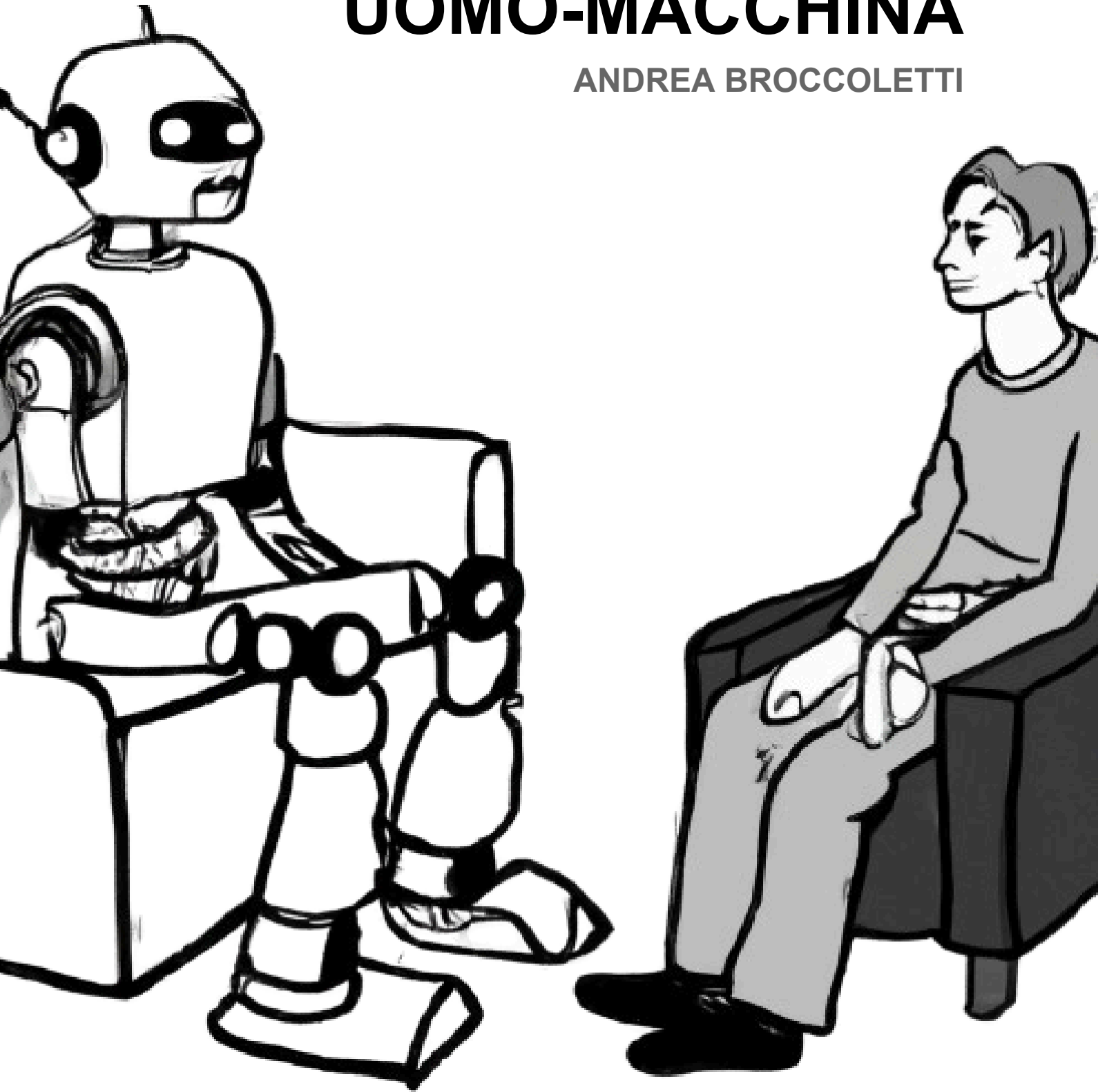


INTERAZIONE UOMO-MACCHINA

ANDREA BROCCOLETTI



CORSO DI LAUREA DI SCIENZE INFORMATICHE
Università degli studi di Milano-Bicocca



CREDITI

Andrea Broccoletti

brockdev.it

Per le fonti e i materiali utilizzati:

Federico Cabitza

prof.re Interazione Uomo-Macchina

Chiara Natali

prof.ssa Interazione Uomo-Macchina

L'INTERAZIONE UOMO MACCHINA

Definizione

L'interazione uomo macchina è una disciplina che studia l'**interazione tra utenti e computer** per la **progettazione** e lo **sviluppo di sistemi interattivi** che siano usabili, affidabili e che facilitino le attività umane.

Per fare ciò, è necessario considerare diversi settori:

- Informatico, **paradigmi di interazione**
- Scienze della progettazione, **ergonomia**, design e project management
- Scienze dell'uomo, **psicologia**, linguistica e scienze cognitive e della comunicazione

Origini

La disciplina nasce intorno agli anni '80, più precisamente nel **1984** quando nasce la prima interfaccia grafica, il **desktop** con l'Apple Macintosh.

parallelamente a tutto questo, si assiste anche alla nascita di diverse discipline legate a quella che verrà poi definita come **software engineering**. In particolare, la prima conferenza dedicata all'interazione, la **ACM CHI**, è del 1983.

Ergonomia cognitiva

Studio delle **interazioni tra l'uomo e gli strumenti che servono per elaborare l'informazione**, quindi si studiano anche tutti i processi cognitivi coinvolti come percezione, attenzione, memoria, pensiero, linguaggio ed emozioni.

Non si associa ad un concetto di comodità materiale, ma significa etimologicamente studio del lavoro, dell'attività di tutti i tipi.

Difficoltà di progettazione

Esistono diverse difficoltà nel progettare in modo che l'interfaccia sia efficace:

- La varietà di **sistemi interattivi**, estremamente diversi tra loro
- La varietà degli **utenti**
- La varietà degli **scopi** e degli **usi**. si può progettare per qualsiasi contesto

Concetti base

- **Design** equivalente alla progettazione

- **Progettare** ideare qualcosa e studiare il modo di attuarla, così da portare un valore nell'ambito in cui viene usata
- **Sistema interattivo** sistema che risponde ad utenti umani e reagisce alle loro azioni in maniera percepibile ad un utente umano
- **Valutazione** processo in cui si considera cosa è di valore per l'utente e si considera tale valore come criterio per capire la qualità di un sistema

Temi dell'interazione

Si trattano criteri, metodi e strumenti per:

- La **progettazione dell'interazione** fra esseri umani e sistemi interattivi
- La **valutazione dell'usabilità** dei sistemi interattivi
- Progettare **nuove tecniche di interazione**
- Valutare l'**impatto dell'automazione** nei contesti umani

Non esiste una vera e propria differenza tra **impatto** e valutazione, in particolare la valutazione è l'impatto che il sistema ha sui singoli individui in un contesto spazio-temporale ben definito e limitato, "qui ed ora".

L'usabilità è un tipo di impatto sugli utenti e non una caratteristica intrinseca del sistema, non si può quindi valutare senza il coinvolgimento degli utenti.

SISTEMA SOCIO-TECNICO

Sistema ST

Un sistema socio-tecnico è un sistema in cui la **componente umana** (sociale) e quella **tecnica** (tecnologica) sono inestricabilmente **legate** ed interdipendenti tra loro.

Un tale sistema, agli occhi di un osservatore esterno, appare come un'entità unitaria ma collettiva, con caratteristiche e comportamento proprio.

L'interazione delle due componenti porta a fenomeni emergenti imprevedibili. Non è possibile prevedere a priori come opererà un sistema ST: le **proprietà** non sono desumibili da quelle delle sue parti prese isolatamente, ma emergono proprio in virtù delle **interazioni** tra quelle in un determinato **ambiente** (analisi locale).

Il sistema gode dell'**invarianza di scala**: si può riconoscere a diversi livelli di grandezza del sistema di interesse (da un small business fino ad un'intera azienda).

Proprietà emergenti

Sono quelle che emergono dall'interazione delle parti:

- **Funzionali** riguardano il **funzionamento** dell'intero sistema una volta che tutte le sue parti funzionano bene
- **Non funzionali** riguardano **quanto bene opera il sistema** in un determinato **contesto**

STS Thinking

È un approccio consapevole del fatto che le due componenti si integrano bene e si ottimizzano solo congiuntamente in **configurazioni sub-ottime**.

L'obiettivo è appunto trovare una delle configurazioni che permetta alle due componenti di integrarsi in maniera ottimale in modo anche da soddisfare le persone coinvolte.

CONSEGUENZE INATTESE

Conseguenze

Un sistema socio-tecnico è un sistema complesso, in cui è impossibile prevedere ogni possibile evoluzione.

Da un punto di vista matematico, significa dalla **evoluzione parzialmente o totalmente imprevedibile**, ma da un punto di vista tecnologico, è il concetto di **conseguenza inattesa delle tecnologie**.

Effetto Peltzman

Effetto per cui non è scontato e agevole determinare la conseguenza di un determinato intervento, che può portare a conseguenze contrarie a quelle attese.

Overliance

Nell'ambito della tecnologia, si pensi ad una tecnologia efficace e che soddisfi veramente gli utenti: se una cosa è buona e si usa, può allora diventare una **dipendenza**, con riscontri anche negativi.

Si distingue in:

- **Overdependence** riguarda **ciò che si fa**, gli utenti dipendono troppo dallo strumento informatico generando: mancanza di **autonomia**, **abuso** dello strumento anche quando non necessario, mancanza o ignoranza di **alternative**
- **Overconfidence** riguarda **ciò che si ritiene giusto o rischioso**, pensando quindi che lo strumento non smetterà mai di **funzionare**, che non genera mai **sbagli** e che non comporti mai **danni**

Definizioni di Parasuraman

- **Complacency** estrema **fiducia** che il sistema computazionale funzionerà sempre per come è stato progettato, e quindi rischia di far abbassare il livello di attenzione quando lo si usa
- **Automation bias** distorsione cognitiva dovuta all'automazione, si verifica quando c'è un'eccessiva fiducia nella risposta del supporto alle decisioni, causando quindi errori di omissione o di azione quando i sistemi sono imperfetti

CONCETTI DI QUALITÀ

Usabilità

Una bassa usabilità porta inevitabilmente a danni e problemi. La bassa usabilità la si può ritrovare in diverse situazioni:

- Gli utenti non capiscono come usare il sistema
- Il sistema presenta un'eccessiva quantità di **funzionalità** (low use)
- Gli utenti non capiscono **cosa stia facendo il sistema** (poca trasparenza)
- Interazioni con tecnologie sono spesso fonte di **emozioni negative**, come frustrazione e ansia, che possono portare l'utente a compiere errori più o meno gravi

È importante non dare per scontato che la responsabilità sia dell'utente nel momento in cui si presenta un problema. Molto spesso infatti vi è un concorso di colpa da parte dei progettisti, che devono quindi cercare di prevenire problemi legati all'usabilità.

The Norman doors

una porta di Norman è una porta progettata in modo da non far capire se si apra tirando o spingendo, rappresenta la tendenza del progettista a **complicare affari semplici** e a pretendere che basti descrivere una **procedura** per renderla ovvia a qualsiasi utente.

The Floyd toilette

È un oggetto che funziona ma che è assurdamente complicato, rappresenta la tendenza del progettista a concepire oggetti o **sistemi che non invitano all'uso** (e all'uso corretto) sulla base di elementi visuali o indicazioni chiare (affordances), affidandosi solo alla memoria, all'esperienza o all'intuizione ed incontrando spesso il disorientamento o fraintendimento dell'utente.

Sistema interattivo

Combinazione di componenti hardware e software che ricevono input da un utente umano, e gli forniscono un output allo scopo di supportare l'effettuazione **di un compito**.

Interfaccia

È l'insieme dei componenti di un sistema interattivo (software o hardware) che forniscono all'utente informazioni e comandi per permettergli di effettuare specifici compiti attraverso il sistema.

L'utente non vede la parte di back-end, per lui **l'interfaccia è il sistema interattivo**.

Questo significa che in realtà molti problemi di usabilità sono in relazione con problemi di interfaccia. L'elemento comune più rilevante tra i due è il concetto di effettuazione di un **compito** (task) con il fine di **raggiungere un obiettivo** ben definito.

Constraints

Sono i vincoli, che possono essere:

- **Passivi o strutturali** come un form con tre campi per vincolare l'inserimento di soli tre tipi di dato
- **Attivi o funzionali** come dei controlli per cui l'utente non può proseguire se non compila tutti i campi o se non li compila correttamente

I vincoli sono strumenti per garantire una **migliore qualità dei dati** e una standardizzazione delle pratiche, ma ovviamente non si deve esagerare.

Workaround

Percorso che l'utente avrebbe voluto fosse stato progettato dal progettista e che viene quindi **percorso** lo stesso a discapito dei **vincoli stabiliti**.

In informatica, è qualsiasi azione relativa all'esecuzione di un processo o di un compito (supportati dal sistema informatico) che non è prevista o descritta nei manuali di uso del sistema informatico e/o nei manuali che descrivono tale processo o procedura e che può bypassare l'uso del sistema o piegarlo ai propri fini.

Affordance

Una qualità o proprietà di un oggetto che definisce i suoi possibili usi o **rende evidente** all'utente **come utilizzarlo**.

Ci sono diversi tipi di affordance:

- **Cognitive** aiutano l'utente ad effettuare azioni cognitive come pensare, decidere, ricordare
- **Fisiche** aiutano l'utente con le sue azioni fisiche come cliccare, toccare, muovere
- **Sensoriali** aiutano l'utente con le azioni sensoriali come vedere, udire, sentire
- **Funzionali** aiutano l'utente in lavori reali affinché l'utente compia azioni che hanno effetti nel mondo
- **Emozionali** aiutano, supportano e facilitano la comunicazione emotiva, con lo scopo di suscitare emozioni nell'utente
- **Sociali** proprietà di un oggetto che invitano e facilitano azioni collaborative, sociali

Queste categorie di affordance non sono mutualmente esclusive, è possibile avere vari aspetti di un oggetto che fanno sia da affordance fisica che cognitiva o sensoriale.

Legge di Fitts

Ha a che fare con la dimensione dei comandi e alla distanza media da un punto a un'area. In ergonomia questa legge rappresenta il modello matematico di un movimento umano, calcolando il tempo impiegato per **muoversi rapidamente da un punto iniziale a un'area con una determinata estensione**. Una buona interfaccia **minimizza** il tempo.

Scheumorfismo

Uno scheumorfismo è un **ornamento** fisico o grafico apposto su un oggetto allo scopo di **richiamare le caratteristiche estetiche di un altro**.

In ambito digitale, è scheumorfismo un oggetto le cui affordance **ricordano quelle di un oggetto reale**, con il quale si pensa l'utente abbia una maggiore familiarità.

Ad oggi non è più utilizzato, preferendo il **flat design**.

User created affordances

Sono le affordances introdotte dall'utente stesso che si appropria di uno strumento al quale manca qualcosa per comprenderne a pieno il funzionamento.

È un segnale al progettista che secondo l'utente **manca qualche indicazione** per l'uso ottimale dello strumento

Mapping

È una caratteristica dell'affordance, rappresenta la relazione tra un controllo e l'effetto che tale azione avrà nel mondo reale.

Esistono i tipi di mapping:

- **Arbitrario** una relazione del tutto immaginata dal progettista e che l'utente adotta solo per prove ed errori e poi impara a memoria, senza riferimenti culturali o indicazioni fisiche
- **Convenzionale** la relazione tra controllo ed effetto è stabile, quasi ovvia, ma legata a delle convenzioni
- **Naturale** la relazione è meno immediata che nel caso diretto, ma ci viene appunto naturale immaginare la relazione esistente tra controllo ed effetto, come ad esempio il puntatore a forma di mano che suggerisce un'azione
- **Diretto** vi è una relazione fisica, visibile, spesso legata alla posizione tra controllo e l'effettore controllato che produce un effetto

Design by affordance

1. Seguire le maggiori **convenzioni** già stabilite per immagini e azioni e adottare il **mapping naturale** ove possibile
2. Se appropriato, utilizzare delle **scritte in aggiunta** ad icone e grafiche
3. Utilizzare **metafore riconoscibili** e ormai standard
4. Essere consistenti e coerenti nell'uso del modello concettuale dietro al design

Progettare una buona interfaccia

Con una buona interfaccia l'utente può essere più **superficiale** e svolgere il suo compito dimenticandosi che sta usando un computer.

Una buona interfaccia dà inoltre benefici in termini di **maggior produttività**, minor turnover, miglior morale, minore frustrazione, stress, insoddisfazione, maggiore soddisfazione nel lavoro che si traducono in minori costi operativi.

Un tipico errore di interfaccia è quello di **giustapposizione**: errori dovuti a quando un'opzione è vicina ad un'altra sullo schermo e quindi capita con facilità che quella sbagliata venga selezionata per errore.

Multifunzionalità

Vista la moltitudine di oggetti che vengono ormai usati, nel corso degli anni si è avuto un **aumento della complessità** nei dispositivi.

Bisogna **progettare per la più bassa complessità d'uso possibile**. Lavorare per l'usabilità significa abbassare la quota degli oggetti con un'elevata complessità:

- **d'uso** come viene effettivamente utilizzato il sistema
- **funzionale** quanti casi d'uso possiede e le sue funzioni
- **strutturale** come è costruito e programmato

Si deve semplificare l'uso perché:

- **Pervasività** della tecnologia nel mondo odierno
- Necessità di semplificare l'utilizzo e di renderla **accessibile a tutti**
- Necessità di comprendere ruoli e possibilità della tecnologia per **migliorare la qualità della vita**

VALUTAZIONE EURISTICA

La valutazione

Si può collegare il concetto di principi al concetto di euristica. Infatti una euristica non è altro che un **insieme di concetti, riferimenti e strategie** che si sono rivelate particolarmente adatte a risolvere un determinato problema.

Si definisce come la **valutazione di usabilità svolta alla luce di un determinato insieme di euristiche ben definite, possibilmente ben note e prese di riferimento, per identificare soluzioni di design che o si conformano o violano una o più euristiche di tale insieme**. Il fine è quindi farsi un'idea sufficientemente completa dei possibili problemi di usabilità, più se ne trovano migliore è il lavoro svolto.

utenti ed esperti

Spesso è una distinzione un po' forzata, con esperto si intende una persona che ha dimestichezza con i sistemi informatici, che sa di cosa lamentarsi e può stabilire se una cosa è facile da usare oppure no.

Nella valutazione si richiedono utenti **esperti sull'usabilità** ed **esperti sul dominio**.

Indicazioni delle indicazioni

- **Generalità** fa riferimento all'ambito di applicazione, più è generale e più potrà essere applicato in diversi domini applicativi
- **Prescrittività** descrive quanto una certa cosa è mandatoria, da seguire

In base alle due dimensioni, si possono trovare diverse indicazioni.

Principi

Indicazioni generali per la progettazione di interfacce utente usabili, basati su evidenza scientifica o sul generale consenso. Derivano dalla conoscenza degli aspetti psicologici, computazionali e sociali e sono indipendenti dalla tecnologia. Sono espressi spesso in forma molto generale.

Linee guida e design patterns

Sono un insieme di raccomandazioni non vincolanti per il progetto dell'interfaccia utente per una particolare classe di sistemi, espresso in modo generale ma meno astratte dei principi. Si dà al progettista la discrezione di applicarli o meno, caso per caso.

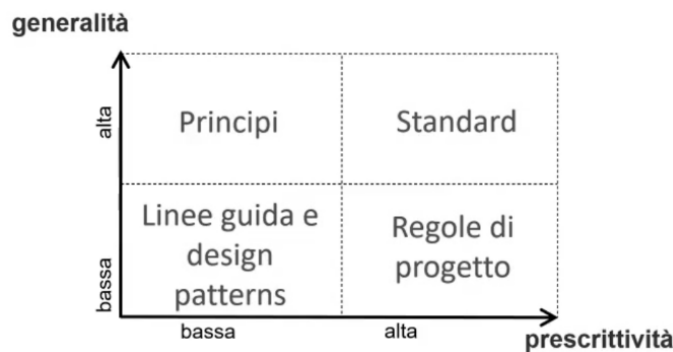
Il concetto di design patterns si introduce per riutilizzare buone pratiche e non doverle reinventare ogni volta che si ripropongono.

Standard

Insieme di regole da applicare nel progetto di una classe di sistemi. Possono essere vincolanti e sono di norma emesse da un ente di standardizzazione.

Regole di progetto

Insieme di regole o specifiche da applicare nel progetto di un particolare sistema; sono vincolanti all'interno del progetto, perché ad esempio dettate da precisi requisiti del cliente.



ISO 9241

Standard chiamato **Ergonomics of human System interaction**, è composto da diverse serie, ciascuna delle quali composta di diversi documenti, in diverso stadio di evoluzione. Attualmente sono 20 documenti.

Autorevolezza

Ci sono diversi **livelli di autorevolezza** o forza dell'evidenza delle indicazioni:

- A. Completamente supportate da risultati di ricerca e dati empirici
- B. Basate su pratica generalmente accettata in modo documentato
- C. Non ben documentate, ma supportate dall'opinione di professionisti esperti
- D. Opinione individuale del progettista

Negli standard sono solitamente recepite solo indicazioni di livello A e B. I principi di Nielsen sono tra B e C.

Insiemi di euristiche

Esistono svariati insiemi di euristiche, si possono scegliere secondo una propria analisi ma anche in base alle **finalità**.

IDX Checklist

Sviluppata da due programmatori che la usano per il proprio contesto aziendale e che, considerandola efficace, l'hanno pubblicata.

Standard ISO 9241-110

Sono 7 principi del dialogo uomo-macchina: adeguatezza del compito, autodescrizione, conformità alle aspettative dell'utente, adeguatezza all'apprendimento, controllabilità, tolleranza verso gli errori, adeguatezza alla personalizzazione.

Bells and whistles bias

Convinzione errata che più opzioni si offrono e più modi di fare si mettono a disposizione, più utenti si soddisfano. In realtà ogni **cosa in più da imparare** è una cosa in più da poter **usare in maniera scorretta**.

Fenomeno del Low Use

Situazione in cui un'applicazione è eccessiva per un utente, riguarda infatti il **rapporto tra funzioni** usate spesso o a volte e quelle che non sono note o che sono note ma usate praticamente mai.

Estetica

Euristiche che riguardano l'importanza della grafica, del layout e del colore:

- Mettere le informazioni che devono essere lette per prime
- Oggetti lampeggianti e caratteri maiuscoli rallentano del 10% la velocità di lettura
- Non usare più di 5-7 colori e mai brillanti per lo sfondo
- Assicurarsi che gli elementi dell'interfaccia si distinguono anche senza i colori

Prevenzione degli errori

Si possono prevenire gli errori:

- Creando **meccanismi di prevenzione**
- Riducendo gli errori tramite **richiesta di conferma** di un'azione
- **Evitando comandi troppo simili** con significati diversi
- Creando **meccanismi di comunicazione** utente-progettista: errori dell'utente rilevati nell'uso portano ad una riprogettazione

Esistono molteplici tecniche di prevenzione degli errori:

- **diversificare le azioni dell'utente** tecnica per prevenire i lapsus ed evitare che l'utente esegua inavvertitamente un'azione al posto di un'altra
- **amodalità** un sistema modale è un sistema che può avere un comportamento modale, ovvero si comporta a seconda dello **stato corrente**. C'è un trade off da mantenere tra l'amodalità e l'interfaccia minima, In alternativa si può evidenziare la modalità corrente
- funzioni **vincolate**
- avvertimenti e richieste di **conferma**
- default **inoffensivi**
- bypass sicuri

Avvertimenti e richieste di conferma

Ci sono tre tipologie di avvertimenti:

- **Note** semplici accorgimenti sullo stato

- **Caution** avvertimenti su qualcosa che si sta facendo
- **Stop** segnalazione di impossibilità

Si dovrebbe sempre chiedere **conferma** prima di effettuare azioni irreversibili o comunque pericolose, con messaggi **precisi** e **costruttivi** tramite la sequenza **alert, identify e direct**: notificare che qualcosa non va, dire cosa non va, suggerire come risolvere il problema.

Infine, è bene non usare messaggi **intimidatori**, evitando termini come fatale, illegale, aborted.

VALUTAZIONE EURISTICA: IL METODO

Il metodo

Una valutazione euristica di usabilità è una tecnica **sistematica** di ispezione predittiva. Un gruppo di valutatori esamina il sistema e identificare eventuali problemi di usabilità rispetto ad un insieme prefissato di **principi**, ossia le **euristiche**.

Il risultato è un elenco di problemi di usabilità, associati ad una o più euristiche, pesati per gravità.

Olistica

Non guidata, basata su una libera navigazione ed esplorazione del sistema, per viste di insieme. Tipico dei sistema **walk-up-and-use**, ovvero quei sistemi così semplici che sono autoesplicativi e possono essere usati senza un'introduzione o training dell'utente.

Task-oriented

Guidata da una task o un **scenario**.

L'ideale sarebbe avere un utilizzo sia olistico sia task-oriented, senza che l'uno escluda l'altro.

Funzionamento del metodo

Segue i punti:

1. Scegliere un'applicazione o un mockup da valutare
2. Concepire dei task, o casi d'uso
3. Ogni membro del gruppo esplora l'applicazione
4. Ogni membro, alla luce delle euristiche, scrive un **elenco di problemi** che ha riscontrato, riconducendoli alla violazione di una o più euristiche
5. Coinvolgere degli esperti di dominio che esploreranno l'applicazione
6. Presentare all'utente un task e farlo eseguire
7. Estrarre collaborativamente un elenco di problemi di usabilità emersi
8. Unire le liste con i problemi inizialmente annotati
9. Predisporre un **questionario** con il quale prioritizzare e far prioritizzare i problemi trovati
10. Associare ad ogni problema una o due euristiche

Concepire i task

Concepire task **realistici**, che siano **eseguibili** evitando di suggerire o descrivere i singoli passi.

Efficienza dei valutatori

Empiricamente si è fissato a **15 valutatori** il limite di valutatori oltre il quale non c'è un effettivo miglioramento nei problemi trovati.

In particolare, si può stabilire che il numero di problemi di usabilità distinti trovati aggregando i rapporti di v valutatori indipendenti con N che indica il numero totale di problemi e p la proporzione di problemi trovati da un singolo valutatore:

$$ProblemsFound(v) = N(1 - v(1 - p))$$

Mentre definendo T il numero totale di problemi, si può stimare il numero totale di problemi trovati da due persone X e Y indipendentemente l'una dall'altra:

$$\frac{X \cdot Y}{X \cap Y}$$

Dove con unione si intende il numero di errori trovati da entrambi.

Valutazione della severità

Un **elenco** del genere deve essere proposto ai valutatori per la valutazione della severità, cioè per ciascun problema i partecipanti alla valutazione della severità devono assegnare un grado di severità.

Si può utilizzare la scala proposta da Nielsen:

0. Non è un problema di usabilità
1. Problema accessorio, non deve essere risolto, a meno che non si abbiano le risorse per farlo
2. Problema secondario, bassa priorità attribuita alla sua risoluzione
3. Problema rilevante, è importante risolverlo e gli si attribuisce alta priorità alla sua risoluzione
4. Catastrofe di usabilità, è imperativo risolverlo prima che il prodotto possa essere rilasciato

Il punteggio dovrebbe essere scelto tenendo conto dei seguenti fattori:

- **Frequenza** di occorrenza
- **impatto** del problema se occorre
- **Persistenza** perchè se succede spesso è più grave del one-time

Risorse necessarie

Un **sistema funzionante** con N valutatori, un osservatore quando si coinvolgono gli utenti e del **tempo**.

PRINCIPI EURISTICI DI JACOB NIELSEN

Le euristiche

Di seguito sono presentate le euristiche di Nielsen, rielaborate dal prof. Mussio in italiano:

1. far vedere lo stato del sistema

Cioè si deve fare in modo che il sistema fornisca un feedback all'utente.

2. adeguare il sistema al mondo reale

Non si deve pretendere che l'utente si adegui al sistema, quindi l'applicazione deve parlare il linguaggio dell'utente.

3. controllo dell'utente e libertà

Lasciare controllo e libertà, in maniera concreta si traduce in flessibilità, che si raggiunge quando un sistema permette all'utente di raggiungere i suoi obiettivi in più di un modo. Si opera anche attraverso l'individualizzazione da non confondere con la profilazione.

4. assicurare consistenza

Qualunque sia la decisione del progettista, si applica in maniera coerente con il sistema e all'interno dell'applicazione, consistente con l'ambiente interno ed esterno.

5. riconoscimento piuttosto che uso della memoria dell'utente

L'utente non deve ricordarsi come fare una certa cosa, ma deve essere intuitivamente portato a farla senza troppi ragionamenti sull'utilizzo.

6. assicurare flessibilità ed efficienza d'uso

Non sono altro che gli acceleratori.

7. visualizzare tutte e sole le informazioni necessarie

Si rifà ad un concetto di minimalità: si devono visualizzare tutte e sole le informazioni necessarie per un certo contesto applicativo, è inutile inondare l'utente di informazioni che non utilizzerà. ha a che fare sia con l'estetica, sia con l'essenzialità dell'architettura dell'informazione.

8. prevenire gli errori

Prima di tutto prevenirli, minimizzare le probabilità che occorran.

9. permettere all'utente di correggere gli errori

Non solo di rilevarli, introducendo funzionalità come gli UNDO.

10. help e documentazione

Nel tempo si è rivalutata la documentazione inventando tutorial, Q&A, help che fanno capire il sistema all'utente. Declinabile quindi anche nelle modalità odierne.

I precedenti punti sono suddivisi per tipologie in percezione per i punti 1, 2, 3; cognizione per i punti 4, 5, 6, 7; errori per i punti 8, 9, 10.

VALUTAZIONE QUANTITATIVA

Sistema di successo

Un sistema può essere considerato di successo se è:

- **Utile** deve fare quello che viene richiesto
- **Usabile** deve consentire di fare quello che viene richiesto in modo naturale, senza aumentare i rischi di errore
- **Usato** deve rendere le persone desiderose di usarlo, quindi essere interessante, piacevole, valorizzare la **UX** e la **IxD**

L'**esperienza d'uso** è ciò che un utente prova quando utilizza un prodotto, un sistema o un servizio. L'UX ha natura **soggettiva** perché riguarda i pensieri e le sensazioni di un individuo nei confronti di un sistema, inoltre è **dinamica** dal momento che si modifica nel tempo al variare di circostanze.

Usabilità

Il grado con cui un prodotto può essere usato da determinati utenti per raggiungere determinati obiettivi con **efficacia**, **efficienza**, **soddisfazione** in un determinato contesto d'uso.

L'efficacia indica il grado di correttezza, accuratezza, completezza in cui l'utente raggiunge i suoi obiettivi.

L'efficienza indica il rapporto tra le risorse usate utili per raggiungere il risultato e quelle effettivamente consumate.

Efficacia è fare la cosa giusta, efficienza è farla nel modo giusto.

Misurare l'usabilità

Si devono introdurre delle differenze per stabilire cosa significa misurare:

- **Misura** è il risultato di una misurazione, assegnazione empirica ed oggettiva di un valore ad un'entità, per caratterizzare un attributo specifico
- **Metrica** comprende la misurazione quantitativa del grado di possesso di uno specifico attributo da parte di un sistema e definisce anche le procedure, modalità e regole di misurazione

Nelle misurazioni, una **differenza praticamente significativa** è una differenza che fa differenza. **Statisticamente significativa** significa che non è dovuta al caso e non dipende dal campione coinvolto.

Estensioni di usabilità

SI estende in:

- **Robustezza** il sistema deve **indurre un basso tasso di errori**, che devono essere facili da correggere e con basso impatto
- **Apprendibilità/Learnability** facilità nell'apprendere il comportamento del sistema
- **Ricordabilità/Memorability** facile da ricordare, anche quando un utente riusa il sistema dopo diverso tempo

Trade-off learnability e soddisfazione

Il **trade-off** è una situazione che implica una scelta tra due o più possibilità, in cui la diminuzione di una quantità costituisce un aumento in un'altra quantità.

A volte conviene progettare sistemi difficili da imparare ma che danno soddisfazione all'utente esperto perchè si crea fidelizzazione. Altre volte invece è la prima impressione che conta, e conviene creare sistemi che piacciono a primo impatto.

Un sistema che abbia entrambi i profili non può esistere, ma bisogna trovare un valido **compromesso tra learnability e alta usabilità** nel lungo periodo.

in generale, i trade-off devono essere risolti definendo quali dimensioni di usabilità sono più importanti per un certo sistema.

Trade-off robustezza e usabilità

Il desiderio di evitare all'utente errori può portare a progettare un numero eccessivo di vincoli.

VALUTAZIONE DI USABILITÀ

La valutazione

La valutazione qualitativa riguarda la **qualità del sistema**, limitando la considerazione a livello semantico, portando a informazioni meno oggettive, ma non per questo meno utili.

Si conduce coinvolgendo gli utenti che devono valutare, alla luce di principi di buona progettazione, la cosiddetta euristica.

Si definisce qualitativa perché si procede secondo un metodo.

La valutazione deve avere un approccio empirico e può essere:

- **Longitudinale** si progetta un sistema e si confrontano le varie **versioni** del sistema nel tempo
- **Trasversale** si confrontano due o più **sistemi diversi**

Attività di raccolta

Si esegue un'attività di raccolta di dati sull'usabilità di un sistema rispetto ad uno specifico gruppo di utenti, nello svolgere una certa attività, in un certo ambiente o in un certo contesto di lavoro.

È un'attività:

- **Empirica** coinvolge utenti veri
- **Concreta** eseguita su sistemi reali
- **Sistematica** da eseguirsi con metodo

Si identificano gli obiettivi rilevanti, quindi i relativi casi d'uso, cioè almeno lo scenario principale.

La valutazione può essere:

- **Formativa** aiuta nello sviluppo, rivolta ad identificare problemi o opportunità di miglioramento
- **Riassuntiva** giudizio sul prodotto, è svolta per avere una misura di usabilità e può essere volta a ottenere un valore assoluto o comparativo
- **Absoluta** descrive quanti problemi ci sono
- **Comparativa** produce un confronto tra sistemi
- **Qualitativa** coinvolge esperti ed è volta ad analizzare quello che essi fanno, euristica
- **Quantitativa** basata su test utente e volta a misurare degli attributi di qualità come tempi di esecuzione e anche l'opinione degli utenti su scale psicometriche

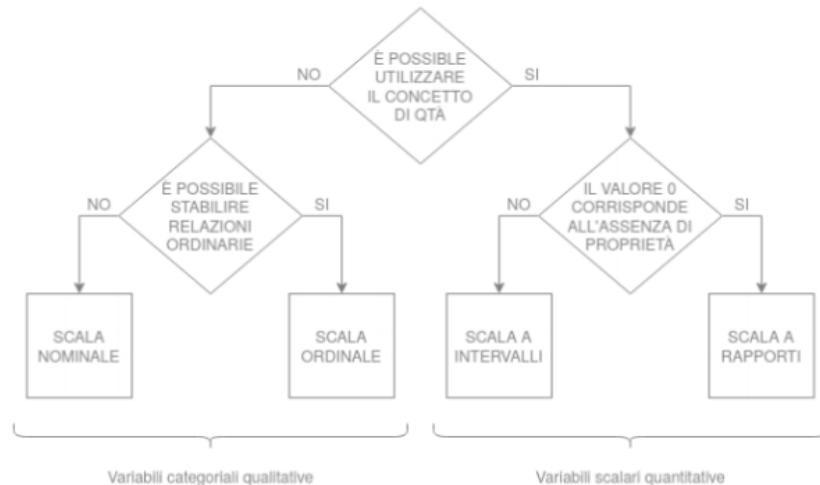
Perché valutare

Per capire il mondo reale, verificare la conformità a degli standard o confrontare progetti e sistemi.

Come si confronta

Si raccolgono dati da un campione di utenti del sistema. I dati sono variabili:

- **numeriche** scalari
- **ordinali** insiemi di numeri in cui vi sono definite relazioni d'ordine totali, ovvero senza uguaglianze
- **categoriche** i valori sono modalità, ossia parole come sì o no, giorno o notte...



Stabiliti quali sono i dati, si confronta come:

1. Si raccolgono i dati
2. Si calcolano le statistiche per descrivere il campione
3. Si cerca di **stimare i parametri** dell'intera popolazione

Non si può conoscere il parametro della popolazione ma solo la statistica di un campione, si può quindi stimare l'intervallo che è ragionevole pensare contenga il parametro vero.

Ciò è definito intervallo di confidenza. Ad esempio, con un **intervallo di confidenza** del 95% c'è una probabilità del 95% che il parametro della popolazione sia compreso tra i due estremi dell'intervallo.

Intervallo di confidenza

È l'intervallo di valori di cui si è abbastanza confidenti che contenga il valore del parametro. Più è alto il livello di confidenza, maggiore è l'intervallo di confidenza t . Maggiore è la numerosità del campione, maggiore è la precisione.

Approccio di Neyman

Si **confrontano gli intervalli di confidenza** di diverse valutazioni quantitative.

Controllando quali intervalli si sovrappongono è possibile capire quali valutazioni hanno carattere simile, ricordandosi sempre l'ampiezza dell'intervallo di confidenza.

Se non si considerano gli intervalli non è possibile dire nulla a livello di significatività né di intera popolazione.

Approccio di Fisher

Si formulano test di ipotesi e si verificano con essi le ipotesi nulle.

L'ipotesi nulla è assunta vera e prevede che non ci sia differenza tra i sistemi valutati o tra lo stesso sistema a tempi diversi o tra strati di utenti.

L'ipotesi nulla indica che **non ci sono differenze statisticamente significative tra A e B**, l'obiettivo è quello di rigettare questa ipotesi attraverso dati campionari.

The null ritual

Procedura statistica per la statistica inferenziale:

1. Si sceglie il test giusto per l'ipotesi e il tipo di dato che si raccoglie
2. Si esegue il test e si calcola il corrispondente livello di significatività osservato dei dati, detto **p-value**
3. Si riportano i risultati in uno stile chiaro e confrontabile, come lo standard APA

Table of error types		Null hypothesis (H_0) is	
		True	False
Decision about null hypothesis (H_0)	Don't reject	Correct inference (true negative) (probability = $1-\alpha$)	Type II error (false negative) (probability = β)
	Reject	Type I error (false positive) (probability = α)	Correct inference (true positive) (probability = $1-\beta$)

Euristica di Cohen

tenere alpha 4 volte più piccolo di beta. Si definisce inoltre **power del test** la capacità di un test di vedere una differenza se c'è, ovvero rigettare l'ipotesi nulla quando è falsa.

P-Value

la probabilità di raccogliere dati come quelli raccolti, assumendo l'ipotesi nulla come **vera**.

Se questa probabilità è **bassa**, è improbabile che la differenza sia solo dovuta al caso:

- se il p-value è inferiore al livello di significatività impostato, allora si può **rigettare** l'ipotesi nulla e ipotizzare che l'ipotesi nulla sia vera
- se il p-value è superiore al livello di significatività il confronto è stato utile lo stesso, ma **non si sono raccolte prove sufficienti** che ci siano reali differenze. Se non si rigetta l'ipotesi nulla, non vuol dire che sia veramente vera

INGEGNERIA DELL'USABILITÀ

Efficienza d'uso

Il sistema deve essere efficiente cosicché sia possibile un alto livello di produttività. Un modo di misurare l'efficienza è **calcolare il tempo necessario ad eseguire il compito correttamente**. Il test è chiamato test T di Student. Si confrontano i tempi medi di esecuzione.

L'ipotesi nulla del t test è che **non c'è differenza significativa** tra le medie di due sotto-campioni, assumendo la loro varianza uguale.

Per poterlo applicare, si dovrebbe avere una **distribuzione normale**, inoltre bisogna decidere se effettuare un test **paired** oppure **unpaired**, se si usano o non si usano gli stessi soggetti.

Per misurare l'efficienza assoluta del sistema A ,bisognerà confrontare la performance degli utenti con l'**optimal execution time** (OET). Si fa il confronto rispetto ad un benchmark. Due valutatori fanno 5 o 6 prove di esecuzione del task ,in modo corretto e a varie velocità (senza esagerare nei tempi), riportando i tempi di esecuzione.

Vengono calcolate **media** (M) e **deviazione standard** (SD). Si effettua un confronto tra la media degli utenti e quella dei valutatori.

Se invece si vuole confrontare più di due sistemi tra loro, si applicherà il **test ANOVA**, che generalizza il test T di Student. Quando i confronti sono a coppia, dividiamo per il numero di confronti che si fanno. Così si diventa più conservativi.

Robustezza o efficacia

Il sistema deve indurre un **basso tasso di errori** (o prevenire proattivamente alcuni), facili da correggere, e con basso impatto.

Un modo di misurare l'efficacia è calcolare il **tasso di errore medio** dei soggetti del campione quando usano il sistema A e il sistema B(A') e confrontarli. Un modo possibile è fare una matrice dove si segnano quali task sono stati completati con successo e quali hanno presentato errori, segnando in quali l'utente è stato aiutato dagli esperti del dominio.

Si confronta per ogni task il numero di errori osservato nel campione di utenti con un tasso ottimale (per convenzione 5% o 1%).

Il numero di errori è una variabile discreta ed è possibile applicare un test delle proporzioni (tabella di contingenza 2x2) per stabilire se il numero di errori di due applicazioni è diverso oltre l'effetto dovuto al caso. Ipotesi nulla: **non c'è differenza tra i valori osservati e i valori**.

Soddisfazione

Il sistema deve essere **gratificante da usare**.

- **Misura oggettiva** si fanno **misurazioni psicofisiche** per misurare livelli di stress o benessere
- **Misura soggettiva** di un osservatore esterno: si conta il **numero di espressioni del viso** legato a frustrazione, incomprensione, perplessità e quelle legate alla soddisfazione.

L'**approccio di Mill** utilizza uno strumento psicometrico, ad esempio un questionario con item a scale ordinali e un uso preparato e responsabile delle tecniche statistiche opportune.

Scala Likert

Questa scala è una tecnica psicometrica di misurazione dell'atteggiamento. Tale tecnica consiste principalmente nel mettere a punto un certo numero di affermazioni, definiti item, che esprimono un atteggiamento positivo e negativo rispetto ad uno specifico oggetto.

Ai rispondenti si chiede su di esse il loro **grado di accordo o disaccordo** con quanto espresso dall'affermazione. Alcuni permettono l'escape option.

Pro

Molto nota e semplice da capire.

Contro

- **Acquiescence bias** misura l'accordo del soggetto con affermazioni dello sperimentatore
- **Central tendency bias** si presenta quando si ha un numero dispari di opzioni, che consentono di fuggire alla valutazione scegliendo opzioni neutre

Le scale a differenziale semantico sono scale ordinali in cui tipicamente vengono segnati gli estremi (ad es. molto buono e molto cattivo) e una serie di opzioni intermedie senza un'ancora.

Si chiamano a differenziale semantico perché, dato un concetto, si prendono gli estremi semantici e le persone possono scegliere qualunque livello intermedio per decidere quanto si sentono più vicini ad un polo o ad un altro.

Una scala Likert è dunque una scala a differenziale semantico in cui sono presenti tutte le ancore (le ancore sono gli elementi della scala).

Net Promoter Score

Indicatore che fa riferimento alla **raccomandabilità di un sistema**.

È estremamente potente e facile e consiste in una sola domanda: "Da 1 a 10 qual è la probabilità che tu raccomandi il prodotto/servizio a un amico o conoscente?"

Il valore del NPS è dato dalla **differenza** tra la percentuale dei promotori, ossia le persone che tendono a rispondere con un numero da 9 a 10, e la percentuale dei detrattori, coloro che danno un voto compreso tra 0 e 6. Come sistema il Net Promoter Score cerca di mitigare abbastanza il bias di accondiscendenza. Notare come il NPS **ignori** tutti i voti 7 e 8.

Pro

Non considera come cardini delle categorie ordinali, inoltre è semplice e permette confronti nel tempo.

Contro

Misura solo la raccomandabilità, e può essere confuso con un voto.

SUMI

Il **Software Usability Measurement Inventory** serve per misurare la qualità di sistemi interattivi dal punto di vista dell'utente cioè la **quality of use** in 5 sottodimensioni: efficiency, affect, helpfulness, controllability, learnability.

Si tratta di 50 affermazioni con cui l'utente può essere d'accordo, in disaccordo o essere indeciso.

Vi possono essere anche delle domande aperte. Tramite un paio di domande di profilazione, il SUMI permette di **stratificare il campione** in persone che si reputano o meno esperte del dominio applicativo.

Pro

Facile da usare, risultati facili da comunicare e ha spazio per i feedback dell'utente.

Contro

non è facile trovare la versione italiana, è piuttosto lungo e richiede un uso prolungato da parte dell'utente.

Il **fatigue bias** è il rumore nelle risposte dovuto alla tendenza di certi rispondenti a rispondere a caso quando sono stanche e il questionario sta durando troppo o facendo domande noiose o inopportune.

Subentra quasi sempre dopo 10-15 minuti in mancanza di qualsiasi incentivo.

AttrakDiff

Questionario per valutare la User Experience. Adotta una **scala a differenziale semantico**, infatti l'utente deve scegliere tra due opposti semantici. Diviso in 4 ambiti:

- **Qualità pragmatica** riguarda l'usabilità in termini di utilità ed efficacia
- **Qualità edonica** riguarda l'estetica, relativo al bello e a ciò che piace
- costrutti di valutazione

Pro

Misura dimensioni difficili e ha una rappresentazione grafica appetibile (snake diagram).

Contro

Non minimizza il centrale tendency bias, non è immediatamente interpretabile e utilizza come indicatore quantitativo la media delle medie dei singoli rispondenti, trasformando posizioni sulla scala in numeri.

UEQ

Lo **User Experience Questionnaire** è un questionario di 26 item a differenziale semantico e 6 sottoscale: attraenza, facilità di comprensione e apprendimento, efficienza e velocità, affidabilità e predicibilità, stimolazione, creatività e innovatività,

Quando lo si usa, spesso si cerca la correlazione tra lo score ottenuto per ogni costrutto e il tempo di esecuzione.

Pro

Breve e con parole semplici, compara aree semantiche importanti ed è ben supportato per lingue e materiale.

Contro

Non minimizza il central tendency bias.

System Usability Scale

Sono scale in cui gli item pari sono l'**opposto** di quelli dispari (voto da 1 a 5).

Si deve quindi fare attenzione nell'attribuire dei punteggi agli item:

- **dispari** il contributo allo score è dato dal valore scelto dall'utente meno 1
- **pari** il contributo è dato dalla differenza tra 5 e il valore scelto dall'utente, questo converte tutte le risposte in un range da 0 a 4.

Poi si sommano tutte le risposte di un singolo utente e moltiplica il risultato per 2.5 per ottenere il valore totale per quell'utente, che varia da 0 a 100. Si fa poi la media di tutti gli score.

Pro

Breve e veloce da compilare, noto ed utilizzato, riesce a minimizzare l'option order bias e l'acquiescence bias.

Contro

Normalizza in centesimi, mescola usabilità e apprendibilità, non si possono fare confronti statisticamente significativi.

Recap sui test

Per osservare differenze statisticamente significative (associate ad un p value inferiore del 5%) anche quando si hanno le stesse mediane si possono utilizzare, in base al contesto, due tipi di test:

- Test parametrici: test T di Student
- Test non parametrici: Mann-Whitney o Wilcoxon (che non confrontano le mediane, il focus dell'analisi è la distribuzione)

Se non si possono confrontare le medie perché i dati sono ordinali si adotta il corrispondente test non parametrico per confrontare le mediane.

I test non parametrici verificano se due distribuzioni sono equivalenti, si focalizzano sui ranghi. Quindi questi test testano la probabilità che dati due punti presi a caso dalle distribuzioni, uno sia più grande dell'altro o viceversa sono uguali.

DARK PATTERNS

Cosa sono

Sono una tipologia di web design che **funzionano contro i bisogni dell'utente**.

Sono elementi UI accuratamente progettati e combinati per **confondere** o **fuorviare** l'utente. In realtà, gli utenti dovrebbero essere messi al centro **user-centered**, anche se questo implica svantaggi in termini di guadagno a breve termine o conversion rate.

Si può creare facilmente un dark pattern "invertendo" le euristiche di Nielsen

Tipologie

1. **Bait and switch** viene cambiato l'effetto di un pulsante al quale l'utente era abituato facendo accadere altro a cui fa comodo all'azienda che venga cliccato
2. **Confirm shaming** è la tattica di far sentire l'utente in colpa per aver selezionato una determinata opzione, che per l'azienda è svantaggiosa
3. **Disguised Ads** gli annunci sono mascherati nella pagina web come se facessero parte della normale navigazione e quindi l'utente può cliccarci sopra pensando non siano annunci
4. **Forced Continuity** avviene quando la prova gratuita di un servizio termina e vengono effettuati addebiti sulla carta di credito senza alcun preavviso; inoltre non viene mai dato un modo semplice per annullare il rinnovo automatico
5. **Friend spam** il sito chiede il permesso per l'email o i social media che utilizzerà poi per trovare amici e contatti, per inviare infine un messaggio che dichiara essere da parte tua
6. **Hidden costs** avviene quando, all'ultimo step del checkout, si scoprono costi che non erano stati dichiarati in precedenza
7. **Misdirection** Design pattern che focalizza l'attenzione dell'utente su una cosa per distorglierla da un'altra, come l'installazione di software aggiuntivi
8. **Price comparison prevention** il sito rende difficile comparare il prezzo di un articolo con un altro, non consentendo una scelta informata
9. **Privacy Zuckering** il design inganna l'utente a condividere inconsapevolmente più dati privati di quanti ne avrebbe voluti condividere
10. **Roach motel** rende molto facile entrare in una pagina o in una situazione, ma molto difficile uscirne
11. **Sneak into basket** durante il percorso d'acquisto, il sito inserisce uno o più articoli aggiuntivi nel carrello senza avvisare l'utente; è un pattern illegale in vari paesi UE
12. **Trick questions** l'utente risponde a domande che a prima vista significano qualcosa ma, se rilette più a fondo, hanno un altro significato, solitamente si usa un linguaggio confuso

Computer as Social Actors

Si ha la tendenza a parlare con le macchine, considerandole un pari umano.

Alcuni esperimenti tra gli anni '90 e 2000 hanno dimostrato che gli utenti sono indotti a comportarsi **come se le macchine fossero umane**, anche se sanno benissimo che non hanno coscienza o non agiscono mosse da motivazioni tipicamente umane.

L'attribuzione di attitudini, interazioni e disposizioni umane alle macchine si chiama **ethopoeia**. Ovviamente non si sta parlando dell'aspetto estetico, non è una antropomorfizzazione.

Basta poco per indurre questa sensazione: messaggi in linguaggio testuale, risposte basate su input predefiniti multipli, attribuzione di ruoli tipicamente umani (come gli assistenti virtuali).

Alcuni comportamenti sociali

Sono di seguito elencati alcuni comportamenti sociali che gli esseri umani riescono ad avere anche con le macchine, e che portano a preferire certe applicazioni.

Reciprocità nella cortesia

La valutazione di un'applicazione non andrebbe mai fatta nel contesto della medesima applicazione, perché l'utente **tende ad essere più gentile** con l'applicazione che l'ha aiutato a raggiungere il suo obiettivo.

Gli utenti si aspettano **educazione e cortesia**, i messaggi di errori devono essere educati e non troppo bruschi.

Infine, si ponga attenzione alla semplice **traduzione**, altrimenti all'utente può sembrare che l'applicazione sia stupida

Reazione a personalità diverse

Gli utenti sono in grado di identificare messaggi che indicano **dominanza o remissività** e sono in grado di avvertire somiglianza tra la personalità dell'applicazione e la loro: gli utenti preferiscono interagire con computer che stabiliscono la stessa tipologia di personalità, infatti il simile apprezza il simile.

Adulazione

Molti sistemi sono normalmente severi o neutri nei loro feedback. potrebbe essere una buona idea **addolcire** il feedback, anche sfiorando l'adulazione, soprattutto nel caso di applicazioni tutorial o per la formazione, in quelle per aumentare la creatività o che impongono compiti ripetitivi e noiosi.

Stereotipi legati al genere

Applicazioni che hanno usato voci femminili per impartire istruzioni in tono perentorio, sono state valutate più negativamente delle stesse applicazioni che usavano voci maschili.

Applicazioni di tutoring con voci maschili sono state valutate più **competenti** rispetto a quelle con voci femminili.

Purtroppo, i **pregiudizi** sono meccanismi umani potentissimi e non si limitano solo al genere: ma bisogna tenerne conto o **sfidarli** apertamente?

Senso di appartenenza

Il comportamento sociale è rivolto al programmatore? No, gli utenti considerano i programmi come **parte del computer**, e non come prodotto di programmatori, e questo rende agli occhi dell'utente il sistema più **usabile**.

INTERVISTE E ANALISI QUALITATIVE

Intervista

Conversazioni dove vengono fatte **domande** per raccogliere **informazioni** su un **argomento** specifico. Sono basate sulla **collaborazione** fra intervistatore e intervistato e sul **rispetto** del tempo, della confidenza e del giudizio, sia esso positivo, negativo o astensionista.

Ha diversi vantaggi:

- Avere informazioni su eventi ai quali non siamo stati partecipi
- Imparare come le persone **percepiscono** le tecnologie
- Integrare o confrontare molteplici **prospettive**

Ma anche svantaggi:

- Le persone non dicono sempre ciò che fanno o pensano
- E non sempre fanno quello che dicono, spesso infatti hanno una **visione idealizzata** di loro stessi, in linea con ciò che l'intervistatore vuole sentirsi dire

A seconda del **contesto**, si possono identificare interviste intense con un focus specifico, talk in action legate ad una specifica attività, informali.

Un'intervista può essere **non strutturata** per esplorare argomenti diversi e definire quello principale, con lo svantaggio di non mantenere spesso la conversazione su argomenti rilevanti, oppure **strutturata** che sono quantitativamente più valide, ma consentono un piccolo margine di azioni a nuovi spunti.

Nelle interviste **semi-strutturate**, invece, l'intervistatore dispone di una guida generale, ma può esplorare altre questioni sollevate dai partecipanti.

Planning

1. Definire il focus
2. Scegliere i partecipanti
 - a. Identificare le categorie di partecipanti possibili
 - b. Identificare i **key informants**, ovvero insider con alta conoscenza
3. Riflettere su location
4. Pensare a quante interviste fare con ogni partecipante
5. Considerare se effettuare registrazioni o trascrizioni
6. inoltre
 - a. reclutare gli intervistati
 - b. Pianificare le interviste
 - c. Trascrivere note o registrazioni
 - d. Analizzare i risultati grezzi

Il copione

Introdurre e far firmare il **consenso informato**, effettuare domande di profilazione e generali, proseguire con quelle più specifiche e semplici verso quelle più controverse, dando infine spazio ai partecipanti per aggiungere le proprie opinioni e idee.

Non bisogna **mai** fornire domande chiuse, più domande contemporaneamente e nascondere giudizi che possono compromettere le risposte.

Focus group

Discussione tra 6-12 partecipanti, per discutere di un argomento specifico, guidata da un **moderatore** che promuove l'interazione e ne mantiene la rilevanza.

L'intervista si limita a pochi specifici argomenti, di solito dura 2 ore circa e permette sia di delineare una concezione comune di un concetto, che sondare opinioni individuali.

Grounded theory

Un metodo generale di analisi comparativa e un insieme di procedure capaci di **generare sistematicamente una teoria fondata sui dati**.

I dati sono **codificati** dalle etichette che raggruppano diverse affermazioni degli intervistati, basati il più possibile sui dati stessi e aggiornati in continuazione.

Core category

La categoria **centrale che rappresenta il concetto organizzatore principale** di un'area di ricerca che può essere individuato induttivamente, procedendo nel lavoro di gerarchizzazione delle categorie emerse dai dati. Corrisponde a quella più **potente analiticamente**.

Processo di analisi

1. **Codifica iniziale** rimane molto legata ai dati, legge e scompone i testi in tutte le direzioni che questi consentono, mira a identificare delle prime etichette concettuali
2. **Codifica focalizzata** mira a trasformare i primi concetti in categorie interpretative ancora abbozzate e tracciare i primi collegamenti di senso fra esse
3. **Codifica teorica** la definizione delle categorie assume la sua forma più compiuta e la teorizzazione procede verso l'identificazione delle categorie centrali, i concetti chiave attorno ai quali si organizzerà la teoria

Dopo aver messo a punto le categorie, poste in relazione gerarchica, e aver individuato la core category, si tratta di integrare la teoria e tematizzarla intorno ai suoi assi concettuali empiricamente emersi.

Impatto dell'AI

Sono dei modelli per valutare l'impatto **etico** e la qualità dei sistemi di supporto decisionale. Un primo esempio di proposta di linee guida si ha con gli **AI Principles** del Future of Life Institute del 2017.

Da allora si ha una proliferazione dei principi per l'IA, che ha portato ad una maggiore riflessione su questi temi, ma anche a molta **vaghezza** e difficoltà di applicazione nella realtà dei principi.

Linee guida etiche per un'IA affidabile

Definite da un gruppo di esperti di alto livello sull'intelligenza artificiale per l'Unione Europea, definiscono in poche parole cosa si intende per IA affidabile: si basa su tre componenti che dovrebbero essere presenti durante l'intero ciclo di vita del sistema:

- **Legalità** l'IA deve ottemperare a tutte le leggi e ai regolamenti applicabili
- **Eticità** l'IA dovrebbe assicurare l'adesione a principi e valori etici
- **Robustezza** dal punto di vista tecnico e sociale poiché, anche con le migliori intenzioni, i sistemi di IA possono causare danni non intenzionali

Intervento e sorveglianza

I sistemi di IA dovrebbero sostenere l'autonomia e il processo decisionale umani, come prescrive il principio del **rispetto dell'autonomia umana**. A tal fine essi devono agire come catalizzatori di una società democratica, prospera ed equa, sostenendo l'intervento degli utenti e promuovendo i diritti fondamentali, e **devono consentire la sorveglianza umana**.

Robustezza tecnica e sicurezza

Per garantire la robustezza tecnica è necessario che i sistemi siano sviluppati con un approccio di **prevenzione dei rischi** e in maniera tale che si comportino in maniera attendibile secondo le previsioni, riducendo al minimo i danni non intenzionali e imprevisti e prevenendo danni inaccettabili. Ciò dovrebbe valere anche per gli agenti umani che possono interagire in maniera **contraddittoria** con il sistema

Si dovrebbe inoltre garantire l'integrità **fisica** e **psichica** degli esseri umani.

Privacy e governance dei dati

Per prevenire danni alla riservatezza occorre tra l'altro un'adeguata governance dei dati che riguardi la **qualità** e l'**integrità** degli stessi.

Trasparenza

Requisito strettamente connesso al **principio dell'esplicabilità**, e comprende la trasparenza degli elementi pertinenti per un sistema di IA.

Diversità, non-discriminazione ed equità

Occorre che l'inclusione e la diversità siano permesse durante l'intero ciclo di vita del sistema. Riguarda anche la **parità di trattamento** e la **parità di accesso**.

Benessere sociale e ambientale

Anche la società in generale, altri **esseri senzienti** e l'**ambiente** dovrebbero essere considerati come portatori di interessi durante l'intero ciclo di vita del sistema di IA.

La sostenibilità e la responsabilità ecologica dei sistemi di IA dovrebbe essere incoraggiata.

Accountability

Occorre mettere in atto meccanismi che garantiscono la **responsabilità**, che ne permettano la **verificabilità** e la riduzione al **minimo** degli effetti negativi.

Valutazione dell'impatto algoritmico

Un processo interdisciplinare che mette in comunicazione gli esperti tecnici con giuristi, sociologi, filosofi e altre figure extra-tecniche.

Si basa su una serie di processi:

- **Descrizione del sistema** in aspetti tecnici e ambito di applicazione
- **Valutazione del rischio** si intende qualcosa di intrinseco al sistema, ovvero la vulnerabilità, presenza di bias, eventuali errori nei modelli, ma anche la **sensibilità** dell'ambito di applicazione
- **Valutazione dell'impatto** richiede una prospettiva ampia e di lungo termine, facendo riferimento alla società, a possibili conseguenze inaspettate
- **Identificazione di misure di mitigazione** riduzione del rischio posto dal sistema attraverso modifiche circa il modello, i dati, le modalità di utilizzo o l'ambito di utilizzo
- **Revisione periodica** elemento fondamentale a causa dell'imprevedibilità dei sistemi basati su autoapprendimento e a causa del continuo cambiamento di dati e contesto di applicazione

Valutazione esterna

Un **audit esterno** è un processo attraverso il quale una terza parte indipendente esamina l'impatto e, per quanto possibile, il funzionamento di un sistema algoritmico al fine di individuare **potenziali anomalie o pratiche che potrebbero essere ingiuste** o dannose nei confronti di gruppi protetti o della società nel suo complesso.

L'impatto viene spesso valutato attraverso **analisi sistematiche delle popolazioni interessate**, fonti secondarie e database di sistemi di raccolta, infatti non hanno a disposizione l'accesso al sistema completo.

VALUTAZIONE DELLA QUALITÀ IA

Accuratezza

Non è sufficiente per questioni legate alla replicabilità, attendibilità dei dati e validità o utilità dei risultati.

L'accuratezza dimostrata sui dati di training spesso **non si traduce** in accuratezze simili con dati del mondo reale. Servono quindi procedure rigorose di **validazione esterna** attraverso vari database tratti dal mondo reale.

Inoltre, l'accuratezza dipende fortemente dalla qualità dei dati di training, che spesso nascondono bias o errori in fase di raccoglimento o **data labeling**. Per ottenere label attendibili, sono necessari più valutatori esperti.

Infine, l'accuratezza non misura di per sé l'effettiva utilità di un sistema di supporto decisionale né il suo **impatto sulle decisioni umane**.

DSS Quality Assessment

Calcolatore online per la valutazione di qualità di sistemi di supporto decisionale IA che considera diverse dimensioni complementari come affidabilità, robustezza, calibrazione, utilità e impatto sull'interazione umana.

Robustezza

Valutare la robustezza vuol dire capire il livello di **adattabilità** e **coerenza** del sistema anche quando utilizzato in ambiti non familiari.

L'obiettivo principale è analizzare come un sistema AI reagisce a dati differenti da quelli di training.

Un **sistema robusto** mantiene un livello di performance comparabile anche quando esposto a dati differenti.

I due pilastri della robustezza sono:

- **Valutazione delle performance** misurata a partire da tre metriche-chiave quali potere di discriminazione, utilità e calibrazione
- **Valutazione della similarità dei dati** se un sistema peggiora la propria performance solo marginalmente di fronte a dati diversi, allora lo si può definire robusto. Spesso è difficile perché i dati usati sono spesso **multi-dimensionali**, inoltre dovrebbero essere né troppo simili né troppo differenti

Calibrazione

L'analisi della calibrazione ha un ruolo cruciale nel **capire se le previsioni del SSD sono sufficientemente allineate con le frequenze osservate**, permettendo quindi al decisore umano di prendere in considerazione gli output del sistema, fare ragionamenti probabilistici e basati sul rischio.

Può essere:

- **Class-wise** deve essere calibrato per ogni classe in cui il SSD opera, spesso un problema per la presenza di **database non bilanciati**
- **Local** riguarda la calibrazione all'interno di specifici range di probabilità, ovvero che non ci sia una tendenza a sottostimare o sovrastimare
- **Sovra-confidenza e sotto-confidenza** bisogna capire anche la natura degli errori di calibrazione, un SSD con eccessiva confidenza potrebbe prevedere degli outcome come molto più probabili di quanto siano effettivamente, d'altra parte una bassa confidenza potrebbe sottostimare i rischi presenti

Utilità

L'utilità è la **capacità** di un sistema di supporto decisionale di **facilitare la presa di decisioni che portino a risultati ottimali a costi ridotti**.

In sostanza, è utile se le sue rivelazioni corrette superano **in modo sostanziale** i suoi eventuali errori. Tuttavia, si devono considerare elementi cruciali:

1. **Gerarchia** si dovrebbe in particolare gli errori che portano a conseguenze più gravi
2. **Assistenza in casi cruciali** le decisioni più significative dovrebbero evitare in particolare gli errori associati con conseguenze più gravi
3. **Confidenza nelle predizioni** un SSD ottimale non dovrebbe tirare a indovinare, ma fornire un supporto coerente e affidabile

Per esempio, un SSD che fornisce supporto nella diagnosi/cura di un paziente. Curare comporta costi (medicine, personale medico) che hanno alti benefici (guarigione). **Trattamenti scorretti** comportano costi (medicine, personale medico) che hanno gravi conseguenze (mancata guarigione e/o danni alla salute), così come le **mancate diagnosi**. Quindi, un SSD utile avrebbe come risultato minori costi (cure ben ponderate, nessun danno) rispetto a scenari dove il SSD non è presente.

Attendibilità

Si deve misurare l'accuratezza degli **annotatori**, comparando i label da loro forniti rispetto al risultato di un ipotetico **voto di maggioranza** fra tutti gli annotatori, ovvero l'**interrater agreement**.

Tre dimensioni sono necessarie per quantificare il livello di attendibilità:

1. **Accordo** quanto un gruppo di annotatori è d'accordo sul label dato ad un caso
2. **Confidenza** quanto uno specifico annotatore è sicuro di uno specifico label
3. **competenza** quanto un annotatore è accurato, ottenuta in base al livello di accordo con gli altri rater

Interazione umana

Approfondimento sul processo decisionale **Human-first**, che segue il processo: prima giudizio umano, esposizione al giudizio dell'IA, decisione umana finale.

Osservare i cambiamenti di decisione pre e post esposizione al giudizio di un sistema IA ci permette di capire **se e quali bias cognitivi** siano presenti, in termini di affidamento al sistema di supporto (i **reliance patterns**, o pattern di affidamento umano su IA).

Bias di automazione

La tendenza a fare eccessivo affidamento sull'IA, anche a discapito del proprio buonsenso.

Avversione algoritmica

La tendenza ad opporsi a ogni giudizio dell'IA, facendo troppo affidamento sul proprio o, in casi estremi, anche a discapito del proprio buonsenso.

Tecnovigilanza

L'idea che sia **necessario** valutare adeguatamente e su più dimensioni di qualità i sistemi che intendiamo introdurre nelle attività umane, ad esempio e soprattutto per le seguenti dimensioni: etica, robustezza, similarità, calibrazione, utilità, affidabilità, interazione umana.