#### L'uso criminale dell'IA

L’ IA può svolgere un ruolo sempre più **essenziale** negli atti criminali in futuro. Esempi chiari di che vengono chiamati “crimini di AI” (**CIA**) sono forniti da due esperimenti di ricerca (teorici).

Nel primo, due scienziati sociali computazionali hanno usato l’AI come strumento per convincere utenti di social media a cliccare su collegamenti di phishing; poiché ogni messaggio è stato costruito con tecniche di ML applicate ai comportamenti passati e ai profili pubblici degli utenti, il contenuto è stato ritagliato su ciascun individuo.

Nel secondo esperimento, tre scienziati informatici hanno simulato un mercato e hanno scoperto che gli agenti di scambio potevano apprendere ed eseguire una “vantaggiosa” campagna di manipolazione del mercato che includeva una serie di falsi ordini ingannevoli.

Di seguito, l'analisi riportata risponde a due domande: 1. Quali sono le minacce fondamentalmente peculiari e plausibili poste dai CIA? 2. Quali soluzioni sono disponibili o possono essere elaborate per affrontare i CIA?

#### Preoccupazioni

Un'analisi iniziale di revisione della letteratura ha filtrato i risultati relativi ad atti o omissioni criminali che: - si sono vericati o probabilmente si vericheranno in base alle attuali tecnologie di (**plausibilità**); - richiedono l’ come fattore essenziale (**unicità**); - sono perseguiti nel diritto nazionale

![[Pasted image 20240118224036.png]]

Ciò ha portato a individuare cinque aree criminali potenzialmente interessate dai CIA: 1. **commercio, mercati finanziari e insolvenza** 2. **droghe** nocive o pericolose 3. reati **contro la persona** (inclusi omicidio doloso o colposo, molestie, stalking, tortura); 4. reati **sessuali** (compresi stupro, aggressione sessuale); 5. **furto e frode**, contraffazione e sostituzione di persona.

Vengono identificate 4 ragioni di preoccupazione:

1. **Emergenza**

   La preoccupazione per l’emergenza si riferisce al fatto che un’analisi superciale del design e dell’implementazione di un agente articiale potrebbe suggerire un tipo particolare di comportamento relativamente semplice, ma la verità è che, a seguito dell’implementazione, l’AI può agire in modi potenzialmente più sosticati che vanno oltre le nostre aspettative iniziali. Pertanto, azioni e piani coordinati possono **emergere autonomamente**.

Il comportamento emergente potrebbe avere implicazioni criminali, nella misura in cui devia dal design originale

2. **Responsabilità**

La preoccupazione relativa alla responsabilità si riferisce al fatto che i CIA potrebbero minare i modelli di responsabilità esistenti, minacciando così il potere dissuasivo e riparatore della legge.

La prima condizione per la responsabilità penale è l'**actus reus**: un atto o un'omissione criminale posta in essere **volontariamente**. Per le tipologie di deniti in modo tale che solo l' può realizzare l'atto o l'omissione criminale, l'aspetto volontario dell'actus reus potrebbe non essere mai soddisfatto poiché l'idea che un possa agire volontariamente è priva di fondamento.

Quando la responsabilità penale è basata sulla colpa, ha anche una seconda condizione, la **mens rea** (una mente colpevole). ^b17b93 ^bdc55f ^8009a2 ^c9f83d

Un agente articiale può essere **responsabile causalmente** di un atto criminale, ma soltanto un agente umano può esserne **moralmente responsabile**.

La complessità dell'AI fornisce un grande incentivo agli agenti umani per evitare di scoprire cosa sta facendo esattamente il sistema di AI, poiché meno gli agenti umani sanno, più saranno in grado di negare la loro responsabilità

In alternativa, i legislatori possono denfiire la responsabilità penale **senza un requisito di colpa**; ciò porterebbe ad attribuire la responsabilità alla persona giuridica senza colpa che ha attivato un nonostante il rischio che possa plausibilmente compiere un'azione o un'omissione criminale. La responsabilità si applica agli agenti che **fanno la differenza** in un sistema complesso in cui i singoli agenti svolgono azioni neutrali che però sfociano in un crimine collettivo.

1. **Monitoraggio**

La preoccupazione per il monitoraggio dei fa riferimento a tre tipi di problemi: **attribuzione, fattibilità e azioni intersistemiche**.

L'**attribuzione** del mancato rispetto della normativa vigente costituisce un problema di monitoraggio degli utilizzati come strumenti di reato, dovuta alla capacità di questa nuova tipologia di agenti smart di operare in modo indipendente e autonomo: due caratteristiche che tendono a confondere ogni tentativo di tracciare la responsabilità riconducendo gli effetti di un'azione all'autore del reato. ^7b4251

Per quanto riguarda la fattibilità del monitoraggio, l'autore di un reato può trarre vantaggio dai casi in cui gli operano a velocità e livelli di complessità che vanno semplicemente al di là della capacità di monitorarne la conformità con le norme.

Le azioni intersistemiche fanno riferimento a un problema per i sistemi di monitoraggio dei con visione a tunnel che si **concentrano solo su un singolo sistema**.

1. **Psicologia**

   La **psicologia** fa riferimento alla preoccupazione che l' possa inuenzare/manipolare negativamente lo stato mentale di un utente no al punto di agevolare o causare (in tutto o in parte) il crimine. Un effetto psicologico si basa sulla capacità degli di ottenere la ducia degli utenti, rendendo le persone vulnerabili alla manipolazione.

## Minacce

#### 1. Commercio, mercati finanziari e insolvenza

Attualmente, sorgono problemi nel caso del coinvolgimento dell' soprattutto in tre aree: **manipolazione del mercato, fissazione dei prezzi e collusione**.

La **manipolazione del mercato** è definita come quelle "azioni e/o operazioni da parte di partecipanti al mercato che tentano di infuenzare artificialmente i prezzi di mercato".

È stato dimostrato che tali forme di inganno emergono da un'implementazione apparentemente conforme di un progettato AA per operare per conto di un utente (cioè un agente articiale di trading). Questo perché un AA, in particolare uno che apprende da osservazioni reali o simulate, può imparare a generare segnali che sono effettivamente ingannevoli: - effettuare ordini senza alcuna intenzione di eseguirli, semplicemente per manipolare gli onesti partecipanti al mercato - acquisire una posizione in uno strumento finanziario, come un titolo, per poi gonare articialmente il titolo tramite la sua promozione fraudolenta

Questo è noto in termini colloquiali come schema "pompa e sgona" (**pump-and-dump**).

La **collusione**, sotto forma di **fissazione dei prezzi**, può emergere anche nei sistemi automatizzati grazie alle capacità di pianificazione e autonomia degli AA: - algoritmi imparano a coordinarsi, ad esempio per tenere un prezzo alto

L'assenza di intenzionalità, l'intervallo decisionale molto breve e la probabilità che la collusione emerga a seguito delle interazioni tra sollevano

anche serie preoccupazioni per quanto riguarda la [[8. Cattive pratiche - l'uso improprio dell'AI per il male sociale#^c9f83d|responsabilità]] e il [[#7b4251|monitoraggio]].

#### 2. Droghe nocive o pericolose

> I crimini che rientrano in questa categoria includono il **traffico, la vendita, l'acquisto e il possesso di droghe vietate.**

In questo caso, l'AI può fungere da strumento per sostenere il traco e la vendita di sostanze illecite. Il traffco business-to-business di droga che utilizza l'AI è una minaccia dovuta ai criminali che adoperano **veicoli senza equipaggio**, che fanno leva sulla pianicazione dell'AI e sulle tecnologie di navigazione autonoma come strumenti per migliorare i tassi di successo del contrabbando.

Poiché le reti di contrabbando vengono fermate dal monitoraggio e dall'intercettazione delle linee di trasporto, l'applicazione delle norme diventa più dicile quando vengono usati veicoli senza equipaggio per trasportare ciò che è contrabbandato.

#### 3. Reati contro la persona

> I crimini che rientrano nella categoria dei reati contro la persona riguardano **molestie** e **torture**.
>
> Le **molestie** comprendono comportamenti intenzionali e ripetitivi che generano allarme o causano disagio a una persona.

Per quanto riguarda i CIA basati sulle molestie, la letteratura fa riferimento ai social bot. Un malintenzionato può avvalersi di un social bot come strumento di molestia diretta o indiretta. La molestia diretta è costituita dalla diffusione di messaggi di odio contro la persona.

> Per quanto riguarda la tortura, si configura quando un pubblico ufficiale infligge intenzionalmente gravi dolori o sofferenze a un altro soggetto nell'esercizio o nel presunto esercizio delle sue funzioni ufficiali.

L'uso dell'IA per l'interrogatorio è motivato dalla sua capacità di rilevare meglio l'inganno, l'emulazione dei tratti umani (come la voce) e la modellazione affettiva per manipolare l'interrogato. Tuttavia, un con queste capacità può imparare a **torturare una vittima**.

L'interrogato probabilmente sa che l'AI non può comprendere il dolore o provare empatia, ed è pertanto improbabile che agisca con pietà e interrompa l'interrogatorio. Senza compassione la semplice presenza di un di interrogatorio può far capitolare il soggetto per paura, il che, secondo il diritto internazionale, potrebbe costituire un **crimine di tortura** (minacciata). Inoltre, chi si avvale di un AA può essere in grado di distaccarsi, **emotivamente e fisicamente**; perciò, diventa più facile ricorrere alla tortura.

#### 4. Reati sessuali

> I reati sessuali discussi in letteratura in relazione all’ sono i seguenti: stupro (cioè sesso penetrativo senza consenso), aggressione sessuale (cioè contatto sessuale senza consenso) e rapporti o attività sessuali con un minore.

Questi crimini coinvolgono l’AI quando, tramite un’interazione avanzata uomo-computer, quest’ultima **promuove l’oggettivazione sessuale o l’abuso e la violenza sessualizzati**, e potenzialmente simula e quindi **aumenta il desiderio sessuale** per i reati sessuali.

#### 5. Furto e frode, contraazione e sostituzione di persona

> Contraffazione e sostituzione di persona sono collegate tramite i CIA a furti e frodi extra-aziendali, con implicazioni anche per l’uso di nelle frodi aziendali.

Per quanto riguarda il furto e la frode extra-aziendale, il processo prevede due fasi. - Inizia con l’utilizzo dell’AI per raccogliere dati personali - usando social bot di social media - phishing - procede con l’utilizzo dei dati personali rubati e di altri metodi di AI per forgiare un’identità che induca le autorità bancarie a eettuare una transazione (ovvero furto e frode bancaria)

## Soluzioni disponibili

#### 1. Affrontare l'emergenza

> Le soluzioni giuridiche possono comportare la limitazione dell’autonomia degli agenti o del loro impiego. - Per esempio, la Germania ha creato contesti deregolamentati in cui è consentita la sperimentazione di automobili a guida autonoma

#### 2. Affrontare la responsabilità

4 modelli:

- **responsabilità diretta**: attribuisce gli elementi fattuali e mentali a un AA
  - un limite fondamentale di questo modello risiede nel fatto che gli AA non hanno personalità giuridica e capacità di agire, e quindi non possiamo ritenere un AA legalmente responsabile
  - porterebbe a una deresponsabilizzazione degli agenti umani dietro l'AA
- **perpetrazione per mezzo di altri**: l'AA è uno strumento il cui orchestratore è il vero autore
  - 3 candidati umani: programmatori, produttori e utilizzatori
    - per essere responsabile, l’operatore di un deve volere la realizzazione del fatto illecito

- **responsabilità di comando**: nei contesti in cui esiste una catena di comando, attribuisce la responsabilità a **qualsiasi ufficiale che sia a conoscenza ma non si adopera per prevenire i crimini**
    - Tuttavia questioni relative a livelli di crescente complessità nella programmazione, relazioni robo-umane e integrazione in strutture gerarchiche, mettono in discussione la sostenibilità di queste teorie.
- **conseguenza naturale e probabile**: concerne i casi di in cui uno sviluppatore o un utente di non intendono né hanno conoscenza a priori di un reato.
    - La responsabilità è attribuita allo sviluppatore o all'utente se il danno è conseguenza naturale e probabile della loro condotta, esponendo gli altri in modo imprudente o negligente al rischio

#### 3. Controllo del monitoraggio

Ci sono quattro meccanismi principali per affrontare il monitoraggio dei CIA.

1) Elaborare predittori dei utilizzando la conoscenza del dominio 2) Utilizzare la simulazione sociale per scoprire schemi ricorrenti di criminalità 3) Affrontare la tracciabilità lasciando indizi rivelatori nelle componenti che costituiscono gli strumenti dei CIA 4) Effettuare monitoraggio intersistemico e avvalersi dell'auto-organizzazione tra sistemi - concepire un sistema che assume il ruolo di paziente morale.

#### 4. Affrontare la psicologia

> Ci sono due preoccupazioni principali relative all'elemento psicologico dei CIA: la manipolazione degli utenti e (nel caso dell' antropomorfica) la creazione in un utente del desiderio di compiere un crimine.

Se gli antropomorci costituiscono un problema, allora possono esserci due approcci: - limitare gli AA antropomorfici che consentono di simulare un crimine - servirsi di antropomorfici come modo per respingere i reati sessuali simulati (incompatibile con il primo)