Interazione Persona-Calcolatore

Rapporto fra HCI ed AI

Prof.ssa Daniela Fogli

Dipartimento di Ingegneria dell'Informazione

Cambi di prospettiva

- Artificial Intelligence
  → Human-Centered Artificial Intelligence

- Human-centered design con tecniche tradizionali
  → Human-centered design con l'aiuto dell'AI

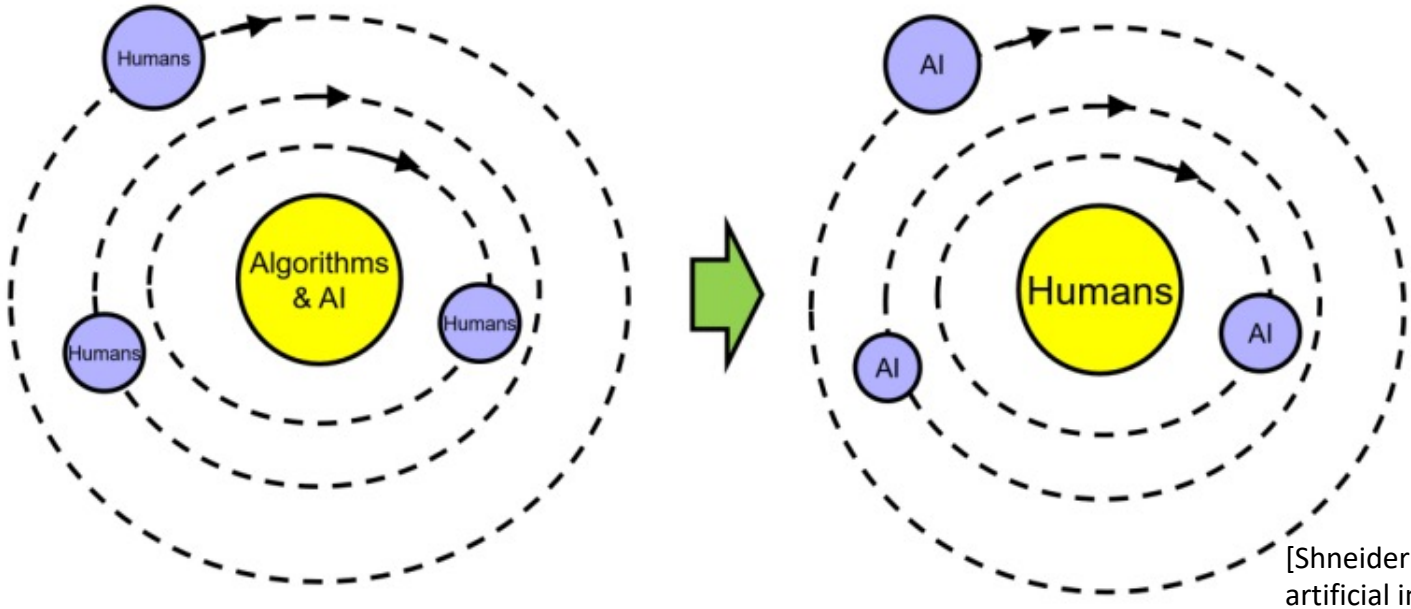
- Human-computer interaction
  → Human-AI interaction

BEN SHNEIDERMAN

Human-Centered AI

Computer scientists should build devices to enhance and empower—not replace—humans.

Human-Centered Artificial Intelligence



[Shneiderman, B. (2020). Human-centered artificial intelligence: Three fresh ideas.]

HCAI aims is to **amplify, augment, enhance,** and **empower** users, by increasing their performance

Human-Centered Artificial Intelligence (HCAI)

Due aspetti chiave

Processo: HCAI si basa sui metodi di HCI per la progettazione di sistemi interattivi come user research, usability testing, processo iterativo and continua valutazione dei sistemi

Prodotto: I sistemi progettati con HCAI sono concepiti come *supertool* che amplificano, aumentano, potenziano, e migliorano le prestazioni umane

Scenario: Consegna di pasti agli anziani

- **AI tradizionale**
  - Robot mobili o droni che portano il cibo in casa
  - Anziani trattati come destinatari passivi dell'automazione

- **HCAI**
  - Addetti alle consegne stabiliscono un contatto sociale con gli anziani
  - Gli algoritmi di AI potrebbero aiutare a far coincidere interessi e personalità di addetti alle consegne e anziani, a pianificare percorsi efficienti, …

AI vs HCAI

- AI tradizionale mira allo sviluppo di **sistemi autonomi**

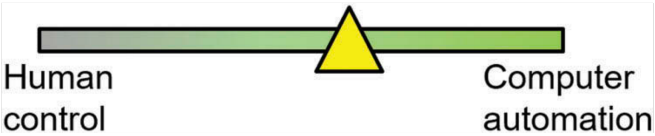
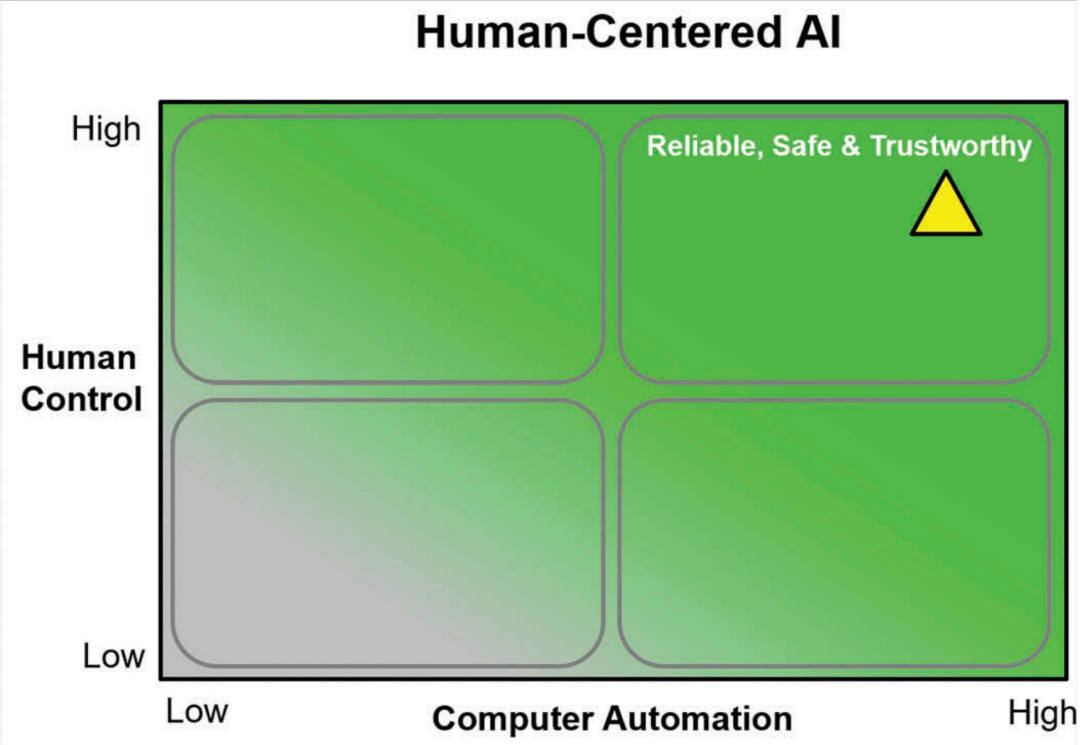


Figure 1. One-dimensional thinking suggest that designers must choose between human control and computer automation.

(Shneiderman, 2020)

- **AI-based systems:** "software designs that give users high levels of understanding and control over their AI-enabled devices" (Shneiderman, 2021)

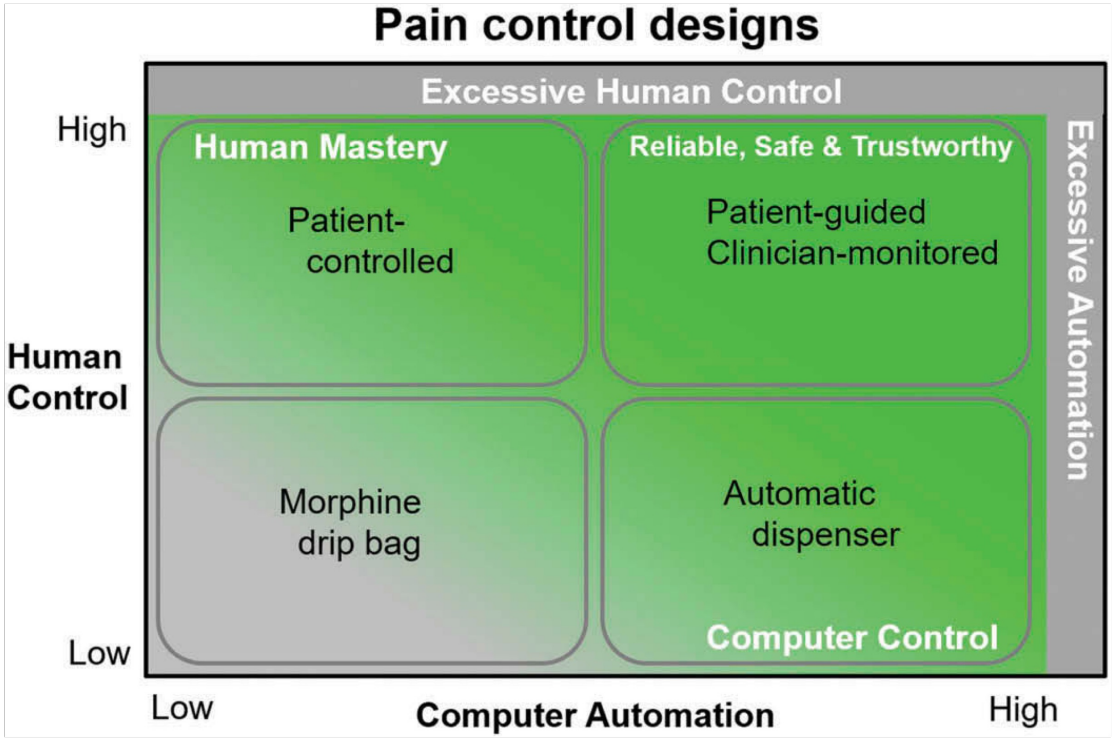
The two-dimensional HCAI framework



(Shneiderman, 2020)

- Un sistema **reliable** produce le risposte attese/corrette
- Un sistema **safe** previene conseguenze catastrofiche: evita un controllo umano o automazione eccessivi
- Un sistema **trustworthy** è un sistema che merita fiducia: è valutato da un ente di supervisione indipendente e autorevole

Esempio nel dominio medico

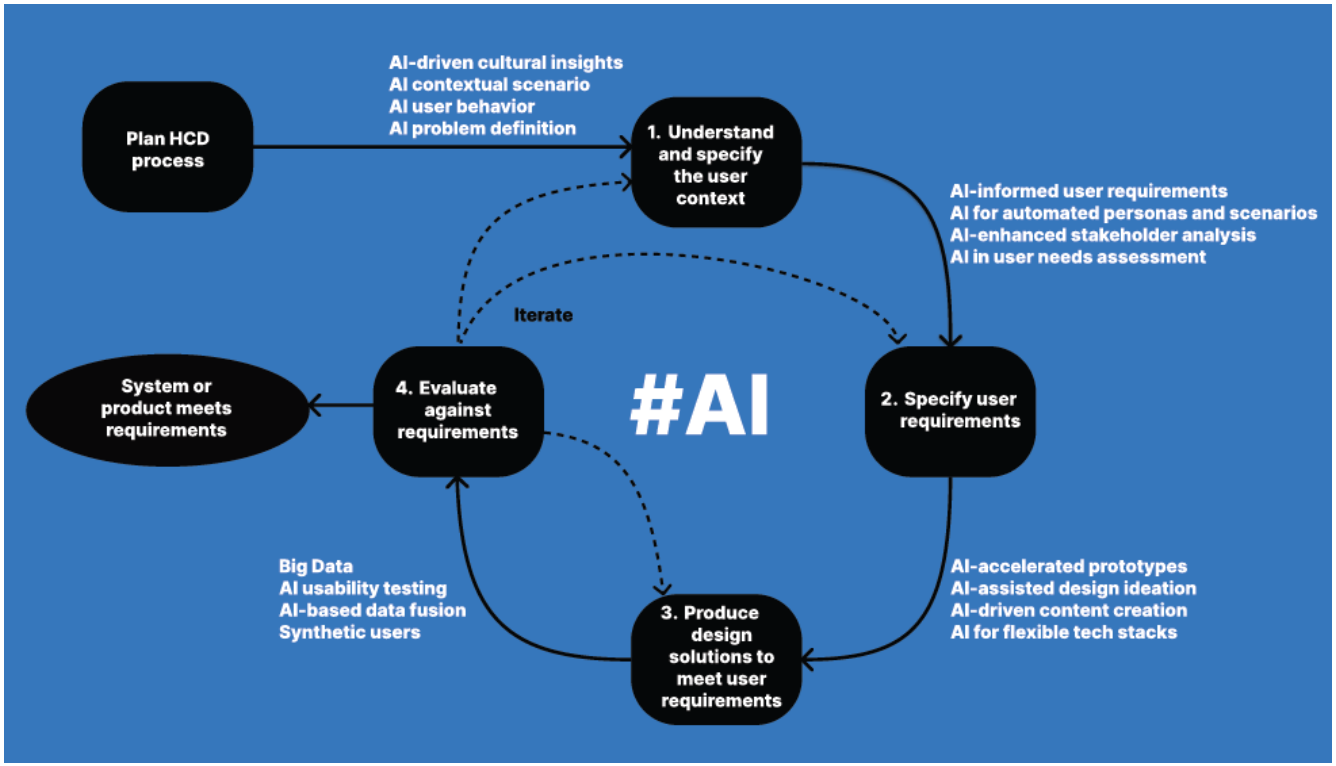


(Shneiderman, 2020)

Artificial Intelligence (AI) vs Intelligence Augmentation (IA)

Quando è utile…	Quando è meglio…
Automation	**Augmentation**
Le persone non sanno come fare una certa cosa	Le persone si divertono a svolgere quel compito o si sentono responsabili
Le persone non possono fare una certa cosa	Situazioni ad alto rischio
Il compito è noioso, ripetitivo, o pericoloso	Preferenze personali complesse

Come potrebbe cambiare la HCD con l'AI



The human-centered design cycle according to ISO 9241-210:2019, with suggestions on how AI may be used in different steps.

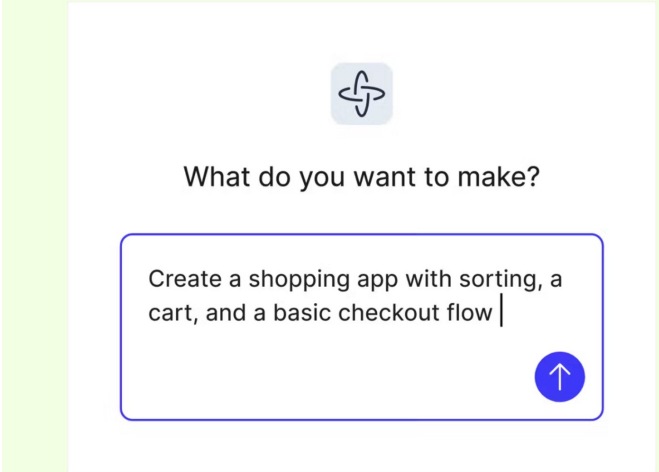
Albrecht Schmidt, Passant Elagroudy, Fiona Draxler, Frauke Kreuter, and Robin Welsch. 2024. Simulating the Human in HCD with ChatGPT: Redesigning Interaction Design with AI. interactions 31, 1 (January - February 2024), 24–31.

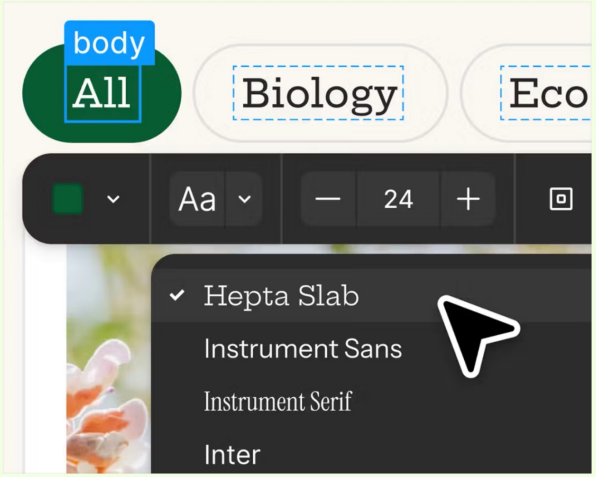
Progettazione dell'interazione

- Includere strumenti di AI come **design partner**, che possono espandere la nostra **creatività**
- Strumenti per **generare persona** e **scenari**: e.g., User Persona Generator (ChatGPT), PersonAI (Figma)
- Strumenti per **generare storyboard**: e.g., AI Storyboard Generator, Dall-E Image Creator
- Inserire un **partecipante virtuale** in un focus group o far generare le domande per mantenere la discussione focalizzata sugli aspetti rilevanti

AI come collaboratore alla prototipazione

- AI per **generare prototipi** a bassa fedeltà (e.g., Figma Make), in modo che sia poi più semplice produrre quelli ad alta fedeltà (prototipazione orizzontale)
- Permette inoltre di creare **diversi prototipi** fra cui scegliere
- AI per **generare codice** (in HTML, CCS, javascript, …) per mostrare una funzionalità (prototipazione verticale)





AI nella valutazione di usabilità

- Usare AI per **generare test di compito e di scenario**
- **Synthetic evaluation**: usare AI generativa per simulare gli utenti
  - Diverse start-up: Synthetic Users, OpinioAI
  - Sono metodi funzionalmente e moralmente appropriati?
    - Chi è a favore sostiene che si risparmia **tempo**, o comunque che consentono di fare **testing a più vasta scala**, permettendo di simulare diversi tipi di partecipanti (**inclusione**), consentendo di valutare in sicurezza **sistemi safety-critical**
    - Altri sostengono che rischiano di **simulare solo certi gruppi** di utenti (WEIRD - Western, Educated, Industrialized, Rich, Democratic) più facili da rappresentare con i dati di training
    - Valutando pro e contro, si può pensare di **combinarli** con metodi tradizionali

Implicazioni etiche

- Valutare bene i **bias e i limiti** di usare AI in HCD
- AI non può avere **intuizione, empatia e comprensione** degli utenti, quindi difficilmente riesce a emulare alcune situazioni relative a certi sistemi (e.g., social VR, empathic computing)
- Adottare approcci **multi-metodo**
- **Combinare** user study a piccola scala (pochi utenti veri) con user study a grande scala (utenti simulati)
- **AI come strumento** per migliorare il lavoro dei progettisti non per sostituirli

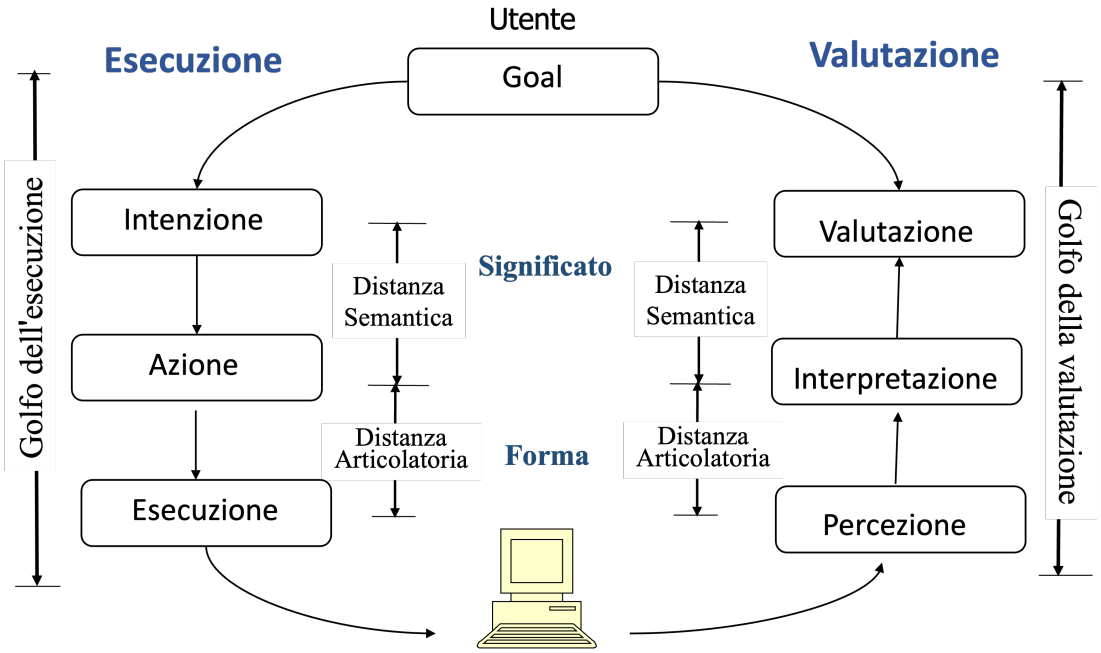
Interazione con i sistemi AI-based

- Combinare
  - **Interazioni implicite**: locazione, identità di oggetti e persone, gesti, espressioni facciali, dati biometrici
  - **Dati**: storia dell'interazione e preferenze dell'utente o del gruppo di utenti
  - **Interazioni esplicite** multi-modali: comandi vocali, testo, disegni a mano libera
- Per **personalizzare** i sistemi e la loro interfaccia utente (**sistemi adattivi**)
- Ma **condivisione di dati sensibili** con il sistema di AI (e.g., stato di disabilità, interazioni passate)
- **Trade-off** fra i benefici della personalizzazione (apprendibilità, accessibilità) e i problemi sociali (e.g., filter bubbles, privacy)

Sistemi AI-based e i golfi …

*Far capire **cosa si può fare con il sistema di AI**, chiarire come esprimere le proprie intenzioni*

*La parola **«intelligenza»** può indurre un errato modello mentale*



*Capire se il sistema ha **completato il task correttamente***

*Può servire un certo livello di **AI literacy***

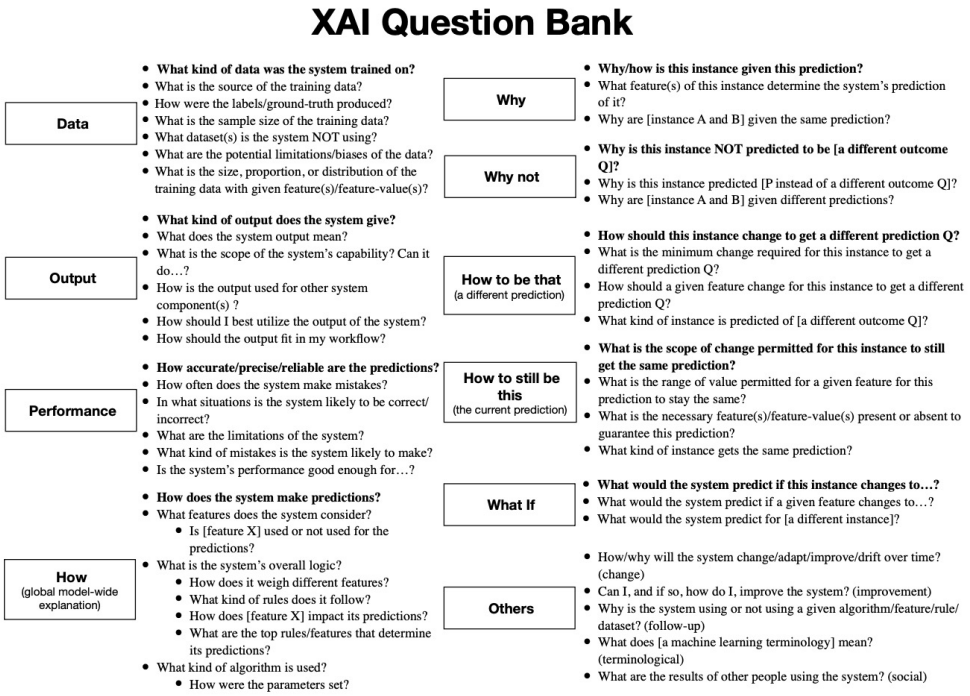
I rischi di certe metafore

- **Antropomorfismo** degli agenti intelligenti, che parlano in linguaggio naturale, eventualmente tramite avatar
- Anche se magari più facili da usare, più piacevoli e coinvolgenti, possono portare a **overtrust** o a coltivare relazioni non salutari con questi tipi di sistemi
- È necessario studiare i loro **impieghi realistici a grande scala** perché alcuni problemi (dipendenze, problemi di salute mentale, …) possono non emergere negli studi di laboratorio (ma questo ha comunque un problema etico)
- Studiare **l'impatto** dell'AI antropomorfica **sui più vulnerabili** (bambini, anziani, persone con disabilità cognitiva, …)
- **Responsible AI** basata sui metodi di HCI e UX design da impiegare lungo tutto il ciclo di vita adottando una **prospettiva socio-tecnica** [Liao et al., 2024]

Prospettiva socio-tecnica della Responsible AI

- *"AI systems are inherently socio-technical in nature, meaning they are influenced by societal dynamics and human behavior.* ***AI risks*** *– and benefits – can emerge from the* ***interplay of technical aspects combined with societal factors*** *related to how a system is used, its interactions with other AI systems, who operates it, and the social context in which it is deployed"* [NIST AI Risk Management Framework]
- Esempio:
  - Per perseguire il principio di **explainability** spesso si adottano soluzioni tecniche che forniscono output complessi su come il modello lavora
  - Occorre creare spiegazioni facili da comprendere (e.g., con testo, grafica, immagini, …), che diminuiscano information overload e permettano agli utenti di accedere progressivamente ai dettagli
  - Il "Question-driven design process for explainability" [Liao et al., 2021] è una **metodologia user-centered** proposta per selezionare le tecniche di explainability e come progettarle

Question-Driven Design Process



Question	Explanations	Example XAI techniques
Global how (global model-wide)	• Describe the general model logic as feature impact*, rules+ or decision-trees• (sometimes need to explain with a surrogate simple model) • If a user is only interested in a high-level view, describe what are the top features or rules considered	ProfWeight*+•, Global Feature Importance*, PDP*, BRCG+ , GLRM+ , Rule List+ , DT Surrogate•
Why	• Describe what key features of the instance determine the model's prediction of it* • Describe rules+ that the instance fits to guarantee the prediction • Show similar examples• with the same predicted outcome to justify the model's prediction	LIME*, SHAP*, LOCO*, Anchors+, ProtoDash•
Why not	• Describe what changes are required for the instance to get the alternative prediction and/or what features of the instance guarantee the current prediction* • Show prototypical examples+ that had the alternative outcome	CEM* , Prototype counterfactual+ , ProtoDash+ (on alternative class)
How to be that (a different prediction)	• Highlight features that if changed (increased, decreased, absent, or present) could alter the prediction* • Show examples with minimum differences but had a different outcome than the prediction+	CEM*, Counterfactuals*, DiCE+
How to still be this (the current prediction)	• Describe features/feature ranges* or rules+ that could guarantee the same prediction • Show examples that are different from the particular instance but still had the same outcome	CEM*, Anchors+
What if	• Show how the prediction changes corresponding to the inquired change	PDP, ALE, What-if Tool
Performance	• Provide performance metrics of the model • Show uncertainty information for each prediction • Describe potential strengths and limitations of the model	Precision, Recall, Accuracy, F1, AUC Uncertainty Qauntification 360 FactSheets, Model Cards
Data	• Document comprehensive information about the training data, including the source, provenance, type, size, coverage of population, potential biases, etc.	FactSheets, DataSheets
Output	• Describe the scope of output or system functions. • Suggest how the output should be used for downstream tasks or user workflow	FactSheets, Model Cards

[Liao et al., 2021]

Valutazione dei sistemi AI-based

- Esistono **metriche** consolidate e metodi per valutare singoli algoritmi di AI
- Ma poca attenzione alla **valutazione di sistemi AI-based in contesti reali** e il loro uso da parte di utenti non esperti di AI
- La generalità e la complessità di certi sistemi è impossibile da valutare esaustivamente, anche a causa della loro **adattività** ed **evoluzione dei modelli** che possono cambiare il comportamento «on-the-fly» (nonché al **non-determinismo** di certi approcci)
- Alcuni sono progettati per collaborare con le persone (Human-AI Teaming / Symbiotic AI) e occorre valutare il loro impatto
- Cambiano le **dimensioni di usabilità**: trasparenza, directability, efficienza del sistema human-AI, conformità agli standard sull'etica

Le sfide della valutazione

- Comprendere come attuare un **monitoraggio** continuo dei sistemi che una volta installati evolvono nel tempo
- Sviluppare framework per valutare la **sicurezza** dei sistemi di AI che fanno azioni nel mondo (agenti)
- Creare metodi per aumentare la **trasparenza** degli algoritmi di machine learning
- Sviluppare metodologie di valutazione che considerano come l'essere umano si rapporta alle capacità dell'AI
  - *"move beyond purely algorithmic considerations and engage in participatory design processes that prioritize the voices and experiences of end users, ensuring that the **AI solutions are truly aligned with their real-world needs and challenges**"*
    [AAAI 2025 Presidential Panel on the Future of AI Research]
- Comprendere i **trade-off** fra le varie dimensioni di usabilità

Strumenti di valutazione ad-hoc

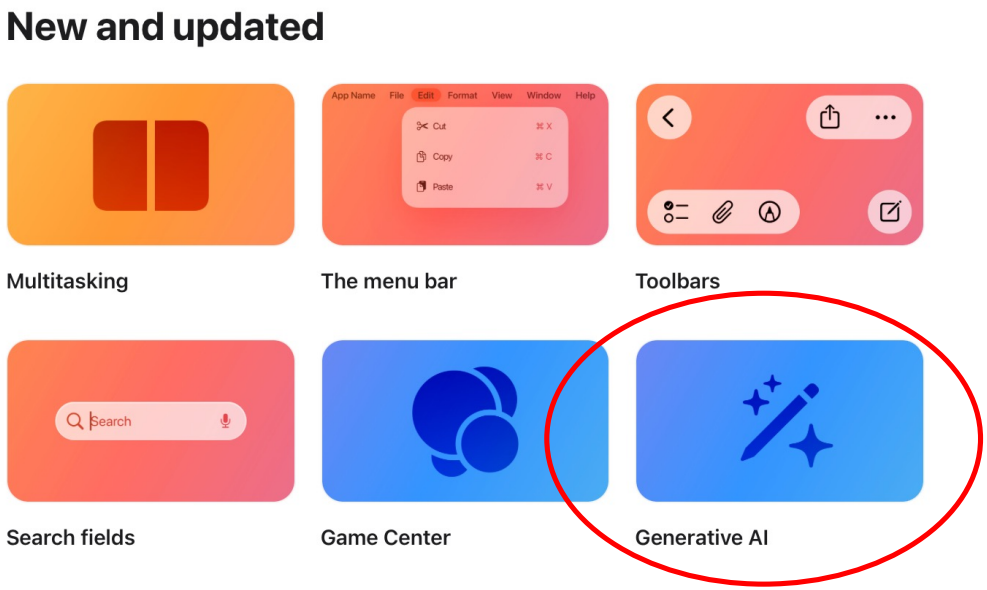
- Esempio: Bot Usability Scale (BUS) [Borsci et al., 2023]

Table 1 The original (English) version of BUS 15. Each item is assessed on a 5-point Likert scale from 1 ("strongly disagree") to 5 ("strongly agree")

Factor	Item
1—Perceived accessibility to chatbot functions	1. The chatbot function was easily detectable
	2. It was easy to find the chatbot
2—Perceived quality of chatbot functions	3. Communicating with the chatbot was clear
	4. I was immediately made aware of what information the chatbot can give me
	5. The interaction with the chatbot felt like an ongoing conversation
	6. The chatbot was able to keep track of context
	7. The chatbot was able to make references to the website or service when appropriate
	8. The chatbot could handle situations in which the line of conversation was not clear
	9. The chatbot's responses were easy to understand
3—Perceived quality of conversation and information provided	10. I find that the chatbot understands what I want and helps me achieve my goal
	11. The chatbot gives me the appropriate amount of information
	12. The chatbot only gives me the information I need
	13. I feel like the chatbot's responses were accurate
4—Perceived privacy and security	14. I believe the chatbot informs me of any possible privacy issues
5—Time response	15. My waiting time for a response from the chatbot was short

Human-AI Interaction

- Linee guida di Microsoft e di Google su Human-AI interaction
- Linee guida di Apple su Generative AI



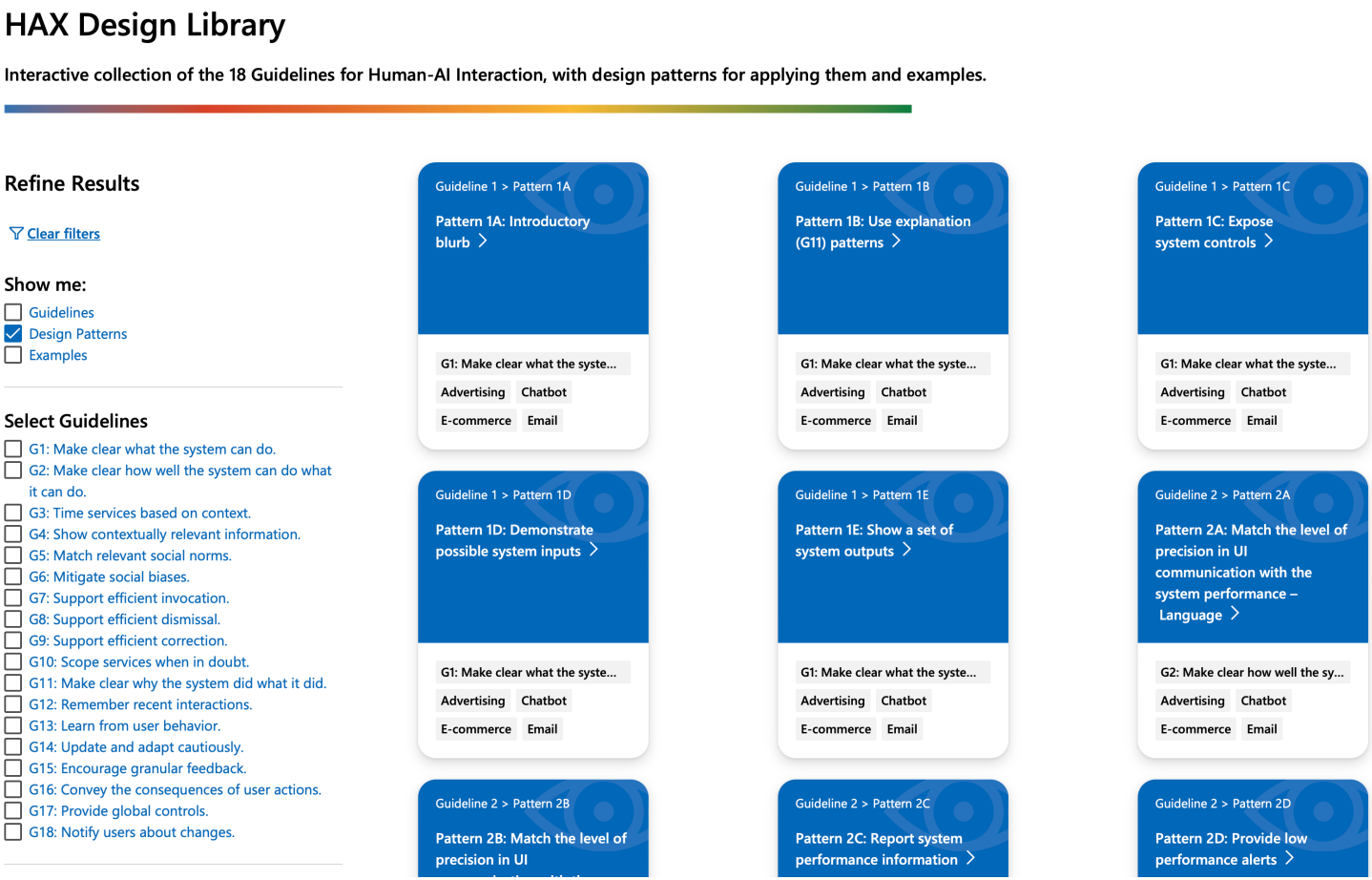
Guidelines for Human-AI Interaction

(Amershi et al., 2019) - CHI 2019 Honorable Mention Award

https://www.microsoft.com/en-us/haxtoolkit/ai-guidelines/

Design pattern per 8 delle 18 guideline

https://www.microsoft.com/en-us/haxtoolkit/library/



Un pattern

Pattern 1B: Use explanation (G11)

Problem

The user needs to understand what the system can do.

Solution

Provide explanations that enable users to gain insights into system capabilities. Explanations help user understanding because they expose relationships between system inputs and outputs (see G11 patterns).

Use when

- The user is already familiar with the type of system or feature.
- The system is capable of generating explanations.
- See patterns G11 A-G for when to use each G11 pattern.

How

Use Guideline 11 patterns to explain why the system did what it did.

Provide insights into what the system can do by using G11-B: Global Explanations.

For providing insights into specific system behaviors, use G11-A, Local Explanations.

User benefits

- Minimizes the number of explanations throughout the system.
- Global explanations enable users to gain insights into both *how* the overall system works and *what* it can do.
- Local explanations support learning by doing: They enable the user to learn about system capabilities in the process of using the system.

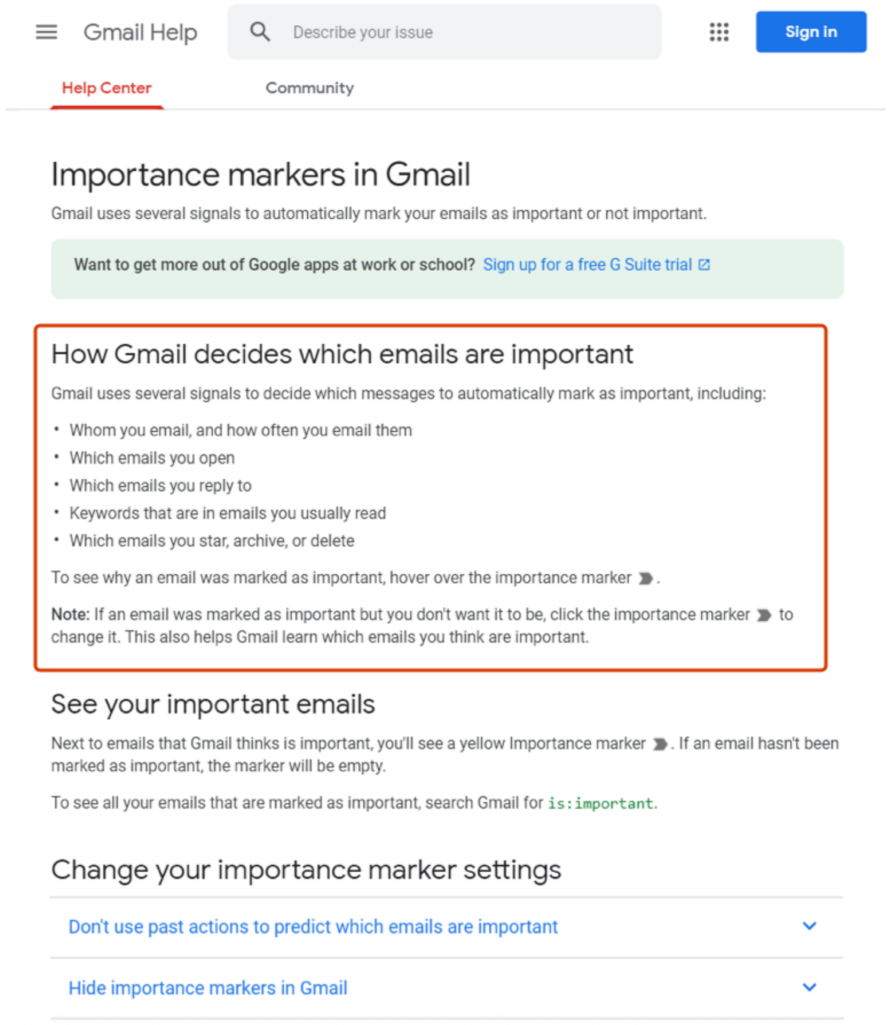
Common pitfalls

- Low discoverability of explanations made available on request.
- Local explanations used alone may not provide a coherent view of what the system can do.
- Global explanations may exclude important system capabilities.
- Explanations that provide insights into a system's granular behaviors may not make clear the system's high-level capability.
- The pattern implementation creates expectations that the system can do more than it is capable.

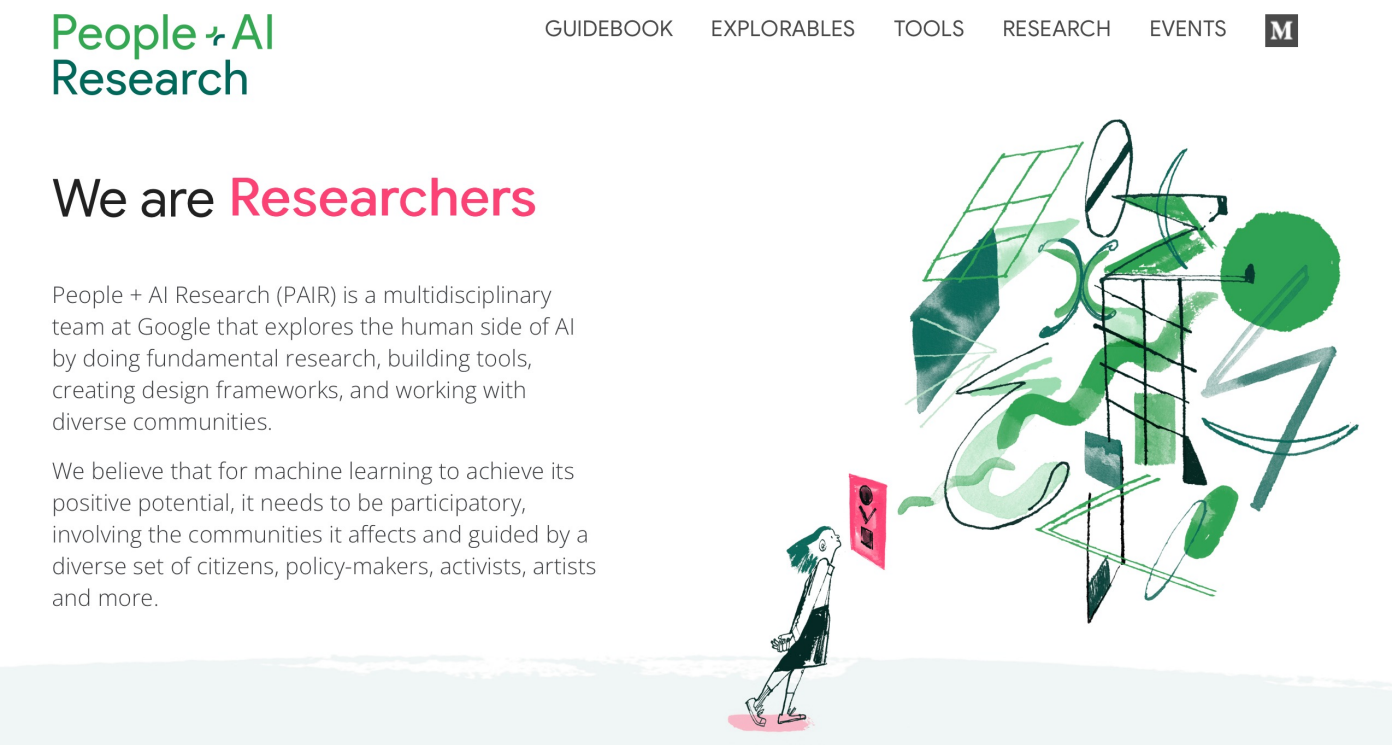
Note: Over-inflated user expectations have been shown to cause frustration and even product abandonment.

Esempio

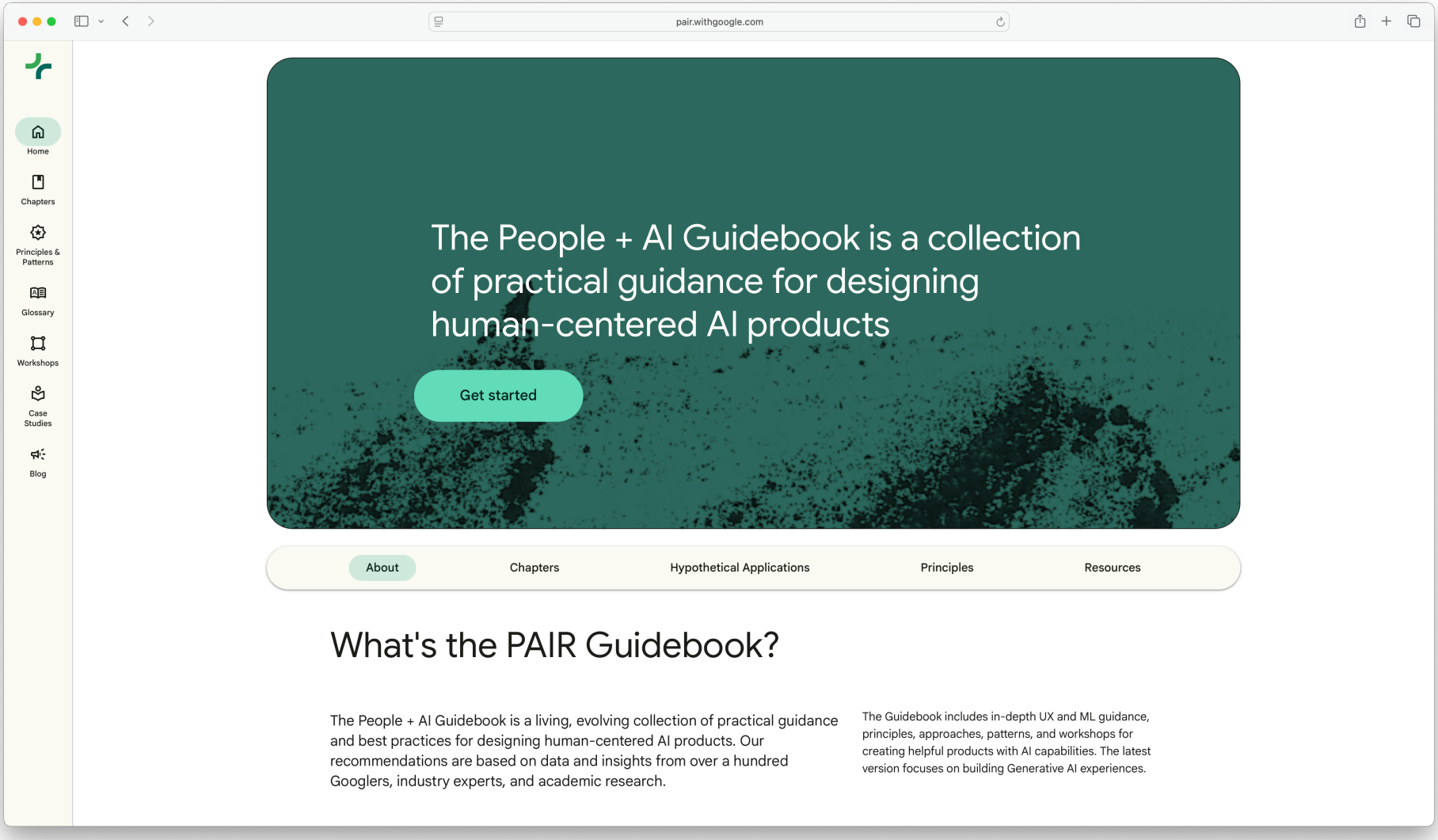
Gmail sets expectations for what the system can do (Guideline 1) by using explanation patterns (Pattern 1B) in its global explanation (Pattern 11B) of how it determines email importance. *Image captured July 2020.*



HCAI in Google



https://pair.withgoogle.com


pair.withgoogle.com
Home
Chapters
Principles & Patterns
Glossary
Workshops
Case Studies
Blog
The People + AI Guidebook is a collection of practical guidance for designing human-centered AI products
Get started
About
Chapters
Hypothetical Applications
Principles
Resources
What's the PAIR Guidebook?
The People + AI Guidebook is a living, evolving collection of practical guidance and best practices for designing human-centered AI products. Our recommendations are based on data and insights from over a hundred Googlers, industry experts, and academic research.
The Guidebook includes in-depth UX and ML guidance, principles, approaches, patterns, and workshops for creating helpful products with AI capabilities. The latest version focuses on building Generative AI experiences.

Principles

USER AUTONOMY

Design for the appropriate level of user autonomy

DATA & MODEL ALIGNMENT

Align AI with real-world behaviors

EVOLVING SAFETY

Treat safety as an evolving endeavor

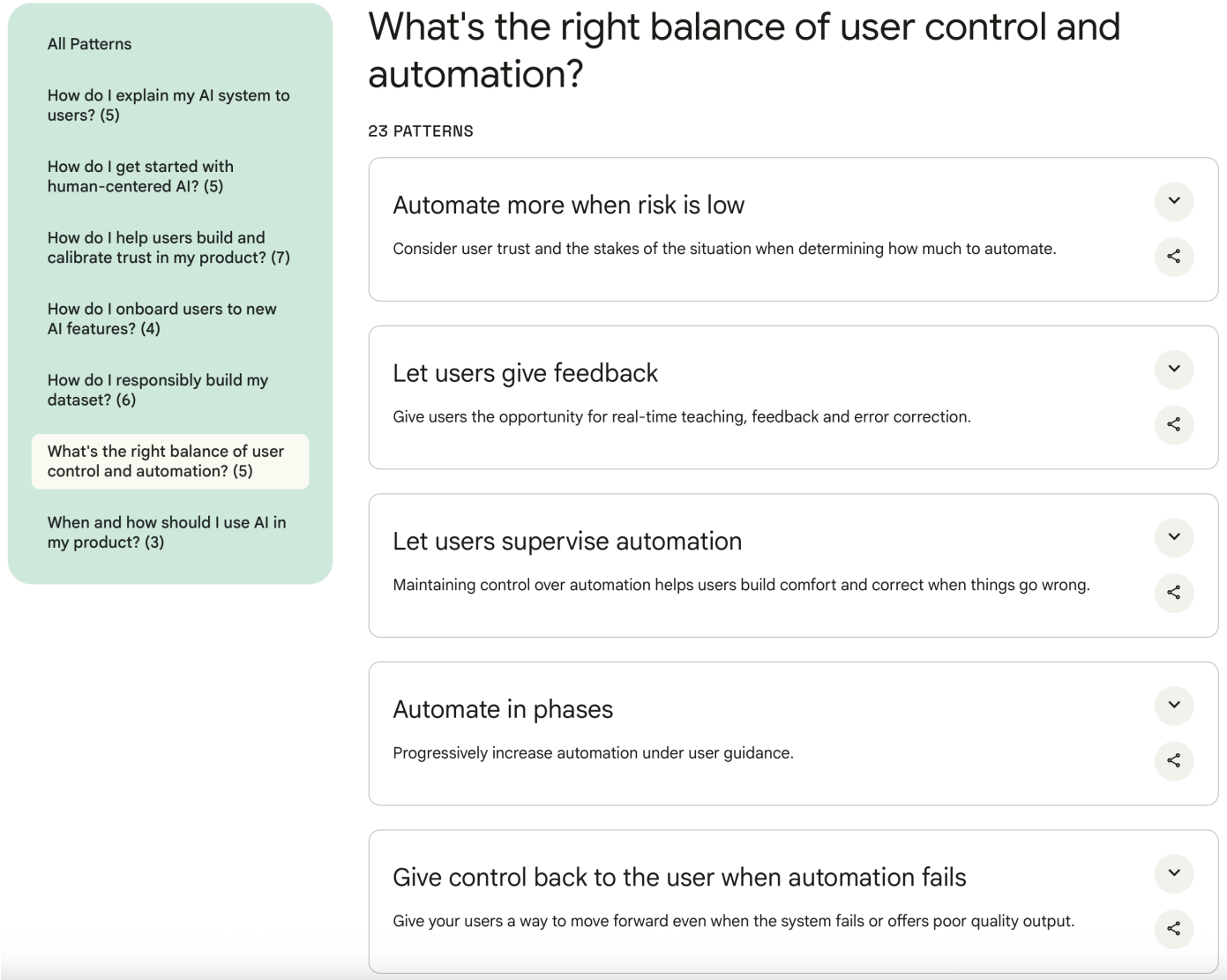
ADAPT WITH FEEDBACK

Adapt AI with user feedback

HELPFUL AI

Create helpful AI that enhances work and play

Pattern


All Patterns
How do I explain my AI system to users? (5)
How do I get started with human-centered AI? (5)
How do I help users build and calibrate trust in my product? (7)
How do I onboard users to new AI features? (4)
How do I responsibly build my dataset? (6)
What's the right balance of user control and automation? (5)
When and how should I use AI in my product? (3)
What's the right balance of user control and automation?
23 PATTERNS
Automate more when risk is low
Consider user trust and the stakes of the situation when determining how much to automate.
Let users give feedback
Give users the opportunity for real-time teaching, feedback and error correction.
Let users supervise automation
Maintaining control over automation helps users build comfort and correct when things go wrong.
Automate in phases
Progressively increase automation under user guidance.
Give control back to the user when automation fails
Give your users a way to move forward even when the system fails or offers poor quality output.

Esempi di pattern

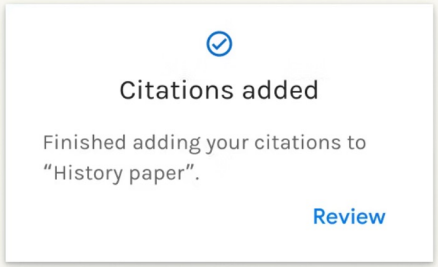
Automate more when risk is low

Consider user trust and the stakes of the situation when determining how much to automate.

When determining how much to automate your product flows, think about the stakes of your product, and the level of comfort that users may have with your type of product. In low risk, well-established products, like content recommendation systems, you might choose to prioritize a more heavily automated product flow where user control is available but optional. However, when onboarding to a new type of product, or in high-stakes situations, errors can be particularly problematic, and can corrode user trust and potentially cause dangerous situations. In such cases, design your system to give users more control over the system.

Aim For

Be more proactive with automation when failure tolerance is higher.



Give control back to the user when automation fails

Give your users a way to move forward even when the system fails or offers poor quality output.

When an AI system fails or gives users a poor prediction, the easiest path forward is often to let the user take over in a non-automated way. Make it as easy and intuitive as possible for users to quickly pick up where the system leaves off. Give them all the information they need to take the reins:

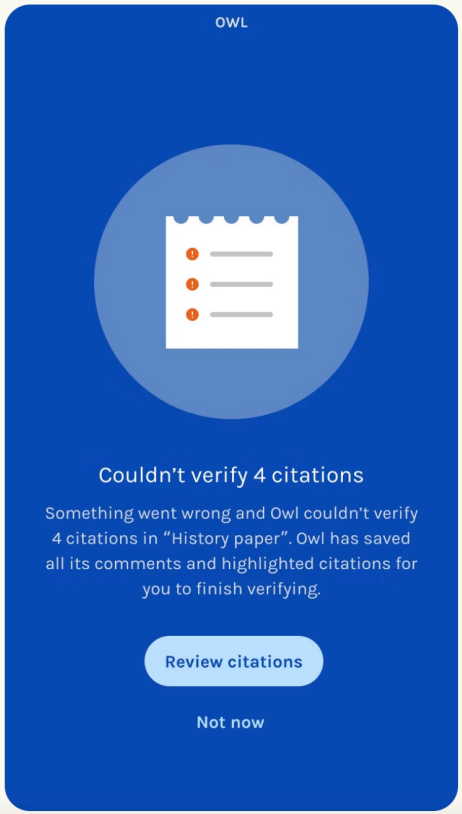
- Awareness of the situation
- What they need to do next
- How to take their next action

In difficult or high-stakes situations, you may even need to redirect users to a human for extra support.

OWL

Aim for

Help users to take over when automation fails.



Riferimenti bibliografici

- Shneiderman, B. Human-Centered AI, Issues in Science and Technology, Winter, 56-61, 2021
- Shneiderman, B. Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy, International Journal of Human–Computer Interaction, 36:6, 495-504, 2020
- Shneiderman, B. Human-Centered AI, Oxford University Press, 2022
- Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., Horvitz, E. Guidelines for Human-AI Interaction. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19). ACM Press, Paper 3, 1–13
- Schmidt, A., Elagroudy, P., Draxler, F., Kreuter, F., Robin Welsch, R. Simulating the Human in HCD with ChatGPT: Redesigning Interaction Design with AI. interactions 31, 1 (January - February 2024), 24–31.
- Meredith Ringel Morris. HCI for AGI. interactions 32, 2 (March - April 2025), 26–32.
- Zijie J. Wang, Chinmay Kulkarni, Lauren Wilcox, Michael Terry, and Michael Madaio. Farsight: Fostering Responsible AI Awareness During AI Application Prototyping. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24). ACM, New York, NY, USA, Article 976, 1–40.
- Borsci, S., Schmettow, M., Malizia, A. et al. A confirmatory factorial analysis of the Chatbot Usability Scale: a multilanguage validation. Pers Ubiquit Comput 27, 317–330 (2023).
- Liao, Q. V., Vorvoreanu, M., Subramonyam, H., Lauren Wilcox, L. UX Matters: The Critical Role of UX in Responsible AI. interactions 31, 4 (July - August 2024), 22–27.