CLAUDIA NICHETTI
DANIEL VENTURA
JÚLIO CARLOS DE SOUZA VAN DER LINDEN
LUCIENE MACHADO
EDITORS

Design, Culture and Innovation

Volume Five

PERSPECTIVES, EXPERIENCES AND CONNECTIONS

UFRGS
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
PGDESIGN UFRGS
LDCI Lab Design Cultura Inovação
marcavisual
editora e projetos culturais

CLAUDIA NICHETTI
DANIEL VENTURA
JÚLIO CARLOS DE SOUZA VAN DER LINDEN
LUCIENE MACHADO
EDITORS

Design, Culture and Innovation

Volume Five

PERSPECTIVES, EXPERIENCES AND CONNECTIONS

UFRGS
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
PGDESIGN UFRGS
LDCI Lab Design Cultura Inovação
marcavisual editora e projetos culturais

D457 Design, culture and innovation: perspectives, experiences and connections – V. V [recurso eletrônico] / organizadores Claudia Nichetti, Daniel Ventura, Júlio Carlos de Souza van der Linden [e] Luciene Machado. – Porto Alegre: Marcavisual, 2026.
109 p. ; digital

ISBN 978-65-6230-017-8

Este livro é uma publicação do Laboratório Design, Cultura e Inovação (LDCI) da Universidade Federal do Rio Grande do Sul (UFRGS).

1. Design. 2. Design - Cultura. 3. Design - Inovação. 4. Design - Educação. 5. Design - Urbanidade. 6. Design – Inteligência artificial. 7. Design - UX. I. Nichetti, Claudia. II. Ventura, Daniel. III. Linden, Júlio Carlos de Souza van der. IV. Machado, Luciene.

CDU 658.512.2

CIP-Brasil. Dados Internacionais de Catalogação na Publicação.
(Jaqueline Trombin – Bibliotecária responsável CRB10/979)

MARCAVISUAL PUBLISHING COMPANY

www.marcavisual.com.br

Editorial Board

Adriane Borda Almeida da Silva
Doctor in Philosophy and Sciences of Education – Zaragoza University/Spain

Airton Cattani
President
Doctor in Informatic on Education – Federal University of Rio Grande do Sul/Brazil

Aline Sanches
Doctor in Philosophie – São Carlos University/Brazil and Paris VII University/France

Celso Carnos Scaletsky (In memoriam)
Doctor in Architectural Sciences – National Polytechnic Institute of Lorraine/France

Denise Barcellos Pinheiro Machado
Doctor in Urbanism – Paris XII University/France

Gustavo Cossio
Doctor in Design – Rio de Janeiro State University/Brazil

Maria de Lourdes Zuquim
Doctor in Architecture and Urbanism – São Paulo University/Brazil



03

Experience design for ai agents: an analysis of human-centered guidelines in llm-owning companies

Fernando Souza Ferreira
Gilberto Consoni
Júlio van der Linden

Abstract

Recent advances in large language models have driven the expansion of artificial intelligence agents, capable of planning and executing actions autonomously to achieve goals, shifting the challenge from technical development to the design of human-centered experiences. In this scenario, the aim of this research was to analyze how LLM-owning companies structure and communicate design principles for AI agents. Through a thematic documentary analysis of public materials from 22 companies, only three organizations (OpenAI, Google, and Microsoft) were identified as providing consolidated guidelines on agent design. The comparative analysis revealed convergences and gaps between these corporate principles and academic frameworks of human-AI interaction. The main results are: (i) the identification of five thematic axes that structure theoretical-practical alignment, AI visibility, contextualized interactions, feedback mechanisms, adaptive evolution, and trust; and (ii) the mapping of significant asymmetries, such as the absence of prototyping guidelines and iterative error treatment in some companies. It is concluded that, although there is progress toward human-centered practices, corporate guidelines remain fragmented and lack integration with consolidated principles from the literature, indicating opportunities to improve the design of agents that are more ethical, understandable, and aligned with human agency.

Keywords: User Experience; AI Agents; Human-AI Interaction; Human-Centered Artificial Intelligence; Large Language Models; Design Principles.

1. Introduction

Recent advances in large language models (LLMs)[1] and in the Transformer architecture[2] have profoundly redefined the field of artificial intelligence (AI), by introducing mechanisms capable of interpreting contextual relationships with speed, precision, and scalability (VASWANI et al., 2023). This technological leap has driven the development of artificial intelligence agents, understood as goal-oriented autonomous systems capable of perceiving and interpreting context, reasoning about alternatives, making decisions, and executing actions adaptively, with minimal human intervention (ACHARYA; KUPPAN & DIVYA, 2025). As a consequence, a transition can be observed in human-computer interactions: from predominantly static interfaces to collaborative, adaptive experiences mediated by agents that continuously learn from users (CAPOSINO, 2025).

In this new paradigm, user experience (UX) becomes even more critical. UX encompasses pragmatic and hedonic dimensions (HASSENZAHL, 2018), varying across individuals, contexts, and over time. Interactions with AI agents, therefore, do not depend solely on the technical quality of the models, but on how these systems are designed to support human needs, communicate intentions, reduce uncertainties, and establish trust (STEVENS, 2025). The emergence of multimodal agents, hyper-personalization, and generative interfaces reinforces that digital experiences are now living products, which evolve after launch and require continuous cycles of design, learning, and adaptation (CAPOSINO, 2025).

At the same time, the expansion of autonomous systems brings ethical, social, and practical challenges, ranging from amplified inequalities to security risks and loss of privacy, which require adequate governance and design approaches aligned with human-centered artificial intelligence (HCAI) (GUENAT et al., 2022). In this context, the field of human-AI interaction (HAII) seeks to integrate human-computer interaction (HCI), human factors, ethics, and computer science to ensure that intelligent systems remain understandable, transparent, and controllable, preserving human agency (XU, 2025).

---

[1] Large Language Models (LLMs) are artificial intelligence systems trained on enormous volumes of text to understand and generate natural language. LLMs can adapt to different uses, such as answering questions, summarizing texts, translating content, maintaining conversations, and supporting reasoning processes. This capability occurs because these models learn language patterns and use this learning to produce coherent responses (Naveed et al., 2025).

[2] The Transformer architecture is a structure capable of observing several parts of a sentence at the same time and identifying which words are most important for understanding the meaning of the text. This mechanism is called attention. Because it is efficient and capable of capturing relationships between distant words, the Transformer has become the basis of many current artificial intelligence models, including large language models (Vaswani et al., 2023).

At the same time, the expansion of autonomous systems brings ethical, social, and practical challenges, ranging from amplified inequalities to security risks and loss of privacy, which require adequate governance and design approaches aligned with human-centered artificial intelligence (HCAI) (GUENAT et al., 2022). In this context, the field of human-AI interaction (HAII) seeks to integrate human-computer interaction (HCI), human factors, ethics, and computer science to ensure that intelligent systems remain understandable, transparent, and controllable, preserving human agency (XU, 2025).

Despite the growing relevance of design for AI agents, important gaps persist in both literature and practice. Studies show that UX frameworks for AI rarely address continuous feedback integration, improved personalization capabilities, or iterative cycles (SCHRÖDER & FISCHER, 2025). Other authors highlight emerging challenges in human-AI interaction, such as generative variability, unjustified dependence, co-creation, and the inherent uncertainty of models, which require new design principles (WEISZ et al., 2024). Research on prototyping also reveals practical difficulties in alternating between technical and experiential levels, visualizing internal states, and incorporating feedback during execution (LIANG et al., 2025). In addition, studies on human-centered design applied to AI emphasize that the construction of effective agents requires empathy, hypothesis formulation, and a deep understanding of the real context of use, steps that are still poorly consolidated in technical teams, but are fundamental to avoiding solutions misaligned with users' expectations, limitations, and values (CEREJO, 2023). This perspective reinforces that designing AI agents is not merely a matter of engineering, but a process of design, critical evaluation, and alignment between the capabilities of algorithms and experiential value.

Despite this scenario, it remains unknown to what extent the major companies developing LLMs, which define global standards for the ecosystem, communicate design principles for human-centered agents, or how such principles relate to accumulated scientific knowledge. Considering that companies that commercialize LLMs tend to provide formal usage guidelines, it is expected that they also make available structured principles on how to design good AI agents.

In view of this, this article investigates what LLM-owning companies publicly communicate about how to design user experiences for AI agents. The study is based on a thematic documentary analysis (MORGAN, 2022) of official materials from companies developing language models in order to answer three questions: (RQ1) how these companies communicate human-centered principles in public materials; (RQ2) which design principles are recommended for building agents; and (RQ3) how such corporate principles align with or diverge from the literature on human-centered artificial intelligence.

Thus, the contribution of this article is twofold. From a scientific standpoint, it offers the first comparison between corporate guidelines and the main theoretical frameworks of UX for AI, revealing convergences and gaps. From an applied standpoint, it contributes to enabling designers, engineers, and organizations to adopt design practices more aligned with human-centered AI, for the creation of agents that are safer, more ethical, more transparent, and more coherent with the needs of human beings.

2. Methodological Procedures

This research aimed to analyze design principles from LLM-owning companies for the user experience of artificial intelligence agents. It is an inductive and applied study, as it starts from the empirical analysis of corporate documents to build interpretations and propose practical improvements in design principles for AI, whose complexity imposes new challenges on designers, arising from uncertainty about the capabilities of models and the breadth and unpredictability of their results (LIANG et al., 2025). In addition, it has a descriptive-exploratory nature, as it seeks to identify and interpret patterns in the design principles present in corporate AI documentation.

For the systematic analysis of public materials, a thematic documentary procedure was adopted (MORGAN, 2022) with the aim of identifying user-centered corporate recommendations and their design relevance for agents. The empirical scope comprises:

1. Definition of research questions;
2. Selection of LLM-owning companies;
3. Selection of documents that describe human-centered design principles;
4. Analysis of convergences and divergences between corporate approaches and the scientific literature on UX for AI.

2.1 Research Questions

Table 1 - Research Questions

ID	Research Question	Focus
RQ1	How do companies communicate human-centered principles for agents in public materials?	Communication Focus
RQ2	What are the design principles communicated by companies for building human-centered agents?	Content of the principles
RQ3	How do companies' design principles align with the principles present in the literature on human-centered artificial intelli	Theorical-pratical comparison

Source: Developed by the authors (2026)

2.2 Company Selection Criteria

The selection criteria included companies:

- CSE1 - that own LLMs;
- CSE2 - that provide LLMs for text outputs rather than video, image, or audio.

Companies that did not meet both criteria were disregarded.

2.3 Document Selection Criteria

The selection criteria included documents:

- CSD1 - that constitute consolidated, public, and accessible corporate knowledge;
- CSD2 - that are authored by the selected companies;
- CSD3 - that describe design principles for agents.

Documents that did not meet all three criteria were excluded. These criteria were adopted to ensure authentic, credible, representative, and meaningful documents for answering the research questions (MORGAN, 2022).

2.4 Content Analysis

The documentary analysis (MORGAN, 2022) was conducted through an interpretive reading to identify how companies communicate AI design principles in their public materials. This interpretive process followed the principles of thematic documentary analysis, allowing the design principles made explicit by each company to be categorized.

For the analysis, summative content analysis was adopted (HSIEH & SHANNON, 2005), combining (i) the identification and comparison of recurring keywords and expressions related to design theories and (ii) the interpretation of the content based on the semantic patterns found. For each emerging category, verbatim excerpts and analytical paraphrases were extracted, ensuring the adopted criterion between the interpretations and the textual evidence present in the corporate documents.

Finally, a synthesis and comparison stage was carried out, relating the content of the corporate principles to the companies' design intentions for building human-centered agents. This stage made it possible to identify convergences and divergences between the corporate principles and the theoretical background of this study, responding in an integrated manner to the three research questions (RQ1, RQ2, and RQ3).

3. Human-AI Interaction and Artificial Intlligence Agents

This section is dedicated to presenting an overview of human-AI interaction and artificial intelligence agents, as well as presenting related works that have addressed design recommendations for intelligent systems.

3.1 Human-Centered Human-Artificial Intelligence Interaction

Human-AI interaction (HAII) is the field that: (i) investigates how to design, evaluate, and improve interactions between humans and intelligent systems and (ii) operationalizes the principles of human-centered artificial intelligence (HCAI), which emerged from the foundations of human-computer interaction (HCI) (XU et al., 2021). While HCI deals with making computational systems usable, efficient, and accessible (GRUDIN, 2018), HCAI expands this scope so that autonomous, adaptive, and unpredictable systems are reliable, transparent, ethical, and aimed at enhancing, rather than replacing, human performance (SHNEIDERMAN, 2020, 2022; XU et al., 2023). Despite advances in HCI and HCAI, Xu (2025) highlights that HAII still does not fully integrate practice and theory, which requires greater articulation among computer science, design, psychology, and ethics in order to develop truly human-centered solutions.

For such articulation, Xu (2025) proposes human-centered human-AI interaction (HC-HAII), which is structured into four main components: objectives, scope and research questions, methodology, and design paradigms. Objectives are guided by the HCAI perspective in order to place the human at the center of research and in final control of AI systems. The scope and research questions are guided by the HCAI perspective in order to address gaps in the design, development, implementation, and use of AI systems. The HC-HAII methodology promotes interdisciplinary collaboration among human factors, HCI, and user experience (UX), as well as integrated innovation. Finally, design paradigms, from the HC-HAII perspective, promote multilevel design, which ranges from simple human-AI interactions to ecosystems (e.g., multi-agent systems, human-AI teams, AI-AI coordination) and sociotechnical dimensions (e.g., organizational, cultural, and ethical systems). In summary, HC-HAII proposes establishing a bridge between technology and design with a focus on ensuring that the advancement of automation remains at the service of human agency.

In the context of the agency of interactions with intelligent systems, Jiang et al. (2024) highlight four central axes: collaboration, competition, conflict, and symbiosis. Human-AI collaboration can manifest at different levels of intelligence - assisted, augmented, cooperative, and autonomous - combining human capacities for ethics, creativity, and empathy with AI competencies in analysis, speed, and prediction. Human-AI competition emerges both in digital games and in work dynamics and presents challenges related to resource equity, explainability of decisions, and preservation of the human role. Human-AI conflict refers to divergences in tasks or relationships, which can be mitigated through visible, understandable, and controllable decision-making processes. Finally, human-AI symbiosis is supported by four pillars: human-in-the-loop (feedback in the AI

learning cycle), explainable AI (XAI), user-friendly AI, and responsible AI. This perspective highlights the importance of AI literacy so that users can understand the capabilities and limitations of the technology, interact effectively, assess its reliability and credibility, use it responsibly and ethically, and adapt to its constant evolution (SÜßE; KOBERT & KRIES, 2021).

3.2 Artificial Intelligence Agents

Artificial intelligence agents are autonomous systems capable of perceiving their environment, reasoning, and acting independently to achieve specific goals with minimal human interference (ACHARYA; KUPPAN & DIVYA, 2025; NISA et al., 2025; SAPKOTA & ROUMELIOTIS; KARKEE, 2026). These systems are also capable of explaining their decisions, describing the alternatives considered, actions executed, adaptations made, and the reasons for each choice (SADO et al., 2023). By connecting reasoning to the autonomous execution of actions, agents become able to use tools, navigate the internet, interact with other applications, and respond dynamically to changes in the environment through reactive intelligence and adaptability (CHACON-CHAMORRO et al., 2025).

The adaptability of AI agents is enabled by pre-trained large language models (LLMs), which perform multiple interactions and interpret instructions in natural language, known as prompts, in order to guide their responses (ZAMFIRESCU-PEREIRA et al., 2023). Among the main approaches to prompt design, the following stand out: (i) language dynamics, which involves communicating intentions and exploring contexts and meanings; (ii) direct task specification, based on clear instructions (zero-shot); (iii) specification by demonstration, which uses explicit examples (few-shot); (iv) specification by memetic proxies, which employs cultural references or characters as semantic shortcuts; (v) behavior restriction, which seeks to reduce ambiguities; (vi) serialization of reasoning, which divides the process into steps to increase precision; and (vii) prompt metaprogramming, which creates prompts capable of generating other prompts, expanding the adaptability and autonomy of the model. The purpose of these approaches goes beyond obtaining responses: it is about shaping the linguistic and cognitive behavior of models, transforming prompt design into an interpretive, design-oriented, and strategic practice aligned with human intentions, contexts, and values (REYNOLDS & MCDONELL, 2021).

In addition, agents can be classified into seven types: reactive, proactive, limited-memory, model-based, goal-oriented, Theory of Mind (ToM), and self-aware—the latter two being capable of inferring human emotions, beliefs, and intentions, and of reasoning about their internal states (NISA et al., 2025). Despite their advances, these systems present two central limitations: hallucinations, plausible but incorrect responses, and high sensitivity to prompts, in which small variations in formulation produce divergent results. Both scenarios require manually designing prompts to ensure the consistency and reliability of interactions (SAPKOTA; ROUMELIOTIS & KARKEE, 2026). Another critical aspect in prompt design is the prevention of prompt leaking, which consists of attacks that seek to extract the system's internal instructions, compromising its security, privacy, and intellectual property (HUI et al., 2024).

3.3 Related Works

The nature of artificial intelligence agents imposes new challenges on designers, which requires the investigation of specific human-AI interaction practices. Principles and prototyping cycles become essential to improving the usability and effectiveness of the human-centered design process in generative systems. In this context, Amershi et al. (2019) synthesized, from guidelines proposed over 20 years of research and professional practice, a set of 18 principles for human-AI interaction design. The principles were refined in four phases by a team of 11 researchers and reviewed by more than 60 designers and usability specialists. The consolidated set was categorized into four moments: initially, during interaction, in case of error, and over time.

- Initially: make clear what the system can do; make clear how well the system can perform.
- During interaction: time services based on context; show contextually relevant information; match relevant social norms; mitigate social biases.
- In case of error: support efficient invocation; support efficient dismissal; support efficient correction; scope services when in doubt; make clear why the system did what it did.
- Over time: remember recent interactions; learn from user behavior; update and adapt cautiously; encourage granular feedback; convey the consequences of user actions; provide global controls; notify users about changes.

Weisz et al. (2024), in turn, developed six design principles for human-generative AI interaction through an iterative process inspired by the principles of Amershi et al. (2019). The initial formulation was based on a literature review on generative AI applications, resulting in seven preliminary principles (WEISZ et al., 2023). The authors then conducted four refinement iterations, which included workshops with designers, a modified heuristic evaluation to test clarity, relevance, and coverage, and practical applications of the principles in two product design projects.

- Responsible design: use a human-centered approach; identify and resolve value tensions; expose or limit emergent behaviors; test and monitor potential harm to the user.
- Design for generative variability: leverage multiple outputs; visualize the user journey; enable curation and annotation; highlight differences among outputs.
- Design for mental models: guide the user regarding generative variability; teach effective use through examples; understand the user's mental model; teach the system about the user.

- Design for co-creation: help the user formulate effective prompts; provide generic input parameters; offer controls relevant to the use case; support the co-editing of outputs.
- Design for appropriate trust and reliance: calibrate trust with explanations; provide justifications for outputs; use friction to prevent over-reliance; indicate the role of AI.
- Design for imperfection: make uncertainty visible; evaluate outputs with domain-specific metrics; offer ways to improve outputs; provide feedback mechanisms.

Liang et al. (2025) sought to understand the key activities performed by engineers during the prototyping of agent user experiences, through a study conducted with 12 participants. The study identified an iterative cycle for agent prototyping: design, execution, debugging, and refinement. Based on the iterative cycle of agent prototyping, the study identified three main design implications:

- Offer understandable and configurable interaction components: interface actions should have clear nomenclature, preview features to facilitate the discovery process and the suggestion of effective prompts, as well as resources that reveal prompts for quick editing.
- Allow alternation between levels of technical information and user experience: the prototype should allow the presentation of different amounts of information during agent execution to be alternated, in order to avoid conflicts between the needs of engineers and users. Engineering professionals require details to understand the agent's behavior, while users may become overwhelmed by excessive information.
- Favor agile cycles of design and testing: it is recommended to offer review resources, such as indicators of the agent's current status, alerts and update messages, and personalization features, such as variation in execution speed and in the size of the agent's response.

Finally, Schröder & Fischer (2025) presented a systematic literature review on user experience frameworks applied to generative AI systems, analyzing 16 studies in terms of their contributions and gaps in relation to consolidated theoretical and practical principles. The results indicate that only three frameworks explicitly address prototyping methods in the design of generative AI systems. In this respect, the analysis reveals five main opportunities for improving user experience frameworks for AI:

- Iterative design: strengthen feedback cycles and incorporate usability testing in early stages, for the early identification of problems.
- User requirements: adopt structured methods for elicitation and continuous feedback integration, in order to align expectations and design goals.
- Holistic experience: consider emotional dimensions of satisfaction and long-term user engagement, ensuring experiences that go beyond functional utility.
- Trust and control: improve transparency for users and increase control over AI decisions, in order to promote ethical collaboration.
- Adaptability and context: apply responsive and context-sensitive design techniques and promote improved personalization capabilities, adjusting interactions to individual preferences and situational conditions.

4. Results and Discussion

The presentation of the results is structured around the three research questions, which address the communication focus, the content of the principles, and the theoretical-practical comparison.

4.1 RQ1 - How do companies communicate human-centered principles for agents in public materials?

Twenty-two LLM-owning companies that enable the construction of AI agents based on textual processing were selected, originating from seven countries. The selection was conducted between November 8 and November 15, 2025, with an exploratory character and without the intention of exhausting or systematically mapping the entire global ecosystem of companies developing language models. The survey was carried out through internet searches using the term "LLM companies" and a consultation with ChatGPT using the prompt: "List LLM-owning companies for text processing. Consider: (i) private or public companies, (ii) companies that are the main developer, (iii) include global companies, and (iv) exclude companies whose primary focus is only image, audio, or video generation." Companies that did not meet the criteria were not selected: regarding CSE1, model-hosting repositories, for example, HuggingFace, or companies that use third-party models, for example, Perplexity; and regarding CSE2, companies whose primary focus is not a textual solution, such as Stability AI, Runway, Midjourney, ElevenLabs, Pika Labs, and Synthesia.

Next, for each selected company, the addresses of the models' technical documentation were identified, to which two search methods were applied: (i) an exploratory verification to locate explicitly presented design principles for AI and (ii) the use of the keywords "design," "guideline," and "experience" in the internal search mechanisms of these documentations. If no guideline was identified in the technical documentation, a complementary internet search was carried out using the expression "{Company Name} design guidelines," with the objective of locating specific portals related to the company's design operations and, within them, identifying possible design principles for AI. This strategy aimed to cover, in a broad and structured manner, the main public institutional sources relevant to the analysis.

As a result of this process, three documents were identified that present design principles directly applicable to agent prototyping, essential elements for guiding human-AI interaction practices, which will be analyzed in light of the principles of human-centered artificial intelligence. For the remaining companies, no proprietary materials were found that (CSD2) were authored by the companies themselves and (CSD3) presented design principles for agents. The following table summarizes these findings.

Table 2 - Technical Documentation and Design Principles of the Models

Company (Country)	Reference Model	Documentation	Design Principles
OpenAI (USA)	GPT	https://developers.openai.com/	https://developers.openai.com/apps-sdk/concepts/design-guidelines
Microsoft (USA)	Copilot	https://learn.microsoft.com/en-us/copilot/microsoft-copilot	https://microsoft.design/articles/ux-design-for-agents/
Google DeepMind (USA/UK)	Gemini	https://ai.google.dev/gemini-api/docs	https://pair.withgoogle.com/guidebook/
Cohere (Canadá)	Command	https://cohere.com/developers	Not found
Meta AI (USA)	Llama	https://llama.meta.com/docs	Not found
Anthropic (USA)	Claude	https://docs.claude.com/en/home	Not found
Salesforce (USA)	XGen	https://www.salesforce.com/agentforce/	Not found
IBM (USA)	Granite	https://www.ibm.com/granite/docs	Not found
Mistral AI (França)	Mistral	https://docs.mistral.ai	Not found
Amazon AWS (USA)	Titan	https://builder.aws.com/learn/topics/agentic-ai	Not found
Aleph Alpha (Alemanha)	Luminous	https://docs.aleph-alpha.com	Not found
Tencent (China)	Hunyuan	https://cloud.tencent.com/document/product/1729	Not found
Alibaba (China)	Qwen	https://qwen.readthedocs.io	Not found
xAI (USA)	Grok	https://docs.x.ai/docs/overview	Not found
Baidu (China)	ERNIE	https://ernie.baidu.com/	Not found
Reka AI (Singapura)	Reka	https://docs.reka.ai/overview	Not found
Snowflake (USA)	Artic	https://docs.snowflake.com/en/arctic	Not found
Databricks (USA)	DBRX	https://docs.databricks.com/aws/en/developers/	Not found
Writer (USA)	Palmyra	https://dev.writer.com/home/introduction	Not found
NVIDIA (USA)	Nemotron	https://docs.nvidia.com/nemo/agent-toolkit/latest/index.html	Not found
DeepSeek (China)	DeepSeek	https://www.deepseek.com/	Not found

Source: Developed by the authors (2026)

As a response to RQ1, it was observed that only 3 of the 22 companies analyzed disclosed design principles aimed at the prototyping of artificial intelligence agents. Among them, OpenAI (2025) concentrates this content directly on pages of the technical documentation of artificial intelligence models, while Google (2025) and Microsoft (2025) organize it in specific addresses related to the companies' design operations. The analysis of this content will be deepened in the responses to RQ2 and RQ3.

4.2 RQ2 - What are the design principles communicated by companies for building human-centered agents?

OpenAI (2025) describes that its design principles for agents should be conversational, intelligent, simple, responsive, and accessible. The guidelines also delimit good and bad uses: agents are appropriate when they support tasks with clear beginnings and endings, generate immediate value, and can be presented in a visually simple way; however, complex flows, long or static content, duplication of system functionalities, exposure of sensitive information, and promotional uses should be avoided. In addition, OpenAI defines rules for tone and proactivity: content should be concise, contextual, and non-promotional, while proactive interventions can occur only when strictly relevant to the user's intent. Unsolicited re-engagements[3] are prohibited, and any intervention must make explicit its reason and timing. The guiding design principles are:

- Conversational: provide an experience that is integrated and natural to the flow of dialogue.
- Intelligent: use context to anticipate intentions and generate relevant responses.
- Simple: concentrate each interaction on a single action or clear outcome.
- Responsive: ensure lightness and speed in execution, without overloading.
- Accessible: provide inclusive support for a wide variety of users.

Google's (2025) design principles, in turn, emphasize that experiences with AI should prioritize users' autonomy and different levels of literacy in the design of AI-supported flows, in order to increase their efficiency without compromising their agency. They should also align data and models with real contexts of use, recognizing that natural interactions with AI vary according to the situation and that systems may be used in unexpected tasks, beyond the intended scope or training data. In addition, safety should be treated as an evolving process, since new risks may emerge with the use of AI in complex tasks and varied contexts, which requires proactive safeguards, continuous safety adaptation, and adjustments to

[3] Re-engagements are attempts by the AI agent to reactivate interaction with the user without the user having requested it. In other words, they are messages, reminders, suggestions, or calls to action that appear to bring the user back to the conversation, resume a task, or promote some use, even when there is no clearly expressed intent from the user.

the models. The company reinforces the need to incorporate feedback mechanisms capable of interpreting results, monitoring changes, and guiding improvements, since human-AI interaction operates as a bidirectional learning cycle in which AI personalizes its responses while users adjust their behaviors. Finally, experiences with AI should be useful and inspire creativity, in order to enhance what people value in work and leisure. In summary, these principles are organized into five central axes:

- Design for the appropriate level of user autonomy
- Align AI with real-world behaviors
- Treat safety as an ever-evolving effort
- Adapt AI based on user feedback
- Create useful AI that enhances work and leisure

Finally, Microsoft (2025) highlights agents as a new chapter in human-computer interaction. The company organizes its design principles into three complementary dimensions: Agent (Space), Agent (Time), and Agent (Core). Agent (Space) is the physical and digital space in which agents operate, connecting people, events, and knowledge without replacing or reducing the human role, in an accessible, multimodal, and unobtrusive way, with fluid transitions between proactive and reactive modes. Agent (Time) is described as designing agents that interact across past, present, and future with other agents or people, using contextual history to inform decisions, adopting a posture of encouraging rather than notifying, and continuously evolving according to users' preferences, behaviors, and needs. Agent (Core), in turn, refers to the central elements of the design of an individual agent, highlighting uncertainty as an inherent characteristic of agents and transparency, control, and consistency as fundamental elements. In summary, Microsoft's principles are organized as follows:

- Agent (Space)
  - Connect, do not collapse: agents should bring people and knowledge closer together, without replacing humans.
  - Easy to access, yet practically invisible: agents should offer voice, text, sound, and image interaction and should perform "invisible" processing for fluid transitions between foreground and background.
- Agent (Time)
  - Reflecting on history beyond states (from past to present): use past memory to contextualize present interactions, connecting events and behaviors to generate relevant responses.
  - Encouraging more than notifying (around "now"): it should proactively engage in a contextual way that is sensitive to the environment and to the user's preferences, gradually empowering users.
  - Adaptation and evolution (from present to future): continuously adjust to different devices, modes, and preferences, as well as to users' behavior, feedback, and accessibility.
- Agent (Core)
  - Embrace uncertainty, but establish an appropriate relationship of trust: the level of uncertainty should always be visible and accessible to the user in order to avoid excessive trust.
  - Transparency, control, and consistency are fundamental elements of all agents: agents should be transparent and controllable so that users can visualize and personalize their actions, preferences, and connections.

As a response to RQ2, the principles communicated by OpenAI, Google, and Microsoft converge toward a central ethical principle: agents should expand human capabilities without reducing their autonomy, with experiences that are transparent, safe, contextualized, and genuinely useful to the user. Although they adopt different narratives - OpenAI emphasizing the conversational experience, Google prioritizing sociotechnical alignment, and Microsoft positioning agents that operate in space, time, and the functional core - all converge in defense of a design that balances utility, simplicity, control, and trust. Next (RQ3), the extent to which these corporate principles align with or diverge from the literature on human-centered artificial intelligence is analyzed.

4.3 RQ3 - How do companies' design principles align with the principles present in the literature on human-centered artificial intelligence?

To analyze the extent to which the design principles for AI communicated by companies align with or diverge from the literature on human-centered artificial intelligence, a synthesis composed of five thematic axes was developed. These are: (i) AI visibility (technical clarity and human values); (ii) contextualized interactions (user control and understanding of multiple outputs); (iii) feedback mechanisms (repair and refinement); (iv) adaptive evolution (human-AI co-creation, continuous personalization); and (v) trust (transparency, explanation of decisions and executions). These themes were derived from the four related works in this study and structure a broad analytical space to guide the interpretation and classification of the verbatim excerpts from the corporate principles. The following table presents the relationship between each theme and the respective reference works.

Table 3 - Thematic Axis of the Related Works

Thematic Axis	Human-AI Interaction Design Guidelines (Amershi et al., 2019)	Generative Design Principles (Weisz et al., 2024)	Practical Prototyping Implications (Liang et al., 2025)	Framework Improvement Opportunities (Schröder and Fischer, 2025)
AI visibility (technical clarity and human values).	Make clear what the system can do and how well it can perform.	Responsible design: align values and mitigate risks.	Provide understandable and configurable components.	Structure user requirements and expectations.
Contextualized interactions (user control and understanding of multiple outputs).	Adjust the timing and context of responses, display relevant information, conform to norms, and mitigate social biases.	Design for mental models and variability: guide the user regarding multiple outputs and teach effective use.	Allow alternation between technical and experiential levels.	Promote control and contextual adaptability.
Feedback mechanisms (repair and refinement).	Facilitate activation, rejection, and correction.	Design for imperfection: make uncertainty visible, offer means for improvement, and provide feedback channels.	Favor agile testing and review cycles.	Strengthen iterative design and continuous feedback.
Adaptive evolution (human-AI co-creation, continuous personalization).	Remember interactions, learn from the user, update cautiously, and encourage feedback.	Design for co-creation: support prompt formulation and the co-editing of outputs.	----	Incorporate a holistic and context-responsive experience.
Trust (transparency, explanation of decisions and executions).	----	Design for appropriate trust and reliance: calibrate trust and justify outputs.	----	Trust and control through transparency and ethics.

Source: Developed by the authors (2026)

4.3.1 AI Visibility (technical clarity and human values)

Only OpenAI and Microsoft present principles that converge with the need to make explicit the purpose, scope, and behavior of AI agents from the beginning of the interaction. OpenAI emphasizes a clear and reliable experience, with transparency regarding the agent's status and the context needed for users to understand the results. Microsoft reinforces this orientation by highlighting the need to make visible, in real time, the agent's status, skills, and connections, even when it operates in the background. These corporate principles correspond to the recommendations to make clear what the system can do and make clear how well the system can perform (AMERSHI et al., 2019), to the commitment to responsibility toward human values (WEISZ et al., 2024), and to the importance of clear nomenclatures for system actions (LIANG et al., 2025). In summary, OpenAI's and Microsoft's principles follow the theory of human-centered AI by treating visibility, technical clarity, and human values as requirements for autonomy and human control, while OpenAI presents a gap in this respect.

4.3.2 Contextualized Interactions (user control and understanding of multiple outputs)

OpenAI, Microsoft, and Google converge with the thematic axis of contextualized human-AI interactions and controls adjusted to users. OpenAI emphasizes that experiences should feel like a natural extension of the system's conversational flow, guiding each interaction to focus on a single clear action and tools to anticipate the user's intent based on context.

Microsoft reinforces this perspective by highlighting that agents should be intuitive, accessible, and multimodal, in order to provide information based on the user's social and cultural environment and to simplify flows and reduce cognitive load. Google, in turn, highlights the need to design for different levels of autonomy, considering tasks, expertise, and the effort required to use the system, recognizing that people engage with AI in different ways depending on the context.

These corporate orientations converge with recommendations from the literature, such as adjusting the timing and presentation of information according to context (AMERSHI et al., 2019), designing interactions that respect users' mental models (WEISZ et al., 2024), adapting interactions to individual preferences (SCHRÖDER & FISCHER, 2025), and alternating between degrees of technical detail and simplicity (LIANG et al., 2025). Therefore, the three companies incorporate central principles of human-centered AI by promoting situated, context-sensitive interactions adjusted to the level of control desired by the user.

4.3.3 Feedback Mechanisms (repair and refinement)

In the thematic axis of feedback mechanisms (repair and refinement), it is observed that only Microsoft and Google present explicit principles, while OpenAI does not detail guidelines on iterative correction or failure treatment. Microsoft emphasizes that it is necessary to signal the likelihood of uncertainty in order to avoid excessive trust by users in the models and to incorporate feedback collection mechanisms that allow the agent to refine its performance based on users' behavior and responses.

Google, in turn, characterizes mechanisms for analyzing interaction so that AI learns from users' behaviors and feedback, which support the interpretation and adjustment of results. Such principles are aligned with theoretical recommendations to facilitate efficient correction or rejection of results (AMERSHI et al., 2019; SCHRÖDER & FISCHER, 2025; WEISZ et al., 2024) and to design for imperfection, providing feedback mechanisms and ways to improve system outputs (WEISZ et al., 2024). They also dialogue with the need for clear infrastructure for editing and iteratively reviewing the agent's actions (LIANG et al., 2025), and with the importance of structured feedback cycles to strengthen the iterative design of generative systems (SCHRÖDER & FISCHER, 2025). In summary, Microsoft and Google demonstrate consistent alignment with the literature on human-centered AI, while OpenAI presents a gap in this respect.

4.3.4 Adaptive Evolution (human-AI co-creation, continuous personalization)

In the theme of adaptive evolution between humans and AI, the three companies converge in recognizing that agents should learn from the user over time, but they do so with different intensities and focuses. OpenAI only highlights that agents should use conversational context to anticipate intentions and support the user with progressively more relevant interactions, without describing formal memory mechanisms.

Microsoft, the only company that formulates a structured temporal model—past, present, and future—recommends considering historical memory, connected databases, and mechanisms of continuous evolution in agent design, in order to inform and interact with current events and tasks. Thus, the agent progressively expands its complexity to empower users in the long term. Google emphasizes real-world-centered evolution, in experiences that enhance people's work and leisure through AI. Guided by continuous feedback, the evolution of human-AI interaction follows a bidirectional cycle in which users adjust their behaviors and AI personalizes responses based on this learning over time.

Taken together, these corporate principles align with the theoretical principles of remembering interactions, learning from user behavior, updating and adapting cautiously (AMERSHI et al., 2019), helping the user formulate effective prompts (WEISZ et al., 2024), and considering emotional dimensions of satisfaction and long-term user engagement (SCHRÖDER & FISCHER, 2025). However, it is noted that only Microsoft presents an explicit vision of co-creation and systemic evolution, while OpenAI and Google focus on incremental adaptation.

4.3.5 Trust (transparency, explanation of decisions and executions)

In the thematic axis of trust, the three companies analyzed converge in stating that human-centered agents should be transparent, understandable, and controllable.

OpenAI presents an operational approach centered on safe use, prohibiting practices that may reduce trust, such as exposure of sensitive data, promotional messages, or unsolicited re-engagements, and reinforcing that proactivity should occur only when contextual and non-intrusive, in order to preserve user agency.

Microsoft goes further by structuring trust as a central principle of agent design, stating that these systems must earn human trust. The company introduces the idea of transparency, in which users can visualize the agent's actions, automations, settings, and states, as well as personalize its functioning. It also highlights the importance of recognizing and communicating agents' natural uncertainties, aligning with the principles of explaining why the system did what it did (AMERSHI et al., 2019) and designing for trust, with explanations and justifications for outputs (WEISZ et al., 2024).

Google reinforces this set by positioning safety as a continuous effort, asking designers to implement safety measures for different levels of risk and to review strategies as the user's context evolves. This view is coherent with the principles of responsibility, ethical governance, and trust calibration (AMERSHI et al., 2019; SCHRÖDER & FISCHER, 2025; WEISZ et al., 2024). Taken together, the three companies emphasize that trust is not only a technical attribute, but a sociotechnical process sustained by explanations, mechanisms of control, and responsibility, which are foundations of human-centered human-AI interaction.

In response to RQ3, the comparative analysis shows that OpenAI, Google, and Microsoft incorporate the foundations of human-centered artificial intelligence in distinct yet convergent ways. While OpenAI prioritizes simple and safe conversational experiences, Google emphasizes sociotechnical alignment, graduated autonomy, and feedback-based evolution, and Microsoft presents the most comprehensive approach, structuring human-AI interaction in the space, time, and functional core of agents. The following table presents the verbatim excerpts classified by thematic axis.

Table 3 - Thematic Axis of the Related Works

Thematic Axis	OpenAI	Microsoft	Google
AI visibility (technical clarity and human values).	"Always show why and when your tool is resurfacing." "Provide enough context so users understand the purpose of the nudge." "The overall experience must always feel like ChatGPT: clear, helpful and trustworthy."	"Agent status, or what the agent is doing, is clearly visible at all times." "Agents' associated knowledge, tools/ skills, and connections with people and other agents are transparent and customizable." "Agents that operate as background processes have a user-facing mechanism to view and control actions and automations."	----
Contextualized interactions (user control and understanding of multiple outputs).	"Experiences should feel like a natural extension of ChatGPT, fitting seamlessly into the conversational flow." "Tools should be aware of conversation context, supporting and anticipating user intent." "Each interaction should focus on a single clear action or outcome." "Tools should feel fast and lightweight, enhancing conversation rather than overwhelming it."	"Transparency, control, and consistency are foundational elements of all agents." "Individual agents and multi-agent systems are easily discoverable, intuitive, and accessible." "Any human-facing agent supports multimodal inputs and outputs." "Agents deliver information based on contextual environment, social and cultural information… and user preferences." "Agents can simplify flows or dynamically generate cues to direct the user's attention at the right moment."	"Design for the appropriate level of user autonomy… consider different user tasks, expertise, and the effort required to steer the AI system." "People engage differently with AI systems based on their context." "People also prefer to remain in control of which tasks tools should execute, how, and to what end."
Feedback mechanisms (repair and refinement).	----	"The level of certainty and reasoning behind a recommendation are visible… to avoid overreliance." "An agent adapts to user behavior and feedback" "An agent uses data in memory to inform and engage with current events and tasks." "Agents use memory and connected databases to provide more relevant results based on richer historical data." "Agent interaction can be gradual, evolving and growing in complexity to empower users over the long term." "An agent is shaped by and evolves through continuous user interaction."	"Adapt your AI system with feedback from people, during individual interactions and over time." "Set up feedback mechanisms that can be used to interpret AI outcomes." "Account for changes introduced by AI." "Align datasets and models to people's interactions with AI systems in real-world contexts." "AI learns from users to personalize their experiences, and users adapt their behaviors and workflows in response to AI outcomes." "Human-AI interactions are a bidirectional feedback loop." "Adapt your AI system with feedback… over time." "Focus on creating experiences that seamlessly integrate and enhances aspects of work and play." "AI can amplify human expertise."
Trust (transparency, explanation of decisions and executions).	"Avoid… surface sensitive or private information directly in a card." "Avoid spam, jargon, or promotional language." "Not allowed: unsolicited promotions, upsells, or repeated attempts to re-engage." "Proactivity should feel like a natural continuation of the conversation, not an interruption."	"Agents require a huge degree of user trus – which must be earned." "Agents are not designed to replace or belittle people." "Humans can customize agent settings… and are in control of when the Agent is on/off." "Embrace uncertainty, but establish trust."	"People expect tools… to improve their efficiency… [and] prefer to remain in control." "Treat safety as an evolving endeavor." "Develop safeguards for different risk levels." "Be prepared to adapt and modify safety strategies… as user context evolves."

Source: Developed by the authors (2026)

5. Results and Discussion

This research aimed to analyze design principles from LLM-owning companies for the user experience of artificial intelligence agents, in light of the literature on human-centered human-AI interaction (HC-HAII). Based on a thematic documentary analysis of public materials, 22 global companies were examined, of which only 3 (OpenAI, Google, and Microsoft) presented specific corporate design principles for AI agents.

The comparison of these materials with theoretical references such as Amershi et al. (2019), Weisz et al. (2024), Liang et al. (2025), and Schröder & Fischer (2025) made it possible to identify convergences and gaps between corporate practice and theory. This scenario suggests that a large share of companies still treats agent design as an operational layer subordinated to the technical development of models, rather than as a strategic axis of human-AI interaction.

When compared with the references of Amershi et al. (2019), Weisz et al. (2024), Liang et al. (2025), and Schröder & Fischer (2025), the corporate principles present convergences, but also gaps. OpenAI presents guidance for conversational experience, simplicity, operational safety, and limits of use (good and bad applications, tone, proactivity), but does not make explicit mechanisms of iterative feedback, an axis that the literature highlights as essential for treating errors as opportunities for joint learning and for structuring iterative design cycles. Google, in turn, emphasizes user autonomy, sociotechnical alignment, evolving safety, and continuous feedback, but does not detail the visibility of the agent's internal behavior to the same degree present in Microsoft's guidelines. Microsoft offers the most systemic vision, closest to the notion of agents as sociotechnical "ecosystems," structuring principles in the space, time, and core of agents and emphasizing memory, progressive adaptation, transparency, control, and uncertainty.

In response to RQ1, the results indicate that the public communication of human-centered principles for agents is still incipient in the LLM ecosystem. Although there is a predominant discourse of responsibility and safe use on companies' websites, only 3 of the 22 organizations analyzed provide explicit design principles for agents in corporate-authored documents. This reinforces Xu's (2025) finding that HAII still does not fully integrate theory and practice: most companies focus on technical documentation for APIs, model performance, and use cases, but not on guiding how to design human-centered experiences with AI agents.

Regarding RQ2, it was observed that OpenAI, Google, and Microsoft converge on the ethical principle that agents should expand human capabilities without reducing their autonomy. OpenAI emphasizes the conversational and proactive experience, in order to avoid intrusion, spam, or loss of trust. Google organizes its principles around calibrated autonomy and alignment with the real world. Microsoft structures agents to simultaneously address historical memory, continuous evolution, and user control. These corporate principles point to an understanding that agents are sociotechnical partners and not automated tools.

In relation to RQ3, the thematic synthesis into five axes showed adherence, but no company fully covers the repertoire of the human-centered AI literature. In the axis of AI Visibility (technical clarity and human values), OpenAI and Microsoft directly approach the recommendations to "make clear what the system can do" and "make clear how well the system can perform" (AMERSHI et al., 2019), as well as value-oriented responsible design (WEISZ et al., 2024). In the axis of contextualized interactions with user control, the three companies converge with the literature by emphasizing context, mental models, graduated autonomy, and adjustment of the experience to users' preferences (WEISZ et al., 2024; LIANG et al., 2025; SCHRÖDER & FISCHER, 2025). In the axes of feedback mechanisms and adaptive evolution, Microsoft and Google are more strongly aligned with theory, by addressing feedback, uncertainty, continuous learning, and bidirectional human-AI cycles, while OpenAI presents gaps. Finally, in the axis of trust, the three companies reinforce the importance of transparency, control, and safety as a continuous process, in line with the proposals of Shneiderman (2020, 2022), Amershi et al. (2019), and Weisz et al. (2024).

The results suggest that the corporate guidelines analyzed incorporate elements of HC-HAII (XU, 2025), such as human agency, final control, transparency, and responsibility, but still in a fragmented way and with limited operational guidance for the daily work of designing and prototyping agents. While theory emphasizes prototyping cycles, explicit mechanisms for error correction, co-creation with users, systematic exploration of generative variability, and iterative testing methods, companies' guidelines tend to remain at a more descriptive and normative level. This reinforces the diagnosis of Schröder & Fischer (2025) that UX frameworks for generative AI still offer little practical guidance on prototyping methods and testing cycles with users.

As a contribution, this study: (i) comparatively maps the current state of public communication of design principles from three of the main global actors in AI agents; (ii) proposes a theoretical-thematic synthesis into five axes that integrates different streams of the HAII and HCAI literature; and (iii) highlights approximations and gaps between this theoretical framework and corporate practice. For designers, researchers, and product managers, the results suggest that adopting "best practices" provided by LLM providers is not enough; it is necessary to complement them with principles of human-centered human-AI interaction.

5.1 Limitations

This research presents limitations that should be considered when interpreting the results. As a descriptive-exploratory study, it does not seek to exhaust the discussion on design principles in the market or to encompass the entire theoretical production on human-centered AI. The empirical field includes 22 LLM-owning companies, but only three provide design principles for agents in consolidated corporate documentation. The analysis relies exclusively on public materials, which prevents assessing the extent to which such principles are actually applied in internal design and development practices. In addition, the use of thematic analysis involves the researcher's interpretations and judgments in the selection of verbatim excerpts, classification into thematic axes, and conceptual approximation with the literature. Finally, the temporal scope concerns a rapidly changing domain; thus, new guidelines or subsequent updates may alter the panorama presented, requiring continuous revisions.

5.2 Future Work

As future work, it is suggested: (i) to conduct empirical studies with designers and product teams to investigate how corporate guidelines are effectively applied in practice; and (ii) to systematically and practically compare design frameworks for agents, both from the literature and from the market, in order to support iterative design cycles and strengthen the development of ethical, understandable, and truly user-centered agents.

6. Conclusion

This research revealed that design principles for AI agents are still not widely disseminated in the LLM ecosystem. Among the 22 companies owning the analyzed models, only three (OpenAI, Google, and Microsoft) publicly provide guidelines applicable to agent design, revealing a mismatch between the discourse of responsibility in AI and the provision of guidance for product teams.

From a practical standpoint, the results show that corporate guidelines only partially cover the requirements for human-centered design, leading designers, engineers, and managers to rely on specialized literature to complement conceptual and operational gaps. At the same time, companies have the opportunity to evolve these guidelines into more complete frameworks, integrating UX principles, prototyping practices, and mechanisms of ethical governance. In summary, this study demonstrates that companies are moving toward more human-centered AI and contributes to guiding the future development of AI agents that are not only functional, but also understandable, reliable, and aligned with human agency.

6.1 Author Note

Generative artificial intelligence (ChatGPT, OpenAI) was used as an auxiliary tool for grammatical and academic writing refinement and for translating the manuscript, originally written in Portuguese, into English. The final text was reviewed and edited by the authors.

References

ACHARYA, D. B.; KUPPAN, K.; DIVYA, B. **Agentic AI: autonomous intelligence for complex goals—A comprehensive survey.** IEEE Access, v. 13, p. 18912–18936, 2025. https://doi.org/10.1109/ACCESS.2025.10849561.

AMERSHI, S. et al. **Guidelines for human-AI interaction.** In: CHI '19: CHI Conference on Human Factors in Computing Systems, 2019, Glasgow, Scotland, UK. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 2019. p. 1–13. https://doi.org/10.1145/3290605.3300233.

CAPOSINO, M. **AI agents and UI design: creating engaging interfaces and top trends.** 2025. Disponível em: https://fuselabcreative.com/ui-design-for-ai-agents/.

CEREJO, J. **The design process of human-centered AI.** Design Bootcamp, 2023. Disponível em: https://medium.com/design-bootcamp/human-centered-ai-design-process-part-2-empathize-hypothesis-6065db967716.

CHACON-CHAMORRO, M. et al. **Cooperative resilience in artificial intelligence multiagent systems.** IEEE Transactions on Artificial Intelligence, 2025. https://ieeexplore.ieee.org/document/10988739.

GOOGLE. **People + AI Guidebook.** 2025. Disponível em: https://pair.withgoogle.com/guidebook/.

GRUDIN, J. **From tool to partner: the evolution of human-computer interaction.** In: CHI Conference on Human Factors in Computing Systems, 2018, New York, NY, USA. Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, 2018. p. 1–3. https://doi.org/10.1145/3170427.3170663.

GUENAT, S. et al. **Meeting sustainable development goals via robotics and autonomous systems.** Nature Communications, v. 13, n. 1, p. 3559, 2022. https://doi.org/10.1038/s41467-022-31150-5.

HASSENZAHL, M. **The thing and I: understanding the relationship between user and product.** In: BLYTHE, M.; MONK, A. (org.). Funology 2. Cham: Springer, 2018. p. 301–313. https://doi.org/10.1007/978-3-319-68213-6_19.

HSIEH, H.-F.; SHANNON, S. E. **Three approaches to qualitative content analysis.** Qualitative Health Research, v. 15, n. 9, p. 1277–1288, 2005. https://doi.org/10.1177/1049732305276687.

HUI, B. et al. **PLeak: prompt leaking attacks against large language model applications.** In: ACM SIGSAC Conference on Computer and Communications Security (CCS '24), 2024, Salt Lake City, USA. Proceedings of the 2024 ACM SIGSAC Conference on Computer and Communications Security, 2024. p. 3600–3614. https://doi.org/10.1145/3658644.3670370.

JIANG, T. et al. **Human-AI interaction research agenda: a user-centered perspective.** Data and Information Management, v. 8, n. 4, p. 100078, 2024.

LIANG, J. T. et al. **AgentBuilder: exploring scaffolds for prototyping user experiences of interface agents.** arXiv, 2025. http://arxiv.org/abs/2510.04452.

MICROSOFT. **UX design for agents.** 2025. Disponível em: https://microsoft.design/articles/ux-design-for-agents/.

MORGAN, H. **Conducting a qualitative document analysis.** The Qualitative Report, 2022. https://nsuworks.nova.edu/tqr/vol27/iss1/4/.

NAVEED, H. et al. **A comprehensive overview of large language models.** ACM Transactions on Intelligent Systems and Technology, v. 16, n. 5, 2025. https://doi.org/10.1145/3744746.

NISA, U. et al. **Agentic AI: the age of reasoning—A review. Journal of Automation and Intelligence,** 2025. https://doi.org/10.1016/j.jai.2025.1000516.

OPENAI. **App design guidelines.** 2025. Disponível em: https://developers.openai.com/apps-sdk/concepts/design-guidelines.

REYNOLDS, L.; MCDONELL, K. **Prompt programming for large language models: beyond the few-shot paradigm.** In: CHI Conference on Human Factors in Computing Systems, 2021, Yokohama, Japan. Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems, 2021. p. 1–7. https://doi.org/10.1145/3411763.3451760.

SADO, F. et al. **Explainable goal-driven agents and robots – a comprehensive review.** ACM Computing Surveys, v. 55, n. 10, 2023. https://doi.org/10.1145/3564240.

SAPKOTA, R.; ROUMELIOTIS, K. I.; KARKEE, M. **AI agents vs. agentic AI: a conceptual taxonomy, applications and challenges.** Information Fusion, 2026. http://arxiv.org/abs/2505.10468.

SCHRÖDER, F.; FISCHER, M. **State-of-the-art UX frameworks for human-centered AI in generative AI systems: a systematic literature review.** In: DEGEN, H.; NTOA, S. (org.). Artificial Intelligence in HCI. Cham: Springer, 2025. p. 435–450. https://doi.org/10.1007/978-3-031-93412-4_27.

SHNEIDERMAN, B. **Human-centered AI.** Oxford: Oxford University Press, 2022. https://academic.oup.com/book/41126.

SHNEIDERMAN, B. **Human-centered artificial intelligence: reliable, safe & trustworthy.** International Journal of Human–Computer Interaction, v. 36, n. 6, p. 495–504, 2020. https://doi.org/10.1080/10447318.2020.1741118.

SÜSSE, T.; KOBERT, M.; KRIES, C. Antecedents of constructive human-AI collaboration: an exploration of human actors' key competencies. In: CAMARINHA-MATOS, L. M.; BOUCHER, X.; AFSARMANESH, H. (org.). **Smart and Sustainable Collaborative Networks 4.0.** Cham: Springer, 2021. p. 113–124. https://doi.org/10.1007/978-3-030-85969-5_10.

VASWANI, A. et al. **Attention is all you need.** arXiv, 2023. http://arxiv.org/abs/1706.03762.

XU, W. et al. **From human-computer interaction to human-AI interaction: new challenges and opportunities for enabling human-centered AI.** arXiv, 2021. http://arxiv.org/abs/2105.05424.

XU, W. **Human-centered human-AI interaction (HC-HAII): a human-centered AI perspective.** arXiv, 2025. http://arxiv.org/abs/2508.03969.

XU, W. et al. **Transitioning to human interaction with AI systems: new challenges and opportunities for HCI professionals to enable human-centered AI.** arXiv, 2023. http://arxiv.org/abs/2105.05424.

WEISZ, J. D. et al. **Design principles for generative AI applications. In: CHI Conference on Human Factors in Computing Systems.** 2024, Honolulu, USA. Proceedings of the CHI Conference on Human Factors in Computing Systems, 2024. p. 1–22. https://doi.org/10.1145/3613904.3642466.

WEISZ, J. D. et al. **Toward general design principles for generative AI applications.** arXiv, 2023. http://arxiv.org/abs/2301.05578.

ZAMFIRESCU-PEREIRA, J. D. et al. **Why Johnny can't prompt: how non-AI experts try (and fail) to design LLM prompts.** In: CHI Conference on Human Factors in Computing Systems, 2023, Hamburg, Germany. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 2023. p. 1–21. https://doi.org/10.1145/3544548.3581388.

COMO CITAR ESTE CAPÍTULO | ABNT

FERREIRA, Fernando Souza; CONSONI, Gilberto; VAN DER LINDEN, Júlio Carlos de Souza. Experience design for AI agents: an analysis of human-centered guidelines in LLM-owning companies. *In:* NICHETTI, C.; VENTURA, D.; VAN DER LINDEN, J. C. S.; MACHADO, L. (org.). **Design, culture and innovation:** perspectives, experiences and connections. Porto Alegre: Marca Visual, 2026. v. 5, p. 62-87.

HOW TO CITE THIS CHAPTER | HARVARD

FERREIRA, F. S.; CONSONI, G.; VAN DER LINDEN, J. C. S. (2026) 'Experience design for AI agents: an analysis of human-centered guidelines in LLM-owning companies'. *In:* Nichetti, C.; Ventura, D.; Van Der Linden, J. C. S.; Machado, L. (eds.). *Design, culture and innovation: perspectives, experiences and connections.* Vol. 5. Porto Alegre: Marca Visual, pp. 62-87.