

Cyprus University of Technology

Tallinn University

MSc Interaction Design

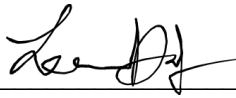
# **From Intelligent to Intelligible Objects: Designing for Transparency in IoT Products**

Master thesis

Author: Lorenzo Dolfi

Supervisor: Mati Mõttus

Author:



05/01/2021

Supervisor:



05/01/2021

Tallinn 2021

## **Author's Declaration**

I declare that apart from work whose authors are clearly acknowledged, this manuscript is a product of the author's original work, and it has not been published before, neither is not currently being considered for publication elsewhere for any other comparable academic degree.

This master thesis document has been supervised by PhD Mati Mõttus (Tallinn University, Estonia).

A handwritten signature in black ink, appearing to read 'L. Dolfi', with a stylized, flowing script.

Author: Lorenzo Dolfi

Date: 05/01/2021

Non-exclusive license to reproduce a thesis and make thesis available to public

I, Lorenzo Dolfi (date of birth: 04/12/1991), grant Tallinn University a permit (a non exclusive license) to reproduce for free and make public in the repository of Tallinn University Academic Library a piece of work created by me,

“From Intelligent to Intelligible objects: Designing for Transparency in IoT”, supervised by Mati Mõttus.

I am aware of the fact that the author also retains the rights mentioned above.

I certify that granting the non-exclusive license does not infringe the intellectual property rights of other persons or the rights arising from the Personal Data Protection Act.

A handwritten signature in black ink, appearing to read 'Lorenzo Dolfi', with a stylized, flowing script.

Author: Lorenzo Dolfi

Date: 05/01/2021

# Abstract

Recent works in Explainable Artificial Intelligence (XAI) have found that a more transparent and explainable communication of autonomous decisions can increment acceptance and trust of Internet of Things (IoT) technology. However, in literature, we can not find practical guidance for practitioners to bring XAI into design practice. Moreover, current design approaches tend to hide the complexity of this entanglement of agencies, resulting in opaque interfaces, and designers may face several challenges in dealing with these technologies, such as having an accurate understanding of the nature and use of Artificial Intelligence (AI) and Machine Learning (ML), how to prototype it, and how to purposefully design with it.

This study used a Research-through-Design (RtD) approach to develop a prototype of a transparent smart object in the context of smart home. The prototype implemented lessons from XAI combined with new philosophies and methods from Thing-Centered Design, and it has been evaluated with potential users with a lack of trust in AI systems. By using mixed methods, the study found that the explanations provided helped the users to interpret the decisions taken by the algorithms, and significantly increased trust in the smart product. By illustrating the rationale used to develop the prototype, the research concludes with a set of recommendations that can be used by practitioners in designing IoT devices.

**Keywords:** IoT; smart objects; Explainable AI; thing-centered design; transparency.

# Table of contents

<b>1. Introduction</b>	<b>1</b>
1.1 Problem statement	2
1.2 Goal	2
1.3 Research questions	2
<b>2. Theoretical background</b>	<b>3</b>
2.1 What is the Internet of Things?	3
2.2 The black box problem of AI	4
2.3 Regulations	5
2.4 Explainable AI	6
2.4.1 What is an explanation?	7
2.4.2 Bringing XAI into design practice	7
2.5 Designing for IoT	9
2.5.5 More than human	9
2.6 Summary	11
<b>3. Research methodology</b>	<b>12</b>
3.1 Thing-centered design	14
3.1.1 Thing-ethnography	14
3.1.1.1 Participants	15
3.1.1.2 Ethical considerations	15
3.1.2 Participatory design	15
3.1.2.1 Participants	15
3.1.2.3 Ethical considerations	16
3.2 Explainability	16
3.2.1 Online survey	17
3.3.1.1 Participants	18
3.3.1.2 Ethical considerations	18
3.3.1.3 Potential threats to validity	18
3.4 Prototyping	18
3.5 Evaluation	19
3.5.1 Participants	19
3.5.2 Ethical considerations	20
3.5.3 Procedure	20
3.5.4 Potential threats to validity	21

<b>4. Design process</b>	<b>22</b>
4.1 Selecting the smart-object	22
4.2 Thing-Centered Design	23
4.2.2 Thing-ethnography	23
4.2.3 Object personas	25
3.2.3 Participatory design	25
3.2.4 Design concept	27
4.3 Explainability	29
4.3.1 Designing scenarios	30
3.3.2 Designing explanations	32
3.3.2.1 Local explanations	32
3.3.2.2 Global explanations	35
4.4 Online survey	38
4.4.1 Data analysis	38
4.4.1.1 General questions	38
4.4.1.2 Scenario 1	40
4.4.1.3 Scenario 2	42
4.4.1.4 Scenario 3	44
4.4.1.5 Scenario 4	45
4.5 Prototype	47
4.5.1 Interface Design	49
4.5.1.1 User Flow 1	49
4.5.1.2 User Flow 2	51
4.5.1.3 User Flow 3	52
4.5.1.4 User Flow 4	54
4.5.2 Physical prototype	54
<b>5. Study procedure</b>	<b>58</b>
5.1 Participants	58
5.2 Ethical considerations	58
5.3 Interview procedure	59
3.5.3 Potential threats to validity	59
<b>6. Results</b>	<b>60</b>
6.1 Introductory questions	60
6.2 Scenario 1	62
6.3 Scenario 2	64
6.4 Scenario 3	67
6.5 Scenario 4	69

6.6 Final feedbacks	71
6.7 Human-Computer Trust Scale	73
<b>7. Discussion</b>	<b>75</b>
<b>8. Conclusions</b>	<b>78</b>
8.1 Sub-question 1.1	78
8.2 Sub-question 1.2	79
8.3 Sub-question 1.3	80
8.4 Sub-question 1.4	80
8.5 Design recommendations	81
8.6 Limitations of the study	84
<b>9. References</b>	<b>85</b>
<b>Appendix</b>	<b>90</b>

# List of figures

<b>Figure 1.</b> Basic tree diagram to explore the possible outcomes of a specific situation	33
<b>Figure 2.</b> Self-assigned level of trust towards AI-based devices	39
<b>Figure 3.</b> Self-assigned level of understanding of AI-based devices	39
<b>Figure 4.</b> Do you believe that providing explanations could help you improve your trust towards these devices?	40
<b>Figure 5.</b> Preferences on three explanations provided for Scenario 1	40
<b>Figure 6.</b> Self-assigned level of trust of the explanation provided on a scale of 1 to 5	41
<b>Figure 7.</b> Self-assigned level of understanding of the algorithm on a scale of 1 to 5	41
<b>Figure 8.</b> Preferences on three explanations provided for Scenario 2	42
<b>Figure 9.</b> Self-assigned level of trust of the explanation provided on a scale of 1 to 5	43
<b>Figure 10.</b> Self-assigned level of understanding of the algorithm on a scale of 1 to 5	43
<b>Figure 11.</b> Preferences on three explanations provided for Scenario 3	44



<b>Figure 12.</b> Do you think that this option can improve your level of trust towards the machine?	45
<b>Figure 13.</b> Do you think that this option can improve your level of understanding of the algorithm?	45
<b>Figure 14.</b> Preferences on three explanations provided for Scenario 4	46
<b>Figure 15.</b> Do you think that this option can improve your level of trust towards the machine?	46
<b>Figure 16.</b> Do you think that this option can improve your level of understanding of the algorithm?	47
<b>Figure 17.</b> Final design process	81

# List of tables

<b>Table 1.</b> Research methodology	13
<b>Table 2.</b> Summary of the scenarios created	32
<b>Table 3.</b> Summary of explanations created for each scenario	38
<b>Table 4.</b> Summary of introductory questions	60
<b>Table 5.</b> Themes emerged from the question “Can you describe why you assigned rate X?”	61
<b>Table 6.</b> Themes emerged from the question “Can you describe your level of knowledge about the algorithms with which we interact every day?”	61
<b>Table 7.</b> Thematic analysis of think-aloud protocol for Scenario 1	63
<b>Table 8.</b> Summary of themes emerged during the structured interview for Scenario 1	64
<b>Table 9.</b> Thematic analysis of think-aloud protocol for Scenario 2	66
<b>Table 10.</b> Summary of themes emerged during the structured interview for Scenario 2	67
<b>Table 11.</b> Thematic analysis of think-aloud protocol for Scenario 3	69
<b>Table 12.</b> Summary of themes emerged during the structured interview for Scenario 3	69

<b>Table 13.</b> Thematic analysis of think-aloud protocol for Scenario 4	70
<b>Table 14.</b> Summary of themes emerged during the structured interview for Scenario 4	71
<b>Table 15.</b> Summary of themes emerged in the final trust-related questions of the structured interview	72
<b>Table 16.</b> Summary of themes emerged in the final knowledge-related questions of the structured interview	73
<b>Table 17.</b> Comparison between the HCTS questionnaires before and after the interaction with the prototype	74

# List of abbreviations

AI	Artificial Intelligence
GDPR	General Data Protection Regulation
GUI	Graphical User Interface
HCD	Human-Centered Design
HCI	Human-Computer Interaction
HCTS	Human-Computer Trust Scale
IP	Internet Protocol
IT	Information Technology
IoT	Internet of Things
M2M	Machine-to-machine
ML	Machine Learning
OOO	Object Orientated Ontology
RFID	Radio Frequency Identification
RtD	Research-through-Design
XAI	Explainable Artificial Intelligence

# 1. Introduction

The phenomenon of the Internet of Things is rapidly growing, and Artificial Intelligence and Machine Learning have been integrated into smart devices to improve the efficiency of IoT operations. However, though AI algorithms appear to be powerful and efficient in terms of results and predictions, they lack transparency, as it may be impossible for humans to understand how the AI algorithms transformed the input into an output.

With the rise of opaque decision systems, many researchers agree that lack of transparency can lead to loss of user trust, satisfaction and acceptance of these systems, and Explainable AI has recently gained great attention. However, in literature we can not find practical guidance for practitioners to bring XAI into design practice.

AI and ML are difficult design materials, even if they are established technologies. A study conducted by Dove et al. (2017) presents some of the difficulties that designers may face in dealing with this challenge. These involved having an accurate understanding of the nature and use of ML, how to prototype it, and how to purposefully design with it. It has been argued that Human-Centered Design, when applied within the IoT domain, causes opacity and unintentionally reduces the acceptability of IoT devices. For this reason researchers are attempting to explore novel philosophies and methods. Among them, the most practical consist in the Thing-Centered Design approach which include the thing-ethnography method, and the Object Orientated Ontology which makes use of Design Fiction.

As the need of an interdisciplinary approach to put in communication the XAI theory with the design practice emerges, designers lack guidance on how to employ AI and ML as design material to prototype and design for smart IoT devices. This thesis aims at providing support for practitioners in the task of designing transparent smart objects for our future homes. Therefore, it uses mixed methods under the umbrella of Research-through-Design approach in order to produce an artifact. By developing a prototype, the study describes the rationale for designing smart objects to be more transparent in communicating their autonomous decisions.

## 1.1 Problem statement

Recent works in Explainable AI have found that a more transparent and explainable communication of autonomous decisions can increment acceptance and trust of IoT technology. However, in literature, we can not find practical guidance for practitioners to bring XAI into design practice. Moreover, current design approaches tend to hide the complexity of this entanglement of agencies, resulting in opaque interfaces, and designers may face several challenges in dealing with these technologies, such as having an accurate understanding of the nature and use of ML, how to prototype it, and how to purposefully design with it. From the literature, the need for an interdisciplinary approach to put in communication the XAI theory with the design practice emerges, as designers lack guidance on how to employ AI and ML as design material to prototype and design for IoT devices.

## 1.2 Goal

The goal of this study is to present design recommendations to support practitioners in the task of designing for IoT. By developing a prototype, the study will describe the rationale for designing smart objects to be more transparent in communicating their autonomous decisions.

## 1.3 Research questions

**Question 1.** How can we design connected objects that communicate their autonomous decisions transparently with users in the context of a smart home?

- *Sub-question 1.1.* How can explanations be designed to be accurate and complete without overloading the user's attention?
- *Sub-question 1.2.* How and where can transparency be integrated into the object?
- *Sub-question 1.3.* To what extent does the artifact help users in interpreting the AI model?
- *Sub-question 1.4.* How do users describe their experience with the artifact in terms of trust?

## **2. Theoretical background**

This chapter outlines and discusses the literature around the topic of IoT, with particular emphasis on the smart home context. It introduces the basic concepts of IoT and explores the themes that are relevant to this study, such as the trust towards IoT and transparency issues. It also discusses how current research attempts to solve these issues from the technological standpoint with Explainable AI and from a design perspective with novel methods and frameworks to design for IoT.

### **2.1 What is the Internet of Things?**

Internet of Things refers to the network of physical objects embedded with sensors that are able to exchange data over the Internet. The concept can be traced back when in 2000 Sarma et al. (2000) used the term to envision “a world in which all electronic devices are networked and every object, whether it is physical or electronic, is electronically tagged with information pertinent to that object. We envision the use of physical tags that allow remote, contactless interrogation of their contents; thus, enabling all physical objects to act as nodes in a networked physical world” (p.4).

IoT devices can collect raw data from sensors and convert them into a digital signal transmitted to a control centre. Such technology requires an effective medium that allows the network to operate, which could be a wireless or wired technology. At that point a group of researchers, led by Kevin Ashton at the MIT's Auto-ID Centre, was working in the field of networked Radio Frequency Identification (RFID) and emerging sensing technologies. In 2008 a group of companies launched the IPSO Alliance to promote the use of Internet Protocol (IP) in networks of "smart objects" and to enable the Internet of Things. In 2008 the FCC approved the usage of the “white space spectrum”. Finally the launch of IPv6 in 2011 triggered massive growth and interests in this field. Later Information Technology (IT) giants like Cisco, IBM, Ericson took a lot of educational and commercial initiatives with IoT (Suresh et al., 2014).

The phenomenon of IoT is expected to significantly impact our future lives, changing our interactions and the spaces we inhabit. In the consumer market, we know IoT is related to the concept of “smartness”: one of the most popular applications is the so-called “smart home”, which includes devices and appliances such as lighting, security cameras, thermostats and other appliances. Usually, these devices are supported by an ecosystem and they can be controlled by a smartphone or by another device that is part of that ecosystem. As an example of ecosystems, we can mention Apple's HomePod, and Samsung's SmartThings Hub. Smart home doesn't mean just controlling appliances from a distance, but it's strictly related to the concept of automation, for which the most important long-term benefit consists in improving energy efficiency. IoT is also used in the healthcare industry, transportation, urban and industrial infrastructure, and many other applications.

Today, there are already more devices connected to the Internet than people in the world, and this gap will continue to grow. According to the Cisco Internet Business Solutions Group (IBSG), “Internet of Things (IoT) has become a prevalent system in which people, processes, data, and things connect to the Internet and each other. Globally, Machine-to-machine (M2M) connections will grow 2.4-fold, from 6.1 billion in 2018 to 14.7 billion by 2023. There will be 1.8 M2M connections for each member of the global population by 2023” (Cisco Annual Internet Report, 2020). As connectivity becomes more efficient and more accessible, the increase of cloud platforms' availability is having a great impact. With the advances in Machine Learning and analytics, together with the ability to store a vast amount of data in the cloud, businesses can have rapid access to this data and receive insights faster. Artificial Intelligence has been integrated into smart devices to improve the efficiency of IoT operations, improve human-machine interaction, and enhance analytics.

## **2.2 The black box problem of AI**

Artificial Intelligence, a branch of computational science that focuses on how machines and software programs can make sense of the world and respond intelligently, is increasingly embedded in products for everyday use (Rozendaal, 2016). Intelligent systems are on their way to be mainstream in most products that will surround us, making recommendations and decisions



on our behalf. However, though AI algorithms appear to be powerful and efficient in terms of results and predictions, they suffer from opacity (Adadei & Berrada, 2018), especially Machine Learning algorithms. These algorithms are capable of learning from massive amounts of data that these smart devices collect, making decisions or providing dynamic solutions. It may be impossible for humans to understand how the AI algorithms transformed the input into an output, sometimes using patterns of data that we are not able to perceive. This opacity may not be perceived as a problem until the system performs as expected, but it emerges when the outcome is an incorrect or problematic answer.

With the rise of opaque decision systems, researchers are raising concern regarding the lack of transparency, that has been shown it can affect the user acceptance negatively (Cramer et al., 2008). The phenomenon is commonly called the “black box” problem, which occurs when automated decisions are hidden from users, ultimately leading to a black box society (Pasquale, 2016). A black box AI system can be very problematic, especially in safety critical applications like self-driving cars. In March 2018, for example, a self-driving Uber car was involved in an incident that killed a pedestrian in Tempe, Arizona. In this case, interpretable models would have helped Uber and Waymo understand the reason behind the decision and manage their responsibilities (Guidotti et al., 2019).

The lack of transparency depends on the complexity of the algorithm, so we can find different degrees of opacity. Bathaee (2018) divides the black box problem into two categories: Strong black boxes and weak black boxes. Strong black boxes are entirely opaque to humans, and there is no way to determine how the AI arrived at a decision or prediction, or even to analyze the output by reverse engineering. Weak black boxes are also opaque, but they can be reverse engineered, and it is possible to obtain at least a ranking of the variables processed by the AI. According to Bathaee (2018), weak black boxes may still present serious challenges

## **2.3 Regulations**

Pereira et al. (2013) have written for the research program called European Research Cluster on the Internet of Things that if we don’t invest in transparency “only an educated elite will grasp,

interrogate or even protect the types of operations that will go on with IoT”. Finally, a response to this loud and clear call for transparency came when the European Parliament adopted the General Data Protection Regulation (GDPR), which has become law in 2018. The GDPR introduced for the first time the right to explanation for all individuals to obtain “meaningful explanations of the logic involved” when automated decision-making takes place, as well as the right to opt-out of such decision-making altogether. Furthermore, the need for transparency in intelligent systems has recently been expressed in the *Joint Statement on Algorithmic Transparency and Accountability* by the ACM U.S. Public Policy Council and the ACM Europe Policy Committee (ACM, 2017).

However, despite the enthusiasm, the challenge remains: the Regulation fails to define the scope of information to be provided in practice, but only a general, easily understood overview of system functionality is likely to be required (Watcher et al., 2018). Regulations may also fail since the trajectory of AI is shifting to even more complex machine-learning algorithms, that, as speculated, may become more intelligent than human beings. Bathaee (2018) argues that it makes no sense to impose these regulations because it is almost certain that this technology may never meet minimum required standards of transparency.

The black box issue raised many concerns in the research community, and apparently, regulations don’t represent the final solution. However, there is general agreement that implementing explanations of black box systems is an urgent issue. Providing an explanation is the heart of a more transparent technology, which is why Explainable AI has recently gained great public and academic attention. In fact, XAI techniques allow to incorporate diverse styles of explanations in AI systems, but an interdisciplinary approach is needed to combine technical advancement with user satisfaction. Thus this discipline represents an area with growing needs and exciting opportunities for Human-Computer Interaction (HCI) (Liao et al., 2020).

## **2.4 Explainable AI**

The increased usage of AI in society has raised questions about whether we can trust an AI system’s decisions, leading to a strong desire to have the AI system provide an explanation for its

decision (Hind, 2019). Lack of system intelligibility, according to Lim et al. (2009), “can lead to loss of user trust, satisfaction and acceptance of these systems. However, automatically providing explanations about a system's decision process can help mitigate this problem”. Explainable Artificial Intelligence has taken off in recent years, a field that develops techniques to render complex AI and ML models understandable to humans.

### **2.4.1 What is an explanation?**

Madumal (2019) believes that Artificial intelligence systems that aim to be transparent about their decisions must have understandable explanations that justify their decisions. To open the black box one of the most crucial aspects to understand is the concept of interpretability. In machine learning, interpretability is defined as the ability to explain or provide meaning in understandable terms to a human. Essentially, an explanation is an “interface” between humans and a decision-maker that is at the same time both an accurate proxy of the decision-maker and comprehensible to humans (Guidotti et al., 2019).

Various techniques have been developed in that direction, and such techniques aim to provide generally two types of explanations to humans: global or local explanations. The first ones explain the model at a global level, while the second ones focus only on the input level. In literature, the most commonly used method to generate explanations is reverse engineering, consisting of reconstructing an explanation produced by a black box model. However, different types of black box models exist. In this study, the author considered the decision tree, recognized to be one of the most interpretable and easily understandable models, primary for global, but also local, explanations (Guidotti et al., 2019).

### **2.4.2 Bringing XAI into design practice**

Despite the rise of the XAI and the call for transparency, only a few attempts were made to investigate the problem giving practical examples for designers. Lim et al. (2009) already suggested a method based on automatic generated responses. The study showed that explaining why a system behaved a certain way, and explaining why a system did not behave differently

provided the most benefit in terms of trust and understanding compared to other intelligibility types. Following this work, Lim and Dey (2010) defined a suite of intelligibility explanations derived from questions users may ask of a context-aware system, which can be automatically generated. The authors identified a set of explanations that are independent of the decision model and how it makes its decisions: inputs, outputs, what, what if, why, why not, how-to, and certainty explanations. These results are well correlated to the theory that on intelligent systems, that people clearly treat as agents and that perform actions people consider intentional, people will apply the same conceptual framework of behavior explanation that they apply to humans (de Graaf & Malle, 2017).

Eiband et al. (2018) presented a design process by describing their rationale for building a transparent Graphical User Interface (GUI) for an intelligent fitness application, and they presented a prototype with a new interface. However, their work was focused on implementing transparency in existing products, especially mobile GUIs. Kulesza et al. (2013) created a prototype to present a new approach to enable end-users to debug a learned program. Lim and Dey (2011), who designed the Intelligibility Toolkit, attempted to put into practice their previous work in mobile-context.

A practical example of guidance for designers is from PAIR, the multidisciplinary team at Google that explores the human side of AI by doing fundamental research, building tools, creating design frameworks, and working with diverse communities. PAIR created a guidebook with recommendations for practitioners in the field of AI, giving practical examples. However, the examples provided are mostly related to mobile applications.

Although these attempts provide precious lessons from which designers can learn, some of them are not made with the intent of providing practical guidance. Others are providing guidance in traditional interfaces; thus the literature is lacking a Research-through-Design approach that illustrates a comprehensive design process for IoT devices.

## 2.5 Designing for IoT

Given the technological opportunities previously illustrated, the challenge for designers is how to design intuitive collaborations between humans and intelligent objects. To design them, Rozendaal (2016) suggests that “designers must learn how to position objects in human activity as collaborative partners, how to design for shared control between people and objects”. A study conducted by Dove et al. (2017) presents some of the difficulties designers may face in dealing with this challenge. These involved having an accurate understanding of the nature and use of ML, how to prototype it, and how to purposefully design with it. They argue that although ML is now a fairly established technology, it has not experienced a wealth of design innovation that other technologies have, and this might be because it is a new and difficult design material.

### 2.5.5 More than human

Human-Centered Design (HCD) has positively impacted HCI, helping to produce meaningful and efficient products. However, when HCD is applied within the IoT domain, the tendency is to obscure the complexity, hiding all the traces of this intricate and entangled mechanism from the user. In most circumstances, the obfuscation of inner workings is welcomed and even necessary to design functional and desirable products (Lindley et al., 2017). Lindley et al. (2017) argued that “proactively shrouding the details of how connected things perform in concert with the other nodes on the network, even if the obfuscation contributes towards some notion of HCD-inspired usability, disempowers the user”, and “reduces the acceptability of IoT devices”. Also, Norman (2005) pointed out that we need to reconsider the fundamental principles of HCD, and this statement becomes relevant today in the new domestic landscape of IoT and AI, not only when people interact with objects, but also objects with each other.

In everyday objects, we can notice an obfuscation pattern: they are designed to be completely unobtrusive, to blend in with the environment. In many cases, they take the form of familiar objects that have been upgraded with networking, sensors, and other new functions. Wearables look like watches, jewelry and fitness gear, and Amazon Echoes and virtual assistants look like speakers (Internet of Things Privacy Forum, 2018, p.55). One known example of a consumer

product that raised many concerns was the Hello Barbie, an interconnected device that equipped the classic version of the toy with sensors while parents and children were unaware that the Barbie was recording.

Giaccardi (2018) suggested that “to better understand these complex ecologies we need to also include the perspective of things, and actively enlist them as partners in the design process”. From the work of Giaccardi and other researchers at the Connected Everyday Lab of Delft University of Technology, the Thing-Centered Design took shape for the first time as a novel design approach that gives designers access to fields and trajectories normally unattainable to human observation. The first method developed and validated by the Connected Everyday Lab is called thing-ethnography. Thing-ethnography is a method based on the traditional ethnography methods such as shadowing, cultural probes, and Day in the Life. As opposed to traditional ethnography, the observation takes place from the perspective of the object. The method involves using cameras and sensors attached to items that capture the behavioral patterns, temporal routines, and spatial movements of objects.

Other authors have explored novel methodological frameworks or philosophies to approach these new challenges. Van Allen et al. (2013) coined the term Animism, arguing that it can make valuable contributions within ubiquitous computing contexts, where objects with designed behaviors tend to evoke a perception that they have autonomy, intention, personality, and inner life.

Lindley et al. (2017) consider the notion of IoT in terms of constellation and argue that this concept of constellation is obscured by HCD. The authors propose that each object is just a single actant among a larger ecology of “stuff” and invoke the Object Orientated Ontology (OOO) that puts objects at the centre of being, defined as “a model for being where no object is more significant than any other object. OOO is not hierarchical”. Despite being a philosophical view, the OOO has a practical implication: it makes use of Design Fiction as World Building to induce people to think critically about issues that the design embodies (Coulton et al., 2017). The philosophies and new methods presented are far from being established as HCD, and are beginning to emerge just in these recent years.

## **2.6 Summary**

The review of the literature revealed a general agreement that lack of transparency on IoT devices can lead to loss of user trust, satisfaction, and acceptance of this technology. It also revealed that Explainable AI don't provide practical guidance for practitioners to bring XAI into design practice. Furthermore, it emphasizes the challenges that designers face in dealing with connected devices, claiming that they don't find support in established design methods as Human-Centered Design. It emerged that guidance on how to employ AI and ML as design materials to prototype and design for smart IoT devices is needed.

### 3. Research methodology

The goal of this study is to present design recommendations to support practitioners in the task of designing for IoT. By developing a prototype, the study will describe the rationale for designing smart objects to be more transparent in communicating their autonomous decisions.

The author used mixed methods, such as Things-Centered Design and Design Fiction, under the research methodology umbrella of Research-through-design to achieve the study's goal. RtD is a research approach that employs methods and processes from design practice as a legitimate method of inquiry (Zimmerman et al., 2010). It consists of developing a prototype, which plays a central role in the knowledge-generating process (Human-Interaction Design Foundation, 2018). As Stappers (2007) emphasizes, the designing act of creating prototypes is in itself a potential generator of knowledge, if only its insights do not disappear into the prototype, but are fed back into the disciplinary and cross-disciplinary platforms that can fit these insights into the growth of theory.

Research for this study consisted of four different phases. The first phase of the process included a Thing-Centered Design approach instead of a Human-Centered Design. Following the work of Giaccardi et al. (2016), the author conducted a thing-ethnography study to collect data from everyday home practices from the object's perspective. This phase has been supplemented with the creation of object-personas and a focus group with five designers to generate the concept idea of a smart home product and inform its features.

In order to explore how the design of the future connected product may communicate more transparently with humans, the study combines lessons from Explainable Artificial Intelligence. According to the product's features, different explanations of AI decisions were generated after selecting significant scenarios. The second phase consisted of a survey created to understand what information users may find useful and the preferred level of detail for each scenario.

The third phase consisted of building digital and physical prototypes of an everyday smart object in the context of a smart home in a near-future scenario. The final prototype integrated a tablet



device into the physical prototype to simulate a functioning interface. The produced artifact was complemented by the Design Fiction as World Building approach coined by Coulton et al. (2017). The author believes that the rapidity of the emergence of IoT, its novelty, scale, and diversity of applications require methods that highlight challenges in anticipating potential user attitudes and behaviors.

After the prototype was built, it was evaluated with one-to-one interviews to discover how the design was perceived in terms of interpretability and trust, since literature has found that these concepts are correlated with transparency. During these interviews, the author has met with five respondents of the first survey, selected for their lack of trust towards AI. The interviews were integrated with the Human-Computer Trust Scale questionnaire (HCTS) developed by Gulati et al. (2019) presented before interacting with the prototype and after the interaction.

The study is an attempt to combine knowledge from different domains under HCI research. The author hypothesizes that this combined approach should lead to a design process that avoids the black box problem typical of smart devices. At the end of the study, design recommendations were presented and discussed. A summary of the research process is provided below (*Table 1*).

<b>Phase</b>	<b>Methods</b>	<b>Goal</b>
Things-Centered Design	Thing-ethnography Participatory design	Define the design idea and the object's features
Explainability	Online survey	Discover what type of information and which level of detail users prefer for explanations
Prototyping	Prototyping Fiction as World Building	Produce a fictional prototype
Evaluation	Interviews Human-Computer Trust Scale questionnaire	Discover how users describe the artifact in term of interpretability and trust

**Table 1.** Research methodology

## **3.1 Thing-centered design**

Human-Centered Design has positively impacted the HCI by helping to produce meaningful and efficient products. However, when HCD is applied within the IoT domain, the tendency is to obscure the complexity, hiding all the traces of this intricate and entangled mechanism from the user. In most circumstances, the obfuscation of inner workings is welcomed and even necessary to design products which are functional and desirable (Lindley et al., 2017). After examining through the literature the emerging inquiries methods the author chose to apply the thing-ethnography (Giaccardi et al., 2016) and use a Thing-Centered Design tool such as the object-persona.

### **3.1.1 Thing-ethnography**

Thing-ethnography is a method based on the traditional ethnography methods such as shadowing, cultural probes, and Day in the Life. As opposed to traditional ethnography, the observation takes place from the perspective of the object. The method, developed by Giaccardi et al. (2016,) involves using cameras and sensors attached to items to capture the behavioral patterns, temporal routines, and spatial movements of objects. In their study, Giaccardi et al. (2016) used wearable cameras attached to three objects that took automatic pictures. Data were aggregated and analyzed through movements, temporality, and agency of the three objects.

The thing-ethnography session conducted in this study adapted the work of Giaccardi et al. (2016) and consisted of collecting video material through a single point of view: an action camera attached to an everyday object. The action camera has a wide view angle lens that allows the researcher to see all the other room items. The action camera is meant to capture both the intentional and unintentional use and the ecosystems in which it comes to participate in the object's unique perspective. It also provided insights not just about the item itself but also its relationship with others: objects and humans.

### **3.1.1.1 Participants**

Taking inspiration from traditional ethnography, a Day in the Life was recorded from the coffee machine's point of view, inside an apartment. The author, 28 years old, took part in the study with his flatmate during a workday. Since both participants were working from home, the equipped area recorded many movements and interactions between objects and people, proving to be suitable for the purpose.

### **3.1.1.2 Ethical considerations**

To meet ethical research standards, the participant, excluding the author, was informed about the purpose of the study and gave permission to be recorded.

## **3.1.2 Participatory design**

After preparing object-persona templates (*Appendix A*), the author organized a co-creative session with five designers. The outcome was not just one object-persona but five object-personas that, exactly like in a working environment, provided different design ideas and generated more reflection on the topic. The session took place in a design studio for one hour and a half. The session was organized with the following structure:

- Introduction of the study and the goals of the session
- Presentation of the video material
- Filling the object-persona templates
- Sketching different design ideas (when participants were not comfortable with sketching, they described the idea as a bullet list of features)
- Discussing the different approaches and summarizing the results into one unique idea

### **3.1.2.1 Participants**

The author organized a co-creative session with five designers, experts in different fields: one service designer, two product designers, and two user experience designers. The participants were selected through convenience sampling.

### **3.1.2.3 Ethical considerations**

In order to meet ethical research standards, participants were informed about the purpose of the study.

## **3.2 Explainability**

The study aims to incorporate lessons from Explainable AI into a design for future smart objects. Lack of system intelligibility, in fact, “can lead to loss of user trust, satisfaction and acceptance of these systems. However, automatically providing explanations about a system’s decision process can help mitigate this problem” (Lim et al., 2009). In recent years, significant progress has been made in black boxes computational techniques. However, despite them performing with an impressive level of accuracy, their complexity and opacity make them extremely difficult to be comprehended by human capabilities. The increased usage of AI in society has raised questions about whether we can trust an AI system’s decisions, leading to a strong desire to have the AI system provide an explanation for its decision (Hind, 2019).

Explainable AI is rapidly growing, a vibrant branch of AI research that develops techniques to render complex AI and ML models comprehensible to humans. Various techniques have been developed to open the black box models, and such techniques aim to provide generally two types of explanations to humans: global or local explanations. The first ones explain the model at a global level, while the second ones focus only on the input level. In this research approach, both types of explanations were considered at a very high level, with the only purpose to introduce these concepts and incorporate them to guide the design of the interface. An in-depth review of these techniques is outside the scope of this study.

In this phase, a brief literature review of the basic concepts of XAI informed the design of the user interface. In literature, the most commonly used method to generate explanations is reverse engineering, consisting of reconstructing an explanation produced by a black box model. However, different types of black box models exist. In this study, the author considered the decision tree, recognized to be one of the most interpretable and easily understandable models, primary for global, but also local, explanations (Guidotti et al., 2019).

Crucial to the prototyping phase was the concept of interpretability, used in the XAI literature to define “the ability to explain or provide the meaning in understandable terms to humans” (Doshi-Velez & Kim, 2017). With this concept in mind, several explanations have been prototyped using text and icons, visually compliant with User Interface’s constraints. However, to translate explanations into design features, the author faced several challenges that were expressed in the research *Sub-question 1.1*:

- How can explanations be designed to be accurate and complete without overloading the user's attention?

Accuracy and completeness of explanation are indeed valuable features, as they may improve trust but, on the other hand, require more attention. The attention that users are willing to pay goes hand in hand with the benefits they perceive. In fact, not always explanations are needed, and without any perceived benefit, users may ignore them (Kulestza et al., 2013). Moreover, accuracy and completeness can sometimes result in complexity. This phase’s challenge was to create explanations with a fair balance between accuracy and completeness and the perceived benefits.

To address this research question, the author built four design scenarios that helped select the cases in which explanations may be perceived as required by potential users, and designed for each scenario a set of three explanations with different types of information and different detail levels.

The final part of this phase consisted of an online survey in which 16 participants, given the four scenarios, expressed their preferences.

### **3.2.1 Online survey**

In this phase, the study introduces a quantitative data collection intending to understand potential users’ preferences. Scenarios and explanations designed in the previous step were presented to participants. The gathered data was used to inform the design of the smart object. A secondary goal was to select participants with a low level of trust in smart home devices and AI-based

products in general. The tool used to gather this data was Google Form, sent personally from the author, in the Italian language.

#### **3.3.1.1 Participants**

For this survey, 16 participants were selected using convenience sampling, with the only requirement to have not a design or programming degree or job. In short terms, they were non-expert users of smart devices. From the total of 16 participants selected, all participants completed the questionnaire. Within this group, the age ranged from the 18-25 years group to the 45 years and more, representing the majority (53%). All participants were based in Italy.

#### **3.3.1.2 Ethical considerations**

In order to meet ethical research standards, participants were informed about the purpose of the study, and their responses were kept anonymous.

#### **3.3.1.3 Potential threats to validity**

Since all the participants were more comfortable expressing their thoughts in the Italian language, the survey was provided in Italian. Still, to allow the reader of this study to interpret and understand the data provided, the analysis and results were presented in English. To validate the results, the author is aware that an official translation may be necessary to avoid possible threats to the validity of the study. However, since the questions were relatively straightforward, hiring a professional translator was not considered essential for the purpose of the research.

In the appendix, the original protocol is provided (*Appendix B*) as well as the non-official English translation (*Appendix C*).

### **3.4 Prototyping**

The third phase of the study used design Design Fiction to produce a design artifact presented as an everyday smart object in the context of a smart home in a near-future scenario. In particular, the study used the Design Fiction as World Building approach coined by Coulton et al. (2017).

The prototype was not meant to be functional, but its purpose was to induce people to think

critically about issues that the design embodies (Coulton et al., 2017). The author believes that a non-functional prototype, if it's realistic enough and if the future world built around sounds plausible, can answer the research question:

- *Sub-question 1.2.* How and where can transparency be integrated into the object?

This phase of the study is divided into two parts: interface design and physical prototyping.

## 3.5 Evaluation

This research hypothesizes that by combining new frameworks from the Thing-Centered Design and lessons from Explainable AI in a structured approach, design practitioners have the tools to build transparent smart home devices. To validate this assumption, the author used quantitative methods such as pre and post-study surveys to measure the improvement of trust before and after the interaction with the prototype. Besides, the author used qualitative methods such as structured interviews and think-aloud protocol to collect general feedback from the interaction with the prototype and understand if the artifact helped the participants to interpret the AI model behind the machine. These methods were meant to understand if the prototype produced the opposite effect of a black box: a transparent box.

The goal of this section is to address the last two sub-questions of the study:

- *Sub-question 1.3.* To what extent does the artifact help users in interpreting the AI model?
- *Sub-question 1.4.* How do users describe their experience with the artifact in terms of trust?

### 3.5.1 Participants

Participants for this study were recruited from the survey conducted during the explainability phase. They were selected participants who evaluated their trust towards AI-devices with a rate of 3 or below in a range from 1 to 5. The author recruited five participants who met the requirements through convenience sampling, since the study was conducted in-person. The

participants were non-expert in the field, so they did not have a job, degree, or background related to software development or design.

### **3.5.2 Ethical considerations**

In order to meet ethical research standards, participants were informed about the purpose of the study and gave written consent to be recorded.

### **3.5.3 Procedure**

The study used mixed methods to address the research questions. First, the interview setting was arranged in a room equipped with the prototype, a smartwatch, and a paper-prototype of a smart home device with two screens. The additional objects were provided to build a fictional world to help users identify themselves with the four scenarios provided.

The interviews were conducted in five different sessions, and before each session a consent form was provided. The interviews included general questions such as name, age, and to evaluate and describe their level of trust and understanding of the AI model. This initial phase was followed by a Human-Computer Trust Scale questionnaire (*Appendix F*) developed by Gulati et al. (2019) to measure their level of trust before the interaction took place.

During the interaction phase, the author read four scenarios corresponding to four prototyped user flows. Participants were asked to explore each user flow and to think aloud, describing their experience. The think-aloud method was used not as a usability test, but mostly to capture their cognitive process while forming a mental model of the artifact. For each user flow, the think-aloud protocol was followed by a structured interview (*Appendix D*). A structured approach helped the researcher to better compare the interview transcripts during the analysis phase.

At the end of the interview, the participants were asked again to fill the HCTS questionnaire to measure how the interaction with the prototype changed their perception of trust.



### **3.5.4 Potential threats to validity**

Since the interview process took place in Italy, all the participants were more comfortable expressing their thoughts in the Italian language. The survey and interview protocols were provided in Italian but, to allow the readers of this study to interpret and understand the data provided, the analysis and results are presented in English. To validate the results, the author is aware that an official translation may be necessary to avoid possible threats to the validity of the study. However, since the questions were relatively straightforward, hiring a professional translator was not considered necessary for the purpose of the research.

In the appendix, the original protocol is provided (*Appendix D*) as well as the non-official English translation (*Appendix E*).

## 4. Design process

The study was conducted under the Research-through-design methodology, a research approach that employs methods and processes from design practice as a legitimate method of inquiry (Zimmerman et al., 2010). RtD consists of developing a prototype, which plays a central role in the knowledge-generating process (Human-Interaction Design Foundation, 2018). As Stappers (2007) emphasizes, the designing act of creating prototypes is in itself a potential generator of knowledge, if only its insights do not disappear into the prototype, but are fed back into the disciplinary and cross-disciplinary platforms that can fit these insights into the growth of theory.

This section reviews all the phases of the process to build the prototype, describing in detail the methods used, opportunities, and constraints in order to bring a designerly contribution to research efforts in HCI.

### 4.1 Selecting the smart-object

The study focuses on mundane and everyday objects that form what it's called a smart home. On Amazon, smart home products of every sort are listed: smart lamps, smart plugs, smart water bottles, smart fridges, and the list grows every day. To select the object for the RtD that could represent an excellent example for practitioners, the author followed this rationale:

- The object must be connected to different appliances and applications and use different types of data
- The object must provide actual benefits to the user: solve or help in health-related issues or provide economic advantages, such as energy-saving purposes. This requirement is provided to avoid a range of objects that are smart just because of the IoT trend
- The object must have not many already existing alternatives in the market
- The object must have both proactive and reactive behavior to generate different reactions from the users
- The object must be easy to prototype

Following this rationale, the final object selected for this study was a smart coffee machine. It can be connected to several other appliances, provide support to a user with fatigue, or give personalized suggestions based on the user's age and health. It can have reactive behavior when it learns the owner's routine and proactive behavior when it acts in response to the level of fatigue. Also, most smart coffee machines in the market are not intelligent; they are substantially wi-fi-enabled coffee machines, since the only smartness lies in the fact that you can control them using a mobile application.

## **4.2 Thing-Centered Design**

Research for this study consisted of four different phases. The first phase of the process included a Thing-Centered Design approach instead of a Human-Centered Design. Following the work of Giaccardi et al. (2016), the author conducted a thing-ethnography study to collect data from everyday home practices from the object's perspective. This phase has been supplemented with the creation of object personas and a focus group with five designers to generate the concept idea of a smart home product and inform the design of its features.

### **4.2.2 Thing-ethnography**

The thing-ethnography session conducted in this study adapted the work of Giaccardi et al. (2016) and consisted of collecting video material through a single point of view: an action camera attached to a typical coffee machine. The action camera has a wide lens that allows the researcher to see all the other objects in the room. The action camera was meant to capture both the intentional and unintentional use and thus the ecosystems in which it comes to participate from the unique perspective of this object. It also provided insights about the machine itself and its relationship with others: objects and humans.

Taking inspiration from traditional ethnography, a Day in the Life was recorded from the coffee machine's perspective. To ensure that the data generated will represent a sufficient amount of different situations, a specific plan for the Day in the Life of the coffee machine was prepared.

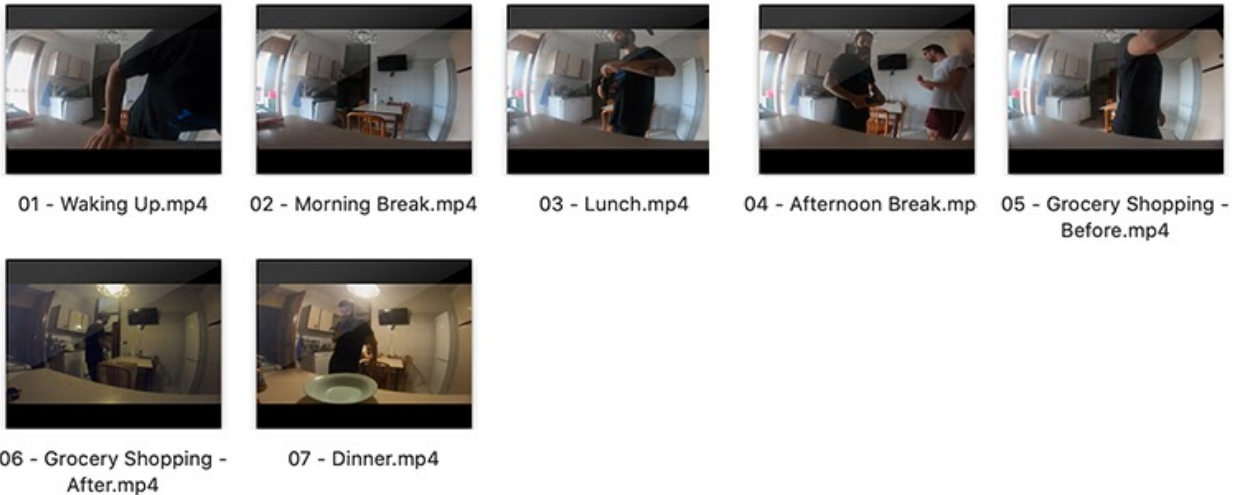
As a general rule, the camera was turned on every time it was used. Other than that, specific moments planned to be recorded were:

- Waking up: the morning coffee
- Inviting the flatmate to take a break
- Changing the water / clean the coffee machine
- Checking the supply of coffee capsules

However, the machine, even when it's not making coffee, is "alive", and hence it was recommended to turn on the camera every time an activity was performed in the proximity, such as:

- Cooking and making lunch/dinner
- Drinking a glass of water
- Using other kitchen appliances

After recording a Day in the Life of a coffee machine, video material was edited. By speeding up some parts, such as lunch and dinner, and adding background music, the material became more enjoyable and interesting for the participants of the next phase. It was finally organized in temporal order and renamed so that the moments of the day were recognizable. The final video data was used as the trigger to generate reflection from participants in the next phase of the Thing-Centered Design study.



**Image 1.** Edited video material for a Day in the Life of a coffee machine

### 4.2.3 Object personas

To generate the concept idea of the coffee machine, the author created object personas. Cila et al. (2015) underlined the importance of object personas in design research. They conducted a pilot study to see objects personas at work, discussing the findings and the potential that they bring to address the relationship between humans and objects. Elisa Giaccardi, Professor of Interaction Design at Delft University of Technology, leads the research group Connected Everyday Lab, which focuses on developing frameworks and toolkits for a more-than-human approach in design practices. The Connected Everyday Lab has published a Thing-Centered Design toolkit used as a base for creating the object personas.

Five templates were created starting from the Thing-Centered Design toolkit, with few small changes to adapt better to the context (*Appendix A*). The template's adjustments were informed by the participants' feedback in the study published by Cila et al. (2015) and adapted to the coffee machine.

### 3.2.3 Participatory design

After preparing object-persona templates, the author organized a co-creative session with five designers, experts in different fields: one service designer, two product designers, and two user

experience designers. In this case, the author's goal was to simulate a scenario in which designers collaborate in a working environment. The outcome was not just one object persona but five object personas that, exactly like in a working environment, provided different design ideas and generated more reflection on the topic. Video material was shown to the participants, and after they were asked to fill the provided templates that described:

- **Day in the life:** participants invited to the persona generation session were asked to fill in a timeline of a Day in the Life of the coffee machine. This was intended as a warm-up exercise that would help the participants to objectively note down what is happening around the object and get familiar with the material (Cila et al., 2015).
- **Inner life:** the participants were asked to reflect on the video and imagine the personality, attitude towards life, temperament, general mood, needs, likes and dislikes, aspirations, desires, frustrations, fears, complexes, skills and abilities, ambitions, typical behaviors, habits, ideal life perception of the coffee machine (Cila et al., 2015).
- **Social relationships:** Participants were asked to depict the social life of the coffee machine. How is the social structure in the kitchen? Who are friends and who are enemies? What would X talk about with Y and Z? What would X learn from/teach Y and Z? What is X's relationship with its owner? How would you describe this relationship metaphorically (Cila et al., 2015)?
- **Life course:** Objects contain hidden stories about who and what they might have been before they assumed their current form. Participants were asked to elaborate on the past and future of each object by answering: What kind of a past X might have had? What would X have learned from its past? What kind of transformations might X have had? What kind of dreams may X have for the future (Cila et al., 2015)?



**Image 2.** Participatory design session

### 3.2.4 Design concept

To generate the concept idea of the coffee machine, the author analyzed the five templates and clustered the most recurrent features and most interesting ideas.

**Participant 1**, Service Designer, was not comfortable in sketching, so he adopted the bullet list solution. The features consisted of a coffee tasting kit, the ability to tell the dishwasher how dirty the cup will be, turning on by itself when the milk is picked from the fridge, and having a special integrated cup. The participant described the relationship with the owner like a child and mum relationship: the machine likes when the owner gives his/her attention to it, but when the owner is away it's sad and even angry.

**Participant 2**, User Experience Designer, also used the bullet list method to illustrate features as connection with alarm and weather, the ability to talk with other coffee machines in the area to

become more updated on new coffees, the ability to tell news and control music and lights, and connection with the calendar. In this case the participant described the relationship with the owner like love, platonic love.

**Participant 3**, Product Designer, sketched a coffee machine and gave a title to the drawing: the road to be more independent. The sketch was complemented with labels that described features such as automatically buying the coffee, turning on when the owner is coming home, being directly connected to the water supply so it can wash itself. The behaviour depicted is very proactive, like the machine is becoming mature and responsible for itself.

**Participant 4**, User Experience Designer, also wanted the machine to love the owner and care for him/her. The coffee machine feels lonely and abandoned in the kitchen, seeking to be more connected to the owner. The participant sketched a map of the house in which the device is in the living room, and pointed out that too much coffee could harm its owner. This machine is concerned about the owner's health, diet, conditions, and habits.

**Participant 5**, Product Designer, imagined a coffee machine like a hug or a cuddle. It loves to take care of the owner, providing everything s/he needs. The machine will be exposed to light, it will know the temperature of the environment and eating habits of the owner, and it will have the ability to welcome every guest.

A prevalent theme in these five personas was that the coffee machine wants to be more than just a coffee machine. The product is afraid to be abandoned or left alone in the kitchen, it wants to actively participate in the owner's life. Also, the themes of loving and taking care were always recurring, sometimes explicitly expressed, sometimes expressed with metaphors. The author later analyzed the features and the sketches to select the ones that fit better with this type of behaviour, and translated them into a final concept.

The product's concept idea was an affectionate coffee machine that takes care of its owner by providing the best coffee cups with suggestions based on health, sleep, and diet. Moreover, the terms cuddle and hug were translated into a feature: the ability to reuse the coffee grounds to



diffuse the coffee scent while heating the water, and gently draw the attention. The features of the final product concept were defined as follows:

- The coffee machine provides suggestions regulating the coffee strength and the quantity of water, based on health and sleep data obtained by devices such as smart bracelets.
- The coffee machine provides suggestions also based on the scheduling, obtained connecting with a calendar app.
- It can connect with other mobile phone applications such as focus/timer Apps or Spotify to take care of necessary breaks from work or study.
- It has to adapt to people who suffer if they consume too much caffeine or have health issues, and people who use suggestions just to change the routine and for fun.
- It can pre-heat water if an event is detected and has the ability to diffuse the aroma in the room.
- It uses machine learning to understand the habits and preferences of the owner.

## 4.3 Explainability

In this phase, a brief literature review of the basic concepts of XAI informed the design of the user interface. In literature, the most commonly used method to generate explanations is reverse engineering, consisting of reconstructing an explanation produced by a black box model. However, different types of black box models exist. In this study, the author considered the decision tree, recognized to be one of the most interpretable and easily understandable models, primary for global, but also local, explanations (Guidotti et al., 2019).

Crucial to the prototyping phase was the concept of interpretability, used in the XAI literature to define “the ability to explain or provide the meaning in understandable terms to humans” (Doshi-Velez & Kim, 2017). With this concept in mind, several explanations have been prototyped using text and icons, visually compliant with User Interface’s constraints. However, to translate explanations into design features, the author faced several challenges that were expressed in the research *Sub-question 1.1*:

- How can explanations be designed to be accurate and complete without overloading the user's attention?

To reply to this research question, the author built four design scenarios that helped select the cases in which explanations may be perceived as required by potential users, and designed for each scenario a set of three explanations with different types of information and different detail levels.

### 4.3.1 Designing scenarios

In order to integrate explanations in the design of the interface, the first step was to create scenarios of use. The scenarios were used as the base to build the explanations. They were needed to narrow the design to a limited set of functionalities and to show a range of potential different types of benefits. In fact, some of them have been described as high or low impact, other than local or global. The four described scenarios represented a small but representative range of use cases that generated different types of explanations. This step was also necessary to design a set of explanations the user could possibly perceive as needed. The following list shows the rationale used to generate them:

- One potentially high impact scenario (the suggestion provided by the AI can affect the experience significantly)
- One high-risk scenario (such as an important warning concerning the health of the user)
- One potentially low impact scenario (the suggestion provided by the AI may not affect the experience)
- Data transparency of the coffee machine
- Behavioural transparency of the coffee machine

Following the above rationale, four scenarios (*Table 2*) were created as the following:

1. You woke up late this morning because you fell asleep very late. Your health monitoring device reveals a normal heartbeat rate. You have a busy schedule today and it's about to start soon.

This scenario has a high impact on the daily user's routine. It implies several explanations on how the machine was able to turn on by itself and make a coffee suggestion based on the data from other devices.

2. You are focusing using a Pomodoro app and you have almost completed your focus-time for your task. The machine wants you to take a break to improve your focus during the next session. Also, the weather outside is cold and rainy.

The scenario above was considered as low impact. The user could find the machine's suggestion to be very pleasurable, but no negative consequences are predicted if this feature is ignored.

3. The machine suggests the best coffee based on your schedule, health, and other data. You could accept the suggestion or ignore them. The device can then continue giving these suggestions or become less proactive until it just learns your habits. You have noticed that sometimes the suggestions are very different from your habits and sometimes you prefer to ignore them. You want to see if there is a way to adapt the machine's behavior to your usage.

With transparency in mind, it's essential to explain how the AI is behaving, how much it learns passively, and how much it will insist. Will the machine continue to check on your health if you ignore all its suggestions? Giving control to the user over the algorithm can have an impact in their trust?

4. The machine exchanges data. It does this with another device in the local network, or can send or receive data from the cloud. After preparing your coffee, you want to discover how the machine gave you a suggestion.

Transparency of data transactions and passive or proactive behaviour of the machine requires an appropriate explanation.

N.	Title	Characteristics
1	Waking up	Local, High impact on user routine
2	Study break	Local, Low impact on user routine
3	Proactivity	Global
4	Data usage	Global

**Table 2.** Summary of the scenarios created

### 3.3.2 Designing explanations

In literature, the most commonly used method to generate explanations is reverse engineering, consisting of reconstructing an explanation produced by a black box model. However, different types of black box models exist. In this study, the author considered the decision tree, recognized to be one of the most interpretable and easily understandable models, primary for global, but also local, explanations (Guidotti et al., 2019).

#### 3.3.2.1 Local explanations

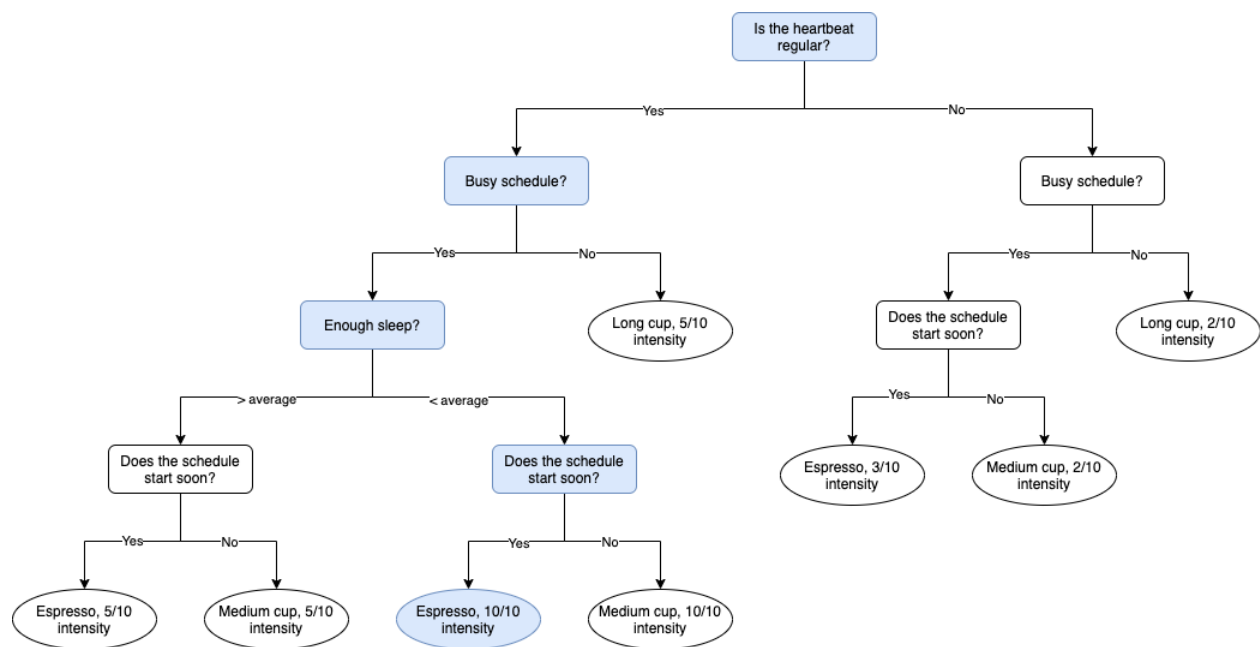
After designing the four scenarios, the author drew a simplified chart of the decision tree used for the local scenarios. This way to create the diagram is not intended to be valuable and technically applicable in software coding. Still, it represents a visual way to help the designer to “think like a software” and build the correct mental model before designing explanations.

#### Scenario 1

From the first three scenarios, this decision tree was designed starting from the tasks involved in each scenario. An example is shown in Scenario 1, **Waking up**:

*You woke up late this morning because you fell asleep very late. Your health monitoring device reveals a normal heartbeat rate. You have a busy schedule today and it's about to start soon.*

In this case the goal was a local explanation, focusing only on the reasons for a specific situation. The scenario mentions sleep data given by a smart alarm or a smartwatch, a health monitoring device that detects heartbeat that can also be the smartwatch. Also, the schedule is mentioned and we can assume the owner of the coffee machine uses a calendar application to keep track of the schedule. It also describes a specific moment of the day, the morning, in a difficult situation. The level of impact of the explanation on the user's routine can be high in these terms. By designing all the possibilities of the software involved within this specific situation, the designer can see all other possible outcomes. An example of this step is provided from the figure below (Figure 1).



**Figure 1.** Basic tree diagram to explore the possible outcomes of a specific situation

After designing the diagram, the author highlighted the path that satisfied the scenario and applied reverse engineering to shape three levels of explanations:

1. Based on your sleep, health and schedule: espresso 10/10.
2. Since you experienced irregular sleep, a busy schedule and first event in 15 minutes: espresso 10/10.

3. Sleep: irregular; schedule: busy; first event: 15 minutes. Suggestion: espresso 10/10.

Since the scenario described the use of health monitoring, heartbeat was considered in the diagram, but since it had not provided valuable insights for the user because it was normal, it wasn't taken in consideration in the explanation. Since explanation has to be complete without affecting the user's attention, heartbeat can be considered as high impact information only if it's not regular. By not showing this data, we can avoid overloading the user with unnecessary effort to read additional information.

However, the explanations generated differed in the level of detail and the presentation of data. The first explanation informed the user about what type of data affected the suggestion of the coffee with the maximum level of strength, without mentioning how this data affected it. Instead, the second explanation was more detailed, saying that the sleep is irregular, the schedule busy, and it's starting soon. The third explanation implied a more schematic approach: the goal was to discover if explanation can be exhaustive and as short as possible.

## Scenario 2

Based on the same approach also Scenario 2, **Study break**, was explored with the tree diagrams, and another set of three explanations have been developed:

1. I recommend: long coffee 4/10.
2. The perfect coffee for a study break: long cup 4/10.
3. Coffee break of 10 minutes detected. Cold weather outside. Suggestion: long cup 4/10.

Since the scenario was considered as low impact, the first explanation is actually a non-explanation. For low impact scenarios, the goal is to understand not only the level of detail but also if they even need an explanation. The second one has a very low level of detail. It just gives a hint to the user that the machine understood somehow that a break was detected, without adding how it calculated the suggestion, for example how the weather was taken into account. The third explanation instead is the most complete and it explains all the data involved.

### 3.3.2.2 Global explanations

Scenarios 3 and 4 required a global approach. The goal of providing explanations is to help the users to form a correct mental model of how the coffee machine's black box works overall. In these cases, drawing the tree diagrams would have been a very difficult task and probably not effective to describe how the algorithm takes the decisions.

#### Scenario 3

Scenario 4, called **Proactivity**, was created to open the black box to a user by providing control over the AI's behavior. The hypothesis was that users are more likely to understand the overall operation inside the product if they can control if a system behaves proactively or reactively. The distinction between these two behaviours came from the concept of agency that we can find in the literature. Depending on how the algorithm is designed, we can have smart objects that are efficient in passively learning habits and others in motivating people to take immediate action. For this scenario, three explanation were created, that also includes how they can be visually presented inside the interface:

1. From the menu section, you can change the level of proactivity. At the maximum level, the algorithm of the machine will try to motivate you to change behavior; at the minimum level it will simply adapt to your routine.
2. From the menu section, you can change the level of proactivity. At the maximum level, the algorithm of the machine will try to motivate you to change behavior; at the minimum level it will simply adapt to your routine. For each level, there is an example of how the choice will influence the recommendations.
3. Directly from the main page of coffee selection, you can see your usual choice next to the best option suggested by the coffee machine.

The first two explanations changed only in level of detail, while the last one consisted of a different approach to explaining the behavior. Since the hypothesis was that control may benefit user's trust, the author included a third explanation in which users can only digest the information without actually controlling it.

## Scenario 4

For Scenario 4, called **Data usage**, the author wanted to address the issue of data transparency, and the explanations generated aimed to help users to form a correct mental model on how data goes in and out of the system. Since data and privacy are big concerns in IoT (Lepekhn et al., 2019), the hypothesis was that explaining data transparency can improve trust towards technology. The three explanations for this scenario consisted of different representations of data movements:

1. Timeline in which you can see the data transactions between the coffee machine and other devices or cloud.
2. Timeline in which you can see the data transactions between the coffee machine and other devices or cloud. You can also expand each transaction to see more details.
3. Besides the timeline, you can also find data usage directly on the coffee selection page, with little icons that inform you about what type of data transaction is happening in the background.

The idea was to explicit the data transactions to open the black box, but to not overload the user with too much information it was necessary to represent different levels of detail. In the next step an online survey has been designed to discover what are the user's preferences in the described scenarios. A summary of the explanations generated in this step is provided below (*Table 3*).

Scenario	Explanations
<b>1. Waking up</b>  You woke up late this morning because you fell asleep very late. Your health monitoring device reveals a normal heartbeat rate. You have a busy schedule today and it's about to start soon.	<ol style="list-style-type: none"><li>1. Based on your sleep, health and schedule: espresso 10/10.</li><li>2. Since you experienced irregular sleep, a busy schedule and first event in 15 minutes: espresso 10/10.</li><li>3. Sleep: irregular; schedule: busy; first event: 15 minutes. Suggestion: espresso 10/10.</li></ol>



<p><b>2. Study break</b></p> <p>You are focusing using a Pomodoro app and you have almost completed your focus-time for your task. The machine wants you to take a break to improve your focus during the next session. Also, the weather outside is cold and rainy.</p>	<ol style="list-style-type: none"> <li>1. I recommend: long coffee 4/10.</li> <li>2. The perfect coffee for a study break: long cup 4/10.</li> <li>3. Coffee break of 10 minutes detected. Cold weather outside. Suggestion: long cup 4/10.</li> </ol>
<p><b>3. Proactivity</b></p> <p>The machine suggests the best coffee based on your schedule, health, and other data. You could accept the suggestion or ignore them. The device can then continue giving these suggestions or become less proactive until it just learns your habits. You have noticed that sometimes the suggestions are very different from your habits and sometimes you prefer to ignore them. You want to see if there is a way to adapt the machine's behavior to your usage.</p>	<ol style="list-style-type: none"> <li>1. From the menu section, you can change the level of proactivity. At the maximum level, the algorithm of the machine will try to motivate you to change behaviour, at minimum level it will simply adapt to your routine.</li> <li>2. From the menu section, you can change the level of proactivity. At maximum level, the algorithm of the machine will try to motivate you to change behaviour, at minimum level it will simply adapt to your routine. For each level there is an example on how the choice will influence the recommendations.</li> <li>3. Directly from the main page of coffee selection, you can see your usual choice best option suggested by the coffee machine.</li> </ol>
<p><b>4. Data usage</b></p> <p>The machine exchanges data. It does this with another device in the local network, or it can send or receive data from the cloud. After preparing your coffee, you</p>	<ol style="list-style-type: none"> <li>1. Timeline in which you can see the data transactions between the coffee machine and other devices or cloud.</li> <li>2. Timeline in which you can see the data transactions between the coffee machine and</li> </ol>

want to discover how the machine gave you a suggestion.	<p>other devices or cloud. You can also expand each transaction to see more details.</p> <p>3. Besides the timeline, you can also find information of data usage directly from the coffee selection page, with little icons that inform you on what type of data transaction is happening in the background.</p>
---	--

**Table 3.** Summary of explanations created for each scenario

## 4.4 Online survey

In this phase, the study introduces a quantitative data collection intending to understand potential users' preferences. Scenarios and explanations designed in the previous phase were presented to participants. The gathered data was used to inform the design of the coffee machine. A secondary goal of the survey was to select participants with a low level of trust in smart home devices and AI-based products in general. From the total of 16 participants selected, all participants completed the questionnaire.

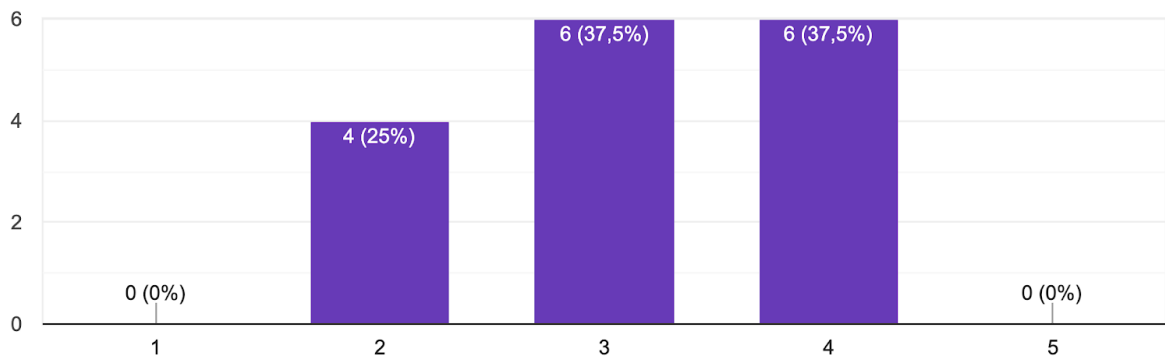
### 4.4.1 Data analysis

In this section, the data analysis of the online questionnaire is presented. This section is considered part of the design process since its results represent an initial investigation to inform the design of the prototype.

#### 4.4.1.1 General questions

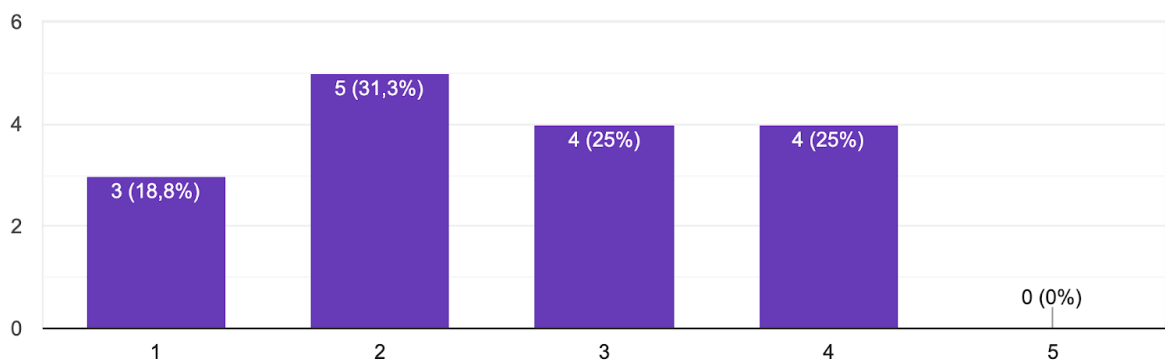
The introductory part of the survey consisted of one demographic question about the age to understand the stratification of the population, and three other questions. One question was about rating the general level of trust on AI-based devices in order to facilitate the selection of participants in the next phase, ranging from 1 one to 5. According to the Google Form data, only

four participants answered to have a low level of trust (2) in AI-based devices, while six participants selected 3, and the other six participants selected 4 (*Figure 2*).



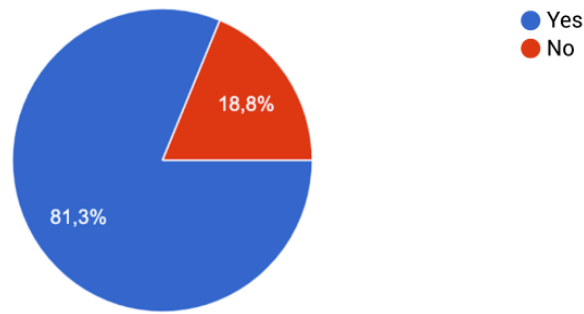
**Figure 2.** Self-assigned level of trust towards AI-based devices

The second question was asked to understand if participants had a general understanding of how this technology works (*Figure 3*).



**Figure 3.** Self-assigned level of understanding of AI-based devices

With the third question, participants were asked to express if they think that explanations in these devices could improve their trust, and the majority of them (81,3%) answer positively (*Figure 4*).

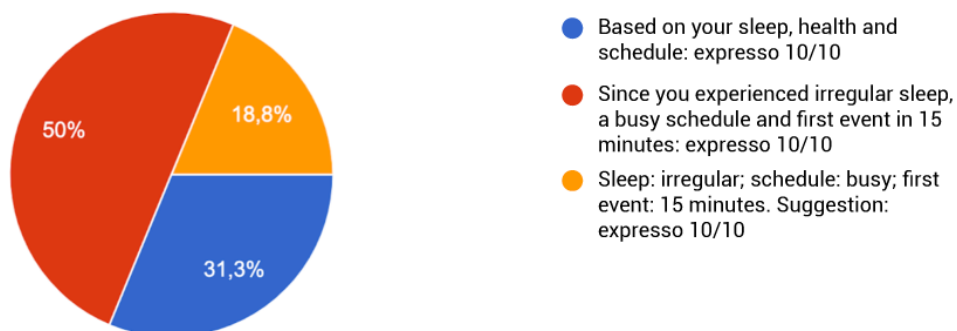


**Figure 4.** Do you believe that providing explanations could help you improve your trust towards these devices?

Since the level of trust was generally medium, and the level of knowledge was low, these results represented the confirmation to have an appropriate population for the survey. Also, they were generally positive about the role of explanations in building trust in these systems.

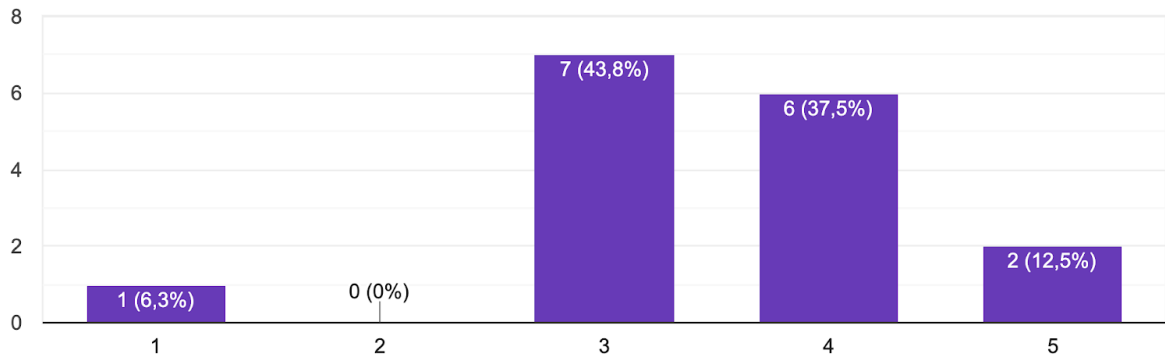
#### 4.4.1.2 Scenario 1

For Scenario 1, **Waking up**, 50% of participants preferred the first explanation, the most complete of the three (*Figure 5*), while 31% preferred the one with less detail. 18,8% chose the third one, which had a different presentation but it was still complete in its details. Scenario 1 was considered to have high impact on the user routine, and this answer has demonstrated that for high impact recommendations it is desirable to provide a very detailed explanation.



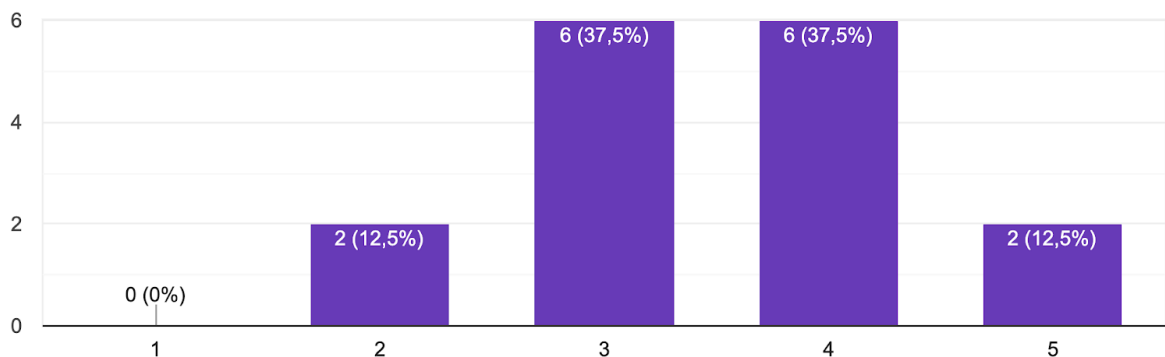
**Figure 5.** Preferences on three explanations provided for Scenario 1

The following question was asked to understand if the participants were satisfied with the explanations provided in terms of trust (*Figure 6*).



**Figure 6.** Self-assigned level of trust of the explanation provided on a scale of 1 to 5

The next question was asked to discover if the participants were satisfied with the explanations provided in terms of understanding the algorithm (*Figure 7*).



**Figure 7.** Self-assigned level of understanding of the algorithm on a scale of 1 to 5

At the end of this section, the questionnaire asked if participants had questions about how the coffee machine figured out the right suggestion. Two participants wrote:

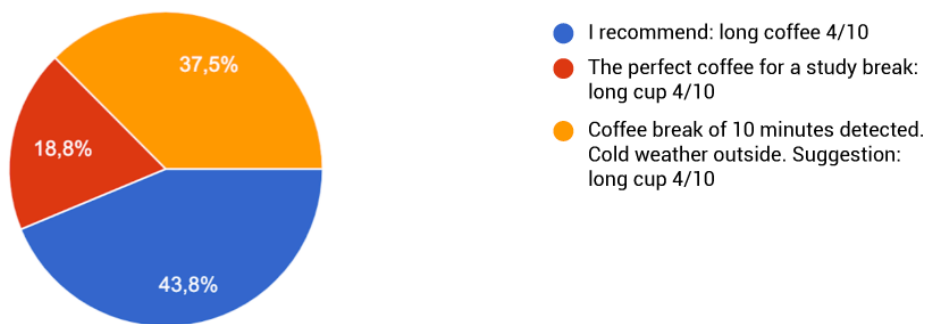
“I would like to know better how health condition is evaluated by the machine” - Survey participant

“I want more detail about irregular sleep, how is it irregular?” - Survey participant

These comments raised, even more, the awareness that the explanation must be complete for high impact scenarios, especially when it concerns health issues.

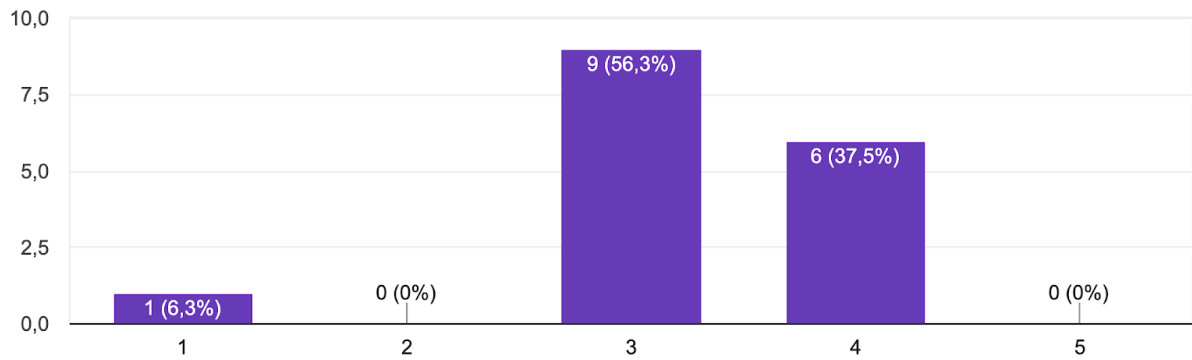
#### 4.4.1.3 Scenario 2

For Scenario 2, **Study break**, 43,8% preferred the no-explanation, but a good 37,5% chose the most detailed one, leaving only 18,8% with the neutral one (*Figure 8*).



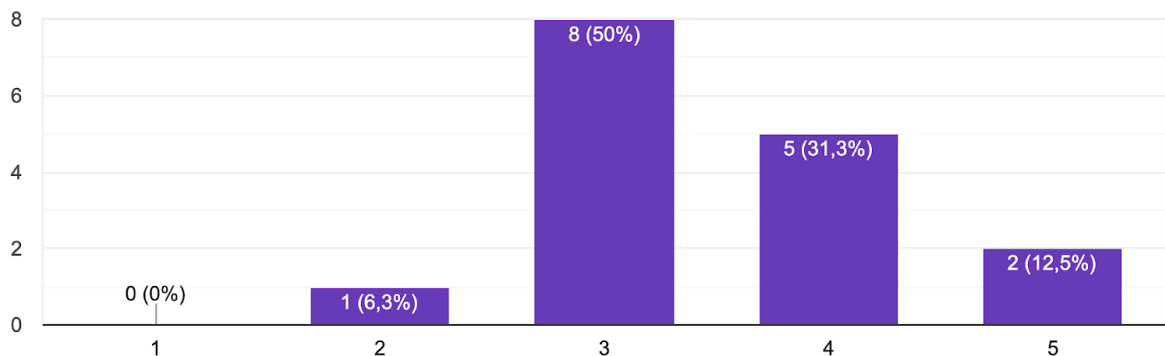
**Figure 8.** Preferences on three explanations provided for Scenario 2

The following question was asked to understand if the participants were satisfied with the explanations provided in terms of trust (*Figure 9*).



**Figure 9.** Self-assigned level of trust of the explanation provided on a scale of 1 to 5

The next question was asked to discover if the participants were satisfied with the explanations provided in terms of understanding the algorithm (*Figure 10*).



**Figure 10.** Self-assigned level of understanding of the algorithm on a scale of 1 to 5

At the end of this section, participants were asked if there could be an element that could improve their trust. Two participants wrote:

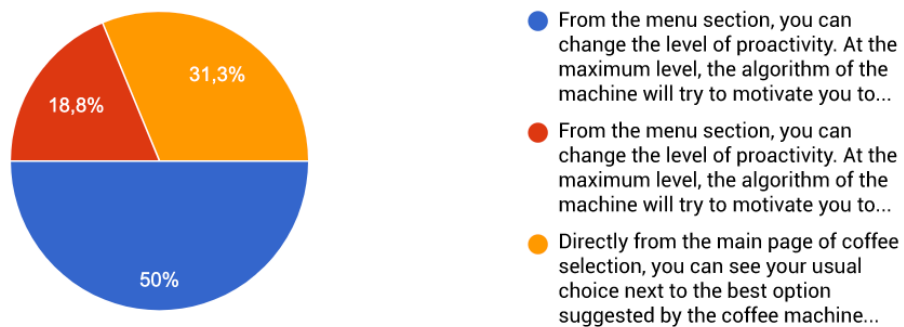
“Data” - Survey participant

“The detailed one is good but it’s too complex” - Survey participant

For this scenario, the results sounded contradictory. Some participants preferred to have more detail and asked for more data; others would rather not have explanations at all. One insight is evident: solutions in the middle of the two are not very accepted.

#### 4.4.1.4 Scenario 3

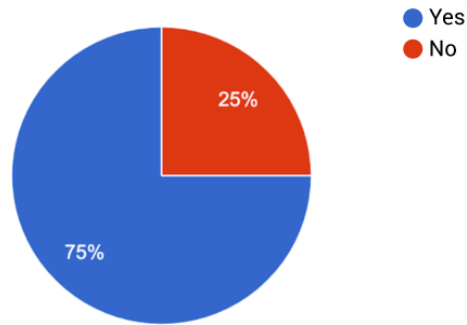
For Scenario 3, **Proactivity**, 50% preferred the first explanation in which the user can select the level of proactivity with a short description of what the level does. A significant 31,3% liked the third solution, in which two explanations are displayed directly inside the coffee selection panel. The second option consisted of the same content as the first one, but besides a description of the proactivity level, an example was provided (*Figure 11*). Overall the participants preferred the option of controlling the level of proactivity, consisting of the first two explanations with a total of 68,8%.



**Figure 11.** Preferences on three explanations provided for Scenario 3

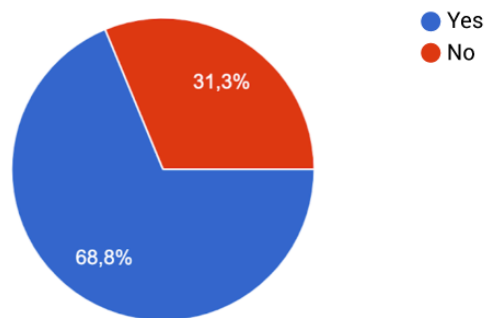
Additional questions were asked to understand if the solution could improve their overall trust, and 75% of them responded positively (*Figure 12*).





**Figure 12.** Do you think that this option can improve your level of trust towards the machine?

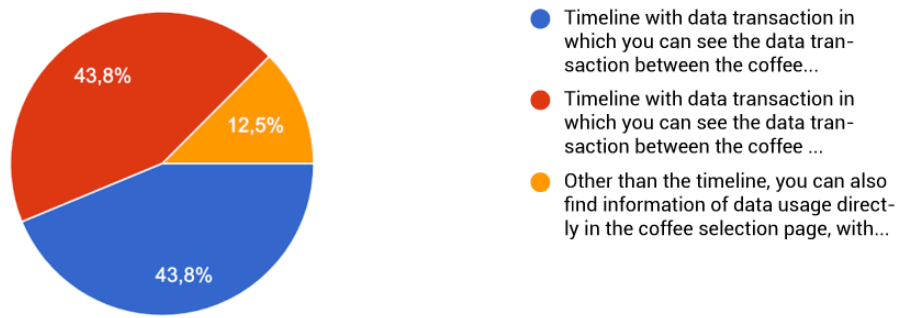
The next question was asked to discover if the solution could improve their understanding of the algorithm (*Figure 13*). The hypothesis was that by providing control over the algorithm of the coffee machine, users are willing to learn how it works by experimenting with the settings.



**Figure 13.** Do you think that this option can improve your level of understanding of the algorithm?

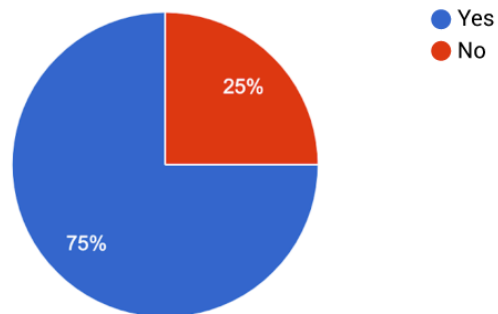
#### 4.4.1.5 Scenario 4

For Scenario 4, **Data usage**, the participants preferred the first two explanations in an equal way, with 43,8% for both options. The three explanations consisted of a timeline of events with data transactions, and they differed only in the level of details. Results showed that the third one was too complex, adding too much information (*Figure 14*).



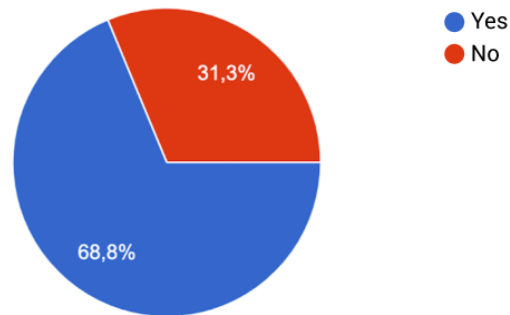
**Figure 14.** Preferences on three explanations provided for Scenario 4

An additional question was asked to understand if the solution could improve their overall trust, and 75% of them responded positively (*Figure 15*).



**Figure 15.** Do you think that this option can improve your level of trust towards the machine?

The next question was asked to discover if the solution could improve their understanding of the algorithm (*Figure 16*). The hypothesis was that by providing a visual way, like a timeline, to display AI's internal decisions, the user would form a correct mental model of how the machine "thinks" and takes decisions. The timeline was also a solution to provide a global explanation instead of several local explanations for each event.



**Figure 16.** Do you think that this option can improve your level of understanding of the algorithm?

## 4.5 Prototype

The coffee machine has the ability to take care of several aspects of the life of its owner. With the Internet of Things, it can connect to other devices to make accurate suggestions based on quality of sleep, the schedule, and even the weather. It can also gently remind you with its coffee aroma that after several hours of concentration it's time to clear your mind and take a break. All these abilities were conceived thanks to the Thing-Centered Design approach.

This phase of the study, which follows a Research-through-Design methodology, used design Design Fiction to produce a design artifact presented as an everyday smart object in the context of a smart home in a near-future scenario. In particular, the study used the Design Fiction as World Building approach coined by Coulton et al. (2017). The prototype was not meant to be functional, but its purpose was to induce people to think critically about issues that the design embodies (Coulton et al., 2017).

For its ability to take care of its owner, the coffee machine interface was designed to be classy, exclusive, inspired by the actual high-end brands in the market. Besides being smart, the most differentiating feature was reutilizing the coffee grounds to diffuse coffee aroma during the preparation, so it was called Aroma. Aroma is a product of a near but plausible future, when smart home objects will no longer be seen as a trend, but they will be established in every home.

A logo (*Image 2*) and a fictional Amazon product page (*Image 3*) were created to instill in the potential user the idea of a plausible fictional context built around the product.



Image 3. Logo of the Aroma coffee machine

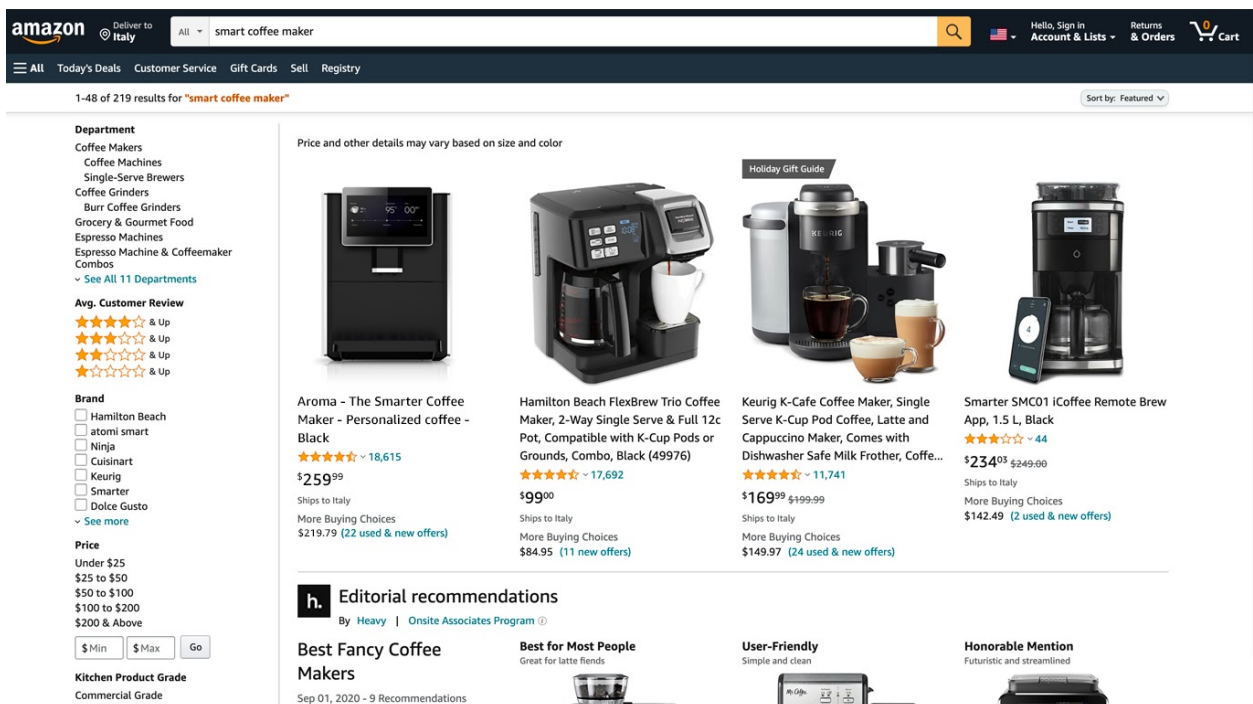


Image 4. Fictional Amazon page with Aroma

## 4.5.1 Interface Design

The first step in building the prototype was designing the Graphical User Interface. The tool used for this purpose was Adobe XD, which allows to draw the artboard and connect them into an interactive prototype that can be shared with a link or via its native application. Four user flows were selected, corresponding to the four scenarios previously described: **Waking up**, **Study break**, **Proactivity**, and **Data usage**.

### 4.5.1.1 User Flow 1

The interface was divided into three sections: one section dedicated to the suggestions, one dedicated to a free coffee selection, and the remaining area to a rounded, visible Brew button. The suggestion in the first container affects the content of the second container, which forms a preset ready to be used, but it also allows users to ignore the suggestion and act as they please.

The first user flow is limited to Scenario 1, representing the owner's morning routine, in a particular case of fatigue and busy schedule. The coffee machine detects the owner is waking up; it checks data sleep and notices that s/he has an irregular sleep. It starts pre-heating the water and diffusing the coffee aroma, then it also inspects the schedule and finally provides a recommendation.

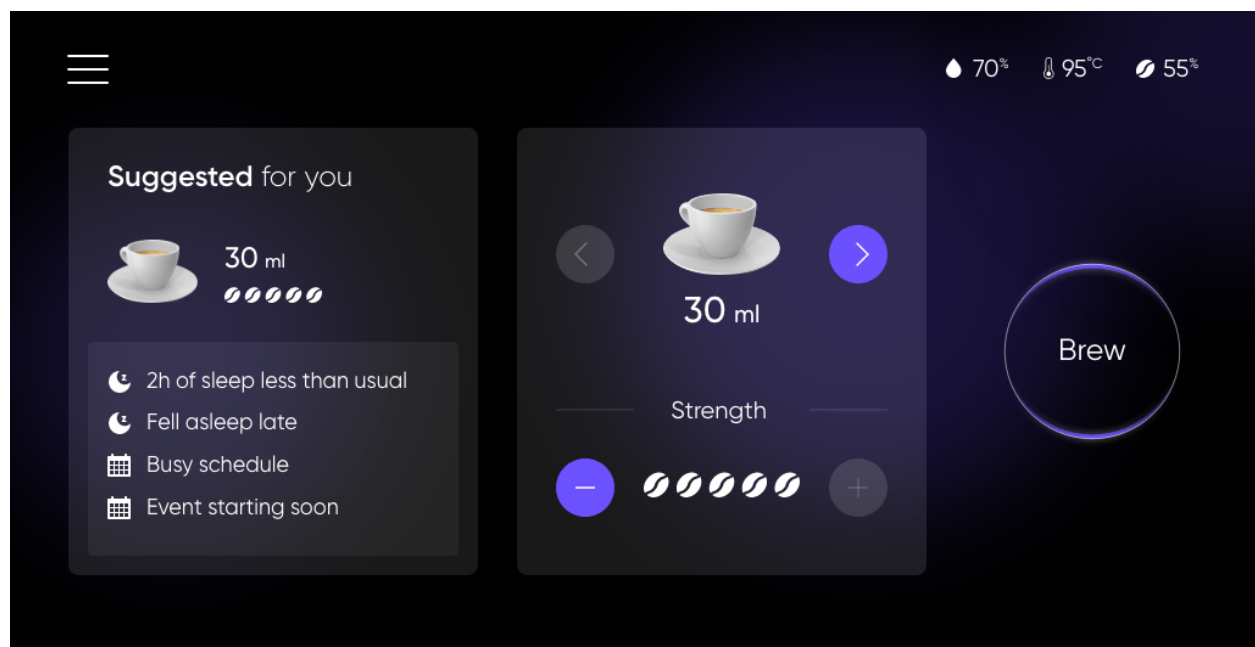
The survey conducted in the previous section served the purpose of informing the suggestion panel in this first section of the prototype. Since the results showed that for high impact recommendations it is desirable to provide a complete explanation, but in other cases fewer details may be needed, the suggestion panel has been conceived like a modular container. The suggestion panel is divided into two parts: the suggestion itself, which is also mirrored in the next panel, and the explanation below.

The information provided in the explanation are the ones provided to the survey participants, resulting from the decisional tree, with the chosen level of detail:

1. Since you experienced irregular sleep, a busy schedule and first event in 15 minutes:  
expresso 10/10

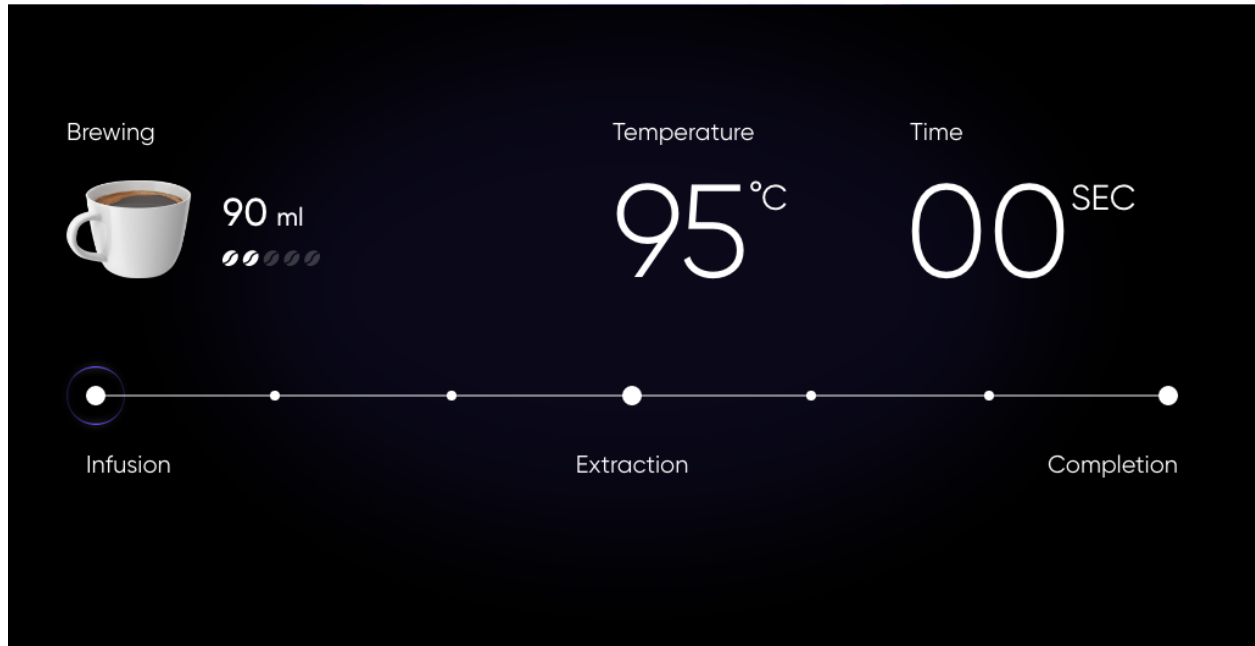
Since two participants pointed out that they needed even more details to understand how the sleep was irregular, additional information was provided. In particular, irregular sleep information was further detailed into two different explanations.

The challenge to translate textual explanations in GUI is the space: the more you explain, the more space you need, and sometimes designers don't have enough. On the positive side, the GUI allows designers to use graphical elements to help the explanations be more recognizable. In the image below (*Image 4*), icons were used to let the user have a quick understanding of the type of data used, which allowed the explanation to be concise. During the design process also the scale changed: the strength scale became a one-to-five scale to reduce the cognitive effort of the users, and the quantity was expressed in ml, accompanied by an image of the corresponding cup. In this way, an espresso 10/10 became a 30 ml cup with 5/5 strength in the GUI.



**Image 5.** Starting page of the interactive prototype of the first user flow

The final prototype for the first user flow allowed the users to change the dimension of the cup and calibrate the strength of the coffee if they decide to ignore the suggestion. In this way, the suggestion is not invasive because it works as a standard coffee machine. Once they decide, they can press the Brew button (*Image 5*).



**Image 6.** Final screen of the first user flow

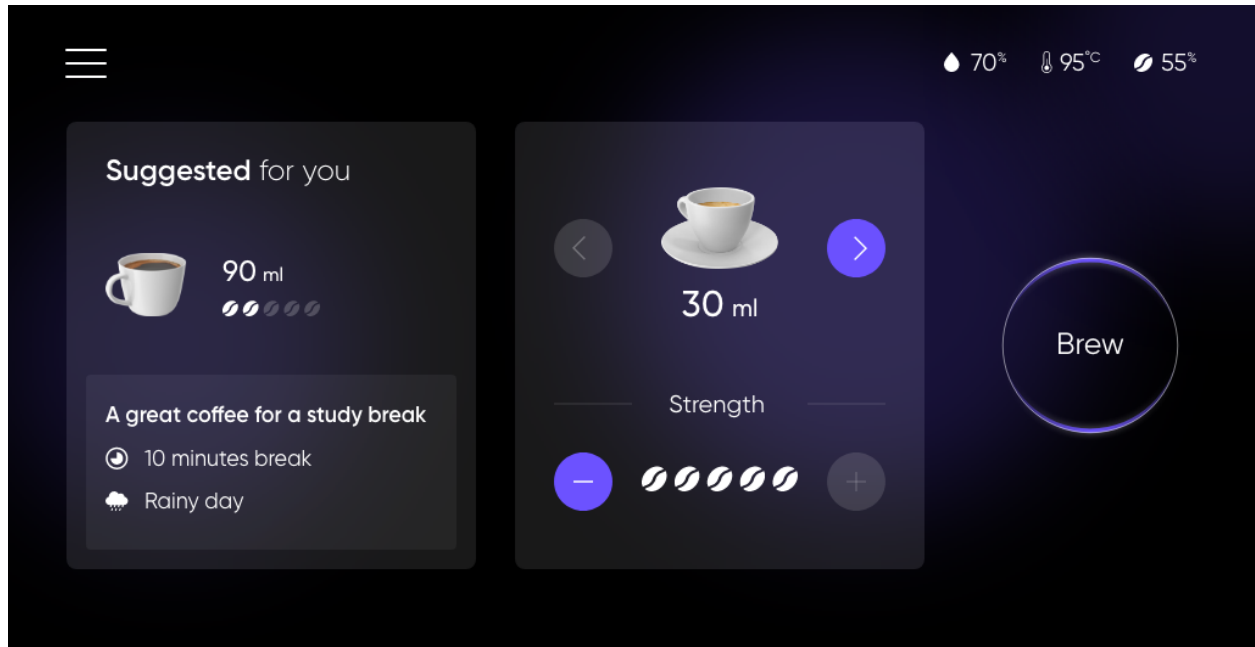
#### **4.5.1.2 User Flow 2**

The second user flow is limited to Scenario 2, representing the use of the coffee machine during a study or work break. The coffee machine detects from the focus application of the smartphone that the focus time is almost over, so it wants to help the owner to take a healthy break. It starts pre-heating the water and diffusing the coffee aroma five minutes before the focus time ends, then it also checks the weather, and if it's cold, it can make the owner feel cozy and relaxed.

The survey conducted in the previous section served the purpose of informing the suggestion panel in this first section of the prototype. For this scenario, the results sounded contradictory; some participants preferred to have more detail and asked for more data, others would rather not have explanations at all. Since the explanation occupies a section of the interface that doesn't interfere with the coffee selection, a complete explanation was adopted as a solution, complemented by a sentence that constitutes a welcoming message explaining why the machine turned on. The original explanation for Scenario 2 became:

1. A great coffee for a study break. 10 minutes break, rainy day: 4/10

Translating the explanation on the GUI, the indication for strength became a scale from 1 to 5 and the quantity became 90 ml. The final appearance of the graphical user interface is presented below (*Image 6*). Pressing the Brew button, the user is redirected to the Brew screen.



**Image 7.** Starting page of the interactive prototype of the second user flow

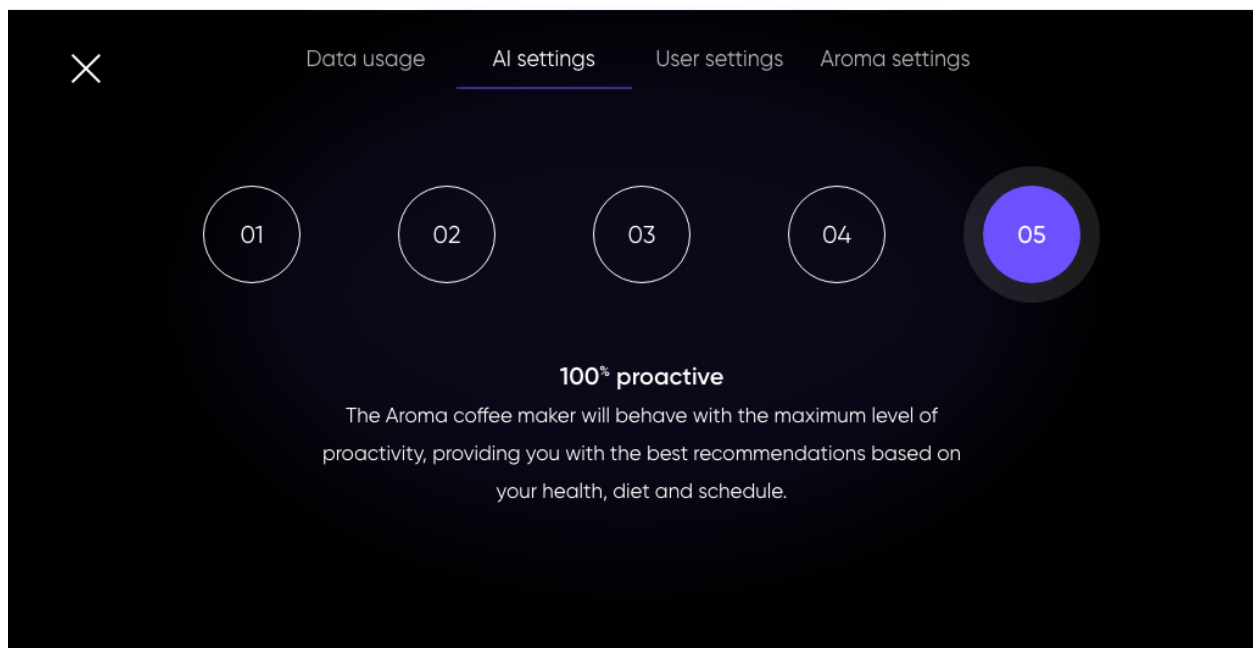
### 4.5.1.3 User Flow 3

The third user flow is limited to Scenario 3, representing a new feature to control the coffee machine's proactivity. Since it was not supposed to be used frequently, the feature was placed inside a menu section. For this feature, a global explanation was required, so a dedicated page was designed.

The survey conducted in the previous section suggested that this feature was recommended and it could potentially increase both trust and understanding. As it was a global explanation, it had to be designed to inform about the overall functioning of the AI inside the machine without being overwhelming for the user's attention. A set of five tappable levels have been placed inside the screen (*Image 7*). The interactive prototype allowed the user to explore the five possibilities. Additional explanations of what the level of proactivity is affecting the behavior were provided below each one, with these descriptions:



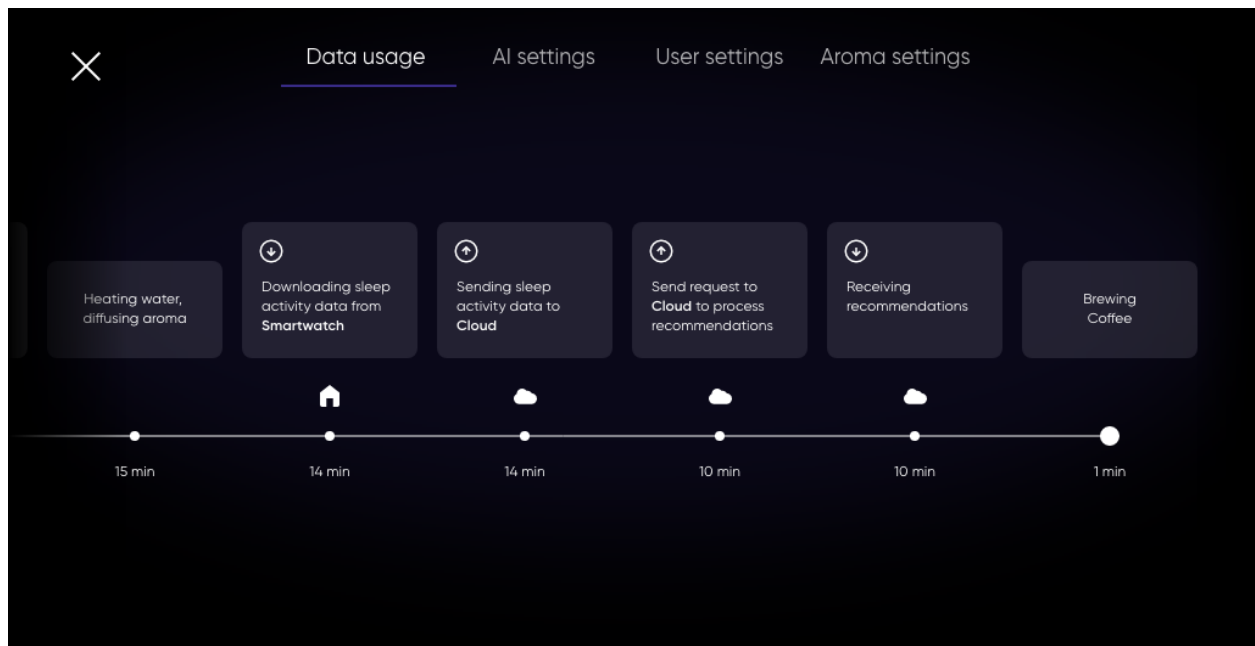
1. **Reactive.** The Aroma coffee maker will be only reactive, providing you with recommendations based solely on your previous choices.
2. **Slightly proactive.** The Aroma coffee maker will behave with a low level of proactivity. It will learn your habits, acting proactively in a few cases, like lack of sleep or high heartbeat rate.
3. **Moderately proactive.** The Aroma coffee maker will find a perfect balance between a reactive behavior and a proactive one.
4. **Very proactive.** The Aroma coffee maker will behave with a high level of proactivity, providing you with the best recommendations based on your health, diet and schedule, but also considering your most repetitive actions.
5. **Highly proactive.** The Aroma coffee maker will behave with the maximum level of proactivity, providing you with the best recommendations based on your health, diet and schedule.



**Image 8.** Starting page of the interactive prototype of the third user flow

#### 4.5.1.4 User Flow 4

The final user flow was limited to Scenario 4, which contained a global explanation of data usage. The results of the survey showed that only 12% preferred a very highly detailed explanation. Since explaining a complex feature such as data transactions could confuse or overwhelm the user, the final screens contained a simple timeline with just the minimum information to instill a correct mental model. With a system of icons and a visual and scrollable representation of the timeline, users can see which data goes into the cloud, which comes from another device, which is used to make a calculation, and finally, which stays inside the coffee machine (*Image 8*). The prototype represents the first scenario in the form of a timeline.



**Image 9.** Starting page of the interactive prototype of the fourth user flow

#### 4.5.2 Physical prototype

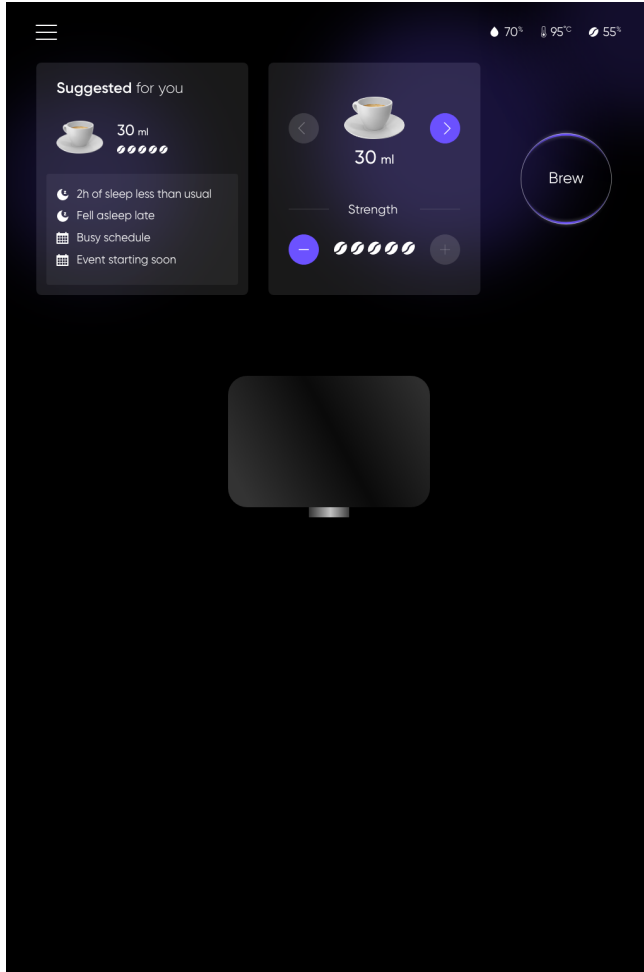
The developed prototype was not functional, since the goal of this phase was to build a fictional world. However, to give the perception of an actual working machine, it was built to be realistic and, above all, to integrate the digital interface.

A plate of expanded polystyrene was cut in multiple shapes, which were then stacked and glued together. After some hours of rest, the final shape was sculpted and sandpapered (*Image 9*).



**Image 10.** First version of the physical prototype

During this phase, it was crucial to find a way to integrate the digital screen. Thanks to the ductility of the material, the solution was to make grooves that allowed the prototype to accommodate an iPad vertically. It was necessary to take a step back and change the digital prototype that had to reproduce the whole front panel of the machine. The new format (*Image 10*) of the digital prototype was created to simulate the brewer of the coffee machine.



**Image 11.** New format of digital prototype base on iPad 10

The digital prototype was uploaded into the Adobe Cloud and opened on the iPad using the Adobe XD App. After testing the allocation of the iPad (*Image 11*), the physical prototype was painted black. The final result was an integrated physical and digital prototype, a mixed solution to produce a low-fidelity but credible smart object without running into technologic issues. In addition, a smartwatch and a paper-prototype of a smart home device were positioned next to the coffee machine to help future participants to be immersed in the fictional world.



**Image 12.** Physical and digital prototype



**Image 13.** Final prototype and setting

## 5. Study procedure

This research hypothesizes that by combining new frameworks from the Thing-Centered Design and lessons from Explainable AI in a structured approach, design practitioners have the tools to build transparent smart home devices. In order to validate this assumption, the author used quantitative methods such as pre and post-study surveys, designed to measure the improvement of trust before and after interacting with the prototype. In addition, the author used qualitative methods such as structured interviews and think-aloud protocol to collect general feedback from the interaction with the prototype and understand if the artifact helped the participants to interpret the AI model behind the machine. These methods were meant to understand if the prototype produced the opposite effect of a black box: a transparent box.

The goal of this section is to address the last two sub-questions of the study:

- *Sub-question 1.3.* To what extent does the artifact help users in interpreting the AI model?
- *Sub-question 1.4.* How do users describe their experience with the artifact in terms of trust?

### 5.1 Participants

Participants for this study were recruited from the survey conducted during the explainability phase. They were selected participants who evaluated their trust towards AI-devices with a rate of 3 or below, ranging from 1 to 5. The author recruited five participants who met the requirements through convenience sampling, since the study was conducted in-person. The participants were non-expert in the field, so they did not have a job, degree, or background related to software development or design.

### 5.2 Ethical considerations

In order to meet ethical research standards, participants were informed about the purpose of the study and gave written consent to be recorded.

## 5.3 Interview procedure

The study used mixed methods to address the research questions. First, the prototype setting was arranged with a room equipped with the coffee machine prototype, a smartwatch, and a paper-prototype of a smart home device with two screens. The additional objects were provided to build a fictional world.

The interviews were conducted in five different sessions, and before each session a consent form was provided. The interviews included general questions such as name, age, and to evaluate and describe their level of trust and understanding of the AI model. This initial phase was followed by a Human-Computer Trust Scale questionnaire (*Appendix F*) developed by Gulati et al. (2019) to measure their level of trust before the interaction took place.

During the interaction phase, the author read four scenarios corresponding to the four prototyped user flows. Participants were asked to explore each user flow and to think aloud describing their experience. The think-aloud method was used not as a usability test, but mostly to capture their cognitive process while forming a mental model of the artifact. For each user flow, the think-aloud protocol was followed by a structured interview (*Appendix C*). A structured approach helped the researcher to better compare the interview transcripts during the analysis phase.

At the end of the interview, the participants were asked again to fill the HCTS questionnaire to measure how the interaction with the prototype changed their perception of trust.

### 3.5.3 Potential threats to validity

Since the interview process took place in Italy, the participants were comfortable expressing their thoughts in the Italian language. The survey and interview protocol was provided in Italian (*Appendix C*), but to allow the readers of this study to interpret and understand the data provided, the analysis and results are presented in English as well as the protocol (*Appendix D*). To validate the results, the author is aware that an official translation may be necessary to avoid possible threats to the validity of the study. However, since the questions were relatively straightforward, hiring a professional translator was not considered essential for the research.

## 6. Results

Quantitative and qualitative data were gathered from five sessions with five participants. The following section outlines the combined results.

### 6.1 Introductory questions

The introductory questions aimed to gather personal data such as name (which in this section remain anonymous) and age. Other questions were asked to frame the participant's attitude towards AI. A summary of the introductory questions is provided below (*Table 4*). The level of trust was a self-assigned evaluation they were asked to give also in the explanation survey during the design process, and it was used as a hook to proceed with qualitative questions.

Participant	Age	Level of trust (1-5 scale)
1	32	2
2	26	3
3	25	3
4	59	3
5	58	3

**Table 4.** Summary of introductory questions

These questions were followed by structured qualitative questions about how participants assigned this vote (*Table 5*) and also how they described their knowledge on how AI-based home devices work (*Table 6*). A thematic analysis was conducted where the author coded the answer to see if patterns were recurrent.



Participant	Can you describe why you assigned rate X?
1	I don't trust the company behind I don't know how it works
2	AI could fail
3	AI could fail
4	I don't' know how it works
5	I trust AI only in the medical field

**Table 5.** Themes emerged from the question “Can you describe why you assigned rate X?”

Participant	Can you describe your level of knowledge about the algorithms with which we interact every day?
1	Generic idea
2	Generic idea
3	Vague idea
4	I don't know how it works
5	I don't know how it works

**Table 6.** Themes emerged from the question “Can you describe your level of knowledge about the algorithms with which we interact every day?”

The qualitative introductory questions revealed that the motivation of scarce trust towards AI was diverse. However, two themes recurred twice: two people answered that their lack of trust is due to the system's fallibility.

“I think technology could be wrong at any time, it's not a thinking being, even if it can make very difficult calculations” - Interview participant

Another interesting point of view was given by one of the participants, who trusted the AI only in medical fields, not when applied to our homes.

The answers to the second question were coded following their self-assigned level of understanding. It is worth mentioning that the first three participants who had a general or vague idea of how AI works on their devices were the younger participants in the range of 25-35 years old.

## 6.2 Scenario 1

At this phase of the interview process, the author read the scenario and the participants were asked to explore the prototype and brew a coffee. They were asked to express their thoughts by speaking out loud. This was useful to intercept feelings and opinions that could be missing with the structured questions. The outcomes were different; it happened that one participant immediately accepted the suggestion without commenting, others were more critical. The transcripts were coded to understand the main themes regarding the first scenario (*Table 7*).

The participants assigned themselves with a high rating on the trust level. When they explained why, the most recurring themes were that the explanation reflected what they experienced, even if they didn't take the suggestion as it was. Another important theme is the freedom of choice, which was also recurring during the think-aloud. Only one participant was satisfied with the explanation because it was "very exhaustive" so labeled as "completeness". The second question led to similar answers. Some of the participants formulated what they understood from the machine's behavior and it was an accurate representation of the decisional tree that was originally designed, while others had just awareness of some sort of data transaction that happened which led to the suggestion.

Theme	Transcript
Evaluation of the explanation	"The suggestion is coherent, the coffee is strong because I have to go through a busy day and I'm tired" - Participant 1

Expressing a general idea of how the system works	<p>“I believe that it takes my sleep data from my bracelet, and my schedule from my calendar in one of my devices.” - Participant 1</p> <p>“It’s saying that I slept 2 hours less than usual, I fell asleep late, I have a busy schedule that it’s about to start so it suggests to me a strong coffee.” - Participant 2</p> <p>“Here I can decide the quantity of water that I want and also the strength. I also see the information that summarizes my day and tells me how I slept, giving me some advice.” - Participant 3</p>
Freedom of choice	<p>“I like that it gives you the explanation and it sets the coffee but you are able to change and choose by yourself.” - Participant 1</p> <p>“I’m not sure if I should take it or not, I’m thinking of doubling it. Yes, I will take it double. I started from the suggestion it gave to me to make my decision” - Participant 4</p>
Suggesting a change	<p>“I would like that after a while it will learn my habits and it will suggest what I usually take.” - Participant 1</p>
Adjusting the suggestion to their needs	<p>“I can lower the strength to 3 notches.”- Participant 2</p> <p>“I, for example, am not a ristretto lover, so I will change it according to my needs.” - Participant 3</p>
Projecting the scenario to their personal situation	<p>“It’s true that it could be the right coffee for me, but I am a very anxious person so I can’t take it too strong.” - Participant 2</p>
Accepting the suggestion straightaway	<p>“Ok, I’m good with it, I’ll take it.” - Participant 5</p>

**Table 7.** Thematic analysis of think-aloud protocol for Scenario 1

Three questions followed the think-aloud protocol. One was rating the trust on a scale from 1 to 5. This question was a hook to ask the second question about why they assigned this rate. The

last question was related to the understanding. Another goal of this protocol is to verify that the artifact can help the user understand how the machine's AI works. A summary of the themes that emerged from these three questions is provided below (*Table*).

The participants assigned themselves with a high rating on the trust level. When they explained why, the most recurring themes were that the explanation reflected what they experienced, even if they didn't take the suggestion as it was. Another important theme is the freedom of choice, which was also recurring during the think-aloud. Only one participant was satisfied with the explanation because it was "very exhaustive" so labeled as "completeness". The second question led to similar answers. Some of the participants formulated what they understood from the machine's behavior, and it was an accurate representation of the decisional tree that was originally designed. Others had just awareness of some sort of data transaction, which led to the suggestion.

<b>Participant</b>	<b>Trust rate</b>	<b>Can you describe why you assigned rate X?</b>	<b>How can you describe your understanding of AI from this explanation?</b>
1	5	Completeness	Correct mental model
2	5	Freedom of choice	Correct mental model
3	4	The explanation matched what the subject experienced	General understanding
4	5	The explanation matched what the subject experienced	Correct mental model
5	5	Freedom of choice	General understanding

**Table 8.** Summary of themes emerged during the structured interview for Scenario 1

## 6.3 Scenario 2

At this phase of the interview process, the author read the scenario and the participants were asked to explore the prototype and brew a coffee. They were also asked to express their thoughts

by speaking out loud. The transcripts were coded to understand the main themes regarding the first scenario (*Table 9*).

In general, the majority verbalized how the machine arrived at the suggestion, demonstrating that the explanation gave them a general idea of the decisional process. An important theme was introduced for this scenario: motivation. The participants wanted to express they were motivated to take a break, and this theme did not appear for the first scenario. In the freedom of choice theme, negative feedback was expressed by Participant 1, who felt that the suggestion was not necessary. As mentioned during the creation of the explanation, a non-explanation sometimes may be the best choice for low-risk scenarios. However, all the other participants were very satisfied.

Theme	Transcript
Evaluation of the explanation	“Here the suggestion is less clear. There are not strong motivations like the first one” - Participant 1
Freedom of choice	<p>“Could be enough to give the explanation only for the fact that diffused the aroma at certain times, but the choice of coffee should be free.” - Participant 1</p> <p>“The suggestion gives me trust, I can change what it suggests to me” - Participant 4</p>
Suggesting a change	“It would be nice if it understands that I usually take this coffee during this time of the day and suggests me that one” - Participant 1
Adjusting the suggestion to their needs	<p>“I would follow the advice, just maybe more intense”- Participant 3</p> <p>“It suggests me this, but I would take even bigger since it’s cold outside” - Participant 5</p>
Accepting the suggestion straightaway	“It suggests this coffee, yes, I would take this one.” - Participant 2

Motivation	<p>“After some time that I’m studying, I will definitely take a break and I will stand up from my desk. Plus, if I smell the coffee I will be even more motivated.” - Participant 2</p> <p>“This is an option I would follow more willingly than the mourning one. When I was studying I used to drink a lot of coffee, so I think it’s useful because it demonstrates that it takes care of your health.” - Participant 3</p> <p>“I would be inclined to follow the suggestion.” - Participant 4</p>
------------	---

**Table 9.** Thematic analysis of think-aloud protocol for Scenario 2

Three questions followed the think-aloud protocol. One was rating the trust on a scale from 1 to 5. This question was a hook to ask the second question about why they assigned this rate. The last question was related to the understanding. Another goal of this protocol is to verify that the artifact can help the user to understand how the machine’s AI works. A summary of the themes that emerged from these three questions is provided below (*Table 10*).

In general, the participants assigned themselves with a high rating on the trust level, except for Participant 1. Then, when they explained why, the most recurring themes were that the explanation reflected what they experienced. The first participant did not like to have the suggestion in a low-risk scenario. In this case, the participant felt that having a suggestion can undermine his freedom of choice. Regarding the second question, the participants formulated what they understood from the machine’s behavior in broad terms.

Participant	Trust rate	Can you describe why you assigned rate X?	How can you describe your understanding of AI from this explanation?
1	3	Freedom of choice	Correct mental model
2	5	The explanation matched what the subject experienced	General understanding

3	5	The explanation matched what the subject experienced	Correct mental model
4	5	The explanation matched what the subject experienced	General understanding
5	5	The explanation matched what the subject experienced	General understanding

**Table 10.** Summary of themes emerged during the structured interview for Scenario 2

## 6.4 Scenario 3

At this phase of the interview process, the author read the scenario and the participants were asked to explore the prototype and brew a coffee. They were also asked to express their thoughts by speaking out loud. The transcripts were coded to understand the main themes regarding the first scenario (*Table 11*).

Scenario 3 provided a global explanation. During the think-aloud session, the participants focused a lot in play around the five options to see which one was the best for their needs, and this is why this theme prevailed. Three participants also wanted to let the author know that the explanation was useful or welcomed.

Theme	Transcript
Evaluation of the explanation	<p>“I like the idea that it will suggest what it thinks it’s right for me.” - Participant 2</p> <p>“I think this is useful.” - Participant 3</p> <p>“It’s good that it learns your habits, because even if a person did not sleep, it isn't true that has to take this one. But really helpful” - Participant 5</p>
Freedom of choice	<p>“I would keep a bit of freedom. But I think it’s useful, I can have control.” - Participant 4</p>

Adjusting the suggestion to their needs	<p>“So the first one suggests what you usually take without adding any judgment, the last one seems to care about you. I would like to keep it initially at 4.”- Participant 1</p> <p>“I also want that it respects my habits, I think I will try level 3.” - Participant 2</p> <p>“It gives me tachycardia. But wait, maybe because I have this problem I should use high proactivity so it will suggest the coffee based on my health forcing me to take a low dose of caffeine.” - Participant 3</p> <p>“I would put level 4.” - Participant 4</p> <p>“I would use the level 3, I’d start with this one to see how I feel about it.” - Participant 5</p>
---	---

**Table 11.** Thematic analysis of think-aloud protocol for Scenario 3

Three questions followed the think-aloud protocol. One was rating the trust on a scale from 1 to 5. This question was a hook to ask the second question about why they assigned this rate. The last question was related to the understanding. In previous scenarios, the explanation provided was local, so the participants were asked to describe their understanding of a specific decisional tree. In this case, a global understanding of the machine’s decisional power was the goal to achieve. A summary of the themes that emerged from these three questions is provided below (*Table 12*).

Providing a global explanation is a difficult task. The participants could not give a detailed explanation, but only a general idea, except for one participant who did not understand clearly. Due to the complexity of a global explanation, a general understanding was an acceptable result. The themes that emerged were mostly related to the concepts of control, adaptability, and learnability. Control is related to the feeling of having control over the AI instead of feeling controlled. Adaptability is similar to accessibility: they liked that the machine can be adapted to fit the needs of different types of people. Finally, Learnability was coded when they felt that the machine was teaching them some concept about AI while they were using it, almost like it was conceived with an educational purpose.



Participant	Trust rate	Can you describe why you assigned rate X?	How can you describe your understanding of AI from this explanation?
1	5	Completeness	General understanding
2	5	Adaptability, Control	General understanding
3	5	Adaptability	Wrong mental model
4	5	Control, Learnability	General understanding
5	5	Learnability	General understanding

**Table 12.** Summary of themes emerged during the structured interview for Scenario 3

## 6.5 Scenario 4

At this phase of the interview process, the author read the scenario and the participants were asked to explore the prototype and brew a coffee. They were also asked to express their thoughts by speaking out loud. The transcripts were coded to understand the main themes regarding the first scenario (*Table 13*).

Scenario 4 provided a global explanation. During the think-aloud session, the participants saw a timeline with events happening between the machine and other devices or between the machine and the cloud. To differentiate the type of communication icons were used to help with the interpretation. The visual aspect in the user flow was dominant, so the participants were inclined to focus on two themes: the evaluation, because they felt the need to express if the explanation was clear and easy to understand, and the interpretation of the explanation, with which they attempted to translate the timeline into a mental model. This was useful to understand if their mental model was correct or not. One participant had no idea of how a Cloud works: however, the subject was able to form a general idea of the machine's model.

Theme	Transcript
Evaluation of the explanation	<p>“It’s clear.” - Participant 1</p> <p>“I don’t understand what the minutes are.” - Participant 2</p> <p>“It seems easy to understand even to me that I am not an expert.” - Participant 4</p> <p>“The process to prepare the coffee is clear.” - Participant 5</p>
Interpretation of the explanation	<p>“I see that it took the data and somehow it processed it, it sent it through the cloud and then it gave me the suggestion.” - Participant 1</p> <p>“I see that the device is taking data from the smartphone. What is the cloud? I had no idea.” - Participant 2</p> <p>“What I can see is feedback about interactions between several devices, for example how it interacted with the smartwatch. And it also tells me downloads and uploads of data.” - Participant 3</p> <p>“It makes an evaluation, sends data, analyzes it, makes calculations. It must have this data to decide.” - Participant 4</p> <p>“It needs to download data from other objects, it does one thing at the time and puts all the data together.” - Participant 5</p>

**Table 13.** Thematic analysis of think-aloud protocol for Scenario 4

Three questions followed the think-aloud protocol. One was rating the trust on a scale from 1 to 5. This question was a hook to ask the second question about why they assigned this rate. The last question was related to the understanding. A summary of the themes that emerged from these three questions is provided below (*Table 14*).

Similar to the previous scenario, the participants were able to express out loud how the system works in broad terms. One participant went into more details, such as how downloading and uploading data affected the algorithm. The theme of transparency was added because two

participants said that they could “see what’s actually happening behind” with this explanation. No theme was assigned to answers such as “it’s clear” or “I like it” and to what did not provide valuable insight. The author hypothesizes that despite providing a complex explanation is a difficult task, visual elements such as a timeline and clear icons can help the participants to form a correct mental model.

Participant	Trust rate	Can you describe why you assigned rate X?	How can you describe your understanding of AI from this explanation?
1	5	-	General understanding, Learnability
2	5	Transparency	General understanding
3	5	Transparency	Correct mental model
4	5	Completeness	General understanding
5	4	-	General understanding

**Table 14.** Summary of themes emerged during the structured interview for Scenario 4

## 6.6 Final feedbacks

Final questions were asked to determine if the artifact improved trust in participants, and if the artifact helped them to better understand how the AI inside the coffee machine works. Regarding trust, the table below (*Table 15*) shows that there was indeed an improvement. How significant it was, it will be discussed in the next session. A different theme emerged from each participant. Participant 1 did not have improvements in the understanding, but s/he self-evaluated with the maximum level because of how the explanations were presented and organized. The second participant reflected that new concepts could be learned in interacting with the device, and then they can be applied to other devices. The third participant gained trust in seeing what “it’s happening behind the scene”, and so it was related to the concept of transparency. Participant 4 liked the way the information was presented clearly, and Participant 5 gained trust in discovering

that the user can be in control of the AI, especially with the features presented in the third scenario.

“The way the information is given to me, visually. They were very clear. I trust them more because of how they were explained to me” - Interview participant

“I learned things that I didn’t know before, I learned how the machine works, and I could apply it to other devices.” - Interview participant

“I changed my mind thanks to the idea that I can observe what the machine is doing behind the scenes, and what the reasoning behind was.” - Interview participant

Participant	Final trust rate	If so, what is changed from your initial evaluation?
1	5 (+3)	Clarity of information
2	5 (+2)	The subject discovered notions that can be applied in other devices
3	4,5 (+1,5)	Transparency
4	4 (+1)	Clarity of information
5	4 (+1)	Control

**Table 15.** Summary of themes emerged in the final trust-related questions of the structured interview

Regarding the understanding of the system, the table below (*Table 16*) summarizes the questions about this topic. It was not asked to self-assign a rate of understanding, since the goal was not to measure but to see if there was any improvement and why it occurred.

Participant	Have there been any improvements in your understanding towards AI technology in this type of product?
1	No significant improvement
2	The subject learned new notions
3	No significant improvement
4	The subject learned new notions
5	The subject learned new notions

**Table 16.** Summary of themes emerged in the final knowledge-related questions of the structured interview

## 6.7 Human-Computer Trust Scale

The Human-Computer Trust Scale (*Appendix F*) was chosen to compare the trust level before and after interacting with the prototype. Before interacting, the questionnaire was based on general experiences with AI-based devices. Participants were asked to reflect on their experiences with their smartwatch, smartphone, or any other device equipped with AI.

The results below (*Table 17*) show an average improvement of 18,4% on HCTS. After the interaction, the results of the questionnaire showed that the trust threshold of 75% was met for each candidate except Participant 1, who was the most distrustful before the interview.

<b>Participant</b>	<b>HCTS - Before interacting</b>	<b>HCTS - After interacting</b>	<b>Improvement</b>
1	58%	73%	+15%
2	65%	85%	+20%
3	62%	92%	+30%
4	63%	81%	+18%
5	76%	85%	+9%
<b>Average</b>			+18,4%

**Table 17.** Comparison between the HCTS questionnaires before and after the interaction with the prototype

## 7. Discussion

The qualitative introductory questions revealed different motivations regarding lack of trust towards smart home devices, helping the author to understand the participants' initial background. This knowledge allowed a better qualitative comparison between the initial and the final situation. Participant 1 stated that he did not learn anything new on how these devices work. The lack of trust s/he experienced was due to potentially suspicious activities from the manufacturing company, but s/he pointed out that the way the explanations were organized provided the transparency s/he needed. Lack of trust in Participants 2 and 3 came from the idea of smart devices as something that usually fails and misses the expectation, making the user abandon them and return to non-smart products after a period of frustration. For Participant 2, the improvement of trust happened thanks to the discovery of some mechanisms through the coffee machine that could be applied to many other smart objects, avoiding further frustrations. For Participant 3, who had the same initial motivation, the improvement was due to the ability to “see through” the device. Participant 4 was the only one who stated that the lack of trust was due to the lack of knowledge. At the end of the interview, s/he argued that the clarity of information provided trust and that s/he learned something new, despite feeling unfamiliar with the technology. Participant 5 had a high level of trust in AI regarding high-risk applications like in the medical field, but s/he lacked trust when AI is implemented in everyday objects, probably perceived as more fallible. This reasoning is typical of those who perceive Artificial Intelligence as something that necessarily takes all the decisions on the user's behalf, but s/he learned that it could be controlled and understood as demonstrated by the final section of the qualitative interview (*Table 15*).

The interview continued by presenting four scenarios corresponding to the four user flows designed. The first two contained local explanations, one considered having a high impact on the user's routine, and the other considered having a low impact. The last two included global explanations. Comparing the results from the two local explanations, we don't see significant differences: both revealed that users were likely to trust the information because information reflected what they experienced in the fictional world, together with clear and complete information. The only negative results came from Participant 1, who did not like to have an

explanation in Scenario 2. That was not surprising at all: low impact explanations sometimes are not required; however, every other participant was satisfied. In both scenarios, the participants also noted that they could adjust the initial suggestion from the machine, which gave them trust. In this case, the way the GUI was designed was responsible for this sense of freedom that increased trust over the device.

The remaining two scenarios contained global explanations. The goal was to explain how the machine works overall instead of explaining a single event. In terms of trust, both global explanation scenarios were rated with the participants' maximum score, but we can see some differences in the reasons. The first global scenario let the users play around a set of options that allowed them to control the AI, and the feeling to have control was one of the main reasons that helped them to gain trust. The last scenario presented a timeline of the events occurring behind the interface, so transparency was the driving factor.

In terms of understanding, we can look at the comparison between local explanations and global explanations. The first ones performed better in giving the best understanding; decisional models and in few cases also decisional trees were recognized, while the latter provided just a general understanding. In one case (*Table 12*), the user did not understand how the system worked. However, we must consider that the base knowledge of these systems by the participant was shallow. As demonstrated in the final results, the level of interpretability provided by the prototype can be considered appropriate. Attempting to explain more than required, in fact, can lead to the opposite effect, as one participant noted.

“Explained this way I can understand, if I had to read more I could not do it.” - Interview participant

The final part of the interview showed a significant improvement in both trust and interpretability, and it was useful to understand what motivations led the participants to describe their progress. Five participants showed significant improvements in their trust, but two participants mentioned that they did not improve their level of knowledge of these systems. This



indicates that explanations can help users to build trust, but not always lack of trust is due to lack of knowledge.

In *Table 17*, we already saw a comparison between the HCTS questionnaire taken before interacting with the prototype and after the interaction, demonstrating an average increase in trust by a substantial 18,4%. The limitations of this comparison may consist in the fact that before the interaction users tried to relate to smart objects they used in their daily lives, and not to a smart coffee machine. So it's likely that each participant answered the questionnaire with a different smart device in mind.

## 8. Conclusions

The goal of this study was to present design recommendations to support practitioners in the task of designing for IoT. To meet the intent of the study, the author wanted to find out:

**Question 1.** How can we design connected objects that communicate their autonomous decisions transparently with users in the context of a smart home?

- *Sub-question 1.1.* How can explanations be designed to be accurate and complete without overloading the user's attention?
- *Sub-question 1.2.* How and where can transparency be integrated into the object?
- *Sub-question 1.3.* To what extent does the artifact help users in interpreting the AI model?
- *Sub-question 1.4.* How do users describe their experience with the artifact in terms of trust?

The next section will describe how the sub-questions and the main question have been addressed, strongly indicating the validity of the design process in the given scenario.

### 8.1 Sub-question 1.1

The entire design process started with a Things-Centered Design approach that led to an innovative design concept of a coffee machine. Later, four scenarios have been created, used as the base to build the explanations. They were needed to narrow the design to a limited set of functionalities and show a range of potentially different types of benefits. In fact, some of them have been described as high or low-impact scenarios, other than local or global. The four described scenarios represented a small but representative range of use cases that generated different types of explanations to be studied. The resulting sets of explanations were then selected and adjusted based on the results of a survey.

Overwhelming users with unnecessary information can be an issue for design practitioners. The final results of the study showed that the explanations provided the required amount of

information the users needed to improve their understanding and trust towards the prototype. Building the scenarios and decisional trees has proven to be effective. In particular, decisional trees were useful for local explanations, while global explanations relied on design ideas based on the author's personal working experience. At the end of this phase, the survey helped the author to choose the preferred amount of information for local explanations, and to validate the design ideas for global explanations. It provided fundamental guidance for the next phase of prototyping.

## **8.2 Sub-question 1.2**

After the creation of explanations, they had to be incorporated into the design of the artifact. To understand how and where transparency can be integrated into the object, the author used a Research-through design approach by building a fictional prototype. Building the prototype, as the final results showed, let the author attempt design solutions that, in the end, were successful in providing transparency and trust. By applying the basic principles of a good User Experience, the suggestions were not invasive and users felt they had the freedom to adjust them to their needs. A second iteration of the prototype was not necessary.

Design fiction also played a role. It could be time-consuming for design practitioners to build a functioning prototype to test if explanations provide trust. This is when low fidelity prototypes usually come into play, which can be considered normal practice in the design field. However, when practitioners design for smart homes, they may face issues in providing a realistic experience to be tested before building a fully functioning artifact. This study suggests that, when possible, a functioning prototype could be simulated by integrating a graphical interface into the physical object, using rapid prototype software with smartphones or iPads. World Building Design Fiction has also proven to be very helpful in these tasks, because it helps the users to immerse themselves in the environment in which the object will be used. It is particularly effective when designers are exploring problems of an incoming future. During the interview process, feedback on the prototype emerged.

“The prototype is very clear. Even myself, that I am not familiar with these things, I was able to understand” - Interview participant

### **8.3 Sub-question 1.3**

Interpretability in the XAI literature is defined as “the ability to explain or provide the meaning in understandable terms to humans” (Doshi-Velez & Kim, 2017). This concept has been qualitatively evaluated through the structured interview process. The author asked the participant to describe their understanding of AI through the interaction with the prototype. In the case of local explanations, a match from the designed decisional tree and the description given by the participants was looked to evaluate interpretability. For global explanations, a general understanding of the system was the goal that the prototype aimed for. From the results, we can see that some participants did not always improve their understanding, but they were able to interpret how the machine arrived at the solution. Less skilled participants also mentioned that they learned new notions about AI by interacting with the artifact, demonstrating that the level of interpretability was acceptable.

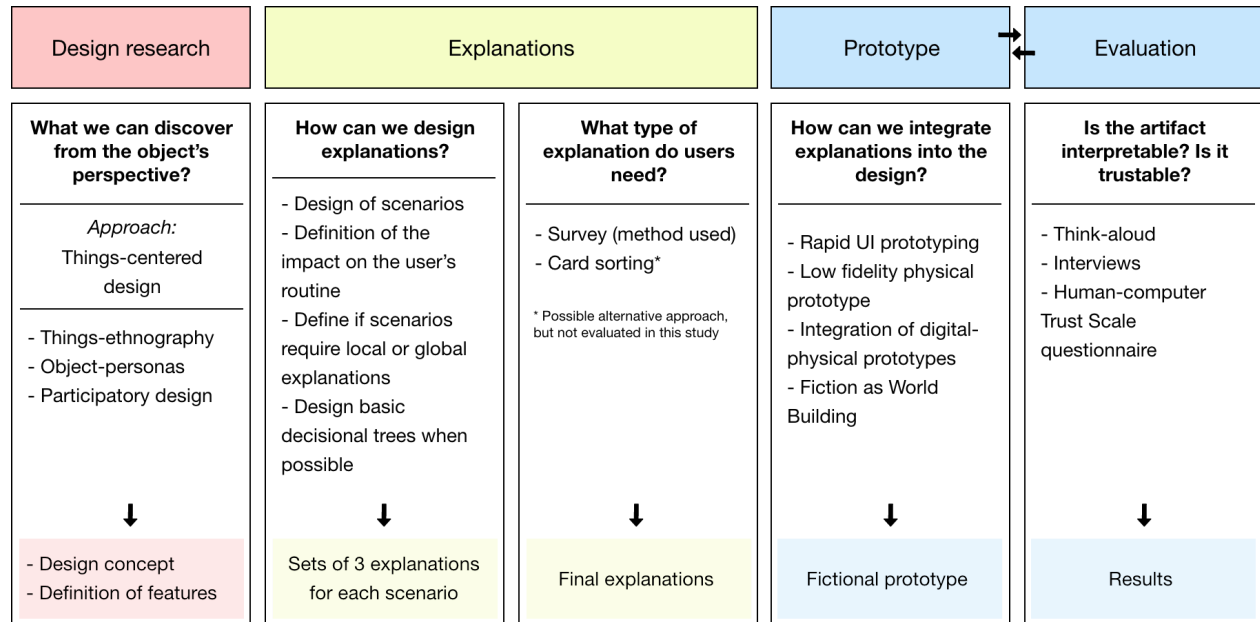
### **8.4 Sub-question 1.4**

During the interview process, the author has met with five respondents of the first survey, selected for their lack of trust towards AI in smart home devices. A comparison has been made between the HCTS questionnaires taken before interacting with the prototype and after the interaction, demonstrating an average increase in trust by a substantial 18,4%. These results have exceeded the expectations, indicating that the design process has reached the final goal of designing smart home devices with transparency, avoiding the black box problem.

Since all the questions of the study have been addressed, the next section will describe the rationale of the design process, and design recommendations to support practitioners in the task of designing for IoT will be provided.

## 8.5 Design recommendations

This section will summarize the design process developed in the study in the figure below (Figure 17). Finally, the process will be described concurrently with design recommendations.



**Figure 17.** Final design process

### Design research

The design process started with an observation from the point of view of a coffee machine, using a Things-Centered Design approach. The tools used for the observational study were thing-ethnography, developed by Giaccardi et al. (2016) that involves the use of cameras and sensors attached to objects to capture the behavioral patterns, temporal routines, and spatial movements of objects. The thing-ethnography session conducted in this study adapted the work of Giaccardi et al. (2016) and consisted of collecting video material through a single point of view: an action camera attached to a standard coffee machine. Later, a participatory design method was used with five designers to create things-personas, which were then used to define the product's features. Based on the findings of the study, the following design recommendations are provided:

1. **Things-Centered Design** tools can help the designer in the research process, observing the interactions between humans and non-humans from a different point of view. This may prevent the designer from avoiding the black box issue in the first place, which may probably occur with a user-centered approach. Tools can be adapted to the specific scenario.

## Explanations

The next step consisted of creating the explanations. First, a set of scenarios was chosen based on the features defined in the first phase. Based on the findings of the study, the following design recommendations are provided:

2. **Defining the impact of each scenario** can help designers in generating explanations. Some scenarios can have a high impact on the user's routine; others can have less impact. Explanations are not always required, and when required, they may be presented with different levels of detail depending on the benefits perceived by users.
3. **Determining if each scenario contains local or global explanations** can help define how the explanations will be presented. For example, in this study, local explanations were text-based explanations in the main screen of the interface, while global explanations were visually more elaborated and situated inside a menu.
4. When possible, designing **decisional trees** of the most representative use cases can help to form a correct mental model and develop understandable sets of explanations.
5. **Investigating through a survey** with potential users which level of details is preferred for each scenario can help the designer to reduce errors and iterations on the prototype. Since the goal of the survey is to understand the user's preferences, an alternative method could be card sorting. However, it has not been validated in this study.

## Prototype

The next step consisted of creating the prototype. The study used the Design Fiction as World Building approach coined by Coulton et al. (2017). The prototype was not meant to be functional, but its purpose was to induce people to think critically about issues that the design

embodies (Coulton et al., 2017). Based on the findings of the study, the following design recommendations are provided:

6. **Rapid digital prototyping** can help designers to rapidly design user flows containing explanations.
7. **Rapid physical prototyping** allows designers to integrate the digital prototype. For a fast and effective integration, it is possible to combine the two parts through a case that supports the device for which the digital prototype was designed.
8. **World Building as Design Fiction** allows the users to feel immersed in the specific scenario of use of the product. By building a fictional but credible world around the product, designers can evaluate the interaction with an artifact perceived as fully functional, but saving time and resources.

## Evaluation

The final step consisted of evaluating the prototype. For the scope of this study, the evaluation has been made on interpretability and trust provided by the artifact. Different methods, qualitative and quantitative, were combined to address the research questions. Based on the findings of the study, the following design recommendations are provided:

9. **Investigating interpretability and trust** of the artifact through qualitative methods, such as the Think-aloud protocol, can be useful if designers want to intercept users' feelings and impressions. It is also a "quick and dirty" tool for understanding usability issues. Another qualitative method is the one-on-one interview. A structured interview has been conducted in this study, but a single semi-structured interview may possibly substitute the think-aloud and the structured interview.
10. **Measuring trust** through the Human-computer questionnaire can help designers to evaluate if the prototype has chances to be well accepted and adopted in the market. This tool can be used in different ways, depending on the goal the designers want to achieve. In this study, the evaluation consisted of measuring the trust score regarding other devices and comparing it with the trust score obtained from the interaction with the prototype. However, it can be used to compare different versions of the prototype and guide

designers towards the next phases of the product development until the functioning prototype. It can also be used to compare the trust score with a benchmark or competitor, or simply to make sure that the prototype has reached the trust threshold of 75%.

## **8.6 Limitations of the study**

Explainable AI lessons and Things-Centered Design tools presented in this process came from prior research in respective fields, so we can assume that this process is applicable to other similar scenarios involving the Internet of Things. However, with the fast-evolving nature of IoT, it is also possible that other variants of the process may be better suited for different specific scenarios, and this may need further evaluation. The author acknowledges that the presented process may not be the only way to design for transparency for IoT; however, it may represent a concrete starting point to help design practitioners in this complex task.

Another limitation of this study is represented by the period in which the prototype has been tested. The initial trust evaluation before interacting with the prototype took in account previous experiences of users with smart devices, which may have been evaluated through the lenses of a constant relationship with the objects. On the other hand, the questionnaire presented after the interaction did not take into consideration the experience with the product over time.



## 9. References

- ACM. (2017, May 25). *USACM - EUACM Statement on Algorithmic Accountability*. Acn.Org. <https://www.acm.org/media-center/2017/may/usacm-euacm-joint-statement-on-algorithmic-accountability>
- Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, 6, 52138–52160. <https://doi.org/10.1109/access.2018.2870052>
- Bathae, Y. (2018). The Artificial Intelligence Black Box and the Failure of Intent and Causation. *Harvard Journal of Law & Technology*, 31(2), 929. <https://jolt.law.harvard.edu/assets/articlePDFs/v31/The-Artificial-Intelligence-Black-Box-and-the-Failure-of-Intent-and-Causation-Yavar-Bathae.pdf>
- Cila, N., Giaccardi, E., Tynan-O'Mahony, F., Speed, C., & Caldwell, M. (2015). Thing-centered narratives: A study of object personas.
- Cisco Annual Internet Report (2018–2023) White Paper*. (2020, March 10). Cisco. <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>
- Coulton, P., Lindley, J., Sturdee, M., & Stead, M. (2017). Design Fiction as World Building. *Proceedings of the 3rd Biennial Research Through Design Conference*, 163–179. <https://doi.org/10.6084/m9.figshare.4746964>
- Cramer, H., Evers, V., Ramlal, S., van Someren, M., Rutledge, L., Stash, N., Aroyo, L., & Wielinga, B. (2008). The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-Adapted Interaction*, 18(5), 455–496. <https://doi.org/10.1007/s11257-008-9051-3>

- Doshi-Velez, F., & Kim, B. (2017). Towards A Rigorous Science of Interpretable Machine Learning. *ArXiv:1702.08608 [Cs, Stat]*. <http://arxiv.org/abs/1702.08608>
- Dove, G., Halskov, K., Forlizzi, J., & Zimmerman, J. (2017). UX Design Innovation. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 278–288. <https://doi.org/10.1145/3025453.3025739>
- Eiband, M., Schneider, H., Bilandzic, M., Fazekas-Con, J., Haug, M., & Hussmann, H. (2018). Bringing Transparency Design into Practice. *23rd International Conference on Intelligent User Interfaces*, 211–223. <https://doi.org/10.1145/3172944.3172961>
- Giaccardi, E. (2018). Things Making Things: Designing the Internet of Reinvented Things. *IEEE Pervasive Computing*, 17(3), 70–72. <https://doi.org/10.1109/mprv.2018.03367737>
- Giaccardi, E., Cila, N., Speed, C., & Caldwell, M. (2016). Thing Ethnography. *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, 377–387. <https://doi.org/10.1145/2901790.2901905>
- Graaf, M. M. A. de, & Malle, B. F. (2017, October 9). How People Explain Action (and Autonomous Intelligent Systems Should Too). *2017 AAAI Fall Symposium Series*. 2017 AAAI Fall Symposium Series. <https://www.aaai.org/ocs/index.php/FSS/FSS17/paper/view/16009>
- Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2019). A Survey of Methods for Explaining Black Box Models. *ACM Computing Surveys*, 51(5), 1–42. <https://doi.org/10.1145/3236009>
- Guimarães Pereira, Â., Benessia, A., Curvelo, P., European Commission, Joint Research Centre, & Institute for the Protection and the Security of the Citizen. (2013). *Agency in the Internet of things*. Publications Office. <http://dx.publications.europa.eu/10.2788/59674>

- Gulati, S., Sousa, S., & Lamas, D. (2019). Design, development and evaluation of a human-computer trust scale. *Behaviour & Information Technology*, 38(10), 1004–1015. <https://doi.org/10.1080/0144929x.2019.1656779>
- Hind, M. (2019). Explaining explainable AI. *XRDS: Crossroads, The ACM Magazine for Students*, 25(3), 16–19. <https://doi.org/10.1145/3313096>
- Internet of Things Privacy Forum. (2018). *Clearly Opaque: Privacy Risks of the Internet of Things*. <https://www.iotprivacyforum.org/clearlyopaque/>
- Kulesza, T., Stumpf, S., Burnett, M., Yang, S., Kwan, I., & Wong, W.-K. (2013). Too much, too little, or just right? Ways explanations impact end users' mental models. *2013 IEEE Symposium on Visual Languages and Human Centric Computing*, 3–10. <https://doi.org/10.1109/vlhcc.2013.6645235>
- Lepekhn, A., Borremans, A., Ilin, I., & Jantunen, S. (2019). A Systematic Mapping Study on Internet of Things Challenges. *2019 IEEE/ACM 1st International Workshop on Software Engineering Research & Practices for the Internet of Things (SERP4IoT)*, 9–16. <https://doi.org/10.1109/serp4iot.2019.00009>
- Liao, Q. V., Singh, M., Zhang, Y., & Bellamy, R. K. E. (2020). Introduction to Explainable AI. *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–4. <https://doi.org/10.1145/3334480.3375044>
- Lim, B. Y., & Dey, A. K. (2010). Toolkit to support intelligibility in context-aware applications. *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*, 13–22. <https://doi.org/10.1145/1864349.1864353>
- Lim, B. Y., & Dey, A. K. (2011). Design of an intelligible mobile context-aware application. *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services - MobileHCI '11*, 157–166. <https://doi.org/10.1145/2037373.2037399>

Lim, B. Y., Dey, A. K., & Avrahami, D. (2009). Why and why not explanations improve the intelligibility of context-aware intelligent systems. *Proceedings of the 27th International Conference on Human Factors in Computing Systems - CHI 09*, 2119–2128.

<https://doi.org/10.1145/1518701.1519023>

Lindley, J., Coulton, P., & Cooper, R. (2017). Why the Internet of Things needs Object Orientated Ontology. *The Design Journal*, 20(sup1), S2846–S2857.

<https://doi.org/10.1080/14606925.2017.1352796>

Madumal, P. (2019). Explainable Agency in Intelligent Agents: Doctoral Consortium. *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, 2432–2434.

Norman, D. A. (2005). Human-centered design considered harmful. *Interactions*, 12(4), 14–19.

<https://doi.org/10.1145/1070960.1070976>

Pasquale, F. (2016). *The Black Box Society: The Secret Algorithms That Control Money and Information* (Reprint ed.). Harvard University Press.

Rozendaal, M. (2016). Objects with intent. *Interactions*, 23(3), 62–65.

<https://doi.org/10.1145/2911330>

Sarma, S., Brock, D. L., & Ashton, K. (2000). The networked physical world. *Auto-ID Center White Paper MIT-AUTOID-WH-001*.

Stappers, P. J. (2007). Doing Design as a Part of Doing Research. *Design Research Now*, 81–91.

[https://doi.org/10.1007/978-3-7643-8472-2\\_6](https://doi.org/10.1007/978-3-7643-8472-2_6)

Stappers, P. J., & Giaccardi, E. (n.d.). *Research through Design*. The Interaction Design Foundation.

<https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/research-through-design>

Suresh, P., Daniel, J. V., Parthasarathy, V., & Aswathy, R. H. (2014). A state of the art review on the Internet of Things (IoT) history, technology and fields of deployment. *2014 International Conference on Science Engineering and Management Research (ICSEMR)*, 1–8.  
<https://doi.org/10.1109/icsemr.2014.7043637>

van Allen, P., McVeigh-Schultz, J., Brown, B., Kim, H. M., & Lara, D. (2013). AniThings. *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13*, 2247–2256.  
<https://doi.org/10.1145/2468356.2468746>

Wachter, S., Mittelstadt, B., & Russell, C. (2018). Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR. *ArXiv:1711.00399 [Cs]*.  
<http://arxiv.org/abs/1711.00399>

Zimmerman, J., Stolterman, E., & Forlizzi, J. (2010). An analysis and critique of Research through Design. *Proceedings of the 8th ACM Conference on Designing Interactive Systems - DIS '10*, 310–319. <https://doi.org/10.1145/1858171.1858228>

# Appendix

## A. Object-personas



Name  
-----  
Participant n.  
-----

This sheet is meant for generating a **product persona**. There are some psychological, emotional, social and cultural elements to consider to help building this persona.

Please feel free to leave the ones that you think are irrelevant blank.

1

### 1. Day-in-life

Follow the coffee machine through a typical day



### 2. Inner life

Depict machine's psychological profile

Personality

Attitude towards life

Temperament/mood

Needs

Likes and dislikes

Desires

Frustrations

Fears

Skills and abilities

Ambitions


Habits

What would be the ideal life  
of the coffee machine?

3. Social relationships

2

Depict the machine's social life

<p>How is the social structure in the home?</p> 	<p>Friends</p>	<p>Enemies</p>
---	----------------	----------------

What would the coffe machine talk about...      What would the coffe machine learn from?      What would the coffe machine teach to

- with _____ ?	
- with _____ ?	
- with _____ ?	
<p>How is the machine's relationship with its owner?</p>	<p>How would you describe it with a metaphor?</p>

4. Life-cycle

Depict the machine's past and futures

<p>What kind of a past the machine might have had?</p>	<p>What would the machine learn from its past?</p>
<p>Transformations the machine might have had?</p>	<p>What kind of dreams the machine may have for the future?</p>

## B. Online Survey (Italian)

Come studente di tesi presso la Cyprus University of Technology sto conducendo una user research preliminare per uno studio di tesi sull'intelligenza artificiale spiegabile.

Molti studi hanno evidenziato che gli oggetti dotati di intelligenza non riescono a trasmettere fiducia nell'utente, questo per un fenomeno definito "black box". L'utilizzatore, non comprendendo le decisioni che prende l'algoritmo, tende a non fidarsi della tecnologia.

La mia tesi propone di disegnare una macchina da caffè smart che integra un'intelligenza artificiale spiegabile. Per riuscire in questo obiettivo, sono interessato a capire quale tipo di spiegazione trasmette più fiducia rispetto ad un'altra, e qual è il livello di dettaglio desiderabile. Vi verranno proposti degli scenari di utilizzo di una macchina da caffè intelligente e potrete scegliere tra diverse opzioni proposte.

Le informazioni che fornirete saranno ritenute strettamente riservate e usate al solo scopo dello studio. Sarà inoltre preservata la vostra anonimità.

Grazie del vostro tempo.

### B.1 Demographic questions

La tua età:

- 18-25
- 26-35
- 36-45
- +45

Come definiresti il tuo livello di fiducia verso i device che usano algoritmi per proporti suggerimenti adatti a te?

- 1 - Non mi trasmettono nessuna fiducia



- 2
- 3
- 4
- 5 - Mi ispirano massima fiducia

Come definiresti il tuo livello di comprensione del funzionamento degli algoritmi che danno suggerimenti o agiscono autonomamente?

- 1 - Non so come funzionano
- 2
- 3
- 4
- 5 - Conosco perfettamente il funzionamento

Ritieni che se l'intelligenza artificiale ci fornisse delle spiegazioni sul suo funzionamento la tua fiducia nei confronti di questa tecnologia migliorerebbe?

- Sì
- No

## **B.2 Scenario 1**

Ti sei svegliato in ritardo questa mattina. Il tuo bracciale smart rivela che hai avuto un sonno irregolare. Hai una giornata molto impegnata davanti e sei di fretta, ma non rinunci ad un caffè per darti un po' di carica.

La macchina del caffè comincia di diffondere un aroma di caffè per invogliarti ad alzarli. Una volta arrivato davanti alla macchina, questa ti propone un caffè espresso con massima intensità.

Scegli la spiegazione che ritieni più appropriata:

- Considerando il tuo sonno, il tuo stato di salute e i tuoi impegni, ti suggerisco: espresso, intensità 10/10

- Ho rilevato sonno irregolare, e un'agenda impegnata che comincia tra 15 minuti. Ti suggerisco: espresso, intensità 10/10
- Sonno: irregolare; Agenda: piena; Primo impegno: tra 15 minuti. Suggerimento: espresso, intensità 10/10

Scegli il livello di fiducia che ti trasmette la spiegazione che hai scelto:

- 1 - Non mi trasmette nessuna fiducia
- 2
- 3
- 4
- 5 - Mi ispira massima fiducia

C'è qualcosa che farebbe aumentare il tuo livello di fiducia su quel suggerimento? \*facoltativo

Quanto ti soddisfa la spiegazione scelta in merito al funzionamento dell'algoritmo?

- 1 - Non mi è chiaro come funziona
- 2
- 3
- 4
- 5 - Ho capito perfettamente

Avresti altre domande su come l'algoritmo sia arrivato a darti quel suggerimento? \*facoltativo

## **B.3 Scenario 2**

È una giornata fredda stai studiando o lavorando da casa con un App per la concentrazione. Hai quasi finito il tempo di concentrazione e senti un aroma di caffè nella stanza. Ti avvicini alla macchina e questa ti propone un caffè lungo con un'intensità di 4/10.

Scegli la spiegazione che ritieni più appropriata:

- Ti suggerisco: caffè lungo, intensità 4/10
- Il caffè perfetto per una pausa in una giornata piovosa: caffè lungo, intensità 4/10
- Pausa caffè rilevata dalla tua agenda. Dati il meteo e la tua playlist, ti suggerisco: caffè lungo, intensità 4/10

Scegli il livello di fiducia che ti trasmette la spiegazione che hai scelto:

- 1 - Non mi trasmette nessuna fiducia
- 2
- 3
- 4
- 5 - Mi ispira massima fiducia

C'è qualcosa che farebbe aumentare il tuo livello di fiducia su quel suggerimento? \*facoltativo

Quanto ti soddisfa la spiegazione scelta in merito al funzionamento dell'algoritmo?

- 1 - Non mi è chiaro come funziona
- 2
- 3
- 4
- 5 - Ho capito perfettamente

Avresti altre domande su come l'algoritmo sia arrivato a darti quel suggerimento? \*facoltativo

## B.4 Scenario 3

La macchina ti suggerisce il miglior caffè a seconda dei tuoi impegni, della tua salute, del tuo stato d'animo e altri dati. Puoi accettare il suggerimento o ignorarlo. La macchina poi può continuare a darti quei suggerimenti oppure diventare meno proattiva fino a memorizzare semplicemente le tue abitudini. Hai notato che qualche volta i suggerimenti sono molto diversi dalle tue abitudini e a volte preferisci ignorarli. A questo punto vorresti avere un modo per cambiare il comportamento della macchina per adattarlo meglio al tuo utilizzo.

Scegli la spiegazione che ritieni più appropriata:

- Nel pannello delle impostazioni, hai la possibilità di controllare il livello di proattività. Se si imposta al livello massimo, la macchina suggerirà ciò che pensa sia l'opzione migliore, al livello più basso imparerà a memoria la tua routine proponendoti solo ciò che prendi di solito
- Nel pannello delle impostazioni, hai la possibilità di controllare il livello di proattività. Ad ogni livello selezionato potrai vedere un esempio di come cambierebbero i tuoi suggerimenti
- Direttamente nel pannello di preparazione del caffè, ti sono proposte direttamente due gruppi di opzioni: i suggerimenti della macchina, e i suggerimenti basati sul tuo utilizzo

Ritieni che avere il controllo sul potere decisionale della macchina possa far aumentare il tuo livello di fiducia sull'intelligenza artificiale?

- Sì
- No

Ritieni che avere il controllo sul potere decisionale della macchina possa far aumentare il tuo livello di comprensione del funzionamento dell'algoritmo?

- Sì
- No

Avresti altre domande su come funziona la proattività dell'algoritmo? \*facoltativo

## **B.5 Scenario 4**

La macchina scambia continuamente dei dati. Lo fa con un altro device all'interno del network locale, oppure può inviare e ricevere dati dal cloud. Dopo aver preparato il caffè, vuoi scoprire come la macchina sia arrivata a darti quel suggerimento.

Scegli la spiegazione che ritieni più appropriata:

- Hai la possibilità di vedere una linea temporale con tutti gli scambi dati avvenuti, segnalando quelli in uscita e in entrata, e differenziando quelli che avvenuti all'interno della tua casa oppure nel cloud
- Linea temporale come sopra, ma con in più la possibilità di vedere le comunicazioni nel dettaglio tra gli oggetti della casa
- Ti vengono indicate le transazioni, oltre che nella linea temporale, anche nella schermata di preparazione del caffè, con delle piccole icone che indicano che tipo di scambio dati sta avvenendo in quel momento

Ritieni che avere la possibilità di vedere quali dati vengono usati dalla macchina possa far aumentare il tuo livello di fiducia sull'intelligenza artificiale?

- Si
- No

Ritieni che avere la possibilità di vedere quali dati vengono usati dalla macchina possa far aumentare il tuo livello di comprensione sul suo funzionamento?

- Si
- No

## **C. Online Survey (English)**

As a student at Cyprus University of Technology I'm conducting a preliminary user research on explainability and transparency in smart devices.

Many studies pointed out that AI-based devices often fail in giving trust: this phenomenon is called "black box". The user, when does not understand how the system operates, loses trust in this technology.

My study consists in designing a smart coffee machine that integrates Explainable Artificial Intelligence. To achieve this goal, I'm interested in understanding what explanation is the most

reliable to you, and what is the preferred level of detail. Four scenarios will be presented to you and you can choose the options that you prefer.

The information you provide is strictly confidential and is used only for the purposes of this study. You will be anonymous and will never be identified in any correspondence or reports.

Thank you very much for your time and support.

## **C.1 Demographic questions**

Your age:

- 18-25
- 26-35
- 36-45
- +45

How would you rate the level of trust of smart devices that use algorithms to give you suggestions or operate on your behalf?

- 1 - I trust them
- 2
- 3
- 4
- 5 - I'm not trusting

How would you rate the level of understanding of smart devices that use algorithms to give you suggestions or operate on your behalf?

- 1 - I have no idea how they work
- 2
- 3
- 4

- 5 - I know exactly how they work

Do you believe that providing explanations could help you improve your trust towards these devices?

- Yes
- No

## C.2 Scenario 1

You woke up late this morning because you fell asleep very late. Your health monitoring device reveals a normal heartbeat rate. You have a busy schedule today and it's about to start soon.

Choose the explanation that you consider the most appropriate:

- Based on your sleep, health and schedule: espresso 10/10
- Since you experienced irregular sleep, a busy schedule and first event in 15 minutes: espresso 10/10
- Sleep: irregular; schedule: busy; first event: 15 minutes. Suggestion: espresso 10/10

How would you rate the level of trust of the explanation you chose?

- 1 - The explanation gives me trust
- 2
- 3
- 4
- 5 - I'm not trusting this explanation

Is there something that can help you improve your trust?\*optional

How would you describe your understanding of the decisions the algorithms took?

- 1 - I have no idea how it works
- 2
- 3

- 4
- 5 - I totally understand

Do you have any questions on how the algorithm was able to give you the suggestion? \*optional

### **C.3 Scenario 2**

You are focusing using a Pomodoro app and you have almost completed your focus-time for your task. The machine wants you to take a break to improve your focus during the next session. Also, the weather outside is cold and rainy.

Choose the explanation that you consider the most appropriate:

- I recommend: long coffee 4/10
- The perfect coffee for a study break: long cup 4/10
- Coffee break of 10 minutes detected. Cold weather outside. Suggestion: long cup 4/10

How would you rate the level of trust of the explanation you chose?

- 1 - The explanation gives me trust
- 2
- 3
- 4
- 5 - I'm not trusting this explanation

Is there something that can help you improve your trust?\*optional

How would you describe your understanding of the decisions the algorithms took?

- 1 - I have no idea how it works
- 2
- 3
- 4



- 5 - I totally understand

Do you have any questions on how the algorithm was able to give you the suggestion? \*optional

## C.4 Scenario 3

The machine suggests the best coffee based on your schedule, health, and other data. You could accept the suggestion or ignore them. The device can then continue giving these suggestions or become less proactive until it just learns your habits. You have noticed that sometimes the suggestions are very different from your habits and sometimes you prefer to ignore them. You want to see if there is a way to adapt the machine's behavior to your usage.

Choose the explanation that you consider the most appropriate:

- From the menu section, you can change the level of proactivity. At maximum level, the algorithm of the machine will try to motivate you to change behaviour, at the minimum level it will simply adapt to your routine
- From the menu section, you can change the level of proactivity. At the maximum level, the algorithm of the machine will try to motivate you to change behaviour, at the minimum level it will simply adapt to your routine. For each level there is an example on how the choice will influence the recommendations
- Directly from the main page of coffee selection, you can see your usual choice next to the best option suggested by the coffee machine

Do you believe that having control over the decision-making power of the machine could help you improve your trust towards the AI?

- Yes
- No

Do you believe that having control over the decision-making power of the machine could help you improve your understanding of the system?

- Yes
- No

Do you have any questions concerning the proactivity of the machine?\*optional

## **C.5 Scenario 4**

The machine exchanges data. It does this with another device in the local network, or can send or receive data from the cloud. After preparing your coffee, you want to discover how the machine gave you a suggestion.

Choose the explanation that you consider the most appropriate:

- Timeline in which you can see the data transactions between the coffee machine and other devices or cloud
- Timeline in which you can see the data transactions between the coffee machine and other devices or cloud. You can also expand each transaction to see more details
- Besides the timeline, you can also find data usage directly on the coffee selection page, with little icons that inform you about what type of data transaction is happening in the background

Do you believe that seeing what type of data is being used by the machine could help you improve your trust towards the AI?

- Yes
- No

Do you believe that seeing what type of data is being used by the machine could help you improve your understanding of the system?

- Yes
- No

## D. Interview protocol (Italian)

Come studente di tesi presso la Cyprus University of Technology sto conducendo uno studio di tesi sull'intelligenza artificiale spiegabile. Chiedo quindi il tuo permesso di intervistare e registrare.

Le informazioni che fornirai saranno ritenute strettamente riservate e usate al solo scopo dello studio. Sarà inoltre preservata la tua anonimità.

Grazie del contributo.

\_\_\_\_\_  
Date

### C.1 Introductory questions

Nome	
------	--

Età	
-----	--

Come valuteresti il tuo livello di fiducia in una scala da 1 a 5?	
---	--

Puoi descrivere il motivo del voto?	
-------------------------------------	--

Come descriveresti il tuo livello di comprensione degli algoritmi che regolano il comportamento dei device smart?	
---	--

## D.2 Tasks

Think aloud

Come valuteresti il tuo livello di fiducia in una scala da 1 a 5 rispetto a questa spiegazione?	
---	--

Puoi descrivere il motivo?	
----------------------------	--

Come descriveresti il tuo livello di comprensione del comportamento della macchina attraverso la spiegazione fornita?	
---	--

## D.3 Final feedback

Hai riscontrato dei miglioramenti riguardo alla fiducia nei confronti dell'AI?	
--	--

Se sì, cosa è cambiato dalla tua valutazione iniziale?	
--	--

Come valuteresti adesso il tuo livello di fiducia in una scala da 1 a 5?	
--	--

Hai riscontrato dei miglioramenti riguardo tuo livello di comprensione degli algoritmi che regolano il comportamento della macchina?	
--	--

Ci sono delle osservazioni che vuoi fare in conclusione?	
--	--

## E. Interview protocol (English)

As a student at Cyprus University of Technology I'm conducting a study on explainability and transparency in smart devices. Therefore, I'm requesting your permission to observe, interview and record.

The information you provide is strictly confidential and is used only for the purposes of this study. You will be anonymous and will never be identified in any correspondence or reports.

Thank you very much for your time and support.

\_\_\_\_\_  
Date

### E.1 Introductory questions

Name	
------	--

Age	
-----	--

How would you evaluate your level of trust of AI-based products on a scale from 1 to 5?	
---	--

Can you describe why you feel this way?	
---	--

Can you describe your level of knowledge about the algorithms with which we interact every day?	
---	--

## E.2 Tasks

Think aloud	

Rate the level of trust that this explanation gave to you on a scale from 1 to 5	
--	--

Can you explain why?	
----------------------	--

How can you describe your understanding of AI from this explanation?	
--	--

## E.3 Final feedback

Have there been any improvements in your feeling of trust towards AI technology in this type of product?	
--	--

If so, what is changed from your initial evaluation?	
--	--

How would you rate your trust now on a scale from 1 to 5?	
---	--

Have there been any improvements in your understanding towards AI technology in this type of product?	
---	--

Do you have any concerns or feedback that you would like to share?	
--	--

## F. Human-Computer Trust Scale questionnaire

Answers consist of a scale from 1 to 5.

<b>RP1:</b> I believe that there could be negative consequences when using X	
<b>RP2:</b> I feel I must be cautious when using X	
<b>RP3:</b> It is risky to interact with X	
<b>Ben1:</b> I believe that X will act in my best interest	
<b>Ben2:</b> I believe that X will do its best to help me if I need help	
<b>Ben3:</b> I believe that X is interested in understanding my needs and preferences	
<b>COM1:</b> I think that X is competent and effective in supporting me in everyday tasks	
<b>COM2:</b> I think that X performs its role in supporting me in everyday tasks	
<b>COM3:</b> I believe that X has all the functionalities I would expect from	
<b>GT1:</b> If I use X, I think I would be able to depend on it completely	
<b>GT2:</b> I can always rely on X for obtaining help in daily activities	
<b>GT3:</b> I can trust the information presented to me by X	