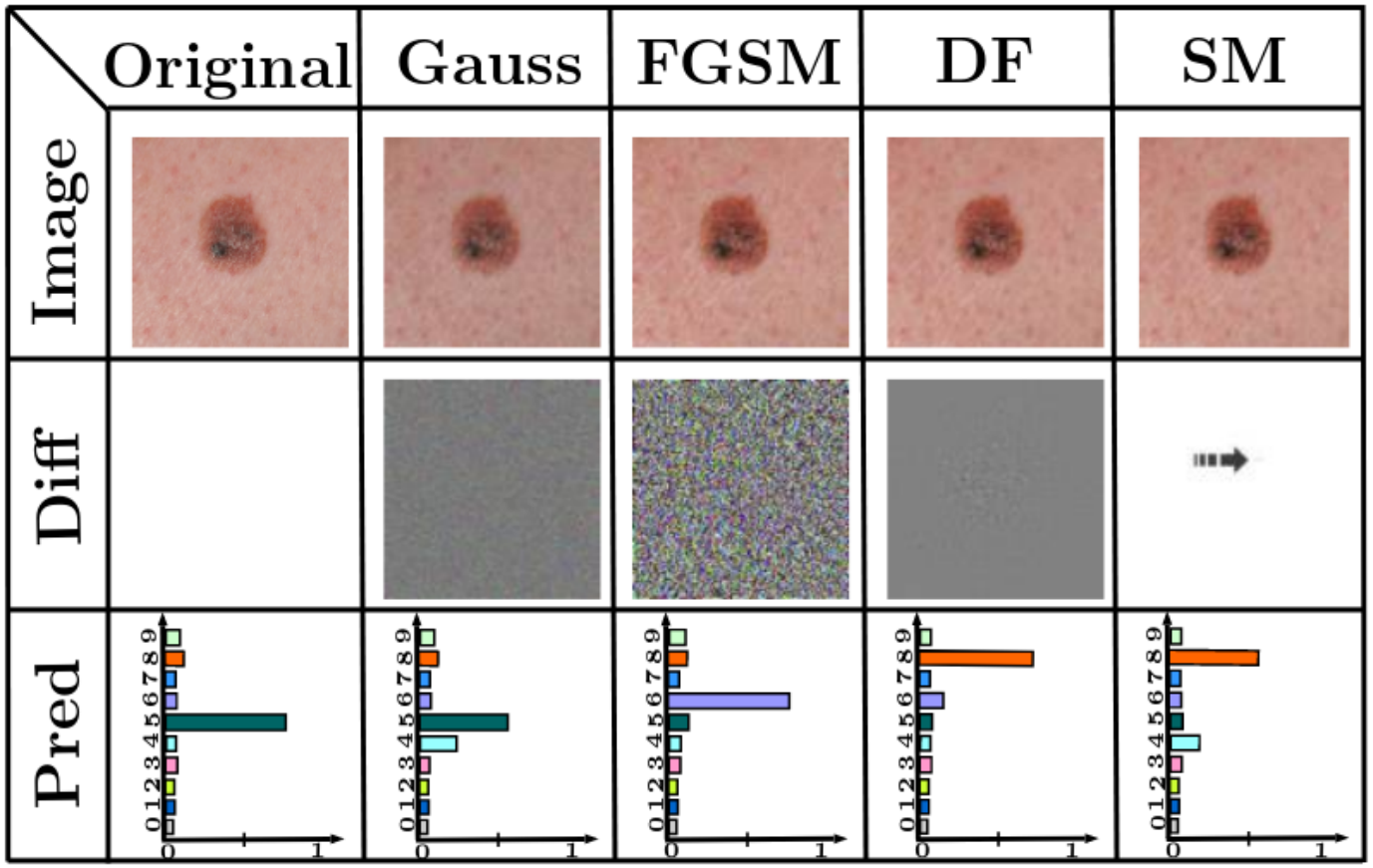



(Yuan et al, 2018) Adversary examples of hiding pedestrians in the semantic segmentation task.
**Left image: original** image; **Middle image**: the segmentation of the original image **predicted** by DNN;
**Right image**: the segmentation of the **adversarial image predicted** by DNN.

# ABITI ANTI-SORVEGLIANZA
# CONTRO IL RICONOSCIMENTO FACCIALE



# Adversarial Example – l'opportunità




Classificazione di lesione della pelle
*From M. Paschali et al, arXiv:1804.00504*

- Gli AE incominciano ad essere utilizzati come benchmark per testing della **robustezza** e generalizzabilità di tool ML, rimpiazzando la necessità di ampi (e costosi) data-set

# Skin Analytics: la prima AI triage approvata senza supervisione umana per il cancro della pelle (UKCA)



https://www.linkedin.com/posts/savevski_skin-analytics-la-prima-ai-triage-approvata-activity-7325047871827771392-pkkk?utm_source=share&utm_medium=member_desktop&rcm=ACoAAATlgE8BLIMGuHcQihKDloof9HxRhzRHkHE

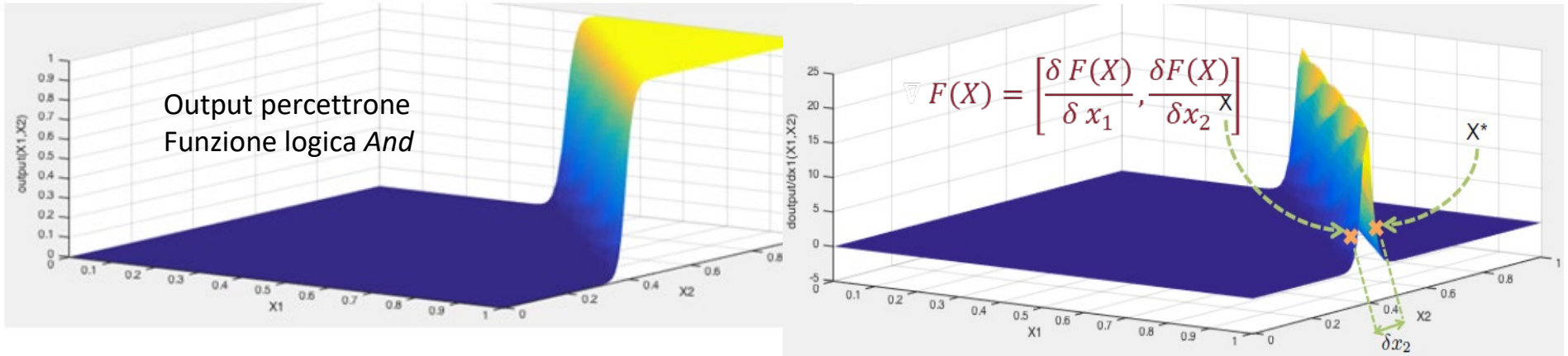
Data una Rete F(X)=Y, trovare la minima perturbazione $\delta_X$ per cui X*=X+ $\delta_X$ e F(X*)= Y*≠ Y

$$arg\ min_{\delta_X} \| \delta_X \|: \ F(X + \delta_X) = Y^*$$

Per la soluzione della precedente relazione esistono differenti metodologie: Gradient-based, Score based , Decision based attacks.

**Jacobian based Saliency Map**

Calcolo della **minima perturbazione** da applicare agli input selezionati per ottenere il cambiamento di classe desiderato

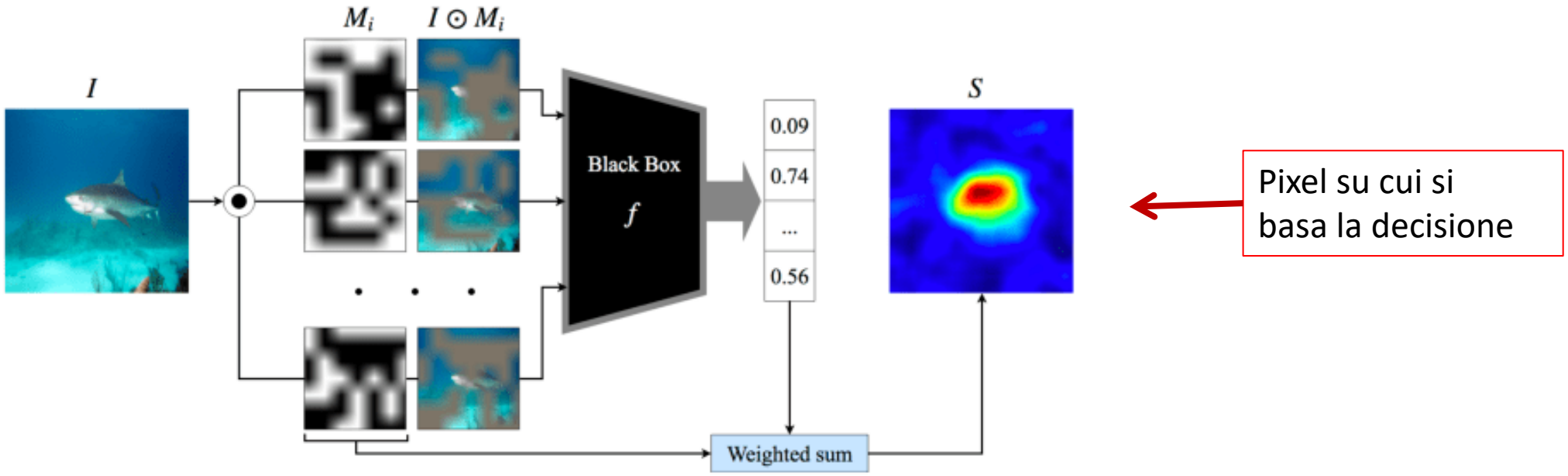



Per piccoli $\delta_{X2}$ ($\delta_{X2}$= 0.05) si hanno grandi cambiamenti dell'output (F(*x1,x2*) passa da uno a zero)

Nicolas Papernot, Patrick McDaniel, Somesh Jhay, Matt Fredriksonz, Z. Berkay Celik, Ananthram Swami. *The Limitations of Deep Learning in Adversarial Settings.* Accepted to the 1st IEEE European Symposium on Security & Privacy, IEEE 2016. Saarbrucken, Germany

# Criticità: explainability e saliency maps

Sono state proposte soluzioni al problema della **opacità** delle reti DL.
Per esempio, si possono generare "**mappe di importanza**" (*saliency maps*), che indicano quanto ogni pixel determina la predizione del modello, senza alcuna conoscenza dell'architettura della rete.
Le mappe di importanza sono costruite sintetizzando delle opportune maschere, che vengono poi applicate ai dati di ingresso della rete, per valutare infine l'output corrispondente.

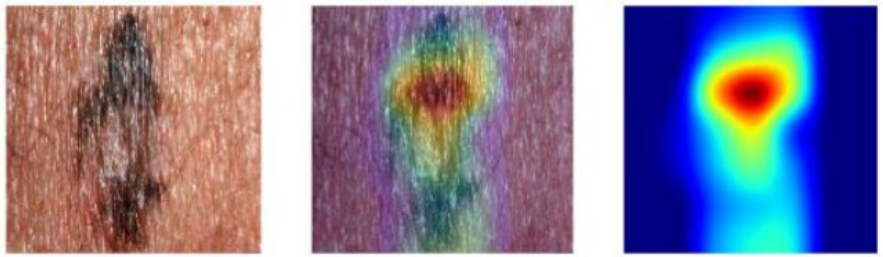



RISE applies **random masks** to inputs and measures the value of each pixel

Petsiuk et al, RISE: Randomized Input Sampling for Explanation of Black-box Models

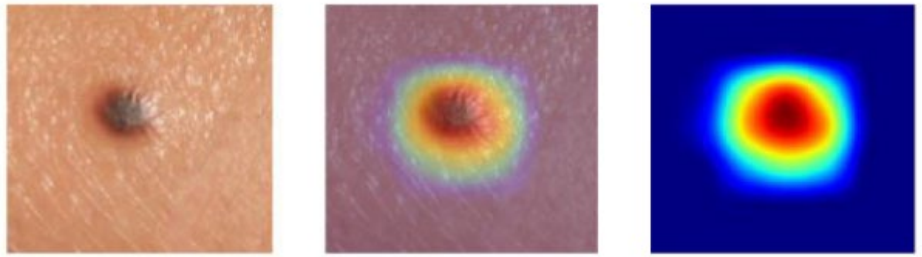
# BISOGNO DI MODELLI SPIEGABILI E INTERPRETABILI

- **Verifica del sistema**: In molte applicazioni non ci si può fidare di un sistema di default, come nell'assistenza sanitaria dove l'uso di modelli che possono essere interpretati e verificati da esperti medici è una necessità assoluta.
- CORRETTA CLASSIFICAZIONE ERRATA CLASSIFICAZIONE



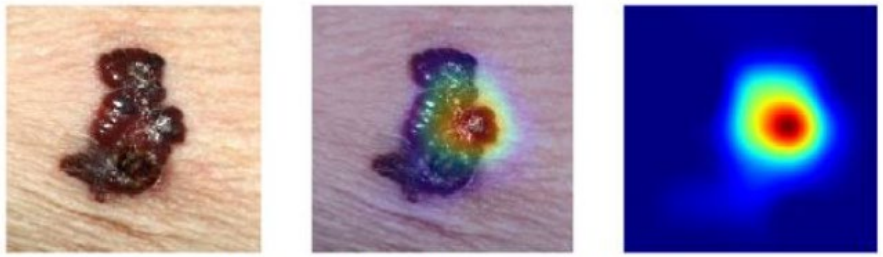
(a) Melanoma with 100% confidence.

Source: Edinburgh dataset.



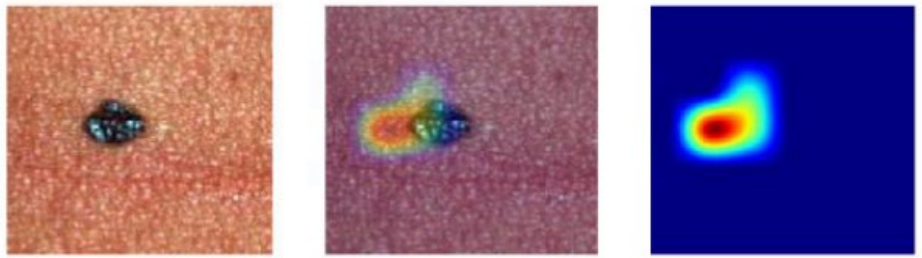
(a) Predicted as Haemangioma with 100% confidence.

Source: Edinburgh dataset.



(b) Melanoma with 100% confidence.

Source: Edinburgh dataset.



(b) Predicted as Basal Cell Carcinoma with 100% confidence.

Source: Edinburgh dataset.

*From M. Paschali et al, arXiv:1804.00504*

# INNOVAZIONE ULTERIORE:
# Chat e gli Algoritmi Generativi

- Rispetto alla ricerca tradizionale sul *machine learning (Applicazioni VERTICALI)*, ultimamente gli strumenti prodotti da quest'ultima stanno diventando sempre più alla portata di tutti.
- Un esempio è dato da ***chatGPT***, che offre **un'interfaccia molto semplice ad un modello NLP**, addestrato con un numero esorbitante di testi di dominio pubblico. Grazie a questo, milioni di utenti hanno potuto avere esperienza delle capacità di tali modelli di risolvere compiti tradizionalmente considerati prerogativa solo dell'essere umano, in quanto richiedono capacità di comprensione e analisi di problemi: **riassunto testi, imitazione di stili di autori letterari (pastiche), creazione di disegni, foto, canzoni, video…**
- ***TRIAGE CHATBOT: Software per supporto psicologico*. L'**AI è la base per costruire **strumenti conversazionali (*chatbot*), in grado di suggerire l'opportunità per il soggetto di intraprendere un percorso di cura (*self-referral*)**. Il Regno Unito ha una app (Limbic) registrata come **DM di classe IIa** UKCA (UK Conformity Assessed).
- La chatbot può classificare disturbi di salute mentale comuni trattati da un programma del NHS "Talking Therapies" , con una accuratezza del 93%.
- Il programma NHS Talking Therapies [https://www.england.nhs.uk/mental-health/adults/nhs-talking-therapies/] è stato sviluppato per migliorare l'erogazione e l'accesso alle terapie psicologiche per la depressione e i disturbi d'ansia, raccomandate dal NICE e basate sull'evidenza, all'interno del NHS.
- Nel 2021/22 quasi 1,2 milioni di persone hanno potuto accedere ai servizi.

# LLM

Embedding

come possono i large language model generare proteine

| ChatGPT | Modello specializzato |
| --- | --- |
| I grandi modelli di linguaggio, come ChatGPT, non sono in grado di generare direttamente proteine. La generazione di proteine è un processo molto complesso e richiede una conoscenza approfondita della biologia molecolare e delle interazioni tra gli amminoacidi.<br><br>I large language model come ChatGPT sono stati addestrati su testi di diverse discipline, inclusa la biologia, ma la loro capacità di generare proteine è limitata alla sintesi di sequenze di testo che potrebbero rappresentare potenziali sequenze di amminoacidi. Tuttavia, queste sequenze generate non avrebbero alcun significato biologico reale a meno che non siano validate e verificate attraverso metodi sperimentali appropriati.<br><br>La progettazione e la generazione di proteine sono solitamente svolte da scienziati specializzati nel campo della biochimica e della biologia molecolare. Utilizzano sofisticati strumenti di progettazione assistita dal computer (CAD) e software di modellazione molecolare per creare nuove sequenze di amminoacidi che abbiano determinate proprietà o funzioni specifiche. Successivamente, queste sequenze possono essere sintetizzate in laboratorio e sottoposte a test e analisi per valutare le loro proprietà e funzioni.<br><br>In sintesi, mentre i large language model possono generare sequenze di testo che potrebbero corrispondere a sequenze di amminoacidi, la progettazione e la generazione di proteine funzionali richiedono competenze e approcci specializzati che vanno al di là delle capacità di questi modelli. | I Large Language Model (LLM) sono strumenti basati sull'intelligenza artificiale che, utilizzando il machine learning, sono in grado di generare sequenze di testo in modo autonomo. Tuttavia, grazie all'avanzamento delle tecnologie, oggi è possibile utilizzarli anche per generare sequenze proteiche.<br><br>La creazione di proteine sintetiche tramite i LLM potrebbe essere un passo importante nella lotta alle malattie infettive come il COVID-19. Infatti, grazie alla capacità dei LLM di generare sequenze di proteine, si potrebbero progettare nuove proteine che agiscano come vaccini o terapie contro il virus.<br><br>Tuttavia, come evidenziato dagli studiosi, il processo di creazione di proteine sintetiche non è privo di rischi. Infatti, potrebbero formarsi grumi di cellule fuse che causerebbero danni ai tessuti e complicanze immunitarie.<br><br>Inoltre, è stato scoperto che la proteina Spike presente nel virus SARS-CoV-2, che viene utilizzata come bersaglio dal vaccino, ha una sequenza simile a una tossina. Ciò potrebbe causare alcune deregolazioni neurologiche nel COVID-19 e, teoricamente, anche nei vaccinati.<br><br>Nonostante questi rischi, i LLM rappresentano una grande opportunità per la creazione di nuove terapie e vaccini contro le malattie infettive. Sarà importante, quindi, continuare a studiare e approfondire le conoscenze su questa tecnologia per garantirne un uso sicuro ed efficace. |

# Legge di scala o distillazione ?

*https://www.wired.it/article/legge-scala-intelligenza-artificiale-low-cost/*

- Come si legge in un lungo report del think tank Bruegel, “dall'inizio dell'attuale generazione di large language model (Llm) nel 2017 fino alla metà del 2024, i principali Llm hanno investito massicciamente nella fase di pre-addestramento dei modelli di intelligenza artificiale, che richiede enormi volumi di dati di addestramento (o token) e una grande potenza di calcolo per alimentare miliardi di parametri interni. Gli Llm hanno seguito la legge di scala: per ottenere prestazioni migliori servono più dati di addestramento e maggiore capacità computazionale. Questo ha spinto gli investimenti in strutture di calcolo iperscalabili, dotate di chip specializzati per l'intelligenza artificiale”.
- Un’alternativa potrebbe essere rappresentata da DeepSeek, la startup cinese che è stata in grado di sviluppare modelli dalle prestazioni vicine a ChatGPT e ai suoi fratelli investendo una frazione dei soldi: “DeepSeek ha utilizzato il proprio Large language model, V3, insieme ad altri Llm come **Qwen** di Alibaba, **Llama** di Meta e probabilmente il modello di ragionamento o1 di OpenAI, per generare 800.000 esempi di ‘chain of thought’ [frasi che mostrano passo dopo passo come arrivare a una risposta, rendendo esplicito il processo di ragionamento del modello, ndr]”, si legge sempre sul sito di Bruegel. “[Questo processo] ha il vantaggio di richiedere molta meno potenza di calcolo rispetto al pre-addestramento iniziale del modello ed è ciò che ha permesso a DeepSeek di spendere solo una frazione rispetto ai grandi Llm più noti”.
- È il processo noto come **distillazione**, che consente ai modelli più piccoli di apprendere da quelli già esistenti e più grandi, trasferendo la conoscenza di un modello linguistico di grandi dimensioni (ribattezzato “**insegnante**”) a uno più piccolo (lo “**studente**”), mantenendo prestazioni simili ma con minori costi computazionali. Nella distillazione, il modello più piccolo viene esposto alle risposte e ai ragionamenti del modello più grande invece che ai soli dati grezzi, apprendendo più rapidamente anche schemi complessi.
- In questo modo, l’intelligenza artificiale potrebbe uscire dai data center delle Big tech per diffondersi in un ecosistema fatto di modelli più piccoli e più accessibili, senza rinunciare alla potenza e abbattendo anche i costi di utilizzo per gli utenti finali. Come segnala sempre Bruegel, potremmo sostituire la legge di scala con la “legge di scala della distillazione”, in cui i progressi sono basati sulla capacità di aumentare le prestazioni e l’efficienza di modelli dalle dimensioni simili, e non più sulla costruzione di modelli sempre più grandi.

# ABILITA' EMERGENTI e Allucinazioni

- Trasforma parole in numeri che popolano vettori per la ricerca (cattura il significato per contesti (Token = parola spezzata, veloce o veloce-mente)
- Selezione del materiale la fa l'umano a favore di categorie specialistiche (Medici e Ricercatori)
- **Embedding** → applicazione **di transfer learning** (o più aggiornate o più verticali)
- **Intelligenza** diversa da quella umana !!!!! Fondata su Induzione statistica (?)
- La modalità di interazione è conversazionale.
- Sei tu che devi capire come lui usa i testi delle domande (**Jail Break** per fargli fare compiti vietati da filtri di policy - Safeguards
- Non è Google per cui non puoi usare le parole chiave (20 anni di uso di internet ci ha abituato male)
- Risposta sempre la stessa, ma può evolvere, dipende da un parametro che si passa (temperatura) determina la creatività usata nella risposta (se alta la risposta la cambia)
- Open AI dice che sceglie la parola più probabile !!! (**collegamento stocastico senza consapevolezza**)
- Ma ci sono le **Abilità Emergenti** (le ha apprese (?) da solo, non programmate)(rilevazioni empiriche – benchmark → modelli valutati con un punteggio)(nessuno sa se è completo)
- Ci sono filtri, ma l'etica non è univoca (paesi e dati di addestramento) → la modalità generalista è meno utile
- **Verificare** le risposte !!!!
- Problematica nelle chat di screening o diagnostiche – anche usato da specialisti? Dati nel prompt possono essere condivisi tra gli utenti

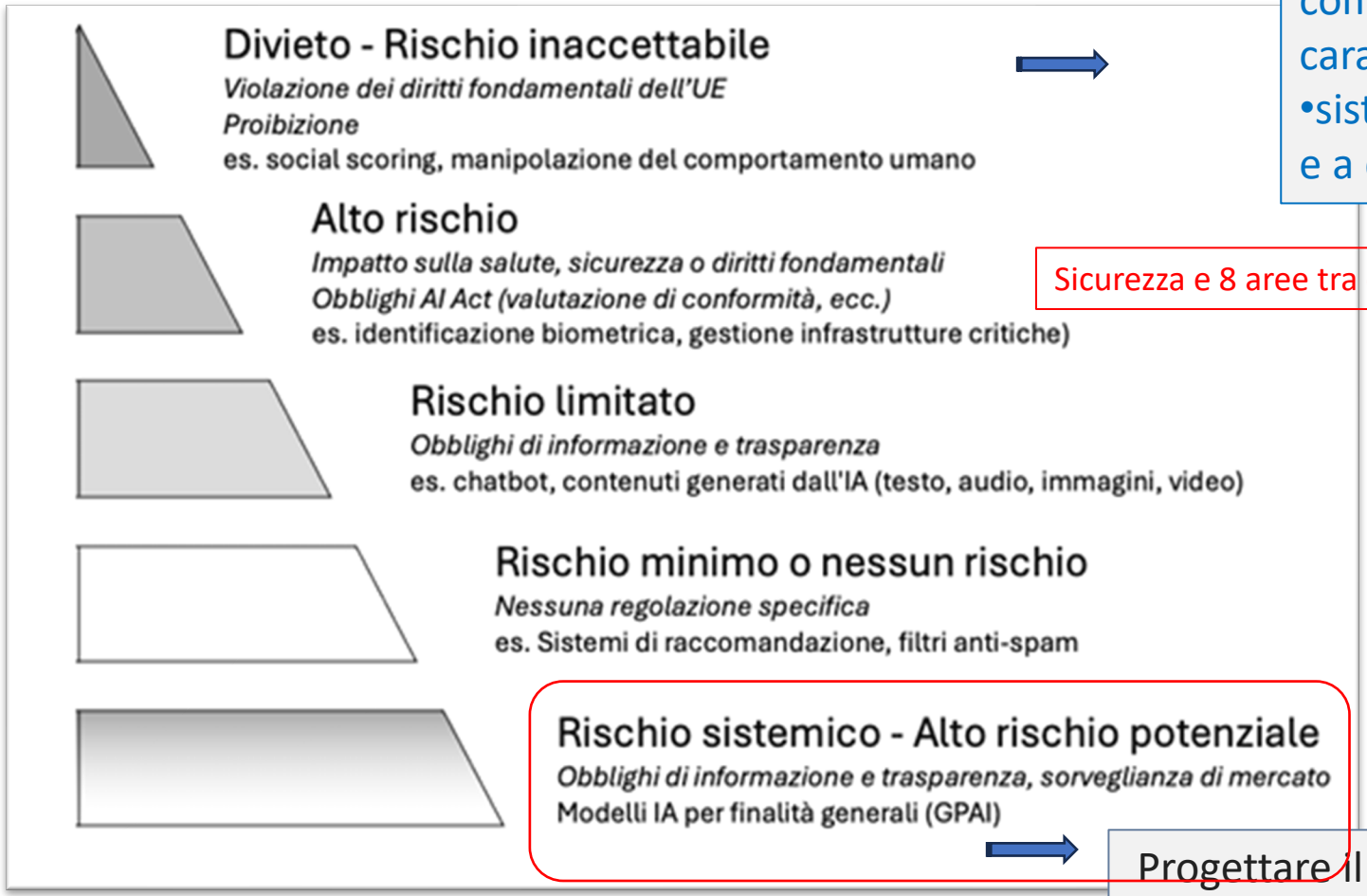
# Personal Generative Chat BOT via LLM

- «You can now create custom versions of *Generative Chat LLM* that combine instructions, extra knowledge, and any combination of skills.”

- Se puoi personalizzarlo, diventa il tuo strumento di lavoro efficiente, sarai tu ad inserire la cultura con la quale vuoi poter ricercare o produrre testi o imparare (self referral)

- Programmi di gestione pazienti dei MMG ???

# AI act e AGID 2025 per la




- **manipolazione comportamentale** cognitiva di persone o gruppi vulnerabili specifici;
- **classificazione sociale** delle persone in base a comportamento, livello socio-economico, caratteristiche personali;
- sistemi di **identificazione biometrica** in tempo reale e a distanza, come il riconoscimento facciale.

Sicurezza e 8 aree tra cui la Salute DM

Progettare il modello GenAI in modo da impedire la generazione di contenuti illegali e pubblicare riepiloghi dei dati con diritti d'autore utilizzati per l'addestramento.

# Big Data (serie ISO 25000)

- **Immagazzinamento** e sicurezza
  - Salvaguardia Privacy (*le immagini sono adeguate?*)
  - Salvaguardia Ownership (*paziente e struttura sanitaria che genera il dato*)
  - Preservabilità dei dati (*robustezza infrastruttura*)
  - Silos o database locali o altre infrastrutture?
- **Accesso** e fruizione dei dati (*service provider data management*)
  - Standardizzazione (standard aperti ?)
  - Robustezza dei protocolli di accesso
  - *Come fronteggiare l'eterogeneità «intrinseca» del sistema sanitario*
- **GDPR** garantisce il diritto alla privacy e il diritto, in caso di diagnosi automatica, alla **spiegabilità** !!!!

Leonard J Kish & Eric J Topol

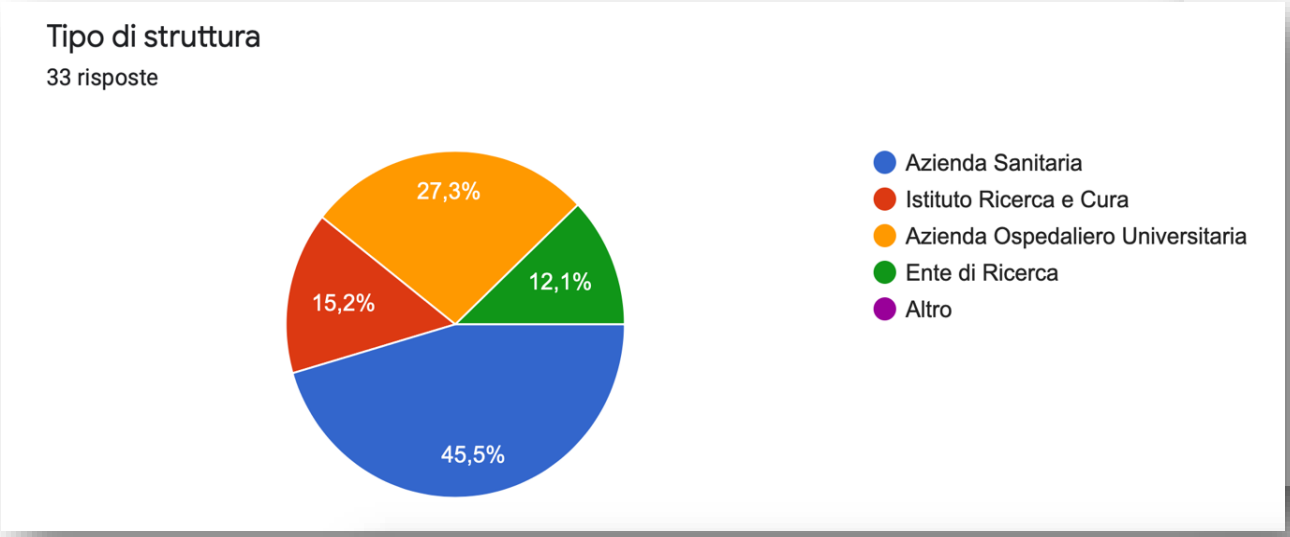
For the benefits of digital medicine to be fully realized, we need not only to find a shared home for personal health data but also to give individuals the right to own them.

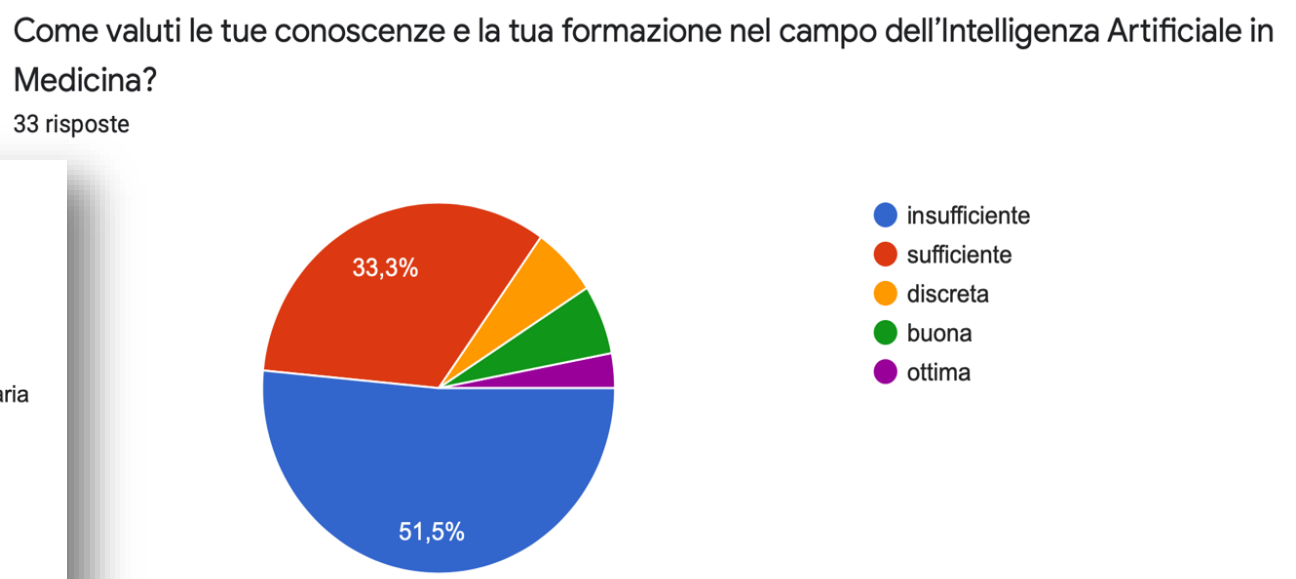
# Gestione dati; esemplificazione: Le Immagini

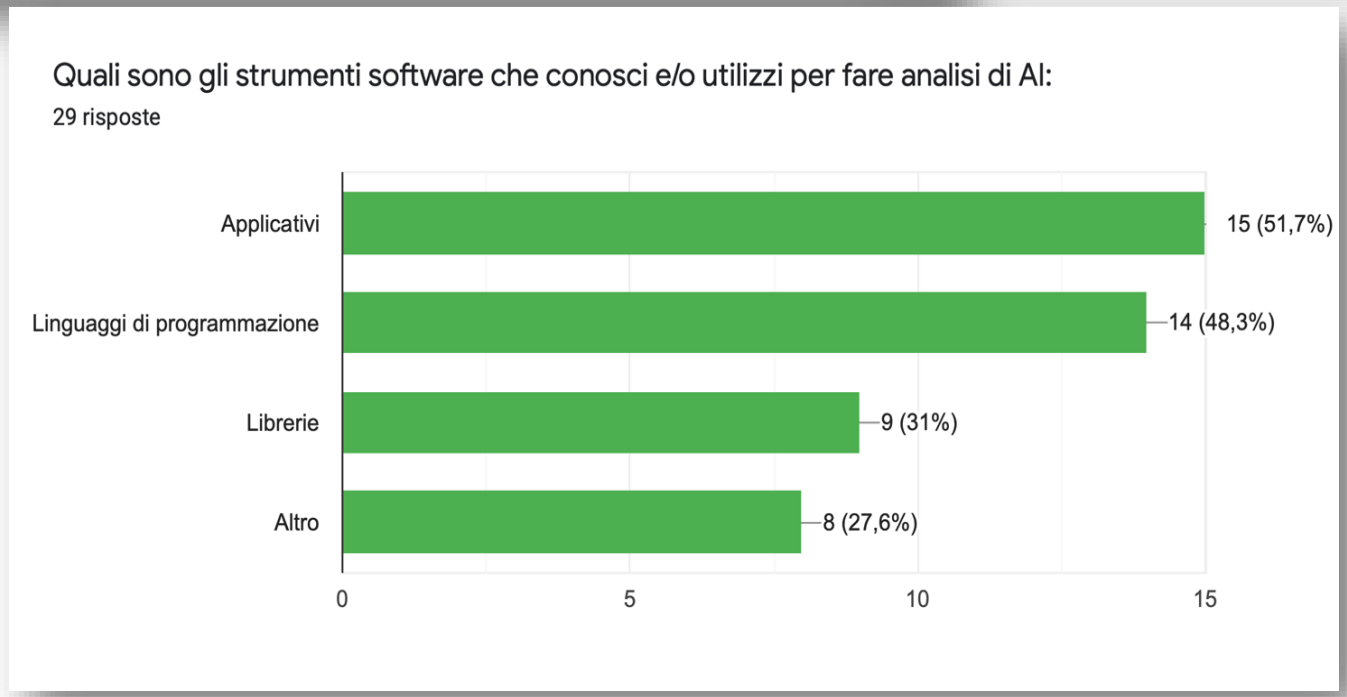
- Su pacs o smartphone (schermi diversi)?, certificazione per la qualità non manomesse (GDPR)
- GDPR minimizzare dati scambiati, non copiare, non manomettere
- Uso del FSE
- **Es.: Uso dei social ?**



# ISS – AIFM (2021)



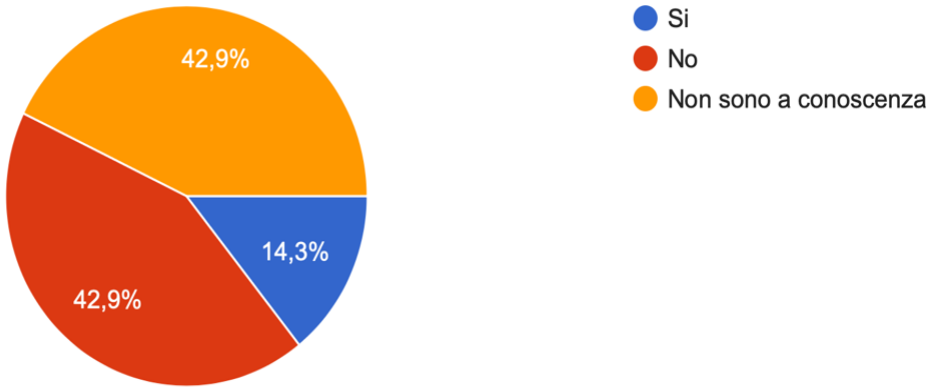




M. Iori and C. Talamonti on behalf of AIFM AI group

# ISS - AIFM

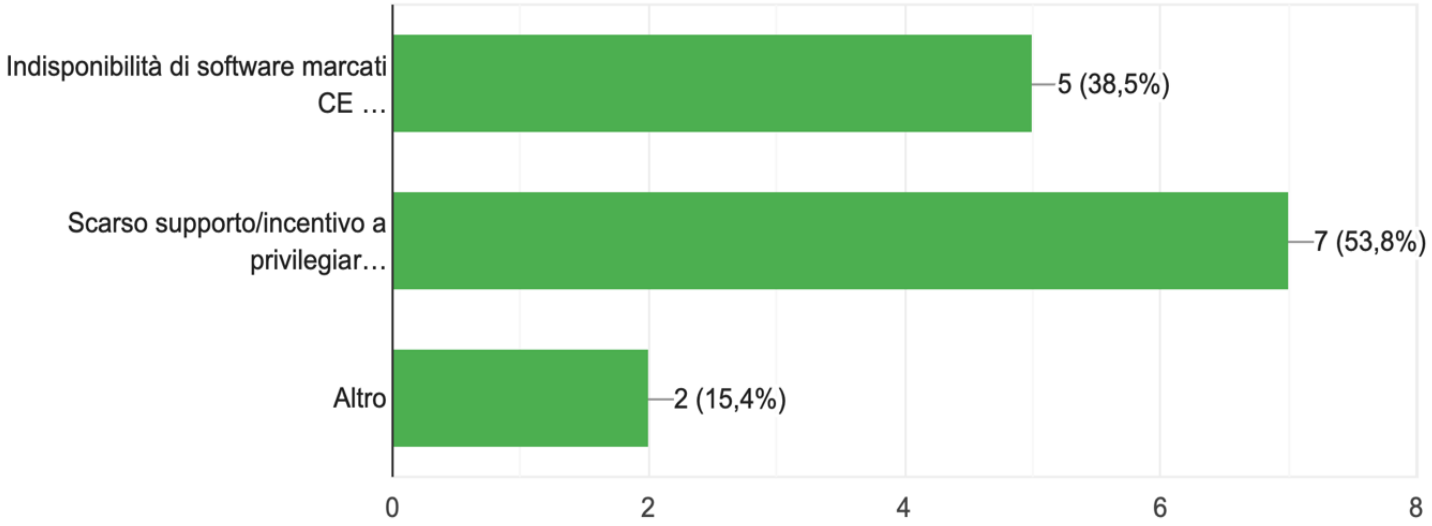
Nella tua struttura sono utilizzati software AI marcati CE e/o FDA in uso clinico?

28 risposte




Se No, indica eventuali motivi del non utilizzo, se ne sei a conoscenza

13 risposte




M. Iori and C. Talamonti on behalf of AIFM AI group

# Consultazione della Commissione europea sulla trasformazione della salute e dell'assistenza nel mercato unico digitale (2018)

*Quali sono i principali **ostacoli all'accesso elettronico** ai dati sanitari?*

- 61,5%: rischi di violazioni della privacy
- 59,4%: mancanza di infrastrutture
- 50,6%: restrizioni legali negli Stati membri

*Quali sono i principali **ostacoli alla condivisione elettronica** dei dati sanitari?*

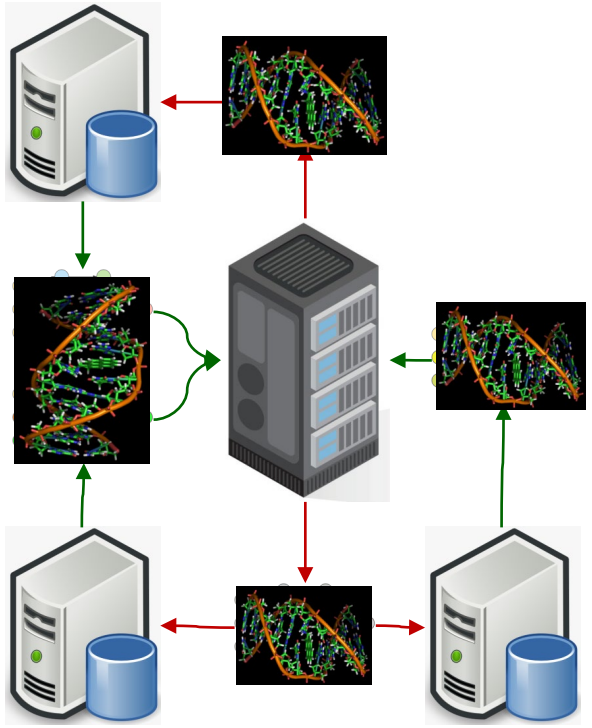
- 70%: eterogeneità delle cartelle cliniche elettroniche
- 62%: mancanza di interoperabilità tecnica
- 52%: rischi di violazione della privacy
- 51%: mancanza di infrastrutture

- 460 responders professionisti o per conto di un'organizzazione

# Infrastrutture - Migliorare la cura grazie ad AI + Big Data

Poter interrogare la rete nazionale (o europea) per un dato genomico particolare permetterebbe di ottenere informazioni sulla cura più efficace per quello specifico paziente

( ad es. Medicina Personalizzata e di Precisione)



# Strutture Federate

Contesto:

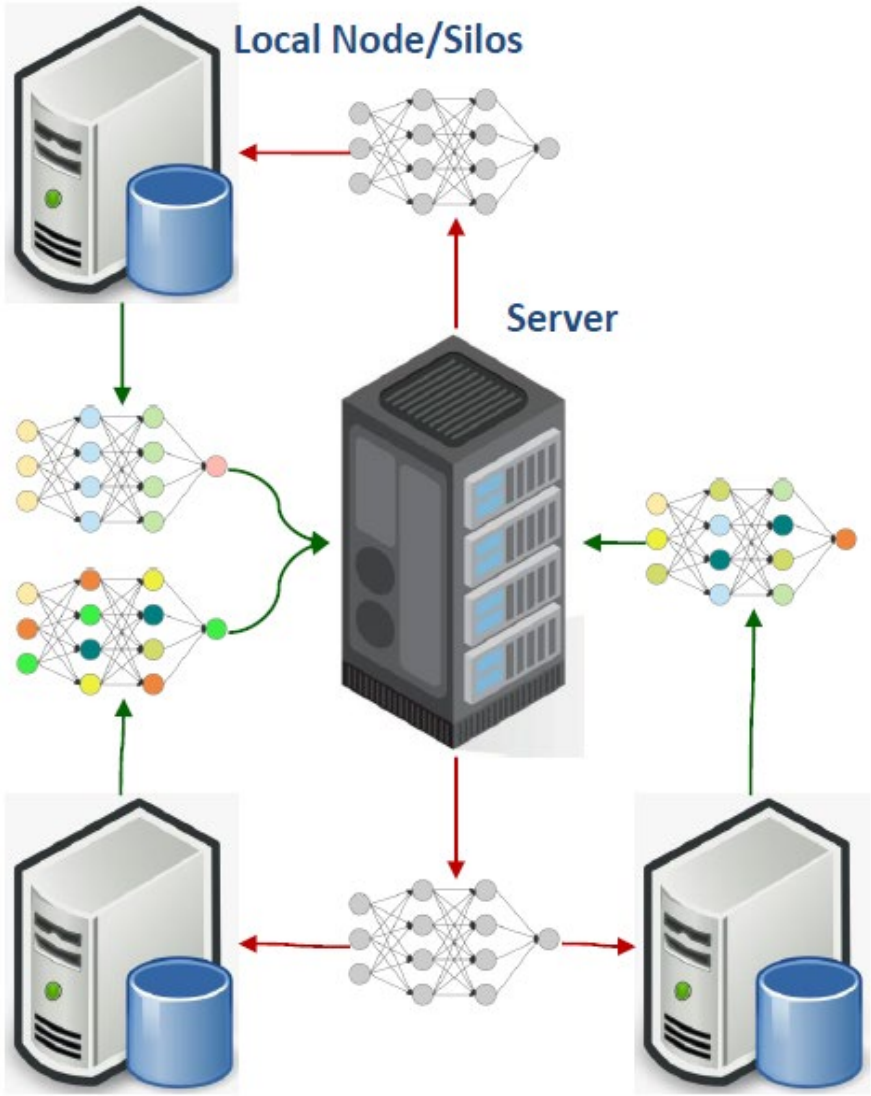
data center isolati distribuiti, requisiti di privacy e sicurezza, barriere normative (ad es. Limitazione dello scambio di informazioni sul genoma)

Come sfruttiamo questi dati?

1. I nodi ricevono il modello da un server
2. I nodi addestrano il modello o lo sfruttano
3. I nodi inviano un modello addestrato al server
4. Il server combina i modelli di formazione locale ricevuti e produce un (nuovo) modello «federato»
5. Il server invia il modello federato ai nodi,... Riavvia da 1.

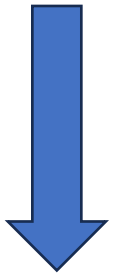
Nessun dato sensibile (ad es. Identificativo personale) lascia i nodi locali;

Utilizzo di più centri (ridurre del Bias, migliorare l'addestramento)



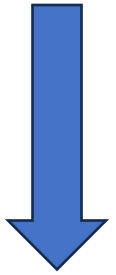
# Problematiche

- Uso clinico



- Responsabilità

- Gestione Dati



- Tutela personale ed eticità

# Leggi, norme, linee guida

## **Strategia nazionale per l'Intelligenza Artificiale** (MISE) 2020

La Strategia Nazionale ha l'obiettivo di delineare un piano coerente per consentire all'Italia di massimizzare i benefici e minimizzare i costi derivanti dal **più importante cambio di paradigma tecnologico dei nostri tempi**.

Essa si inserisce nel quadro del **Piano Coordinato Europeo per l'Intelligenza Artificiale**, e va collocata nell'ambito di **un'azione sinergica tra Stati membri dell'Unione europea e istituzioni comunitarie**.

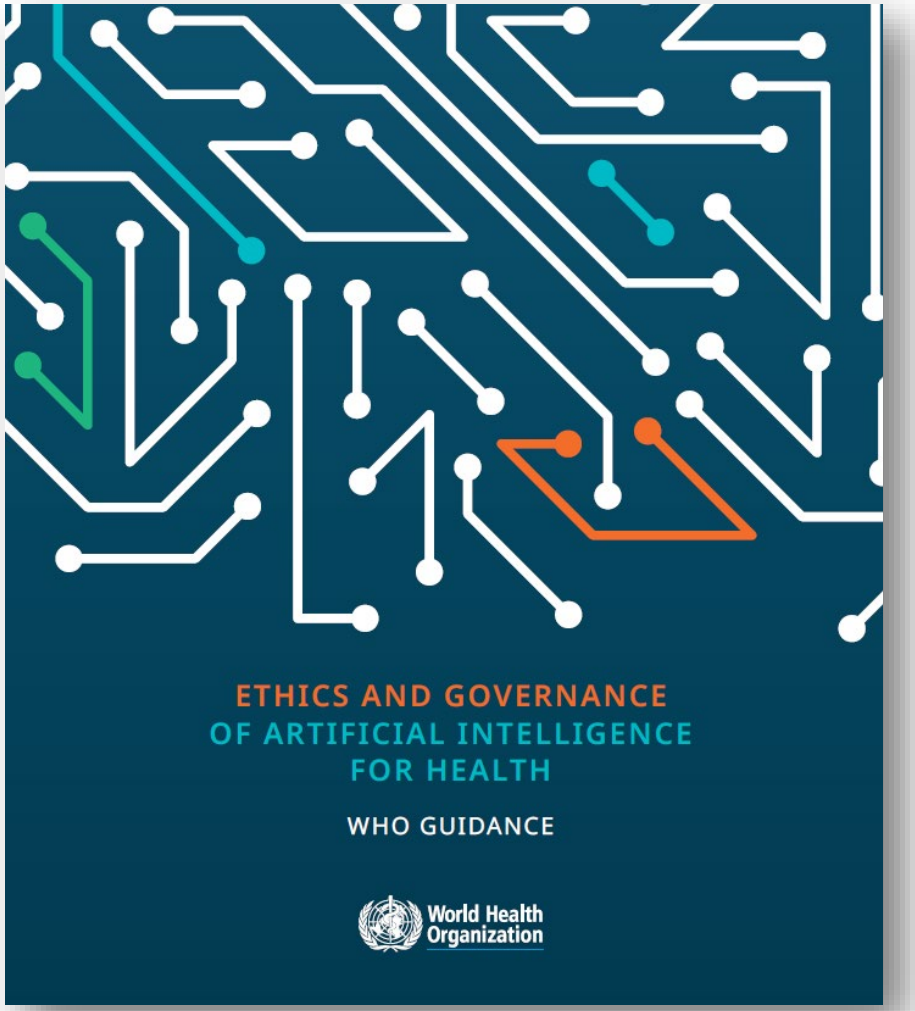
I settori prioritari considerati sono: Industria e manifattura, Agroalimentare, Cultura e turismo, Salute e benessere, Ambiente Infrastruttura e Reti, Città intelligenti, Pubblica Amministrazione

## **L'intelligenza Artificiale al servizio del cittadino**: sfide e opportunità (AgID) 2018

Il libro bianco sull'IA contiene proposte e raccomandazioni per utilizzare al meglio l'Intelligenza Artificiale nel rispetto dei diritti dei cittadini. Esso illustra le linee guida e le raccomandazioni per l'utilizzo sostenibile e responsabile dell'Intelligenza Artificiale nella Pubblica amministrazione e rappresenta l'inizio di un percorso progettuale. Si tratta del primo documento indirizzato alle **amministrazioni pubbliche - scuole, strutture sanitarie, Comuni, Tribunali, Ministeri** - che contiene raccomandazioni e indicazioni su come sfruttare al meglio le opportunità offerte dall'Intelligenza Artificiale, limitandone criticità e aspetti problematici, per sviluppare **servizi pubblici sempre più a misura di cittadino.**

**Versione aggiornata aprile 2025**

# OMS





**2. Artificial intelligence** ..... 4

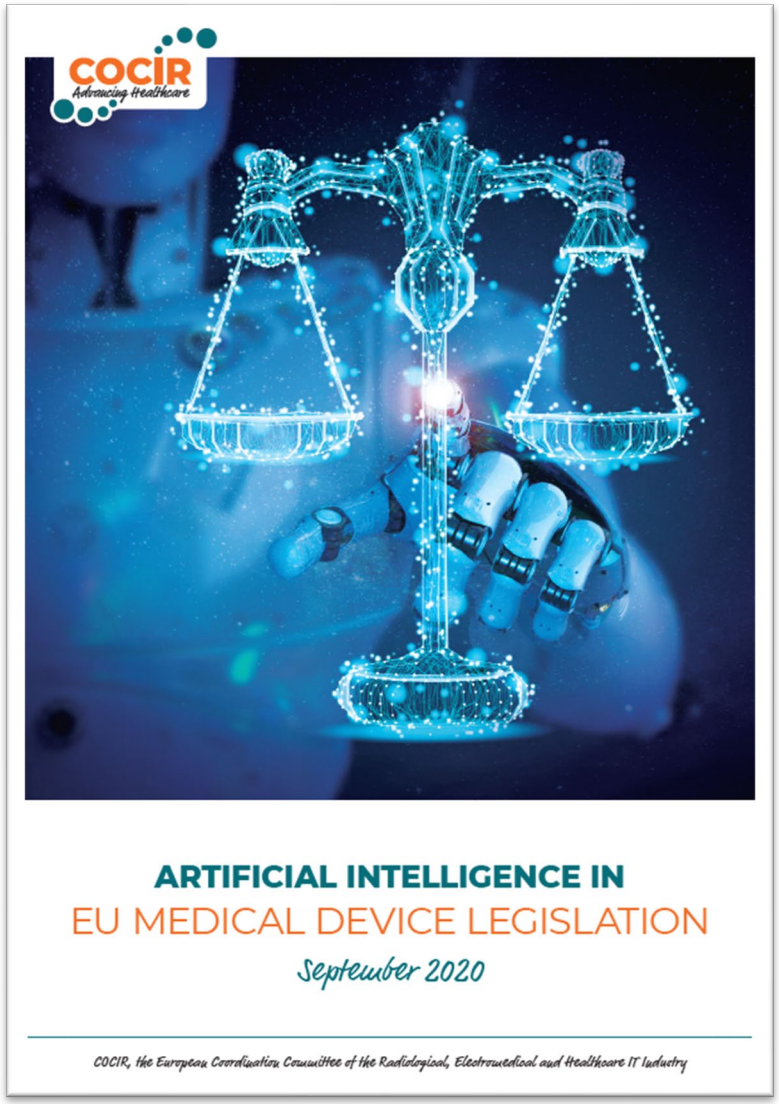
**3. Applications of artificial intelligence for health** ..... 6
3.1 In health care ..... 6
3.2 In health research and drug development ..... 11
3.3 In health systems management and planning ..... 12
3.4 In public health and public health surveillance ..... 12
3.5 The future of artificial intelligence for health ..... 15

**4. Laws, policies and principles that apply to use of artificial intelligence for health** ..... 17
4.1 Artificial intelligence and human rights ..... 17
4.2 Data protection laws and policies ..... 19
4.3 Existing laws and policies related to health data ..... 19
4.4 General principles for the development and use of artificial intelligence ..... 20
4.5 Principles for use of artificial intelligence for health ..... 21
4.6 Bioethics laws and policies ..... 21
4.7 Regulatory considerations ..... 22

**5. Key ethical principles for use of artificial intelligence for health** ..... 23
5.1 Protect autonomy ..... 25
5.2 Promote human well-being, human safety and the public interest ..... 26
5.3 Ensure transparency, explainability and intelligibility ..... 26
5.4 Foster responsibility and accountability ..... 28
5.5 Ensure inclusiveness and equity ..... 29
5.6 Promote artificial intelligence that is responsive and sustainable ..... 30

**6. Ethical challenges to use of artificial intelligence for health care** ..... 31
6.1 Assessing whether artificial intelligence should be used ..... 31
6.2 Artificial intelligence and the digital divide ..... 34
6.3 Data collection and use ..... 35
6.4 Accountability and responsibility for decision-making with artificial intelligence ..... 41
6.5 Autonomous decision-making ..... 45
6.6 Bias and discrimination associated with artificial intelligence ..... 54
6.7 Risks of artificial intelligence technologies to safety and cybersecurity ..... 57

# Cocir




COCIR è il comitato di coordinamento europeo dell'industria informatica radiologica, elettromedicale e sanitaria


3. DEFINITION OF ARTIFICIAL INTELLIGENCE 5

4. ATTRIBUTES OF ARTIFICIAL INTELLIGENCE 6
4.1. CONTROLLABILITY – HUMAN OVERSIGHT 7
4.2. CHANGE THROUGH MACHINE LEARNING 8
4.2.1. PRE-DETERMINED CHANGE 8
4.2.2. CHANGE DYNAMICS - HUMAN CONTROL OF CHANGE 8
4.2.3. CHANGE BOUNDARIES AND ALGORITHM CHANGE PROTOCOL 10
4.2.4. CHANGE ACCORDING TO EU MDR AND EU IVDR 10

5. ETHICS 12

6. TRUST 15
6.1. TRANSPARENCY - EXPLICABILITY 15
6.1.1. TRANSPARENCY ON THE USE OF AUTOMATED DECISION MAKING 16
6.1.2. TRANSPARENT DECISION-MAKING 16
6.1.3. TRANSPARENT ALGORITHM CHANGE 17
6.2. CONFIDENTIALITY & DATA PROTECTION 18
6.3. ELEMENTS OF TRUST REQUIRED BY EU MDR AND EU IVDR 19
6.3.1. PERFORMANCE 19
6.3.2. MINIMIZE BIAS 20
6.3.3. REPEATABILITY 21
6.3.4. RELIABILITY 22
6.3.5. SAFETY 23
6.3.6. ELIMINATE OR REDUCE RISK OF USE ERROR 26
6.3.7. SECURITY 27

7. COMPLIANCE AND ENFORCEMENT UNDER EU MDR AND EU IVDR 28
7.1. RISK-BASED 28
7.2. OVERSIGHT DURING ENTIRE DEVICE LIFETIME 30
7.3. GOVERNANCE 31
7.4. NOTIFIED BODIES 32
7.5. TESTING 32
7.6. ADDRESSEES 34
7.6.1. MANUFACTURERS OF AI-BASED MEDICAL DEVICES AND THEIR ACCESSORIES 34
7.6.2. HEALTH INSTITUTIONS MANUFACTURING AI-BASED MEDICAL DEVICES FOR IN-HOUSE USE 34
7.6.3. MANUFACTURERS OF AI-BASED MEDICAL DEVICES SOLD AS A SERVICE 34
7.6.4. MANUFACTURERS OF CUSTOM-MADE AI-BASED MEDICAL DEVICES 34
7.6.5. MANUFACTURERS OF AI-BASED PARTS OR COMPONENTS FOR A MEDICAL DEVICE 34
7.6.6. ECONOMIC OPERATORS 36
7.7. KEEPING OF RECORDS AND DATA 36

8. INTERNATIONAL COMPARISON 37

9. STANDARDS 38

10. VOLUNTARY LABELLING SYSTEM FOR HEALTH APPS 40

11. SUMMARY OF RECOMMENDATIONS 42

European Commission

EN English

Shaping Europe's digital future

Home Policies Activities News Library Funding Calendar Consultations

Home > Library > Ethics guidelines for trustworthy AI

REPORT / STUDY | Publication 08 April 2019

# Ethics guidelines for trustworthy AI

On 8 April 2019, the High-Level Expert Group on AI presented Ethics Guidelines for Trustworthy Artificial Intelligence. This followed the publication of the guidelines' first draft in December 2018 on which more than 500 comments were received through an open consultation.

https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

# ORIENTAMENTI ETICI PER UN'IA AFFIDABILE

- L'obiettivo degli orientamenti è promuovere un'IA affidabile.
- Un'IA affidabile si basa su **tre componenti** che dovrebbero essere presenti durante l'intero ciclo di vita del sistema: a) **legalità**, l'IA deve ottemperare a tutte le leggi e ai regolamenti applicabili, b) **eticità**, l'IA deve assicurare l'adesione a principi e valori etici, e c) **robustezza**, dal punto di vista tecnico e sociale poiché, anche con le migliori intenzioni, i sistemi di IA possono causare danni non intenzionali.
- Ciascuna componente in sé è necessaria ma non sufficiente per realizzare un'IA affidabile. Idealmente le tre componenti operano armonicamente e si sovrappongono; qualora, nella pratica, si dovessero verificare tensioni tra di esse la società dovrebbe adoperarsi per risolverle.

# ORIENTAMENTI ETICI PER UN'IA AFFIDABILE

Attingendo al capitolo I, il capitolo II fornisce indicazioni su come realizzare un'IA affidabile elencando **sette requisiti che i sistemi di IA dovrebbero soddisfare e per la cui attuazione possono essere utilizzati metodi tecnici e non tecnici**.

- Garantire che lo **sviluppo**, la **distribuzione** e **l'utilizzo** dei sistemi di IA soddisfino i requisiti di un'IA affidabile: **1) intervento e sorveglianza umani, 2) robustezza tecnica e sicurezza, 3) riservatezza e governance dei dati, 4) trasparenza, 5) diversità, non discriminazione ed equità, 6) benessere sociale e ambientale e 7) accountability (**responsabilità**).**
- Prendere in considerazione **metodi tecnici e non tecnici** per garantire l'attuazione di tali requisiti.
- **Favorire la ricerca e l'innovazione** a sostegno della valutazione dei sistemi di IA e del raggiungimento del rispetto dei requisiti; **diffondere i risultati** e le domande aperte al grande pubblico e formare sistematicamente una nuova generazione di esperti in etica dell'IA.
- **Informare** in modo chiaro e proattivo i portatori di interessi in merito alle **capacità e ai limiti** del sistema di IA, creando così aspettative realistiche, e in merito ai modi in cui i requisiti sono attuati. Essere **trasparenti** circa il fatto che si sta interagendo con un sistema di IA.
- **Agevolare la tracciabilità e la verificabilità** dei sistemi di IA, in particolare in contesti o situazioni critiche.
- Coinvolgere i portatori di interessi durante l'intero ciclo di vita del sistema di IA. Promuovere la **formazione e l'istruzione** affinché tutti i portatori di interessi siano formati e informati in merito all'IA affidabile.
- Essere consapevoli che vi potrebbero essere **tensioni fondamentali** tra i diversi **principi** e i diversi **requisiti**. **Individuare, valutare, documentare e comunicare costantemente le soluzioni di compromesso**.

# ORIENTAMENTI ETICI PER UN'IA AFFIDABILE

**Principi etici nel contesto dei sistemi di IA**

I sistemi di IA devono migliorare il benessere individuale e collettivo. Questa sezione elenca **quattro principi etici**, radicati nei diritti fondamentali ai quali occorre aderire per garantire che i sistemi di IA siano sviluppati, distribuiti e utilizzati in modo affidabile. Sono definiti come **imperativi etici** affinché gli operatori del settore dell'IA si adoperino sempre per aderirvi. Senza imporre alcuna gerarchia, i principi sono elencati di seguito secondo l'ordine di apparizione dei diritti fondamentali su cui si basano all'interno della Carta dei diritti fondamentali dell'Unione europea.

I principi sono i seguenti:

**i) rispetto dell'autonomia umana**

**ii) prevenzione dei danni**

**iii) equità**

**iv) esplicabilità**

Anche se molti obblighi giuridici riflettono principi etici, l'adesione ai principi etici va oltre il rispetto formale del diritto vigente.

# AI act - Explanatory memorandum

- La proposta si basa sui **valori e sui diritti fondamentali** dell'UE e mira a dare alle persone e agli altri utenti la fiducia necessaria per adottare soluzioni basate sull'IA, incoraggiando nel contempo le imprese a svilupparle.

- L'IA dovrebbe essere uno strumento per le persone ed essere una forza positiva nella società con l'obiettivo finale di aumentare il benessere umano.

- Le norme per l'IA disponibili sul mercato dell'Unione o che incidono in altro modo sulle persone nell'Unione dovrebbero pertanto essere antropocentriche, in modo che le persone possano avere fiducia nel fatto che la tecnologia sia utilizzata in modo sicuro e conforme alla legge, compreso il rispetto dei diritti fondamentali.

# AI act - Explanatory memorandum

La proposta risponde inoltre a richieste esplicite del Parlamento Europeo (PE) e del Consiglio Europeo, che hanno ripetutamente espresso la richiesta di un'azione legislativa per garantire il buon funzionamento del mercato interno dei sistemi di intelligenza artificiale ("sistemi di IA"), in cui sia i **benefici** che i **rischi** dell'IA siano adeguatamente affrontati a livello dell'Unione.

Sostiene l'obiettivo dell'Unione di essere un **leader mondiale nello sviluppo di un'intelligenza artificiale sicura, affidabile ed etica**, come affermato dal Consiglio Europeo, e **garantisce la protezione dei principi etici**, come specificamente richiesto dal Parlamento europeo.

# Regolamento IA - Explanatory memorandum & AIMs

- Le conclusioni più recenti del 21 ottobre 2020 invitavano inoltre ad affrontare **l'opacità, la complessità, le distorsioni, un certo grado di imprevedibilità e il comportamento parzialmente autonomo** di alcuni sistemi di IA, a garantirne la compatibilità con i diritti fondamentali e ad agevolare l'applicazione delle norme giuridiche.

*Obiettivi specifici*

- assicurare che i sistemi di IA immessi sul mercato dell'Unione e utilizzati siano **sicuri** e rispettino la normativa vigente in materia di diritti fondamentali e i valori dell'Unione;
- assicurare la **certezza del diritto** per facilitare gli investimenti e l'innovazione nell'intelligenza artificiale;
- migliorare la **governance** e l'applicazione effettiva della normativa esistente in materia di diritti fondamentali e requisiti di sicurezza applicabili ai sistemi di IA;
- facilitare lo **sviluppo di un mercato unico** per applicazioni di IA lecite, sicure e affidabili nonché prevenire la frammentazione del mercato.

# Regolamento IA

(Dalla Relazione associata alla proposta di Regolamento)

**1.3. Coerenza con le altre normative dell'Unione**

La proposta fa parte di un pacchetto più ampio di misure destinate ad affrontare i problemi posti dallo sviluppo e dall'utilizzo dell'IA, come esaminato nel Libro bianco sull'intelligenza artificiale. Sono pertanto garantite la **coerenza** e la **complementarietà** rispetto ad altre **iniziative** in corso o previste della Commissione, volte anch'esse ad affrontare tali problemi, comprese **la revisione della normativa settoriale sui prodotti** (ad esempio la **direttiva macchine**, la direttiva sulla **sicurezza generale dei prodotti**) e le iniziative che affrontano le questioni connesse alla **responsabilità** in relazione alle nuove tecnologie, compresi i sistemi di IA. Tali iniziative si baseranno sulla presente proposta e la integreranno al fine di apportare chiarezza giuridica e favorire lo sviluppo di un ecosistema di fiducia nei confronti dell'IA in Europa.

# Regolamento IA (prima versione)

Art. 3 - Definizioni

"sistema di intelligenza artificiale" (sistema di IA): un software sviluppato con una o più delle tecniche e degli approcci elencati nell'allegato I, che può, per una determinata serie di obiettivi definiti dall'uomo, generare output quali contenuti, previsioni, raccomandazioni o decisioni che influenzano gli ambienti con cui interagiscono;

Allegato I - TECNICHE E APPROCCI DI INTELLIGENZA ARTIFICIALE

a) Approcci di apprendimento automatico, compresi l'apprendimento supervisionato, l'apprendimento non supervisionato e l'apprendimento per rinforzo, con utilizzo di un'ampia gamma di metodi, tra cui l'apprendimento profondo (*deep learning*);

b) approcci basati sulla logica e approcci basati sulla conoscenza, compresi la rappresentazione della conoscenza, la programmazione induttiva (logica), le basi di conoscenze, i motori inferenziali e deduttivi, il ragionamento (simbolico) e i sistemi esperti;

c) approcci statistici, stima bayesiana, metodi di ricerca e ottimizzazione.

# Regolamento IA (versione 2024)

- Articolo 3
- Definizioni

- Ai fini del presente regolamento si applicano le definizioni seguenti:

- 1) «sistema di IA»: un sistema automatizzato progettato per funzionare con livelli di autonomia variabili e che può presentare adattabilità dopo la diffusione e che, per obiettivi espliciti o impliciti, deduce dall'input che riceve come generare output quali **previsioni, contenuti, raccomandazioni o decisioni** che possono influenzare ambienti fisici o virtuali;

- 2) «rischio»: la combinazione della probabilità del verificarsi di un danno e la gravità del danno stesso;

# Key principles of the AI Act



Horizontal approach



Product safety legislation (AI systems are products) + fundamental rights' layer



Risk-based



Holistic governance and enforcement framework



Trust and Innovation

# Risk-based approach

ISO 14971 Risk Management for Medical Device
Risk Class 2a, 2b, 3



*Not mutually exclusive

**Unacceptable risk**
e.g. social scoring, untargeted scraping
— **Prohibited**

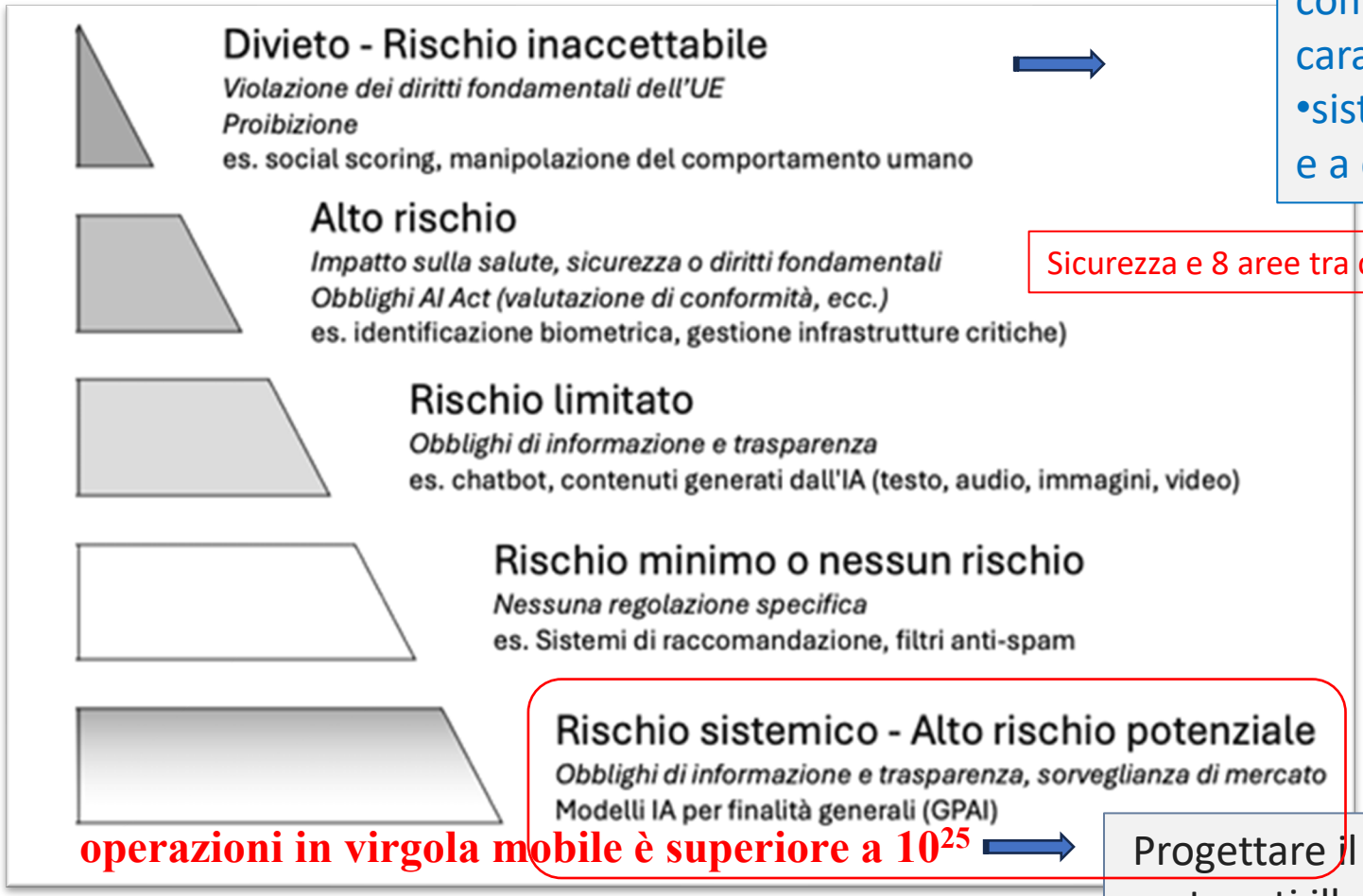
**High risk**
e.g. recruitment, medical devices
— **Permitted** subject to compliance with AI requirements and ex-ante conformity assessment

**'Transparency' risk**
'Impersonation' (chatbots), deep fakes
— **Permitted** but subject to information/transparency obligations

**Minimal or no risk**
— **Permitted** with no restrictions, voluntary codes of conduct possible

European Commission

# AI act e AGID 2025 per la




- **manipolazione comportamentale** cognitiva di persone o gruppi vulnerabili specifici;
- **classificazione sociale** delle persone in base a comportamento, livello socio-economico, caratteristiche personali;
- sistemi di **identificazione biometrica** in tempo reale e a distanza, come il riconoscimento facciale.

Sicurezza e 8 aree tra cui la Salute (DM) → regolamenti per la conformità

«rischio sistemico»: un rischio specifico per le capacità di impatto elevato dei modelli di IA per finalità generali, avente un impatto significativo sul mercato dell'Unione a causa della sua portata o di effetti negativi effettivi o ragionevolmente prevedibili sulla salute pubblica, la sicurezza, i diritti fondamentali o la società nel suo complesso, che può propagarsi su larga scala lungo l'intera catena del valore;

**operazioni in virgola mobile è superiore a $10^{25}$**

Progettare il modello GenAI in modo da impedire la generazione di contenuti illegali e pubblicare riepiloghi dei dati con diritti d'autore utilizzati per l'addestramento. *(fornire info su consumo energetico)*

# Regolamento IA

Art. 3 - Definizioni

"sistema di intelligenza artificiale" (sistema di IA): un software sviluppato con una o più delle tecniche e degli approcci elencati nell'allegato I, che può, per una determinata serie di obiettivi definiti dall'uomo, generare **output** quali contenuti, previsioni, raccomandazioni o decisioni che influenzano gli ambienti con cui interagiscono;

*Articolo 6 - Regole di classificazione per i sistemi di IA ad alto rischio*

1. A prescindere dal fatto che sia immesso sul mercato o messo in servizio in modo indipendente rispetto ai prodotti di cui alle lettere a) e b), un sistema di IA è considerato ad alto rischio se sono soddisfatte entrambe le condizioni seguenti:

a) il sistema di IA è destinato a essere utilizzato come **componente di sicurezza** di un prodotto, o è esso stesso un prodotto, disciplinato dalla normativa di armonizzazione dell'Unione elencata nell'allegato II;

b) il prodotto, il cui componente di sicurezza è il sistema di IA, o il sistema di IA stesso in quanto **prodotto** è soggetto a una valutazione della conformità da parte di terzi ai fini dell'immissione sul mercato o della messa in servizio di tale prodotto ai sensi della normativa di armonizzazione dell'Unione elencata nell'allegato II. Tali prodotti sono, in particolare, macchine, giocattoli, ascensori, [..], dispositivi medici e dispositivi medico-diagnostici in vitro. (considerando 30)

2. Oltre ai sistemi di IA ad alto rischio di cui al paragrafo 1, sono considerati ad alto rischio anche i sistemi di IA di cui all'allegato III.

Non si applica ai DM

# High-risk Artificial Intelligence Systems (Title III, Chapter 1 & Annexes I and III)




## 1 AI SAFETY COMPONENTS OF REGULATED PRODUCTS (ANNEX I)

Only if the products are subject to third-party conformity assessment under the relevant sectorial Product Safety Legislation

**Section A** – Product Safety Legislation based on the New Legislative Framework (e.g. machinery, medical device, toys)

- AI Act applies directly in conjunction with the sectorial Union law

**Section B** – Products Safety Legislation based on the Old Approach (e.g. aviation, automotive)

- AI Act does not apply directly; needs to be transposed through amendments in the sectoral legislation

## 2 CERTAIN (STAND-ALONE) AI SYSTEMS (ANNEX III)

European Commission

# High-risk AI systems will have to comply with certain rules

## 2 High-risk (stand-alone) use cases listed in Annex III:

- **Biometrics:** e.g. **r**emote biometric identification, emotion recognition;
- **Critical infrastructures**: e.g. safety components of digital infrastructure, road traffic
- **Education**: e.g. AI to evaluate learning outcomes, assign students in educational institutions
- **Employment:** e.g. AI to analyse job applications or evaluate candidates, promote or fire workers
- **Essential private and public services**: AI determining eligibility to essential public benefits and services; credit-scoring and creditworthiness assessment, risk assessment and pricing in health and life insurance
- **Law enforcement:** e.g. AI to predict risk of offenders/victims
- **Border management:** e.g. AI evaluating eligibility
- **Administration of justice and democratic processes**

**Filter mechanism**: Excludes systems from the high-risk list that:

- perform narrow procedural tasks,
- improve the result of previous human activities,
- do not influence human decisions or
- do purely preparatory tasks,

NB. Profiling of natural persons always high-risk

**NB!** Only the use cases explicitly listed in Annex III are high-risk;
The Commission can add more through delegated acts to keep the list future proof.

# Obligations for high-risk AI (Annex I and III)




**Provider's obligations**

- **Risk management system** to minimise risks to health, safety and fundamental rights
- **Requirements for Trustworthy AI systems:**
  - Data quality and governance
  - Documentation and traceability (logging capabilities)
  - Transparency and information to deployers,
  - Human oversight
  - Accuracy, cybersecurity and robustness
- **Conformity assessment** to demonstrate compliance prior to placing on the market
  - Self-assessment except for Annex I products and remote biometric systems subject to third party conformity assessment
- **Quality management system**
- **CE marking** and declaration of conformity
- **Register** standalone AI system (Annex III) in EU database
- Conduct **post-market monitoring** and report **serious incidents**
- Non-EU providers to appoint **authorized representative in the EU**

# Obligations for high-risk AI systems (Annex I and III)




## Deployer's obligations

- Operate high-risk AI system in accordance with **instructions of use**
- Ensure **human oversight**: persons assigned must have the necessary competence, training and authority
- **Monitor** for possible risks and **report problems and any serious incident** to the provider or distributor or national market surveillance authorities
- Public authorities to **register the use in the EU database**
- **Inform affected workers** and their representatives
- **Inform people** subjected to decisions taken or informed by high-risk AI systems in Annex III and, upon request, provide them with **an explanation**
- **Fundamental rights impact assessment** (only Annex III systems except critical infrastructure) :
  - If deployers are public entities, private entities providing public services and deployers in the area of insurance and credit scoring

# Support measures for innovation

## AI Regulatory Sandboxes

Definition:

- a controlled framework ***set up by a competent authority*** where providers can develop, train, validate, and test their AI systems (incl. in real-world conditions) pursuant to a sandbox plan for a limited time under regulatory supervision

Objective:

- foster AI innovation and competitiveness, improve legal certainty for innovators and facilitate regulatory compliance with the AI Act

**SMEs have priority access**

Obligations, each Member State to:

- establish (at least) one sandbox or
- establish a joint sandbox with another MS or
- participate in an already existing one

EU Governance:

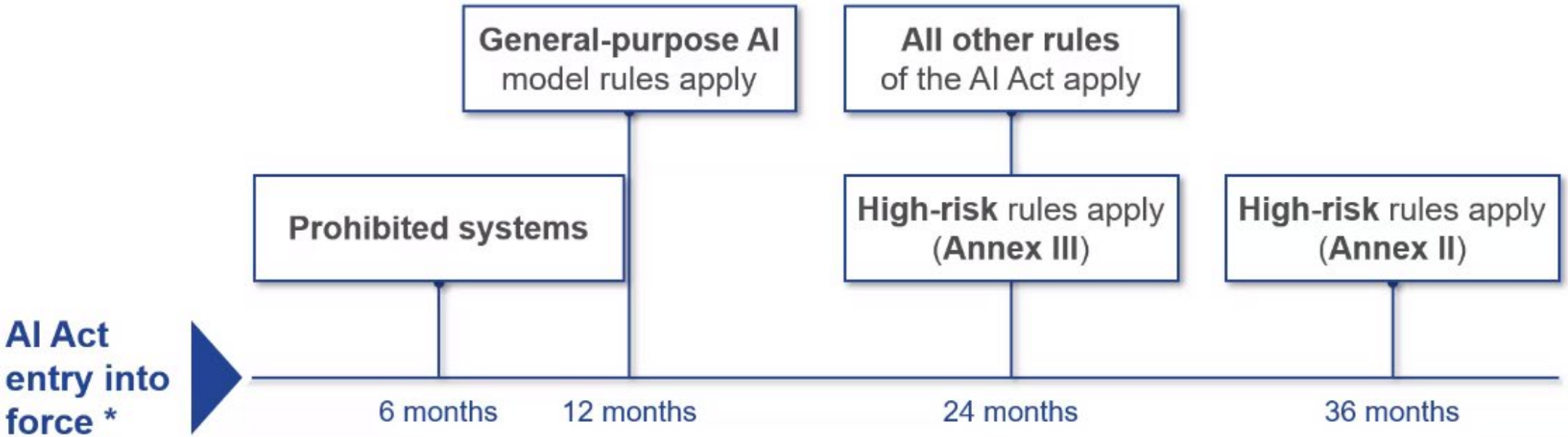
✓ EU to provide common implementing rules and supporting tools
✓ Cooperation through the AI Board
✓ Links with the ecosystems (e.g. TEFs, EDIHs, dataspaces)

## Testing in real-world conditions (outside of sandboxes)

For Annex III systems, under certain safeguards:

- explicit consent of subjects
- real-world testing plan submitted to the market surveillance authority
- register in a dedicated section in the EU-wide database
- set limitations on the period
- additional safeguards for persons belonging to certain vulnerable groups
- a written agreement defining the roles and responsibilities of prospective providers and deployers
- effective oversight by competent personnel
- predictions, recommendations or decisions must be effectively reversed and disregarded
- data protection safeguards
- incident reporting

# The AI Act enters into application in a gradual approach




*Following its adoption by the European Parliament and the Council, the AI Act shall enter into force on the twentieth day following that of its publication in the official Journal.

DIGITAL COMMISSION ESSENTIALS

European Commission

# Articolo 113 Entrata in vigore e applicazione

Il presente regolamento entra in vigore il ventesimo giorno successivo alla pubblicazione nella Gazzetta ufficiale dell'Unione europea.

Si applica a decorrere dal 2 agosto 2026.

Tuttavia:

a) I capi I e II si applicano a decorrere dal 2 febbraio 2025;

b) Il capo III, sezione 4, il capo V, il capo VII, il capo XII e l'articolo 78 si applicano a decorrere dal 2 agosto 2025, ad eccezione dell'articolo 101;

c) L'articolo 6, paragrafo 1, e i corrispondenti obblighi di cui al presente regolamento si applicano a decorrere dal 2 agosto 2027.

Il presente regolamento è obbligatorio in tutti i suoi elementi e direttamente applicabile in ciascuno degli Stati membri.
Fatto a Bruxelles, il 13 giugno 2024

# MDSW basato su tecnologie AI

- Un dato software, e in particolare un MDSW, può essere progettato e realizzato facendo uso di tecnologie AI

- Questa classe di MDSW innovativi, oltre ai requisiti MDR (opp. IVDR), deve essere conforme anche ad altri dispositivi regolatori: *AI Act*

- (Art. 113) Il regolamento AI entra in vigore il ventesimo giorno successivo alla pubblicazione nella Gazzetta ufficiale dell'Unione europea. Si applica a decorrere dal 2 agosto 2026.



Gazzetta ufficiale dell'Unione europea

IT

Serie L

2024/1689

12.7.2024

**REGOLAMENTO (UE) 2024/1689 DEL PARLAMENTO EUROPEO E DEL CONSIGLIO**

**del 13 giugno 2024**

**che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n, 300/2008, (UE) n, 167/2013, (UE) n, 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale)**

# Regolamento IA

**ALLEGATO II ELENCO DELLA NORMATIVA DI ARMONIZZAZIONE DELL'UNIONE Sezione A – Elenco della normativa di armonizzazione dell'Unione in base al nuovo quadro normativo**

1. Direttiva 2006/42/CE del Parlamento europeo e del Consiglio, del 17 maggio 2006, relativa alle macchine e che modifica la direttiva 95/16/CE (GU L 157 del 9.6.2006, pag. 24) [abrogata dal regolamento sui prodotti macchina];

[…]

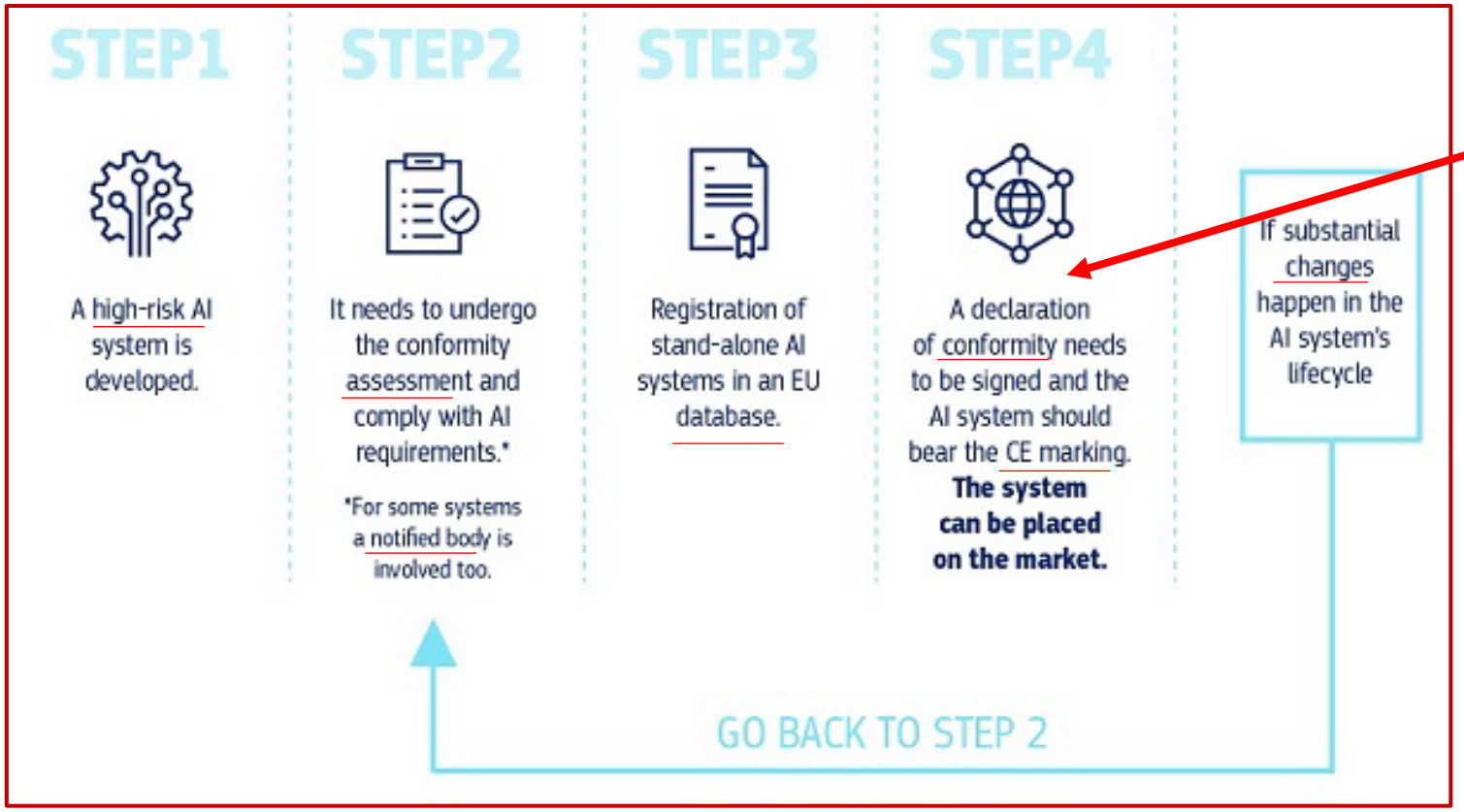
11. regolamento (UE) 2017/745 del Parlamento europeo e del Consiglio, del 5 aprile 2017, relativo ai dispositivi medici, che modifica la direttiva 2001/83/CE, il regolamento (CE) n. 178/2002 e il regolamento (CE) n. 1223/2009 e che abroga le direttive 90/385/CEE e 93/42/CEE del Consiglio (GU L 117 del 5.5.2017, pag. 1);

12. regolamento (UE) 2017/746 del Parlamento europeo e del Consiglio, del 5 aprile 2017, relativo ai dispositivi medico-diagnostici in vitro e che abroga la direttiva 98/79/CE e la decisione 2010/227/UE della Commissione (GU L 117 del 5.5.2017, pag. 176).

# Ciclo di vita dei sistemi AI ad alto rischio

https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

**How does it all work in practice for providers of high-risk AI systems?**




Dichiarazione di conformità secondo il Regolamento AI

# Aspetti regolatori AI Act

(Art. 3) "marcatura CE di conformità" (marcatura CE): una marcatura mediante la quale un fornitore indica che un sistema di IA è conforme ai requisiti stabiliti al titolo III, capo 2, del presente regolamento e in altre normative applicabili dell'Unione che armonizzano le condizioni per la commercializzazione dei prodotti ("normativa di armonizzazione dell'Unione") e che ne prevedono l'apposizione;

- *Requisiti per la conformità*
- *Valutazione della conformità (Art. 19)*
- *Certificazione*

✓ (All. VII - Conformità basata sulla valutazione del sistema di gestione) Se il sistema di IA è conforme ai requisiti di cui al titolo III, capo 2, l'organismo notificato **deve rilasciare un certificato di valutazione UE della documentazione tecnica.** Tale certificato deve indicare il nome e l'indirizzo del fornitore, le conclusioni dell'esame, le eventuali condizioni di validità e i dati necessari per identificare il sistema di IA.

✓ Il certificato e i suoi allegati devono contenere tutte le informazioni pertinenti per consentire la valutazione della conformità del sistema di IA e il controllo del sistema di IA durante l'uso, ove applicabile.

# Aspetti regolatori

*Articolo 19 - Valutazione della conformità*

1. I fornitori dei sistemi di IA ad alto rischio garantiscono che il sistema di IA ad alto rischio sia sottoposto alla pertinente procedura di valutazione della conformità di cui all'articolo 43 prima della sua immissione sul mercato o messa in servizio. Se in seguito a tale valutazione i sistemi di IA risultano conformi ai requisiti di cui al capo 2 del presente titolo, i fornitori redigono una dichiarazione di conformità UE a norma dell'articolo 48 e appongono la marcatura CE di conformità a norma dell'articolo 49.
2. Per i sistemi di IA ad alto rischio di cui all'allegato III, punto 5, lettera b), immessi sul mercato o messi in servizio da fornitori che sono enti creditizi disciplinati dalla direttiva 2013/36/UE, la valutazione della conformità è effettuata nell'ambito della procedura di cui agli articoli da 97 a 101 di tale direttiva.

# MDSW basato su tecnologie AI

CAPO III

**SISTEMI DI IA AD ALTO RISCHIO**

*SEZIONE 1*

***Classificazione dei sistemi di IA come «ad alto rischio»***

*SEZIONE 2*

***Requisiti per i sistemi di IA ad alto rischio***

*SEZIONE 3*

***Obblighi dei fornitori e dei deployer dei sistemi di IA ad alto rischio e di altre parti***

*SEZIONE 4*

***Autorità di notifica e organismi notificati***

*SEZIONE 5*

***Norme, valutazione della conformità, certificati, registrazione***

# Regolamento IA

**REQUISITI PER I SISTEMI DI IA AD ALTO RISCHIO**

- Conformità ai requisiti (Art. 8) *(Sicurezza e prestazione)*
- Sistema di gestione dei rischi (Art. 9)*(ISO 14971)*
- Dati e governance dei dati (Art. 10)
- Documentazione tecnica (Art. 11)
- Conservazione delle registrazioni (Art. 12)
- Trasparenza e fornitura di informazioni agli utenti (Art. 13)
- Sorveglianza umana (Art. 14)
- Accuratezza, robustezza e cybersicurezza (Art. 15)

# Regolamento per l'IA – Overlapping con MDR 2017/745

*"several stakeholders warn the Commission to **avoid duplication**, conflicting obligations and overregulation".*

- si potrebbe pensare che al momento sembra esserci una possibile duplicazione di molte attività richieste sia da **MDR** che da **RIA** (considerando la proposta RIA accettata così com'è).
- Ciò potrebbe essere evitato in futuro tramite una revisione del MDR con incorporazione di requisiti RIA.

(v. lista di overlap:

- Risk management system (similar to MDR and IVDR) (article 9)
- Data governance and data management practices (similar to MDR/IVDR and GDPR) (article 10)
- Technical documentation (similar to MDR/IVDR article 11)
- Logging capabilities (similar to GDPR) (article 12)
- Transparency and information to users (similar to GDPR) (article 13)
- Human oversight requirements (similar to MDR/IVDR) (article 14)
- Accuracy, robustness and cybersecurity (similar to MDR/IVDR and GDPR) (article 15)
- Obligations very much like article 10 MDR/IVDR (device manufacturer obligations plus QMS) (articles 16 and 17)
- Economic operator requirements (similar to MDR/IVDR) (articles 25 to 28)
- MDR and IVDR PMS systems must integrate AIA PMS elements (Article 61 (4))

- l'avvocato Erik Vollebregt impressione: al momento il certificato IA è addizionale rispetto al certificato MDR –
- https://medicaldeviceslegal.com/2021/05/03/the-new-eu-ai-regulation-proposal-medical-devices-and-ivds/

# Overlapping

| Comparison | | |
|---|---|---|
| **Aspect** | **MDR** | **AI Act** |
| **Scope** | Medical devices, in vitro diagnostics, accessories | All AI systems, with sector-specific provisions |
| **Risk Classification** | Classes I, IIa, IIb, III based on health risk | Four tiers: Unacceptable, High, Limited, Minimal risk |
| **Key Risks** | Patient safety, clinical performance | Fundamental rights (bias, discrimination), safety, societal impacts |
| **Conformity Assessment** | Required for higher-risk classes, often via Notified Body | Required for high-risk AI systems; may involve third-party checks |
| **Technical Documentation** | Detailed design, manufacturing, risk analysis, clinical evaluation | Documenting data sets, model training, risk management, system design |
| **Post-Market Surveillance** | Ongoing monitoring, incident reporting, device updates | Continuous performance monitoring, logging, corrective measures for AI systems |
| **Governance** | European Commission, Competent Authorities, EUDAMED | European Commission, national supervisory authorities, European AI Board |
| **Penalties** | Recalls, fines (scale depends on national laws) | Fines up to 6% of global turnover for severe breaches |
| **Sector** | Healthcare | Cross-sector (finance, employment, healthcare, etc.) |
| **Update Cycle** | Iterative, but typically slower than AI | Emphasises iterative model updates and re-checks |

# MDSW basato su tecnologie AI

## Articolo 6 - Regole di classificazione per i sistemi di IA ad alto rischio

1. A prescindere dal fatto che sia immesso sul mercato o messo in servizio indipendentemente dai prodotti di cui alle lettere a) e b), un sistema di IA è considerato ad alto rischio se sono soddisfatte entrambe le condizioni seguenti:

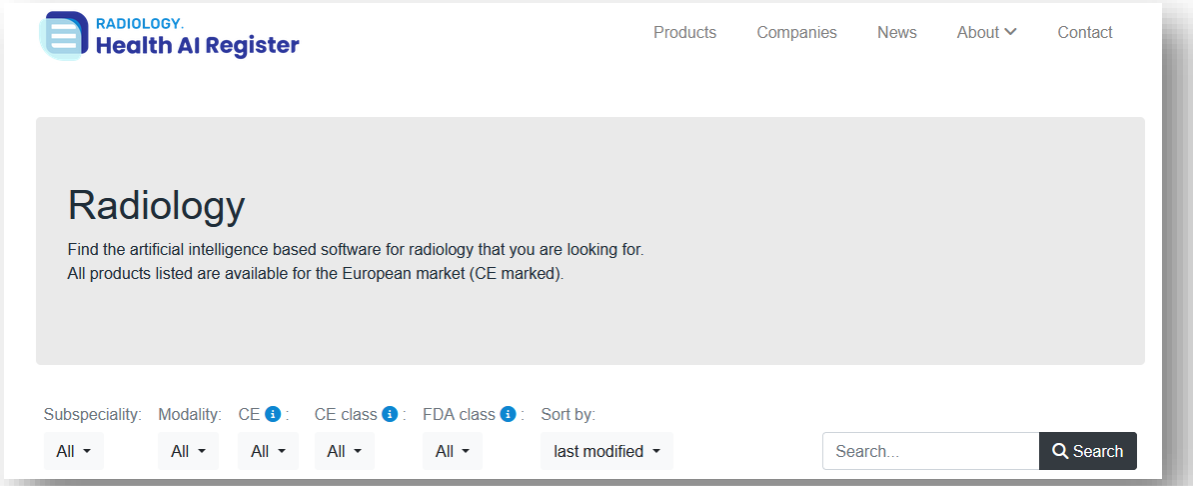
a) il sistema di IA è destinato a essere utilizzato come componente di sicurezza di un prodotto, o il sistema di IA è esso stesso un prodotto, disciplinato dalla normativa di armonizzazione dell'Unione elencata nell'allegato I;

b) il prodotto, il cui componente di sicurezza a norma della lettera a) è il sistema di IA, o il sistema di IA stesso in quanto prodotto, è soggetto a una valutazione della conformità da parte di terzi ai fini dell'immissione sul mercato o della messa in servizio di tale prodotto ai sensi della normativa di armonizzazione dell'Unione elencata nell'allegato I.

# MDSW basato su tecnologie AI

Nel database www.HealthAIregister.com, sono elencati a oggi 82 prodotti dichiarati conformi a MDR.
Di questi, solo uno è in classe I -> classe di rischio minore, non richiede la valutazione di conformità da parte di un ente terzo (Organismo notificato)

| **Classe di rischio secondo MDR** | **# prodotti** |
|---|---|
| I | 1/214 |
| IIa | 51/214 |
| IIb | 30/214 |
| III | 0/214 |




van Leeuwen, K.G. et al. Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. Eur Radiol 31, 3797–3804 (2021). https://doi.org/10.1007/s00330-021-07892-z

Secondo l'art. 6 del regolamento IA, **81 software su 82** sono classificabili come «sistemi di IA ad alto rischio»!

# Requisiti per AI - MDSW

## Capo 3 - SEZIONE 2
## Requisiti per i sistemi di IA ad alto rischio

*Articolo 8*

**Conformità ai requisiti**

1. I sistemi di IA ad alto rischio rispettano i requisiti stabiliti nella presente sezione, tenendo conto delle loro previste finalità nonché dello stato dell'arte generalmente riconosciuto in materia di IA e di tecnologie correlate all'IA. Nel garantire conformità a tali requisiti si tiene conto del sistema di gestione dei rischi di cui all'articolo 9.

2. Se un prodotto contiene un sistema di IA cui si applicano i requisiti del presente regolamento e i requisiti della normativa di armonizzazione dell'Unione elencata nell'allegato I, sezione A, i fornitori sono responsabili di garantire che il loro prodotto sia pienamente conforme a **tutti i requisiti applicabili previsti** dalla normativa di armonizzazione dell'Unione applicabile. ***(MDR per noi)***

Nel garantire la conformità dei sistemi di IA ad alto rischio di cui al paragrafo 1 ai requisiti di cui alla presente sezione e al fine di garantire la coerenza, evitare duplicazioni e ridurre al minimo gli oneri aggiuntivi, i fornitori possono scegliere di integrare, se del caso, i necessari **processi di prova e di comunicazione** nonché le **informazioni e la documentazione** che forniscono relativamente al loro prodotto nella documentazione e nelle procedure esistenti e richieste in conformità della normativa di armonizzazione dell'Unione elencata nell'allegato I, sezione A.

# Requisiti sui dati

*Articolo 10* - **Dati e governance dei dati**

1. I sistemi di IA ad alto rischio che utilizzano tecniche che prevedono l'uso di **dati per l'addestramento di modelli** di IA sono sviluppati sulla base di set di dati di addestramento, convalida e prova che soddisfano i criteri di qualità di cui ai paragrafi da 2 a 5 ogniqualvolta siano utilizzati tali set di dati.

2. I set di dati di addestramento, convalida e prova sono soggetti a **pratiche** di **governance** e **gestione** dei **dati** adeguate alla finalità prevista del sistema di IA ad alto rischio. Tali pratiche riguardano in particolare:

a) le scelte progettuali pertinenti;
b) i processi di raccolta dei dati e l'origine dei dati, nonché la finalità originaria della raccolta nel caso di dati personali;
c) le operazioni di trattamento pertinenti ai fini della preparazione dei dati, quali annotazione, etichettatura, pulizia, aggiornamento, arricchimento e aggregazione;
d) la formulazione di ipotesi, in particolare per quanto riguarda le informazioni che si presume che i dati misurino e rappresentino;
e) una valutazione della disponibilità, della quantità e dell'adeguatezza dei set di dati necessari;
f) un esame atto a valutare le possibili **distorsioni** suscettibili di incidere sulla salute e sulla sicurezza delle persone, di avere un impatto negativo sui diritti fondamentali o di comportare discriminazioni vietate dal diritto dell'Unione, specie laddove gli output di dati influenzano gli input per operazioni future;
g) le misure adeguate per individuare, prevenire e attenuare le possibili distorsioni individuate conformemente alla lettera f);
h) l'individuazione di lacune o carenze pertinenti nei dati tali da pregiudicare il rispetto del presente regolamento e il modo in cui tali lacune e carenze possono essere colmate.

# Requisiti sui dati

*Articolo 10* - **Dati e governance dei dati**

3. I set di dati di addestramento, convalida e prova sono pertinenti, sufficientemente rappresentativi e, nella misura del possibile, esenti da errori e completi nell'ottica della finalità prevista. Essi possiedono le **proprietà statistiche** appropriate anche, ove applicabile, per quanto riguarda le persone o i gruppi di persone relativamente ai quali il sistema di IA ad alto rischio è destinato a essere usato. Queste caratteristiche dei set di dati possono essere soddisfatte a livello di singoli set di dati o a livello di una combinazione degli stessi.

4. I set di dati tengono conto, nella misura necessaria per la finalità prevista, delle caratteristiche o degli elementi particolari dello specifico ambito geografico, contestuale, comportamentale o funzionale all'interno del quale il sistema di IA ad alto rischio è destinato a essere usato.

5. Nella misura in cui ciò sia strettamente necessario al fine di garantire il rilevamento e la correzione delle distorsioni in relazione ai sistemi di IA ad alto rischio in conformità del paragrafo 2, lettere f) e g), del presente articolo, i fornitori di tali sistemi possono eccezionalmente **trattare categorie particolari di dati personali**, fatte salve le tutele adeguate per i diritti e le libertà fondamentali delle persone fisiche.

**Definizione dati Particolari**

Riguardano i dati che rivelano origine razziale o etnica, convinzioni religiose, filosofiche o politiche, appartenenza a sindacati, oppure riguardano la **salute** o la vita sessuale.

Il trattamento di questi dati è regolamentato dall'articolo 9 del GDPR, che impone condizioni più severe rispetto al trattamento dei dati personali in general

# Requisiti sui dati

*Articolo 10* - **Dati e governance dei dati**

(segue) 5. Oltre alle disposizioni di cui ai regolamenti (UE) 2016/679 e (UE) 2018/1725 e alla direttiva (UE) 2016/680 devono essere soddisfatte, affinché tale trattamento avvenga, tutte le condizioni seguenti:

a) il rilevamento e la correzione delle distorsioni non possono essere realizzati efficacemente mediante il trattamento di altri dati, compresi i dati sintetici o anonimizzati;

b) le categorie particolari di dati personali sono soggette a limitazioni tecniche relative al riutilizzo dei dati personali, nonché a misure più avanzate di sicurezza e di tutela della vita privata, compresa la pseudonimizzazione;

c) le categorie particolari di dati personali sono soggette a misure tese a garantire che i dati personali trattati siano resi sicuri e protetti nonché soggetti a garanzie adeguate, ivi compresi controlli e documentazione rigorosi dell'accesso, al fine di evitare abusi e garantire che solo le persone autorizzate e sottostanti a opportuni obblighi di riservatezza abbiano accesso a tali dati personali;

d) le categorie particolari di dati personali non devono essere trasmesse, trasferite o altrimenti consultate da terzi;

e) le categorie particolari di dati personali vengono cancellate dopo che la distorsione è stata corretta oppure i dati personali hanno raggiunto la fine del loro periodo di conservazione, a seconda di quale delle due condizioni si verifica per prima

I tre documenti legislativi menzionati, il Regolamento (UE) 2016/679 (GDPR), il Regolamento (UE) 2018/1725 e la Direttiva (UE) 2016/680, riguardano tutti la protezione dei dati personali, ma in contesti diversi. Il **GDPR** disciplina il trattamento dei dati personali sia nel settore pubblico che in quello privato, mentre il Regolamento (UE) 2018/1725 riguarda specificamente le **istituzion**i e gli organi dell'Unione Europea. La Direttiva (UE) 2016/680 invece, si concentra sulla protezione dei dati personali nelle attività di **polizia e giudiziarie**.

# Bias - Toolkit

**Aequitas: A Bias and Fairness Audit Toolkit**

Pedro Saleiro, Benedict Kuester, Loren Hinkson, Jesse London, Abby Stevens, Ari Anisfeld, Kit T. Rodolfa, Rayid Ghani

https://doi.org/10.48550/arXiv.1811.05577

Recent work has raised concerns on the risk of unintended bias in AI systems being used nowadays that can affect individuals unfairly based on race, gender or religion, among other possible characteristics. While a lot of bias metrics and fairness definitions have been proposed in recent years, there is no consensus on which metric/definition should be used and there are very few available resources to operationalize them. Therefore, despite recent awareness, auditing for bias and fairness when developing and deploying AI systems is not yet a standard practice. We present **Aequitas**, an open source bias and fairness audit toolkit that is an intuitive and easy to use addition to the machine learning workflow, enabling users to seamlessly test models for several bias and fairness metrics in relation to multiple population sub-groups. Aequitas facilitates informed and equitable decisions around developing and deploying algorithmic decision making systems for both data scientists, machine learning researchers and policymakers.

# The legal viewpoint – Directive of EU Parl. & Council, on Liability; Adapting non contractual civil liability rules to AI

**«E' in corso di definizione la possibilità di ricondurre la responsabilità di diagnosi errate (ad esempio) agli stakeholder principali (fabbricante e operatore) se non è possibile definire una responsabilità di un algoritmo»**

Article 10 (16) of the **MDR** states:

*"Natural or legal persons may claim compensation for damage caused by a defective device in accordance with applicable Union and national law."*

There was no such regulation in the Medical Device Directive

The demonstration of the defectiveness of the device is for AI only possible when the decisions can be explained.

# AI Act vs USA e Cina

L'**Artificial Intelligence Act (AI Act)** dell'Unione Europea è una proposta legislativa volta a regolamentare l'uso dell'intelligenza artificiale (IA) all'interno degli Stati membri, adottando un approccio basato sul rischio per garantire la sicurezza e la tutela dei diritti fondamentali dei cittadini europei. Tuttavia, questa iniziativa ha suscitato critiche sia dagli Stati Uniti che dalla Cina, sebbene con motivazioni differenti. https://futuranetwork.eu/

## Critiche degli Stati Uniti:

1. **Impatto sull'Innovazione e sulla Competitività:** l'AI Act può ostacolare l'innovazione tecnologica e la competitività delle imprese, sia europee che americane. L'obbligo di conformarsi a requisiti stringenti potrebbe **disincentivare le startup e le piccole imprese** dall'entrare nel mercato europeo, riducendo gli investimenti in ricerca e sviluppo nel settore dell'IA.

2. **Trasparenza e Protezione dei Dati:** La richiesta di divulgare dettagli sui dataset utilizzati per l'addestramento dei modelli IA potrebbe entrare in conflitto con le leggi statunitensi sulla **proprietà intellettuale e la riservatezza commerciale**.

3. **Disallineamento Normativo**: Esiste il timore che si possano **creare barriere commerciali** e complicare la cooperazione internazionale nel settore dell'IA. necessità di un'armonizzazione delle normative per evitare **frammentazioni che potrebbero danneggiare l'industria tecnologica globale**.

## Critiche della CINA:

1. **Differenti Priorità**: controllo degli algoritmi e sulla gestione delle informazioni, con l'obiettivo di mantenere la stabilità sociale e politica, l'approccio europeo, che enfatizza la tutela dei diritti individuali e la prevenzione dei rischi associati all'IA

2. **Sovranità Tecnologica**

3. **Competizione Geopolitica**

# Grazie per l’attenzione



# *GRAZIE dell'attenzione !*

