
ANNO ACCADEMICO 2024/2025

Etica, Società e Privacy

Teoria

Altair's Notes



UNIVERSITÀ
DI TORINO



DIPARTIMENTO DI INFORMATICA

CAPITOLO 1	PRIVACY - INTRODUZIONE	PAGINA 5
1.1	Il Corso in Breve...	5
1.2	Privacy e Leggi sulla Privacy	5
	Che Cos'è la Privacy? — 5 • Perché la Privacy è Così Importante? — 6 • Privacy negli Stati Uniti — 8 • Privacy nell'Unione Europea — 8 • Regolamenti Transazionali — 13	
CAPITOLO 2	ETICA - INTRODUZIONE	PAGINA 16
2.1	Il Corso in Breve...	16
	Perché Questo Corso? — 17	
2.2	La Costruzione della Realtà Sociale	19
	Come l'AI Influenza la Realtà Sociale — 19 • L'AI Rappresenta Effettivamente il Mondo? — 20 • I Pericoli dei Language Models — 21	
CAPITOLO 3	ANCORA SULL'AI	PAGINA 24
3.1	AI or not AI This is the Dilemma	24
	Il Watermark — 24 • Il Copyright — 25	

Premessa

Licenza

Questi appunti sono rilasciati sotto licenza Creative Commons Attribuzione 4.0 Internazionale (per maggiori informazioni consultare il link: <https://creativecommons.org/version4/>).



Formato utilizzato

Box di "Concetto sbagliato":

Concetto sbagliato 0.1: Testo del concetto sbagliato

Testo contenente il concetto giusto.

Box di "Corollario":

Corollario 0.0.1 Nome del corollario

Testo del corollario. Per corollario si intende una definizione minore, legata a un'altra definizione.

Box di "Definizione":

Definizione 0.0.1: Nome delle definizioni

Testo della definizione.

Box di "Domanda":

Domanda 0.1

Testo della domanda. Le domande sono spesso utilizzate per far riflettere sulle definizioni o sui concetti.

Box di "Esempio":

Esempio 0.0.1 (Nome dell'esempio)

Testo dell'esempio. Gli esempi sono tratti dalle slides del corso.

Box di "Note":

Note:-

Testo della nota. Le note sono spesso utilizzate per chiarire concetti o per dare informazioni aggiuntive.

Box di "Osservazioni":

Osservazioni 0.0.1

Testo delle osservazioni. Le osservazioni sono spesso utilizzate per chiarire concetti o per dare informazioni aggiuntive. A differenza delle note le osservazioni sono più specifiche.

1

Privacy - Introduzione

1.1 Il Corso in Breve...

Obiettivi:

- Riconoscere problemi di privacy nella modellazione e nell'analisi dei dati.
- Conoscenza di base su metodi per preservare la privacy.

Syllabus:

1. Il concetto di privacy e le leggi sulla privacy in differenti paesi.
2. Le sfide della privacy nell'era dei Big Data.
3. Sistemi di informazione.
4. Modelli statistici.
5. Attacchi alla privacy e modelli di anonimizzazione in database statistici.
6. Privacy differenziale.
7. Separazione dei dati.

1.2 Privacy e Leggi sulla Privacy

1.2.1 Che Cos'è la Privacy?

Domanda 1.1

Che cos'è la privacy?

Definizione 1.2.1: Privacy 1

La *privacy* può essere definita come il diritto a stare da soli.

Warren and Brandeis (1890), *The Right to Privacy*:

- Una delle tesi più influenti nella storia americana.
- GLi autori tentarono di trovare un modo di descrivere legalmente la privacy.

Note:-

Esempio: il diritto di una persona a scegliere la seclusione dalle attenzioni altrui, il diritto di non essere osservati nella sfera privata.

Definizione 1.2.2: Privacy 2

La *privacy* può essere definita come accesso limitato alle informazioni.

- Una persona deve essere libera di scegliere in che misura partecipare alla società senza che gli altri debbano sapere.
- Godkin (1880): "nothing is better worthy of legal protection than private life, or, in other words, the right of every man to keep his affairs to himself, and to decide for himself to what extent they shall be the subject of public observation and discussion."
- Bok (1989): la privacy è "the condition of being protected from unwanted access by others—either physical access, personal information, or attention."

Definizione 1.2.3: Privacy 3

La *privacy* può essere vista come controllo sull'informazione.

- Westin and Blom-Cooper (1970): "privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others."
- Fried (1968): "Privacy is not simply an absence of information about us in the minds of others; rather it is the control we have over information about ourselves."

Definizione 1.2.4: Privacy FIPS PUB 41

Il diritto di un'entità ... a determinare il grado con il quale interagire con il proprio ambiente, compreso il grado con cui un'entità voglia condividere informazioni personali con gli altri.

Definizione 1.2.5: Privacy ISO

Il diritto di un individuo a controllare o influenzare quali informazioni collegate a loro possono essere collezionate e salvate e da chi e a chi queste informazioni possono essere accedute.

Osservazioni 1.2.1

Definizioni derivabili:

- La privacy è l'abilità di una persona di controllare la disponibilità di *informazioni* e la sua *esposizione*.
- È collegata a essere abili a funzionare in una società *anonimamente*.

Edward Snowden files: viene pubblicato il file contenente le informazioni riguardo i dati raccolti dal NSA riguardo le chiamate di milioni di privati cittadini.

1.2.2 Perché la Privacy è Così Importante?

Domanda 1.2

Perché la privacy è così importante?

La privacy è importante perché:

- *Sentimenti individuali:*
 - Non confortabile: possesso dell'informazione.
 - Non sicuro: l'informazione può essere usata in modo improprio (furto d'identità).
- *Le aziende hanno bisogno di:*
 - Far sì che i loro clienti si sentano al sicuro.
 - Mantenere una buona reputazione¹.
 - Proteggere sé stesse da ogni disputa legale.
 - Obbedire alle leggi.

Tipi di privacy:

- Political privacy.
- Consumer Privacy.
- Medical privacy.
- Private property.
- Information/Data privacy.

Definizione 1.2.6: Data Privacy

Il problema della *data privacy* emerge quando dei dati unicamente associabili a un individuo sono collezionati e salvati.

Le fonti più comuni di dati affetti da data privacy:

- Record sanitari.
- Investigazioni e processi criminali.
- Transazioni finanziarie.
- Trattati biologici (e. g. materiale genetico).
- Residenza e posizione geografica.
- Geolocalizzazione.
- Utilizzo del web.

Osservazioni 1.2.2

- La sfida della data privacy è trovare un modo per *condividere dati* proteggendo le informazioni che permettono di identificare una determinata persona.
 - Per esempio negli ospedali i dati vengono trasferiti in maniera aggregata.
 - L'idea di condividere i dati in forma aggregata garantisce che solo dati non identificabili sono condivisi.
- La protezione legale del diritto alla privacy *cambia drasticamente a seconda della nazione*.

¹Beh, visto che merda sono Twitter e META non credo gliene fregghi qualcosa.

1.2.3 Privacy negli Stati Uniti

La data privacy *non è molto regolata* negli USA:

- Non c'è una legge che controlli tutti gli aspetti dei dati (acquisizione, collezione, trattamento e uso).
- Se un'azienda colleziona dei dati (anche senza permesso) ha il diritto di utilizzarli.
- Gli istituti possono informarsi sulle condizioni finanziarie di una persona (banche, assicurazioni, etc.) chiedendo report a terze parti.
- Ci sono alcune eccezioni:
 - Dati sanitari (HIPAA).
 - Dati dei bambini sotto i 13 anni online (COPPA).
 - Richieste di prestito (FCRA).
 - Sicurezza informatica (ECPA, PATRIOT², etc.).

1.2.4 Privacy nell'Unione Europea

La data privacy nell'UE *è pesantemente regolata*:

- l'articolo 8 della convenzione europea sui diritti umani (ECHR) prevede il diritto al rispetto della privacy di una persona (poter disporre di una sfera privata, una vita familiare, un domicilio e della propria corrispondenza).
- La corte dei diritti umani ha dato a quest'articolo varie interpretazioni, con le seguenti eccezioni:
 - Ottenere informazioni per censimenti ufficiali.
 - Raccogliere impronte digitali e fotografie per attività di polizia.
 - Collezionare dati medici e spese personali.
 - Implementare sistemi di identificazione personale (un documento di identificazione).

Note:-

Spesso questa maggiore tutela della privacy viene criticata dalle aziende che la vedono come un freno al progresso.

Ci sono state due ere della privacy:

- 1995-2018: Data Protection Directive (DPD).
- 2018-presente: General Data Protection Regulation (GDPR³).

Definizione 1.2.7: Data Protection Directive

La Data Protection Directive (DPD) fu adottata nel 1995 dal parlamento europeo:

- Come tentativo di armonizzare la protezione dei dati nell'unione europea.
- Doveva essere regolamentata da leggi nazionali nel 1998.

²Nel 2001, in seguito a un certo evento terroristico.

³Promulgato nel 2016, ma diventato effettivo dal maggio del 2018

Gli 8 principi di base del DPD:

1. I dati personali devono essere processati a norma di legge e in maniera corretta.
2. I dati personali devono essere processati solo per certi scopi limitati.
3. I dati personali devono essere processati in modo adeguato, rilevante e non eccessivo.
4. I dati personali devono essere processati accuratamente.
5. I dati personali devono essere trattenuti solo per il tempo strettamente necessario.
6. I dati personali devono essere processati in accordo con i diritti del soggetto interessato.
7. I dati personali devono essere processati in modo sicuro.
8. I dati personali devono essere trasferiti solo a nazioni con una protezione adeguata.

Definizione 1.2.8: Personal Data - DPD

Ogni informazione relativa a una persona identificata o identificabile. Una persona identificabile è una persona che può essere identificata, direttamente o indirettamente, in particolare identificando un particolare numero di identificazione o uno o più fattori fisici, psicologici, mentali, economici, culturali o sociali.

Note:-

Questa definizione viene rafforzata nel GDPR.

Definizione 1.2.9: Data Processing - DPD

Ogni operazione o insieme di operazioni che viene effettuata su dati personali, tramite un mezzo automatizzato o meno, come la collezione, registrazione, organizzazione, immagazzinamento, adattamento o alterazione, recupero, consultazione, uso, trasmissione, disseminazione o altra distribuzione, allineamento o combinazione, blocco, cancellazione o distruzione.

Note:-

Vengono effettuati cambiamenti minori nel GDPR.

Definizione 1.2.10: Responsabile - DPD

Il responsabile può essere una persona fisica o giuridica, un'autorità pubblica o un ente che deve garantire l'integrità dei dati trattati.

Corollario 1.2.1 Processore

Il processore dei dati è una persona fisica o giuridica, un'autorità pubblica, un'agenzia o qualunque altro ente che processa i dati personali per conto del responsabile.

Note:-

Non ci sono cambiamenti nel GDPR.

Principi della gestione dei dati. I dati personali non dovrebbero essere processati a meno che non riguardino queste categorie:

- **Trasparenza:** il soggetto dei dati ha il diritto di essere informato da un'azienda che sta elaborando i suoi dati personali.
- **Scopo legittimo:** i dati personali possono essere processati solo per scopi legittimi e non usati per altri scopi non pertinenti.
- **Proporzionalità:** i dati personali processati devono essere adeguati, rilevanti e non eccessivi in relazione allo scopo per quale i dati vengono collezionati e ulteriormente processati.

Trasparenza:

- Il soggetto dei dati deve dare il *consenso*.
- Il processamento era necessario per l'*esecuzione di un contratto*.
- Il processamento era necessario per *assolvere un obbligo legale*.
- Il processamento era necessario per *proteggere gli interessi vitali* del soggetto dei dati.
- Il processamento era necessario per un'operazione di *interesse pubblico*.
- Il processamento era necessario per *scopi di interesse legittimato dal controllore, eccetto quando quegli interessi sono sovrascritti dagli interessi dei diritti fondamentali e della libertà del soggetto*.

Osservazioni 1.2.3

Il soggetto ha il diritto a:

- Accedere a tutti i dati su di lui/lei/loro.
- Domandare la rettifica, cancellazione o blocco dei dati se sono incompleti, inaccurati o non sono processati come previsto dalle regole sulla protezione dei dati.

Proporzionalità:

- I dati devono essere accurati e, quando necessario, aggiornati.
- I dati non devono essere mantenuti in una forma che permette l'identificazione del soggetto per più tempo del necessario.
- Gli stati membri hanno la possibilità di memorizzare i dati personali per fini statistici o scientifici.
- Vengono applicate tutele aggiuntive per la gestione dei *dati sensibili* (credenze religiose, opinioni politiche, salute, orientamento sessuale, gruppo etico, appartenenza a organizzazioni).
- Il soggetto può sempre chiedere la cancellazione dei suoi dati usati per fini pubblicitari⁴.
- Qualunque decisione automatizzata sulla persona non deve essere fatta in maniera completamente automatica.
- L'individuo può *fare ricorso* su qualsiasi decisione automatica in cui vengono processati i propri dati.

Autorità della privacy:

- Ogni stato membro deve eleggere un'*autorità di supervisione* che:
 - Deve *monitorare la protezione dei dati* in quello stato membro.
 - Dare *avvisi al governo riguardo le misure amministrative e i regolamenti*.
 - *Iniziare procedure legali* quando il regolamento sulla protezione dei dati viene violato.
- Il *controllore* deve notificare all'autorità di supervisione le seguenti informazioni:
 - Il nome e l'indirizzo del controllore.
 - Lo scopo del processamento.
 - Una descrizione delle categorie dei dati del soggetto.
 - Il recipiente a cui i dati possono essere inoltrati.
 - Proposte di trasferimento dei dati a stati terzi.
 - Una descrizione generale delle misure prese per garantire la sicurezza dei dati processati.
- Le informazioni sono tenute in un *registro pubblico*.

⁴Contrariamente agli stati uniti.

Leggi della privacy in Italia:

- Il DPD è stato implementato con il decreto legislativo 196/2003 (Codice in materia di protezione dei dati personali).
- Inoltre il Garante per la protezione dei dati limitati doveva applicare misure appropriate in determinati campi (video sorveglianza, dati biometrici, dati sanitari, notifiche di data breach, informazioni bancarie, profili online, processamenti fatti da amministratori di sistema, processamenti a fini di marketing e profiling, pagamenti elettronici, cookies).

Definizione 1.2.11: General Data Protection Regulation

La General Data Protection Regulation (GDPR) è un regolamento attraverso il quale l'unione europea intende aumentare la forza e unificare la protezione dei dati per tutti gli individui all'interno dell'unione europea.

Note:-

L'obiettivo principale del GDPR è quello di offrire un controllo semplificato agli appartenenti ai paesi membri dell'unione europea.

Definizione 1.2.12: Personal Data - GDPR

Uguale al DPD, ma aggiunge come caratteristiche di identificazione i dati di locazione, gli identificatori online, lo stato genetico, economico e culturale.

Definizione 1.2.13: Data Processing - GDPR

Uguale al DPD, ma viene inclusa la "strutturazione" dei dati processati.

Note:-

Inoltre il GDPR include altre definizioni.

Definizione 1.2.14: Pseudoanimizzazione

La pseudoanimizzazione è il processamento in cui i dati personali non possono più essere attribuiti univocamente a un soggetto senza informazioni aggiuntive. Le informazioni aggiuntive sono tenute separate e soggette a misure tecniche e organizzative per assicurare che i dati siano anonimi.

La pseudoanimizzazione:

- Se viene effettuata con politiche adeguate non è soggetto a controlli e penalità.
- La regolamentazione non coinvolge dati usati per statistiche o ricerche.
- Le politiche e le misure che raggiungono la privacy by Design e la privacy by Default sono adeguati.

Definizione 1.2.15: Personal Data Breach

Un leak accidentale nella sicurezza per cui il dato personale viene distrutto, perso, alterato, rubato o acceduto.

Diritti del soggetto:

- L'individuo deve dare un *consenso chiaro* per il trattamento dei propri dati.
- Il soggetto ha un *facile accesso* ai propri dati.
- Viene evidenziato il diritto alla rettifica, alla cancellazione e all'oblio⁵.

⁵Cosa non facile, soprattutto con i social.

- Il diritto di obiezione, incluso l'utilizzo dei propri dati per "profiling".
- Il diritto alla portabilità dei dati da un servizio a un altro.

Privacy *by Design* e *by Default*:

- Viene richiesto che la protezione dati sia presente fin dall'inizio nel sistema informativo.
- Le impostazioni della privacy devono essere di alto livello.
- La privacy deve essere presente per tutto il ciclo di vita del processamento dei dati.
- Come già detto i dati personali devono essere processati solo quando è necessario per ogni specifico scopo.

Definizione 1.2.16: Privacy by Design - Ann Cavoukian

Un approccio all'ingegneria dei sistemi che tiene in considerazione la privacy durante tutto il ciclo di vita. Si basa su 7 principi fondamentali:

1. Proattività, non reattività (prevenzione, non rimedio).
2. Privacy come impostazione di Default.
3. Privacy integrata nel Design.
4. Completamente funzionale.
5. End-to-end security.
6. Visibilità e trasparenza.
7. Rispetto per la privacy degli utenti.

Responsabilità:

- Il titolare del trattamento del dato deve dimostrare che tutte le operazioni messe in atto per garantire la sicurezza del dato siano aderenti alla legge.
- È responsabilità del controllore dei dati di implementare misure effettive e di dimostrare la *compliance* alle attività di processamento.
- L'utente deve essere chiaramente informato riguardo le finalità del trattamento, alla legge usata come base, all'intervallo temporale del trattamento, se i dati vengono trasferiti a terze parti e se avvengono delle decisioni automatizzate.
- Viene introdotto il *Data Protection Officer*, un esperto tecnico del dato e legale del GDPR.
- Se avvengono eventi di rischio bisogna fare un *Data Protection Impact Assessments* (DPIA) per valutarne l'estensione.
- La valutazione e la mitigazione del rischio sono necessarie e approvazione delle autorità nazionali di protezione dei dati (DPA) è necessario per rischi elevati.
- I records delle attività di trattamento devono essere conservate in modo includere le finalità del trattamento, le categorie coinvolte e Termini previsti.
- I records devono essere disponibili all'autorità di supervisione, su richiesta.

Definizione 1.2.17: Data Protection Officer

Il GDPR stabilisce la figura del Data Protection Officer (DPO), una persona con conoscenze specialistiche in materia di protezione dei dati e pratiche che dovrebbero aiutare il titolare del trattamento o l'incaricato del trattamento a monitorare conformità interna al presente regolamento. The DPO are also expected to be proficient at managing IT processes, data security (including dealing with cyber-attacks) and other critical business continuity issues around the holding and processing of personal and sensitive data.

Note:-

Autorità pubbliche e imprese le cui attività principali sono trattamento regolare o sistematico dei dati personali, sono necessario per assumere un DPO.

Definizione 1.2.18: Data Breach

Il controllore dei dati ha l'obbligo legale di notificare l'autorità di supervisione senza ritardi non necessari, a meno che sia improbabile che il data breach causi rischi alla libertà e ai diritti dell'individuo. C'è un massimo di 72 ore dopo aver scoperto il data breach per notificare.

Le seguenti sanzioni possono essere imposte:

- Nel primo e non intenzionale caso di noncompliance viene emesso un avviso.
- Una multa fino a 10 milioni di euro o fino al 2% dell'annualità mondiale fatturato dell'esercizio finanziario precedente nel caso di un'impresa, se vi è stata una violazione di alcuni obblighi.
- Una multa fino a 20 milioni di euro o fino al 4% dell'annuo in tutto il mondo fatturato dell'esercizio finanziario precedente nel caso di un'impresa, se vi sono state gravi violazioni dei principi.

1.2.5 Regolamenti Transazionali

- Safer Harbor privacy principles (fino al 2015).
- EU-US Privacy Shield (2016-2020⁶).
- EU-US Data Privacy Framework (2023-presente).

Definizione 1.2.19: Safe Harbor

Sviluppato tra il 1998 e il 2000 al fine di impedire alle organizzazioni private di l'UE o gli Stati Uniti dalla divulgazione accidentale o dalla perdita di informazioni personali.

Le compagnie USA che tengono dati devono aderire a questi 7 principi:

1. **Notifica:** gli individui devono essere informati che i loro dati stanno venendo collezionati e come verranno usati.
2. **Scelta:** gli individui possono scegliere di rimuovere i propri dati rinunciando ai servizi⁷.
3. **Trasferimento:** la trasmissione dei dati a terzi può avvenire solo ad altre organizzazioni che seguono principi di protezione dei dati adeguati.
4. **Sicurezza:** bisogna evitare la perdita di informazioni collezionate.
5. **Data Integrity:** i dati devono essere rilevanti e affidabili per lo scopo per cui sono raccolti.
6. **Accesso:** gli individui devono essere in grado di accedere, correggere e cancellare i propri dati.
7. **Applicazione:** devono esistere mezzi efficaci per far rispettare queste norme.

⁶Caduto a causa del primo governo Trump.

⁷Google merda.

Breve storia di Safe Harbor:

- Nel 2000 la Commissione europea ha deciso che i principi degli Stati Uniti erano conformi alla direttiva dell'UE (la cosiddetta "Decisione Safe Harbor").
- Dopo che un cliente si è lamentato del fatto che i suoi dati di Facebook non erano sufficientemente protetti dalla Corte di Giustizia della Commissione europea, nel 2015, dichiara la decisione di Safe Harbor invalida.
- La Commissione ha tenuto ulteriori colloqui con gli Stati Uniti autorità competenti verso "un quadro rinnovato e solido per flussi di dati transatlantici".
- La Commissione europea e gli Stati Uniti hanno convenuto di istituire un nuovo quadro per i flussi transatlantici di dati il 2 febbraio 2016, noto come "EU-US Privacy Shield".

Definizione 1.2.20: EU-US Privacy Shield

Lo scudo UE-USA per la privacy è un quadro per gli scambi transatlantici di dati per finalità commerciali tra l'Unione Europea e gli Stati Uniti. Uno dei suoi scopi è quello di consentire alle aziende statunitensi di ricevere più facilmente dati provenienti da entità dell'UE ai sensi delle leggi sulla privacy dell'UE volte a proteggere i cittadini dell'unione europea.

Osservazioni 1.2.4

- La Commissione europea ha adottato il quadro il 12 luglio 2016 ed è entrato in vigore lo stesso giorno.
- Il presidente degli Stati Uniti Donald Trump ha firmato un ordine esecutivo intitolato "Enhancing Public Safety", in cui si afferma che le protezioni della privacy degli Stati Uniti non saranno estese oltre i cittadini o residenti statunitensi.
- Nel luglio 2020 lo scudo UE-USA per la privacy è stato abrogato dalla Corte europea giustizia in quanto non forniva tutele adeguate ai cittadini dell'UE rispetto allo Snooping governativo americano^a.

^aIn poche parole gli americani si comportano da americani.

Definizione 1.2.21: AI Act

La legge sull'IA mira a garantire che i sistemi di IA utilizzati nell'UE siano sicuri, trasparenti e rispettosi dei diritti fondamentali.

L'AI Act classifica i sistemi di IA in base al livello di rischio che rappresentano:

- **Rischio inaccettabile**: i sistemi di IA che rappresentano una chiara minaccia per la sicurezza o i diritti fondamentali sono vietati. Ciò include i sistemi che implementano tecniche subliminali per manipolare il comportamento, sfruttare le vulnerabilità di gruppi specifici o abilitare il punteggio sociale da parte dei governi.
- **Alto rischio**: sistemi di intelligenza artificiale utilizzati in aree critiche come la sanità, l'istruzione, l'occupazione, il diritto, l'applicazione delle norme e i servizi essenziali sono soggetti a obblighi rigorosi. Questi includono rigorosi valutazioni dei rischi, misure di governance dei dati, supervisione umana e monitoraggio continuo per garantire la conformità.
- **Rischio limitato**: i sistemi di intelligenza artificiale a rischio limitato, come i chatbot e i deepfake, sono soggetti a obblighi di trasparenza. I fornitori e i distributori devono informare gli utenti che stanno interagendo con un sistema di intelligenza artificiale.
- **Rischio minimo**: la maggior parte dei sistemi di intelligenza artificiale, come i videogiochi abilitati all'intelligenza artificiale o i filtri antispam, cadono rientrano in questa categoria e sono in gran parte non regolamentate, incoraggiando l'innovazione e lo sviluppo,

2

Etica - Introduzione

2.1 Il Corso in Breve...

Il corso ha un aspetto prettamente *umanistico* e *interdisciplinare*. Si andranno ad affrontare diverse prospettive che impattano su diverse dimensioni della propria vita:

- Innovazione (brevetti).
- Economia (monopoli).
- Politica.
- Giuridica.
- Tecnologica.
- Mercato dell'attenzione.
- Retorica.

Definizione 2.1.1: Retorica

L'arte del dialogare per convincere le persone a fare quello che si vuole.

Note:-

Nei giornali, nella politica, nella pubblicità^a, etc.

^aPer questo esistono gli AdBlocker.

Esempio 2.1.1 (Retorica)

- La cimice asiatica sostituirà la cimice verde nei nostri prati.
- I robot sostituiranno gli esseri umani nei posti di lavoro.

Due frasi grammaticalmente identiche, ma non sono la stessa cosa: la prima ha un significato letterale, la seconda no. Perché la seconda non intende che i robot andranno a prendere e buttare via gli esseri umani dal posto di lavoro, ma che i padroni sostituiranno i lavoratori con dei robot. Questa è una frase strumentale, per nascondere il ruolo dell'essere umano.

Note:-

Qualsiasi affermazione di una società è puramente strumentale. Punta a manipolare le persone per ottenere un ritorno economico.

2.1.1 Perché Questo Corso?

Le tre missioni dell'università:

- Didattica.
- Ricerca.
- Terza missione: esportare le conoscenze tecnologiche alle aziende¹ e rendere coscenti le persone di quello di cui si occupa l'università (opportunità e rischi).

I messaggi del corso:

- La tecnologia e l'impatto che ha nella società sono costruzioni sociali, senza nessuna inevitabilità.
- La tecnologia non può risolvere i problemi che crea (e. g. i bias)².
- L'AI non è solo una tecnologia.
- L'essere umano ha un ruolo importante nel determinare l'impatto delle tecnologie digitali, sia nella propria professione sia come cittadini.



Figure 2.1: Uno dei libri possibili per l'esame.

Che cosa vuol dire che l'AI non è solo una tecnologia?

- In "Atlas of AI", Kate Crawford, propone un esperimento.
- Prendete Google Search (schifezza) e scrivete ARTIFICIAL INTELLIGENCE.
- I risultati sono immagini, principalmente di cervelli, con sfondo blu.

¹Che spreco...

²Fuck Silicon Valley and fuck Marc Andreessen.

- Cos'è che non viene mostrato?
 - L'inquinamento delle miniere di litio (con cui sono fatte le batterie).
 - Le proteste delle popolazioni la cui vita è stata distrutta dalle miniere di litio.
 - Le terre rare (metalli che servono per i microprocessori) che vengono estratte in paesi sottosviluppati, zone di guerra, da minori (spesso in Cina).
 - Le condizioni di schiavitù in cui lavorano centinaia di persone allo sviluppo dei dispositivi elettronici.
 - Gli enormi datacenter alimentati da fonti non rinnovabili.
 - Per allenare i sistemi di AI vengono usate foto su cui non si hanno i diritti.
 - The cleaners: la gente, sottopagata, che si occupa di pulire tutto lo schifo umano (decapitazioni, pedopornografia, violenza su donne, uomini, bambini, animali, etc.) in modo che i modelli di ML e i Social Media siano puliti.
 - L'ossessione per la produttività a cui sono soggetti i lavoratori (dal Taylorismo in avanti).
 - In alcuni paesi il riconoscimento facciale è usato per discriminare minoranze.
 - Le varie challenge su TikTok che procurano morti.
 - Molly Russell: ragazzina di 14 anni che si è suicidata dopo che Instagram e YouTube, con algoritmi AI di personalizzazione, le hanno mostrato più di 2000 contenuti di istigazione al suicidio e all'autolesionismo.
 - Facebook è stato accusato di genocidio in Etiopia, Bangladesh e Myanmar per persecuzioni di minoranze.
 - Assalto a Capitol Hill da parte dei sostenitori di Trump, anche a causa dei Social Media.

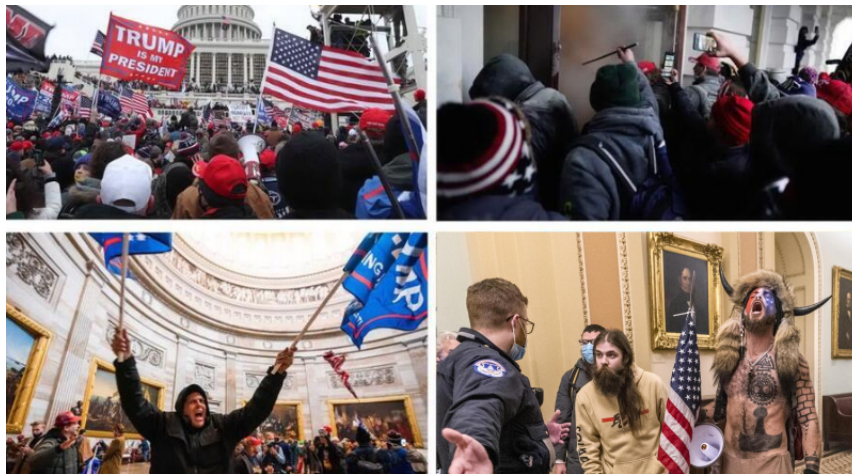


Figure 2.2: Assalto a Capitol Hill (6 Gennaio 2021).

- QAnon nasce su 4chan: teoria secondo la quale il "deep state" sia composto da pedofili che rubano i bambini per estrarne il loro sangue e produrre una droga contro l'invecchiamento.

Sir our plan to replace the entire government with 4chan freaks has run into a few snags

Figure 2.3: Elon moment.

- "You have blood on your hands" disse il senatore Graham a Mark Zuckerberg, "You have a product that's killing people."

Social Network vs. Social Media:

- Un Social Network serve per mettere in connessione le persone.
- Un Social Media serve per distribuire contenuti.
- Diversi Social Network sono diventati Social Media (e. g. FaceBook).

Domanda 2.1

Perché si è andati in questa direzione?

Risposta: ultimamente si mostrano i contenuti più virali, ritenuti interessanti al singolo, per creare dipendenza e aumentare gli introiti dovuti alla pubblicità. "If you're not paying for it, you're not the customer; you're the product being sold".

2.2 La Costruzione della Realtà Sociale

"La costruzione della realtà sociale" è un libro scritto da John Searle in cui sostiene che la realtà sociale sia frutto di una costruzione³. Searle racconta il seguente aneddoto: io, John Searle, cittadino americano, mi trovo in un bistrot lungo la Senna che mi bevo una birra⁴. Ma com'è possibile che ciò sia avvenuto? Sono un cittadino americano, sono salito su un aereo, sono arrivato a Parigi, sono entrato nel bistrot che vende birra, parlo con il cameriere e gli chiedo una birra, ovviamente pagando. Sembra tutto semplice, ma dietro a ogni punto si nasconde una rete concettuale che le persone danno per scontata e accettano. Perché mi fanno salire sull'aereo? Perché sono un cittadino americano e ai cittadini americani è permesso fare turismo in europa. E come provo che sono cittadino americano? Esibisco il passaporto. Come possibile che il bistrot venda birra? Ci vuole la licenza. E com'è possibile che il cameriere sia lì, mi porti la birra e si aspetti di essere pagato? E com'è che vengono accettati pezzi di metallo il cui valore corrispettivo è quello della birra consumata? Dietro tutto questo c'è una *realtà sociale* fatta di norme, ma anche di *consuetudini* che rendono possibili le interazioni e che collegano *fatti bruti* (il passaporto, le monete, etc.) con un *livello astratto* (regole costitutive).

Al giorno d'oggi la tecnologia e, come esempio estremo, il metaverso, va a influenzare e modificare la realtà sociale. Inoltre la realtà sociale è costituita da parole e attualmente si hanno i LLM che "parlano" e possono partecipare alla realtà sociale.

2.2.1 Come l'AI Influenza la Realtà Sociale

Nel 1989, durante la rivolta studentesca di piazza Tiananment, un ragazzo si mise davanti a una colonna di carriarmati e venne portato via. Non sapendo il suo nome viene soprannominato "Tank Man". Poco fa è diventata famosa una "foto" generata dall'AI di Tank Man, quest'immagine fu, per un po' di tempo, indicizzata da Google. Google la rimosse, ma dato che è stato fatto un articolo di debunking è attualmente indicizzata (con il link al relativo articolo in cui viene spiegato in dettaglio il fatto che l'immagine è fake).

Note:-

Oltre a questo ci sono stati altri casi, uno recente è Trump Gaza: come Trump si reimmagina la striscia di Gaza come resort. È questo contenuto è stato condiviso dallo stesso Trump sulla sua piattaforma Truth.

Domanda 2.2

Ma la disinformazione c'è sempre stata. Che cambia tra Tank Man, Trump Gaza e altre fake news non AI-Generated?

Risposta: sebbene la disinformazione ci sia sempre stata, al giorno d'oggi è molto più facile e veloce grazie all'AI. Una volta la disinformazione era solo su giornali e poteva portare conseguenze ai giornalisti, ora le regole per il materiale AI-Generated sono molto vaghe.

³Curioso visto quello di cui viene accusato, anche il consenso è una costruzione sociale?

⁴Average americano

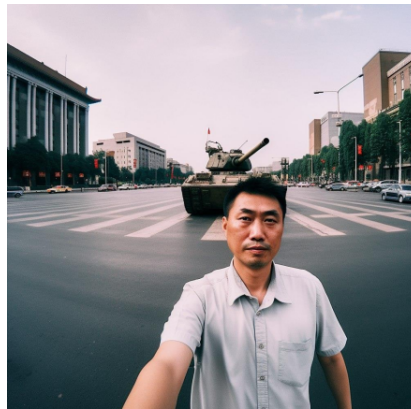


Figure 2.4: Immagine AI-Generated di Tank Man.

Domanda 2.3

I social media sono una piazza pubblica?

Risposta: è un tema molto complesso. Da una parte in una piazza pubblica se si urla qualcosa finisce lì, mentre sui social media *potenzialmente tutti* possono accedere e vedere/sentire. Quanto conta l'algoritmo di personalizzazione in tutto questo? Per esempio Gonzales vs. Google in cui i genitori di una ragazzina morta negli attentati terroristici in Francia hanno fatto causa a Google perché i terroristi si erano radicalizzati con contenuti di YouTube offerti loro dall'algoritmo di personalizzazione. Tuttavia persero la causa.

2.2.2 L'AI Rappresenta Effettivamente il Mondo?

Se si utilizza un generatore di immagine AI per creare foto di giudici, ingegneri, medici, politici e CEO si vedrà una grande maggioranza di maschi bianchi. Al contrario se gli si chiedono immagini di lavapiatti, lavoratori di fastfood e bidelli si troveranno persone di pelle scura. Se gli si chiedono insegnanti, cassieri, lavoratori dei servizi sociali e domestici si troveranno più donne. Ma è così la realtà? Sì e no, in realtà le proporzioni sono più bilanciate di quelle proposte dall'AI generativa che funge da "cassa di risonanza".



Figure 2.5: Esempio.

Domanda 2.4

C'è una soluzione al problema?

La mossa di Google: con Gemini ha utilizzato un dataset "politically correct", però quando gli si chiedeva di rappresentare un soldato tedesco nella WWII sputava fuori persone di colore, asiatiche e donne. Con l'immagine di un papa uscivano papi neri (tecnicamente possibili) e papesse (beh, è una carta dei tarocchi). I padri fondatori, in un'epoca di fervente schiavismo, diventavano di colore. Infine i fondatori di Google stessa diventavano asiatici.

La soluzione di Google: semplice.

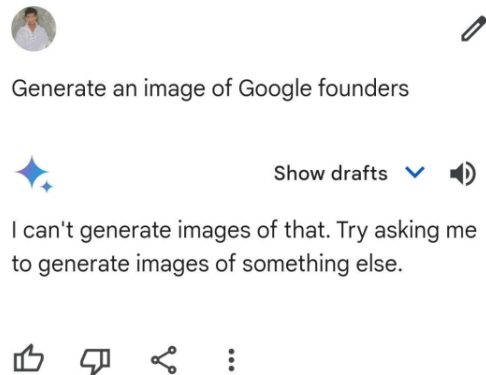


Figure 2.6: Non posso onii-chan :3

Osservazioni 2.2.1

Brevi osservazioni sulla censura nei vari modelli:

- Deepseek: se parli di Taiwan o di quello che è successo a piazza Tiananment.
- ChatGPT: se parli di Noam Chomsky (attivista anarchico e sostenitore di Pol Pot).
- Grook: se parli di fake news in relazioni a Trump e Musk.

Note:-

Queste soluzioni sono solo tecniche, ma il problema è politico.

2.2.3 I Pericoli dei Language Models

Un anno prima dell'uscita di ChatGPT furono licenziate da Google Margaret Mitchell e Timnit Gebru. Esse misero in guardia sul pericolo dei pappagalli stocastici: i language models non sanno quello che dicono, ma semplicemente ripetono delle parole che non comprendono.

Unfathomable Training Data:

- La grandezza del dataset non garantisce la diversità: gran parte del web è scritto in inglese, da uomini bianchi.
- Poiché passano mesi o anni dall'inizio dell'apprendimento il dataset è già vecchio. Per esempio la prima release di ChatGPT non era a conoscenza della guerra in Ucraina. Anche dal punto di vista sociale sono lo specchio di una società più vecchia.
- Encoding Bias.

Le relazioni parasociali: questi language models possono influenzare gli esseri umani nel dialogo reciproco. Si attribuisce a questi sistemi più di quello che realmente c'è. Nel 1966 fu creata Eliza, uno dei primi chatbots, che interpretava uno psicoterapeuta che faceva domande con un particolare stile. Un aneddoto famoso riguarda una segretaria che si convinse di parlare realmente con uno psicoterapeuta (pur sapendo che cos'era Eliza).



Figure 2.7: Eliza.

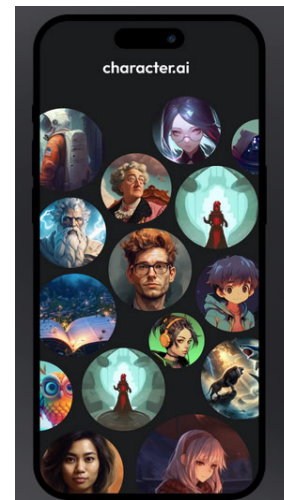


Figure 2.8: La moderna versione di Eliza.

Definizione 2.2.1: Allucinazioni

Se il chatbot non sa qualcosa se lo inventa con un sistema probabilistico. Se le probabilità sono labili viene restituito in output qualcosa che sembra coerente, ma non ha senso (allucinazioni).

La missione di OpenAI: sviluppare la Artificial General Intelligence (AGI) per sostituire l'essere umano in qualsiasi compito che abbia *valore economico*.

Il problema con la gente contraffatta: da quando hanno inventato il denaro chiunque lo contraffae è punito con le pene più severe. Il problema si riconduce alla realtà sociale di Searle e che contraffarre il denaro è un crimine perché fa venire meno la fiducia nel denaro, minando la società stessa. Se vale l'analogia tra denaro contraffatto e umano contraffatto (chatbot che si spaccia per un essere umano) allora si è di fronte al più grande crimine della storia, interagendo con persone finte. Da questo deriva anche il problema del deepfake pornografico AI-Generated.

Altri problemi:

- Le guide generate dall'AI: sempre più libri guida vengono generati dall'intelligenza artificiale. Nel migliore dei casi sono libri turistici inaccurati, ma può capitare di imbattersi in libri più pericolosi (e.g. guide per distinguere i funghi velenosi). Questo fenomeno ha dato origine al termine *sloop internet* per indicare i rifiuti prodotti dall'AI generativa.
- La disinformazione causata dall'AI è personalizzata all'utente. Ciò può fungere da gigantesca echo chamber per convincere determinate persone o per dissuaderne altre.
- Gli scam online: gente a cui telefonano dei "parenti" la cui voce è replicata dall'AI (e.g. Jennifer DeStefano).

3

Ancora sull'AI

3.1 AI or not AI This is the Dilemma

3.1.1 Il Watermark

Definizione 3.1.1: Watermark

Il watermark consiste in cambiamenti non visibili in alcuni pixel chiavi che possono essere individuati dalla macchina.

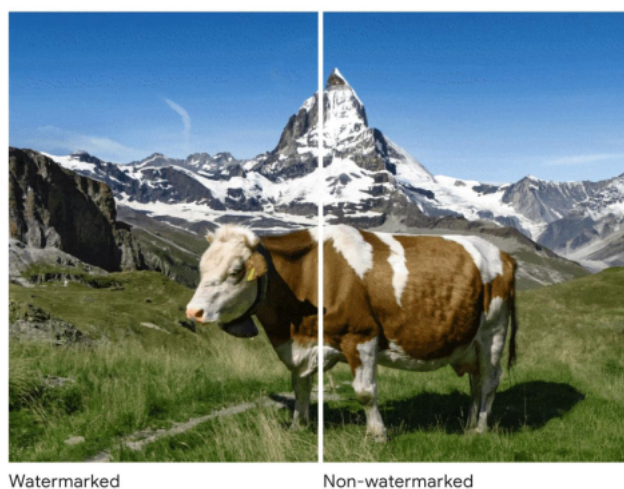


Figure 3.1: Immagine con watermark.

Note:-

Uno dei provvedimenti dell'amministrazione Biden per "difendersi" da watermark è stato quello di dire: inseriamo del watermark che indichi se un'immagine è AI-generated.

Domanda 3.1

Questa è una soluzione ragionevole?

- Se è possibile introdurlo allora esisterà un AI in grado di rimuoverlo.

- Questa proposta garantisce Microsoft, openAI, Google che se ne lavano le mani. Però le aziende non americane non hanno alcun obbligo di applicare il watermark.
- Inoltre il fatto che una foto abbia il watermark, in un'epoca di cospirazionismo, non significa nulla¹.

3.1.2 Il Copyright

Tutti i vari modelli di AI generativa sono stati addestrati su tutti i dati del mondo, principalmente il web (anche su materiale soggetto a copyright). Per esempio alcuni siti come z-library o il web archive² in cui sono presenti contenuti piratati. I vari LLM hanno probabilmente assorbito anche quei contenuti. Ovviamente quella merda che è Google ha detto: eh, ma se è lì sul web allora lo posso prendere, se non lo volevano non lo dovevano pubblicare (a parte che è l'esatto opposto del meccanismo di copyright, ma forse non si rendono conto che è la stessa argomentazione che utilizzano gli stupratori nei confronti della loro vittima). Oppure openAI che chiede alle persone di indicare quali pagine non vogliono sia preso dal crawler. Tuttavia questo non funziona, per esempio con i siti di mirroring.



Figure 3.2: Protestante con cartello.

¹God, how much I hate humanity.

²Entrambi consigliatissimi, da visitare almeno una volta nella vita, sono la nuova biblioteca di Alessandria.

