

CODICE ETICO

per la creazione di valore basato sui dati



RACCOMANDAZIONI

SOMMARIO

Il «Codice Etico per la creazione di valore basato sui dati» è costituito dai seguenti documenti: 1) panoramica; 2) nozioni di base; 3) raccomandazioni; 4) implementazione; 5) contesto.

Il documento **RACCOMANDAZIONI** fornisce raccomandazioni concrete per la creazione di valore etico basato sui dati, strutturate a partire dalle quattro fasi del ciclo di vita dei dati. Le raccomandazioni si basano sui tre orientamenti etici di base e sui tre valori procedurali, fornendo sei gruppi di raccomandazioni per ogni fase. Per ogni gruppo vengono spiegate per prime le intenzioni delle raccomandazioni, seguite poi dalle raccomandazioni concrete. Ogni gruppo termina con un esempio che illustra come possono essere implementate le raccomandazioni. Le spiegazioni del ciclo di vita dei dati e degli orientamenti di base sono disponibili nel documento Nozioni di base. Il documento Raccomandazioni è disponibile in tedesco, inglese, francese e italiano.

COPYRIGHT

Il «Codice Etico per la creazione di valore basato sui dati» è stato redatto dal gruppo di esperti «Data Ethics» di Swiss Alliance for Data-Intensive Services. Team editoriale: Markus Christen, Christoph Heitz, Tom Kleiber, Michele Loi (caporedattore). Traduzione francese: Jean-Gabriel Piguet. Grafica: Ana Nicolasa Caduff. Edizione 2020.

© Swiss Alliance for Data-Intensive Services, 2020.

ISBN 978-3-9522703-2-5; www.data-service-alliance.ch/codex

Licenza: Attribuzione 4.0 Internazionale (CC BY 4.0).

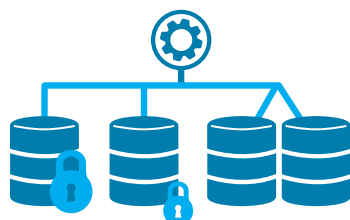


FASE 1

Generazione e acquisizione dei dati

Pag. 4

Orientamenti etici di base
Valori procedurali



FASE 2

Archiviazione e gestione dei dati

Pag. 10

Orientamenti etici di base
Valori procedurali



FASE 3

Analisi dei dati e generazione di conoscenze

Pag. 16

Orientamenti etici di base
Valori procedurali



FASE 4

Prodotti e servizi

Pag. 22

Orientamenti etici di base
Valori procedurali

FASE 1: GENERAZIONE E ACQUISIZIONE DEI DATI

ORIENTAMENTI ETICI DI BASE

PREVENZIONE DEL DANNO

La prevenzione del danno esige assicurarsi che il modo in cui vengono raccolti i dati non causi alcun danno all'utente. Possono sorgere domande su possibili danni se l'interazione con l'utente è progettata per incoraggiare l'utente a rivelare più dati su se stesso (tali progetti possono utilizzare i risultati di ricerche comportamentali e possono essere noti come «nudges», «spinte leggere»). Se un'interfaccia utente è deliberatamente progettata per costringere l'utente a rivelare informazioni che sono contrarie ai suoi interessi, si parla di «dark pattern».



Raccomandazioni

- Evitare di sfruttare l'inerzia, la disattenzione o altre debolezze psicologiche degli individui per ottenere da loro più informazioni personali di quante siano necessarie per fornire il servizio.
- Controllare il web design e user experience per verificare la presenza di dark pattern e rimuoverli.
- Se si utilizzano i nudges, è necessario dimostrare in anticipo o documentare successivamente che essi servono anche gli interessi dell'utente (p.es., aumentando la sicurezza informatica o migliorando il servizio fornito).

ESEMPIO - EVITARE I «DARK PATTERNS»

I «dark patterns» vengono utilizzati per fuorviare l'utente e - nel peggiore dei casi - indurlo a compiere un'azione che non voleva. Per innescare questo «comportamento scorretto», vengono sfruttati i modelli di comportamento appresi dagli utenti: gli utenti vengono manipolati da pulsanti posizionati in modo intelligente o da menu a tendina ingannevoli. Ad esempio, una società vuole vendere una polizza di assicurazione di viaggio a un cliente che ha appena prenotato un viaggio. L'opzione «Nessuna assicurazione» è nascosta (sotto la «N») in un elenco alfabetico dei paesi a cui l'assicurazione potrebbe essere applicata. Le interfacce dovrebbero invece supportare l'usabilità organizzando le funzionalità interattive in modo trasparente e secondo una logica. Nel caso dell'assicurazione di viaggio, la decisione iniziale («sì» contro «no») dovrebbe essere separata dalle possibili opzioni che un «sì» implica.



GIUSTIZIA

Le raccomandazioni basate sulla giustizia hanno lo scopo di garantire la reciprocità, in modo che, al momento della raccolta dei dati, ai soggetti interessati vengano offerti vantaggi proporzionati ai dati richiesti.

Raccomandazioni

- Rendere chiaro ai propri clienti il valore restituito per la divulgazione dei loro dati (ad es. un trattamento più personalizzato).
- Qualora la raccolta di determinati dati sia richiesta per legge, occorre chiarirlo.



ESEMPIO - DOSSIER SANITARIO ELETTRONICO (DSE)

Un medico che visita in privato si è iscritto al sistema DSE in modo da migliorare la qualità della cura per i propri pazienti. Una volta che il paziente ha creato un DSE individuale, il medico può inserire i suoi dati sanitari e di trattamento. Tuttavia, il paziente U. non è più sicuro di voler aprire il DSE ed è preoccupato che i suoi dati possano essere utilizzati in modo improprio. Con riferimento all'orientamento di base della giustizia, il medico può convincere il paziente U. come segue:

- mostra al paziente il beneficio individuale: con il DSE, l'accesso ai documenti rilevanti per il trattamento è garantito in ogni momento, e si possono evitare trattamenti doppi o non necessari.
- il medico chiarisce al paziente che può determinare chi ha accesso ai dati sanitari e nello specifico a quali in particolare. Solo in caso di emergenza è possibile accedere a DSE senza il consenso del paziente. La memorizzazione dei dati è decentralizzata e vengono utilizzati solo server svizzeri. Lo scambio di informazioni avviene solo attraverso un collegamento sicuro, con tutti i processi appositamente certificati.
- fa notare che i pazienti possono mettere facilmente a disposizione i loro dati sanitari per la ricerca, se lo desiderano.

AUTONOMIA

Orientarsi verso l'autonomia significa dare agli utenti una sufficiente autodeterminazione informativa.

Raccomandazioni

- Fornire ai propri clienti strumenti altamente utilizzabili per prendere decisioni consensuali sull'utilizzo dei dati.
- A tal fine, sviluppate forme di consenso informato intuitivamente plausibili e, se necessario, selezionabili dall'utente.
- Consentire agli utenti di adattare o revocare il loro consenso in un secondo momento.
- Quando le circostanze lo permettono, consentire preferenze granulari in merito allo scopo del trattamento dei dati (ad es. evitare di raggruppare troppi scopi simili in un unico consenso).



ESEMPIO - DIVERSI LIVELLI DI CONSENSO

Una piattaforma web fornisce informazioni sulla ricerca medica su patologie specifiche, opportunità di social networking e sui servizi di auto-tracking a gruppi di pazienti. Raccoglie, analizza e commercializza i dati dei suoi utenti, ma li utilizza anche per le sue raccomandazioni personalizzate, che sono fondamentali per la funzionalità e l'attrattiva della piattaforma. Il modello di business della piattaforma è quello di rendere disponibili alle aziende farmaceutiche sia i dati anonimi che quelli personali dei pazienti, sia per il reclutamento di pazienti, sia per la sperimentazione di nuovi farmaci. Inoltre l'azienda collabora sia con l'industria, sia con i ricercatori universitari per la ricerca comportamentale (ad esempio, sugli atteggiamenti nei confronti delle terapie e dei loro effetti collaterali). A tal fine:

- l'azienda fornisce diversi livelli di consenso: i pazienti possono acconsentire all'utilizzo dei loro dati per ottenere raccomandazioni senza che i dati vengano condivisi con l'industria.
- gli utenti possono rivedere il loro consenso e condividere le loro preferenze, in qualsiasi momento. Queste informazioni sono facilmente raggiungibili dal cruscotto del paziente.
- gli utenti possono decidere di interrompere la condivisione dei dati con l'industria in qualsiasi momento - mentre i dati precedentemente condivisi in forma anonima non saranno ritirati, l'accesso ai dati personali sarà bloccato nel momento in cui il consenso sarà revocato elettronicamente.



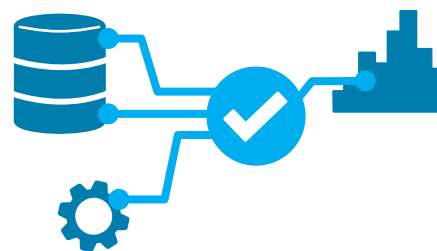
VALORI PROCEDURALI

CONTROLLO

Le raccomandazioni relative al controllo hanno lo scopo di garantire che, quando i dati vengono generati o acquisiti, vengano raccolte tutte le informazioni necessarie per la valutazione etica dell'uso dei dati.

Raccomandazioni

- Documentare le proprie fonti di dati ed eventuali restrizioni sull'uso appropriato dei dati.
- Qualora altre società utilizzino le informazioni, fornire loro le informazioni necessarie per valutare le origini dei dati e le restrizioni d'uso.



ESEMPIO - «OPEN DATA» PER NUOVI MODELLI DI BUSINESS

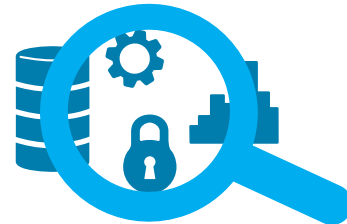
Un grande fornitore di servizi di trasporto decide di mettere a disposizione di terzi i dati raccolti per un uso secondario gratuito in linguaggio macchina (Open Data). L'obiettivo è quello di sostenere un mercato per operatori emergenti di informazioni e servizi. Si spera che in questo modo si promuovano nuovi modelli di business basati sui dati aumentando anche l'attrattiva dell'ambiente di mercato dell'azienda. L'azienda definisce le condizioni di utilizzo, che possono includere l'obbligo per l'utente di indicare l'origine dei dati, di aggiornare i dati grezzi quando l'azienda notifica nuovi dati (per evitare che i modelli di business si basino su dati obsoleti) e di rispettare le norme sulla protezione dei dati (in particolare, di astenersi da qualsiasi tentativo di estrarre dai dati informazioni non anonime sui clienti dell'azienda).

TRASPARENZA

La trasparenza esige che venga garantito che il consenso informato possa essere ottenuto nel modo più semplice, chiaro e accessibile possibile.

Raccomandazioni

- Trovare dei modi semplici per spiegare quali dati sono raccolti, la politica sulla privacy applicata e come vengono utilizzati i dati dei clienti (cioè gli scopi e i risultati attesi).
- Utilizzare i riassunti e altri modi per spiegare i punti chiave dei propri «termini legali e condizioni».
- Collaborare con altre aziende per creare nuovi standard per rendere termini legali e condizioni più leggibili e comprensibili.



ESEMPIO - FOTOGRAFO

Un fotografo offre servizi fotografici per eventi speciali e servizi fotografici personali, oltre a varie opzioni per l'acquisto delle immagini risultanti. Le sue condizioni generali di contratto (CGC) prevedono che le foto possano essere messe a disposizione dei clienti in formato digitale tramite un fornitore di servizi online esterno, in modo che i clienti possano scaricarle. Il fotografo può conservare le foto per un anno dopo l'evasione dell'ordine, in modo che i clienti possano riordinare le stampe e ha anche il diritto di utilizzare le foto per scopi pubblicitari (su un sito web specifico, su manifesti, ecc.). Il copyright rimane di proprietà del fotografo. Il cliente desidera ricevere informazioni sulle condizioni generali di contratto. Il fotografo può rispondere a questa richiesta come segue:

- il fotografo consegna le CGC al cliente prima della conclusione del contratto. Egli riassume gli scopi per i quali i dati saranno utilizzati e si impegna a non utilizzare i dati per altri scopi.
- il fotografo informa il cliente che può cancellare le fotografie dopo la conclusione dell'ordine, ma che il cliente dovrà ricaricarle se desidera effettuare un altro ordine di stampe.
- egli chiede esplicitamente al cliente se può utilizzare le fotografie per i propri scopi pubblicitari.
- infine, il fotografo si assicura che il fornitore esterno di servizi online si attenga alle stesse regole per quanto riguarda il trattamento dei dati (in particolare, non utilizzare i dati per altri scopi) e ne informa il cliente.



RESPONSABILITÀ

Le raccomandazioni in materia di responsabilità sono state concepite per garantire l'utilizzo dei dati raccolti da terzi in modo responsabile.

Raccomandazioni

- Qualora il prodotto dipendesse da dati di altre aziende, assicurarsi che questi dati siano stati raccolti in modo etico.
- Non utilizzare dati provenienti da fonti non attendibili.
- Acquisire dati solo da partner che siano trasparenti sia sulle loro pratiche di raccolta dei dati che sulle eventuali restrizioni associate all'uso dei dati.



ESEMPIO - MONETIZZAZIONE DEI DATI DEI SOCIAL NETWORK

Un'azienda farmaceutica prende in considerazione l'acquisto di dati da una piattaforma web che fornisce opportunità di social networking e di auto-tracciamento a gruppi di pazienti. L'azienda farmaceutica ottiene informazioni sulle pratiche di raccolta dati della piattaforma web. I dati ottenuti sono ottenuti in modo tale che lo scopo della raccolta dati e il collegamento con l'industria, compresa la monetizzazione dei dati, siano comunicati all'utente in modo trasparente e aperto? In caso contrario, l'azienda farmaceutica non acquisisce dati da questa piattaforma e non utilizza i suoi servizi.

FASE 2: ARCHIVIAZIONE E GESTIONE DEI DATI

ORIENTAMENTI ETICI DI BASE

PREVENZIONE DEL DANNO

Le raccomandazioni per la prevenzione del danno, in questa fase, hanno lo scopo di contribuire a garantire che nessun danno possa essere provocato con i dati memorizzati.

Raccomandazioni

- Implementare adeguate misure di sicurezza informatica e applicarle in modo coerente in tutte le forme di archiviazione dei dati (ad esempio basato su cloud, hybrid).
- Sviluppare piani di emergenza per ridurre al minimo l'entità dei danni in caso di violazione dei dati.
- Definire le regole per la cancellazione dei dati con l'obiettivo di ridurre il rischio di danni dovuti ad accessi non autorizzati.



ESEMPIO - DATI SANITARI NEL CLOUD

Un'azienda di analisi dei dati sanitari sviluppa strumenti di raccolta e analisi dei dati personalizzati per i grandi ospedali. L'azienda fornisce agli ospedali un servizio cloud per ospitare tutti i dati, consentendo agli ospedali di concentrarsi sulla cura del paziente e sul miglioramento dei risultati. Al fine di garantire all'ospedale la necessaria sicurezza informatica, l'azienda di analisi dei dati implementa i più elevati standard di sicurezza informatica e si impegna a rispettare rigorosamente tali standard (ad esempio, i dati degli ospedali sono accessibili solo a dipendenti selezionati della società di software che utilizzano l'autenticazione a due fattori, e questi dipendenti sono addestrati a riconoscere gli attacchi di social engineering).



GIUSTIZIA

La giustizia esige che sia reso possibile per clienti utilizzare i propri dati a beneficio di terzi e del bene comune.

Raccomandazioni

- Supportare i clienti nell'utilizzo efficace del diritto alla portabilità dei dati.
- Consentire ai clienti di ricevere e utilizzare facilmente i propri dati (a condizione che non siano stati resi anonimi) per applicazioni di terze parti.



ESEMPIO - DONAZIONE DI DATI

Un'azienda che raccoglie ed elabora i dati personali dei clienti per fornire servizi vuole facilitare ai propri clienti la donazione dei propri dati (se lo desiderano) per scopi moralmente rilevanti (ad esempio, progetti di ricerca senza scopo di lucro o progetti della pubblica amministrazione). Con il riutilizzo dei dati, l'azienda vuole anche contribuire alla sostenibilità digitale. In questo caso potrebbero essere utili le seguenti misure:

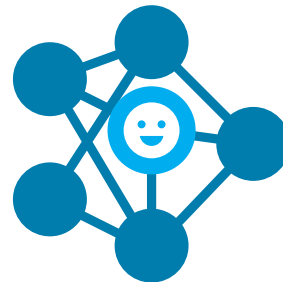
- Creazione di un sistema di gestione delle informazioni sulla piattaforma dell'azienda che consenta ai suoi clienti di gestire autonomamente i loro dati personali. Ciò comprende, in particolare, un'opzione user-friendly per scaricare tutti i dati o alcuni dati selezionati in linguaggio macchina ed esportarli in altri sistemi.
- La creazione di un'interfaccia tecnica con altri sistemi di gestione delle informazioni e piattaforme di deposito a garanzia dei dati, in modo che i clienti dell'azienda possano mettere i loro dati personali direttamente a disposizione di altri fornitori di servizi.

AUTONOMIA

Le raccomandazioni relative all'autonomia hanno lo scopo di garantire che l'archiviazione e la gestione dei dati non comportino alcun rischio per la privacy dei clienti.

Raccomandazioni

- Esaminare se le misure tecniche per proteggere la privacy dei clienti sono appropriate e proporzionate.
- Tenersi aggiornati sugli sviluppi delle tecnologie di sicurezza pertinenti.
- Assicurarsi che i dati memorizzati in forma anonima non possano essere ricondotti alle persone a cui si riferiscono.
- Offrire ai clienti opzioni semplici per limitare la memorizzazione dei loro dati.



ESEMPIO - VISUALIZZAZIONE DELL'AUTODETERMINAZIONE DIGITALE

Un'azienda ha messo l'autodeterminazione digitale dei suoi utenti al centro delle sue attività commerciali. Essa fornisce un sistema di gestione delle informazioni personali e un mercato per i dati personali. La piattaforma consente ai suoi utenti di mettere a disposizione di singole aziende, amministrazioni pubbliche o organizzazioni senza scopo di lucro, ad un certo prezzo o gratuitamente, il loro profilo di dati o estratti selezionati di essi; l'azienda ottiene una quota del prezzo pagato. Gli utenti hanno sempre una visione d'insieme di quali dati vanno a quali destinatari a quale prezzo e possono anche annullare queste offerte. Il valore aggiunto della piattaforma per l'utente consiste nel rafforzamento della sua autodeterminazione digitale attraverso i seguenti elementi:

- i dati degli utenti, come i dati di acquisto, i dati di mobilità o i dati sanitari, vengono aggregati e visualizzati. Ad esempio, gli utenti possono vedere quanto sono stati salutari o sostenibili i loro acquisti o come è stato il loro comportamento di mobilità (ad esempio, il numero di passi) negli ultimi mesi.
- i dati sono memorizzati in forma criptata e non possono essere visualizzati dall'azienda stessa.
- se un cliente vuole trasmettere a terzi parti del suo set di dati, viene effettuato un controllo automatico per determinare se il set di dati così definito può essere ricondotto al cliente. Se il rischio è elevato, il cliente viene messo in guardia contro la divulgazione e deve acconsentire esplicitamente.



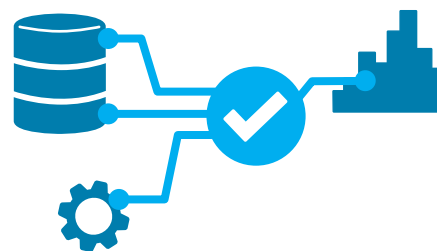
VALORI PROCEDURALI

CONTROLLO

Le raccomandazioni relative al controllo hanno lo scopo di garantire che nessuna informazione rilevante per la valutazione dell'uso etico dei dati venga persa nella gestione dei dati (p.es., la pulizia dei dati, la cancellazione dei dati, ecc.).

Raccomandazioni

- Documentare eventuali modifiche e adeguamenti ai requisiti di gestione dei dati durante lo spostamento dei dati in posizioni diverse, soprattutto quando i servizi sono esternalizzati nel cloud.
- Implementare i requisiti tecnici per garantire la tracciabilità delle modifiche ai dati memorizzati.
- Considerare in anticipo cosa fare con i dati degli account inattivi o dei clienti precedenti.



ESEMPIO - TRACCIABILITÀ DELL'UTILIZZO DEI DATI

Un'azienda utilizza i dati socio-demografici relativi ai propri clienti, nonché i dati derivanti dall'interazione con i clienti (comportamento durante l'ordine, scambi email, ecc.) per classificare i propri clienti in diversi gruppi. Fa ciò mediante modelli analitici dei dati e per prevedere il comportamento futuro dei clienti. Questi modelli sono utilizzati nella gestione delle relazioni con i clienti. Inoltre, alcuni risultati delle previsioni basate su modelli sono condivisi con altre società dello stesso gruppo. Questi dati sono soggetti a continui cambiamenti. Per poter tracciare in qualsiasi momento quali dati sono stati utilizzati per l'ulteriore elaborazione dei modelli di previsione e di analisi, ogni set di dati è arricchito con metadati che descrivono da dove provengono i dati, quando i dati sono stati trasmessi agli analisti o ai model builder, quali restrizioni d'uso sono legate ai dati, ecc. In questo modo è possibile tracciare ogni singolo dataset che è stato utilizzato per l'analisi o la modellistica, assicurando al contempo che gli analisti e i modellisti utilizzino i dati in modo appropriato.

TRASPARENZA

La trasparenza esige che i clienti siano informati su tutti gli aspetti del trattamento dei dati, della memorizzazione, della gestione e della protezione, creando così fiducia (ben riposta).

Raccomandazioni

- Far capire ai clienti come i loro dati sono protetti e come viene registrato l'accesso.
- Se è necessario memorizzare i dati per un tempo molto lungo (p.es., per motivi legali), occorre spiegarlo ai clienti nel momento della raccolta dei dati.



ESEMPIO - INFORMAZIONI SULLA CONSERVAZIONE DEI DATI

La legge federale sulla sorveglianza della posta e delle telecomunicazioni obbliga gli operatori di telecomunicazione in Svizzera a conservare e a mantenere per sei mesi i metadati relativi alle telecomunicazioni dei loro clienti. Ciò ha lo scopo di facilitare il chiarimento dei dati penali e la ricerca delle persone scomparse. Tutti i cittadini svizzeri hanno inoltre il diritto di accedere ai dati conservati. Un fornitore di servizi di telecomunicazione richiama l'attenzione su questo fatto in modo adeguato e dettagliato, indicando quali dati sono conservati, tra l'altro:

- chi è stato su quale sito web e quando e quali applicazioni sono state utilizzate;
- con chi, quando e da dove è avvenuta la comunicazione;
- chi si è connesso a Internet, quando e per quanto tempo;
- Chi ha inviato a chi un'e-mail o un altro messaggio di testo e quando;
- Dove si trova attualmente un utente di telefonia mobile e la cronologia della posizione del telefono negli ultimi sei mesi.



RESPONSABILITÀ

Le raccomandazioni in materia di responsabilità hanno lo scopo di garantire che i ruoli di tutte le persone coinvolte nella gestione dei dati siano stati sufficientemente chiariti.

Raccomandazioni

- Definire a livello organizzativo chi è coinvolto nella valutazione, nel reporting e nell'implementazione degli standard etici nella gestione dei dati.
- Definire regole chiare in merito all'accesso ai dati memorizzati (chi può avere accesso, quando, in quali condizioni e come l'accesso viene registrato o tracciato).



ESEMPIO - IL RUOLO DEL RESPONSABILE DEL TRATTAMENTO DEI DATI

In molte organizzazioni il ruolo del Chief Data Officer (CDO) è ormai consolidato. Il CDO sorveglia l'uso dei dati per garantire che un'organizzazione tragga il massimo beneficio dai suoi dati mantenendo i propri standard. Questi standard stabiliscono tipicamente l'intero processo di gestione dei dati e quindi garantiscono la qualità dei dati. Sia la gestione che la qualità dei dati sono oggetto della presente raccomandazione. Sebbene la carica di Responsabile del trattamento dei dati sia solitamente istituita per motivi di business piuttosto che per motivi etici, è comunque uno dei doveri del Responsabile del trattamento dei dati stabilire e monitorare gli standard etici dei dati.

ORIENTAMENTI ETICI DI BASE

PREVENZIONE DEL DANNO

Le raccomandazioni sulla prevenzione del danno, in questa fase, hanno lo scopo di garantire che il potenziale di rischio dei modelli predittivi, di raccomandazione o decisionali e di altre analisi basate sui dati, sia identificato con sufficiente anticipo nel processo di analisi e generazione dei modelli.

Raccomandazioni

- Valutare le prevedibili possibilità di un uso improprio dei modelli.
- Nel corso dello sviluppo del modello, evitare in particolare qualsiasi misura che possa portare alla deanonimizzazione dei dati utilizzati.
- Integrare i meccanismi di controllo nei modelli per monitorare e limitare i potenziali danni che potrebbero derivare dal loro uso pratico; ciò vale in particolare per le tecnologie che possono essere utilizzate per violare la privacy o portare a decisioni errate a causa di imprecisioni nel modello, dannose per gli individui sottoposti a tali decisioni.
- Considerare la possibilità di imporre dei limiti alla distribuzione del prodotto, se e quando è appropriato.



ESEMPIO - VALUTAZIONE DELLA SICUREZZA DELLA AI

Uno degli attori chiave nel campo dell'AI (Deep Mind) ha sviluppato un quadro teorico per guidare la valutazione della sicurezza dei suoi prodotti basati sui dati. Il quadro ha tre elementi: (1) Specificazione (vedi l'esempio del Passaggio 3, Controllo, sotto); (2) Robustezza, che richiede l'identificazione di ambienti non rappresentati nel dataset che possono portare l'AI a comportarsi male (es, un robot per la pulizia della casa testato in una casa senza animali domestici può funzionare in modo diverso in un ambiente che include peli di animali domestici) e testare la possibilità di costruire attacchi avversari; (3) Garanzia, che consiste di due elementi: sistemi di ispezione (statistiche riassuntive, ispezioni automatizzate, e strumenti progettati per rendere le decisioni dell'IA spiegabili o interpretabili dopo l'uso) e meccanismi di applicazione (ad esempio, la possibilità di interrompere, cioè un interruttore «off» affidabile). Fonte: <https://deepmind.com/blog/article/specifying-ai-safety-problems>



GIUSTIZIA

Le raccomandazioni sulla giustizia affrontano il problema della discriminazione indiretta in relazione ai modelli e alle previsioni basate sui dati. La discriminazione indiretta si verifica quando una regola decisionale viene applicata a tutti nello stesso modo, ma ha un effetto non giustificabilmente diverso su diversi gruppi di persone (p.es., donne vs. uomini); ad esempio, gli uomini sono significativamente più propensi delle donne a essere selezionate per un determinato lavoro. Ciò è particolarmente rilevante per i modelli e le previsioni sulla base dei quali vengono prese automaticamente le decisioni o che supportano il processo decisionale umano (p.es., raccomandazioni).



Raccomandazioni

- Assicurarsi che qualsiasi set di dati utilizzato per l'apprendimento automatico sia sufficientemente diversificato per evitare distorsioni dovute a dati di training insufficienti.
- Spiegare quali tipi di discriminazione sono previsti e incorporati nel modello; quali tipi di discriminazione indiretta indesiderata, se presente, sono stati rimossi dal modello e quali forme di discriminazione indiretta sono ancora presenti, se del caso.
- Verificare la presenza di discriminazione indiretta con i dati di test.
- Sviluppare competenze interne su come affrontare la discriminazione indiretta che può derivare dall'apprendimento automatico; questo dovrebbe includere anche metodi tecnici (utilizzando la statistica e l'informatica) per eliminare le ingiustizie, se così generate.
- Se del caso, introdurre tecnologie pertinenti per prevenire la discriminazione indiretta.

ESEMPIO - STRUMENTO DIAGNOSTICO PARZIALE

Un team interdisciplinare di scienziati medici produce un modello predittivo delle risposte dei pazienti al trattamento basato sui dati della genomica, dell'epigenomica e della proteomica. Il modello può fornire una previsione, per ogni dato paziente e trattamento proposto, del fatto che il trattamento avrà successo per quel paziente, sulla base dei dati genomici (ecc.) del paziente. Tuttavia, a causa dei limiti delle risorse, si decide di dare la priorità ai pazienti che più probabilmente ne trarranno beneficio. Si scopre che il 90% degli uomini di origine europea, ma solo il 10% degli uomini di origine africana si prevede che beneficeranno del trattamento, quindi c'è una sproporzione tra le persone a cui si raccomanda di ricevere il trattamento. Questo solleva la seguente domanda: i neri hanno una ragionevole lamentela contro questa disuguaglianza? Inizialmente si sostiene che la disparità di possibilità di beneficiare del trattamento giustifica questa disuguaglianza. Tuttavia, uno scienziato dei dati sottolinea che è possibile aumentare la sensibilità della previsione per i neri (cioè una maggiore proporzione di pazienti neri che saranno guariti sarà identificata e riceverà la cura), con l'effetto collaterale che questo riduce la sua precisione (cioè una maggiore proporzione di pazienti neri che non saranno guariti riceverà la cura per errore). Per valutare questo trade-off, viene organizzato un panel di stakeholder che coinvolge i pazienti e le organizzazioni per la parità. Il panel conclude che l'algoritmo modificato è nell'interesse dei pazienti neri, per i quali (a differenza dei pazienti bianchi) non sono disponibili altre cure efficaci.

AUTONOMIA

Un orientamento verso l'autonomia esige che le parti interessate al di fuori dell'azienda siano in grado di effettuare valutazioni adeguate degli aspetti etici della costruzione del modello, qualora questi modelli abbiano un'influenza sull'autonomia di tali parti.

Raccomandazioni

- Assieme agli stakeholders, individuare le questioni etiche rilevanti per l'uso dei modelli basati sui dati o le conoscenze acquisite dai dati.
- Consentire agli stakeholders o ad altre parti interessate di fornire commenti sullo sviluppo del modello, in particolare per quanto riguarda gli obiettivi incorporati nel modello (p.es., l'obiettivo di classificazione o la funzione di errore).



ESEMPIO - WORLD CAFÉ CON GLI STAKEHOLDER

Un'azienda svizzera di telecomunicazioni mantiene un dialogo con i suoi più importanti gruppi di stakeholder in merito alla sua strategia di sostenibilità. I rappresentanti degli stakeholder esterni hanno discusso con l'azienda i temi chiave della sostenibilità in un World Café (un formato per lo svolgimento di discussioni in piccoli gruppi). Sono stati individuati tre temi principali:

- informazioni sui rischi dello standard di telefonia mobile G5,
- formazione di insegnanti e PMI sui temi della digitalizzazione attraverso la creazione di laboratori di digitalizzazione per le PMI e le scuole,
- creazione di maggiori incentivi per i fornitori a ridurre le emissioni di CO2.

I partecipanti al World Café hanno apprezzato l'opportunità di scambiare direttamente con l'azienda. Sulla base dei feedback positivi ricevuti, l'azienda intende organizzare ulteriori eventi di discussione che coinvolgano i propri stakeholder a complemento del precedente sondaggio tra gli stakeholder. Questo tipo di dialogo con gli stakeholder può essere utilizzato anche per discutere in modo specifico i modelli basati sui dati, ad esempio il rischio che tali modelli generino discriminazioni.



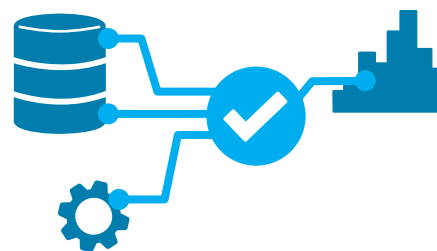
VALORI PROCEDURALI

CONTROLLO

Le raccomandazioni sul controllo hanno lo scopo di garantire che gli aspetti etici dell'analisi dei dati siano sistematicamente documentati.

Raccomandazioni

- Giustificare l'adeguatezza degli algoritmi utilizzati in relazione al set di dati specificato e all'attività di analisi specifica; la giustificazione dovrebbe includere una documentazione relativa alla valutazione dei rischi e delle misure adottate per mitigare i rischi alla luce dei principi di prevenzione del danno, equità e autonomia di cui sopra.
- Garantire che qualsiasi tipo di conoscenza acquisita o modello generato possa essere ricondotto ai dati originali e all'algoritmo utilizzato per crearlo.
- Se l'analisi si basa su dati e/o modelli forniti da altri, assicurarsi che questo input sia stato prodotto in modo conforme ai propri standard etici.



ESEMPIO - CONTROLLO DEI MODELLI DI AI

Uno degli attori chiave nel campo dell'AI (Deep Mind) ha sviluppato un quadro teorico per guidare la valutazione della sicurezza dei suoi prodotti. Il quadro ha tre elementi. Il primo elemento, Specificazione, implementa il valore del Controllo. Il progettista distingue la «specifica ideale» (i desideri del progettista), la «specifica di progetto» (ad esempio, la funzione di ricompensa utilizzata in un sistema di apprendimento di rinforzo), e la «specifica rivelata» (il comportamento effettivo del sistema). Il primo elemento del Controllo è documentare chiaramente la specifica ideale (lo scopo che si vuole raggiungere) e la specifica di progetto (come tale scopo viene tradotto in codice). Per la specifica rivelata, viene proposto un metodo guidato dai dati (ad esempio, apprendimento inverso del rinforzo o scenari simulati) per identificare come la specifica rivelata si discosta dalla specifica ideale (cioè, quando l'IA si discosta dai desideri del creatore). Fonte: <https://deepmind.com/blog/article/specifying-ai-safety-problems>. Un approccio simile è definito in ABOUT ML, la metodologia sviluppata da Partnership for AI che prevede la trasparenza attraverso la documentazione. Vedi: <https://www.partnershiponai.org/wp-content/uploads/2019/07/ABOUT-ML-v0-Draft-Chapter-2.pdf>

TRASPARENZA

Le raccomandazioni sulla trasparenza, in questa fase, sono volte a garantire che sia gli utenti che i revisori esperti delle analisi e dei modelli siano in grado di valutarli in termini di caratteristiche eticamente rilevanti. Ciò vale in particolare per i limiti dei modelli, dove la comunicazione degli stessi deve essere espressa con concetti adeguati al gruppo target che deve utilizzare il modello.



Raccomandazioni

- Spiegare la logica dei modelli e le decisioni che li hanno modellati in un modo che gli utenti del modello possano comprenderli.
- Rendere chiari i limiti dei modelli (p.es., intervalli di confidenza) agli utenti dei modelli.
- Spiegare agli utenti le possibili conseguenze di tali limitazioni (compresa la discriminazione indiretta).
- Documentare le misure adottate per soddisfare i propri standard etici in materia di modellazione basata sui dati: ad esempio, misure adottate per eliminare o ridurre la discriminazione indiretta nel modello.
- Informare gli utenti di queste misure in modo appropriato e mirato; informare il pubblico della logica e delle limitazioni del modello se vengono utilizzati in contesti decisionali che possono influenzare i diritti umani – quando è possibile che queste spiegazioni siano correttamente comprese.

ESEMPIO – INFORMAZIONI SU UN MODELLO DI IA ADATTATO A COLUI CHE RICEVE LA SPIEGAZIONE

Supponiamo che uno data scientist addestri un modello di apprendimento automatico con informazioni sui richiedenti di una linea di credito per prevedere se effettueranno pagamenti puntuali nell'arco di due anni. Questa previsione di machine learning viene utilizzata per decidere se un richiedente si qualifica per una linea di credito. Tre gruppi diversi potrebbero desiderare di ricevere una spiegazione del modello:

- 1) i «data scientist», i quali vorrebbero capire il comportamento del modello nel suo complesso, non solo in casi specifici;
- 2) i funzionari del credito, che hanno bisogno di capire perché una data domanda è stata accettata o respinta quando la si confronta con altre domande;
- 3) richiedenti che non si sono qualificati per una linea di credito e che vorrebbero capire perché, e cosa potrebbero fare per ribaltare questo giudizio.

Per ogni stakeholder viene fornito un diverso tipo di spiegazione. La suite AI Explainability 360 offre:

- 1) software che produce un'approssimazione del prodotto dell'algoritmo sotto forma di modelli direttamente interpretabili, basati su regole, che forniscono una comprensione globale del comportamento dell'algoritmo;
- 2) software che genera profili prototipici di richiedenti di successo;
- 3) software in grado di identificare e comunicare ad un richiedente quale cambiamento individuale nel suo profilo avrebbe cambiato la decisione del modello di IA, a parità di tutto il resto.

Fonte: Pagina dei tutorial del Toolkit di AI Explainability 360 Toolkit, consultabile all'indirizzo aix360.mybluemix.net.



RESPONSABILITÀ

Responsabilità significa garantire che nel complesso processo di analisi dei dati e generazione di modelli siano rispettate le responsabilità per l'adesione alle raccomandazioni etiche.

Raccomandazioni

- Assegnare responsabilità ben definite alle persone coinvolte nello sviluppo del modello per garantire il rispetto degli standard etici per l'analisi dei dati e la generazione di conoscenze.
- A livello organizzativo, assicurarsi che siano rispettati gli standard etici per l'analisi dei dati e la generazione di conoscenze nonché l'identificazione dei problemi etici emergenti.



ESEMPIO - DATA GOVERNANCE BOARD

Una società di media interessata a sviluppare nuovi modelli di business con i dati ha istituito una Data Governance Board. Questo consiglio si riunisce almeno trimestralmente (con riunioni ad hoc a seconda delle necessità) e coinvolge i rappresentanti di tutte le aree di business, in particolare i team attivi nello sviluppo dei prodotti e l'ufficio legale. Questo consiglio discute le questioni etiche delle nuove idee di business apertamente e sotto la guida di una persona che ha ricevuto una formazione in materia di etica. Le discussioni vengono registrate sotto forma di verbale scritto. L'esperienza di apprendimento comune aumenta la consapevolezza delle questioni etiche nella vita professionale quotidiana. In casi critici, queste domande raggiungono anche il management superiore e vengono discusse in questa sede. I verbali delle discussioni di gruppo aiutano a presentare il problema nel modo più preciso possibile, che la direzione apprezza come modalità di comunicazione efficiente.

FASE 4: PRODOTTI E SERVIZI

ORIENTAMENTI ETICI DI BASE

PREVENZIONE DEL DANNO

Le raccomandazioni per la prevenzione del danno, in questa fase, mirano a garantire il riconoscimento del potenziale di abuso di prodotti e servizi basati sui dati.

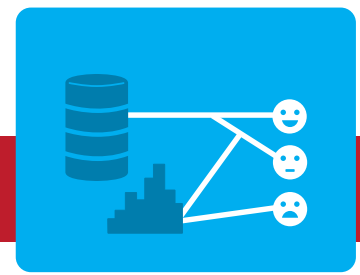
Raccomandazioni

- Non distribuire prodotti e servizi basati sui dati fino a quando non si è verificato in che misura gli utenti non qualificati possano danneggiare i prodotti e/o gli utenti qualificati possano abusarne.
- Per quanto possibile, verificare se i prodotti e i servizi basati sui dati siano effettivamente oggetto di abuso da parte di terzi (p.es., se il software viene utilizzato per scopi diversi da quelli per i quali è stato progettato e testato).
- Determinare se il prodotto può essere utilizzato per deanonimizzare dati anonimi o per raccogliere dati sensibili da dati personali originariamente non problematici.



ESEMPIO - INTERAZIONE CON BOT ABUSIVO

Un'azienda utilizza l'intelligenza artificiale per sviluppare un bot informativo progettato per rispondere a domande di carattere generale (simile al servizio informativo accessibile per telefono, una volta ampiamente disponibile). Il software bot apprende dalle interazioni con gli utenti quanto gli utenti siano soddisfatti delle risposte che fornisce; è stato addestrato con fonti estese e accessibili al pubblico, come Wikipedia. Nel caso di argomenti controversi, il bot fornisce risposte che contengono tutti gli aspetti della controversia al fine di mantenere la neutralità. Ben presto è diventata evidente che il software bot fornisce risposte sempre più «neutrali» anche quando le informazioni sono di fatto sicure: ad esempio, lascia aperta la possibilità che la terra possa essere piatta. La ragione di ciò è un gran numero di sostenitori delle teorie del complotto esprime insoddisfazione per le risposte del bot, influenzandone così l'apprendimento. Gli sviluppatori avevano trascurato di pianificare sistematiche interazioni abusive con il bot. Una volta chiarita l'esistenza e la natura dell'abuso, è stato possibile prendere delle contromisure.



GIUSTIZIA

La giustizia esige di garantire che l'uso di prodotti e servizi basati sui dati non comporti effetti sociali indesiderati come la discriminazione indiretta, la stigmatizzazione o l'esclusione di gruppi di persone. Il dovere etico di valutare e affrontare gli impatti sociali indiretti è proporzionale alle capacità finanziarie e tecnologiche di un'azienda e alla misura in cui i suoi prodotti e servizi basati sui dati influiscono sulla società.



Raccomandazioni

- Evitare di utilizzare prodotti e servizi basati sui dati i cui effetti discriminatori indiretti non sono stati studiati e quantificati, in particolare laddove vi sia il rischio di una significativa discriminazione indiretta.
- Evitare di utilizzare prodotti e servizi basati sui dati che possono inavvertitamente stigmatizzare interi gruppi di persone o minacciare la loro reputazione.
- Valutare se i prodotti e i servizi basati sui dati escludono i gruppi vulnerabili ed evitare tale esclusione, se possibile.
- Monitorare i sistemi decisionali automatizzati o i sistemi di raccomandazione per la discriminazione involontaria anche se il prodotto dei dati è stato testato per la discriminazione indiretta; i dati del test possono avere proprietà statistiche diverse rispetto ai dati reali e possono sorgere nuove forme di discriminazione.

ESEMPIO – SCAMBI DI VALORE NELLE COVID TRACING APPS

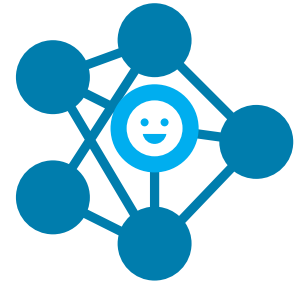
Molte applicazioni europee di tracciamento digitale dei contatti, come Swiss Covid (utilizzata in Svizzera), Immuni (Italia) e Corona Warn App (Germania), hanno scelto un'architettura decentralizzata implementata da un'API Google/Apple che limita notevolmente i dati che possono essere raccolti per ulteriori analisi. Tuttavia, l'architettura decentrata rende molto difficile accertare se alcuni gruppi sociali sono influenzati in modo ineguale dalle notifiche e dai falsi positivi. Se l'app fosse resa obbligatoria, o comunque legata a significativi vantaggi sociali (come il diritto di utilizzare i mezzi di trasporto pubblico), non sapere quali gruppi vengono notificati e le disuguaglianze nel tasso di falsi positivi implicherebbe ignorare importanti elementi di giustizia nella diffusione del danno dei falsi positivi. Pertanto, l'app così come è stata progettata protegge la privacy, ma impedisce la raccolta di informazioni rilevanti per le implicazioni di giustizia dell'uso e della funzionalità dell'app. C'è quindi una tensione tra giustizia e privacy nello sviluppo dell'app. Pertanto, se l'app dovesse essere resa obbligatoria, Google/Apple dovrebbe modificare il protocollo per consentire la raccolta sistematica di dati anonimi, consentendo di valutare le correlazioni tra il tasso di notifica (e di false notifiche positive) e le diverse caratteristiche demografiche (sesso, etnia, reddito, ecc.), al fine di identificare accuratamente le popolazioni a rischio in modo sproporzionato.

AUTONOMIA

Le raccomandazioni sull'autonomia hanno lo scopo di garantire che il feedback che segnala gli effetti negativi dei prodotti e dei servizi basati sui dati sia preso sul serio.

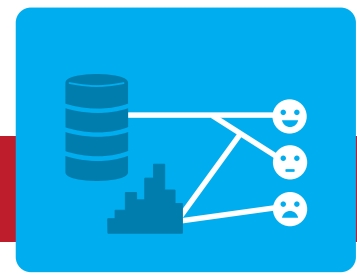
Raccomandazioni

- Implementare e seguire i processi che consentono alle parti interessate di segnalare risultati negativi dei prodotti di dati.
- Seguire i dibattiti degli esperti nei vari contesti applicativi dei prodotti di dati, che forniscono prove di deviazioni sistematiche nei processi decisionali, sulla base della conoscenza dei rispettivi esperti.
- Imparare a riconoscere le decisioni automatizzate, che si discostano da quelle che gli esperti consapevoli della logica di un determinato settore o ambito di competenza considererebbero accettabili.



ESEMPIO - GIURIA DEI CITTADINI

Deep Mind ha collaborato con la (UK) Royal Society for the encouragement of Arts, Manufactures and Commerce (RSA) su un progetto per promuovere un reale coinvolgimento del pubblico sull'impatto reale dell'IA. La RSA ha presieduto una giuria di cittadini per esplorare l'uso dell'IA nel processo decisionale. Questa è stata utilizzata per discutere questioni generali sull'IA, per esempio, che tipo di partecipazione il pubblico deve avere per sentirsi a proprio agio con l'IA e quali sono le «linee rosse» dei cittadini (cioè, domande rilevanti prima dell'implementazione e in fase di progettazione; vedi Fase 3 - Autonomia). Per implementare il principio di autonomia qui descritto, lo stesso mezzo (un forum dei cittadini) potrebbe essere usato per facilitare la discussione degli aspetti negativi dell'implementazione di una tecnologia dopo l'implementazione. Fonte: <https://www.thersa.org/action-and-research/rsa-projects/economy-enterprise-manufacturing-folder/tech-and-society/forum-for-ethical-ai>



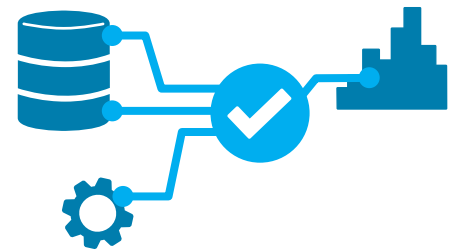
VALORI PROCEDURALI

CONTROLLO

Le raccomandazioni sul controllo mirano a garantire, a livello generale, che i valori di un'azienda si riflettano adeguatamente nei prodotti e servizi basati sui dati che l'azienda intende offrire sul mercato.

Raccomandazioni

- Determinare le priorità di etica dei dati per l'azienda (queste devono essere determinate dai doveri etici e dalla missione dell'azienda); e ricavare i requisiti corrispondenti per i prodotti e servizi basati sui dati.
- Valutare se gli standard etici sono stati seguiti in ogni fase del processo che ha portato al prodotto o servizio (sia esso realizzato all'interno dell'azienda o da altri) e se sono coerenti con la missione, i valori e l'immagine pubblica dell'azienda.



ESEMPIO - PROBLEMI DI APPLICAZIONE DELLE LINEE GUIDA AI

L'ONG AlgorithmWatch ha iniziato a compilare le linee guida per l'«IA etica». Tuttavia, delle oltre 160 linee guida ora incluse nell'inventario globale di AI Ethics Guidelines (2020), solo una piccola parte (10) dispone di adeguati meccanismi di supervisione e applicazione. Revisioni regolari di tali linee guida, come quelle condotte da AlgorithmWatch, e discussioni critiche ricorrenti potrebbero creare una pressione pubblica per rendere controllabili le linee guida risultanti dagli standard etici dei dati.

TRASPARENZA

La trasparenza esige di fornire un'adeguata informazione sul rispetto delle raccomandazioni contenute nel presente Codice e di contribuire al mantenimento delle pratiche etiche.

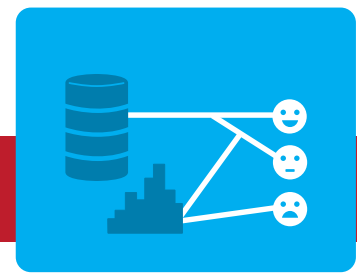
Raccomandazioni

- Comunicare gli standard e pratiche etiche a chiunque possa essere influenzato dai risultati del prodotto di dati.
- Se l'azienda produce dati, modelli o prodotti di dati per terze parti, fornire tutte le informazioni necessarie per l'uso etico di questi dati/modelli/prodotti; ad esempio, se una terza parte desidera costruire un software basato sui dati raccolti dall'azienda, fornire informazioni sufficienti sul set di dati per consentire al progettista dell'applicazione di evitare di introdurre involontariamente errori nel modello.



ESEMPIO - RIVENDITA DI CONOSCENZE GENERATE DA MODELLI

Un'azienda utilizza i dati per generare nuove informazioni attraverso la loro analisi, che a loro volta vengono messi a disposizione di altri per l'ulteriore elaborazione. Nei dati che l'azienda utilizza, ha specificato gli scopi per i quali può utilizzare i dati e li ha memorizzati in metadati. Le linee guida etiche dell'azienda definiscono valori fondamentali, quali la non discriminazione, il rispetto delle persone, l'autodeterminazione, la trasparenza, la responsabilità, ecc. In particolare, l'azienda si è impegnata a garantire la tracciabilità dei dati, il che significa che l'origine e lo scopo dei dati rimangono noti lungo tutta la catena del valore. Questi principi sono allegati alle informazioni vendute. Attraverso indagini sporadiche, l'azienda assicura che i terzi che basano i loro prodotti sui risultati e sui modelli dell'azienda siano informati sulla finalità dei dati così come sono stati originariamente utilizzati dall'azienda.



RESPONSABILITÀ

Le raccomandazioni sulla responsabilità mirano a garantire che i top management dell'azienda possano assumersi in modo appropriato la responsabilità del prodotto.

Raccomandazioni

- Prevenire la diluizione di responsabilità nel complesso processo di sviluppo di prodotti e servizi basati sui dati.
- Implementare processi che garantiscano il mantenimento degli standard etici nella fornitura del prodotto di dati.
- In caso di problemi etici con i prodotti di dati, accettare la responsabilità e non creare capri espiatori.
- Essere consapevoli del fatto che le norme etiche del proprio prodotto di dati possono essere fortemente influenzati dagli standard etici di altre aziende con cui si interagisce (p.es., perché hanno raccolto i dati utilizzati).
- Tenere conto di queste influenze nella valutazione degli standard etici dei prodotti di dati e assumersi la responsabilità del prodotto finale.



ESEMPIO - RESPONSABILE DELLA SICUREZZA INFORMATICA

Parte della responsabilità della direzione è di garantire la sicurezza dei dati (questo è un elemento della fase 2). Il ruolo del responsabile della sicurezza dell'informazione («Chief Information Security Officer,» CISO), che è stato introdotto in molte aziende, può servire da esempio. Nell'istituire tale funzione occorre evitare quanto segue:

- mancanza di una vera e propria autorità investita di tale funzione, in modo che il CISO possa svolgere il suo compito solo formalmente, ma non in modo efficace;
- l'installazione del CISO come capro espiatorio da sacrificare quando si commette un errore.

Al contrario, le seguenti caratteristiche di un CISO possono contribuire al successo di tale figura:

- concentrarsi sulla gestione del cambiamento e sulla trasformazione, se un'azienda vuole svilupparsi in modo sostanziale;
- focalizzazione su questioni operative e tecniche, al fine di sviluppare soluzioni praticabili per le operazioni in corso;
- focalizzazione su questioni normative e di conformità, laddove queste siano di grande importanza per l'azienda;

Queste differenziazioni sono utili non solo per la sottoarea della sicurezza delle informazioni, ma anche per le altre aree menzionate ai punti 1, 3 e 4.

