EUSAiR
BRIDGING INNOVATION AND COMPLIANCE IN AI

Funded by the European Union

WORKSHOP GARR 2026
NET MAKERS

AI regulatory sandbox: il progetto EUSAiR

Le sandbox regolatorie per l'IA alla prova

Marco Billi
Postdoc @ Università di Bologna

Il percorso pilota della sandbox EUSAiR

Funded by the European Union

Fase 0	**Candidatura**	I provider si candidano tramite il modulo web EUSAiR, spesso con il supporto di EDIH e iniziative nazionali sull'IA.
Fase 1	**Triage e pianificazione**	Uno o due incontri di un'ora per classificare la tecnologia e individuare le sfide regolatorie critiche.
Fase 2	**Sperimentazione regolatoria**	Revisione documentale in 3–5 incontri da 1–2 ore su report di test e descrizioni dei dataset. Checklist su supervisione umana, logging, trasparenza e governance dei dati.
Fase 3	**Test avanzati (opzionale)**	Focalizzati su sfide regolatorie specifiche; capacità limitata.
Fase 4	**Report scritto finale**	Exit report con analisi e raccomandazioni formali (simulazione).

Partecipazione: domanda oltre le attese

Funded by the European Union

I pilot hanno superato il numero di test previsti dalla proposta: un segnale forte di domanda di orientamento regolatorio in vista degli obblighi per i sistemi ad alto rischio.

91

Candidature ricevute

casi d'uso da 87 istituzioni

78

Casi coinvolti

almeno un incontro svolto

52

Casi completati

da 48 istituzioni

13

senza primo incontro: nessuna risposta, ritiro o modulo di candidatura mancante

14

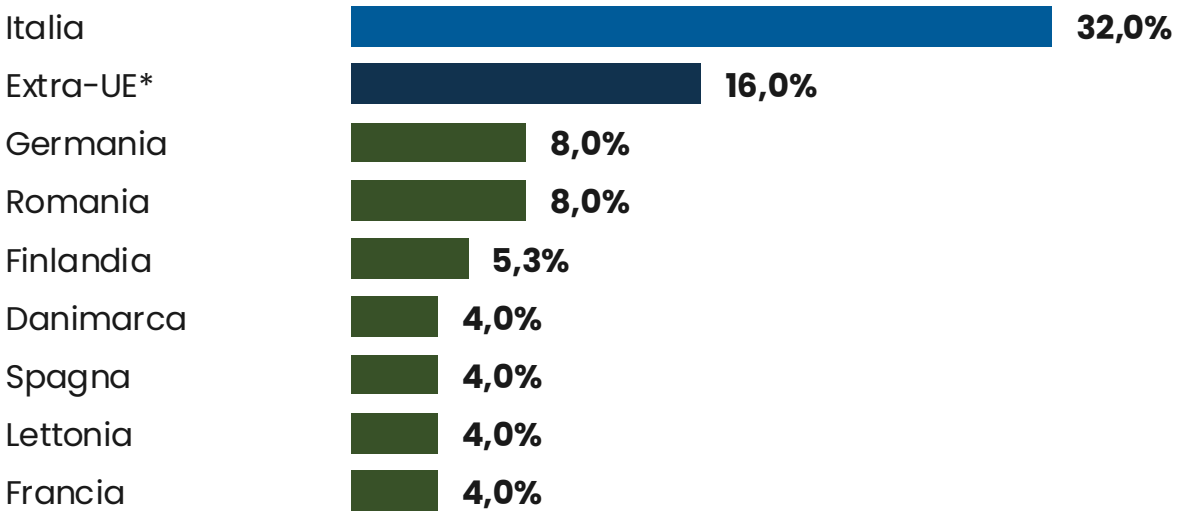
terminati in anticipo: non applicabilità, complessità degli NDA, disimpegno del provider

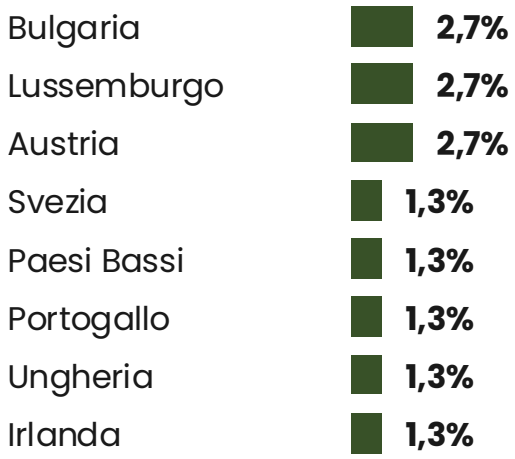
26

in corso o in fase di chiusura al 21 settembre 2026

Copertura geografica

Funded by the European Union





***Extra-UE:** Marocco, Singapore, Regno Unito, Stati Uniti, Cina — segno dell'interesse internazionale per la conformità all'AI Act e per il modello UE di governance delle sandbox.

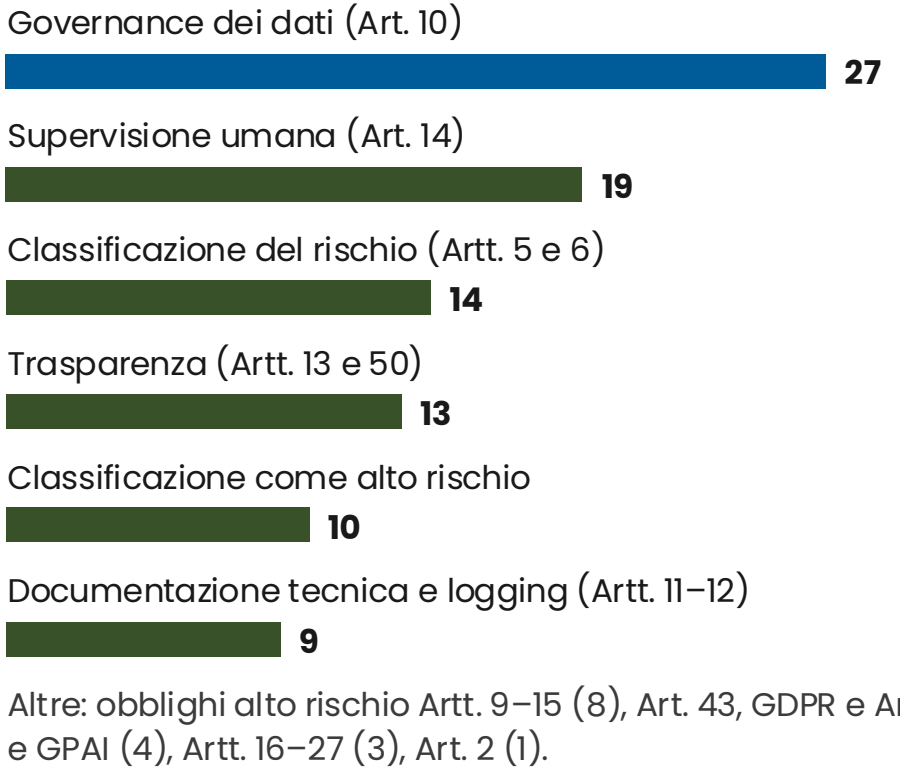
Settori e sfide regolatorie

Funded by the European Union

Settori dei casi d'uso

- Sanità e dispositivi medici
- Lavoro e selezione del personale
- Istruzione
- Finanza (AML, frodi)
- Pubblica amministrazione e servizi pubblici
- Infrastrutture critiche (energia, acqua)
- Moderazione dei contenuti e media
- Strumenti legali e di compliance
- Agricoltura e robotica
- Ingegneria ferroviaria
- Spazio
- Cybersicurezza e verifica dell'identità

Sfide più frequenti (n. di casi d'uso)



Altre: obblighi alto rischio Artt. 9–15 (8), Art. 43, GDPR e Art. 9 (5 ciascuno), Art. 15 e GPAI (4), Artt. 16–27 (3), Art. 2 (1).

Dieci lezioni dai piloti(1–5)

Funded by the European Union

01	**Classificare il deployment, non il prodotto**	Verifica dei media, videoanalisi ferroviaria, monitoraggio delle folle, career coaching e telecamere industriali cambiavano classe di rischio a seconda di dove e come erano impiegati: l'unità di analisi è lo scenario d'uso.
02	**L'autovalutazione trova le lacune, non certifica**	All'inizio molte risposte erano "non ancora" o "da definire". I provider hanno apprezzato trasformarle in azioni tracciabili: un QMS, una dichiarazione pubblica, il coinvolgimento del DPO.
03	**Ingegneria matura ≠ documentazione da AI Act**	Le autovalutazioni scendevano a "parzialmente soddisfacente" quando la domanda passava da "esiste?" a "ha un responsabile, è approvato e revisionato?".
04	**Dalle risposte sì/no alle metriche**	La sandbox ha spinto i provider verso metriche concrete: latenza dell'override, tasso di contestabilità, falsi negativi, risultati del red teaming.
05	**Team legali e tecnici insieme**	Scelte architetturali (blurring prima della registrazione, aggregazione on-edge) hanno risolto tensioni giuridiche, aggiungendo valore alla "compliance by design".

Dieci lezioni dai casi d'uso (6–10)

Funded by the European Union

06	**L'interazione tra normative è la regola**	Quasi ogni caso ha toccato il GDPR; altri MDR, Regolamento Macchine, NIS2, DORA, Regolamento AML, Direttiva Copyright DSM e legge italiana sull'IA
07	**L'Articolo 50 è l'obbligo più vicino**	Per i molti sistemi non ad alto rischio, la trasparenza è il requisito più concreto. Un'informativa a più livelli (avviso all'onboarding più un elemento persistente nell'interfaccia) è compatibile con la user experience.
08	**I sistemi cambiano: la conformità deve seguirli**	Cambiare o aggiornare il foundation model modifica comportamenti rilevanti senza cambiare le funzionalità. Servono confini di cambiamento predefiniti; manca una guida su cosa sia una "modifica sostanziale".
09	**Il coinvolgimento va gestito anche sul piano amministrativo**	I casi incompleti avevano cause comuni: documentazione tecnica assente, NDA non firmati, mancanza di capacità del provider per le attività della sandbox.
10	**I partner dell'ecosistema son necessari**	Le AI Factories hanno fornito test, come benchmark di jailbreak e di data leakage, che una microimpresa non avrebbe potuto progettare da sola.

Sistemi di IA, modelli GPAI e agenti

Funded by the European Union

Sistema di IA o modello GPAI?

Gruppo più numeroso: integratori a valle di modelli GPAI (chatbot turistici, recommender, analisi legislativa, istruzione, moderazione).

- Gli obblighi dei provider GPAI (Artt. 53–55) non passano a chi integra via API senza fine-tuning o modifiche sostanziali.
- L'integratore resta provider del proprio sistema: deve chiedere e usare la documentazione a monte (Allegato XII, Codice di condotta GPAI).
- Un prodotto può essere deployer di modelli terzi e provider del proprio layer fine-tuned.
- Molte domande "GPAI" riguardavano in realtà copyright, GDPR o l'Art. 50.

Sistemi agentici

AML, recupero crediti, redazione di DPIA, consulenza su bandi, audit di compliance, AI assurance.

- Imporre la finalità prevista a livello di runtime: compiti fuori scopo bloccati e registrati.
- LLM non deve prendere decisioni potenzialmente lesive dei diritti fondamentali, pertanto son necessari gate decisionali umani obbligatori prima di azioni esterne.
- Sicurezza degli agenti (content injection, accessi basati sull'identità, audit journal) come parte dell'Art. 15.

Questione aperta: in che modo I sistemi agentici sono contemplati dall'AI Act? E come si dimostra la conformità?

Allegato I: normativa di armonizzazione

Funded by the European Union

Il regime settoriale decide il percorso nell'AI Act: i due vanno gestiti insieme.

Dispositivi medici

Il test decisivo è la finalità medica secondo l'MDR, non il grado di automazione. Un dispositivo già in valutazione di Classe IIa rientra nell'AI Act; per un tool diagnostico interno di un ospedale pubblico conta l'eccezione in-house dell'Art. 5(5) MDR.

Standard

ISO 13485, ISO 14971, IEC 62304 e IEC 62366-1 coprono alcuni dei requisiti dell'AI Act e possono essere utili in attesa degli standard armonizzati CEN-CENELEC JTC 21.

Macchine

Essere componente di sicurezza (es. rilevamento ostacoli su una macchina agricola autonoma) rende il prodotto ad alto rischio: si raccomanda il disaccoppiamento architetturale. Telecamere industriali son diverse da varianti su macchina via Art. 6(1).

Fornitori di sola IA

Robot di magazzino e guida autonoma: come si ripartisce la responsabilità quando l'IA è fornita a parte rispetto a un prodotto certificato da altri. Tre percorsi (retrofit da operatore, integrazione da costruttore, provider dell'intero veicolo) con esiti giuridici diversi.

Classificazione del rischio

Funded by the European Union

La sfida più comune: è il prerequisito per definire un piano di sandbox adeguato.

Ruoli preparatori e human-in-the-loop	Moderazione che prepara decisioni, analisi legislativa per professionisti, bozze di DPIA riviste dai DPO: **non alto rischio. MA** validare un alert non elimina il rischio.
Sovra-conformità	Un servizio di career coaching era stato classificato ad alto rischio per prudenza: l'analisi per finalità mostra che lo stesso sistema può essere a rischio limitato in un'istituzione e ad alto rischio in un'altra.
Esenzioni condizionate	AML non ad alto rischio (Considerando 58) solo con una separazione ingegnerizzata e dimostrabile da usi creditizi o di eleggibilità. Nel Settore bancario, la forma giuridica dell'utente va considerata: determina l'Allegato III 5(b).
Zone grigie irrisolte	Analisi comportamentale di alunni da video senza identificazione: sospesa tra analisi comportamentale, trattamento biometrico e possibili divieti dell'Art. 5. Un concept di riconoscimento delle emozioni si è chiuso anticipatamente.
Alto rischio confermato	Colloqui di selezione automatizzati: valutare l'intera pipeline. Errori nell'estrazione dei CV o nel riconoscimento vocale si propagano nella valutazione LLM; il word error rate va verificato su accenti diversi.

Art. 10: cosa significa “rappresentativo”

Funded by the European Union

L’obbligo analizzato più spesso in profondità nei pilot.

Accuratezza ≠ rappresentatività

Un dataset clinico di un solo ospedale era privo di errori, versionato e completo, ma solo in parte rappresentativo: la rappresentatività dipende da popolazione e contesto. Lo sbilanciamento tra classi diventa un dovere di monitoraggio continuo.

Oltre la demografia

Per i dataset video orientati alla sicurezza, rappresentatività significa copertura degli scenari operativi (luce, meteo, occlusioni, angolo di ripresa), misurata con i tassi di eventi mancati e di falsi allarmi.

Ground truth

Dove manca uno standard di riferimento definitivo, come per le lesioni ovariche non chirurgiche, il protocollo di etichettatura per consenso di esperti va documentato nel fascicolo di conformità.

Sottogruppi

Lo screening dermatologico richiede report per fototipo, sede corporea e dispositivo, con le lacune dichiarate nelle istruzioni per l’uso. Nella salute mentale i dati mancanti sono un segnale di rischio: chi peggiora spesso smette di usare l’app.

In sintesi: cosa ci dicono i pilot

Funded by the European Union

1. La classificazione del rischio viene prima di tutto, e va fatta per **scenario di deployment**, non per prodotto.
2. AI Act, GDPR e regimi settoriali vanno letti insieme: servono **mappe dell'interazione normativa**.
3. Contano le **evidenze misurabili** e una documentazione con responsabili, approvazioni e revisioni.
4. Servono indicazioni su **aggiornamenti dei modelli e modifiche sostanziali**, oltre che su Art. 10 e sistemi agentici.
5. Una sandbox funziona se è **ben gestita** (NDA, documentazione, capacità del provider) e se attiva partner come le AI Factories.

Grazie

Marco Billi · Università di Bologna

Consortium

Funded by the European Union

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. Grant Agreement number 101195535.

WORKSHOP GARR 2026

NET MAKERS