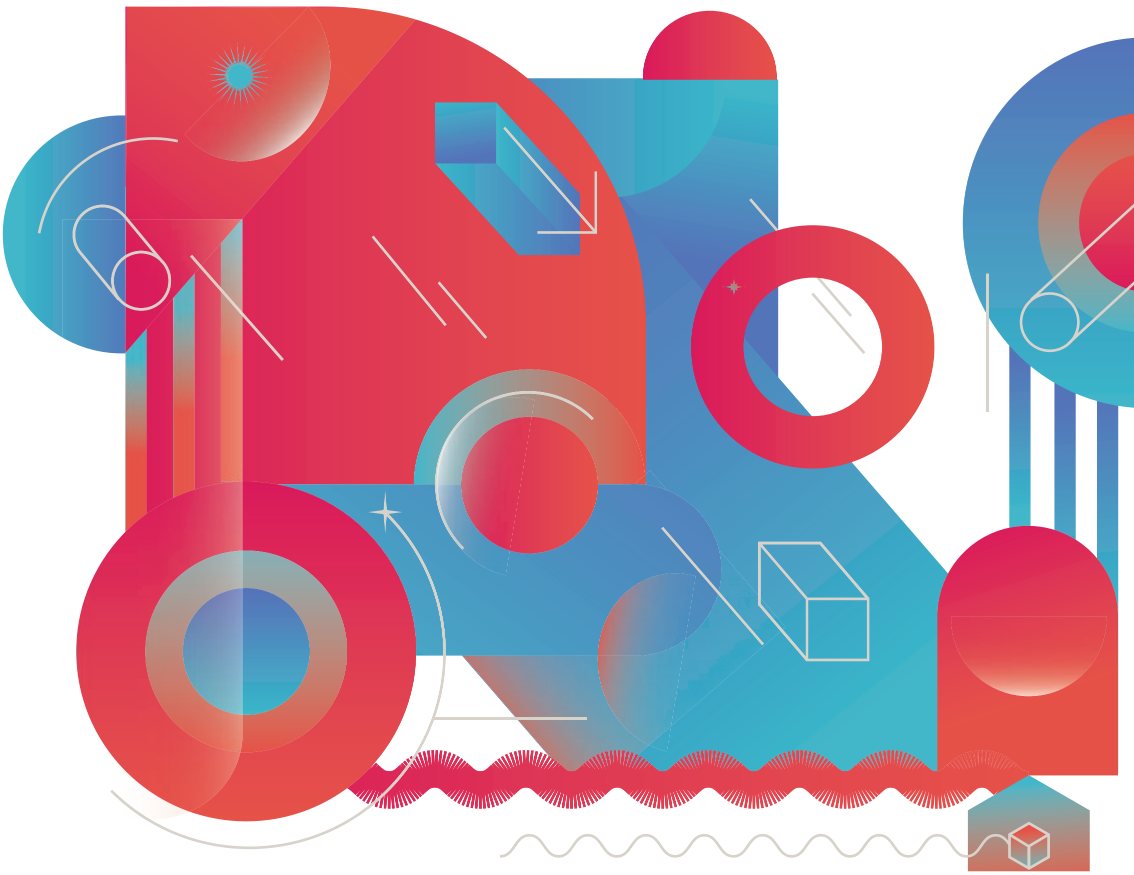
EMBEDDING INTELLIGENCE

Designerly reflections on AI-infused products

edited by Davide Spallazzo, Martina Sciannamè



D.I. FrancoAngeli
DESIGN INTERNATIONAL

The Design International series is born in 2017 as a cultural place for the sharing of ideas and experiences coming from the different fields of design research, becoming a place in which to discovering the wealth and variety of design, where different hypotheses and different answers have been presented, drawing up a fresh map of research in international design, with a specific focus on Italian design.
Different areas have been investigated through the books edited in these years, and other will be explored in the new proposals.
The Scientific Board, composed by experts in fashion, interior, graphic, communication, product and industrial, service and social innovation design, interaction and emotional design, guarantee the high level of the accepted books. After the first selection by the Scientific Board, the proposals are submitted to a double review by other international experts.

EMBEDDING INTELLIGENCE

Designerly reflections on AI-infused products

edited by Davide Spallazzo, Martina Sciannamè

D.I. FrancoAngeli DESIGN INTERNATIONAL
OPEN ACCESS

Cover image by Sara Sciannamè

ISBN e-book Open Access: 9788835141914

Date of first publication: September 2022

Contents

2. User Experience and AI-infused products. A wicked relationship

Davide Spallazzo
Department of Design, Politecnico di Milano

The current panorama of AI-infused devices portrays a significant dominance of first-party smart speakers, which appear to be the first massive embodiment of AI in the domestic landscape. These devices are nothing more than discreet ornaments, looking at their simple physical appearance. Although, the simple appearance betrays a complexity determined by numerous features that make such products challenging to analyze from a UX point of view.

The main evident characteristic is that they are not just “simple products” but ecosystems consisting of several interfaces and touchpoints. Most of them integrate multiple interfaces – namely physical, digital, conversational – sometimes overlapping.

The second element of complexity resides in their technological core, based on learning algorithms. Therefore, the same device can provide different outputs at the same input over time, a condition that can affect the user experience.

To increase the complexity of these devices, at least from a UX standpoint, there is the fact that their real potential is rarely exploited by most users, which mainly uses routine actions such as reading news, weather forecasting, and controlling simple home appliances.

Accordingly, the chapter frames the wicked relationship between user experience and AI-infused products. Moving from the three identified elements of the complexity of AI-infused products, it advances reflection on how it could be possible to analyze these products from a UX standpoint.

2.1 Introduction

Continuous technological advancement has made it feasible to design ubiquitous systems capable of emulating human behavior, resulting in personal assistants that find noteworthy application in many sectors (e.g., medicine, security, transport, education). Conversational AI-based agents such as Amazon Alexa are spread in millions of houses, populating the domestic environment with products powered with Artificial Intelligence (AI) capabilities. AI-infused products can adapt autonomously to the context and users' needs, entertain a dialogue with them, recognize them, and track/anticipate their behavior.

This market has already known a great success. However, it is expected to exponentially grow in the next few years, reaching billions of products integrating smart assistants, being them first-party or third-party hardware (*Canalys Newsroom – Rise of Alexa, Google Assistant and Siri to drive US smart assistant-compatible device base to 1.6 billion in 2022*, no date). Even if not so significant in market share, we may also add to the count personal and domestic robots (e.g., the renowned robot Pepper) provided with AI-based conversational agents, which accompany users in daily tasks.

These goods, which can be considered emblematic of the first wave of AI embodiment in the domestic setting, drew much interest. Nonetheless, the HCI community and the design discipline at large have only recently developed a comparable interest.

The design world, and academia, in particular, has yet to address this issue thoroughly, although AI has been touted as a new material for designers (Holmquist, 2017; Antonelli, 2018). Some studies have analyzed machine learning as a design subject (Dove *et al.*, 2017), the use of virtual assistants in everyday life (Sciuto *et al.*, 2018) and their aesthetic-functional dimensions in the domestic realm (Spallazzo, Sciannamè and Ceconello, 2019). Similarly, other studies coped specifically with conversational interfaces (Vitali and Arquilla, 2019) or reviewed ways for evaluating the user experience enabled by voice-based interactions (Kocaballi, Laranjo and Coiera, 2018).

However, studies investigating the user experience enabled by AI-infused products are needed. Indeed, these devices are frequently perceived as gadgets or toys (Levinson, 1977) that tickle the users' craving novelties rather than playing a significant role in their lives.

Likewise, they are not exceptionally noteworthy in terms of interaction quality, sometimes generating frustration and an essential use unable to unlock their potential entirely (Sciuto *et al*., 2018).

To the best of our knowledge, no systematic attempt has been made to frame the UX entailed by AI-infused products, coping with these devices' complex nature holistically. Accordingly, we consider it relevant to understand and frame the User Experience (UX) they entail and review current UX assessment methods to understand their adequacy for AI-infused systems.

2.1.1 Understanding AI-infused systems

Personal assistants, robots, and self-driving cars are entirely changing the interaction paradigm we are used to. They are no more tools used by humans as extensions of their bodies or minds. AI-infused systems are perceived as *counterparts* (Hassenzahl *et al*., 2020). As such, they introduce a shift from the paradigm of embodiment (Dourish, 2001) to *alterity*, as stated by Hassenzahl and colleagues who defined these products *otherware* (Hassenzahl *et al*., 2020).

Perceiving interactive products as others and not as users' extensions entails introducing new interaction paradigms. We do not interact with voice assistants; we converse with them. Relying on the robustness of the system is no more enough. We must trust them since most of the operations (deep-learning) AI-infused products perform are frequently opaque to end-users, and their outputs are somehow unpredictable. Novel forms of interaction imply a different relationship between humans and machines and necessarily may impact the perception of machines and the user experience they enable.

AI-infused systems learn, reflect, talk, and clearly show their agency to end-users through proactivity. Referring to the three kinds of agency – conditional, need-based, delegated – proposed by Kaptelinin and Nardi (Kaptelinin and Nardi, 2009), these systems go beyond the delegated agency. Indeed, they can betray users' expectations or act autonomously, or, better, choose the best solution for a problem they have been delegated to solve.

Products integrating AI capabilities (e.g., Amazon Echo family) may appear nothing more than inobtrusive ornaments. However, it is evident they betray an intrinsic complexity, increased by features that make such products challenging to analyze from a UX point of view.

The primary distinguishing feature is that they are not only "simple goods", but rather ecosystems with several interfaces and contact points. Most of them combine various interfaces – physical, digital, and conversational – that occasionally overlap in terms of functionality. Being an ecosystem also entails communicating with – and even controlling – other devices. This is especially true for domestic devices, which frequently act – or are supposed to act – as a control hub for home entertainment and an ever-increasing number of connected objects in complex dynamic environments like our homes.

Secondly, their technical heart, based on learning algorithms and neural networks, adds a second layer of complexity. It implies that the same device might provide multiple outputs at the same input over time, a circumstance that could influence UX evaluation in traditional methods.

To add to the complexity of these products, at least from the standpoint of UX evaluation, their true potential is rarely explored by most users. They primarily perform ordinary behaviors such as reading news, weather forecasting, and operating simple home appliances (Sciuto *et al.*, 2018; White, 2018). A condition that may prompt, on the one hand, a conversation on discoverability, as advocated by White (2018), and, on the other, a more in-depth assessment of the role of such systems in our daily lives.

2.1.2 Aims and methodological approach

In line with the reflections expressed above, the hypotheses here advanced are that (i) the user experience enabled by such systems can hardly be described and framed with current assessment methods, and (ii) new qualities of the user experience must be considered. Consequently, a (iii) novel assessment method is required to evaluate the experience enabled by AI-infused systems holistically. Accordingly, the following sections aim to answer three main research questions: *[RQ1] Are current UX assessment methods enough for AI-infused products? [RQ2] Are new UX dimensions needed for these products? [RQ3] What characteristics the new method could have?*

Secondary research has been performed to answer these questions with the specific aims of (i) tracking the evolution of UX assessment and (ii) understanding current methods and their adequacy in evaluating AI-infused systems. The first analysis is confined to works published between 2000 and 2020 in the ACM Digital Library and Springer Link, retrieved through the query strings *UX AND evaluation* and *UX AND assessment*. It focused on the few articles tracing the evolution of UX and its assessment over time summarized in Figure 2.1.

Authors	**Sample of UX eval. meth.**
(Vermeeren *et al.*, 2010)	96
(Bargas-Avila and Hornbæk, 2011)	66
(Lachner *et al.*, 2016)	84
(Rivero and Conte, 2017)	227
(Pettersson *et al.*, 2018)	100

Fig. 2.1 – Most relevant articles tracing the evolution of UXEM and relative sample of UX studies.

A subsequent and more comprehensive study followed a snowball sampling approach moving from the articles listed above, to deepen specific methods. At the same time, the long list of methods collected by the researchers of the Allaboutux.org website served as a reference point to have a broad overview of UX evaluation methods.

2.2 User Experience and its assessment

Since the term User Experience (UX) was introduced around the turn of the millennium, academics have sought novel approaches to understanding and assessing the quality-in-use of interactive systems (Bargas-Avila and Hornbæk, 2011). Leading this change, which began in the HCI sector, was the perception that contemporary assessment techniques were overly focused on system usability and efficiency, missing more general but no less essential elements such as quality, pleasantness, and meaning.

This turn has been recalled by Hassenzahl, Burmester and Koller in a recent work (Hassenzahl, Burmester and Koller, 2021) that traces back twenty years of reflection on UX, moving from their seminal article (Burmester, Hassenzahl and Koller, 2002), which challenged the univocal association between usability and user experience.

Throughout the years, numerous studies have advanced alternative methodologies for understanding, conceptualizing, and supporting the creation of meaningful experiences with interactive systems (Hassenzahl and Tractinsky, 2006), awakening a discussion based on differing epistemological perspectives. Simultaneously, the scope of the studies steadily expanded from the exclusive instrumental and task-oriented evaluation of usability to encompass, to mention a few, pleasure (Jordan, 2000), positive emotions (Norman, 2004), and aesthetics (Tractinsky, Katz and Ikar, 2000).

The academic world actively adopted this unique perspective, either alone or in collaboration with companies and consulting firms, resulting in many publications dealing with UX evaluation of interactive systems, whether they are industrial products, digital interfaces, or services.

The proposed solutions ranged from specific tests (Schmettow, Noordzij and Mundt, 2013) to the deployment of evaluative tools (Lugmayr and Bender, 2016; Minge *et al.*, 2017; Sivaji, Nielsen and Clemmensen, 2017; Almeida *et al.*, 2018; Zhou *et al.*, 2019; Maguire, 2019) to more holistic and broad methods (Obrist, Roto and Väänänen-Vainio-Mattila, 2009; Kujala *et al.*, 2011; Jiménez *et al.*, 2012; Otey, 2017).

In 2010, Vermeeren *et al.* (2010) gathered 96 relevant and unique UX assessment techniques from academia and industry and classified them according to various criteria. The investigated approaches were generally used on digital interfaces (such as websites and mobile applications) and fully functional systems or working prototypes, interfering at an advanced stage of development.

Bargas-Avila and Hornbæk (2011) did a similar wide-ranging critical study concentrating on empirical methodologies for the evaluation of the UX, with an in-depth investigation of 51 publications out of 1254 arising from digital library research. Their research revealed a high frequency of digital interfaces as examined systems and a propensity to evaluate the overall UX and UX aspects such as affect, emotion, enjoyment, aesthetics, attractiveness and hedonic attributes. They also clearly demonstrated that researchers generally utilized self-developed questionnaires and that classic inquiry techniques like questionnaires, interviews, and diaries were the most often used to obtain qualitative data.

Rivero and Conte (2017) conducted a thorough review of UX evaluation methodologies for digital interfaces in 2017. Their findings are similar to those of the previous studies: the study methods are largely traditional (e.g., questionnaires, observation) and used with groups of final users in a controlled setting during or after the usage experience. Differentiating from the previous ones, this study emphasizes the importance of quantitative data as the primary output of the UX assessment.

One year later, Pettersson and colleagues performed a comprehensive review of 100 academic articles published between 2010 and 2016 describing empirical investigations on user experience evaluation focusing on reliability. The authors explicitly addressed the triangulation of techniques in user experience assessment (Pettersson *et al.*, 2018). The review of the state-of-the-art highlights findings that are comparable to those found in the studies mentioned above. The most usually addressed UX characteristics are overall UX and pragmatic features (usability), digital interfaces are the most commonly evaluated items, and self-developed surveys are frequently used as an inquiry method. Additionally, the authors highlight a current trend toward employing – and triangulating data from – various ways of inquiry to better comprehend and contextualize the results. The study emphasizes four open topics for future UX research, listing, among others, the necessity of adapting to ever-evolving technologies and non-human intelligence. The open question is how to perform this adaptation. Accordingly, in the following section, we analyze existing UX assessment methods that may adapt or not to the unique peculiarities of AI-infused systems and try to envision general traits of a novel method.

2.3 Coping with complexity

A first point to address is AI-infused systems' unique nature: ecosystems rather than single, self-standing products. Taking domestic smart speakers as a reference for the discussion, we may state that several devices – e.g., light bulbs, doorbells, cameras, and thermostats – can currently be linked and controlled, creating an integrated system with the smart speaker as its hub.

The user experience allowed by such an ecosystem (i) will evolve and necessarily (ii) involve a variety of touchpoints, whether physical, digital, or based on conversational interfaces.

Beyond this multiplicity of touchpoints in the ecosystem, the devices themselves can be considered complex artifacts from an interaction standpoint. Usually, smart speakers integrate physical buttons (e.g., volume, mute, activation) and are all equipped with conversational interfaces that duplicate more or less the same functions.

In addition, a companion smartphone app enables users to configure the device and conduct a variety of tasks, such as broadcasting music or controlling linked devices.

To add complexity from a UX standpoint, these products usually provide feedback on their current state through different means, whether with colored LED or the assistant's voice.

Looking at the current panorama of UX assessment methods, we could investigate different aspects separately. SUS (Brooke, 1995), Kansei (Schütte *et al.*, 2006) and AttrakDiff (Hassenzahl, Burmester and Koller, 2003) can be used to assess the product's usability/pragmatic or hedonic aspects on the hardware side. Similarly, we may use the well-known Nielsen and Molich heuristics (1990) to evaluate the companion app's user interface and eventually rely on specific methodologies to evaluate the user experience of conversational interfaces (Pyae and Joelsson, 2018; Maguire, 2019).

Still, it would be problematic to conduct a comprehensive user experience assessment to render a holistic view of the quality of the interaction enabled by the devices and their ecosystem.

This first insight may suggest the need for a novel approach to deal with the complexity that may consist in (i) a custom technique integrating existing ones or (ii) a generic method capable of capturing the spirit of the whole experience. The first approach may require the evaluation of the peculiar characteristics of distinct interfaces and touchpoints and, consequently, a broad and modular method to be tailored to each unique scenario under assessment.

The second option appears to be the most popular, based on critical reviews of existing UX rating methodologies. According to Bargas-Avila and Hornbæk's study, 41% of techniques evaluate generic user experience (UX) (Bargas-Avila and Hornbæk, 2011), and the ratio increases to 56% in a more recent examination by Pettersson *et al.* (2018).

The second point of complexity is that AI-infused products rely on learning algorithms. Accordingly, the same system might produce different outputs for the same input over time.

So, time acquires a great relevance in the user experience since the more users interact with the system, the more accurate it should become in anticipating/answering needs and requests.

As a result, assessing the experience over single episodes or even a whole day may produce unreliable data. In contrast, longitudinal research could presumably provide a more accurate picture of the user experience.

The study by Vermeeren and colleagues (2010) outlines a variegated picture of UX assessment methods regarding the periods of the experience objects of analysis. They range from unique snapshots/episodic activities to long-term usage through episodes/specific tasks. Looking at the percentage emerging from their analysis, 36% of the methods object of study already assesses systems' performances in everyday life over a long-term interaction. A different result is highlighted in the more recent work by Rivero and Conte, who point out a much more limited percentage – 6.6% – of methods assessing long-term interactions (Rivero and Conte, 2017). Regardless of the percentage, evaluating the user experience with a longitudinal approach is standard practice.

This second insight may indicate the need to choose/create a method to assess an experience over a long time and, eventually, track how the quality of the experience may evolve (Karapanos *et al.*, 2009).

The third level of complexity in assessing the UX of AI-infused systems is that they are relatively new and may present inherent flaws in the usage. Two recent studies focusing on Amazon's smart speakers highlight, for example, that users rarely acknowledge and understand the actual potential of such devices. This condition may be due to poor discoverability and a lack of proactivity (Sciuto *et al.*, 2018; White, 2018).

This condition raises further doubts on the ability of current UX evaluation methods to assess AI-infused systems holistically: if the user does not fully recognize the systems' agency and potential, how can she fully perceive her experience as unsatisfactory or limited?

Furthermore, an assessment approach focused exclusively on using existing systems/working prototypes, even if correctly developed, may disclose just a few issues that can be addressed through the iterations of the last stages of the design process.

Going back to the studies on Amazon's devices, proactivity can soothe some of the difficulties associated with discoverability (White, 2018). Nevertheless, adding proactivity to a system is not a simple task that can be performed in the last adjustments before the product's launch on the market. It is also true that using potentially powerful devices such as smart speakers for simple weather forecasts or as music players may reveal weaknesses in the reflection on the product's intended usage in the domestic environment. These reflections must be done early in the design process.

Looking again at the study by Vermeeren *et al.*, it emerges that only a low percentage – 25% – of current UX evaluation methods may be used in the early stages of product development. The majority of methods focus on working systems (Vermeeren *et al.*, 2010). Accordingly, a method to be applied at different stages of the design process may be considered a good option for AI-infused systems.

The last element of complexity here discussed relies on the very nature of AI-infused products, that of being *otherware* (Hassenzahl *et al.*, 2020).

The question is whether current UX evaluation methods are equipped to assess something not perceived as an extension of users but rather as a counterpart. Looking back at the comprehensive analyses of UX evaluation methods, they provide a very traditional portrait in terms of objects of study.

The analysis by Vermeeren *et al.* (2010) highlights an important focus on web services (81%), mobile software (77%) and PC software (76%). Very similar results emerge from the study by Bargas-Avila and Hornbæk (2011), which points out a tendency to assess products such as mobile applications and phones, audio, video, TV applications. Interestingly they also underline that 21% of the assessed methods regard art (e.g., audio photography, interactive canvas). These two studies come far before the spread of AI-infused products and do not track any methods evaluating specifically such systems.

A more recent study by Rivero and Conte (2017) highlights a significant percentage of methods broad enough to assess any type of interface (33.9%). Their results further portray attention to web (13.7%) and mobile (8.8%) applications.

One year later, Pettersson *et al.* (2018) provide a fragmented image of products studied in UX research. In their ranking the most assessed systems are mobile apps (15%), interactive games (13%) – that make their first appearance –, webtools (12%) and websites (10%). At the same time, a small percentage (4%) of methods address connected/IoT services.

The analysis of the criticalities of AI-infused systems against current UX evaluation methods suggests a clear answer to [RQ1] *(Are current UX assessment methods enough for AI-infused products).* So far, UX research has not explicitly addressed AI-infused systems and the creation of a novel, bespoke method to address such systems is explicitly needed.

2.4 Understanding UX dimensions

The reflections above indicate the necessity of a novel UX evaluation method to assess the experience enabled by AI-infused systems. Following this line, it is essential to understand if the dimensions of the user experience assessed so far are enough for these systems or if new dimensions must be elicited.

UX dimensions are at the heart of any evaluation process. More than any other characteristic, they evolve through time, reflecting an ever-changing form of comprehending and framing user experience.

Looking back at systematic reviews of UX evaluation methods (Bargas-Avila and Hornbæk, 2011; Pettersson *et al.*, 2018), we can assert that generic UX is by far the most researched dimension, despite its ambiguous description. According to Bargas-Avila and Hornbæk (2011), generic user experience is generally evaluated qualitatively and refers to an overall perception of the experience. Similarly, Pettersson *et al.* (2018) describe generic UX as a broad concept that may be evaluated holistically. Focusing on the whole experience may help manage complexity in multi-interface and multi-touchpoint ecosystems, like those we are studying. However, it may also be constraining and incapable of providing valuable inputs throughout the design process and go to the core of AI-infused systems.

Pragmatic aspects – namely, usability – rank second in the research by Pettersson *et al.* (2018). However, they were not even included as a

UX dimension seven years before in Bargas-Avila and Hornbæk (2011), underlining a deliberate early dissociation from usability concerns to distinguish the new methods. Nevertheless, it must be noted that some of the most popular AI-infused systems do not fully address usability concerns due to a lack of in-depth reflection on basic interaction design principles such as input, output, and feedback modalities. For instance, a recent analysis of AI-infused devices (Spallazzo, Sciannamè and Ceconello, 2019) shows that these devices provide almost no inherent feedback, as physical actions are required in a limited manner; thus, only functional and augmented feedbacks appear to characterize current domestic assistants. Accordingly, pragmatic questions cannot be given for granted and should be evaluated, particularly in the early phases.

Nonetheless, there is little question that usability, emotion/affect, and enjoyment play a significant role in the UX. Not surprisingly, these aspects are the most often examined after general UX (Bargas-Avila and Hornbæk, 2011; Pettersson *et al.*, 2018). Moreover, reflecting on the emotional reaction to objects/environments asks for an examination of the aesthetics/hedonic aspects of items.

Previous studies on the embodiment of AI capabilities in the domestic environment identify an immature perspective on the embodiment of AI in the domestic realm (Spallazzo, Sciannamè and Ceconello, 2019) and the resulting necessity to incorporate an aesthetic dimension into the evaluation technique, particularly during the early prototype phases.

As demonstrated in Bargas-Avila and Hornbæk (2011) and in Pettersson *et al.* (2018), the UX dimensions discussed above are more or less handled by the methodologies current assessment methodologies do not cover dimensions likely relevant to AI-infused systems.

At the same time the impression is that these dimensions, despite essential, seem unable to track the very nature of these devices. So, coming to *[RQ2] Are new UX dimensions needed for these products,* we may state that traditional UX dimensions must be included in the analysis, but they are not enough to assess the core qualities of AI-infused systems and their impact on the user experience.

Accordingly, in the next section, we outline the characteristics that a new tool could have in general terms.

2.5 Defining traits of a new UX evaluation method

The answer to *[RQ3] What characteristics the new method could have?* requires deep reflection and a structured research activity. Nevertheless, we consider it beneficial to stimulate a fruitful discussion advancing broad defining characteristics of a novel UX evaluation method for AI-infused products.

Following Bargas-Avila and Hornbæk' suggestion (Bargas-Avila and Hornbæk, 2011), we define a new evaluation method by describing the methodology to be used and the dimensions of the user experience and the types of goods to be studied.

Given for granted that the objects of study are AI-infused systems, we may firstly summarize some hints regarding the methodological approach.

The discussion on the complexity of such systems already highlighted four main traits for the novel method that we may summarize as:

1. Flexible and holistic, to capture the nature of ever-evolving ecosystems.
2. Able to assess an experience over a long time and, eventually, track how the quality of the experience may evolve.
3. Appliable at different stages of the design process, even at the early stages.
4. Capable of capturing the very core of AI-infused systems as *otherware* rather than users' extensions.

Defining a methodology means also taking a position in the long-lasting debate between the quantitative and qualitative approaches, that move from different epistemological standpoints, one considering UX as something quantifiable and the other considering simplistic the idea of measuring the experience (Bargas-Avila and Hornbæk, 2011; Pettersson *et al*., 2018).

Vermeeren and colleagues highlight a fair distribution of approaches that use solely quantitative data, purely qualitative data, or both in their study (Vermeeren *et al*., 2010). These percentages contrast with those obtained by Bargas-Avila and Hornbæk. They found a majority of pure qualitative approaches (50%) over quantitative ones (33%), as well as those combining the two (17%). The numbers are inverted in the research by Rivero and Conte, with around 58 percent of techniques collecting quantitative data, 14 percent collecting only qualitative data, and lastly, 28 percent collecting both forms of data (Rivero and Conte, 2017). Very varied results may be explained by the different methods used by the researchers to select the available papers for the study. These results do not provide an excellent basis to consider in the definition of the new method. Nonetheless, we can see the need for triangulation of different methods in the divergences of the studies mentioned above, as proposed by (Vermeeren *et al.*, 2010) and investigated by (Pettersson *et al.*, 2018). They highlight a growing trend toward using more than one method to increase the reliability of the results.

We believe that this is a potential direction to pursue to understand the complicated nature of the user experience generated by AI-enhanced home devices. Thus, it is unsurprising that one of the first research focused on Amazon Echo (Sciuto *et al.*, 2018) used a mixed-methods approach that included logs and qualitative interviews.

The second element to be considered in defining a new evaluation method regards the UX dimensions they assess. As stated in the previous section, the current UX evaluation methods mainly focus on traditional UX dimensions that appear unable to fully frame the experience enabled by AI-infused systems.

They do not assess the meaning of engaging with other intelligences (Hassenzahl *et al.*, 2020) embedded in AI-infused systems. Understanding meaning as a definable – and even quantifiable – element in the user experience could play a prominent role in products whose potential is not fully exploited (White, 2018). Indeed, the meaningfulness of a user experience requires reflection on both the physical environment, which is frequently overlooked in existing paradigms and the social context. Additionally, it entails a close relationship with the users' history and motivations and their ethical viewpoints.

At the same time, it becomes mandatory to reflect on the perception of the other intelligence itself and the quality of the relationship it may establish with the user. The study on the UX of conversational interfaces directly consequences the previous statement since voice interaction is becoming a common trait of AI-infused systems (Kocaballi, Laranjo and Coiera, 2018).

Furthermore, the ethical implications of systems that collect and manage a massive amount of data – even personal ones – to work correctly should be somehow considered. Instead of or in addition to reliability, assessing trustworthiness could be relevant to such systems.

2.6 Conclusions

AI-infused products are game-changing, and the UX research appears unprepared to cope with novel paradigms of interaction, entailing a very different way of understanding interactive systems and their agency.

The research discussed so far confirms the assumption that current UX evaluation methods are not entirely adequate to holistically assess the user experience enabled by AI-infused systems. Furthermore, the UX dimensions commonly assessed provide a general understanding of the user experience, but do not analyze the very core of such systems.

Accordingly, a novel UX evaluation method specifically addressed to AI-infused system is needed to provide guidance in the development and improvement of such system. In the last section we highlighted broad traits of a method to be, with the aim of stimulating discussion within the design discipline.

References

Almeida, R.L.A. *et al.* (2018). "Towards Developing a Practical Tool to Assist UX Evaluation in the IoT Scenario". *Anais Estendidos do Simpósio Brasileiro de Sistemas Multimídia e Web* (WebMedia), pp. 91-95. doi:10.5753/webmedia.2018.4575.

Antonelli, P. (2018). *AI Is Design's Latest Material.* Available at: https://design.google/library/ai-designs-latest-material/ (Accessed: 6 February 2019).

Bargas-Avila, J.A. and Hornbæk, K. (2011). "Old Wine in New Bottles or Novel Challenges: A Critical Analysis of Empirical Studies of User Experience". In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM (CHI '11), pp. 2689-2698. doi:10.1145/1978942.1979336.

Brooke, J. (1995). "SUS: A quick and dirty usability scale". *Usability Eval. Ind.*, 189.

Burmester, M., Hassenzahl, M. and Koller, F. (2002). "Usability ist nicht alles – Wege zu attraktiven Produkten (Beyond Usability – Appeal of interactive Products)". *i-com*, 1(1), p. 32. doi:10.1524/icom.2002.1.1.032.

Canalys Newsroom – *Rise of Alexa, Google Assistant and Siri to drive US smart assistant-compatible device base to 1.6 billion in 2022* (no date). Available at: www.canalys.com/newsroom/rise-of-alexa-google-assistant-and-siri-to-drive-us-smart-assistant-compatible-device-base-to-16-billion-in-2022 (Accessed: 14 January 2022).

Dourish, P. (2001). *Where the Action Is: The Foundations of Embodied Interaction*. New Ed edition. Cambridge, MA, USA: The MIT Press.
Dove, G. *et al.* (2017). "UX Design Innovation: Challenges for Working with Machine Learning As a Design Material". In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM (CHI '17), pp. 278-288. doi:10.1145/3025453.3025739.
Hassenzahl, M. *et al.* (2020). "Otherware: how to best interact with autonomous systems". *Interactions*, 28(1), pp. 54-57. doi:10.1145/3436942.
Hassenzahl, M., Burmester, M. and Koller, F. (2003). "AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität". In Szwillus, G. and Ziegler, J. (eds.). *Mensch & Computer 2003: Interaktion in Bewegung*. Wiesbaden, Germany: Vieweg+Teubner Verlag (Berichte des German Chapter of the ACM), pp. 187-196. doi:10.1007/978-3-322-80058-9_19.
Hassenzahl, M., Burmester, M. and Koller, F. (2021). "User Experience Is All There Is: Twenty Years of Designing Positive Experiences and Meaningful Technology". *i-com*, 20, pp. 197-213. doi:10.1515/icom-2021-0034.
Hassenzahl, M. and Tractinsky, N. (2006). "User experience – a research agenda". *Behaviour & Information Technology*, 25(2), pp. 91-97. doi:10.1080/01449290500330331.
Holmquist, L.E. (2017). "Intelligence on tap: artificial intelligence as a new design material". *Interactions*, 24(4), pp. 28-33. doi:10.1145/3085571.
Jiménez, C. *et al.* (2012). "Evaluating a Methodology to Establish Usability Heuristics". In *2012 31st International Conference of the Chilean Computer Science Society. 2012 31st International Conference of the Chilean Computer Science Society*, pp. 51-59. doi:10.1109/SCCC.2012.14.
Jordan, P.W. (2000). *Designing Pleasurable Products: An Introduction to the New Human Factors*. CRC Press. doi:10.1201/9780203305683.
Kaptelinin, V. and Nardi, B.A. (2009). *Acting with Technology: Activity Theory and Interaction Design*. Cambridge, MA, USA: The MIT Press.
Karapanos, E. *et al.* (2009). "User experience over time: an initial framework". In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Boston, MA, USA: Association for Computing Machinery (CHI '09), pp. 729-738. doi:10.1145/1518701.1518814.
Kocaballi, A.B., Laranjo, L. and Coiera, E. (2018). *Measuring User Experience in Conversational Interfaces: A Comparison of Six Questionnaires*. doi:10.14236/ewic/HCI2018.21.
Kujala, S. *et al.* (2011). "UX Curve: A method for evaluating long-term user experience". *Interacting with Computers*, 23(5), pp. 473-483. doi:10.1016/j.intcom.2011.06.005.

Lachner, F. *et al.* (2016). “Quantified UX: Towards a Common Organizational Understanding of User Experience”. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*. New York, NY, USA: Association for Computing Machinery (NordiCHI ’16), pp. 1-10. doi:10.1145/2971485.2971501.

Levinson, P. (1977). “Toy, Mirror, and Art: The Metamorphosis of Technological Culture”. *ETC: A Review of General Semantics*, 34(2), pp. 151-167.

Lugmayr, A. and Bender, S. (2016). “Free UX Testing Tool: The LudoVico UX Machine for Physiological Sensor Data Recording, Analysis, and Visualization for User Experience Design Experiments”. In *Proceedings of the SEACHI 2016 on Smart Cities for Better Living with HCI and UX*. San Jose, CA, USA: Association for Computing Machinery (SEACHI 2016), pp. 36-41. doi:10.1145/2898365.2899801.

Maguire, M. (2019). “Development of a Heuristic Evaluation Tool for Voice User Interfaces”. In Marcus, A. and Wang, W. (eds.). *Design, User Experience, and Usability. Practice and Case Studies*. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 212-225. doi:10.1007/978-3-030-23535-2_16.

Minge, M. *et al.* (2017). “The meCUE Questionnaire: A Modular Tool for Measuring User Experience”. In Soares, M., Falcão, C., and Ahram, T.Z. (eds.). *Advances in Ergonomics Modeling, Usability & Special Populations*. Cham: Springer International Publishing (Advances in Intelligent Systems and Computing), pp. 115-128. doi:10.1007/978-3-319-41685-4_11.

Nielsen, J. and Molich, R. (1990). “Heuristic evaluation of user interfaces”. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Seattle, Washington, USA: Association for Computing Machinery (CHI ’90), pp. 249-256. doi:10.1145/97243.97281.

Norman, D.A. (2004). *Emotional Design: Why We Love (or Hate) Everyday Things*. New York, NY, USA: Basic Books.

Obrist, M., Roto, V. and Väänänen-Vainio-Mattila, K. (2009). “User experience evaluation: do you know which method to use?”. In *CHI ’09 Extended Abstracts on Human Factors in Computing Systems*. Boston, MA, USA: Association for Computing Machinery (CHI EA ’09), pp. 2763-2766. doi:10.1145/1520340.1520401.

Otey, D.Q. (2017). “A methodology to develop usability / user experience heuristics”. In *Proceedings of the XVIII International Conference on Human Computer Interaction – Interacción ’17. the XVIII International Conference*, Cancun, Mexico: ACM Press, pp. 1-2. doi:10.1145/3123818.3133832.

Pettersson, I. *et al.* (2018). "A Bermuda Triangle? A Review of Method Application and Triangulation in User Experience Evaluation". In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Montreal QC, Canada: Association for Computing Machinery (CHI '18), pp. 1-16. doi:10.1145/3173574.3174035.

Pyae, A. and Joelsson, T.N. (2018). "Investigating the usability and user experiences of voice user interface: a case of Google home smart speaker". In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*. Barcelona, Spain: Association for Computing Machinery (MobileHCI '18), pp. 127-131. doi:10.1145/3236112.3236130.

Rivero, L. and Conte, T. (2017). "A Systematic Mapping Study on Research Contributions on UX Evaluation Technologies". In *Proceedings of the XVI Brazilian Symposium on Human Factors in Computing Systems*. Joinville, Brazil: Association for Computing Machinery (IHC 2017), pp. 1-10. doi:10.1145/3160504.3160512.

Schmettow, M., Noordzij, M.L. and Mundt, M. (2013). "An implicit test of UX: individuals differ in what they associate with computers". In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*. Paris, France: Association for Computing Machinery (CHI EA '13), pp. 2039-2048. doi:10.1145/2468356.2468722.

Schütte, S. *et al.* (2006). "Developing Software Tools for Kansei Engineering Processes: Kansei Engineering Software (KESo) and a Design Support System Based on Genetic Algorithm". In *9th International Quality Management for Organizational Development (QMOD) Conference*, August 9-11, Liverpool, England. 9th International Quality Management for Organizational Development (QMOD) Conference, August 9-11, Liverpool, England. Available at: http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-15720 (Accessed: 7 February 2020).

Sciuto, A. *et al.* (2018). "'Hey Alexa, What's Up?': A Mixed-Methods Studies of In-Home Conversational Agent Usage". In *Proceedings of the 2018 Designing Interactive Systems Conference*. New York, NY, USA: ACM (DIS '18), pp. 857-868. doi:10.1145/3196709.3196772.

Sivaji, A., Nielsen, S.F. and Clemmensen, T. (2017), "A Textual Feedback Tool for Empowering Participants in Usability and UX Evaluations". *International Journal of Human-Computer Interaction*, 33(5), pp. 357-370. doi:10.1080/10447318.2016.1243928.

Spallazzo, D., Sciannamè, M. and Ceconello, M. (2019). 'The Domestic Shape of AI: A Reflection on Virtual Assistants". In *DeSForM 2019 Beyond Intelligence*, pp. 52-59. Available at: https://re.public.polimi.it/handle/11311/1121253#.XiheQBNKh0s (Accessed: 22 January 2020).

Tractinsky, N., Katz, A.S. and Ikar, D. (2000). "What is beautiful is usable". *Interacting with Computers*, 13(2), pp. 127-145. doi:10.1016/S0953-5438(00)00031-X.

Vermeeren, A.P.O.S. *et al*. (2010). "User experience evaluation methods: current state and development needs". In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*. Reykjavik, Iceland: Association for Computing Machinery (NordiCHI '10), pp. 521-530. doi:10.1145/1868914.1868973.

Vitali, I. and Arquilla, V. (2019). "Conversational smart products: a research opportunity, first investigation and definition". In DeSForM. *Beyond Intelligence*, pp. 60-68. Available at: https://re.public.polimi.it/handle/11311/1115471#.XhdPpBdKh0s (Accessed: 9 January 2020).

White, R.W. (2018). 'Skill Discovery in Virtual Assistants". *Commun. ACM*, 61(11), pp. 106-113. doi:10.1145/3185336.

Zhou, Z. *et al*. (2019). "ML-Process Canvas: A Design Tool to Support the UX Design of Machine Learning-Empowered Products". In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. Glasgow, Scotland Uk: Association for Computing Machinery (CHI EA '19), pp. 1-6. doi:10.1145/3290607.3312859.

Authors

Marco Ajovalasit. Associate Professor in Human-Centered Design in the Design Department at the Politecnico di Milano, Italy, where he teaches modules in User Experience Design, and Digital Interaction Design. His research expertise is in the field of design for meaning and human-centred design methods leading to products, systems and services which are physically, perceptually, cognitively and emotionally intuitive to their users. Research focuses on the "design for meaning" approach defining human-centered design methods for designing products, systems and services based on new meanings for the consumers, including the use of data, designing ethnography, real fiction and co-creation. A particular emphasis of the research is on the development of a toolkit for designers for organizing the consideration of the intended meaning of the designed artefacts. Focus is placed on the thinking process, dialogue and use of semantics consumers typically associate to the understanding of the designed artefact. He has been the Principal Investigator (PI) (2013-2016) of a £3.2M FP7 EU Grant funded project "Light.Touch.Matters: Design-driven development of touch sensitive luminous flexible plastics for applications in care & well-being", which put together a multi-disciplinary team of designers and material scientists and industry leaders involving four Universities and thirteen SMEs designers from 9 different countries. Within the EU project he has developed design guidelines for meaningful interactions for the consideration of the effects of product properties on the perceptual experiences which occur at a product interface.

Venanzio Arquilla. Designer, Co-Founder of the StartUp Narvalo, Spin-Off of the Politecnico di Milano, is an Associate Professor at the Design Department, Politecnico di Milano. He is the Coordinator of the BSc in Product Design and the MSc in Integrated Product Design at the POLIMI

Design School. His research activities deal with user experience, strategic and service design, smart and connected products. He is the founder and coordinator of the Experience Design Academy that comprise the UX Design and the APP Design and Development higher education courses at POLI.design. He is the co-director of the Master of User Experience Psychology held by Università Cattolica del Sacro Cuore and Politecnico di Milano/POLI.design.

Mauro Ceconello. Architect and Associate Professor, Product Design, School of Design, Politecnico di Milan. He focuses his research activity on tools and apps to enhance cultural heritage and tourism using mobile technology, serious games, and interactive systems. He held courses about augmented reality and mobile experience. He is a scientific coordinator of many research projects concerning the valorization of culture through digital technologies and interaction tools.

Alice Paracolli. Ph.D. student in Design at Politecnico di Milano. She is specialised in User Experience, and her research focuses on Sustainable UX applied to AI-infused Objects. Ubiquitous computing elicits an endless range of internet-enabled devices, offering the potential to the user to be more diligent in energy use, except that it creates an ever-growing web of data-consuming objects that stay on forever. The Ph.D. aims to set design guidelines for sustainable, connected objects, proposing a framework to build continuously incremental sustainable innovation in technology.

Martina Sciannamè. Ph.D. student in Design at Politecnico di Milano. She investigates how AI and ML can be meaningfully included in Design education with a multidisciplinary approach. Her aim is to produce theoretical knowledge and practical tools to enable designers to exploit the opportunities offered by this technology and to communicate with ML experts, being aware of its actual capabilities and limitations. Her hope is that the intersection of Design and ML, within the frame of Ethics, will finally steer the development of ML systems towards real-world problems and responsible innovation for the flourishing of humans.

Davide Spallazzo. Ph.D. in Design and Associate Professor at the Department of Design of Politecnico di Milano. Active in Interaction Design and Human-Computer Interaction, his main research interest is the human-centred approach to digital innovation and meaning-making, applied to diverse fields of inquiry. Over the years, he coordinated and took part in several national and international research projects concerning digital technologies in the Cultural Heritage field, serious gaming, and AI. He is now coordinating the

Meet-AI research project, aimed at creating a new UX assessment method for AI-infused products. He is deputy coordinator of the Ph.D. program in Design and secretary of the MSc in Digital & Interaction Design.

Ilaria Vitali. Product designer and Ph.D. in Design, with honors, at Politecnico di Milano with a thesis titled “Conversation on conversational smart products”, graduated with a Master’s degree in Product Design for Innovation and a dual honors degree from Alta Scuola Politecnica. Her research focuses on smart connected products and devices with conversational interfaces and explores how to design them, creating guidelines and tools for didactic and professional activities. In particular, she developed the Mapping the IoT Toolkit (mappingtheiot.polimi.it), an accessible kit to aid in the design of IoT devices.

Emma Zavarrone. Ph.D. in Strategy, Management and Quantitative Methods in Business (Univ. Bicocca), she is Associate Professor in Social Statistics and Demography at Iulm University. She teaches statistics and textual analysis laboratory for social media. She is the Rector’s delegate for Third Mission and a member of the Evaluation Committee (Nucleo di Valutazione). She is the coordinator of two European projects on the issues of sustainability and dissemination of STEM among young women. She is the scientific director of Humanities Analytics Lab, structure based on textual and AI language analysis. Her main research interests are related to structural models and textual methodologies. She is the author of several international publications.

Artificial intelligence is more-or-less covertly entering our lives and houses, embedded into products and services that are acquiring novel roles and agency on users.

Products such as virtual assistants represent the first wave of materialization of artificial intelligence in the domestic realm and beyond. They are new interlocutors in an emerging redefined relationship between humans and computers. They are agents, with miscommunicated or unclear properties, performing actions to reach human-set goals.

They embed capabilities that industrial products never had. They can learn users' preferences and accordingly adapt their responses, but they are also powerful means to shape people's behavior and build new practices and habits. Nevertheless, the way these products are used is not fully exploiting their potential, and frequently they entail poor user experiences, relegating their role to gadgets or toys.

Furthermore, AI-infused products need vast amounts of personal data to work accurately, and the gathering and processing of this data are often obscure to end-users. As well, how, whether, and when it is preferable to implement AI in products and services is still an open debate. This condition raises critical ethical issues about their usage and may dramatically impact users' trust and, ultimately, the quality of user experience.

The design discipline and the Human-Computer Interaction (HCI) field are just beginning to explore the wicked relationship between Design and AI, looking for a definition of its borders, still blurred and ever-changing. The book approaches this issue from a human-centered standpoint, proposing designerly reflections on AI-infused products. It addresses one main guiding question: what are the design implications of embedding intelligence into everyday objects?

FrancoAngeli
La passione per le conoscenze