Norme tecniche e valutazione della conformità accreditata per lo sviluppo dei sistemi di Intelligenza Artificiale

Daniele Nardi

Dipartimento di Ingegneria Informatica Automatica e Gestionale Antonio Ruberti

Sapienza Università di Roma

consorzio interuniversitario nazionale per l'informatica

Artificial Intelligence and Intelligent Systems

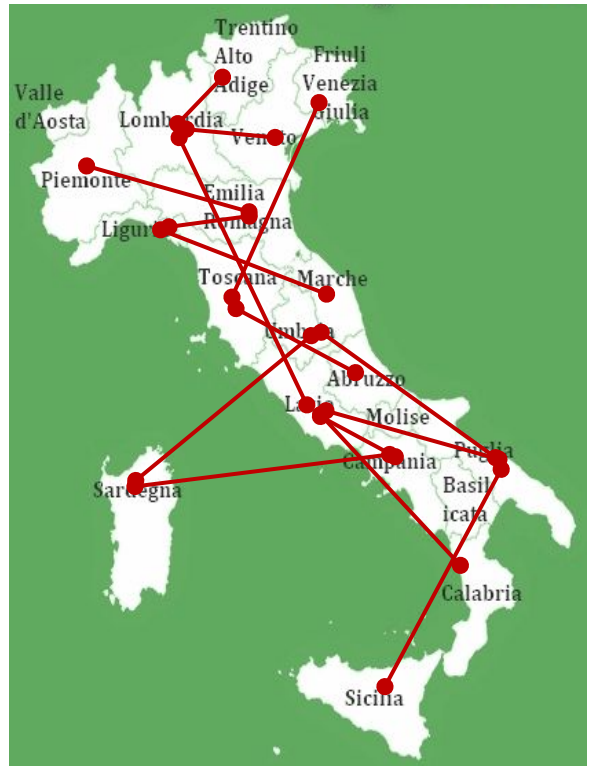
cini National Lab

Roma , 16 ottobre 2024

CINI e CINI-AIIS

Il CINI è un Consorzio di 54 Università in Informatica ed Ingegneria Informatica

Il laboratorio CINI “Artificial Intelligence and Intelligence Systems” è una rete di ricercatori in AI.



Il progetto con Accredia è stato svolto da

- Sapienza (Nardi, Bloisi*, Bisconti Lucidi**)
- Federico II (Sansone, Marrone, Marassi)

AI Act: origini e percorso

Prima legislazione mondiale sull'Intelligenza Artificiale, un punto di equilibrio tra innovazione tecnologica e diritti fondamentali.

2021 prima bozza

2024 approvazione finale

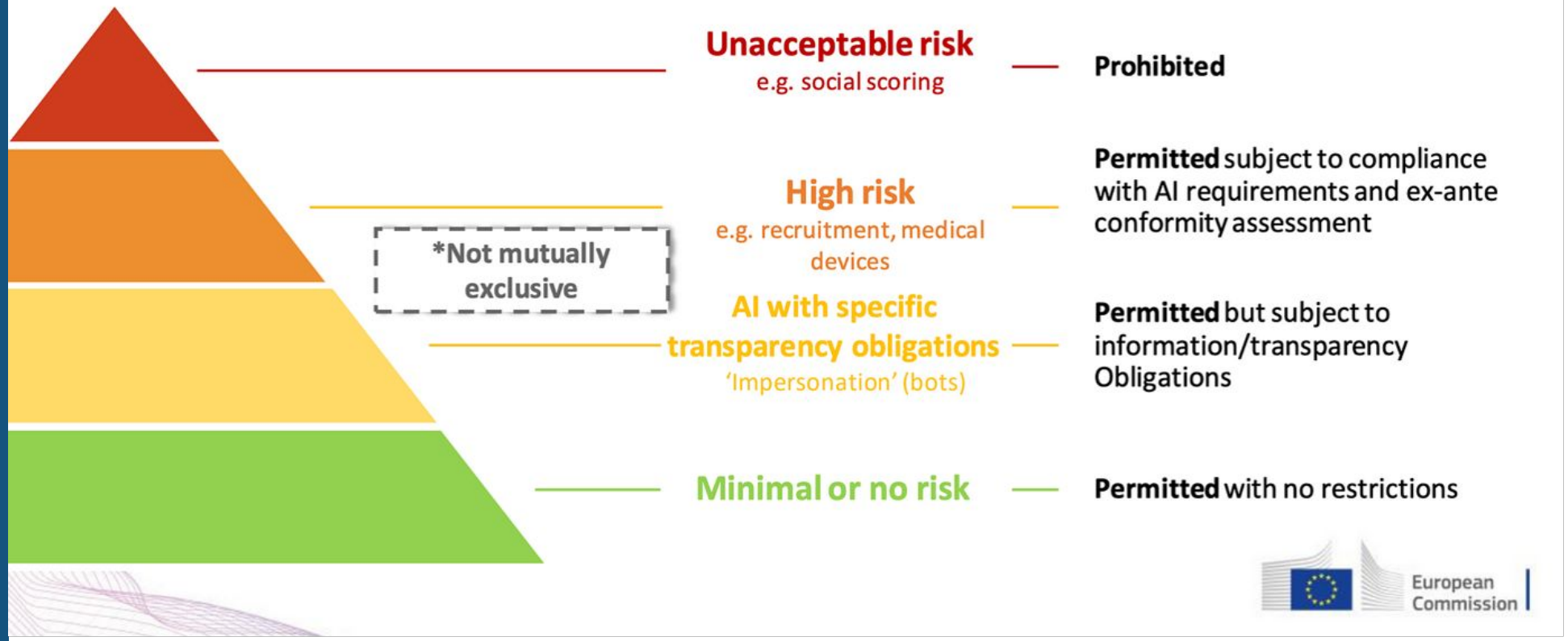
Principi guida AI ACT: (HLEG)

- Human agency and oversight
- Technical Robustness and safety
- Privacy and data governance
- Transparency
- Diversity, non-discrimination and fairness
- Societal and environmental well-being
- Accountability

AI-Act: oggetto

"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;

L’approccio basato sul rischio



Compliance HRS

- Data Quality
- Human Oversight
- Transparency
- Documentation
- Accuracy
- Robustness
- Security

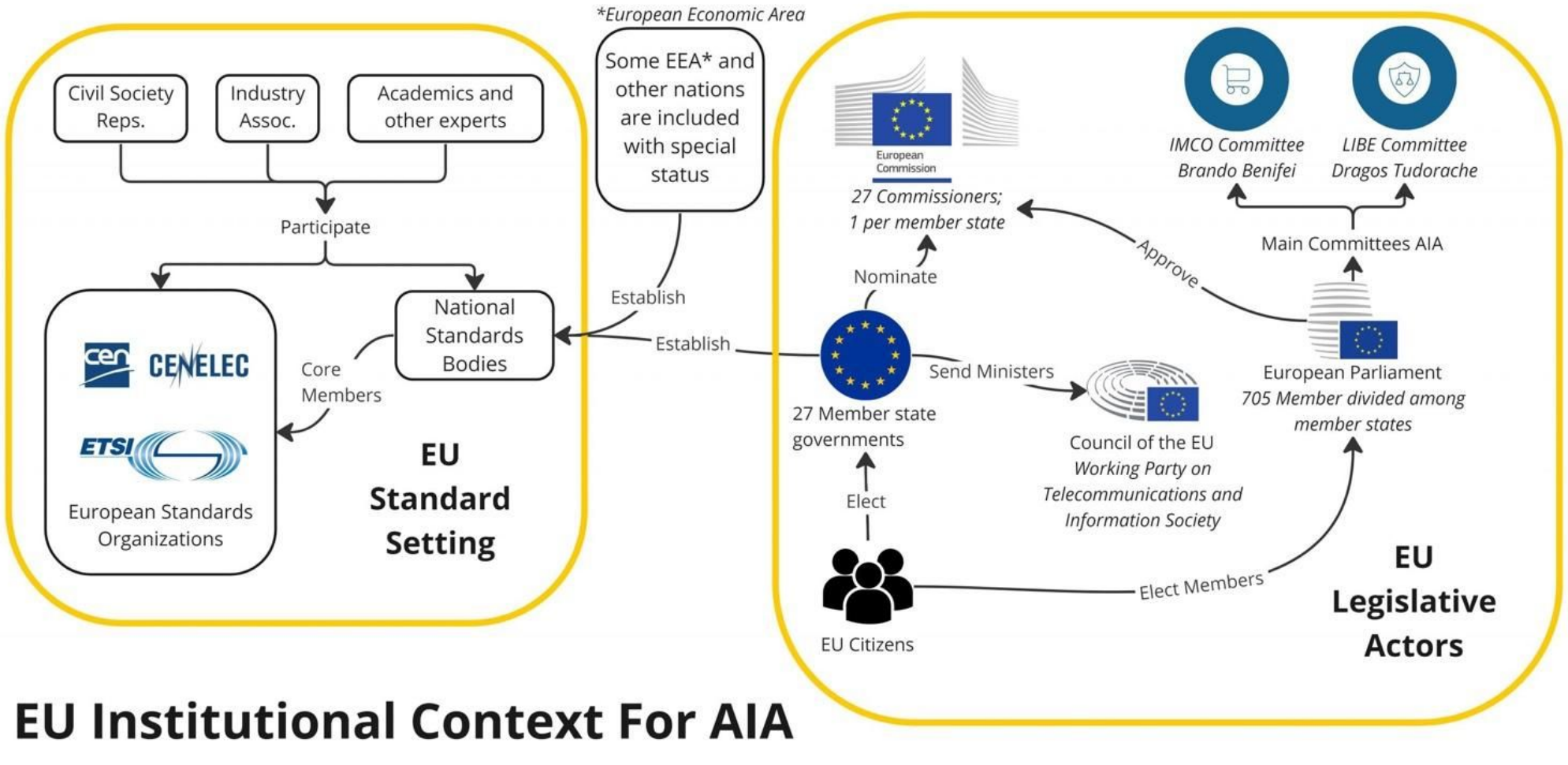
Sistemi ad alto rischio

Il sistema di IA è destinato a essere utilizzato come **componente di sicurezza** di un prodotto, o è esso stesso un prodotto, coperto dalla legislazione di armonizzazione dell'Unione elencata nell'Allegato I;

Settori critici: Sistemi IA che operano in settori fondamentali per sicurezza e diritti, tra cui:

- **Sanità**: Diagnosi medica, dispositivi di assistenza sanitaria.
- **Istruzione**: Sistemi di valutazione che influenzano l'accesso all'istruzione.
- **Occupazione**: Sistemi di reclutamento e monitoraggio delle performance dei lavoratori.
- **Finanza e Assicurazioni**: Valutazioni di credito, gestione del rischio finanziario, polizze vita e salute.
- **Infrastrutture critiche**: Sistemi che gestiscono energia, trasporti, acqua e telecomunicazioni.
- **Applicazione della legge**: Sistemi di sorveglianza e rilevazione di comportamenti illeciti.
- **Gestione della migrazione**: Valutazione del rischio alle frontiere e richieste di asilo.

Il percorso dell'AI-Act



EU Institutional Context For AIA

L’accreditamento

Autorità di notifica

Organismi notificati

Conformità:

- procedura di valutazione della conformità prevista dalla norma
- verifica da parte di un organismo notificato

Situazioni particolari in cui è comunque prevista una verifica di un organismo notificato

Nel mondo

Raccomandazioni

- US (Linee guida del NIST - Basate sul rischio)
- UK (Molto leggero)

- OECD (OCSE) Principi e linee guida

Regolamenti

- Cina

PoC (Proof of Concept)

Artificial Intelligence and Intelligent Systems
cini National Lab

Obiettivo: comprendere l'efficacia e le modalità applicative di alcune norme specifiche in vista di uno schema di certificazione e accreditamento.

Due PoC in ambito medicina:

- detection del melanoma (Sapienza)
- stratificazione dei pazienti con sclerosi multipla (Federico II)

Un PoC sulla governance dei sistemi AI, in collaborazione con INAIL

I PoC in medicina

Detection del melanoma per dispositivi portatili

Stratificazione dei pazienti con sclerosi multipla a supporto del medico specialista

Norma di riferimento: ISO 24027:2021, “Bias in AI systems and AI aided decision making”.

Bias/pregiudizio:

- dati
- cognitivo
- algoritmico

-

Bias	Metodo di verifica	Azione intrapresa	Note	Giudizio di conformità
Automation				
Group Attribution				
Implicit				
Confirmation				
In-Group				
Out-Group homogeneity				
Societal				
Rule-Based				
Requirement				

Distillato dai PoC in medicina

- Nella Detection del Melanoma l'acquisizione del dato può creare dei bias che vanno controllati (BIAS sui dati)

- L'utilizzo di approcci standard alla valutazione della presenza di bias sui dati si lega alla qualità dei dataset, che favorisce a lungo termine l'interoperabilità

- L'analisi del bias cognitivo varia in relazione all'utilizzatore finale, in relazione alla spiegabilità

- Il rapporto tra standard su AI e standard su dispositive medici non è ancora sufficientemente indagato e chiarito.

PoC con INAIL

Obiettivo: comprendere la complessità dell'implementazione della ISO 42001 all'interno di una organizzazione pubblica come INAIL

Norma di riferimento: ISO 42001, che assicura che vengano instaurate policy, misure di governance e di formazione all'interno dell'organizzazione relativamente ai sistemi di AI

- Il sistema di gestione della qualità è propedeutico per l'impostazione di misure di governance efficaci, e per la formazione degli stakeholder interni.
- L'implementazione della 42001 implica la scelta di quale delle differenti componenti di una organizzazione debba prendersi carico di quale azione.

Conclusioni

Inizio di un percorso

Attività interdisciplinare (CINI-AIIS, Sapienza, Federico II)

Educazione di tutti gli stakeholders:

- cittadini
- operatori specializzati
- pubbliche amministrazioni
- erogatori di servizi
- fornitori