Pedro Costa [1]
pedro.carv.c@gmail.com

Luísa Ribas [1,2]
l.ribas@belasartes.ulisboa.pt

[1] Faculty of Fine Arts, University of Lisbon
Portugal

[2] CIEBA

Conversations with ELIZA: on Gender and Artificial Intelligence

Keywords: Artificial Intelligence; Chatbots;
Anthropomorphization; Gender; Femininity;
Stereotypes.

Abstract:

This paper aims to explore the relationship between gender and artificial intelligence. It begins by addressing AI and the possibilities and questions that emerge with its evolution and integration in daily life. It then examines gender in light of a binary framework, in order to understand its role in social, cultural and work related contexts. These topics are then related, seeking to understand how and why chatbots and digital assistants such as Siri, Alexa or Cortana tend to display feminine attributes. Complementing this discussion, the project *Conversations with ELIZA* is presented as an exploration of femininity in AI through the development of four chatbots integrated into a web-based platform. Each of these bots performs specific tasks that seem to highlight particular gender stereotypes or even reflect common assumptions about femininity back to its user. In this manner, this study aims to question whether artificial intelligence tends to reinforce traditional and normative notions of gender and femininity.

1. INTRODUCTION

Artificial intelligence is often associated with fictional and futuristic scenarios in popular culture, even though it has already become part of our daily life. In fact, we frequently interact with AI systems without even realizing it, namely with chatbots whose ubiquity often goes unnoticed. These personal digital assistants are now embedded into our mobile devices and web-based services and platforms. The former can be illustrated by Siri, Alexa or Cortana as personalized services, while the latter refers to online contexts where it has become more and more common to find a bot that aims to assist us in specific tasks (Dale 2016). Regardless of their complexity, they share the goal of assisting users by performing tasks in a kind, helpful and compliant fashion.

However, the process of anthropomorphizing theses assistants by assigning them human-like traits or features seems to be accompanied by a tendency for them to display feminine attributes. These digital entities are often feminized through their name, voice or avatar, while they also execute tasks associated with jobs which are historically performed by women. As such, they seem to behave according to gender stereotypes and reinforce traditional assumptions of femininity (Weber 2005; Hester 2016).

This paper aims to explore how gender relates to AI, while also seeking to understand why most chatbots and digital assistants appear to be female. To this end, it begins by providing an overview of artificial intelligence, addressing questions that arise with their integration in our daily lives. It then approaches the concept of gender in light of an historically patriarchal and heteronormative society that promotes a binary frame (Butler 1990; Haraway 1991), particularly focusing on women and femininity. Artificial intelligence and gender are then related, paying particular attention to Siri, Alexa and Cortana, in order to shed some light on how chatbots and digital assistants appear to be mostly female.

Complementing this discussion and in order to illustrate the topic, the project *Conversations with ELIZA* consists of four chatbots which were developed and integrated in a webpage, seeking to simulate a specific personality with the purpose of emphasizing feminine roles and stereotypes. In this manner, this study thus seeks to question traditional notions of femininity and their significance in AI. It seeks to explore and understand how this relationship takes place, why femininity seems to be often present in AI and which gender roles or stereotypes are reinforced in this process.

2. OVERVIEW

2.1. Artificial Intelligence

2.1.1. A chatbot named ELIZA

Artificial intelligence nowadays encompasses different areas of study, but all of them announce its integration into our daily lives. We now have a more direct contact with this type of technology namely through chatbots that play the role of personal digital assistants embedded into our devices and that engage in conversations with us through natural language. They have also become a natural part of the asynchronous simultaneous conversations we carry out, based in short type interactions. As Robert Dale points out, "chatbots have been around for a long time" and are thus returning, instead of emerging as something new (Dale 2016, 814).

In 1966 Joseph Weizenbaum introduced ELIZA, a computer program capable of analyzing written inputs from users and answer accordingly by using a set of rules, thus establishing a "human" dialogue.[1] Weizenbaum chose a script with which ELIZA

1. To do so, the program searched the inputs for the presence of a keyword, and produced responses "by transforming sentences according to a rule associated with said keyword" (Weizenbaum 1966, 37), and also by replacing certain words or expressions. For example, if a user said something along the lines of "I am upset because of my mother", ELIZA would answer with "Why do you think you are upset because of your mother?". In this sense, ELIZA wasn't restricted to a particular set of responses, although it was limited to a pre-determined set of rules and "adaptable" sentences, and if an input wasn't recognizable or didn't contain any keywords, it failed to have "the provision of a mechanism that would permit ELIZA to respond intelligently" (Weizenbaum 1966, 37).

2. This decision solved a lot of issues regarding ELIZA's "unawareness" about her surroundings or inability to talk about topics out of its framework because the psychiatric interview style allowed a "categorized dyadic natural language communication in which one of the participating pair was free to assume a pose of knowing almost nothing of the real world" (Weizenbaum 1966, 42).

3. Introduced in 1950 by Alan Turing, "the Turing test demands that a human subject decide, based on replies given to her or his questions, whether she or he is communicating with a human or a machine. When the respondents fail to distinguish between human and machine responses, the computer may be considered intelligent" (Halberstam 1991, 442). As such, ELIZA demonstrated how the Turing test poses human intelligence in a somewhat narrow way, since it was considered intelligent simply by being able to follow a logical script and appearing human.

4. They are also able to play music, play videos, search the web, translate sentences, open apps, give directions, announce the weather and even control automation-enabled home systems.

5. According to Morozov, nowadays there is a constant need to attempt to root out any "imperfection, ambiguity, opacity, disorder and opportunity to err, sin or do the wrong thing" (Morozov 2013b, Int. par. 15), which is closely linked to what chatbots aim to do. Personal, digital assistants are now at our disposal, constantly present, ready to help us quickly solve our problems, while tracking our habits and user preferences, leaving little to no room for imperfection — all of this just a touch away.

acted as a Rogerian psychotherapist, since it was "easy to imitate because much of this technique consists of drawing his patient out by reflecting the patient's statements back to him" (Weizenbaum 1976, 3).[2]

ELIZA was one of the first "natural language process applications" that was able to trick some of its users into thinking it "was a person rather than a machine" (Dale 2016, 814), and this was an important point in AI progress. This idea of having a machine talk to us as if it were human, leading us to believe we are speaking to another human being, conveys the purpose of the Turing test.[3]

By shifting human-machine interaction from a purely robotic, rational nature one to a more social one, ELIZA marked a significant development in AI, which relates to the way chatbots evolved towards emulating human behavior. Nowadays, chatbots are a natural component of our technologies as "just another facet of today's always-connected multi-tasking world, where we participate in multiple conversations in parallel, each one at a pace of our choosing" (Dale 2016, 815).

2.1.2. From chatbots to solutionist assistants

William Meisel distinguishes two groups of chatbots: "general personal assistants", which refer to more developed and complex assistants like Siri, Alexa or Cortana, and "specialized digital assistants", which refer to a "tsunami of more narrowly focused chatbots" (Dale 2016, 812).

AI systems of the first type can help with "some subset of the standard virtual assistant skill portfolio", which mainly includes reading, writing, sending emails, scheduling meetings, checking calendars and setting appointments, making calls, sending messages, taking notes, setting reminders, etc (Dale 2016, 812).[4] Usually, general digital assistants are integrated directly into our devices, like Siri in Apple devices, and assist us in a more personalized way.

Specialized digital assistants "operate in very specific domains or help with very specific tasks", usually in web-based platforms or apps that serve specific areas, and their tasks can range from "booking a flight, buying some shoes, taking issue with a parking fine" to sending daily weather forecasts, helping with online shopping payment processes or even just telling jokes (Dale 2016, 812-813).

By performing these tasks, chatbots work towards an amelioration of our daily life, assuring that nothing is left unorganized, forgotten or undone; they make sure that we are as productive as possible by "promoting efficiency, transparency, certitude and perfection — and, by extension, eliminating their evil twins friction, opacity, ambiguity and imperfection" (Morozov 2013b, Int. par. 14). This need to ameliorate our life and maximize production conveys Evgeny Morozov's concept of solutionism, defined as "an intellectual pathology that recognizes problems as problems based on just one criterion: whether they are 'solvable' with a nice and clean technological solution at our disposal" (Morozov 2013a).[5]

2.1.3. From anthropomorphization to companions

Embedded into our cellphones, laptops or tablets, as well as websites, apps or other types of web based services, artificial intelligence is simultaneously ubiquitous and subtle. This growing presence conveys how chatbots are no longer mere tools; they are also "imagined to become friends and companions" (Richardson 2015, 15). This sense of companionship develops alongside with the anthropomorphization of artificial intelligence, as chatbots are endowed human attributes or traits and evolve from assistants to companions that become closer to us.

Anthropomorphization takes place on a more superficial, physical level, through names, voices, avatars, or other kinds of attributes that move away from a purely mechanized presentation. However, it also concerns dialogue and interaction. In this

6. This effect relates to a belief that "the responses which appeared on his typewriter were generated by a human" (Weizenbaum 1966, 42).

sense, Jutta Weber considers that, with anthropomorphization, there is a significant shift from rational-cognitive processes and problem solving to a socio-emotional interaction, which emphasizes this intention of turning our interaction with this type of machines into a more social one (2005, 209). Therefore, this type of technologies are "supposed to mimic or even learn those abilities and characteristics which were, until recently, regarded as purely and typically human and beyond the grasp of machines" (Weber 2005, 213).

Although it becomes more evident among the current myriad of personal digital assistants, anthropomorphization goes back to ELIZA and the fact that "people were conversing with the computer as if it were a person who could be appropriately and usefully addressed in intimate terms" (Weizenbaum 1976, 7).[6]

For Weizenbaum, this raised some problems, which he addressed under the notion of "computationalism", relating to the belief that "the functional relations between mental inputs, outputs and internal states are computational" (Piccini 2004, 814). However, Weizenbaum considered that not every part of the human thought could be reduced to logical formalisms and that "there are some acts of thought that ought to be attempted only by humans" (Weizenbaum 1976, 13). Hofstadter later corroborates this idea stating that "no program in the world understands even one concept at the same level of complexity as an ordinary person does" (Hofstadter 1995, 160).

On the other hand, Weizenbaum also raises the question of human-machine relationships, observing that when we interact with machines as if they were human, we start developing emotional bonds, a sense of empathy and of being understood. In fact, he observed "how quickly and how very deeply people conversing with ELIZA became emotionally involved with the computer" (Weizenbaum 1976, 6).

2.1.4. The ELIZA effect

This illusion, which he considered dangerous, is known as the ELIZA effect, describing:

> the susceptibility of people to read far more understanding than is warranted into string of symbols strung together by computers (…) and the idea that computers "understand" the physical world, reason abstractly, make scientific discoveries, are insightful cohabiters of the world with us". (Hofstadter 1995, 157)

Consequently, people start getting attached to these technologies (and to the entities contained within them). By creating anthropomorphized digital assistants, giving users the false sense they are talking to another human being, human-machine interaction is influenced by feelings of intimacy, closeness and empathy. In this process, personal digital assistants engage in conversations with us, evoking a not-so-far-away world "where some of those conversational partners we'll know to be humans, some we'll know to be bots, and probably some we won't know either way, and may not even care" (Dale 2016, 815).

2.2. Gender and stereotypes

When chatbots are anthropomorphized, whether through their voice, name, or the way they interact, they tend to portray gender related features. In order to understand this phenomenon, we need to take a closer look at gender as one of the aspects through which we socially develop and establish relationships, whether with each other or with ourselves.

Gender constitutes a part of our identity that regulates the type of behavior or acts we establish socially "by managing situated conduct in light of normative conceptions of attitudes and activities appropriate for one's sex category" (West and Zimmerman 1987, 127). In this sense, Judith Butler introduced the idea that gender

has a performative nature, given that gender identity is a repetition of acts stylized through time, manifesting a "cultural interpretation or signification of that [biological] facticity" (Butler 1990, 522).

2.2.1. Binary framework

Simone de Beauvoir once said that "one is not born, but rather becomes, a woman" since 'woman' (as a concept) is a "historical idea and not a natural fact" (in Butler 1990, ch. 1 sec. III par. 3, 1988, 522). These words suggest how gender is not something we are born with and, instead, is something we internalize through performative acts, over time. To be female or male is a matter of sex; but to be a man or a woman is a matter of gender. Gender is also seen as something polar, as seen through a "binary framework" in which there is a "mimetic relation of gender to sex whereby gender mirrors sex or is otherwise restricted by it" (Butler 1990, 88).[7]

Consequently, there is a normalization of what is considered to be feminine or masculine behavior, which becomes predetermined in a foreclosed historically sedimented structure. This establishes a set of expected behaviors and acts according to which we are compelled to act. That expectation is based on the perception that others have of our sex, which is presumed through the "factic datum of primary sexual characteristics" (Butler 1988, 528). In other words, through this "need to routinize (...) behavior in accord with pre-established conceptualizations and behavioral patterns" (Deaux and Major 1987, 370), certain attributes and acts are identified as specifically feminine or masculine and are supposed to imply someone's preferences and behaviors. As we grow up, and are categorized as men or women (or, instead, boys or girls) we are expected to comply to "normative conceptions of appropriate attitudes and activities" that are determined by "institutionalized frameworks through which natural, 'normal sexedness is enacted'" (Goffman, 1977 in West and Zimmerman 1987, 137).

As Prentice and Carranza put it, "prescriptive gender stereotypes" define "the qualities [ascribed] to women and men (...) that are required of women and men" (2002, 269).[8] These stereotypes imply that a gender belief system imposes expectations and gender behavior patterns, as internalized and socially reinforced stereotypes. Butler expands on this, stating that "gender performances (...) are governed by (...) punitive and regulatory social conventions" (Butler 1988, 527) that reject the acts or behaviors that convey some kind of deviation from the norm.

2.2.2. Feminized labour

Gender roles and characteristics deemed as specifically feminine or masculine also imply a structural hierarchization of labour. In other words:

> If, in doing gender, men are also doing dominance and women are doing deference (cf. Goffman 1967, pp. 47-95), the resultant social order, which supposedly reflects 'natural differences', is a powerful reinforcer and legitimator of hierarchical arrangements". (West and Zimmerman 1987, 146)

This means that feminine and masculine behavior is also used to segregate and structure labour accordingly. The workplace and its relationships change since, according to Kelly, when we interact within these contexts "social labels, beliefs and attributions may serve as grounds for predictions and generate behavior designed to validate or invalidate these beliefs" (in Snyder 1977, 8). In fact, a lot of service work is seen as feminized labour or "associated with qualities traditionally coded as feminine" (Hester 2016, 47).[9]

7. According to Judith Butler, gender is "radically independent of sex" and, instead, is a "free-floating artifice", while sex is defined as a "biological facticity" (Butler 1988), which means it is a biological criterion that distinguishes solely between female and male. As Butler puts it, gender "is neither the causal result of sex nor as seemingly fixed as sex" (Butler 1990, ch.1 sec. II par. 1). Therefore, gender is not something inherent "because gender is not a fact, the various acts of gender creates the idea of gender, and without those acts, there would be no gender at all" and gender is shaped and socially defined according to a "tacit collective agreement to perform, produce and sustain discrete and polar genders as cultural fictions" (Butler 1988, 522).

8. Some of these stereotypes, presented by Bem (1981 in Prentice and Carranza 2002, 269), describe feminine characteristics as "affectionate, cheerful, childlike, compassionate, does not use harsh language, eager to soothe hurt feelings, feminine, flatterable, gentle, gullible, loves children, loyal, sensitive to the needs of others, shy, soft-spoken, sympathetic, tender, understanding, warm, yielding". On the other hand, masculine characteristics are described as "acts as a leader, aggressive, ambitious, analytical, assertive, athletic, competitive, defends own beliefs, dominant, forceful, has leadership abilities, independent, individualistic, makes decisions easily, masculine, self-reliant, self-sufficient, strong personality, willing to take a stand, willing to take risks".

9. This is tied to "women's practices (...) within the terms of some more dominant cultural formation (Butler 1990, ch. 1 sec. I par. 8) and to what are historically considered women's places, "idealized social locations seen primarily from the point of view of advanced capitalist societies: Home, Market, Paid Work Place, State, School, Clinic-Hospital and Church" as Donna Haraway explains it (Haraway 1991, 307).

In other words, by expecting certain acts (deemed as feminine) from women, we expect them to occupy jobs and perform tasks associated with these attributes, thereby creating a category of feminine labour. To give a concrete example, historically women have a significant presence in the telecommunications industry, where they filled the role of assisting and establishing calls and communications, which rendered "female operators (...) inferior, subordinate, and knowable" (Zost 2015, 3). In other cases, women fill the role of secretaries, assistants, nurses or even flight attendances. These type of jobs convey, in a way, an "assumption that women possess a natural affinity for service work and emotional labour" (Hester 2016, 47).

This asymmetry also affects the private sphere, namely domestic work. As West and Zimmerman explain, household and child care tasks are considered women's work as a consequence of "normative conceptions of appropriate attitudes and activities for sex category" (West and Zimmerman 1987, 139). The heterosexual framework contributes to this asymmetry since it reinforces the "embodiment of wifely and husbandly roles, and derivatively, of womanly and manly conduct" (Beer 1983 in West and Zimmerman 1987, 144).

Additionally, and according to Donna Haraway, domestic work is transformed into capitalized labour out of the private sphere, through jobs such as office work, nursing or service work. Borrowing from Richard Gordon, Haraway considers that, with new media, a "homework economy" emerges, defined as a "restructuring of work that broadly has the characteristics formerly ascribed to female jobs, jobs done only by women" (Haraway 1991, 304).[10]

Therefore, even outside the domestic sphere, women still ensure domestic tasks: "partly as function of their enforced status as mothers" as well as working in an "integrated circuit (...) in advanced industrial societies [where] these positions have been restructured (...) by social relations mediated and enforced by the new technologies" (Haraway 1991, 305-307). This reflects traditional conceptions of gender derived from a patriarchal heteronormative society where women perform domestic and assistant-like roles, while it also reveals how gender standardization and normalization has consequences at a social, personal and structural level.

2.3. Gendered AI

2.3.1. Feminized labour automated

Gender norms and stereotypes in artificial intelligence take form in various ways, not only through anthropomorphization, but also when tasks performed by chatbots begin to mirror traditional feminine labour. As Halberstam explains, what we observe is a "gender automatization" given that tasks traditionally and historically considered female become a part of technology (Halberstam 1991, 451).

Hence, AI performs tasks considered feminine and does so as a natural part of its system programming. It reflects:

> (...) our assumptions about feminized labour and our existing relationship to socially gendered caring and service behaviors, tapping into those elements of femininity that have historically enabled caregiving or service-providing subjects to better undertake specific obligations, activities, and tasks. (Hester 2016, 50)

The author adds that "we are witnessing the protocols of femininity being programmed into machines" (Hester 2016, 48) as labour previously deemed as feminine becomes technologic. Accordingly, we can observe how general or specialized chatbots automate work that is coded as female (Hester 2016), given that they mainly operate in service or assistance related contexts, acting as personal assistants, secretaries and the like.[11] By operating in contexts of service, and following these standardized behaviors, chatbots also end up emulating attitudes that resemble, as Gustavsson puts

10. For example, a personal assistant conducts "a form of corporate care work, including providing sustenance of the body in the form of teas, coffees and lunch orders, as well as making dentists' appointments, picking up dry cleaning, paying personal bills, and so on" (Hester 2016, 49).

11. General personal assistants such as Siri, Cortana or Alexa perform traditionally feminine tasks by acting as assistants (searching the web, translating sentences or controlling automation-enabled home systems), secretaries (registering information, sending emails or setting up appointments) or even telecommunication operators (making calls, sending messages and establishing communications in general). Similarly, specialized personal assistants also perform tasks aimed at helping us with services, such as online shopping payment processes, acquiring travelling tickets or even looking through a shop's online catalog.

it, a "stereotyped image of female service providers" (in Hester 2016, 47). They display feminine attributes because these characteristics have "its basis in the stereotyped image of female qualities. (...) Such a stereotypical female image of caring, empathy and altruistic behaviour has become a standard component in a service script" (Gustavsson 2005, 402 in Hester 2016, 47).

2.3.2. Gendered interaction

12. There are chatbots that display masculine traits, namely specialized digital assistants that perform specific tasks, mostly in contexts associated with manly labour. However, this paper focuses on general personal assistants rather than specialized, and thus addresses the observed predominance of female traits in this type of assistants.

However, it is not only trough the human attributes displayed, but also the dialogue and tasks it performs, that a chatbot becomes a gendered entity. So beyond the physical level of anthropomorphization, gender is also revealed at the performative interaction level. In other words, gender is visible in features like voice, name or, in some cases, avatar. By default, Siri, Alexa and Cortana display feminine voices, and only Siri has a masculine option, limited to a particular set of languages.[12] Siri's name, in nordic, translates to "beautiful woman who leads you to victory" (Fessler 2017). And Cortana's name is inspired on a character from the videogame Halo, whose avatar is a woman.

These aspects are defined prior to any interaction, but their dialogue and interaction also reinforces this feminization since, beyond their service and assistance, chatbots frequently display feminine characteristics through socio-emotional based dialogues. Siri, for example, presents itself as an entity that, in her words, "lives to serve" and please its users, thus fulfilling a submissive role.

As Weber puts it, these gendered dialogues imply a "reduction of social interaction to stereotypical and gendered behavior patterns" (Weber 2005, 215). Consequently, human like behavior in social machines becomes standardized and gender stereotypes are instrumentalised to manage our relationship with chatbots, reproducing and reinforcing social clichés (Weber 2005, 214). Often, the behavior of chatbots confirms expectations regarding gender, when following socially established feminine behavioral patterns.

2.3.3. Digital moms, caregivers and femmebots

Their tasks also resemble "traditional care giving activities associated with domesticity" (Hester 2016, 49) and, in fact, a lot of their functions consist in ensuring our well being, reflecting upon motherly acts. For Weber, this maternal attitude conveys one of the aspects that mainly defines our relationship with machines, since this interaction follows a "caregiver-infant logic" (Weber 2005, 214). Given that "sociality and emotionality have been deeply gendered categories in western thought that have hitherto been assigned to the feminine realm" (Weber 2005, 213), instead of seeing a machine, we start looking at chatbots as feminine entities that look after us. By fulfilling these roles, chatbots, begin to develop relationships with us that might go beyond mere daily assistance, since they start simultaneously emulating attributes that are not only related to historically feminine labour but also to motherly acts.

According to Snyder, social stereotypes constitute "(...) pieces of information [which] are usually the first to be noticed in social interaction and can gain high priority for channeling subsequent information processing and even social interaction" (Snyder 1977, 2).

Thus, when chatbots relate to us by simulating social norms and gender stereotypes, they establish expectations and possible approaches regarding user interaction, such as the idea that "all women are dependent and conforming" as Snyder suggests (Snyder 1977, 2). Adding to this idea, Hester states that "when technologies 'do gender' it is obviously not natural, but is instead visible as the product of deliberate choices about how best to relate, assist, or persuade the imagined technology user" (Hester 2016, 50).

The ELIZA effect already identified an attachment that derives from approaching machines as if they were human, and the fact that we might develop emotional bonds and a sense of empathy with them. In the context of daily interaction with ubiquitous chatbots, these social and affective effects become more evident.

When this attachment is felt towards entities that appear to empathize and understand their users, and whose role consists of assisting and simultaneously look after them, it also reinforces the idea that emotionality and ensuring someone's well being are feminine features. When bots interact in a motherly logic, attachment also conforms to expectations and stereotypes that associate femininity with emotional and domestic caregiving acts.

Interacting with artificial intelligence systems on a daily basis, makes us look at them not only as mere machines, but also as "mirrors or substitutes" with gendered attributes that match socially established expectations (Weber 2005, 216). As they try to become closer to reality, it is from reality itself that they draw rules for their interaction and presentation, thus reproducing and automating historically feminine jobs and tasks, but also articulating these roles with female voices, names, avatars and social behaviors.

When interacting with these humanized chatbots, we engage in communication processes similar to those we establish with human beings. Consequently, the way we relate to our peers starts influencing how we relate to artificial intelligence and vice-versa. And when we look at these gendered digital personal assistants as substitutes, there is a risk that they might affect the way we feel, perceive, interpret and even describe reality, gender and women.

This results in a somewhat questionable relation between femininity and artificial intelligence that appears to conform to normalized ideas of gender, reflecting these ideas back to reality.

3. CONVERSATIONS WITH ELIZA

Taking on the ideas addressed, and in order to complement their discussion, the project *Conversations with ELIZA* seeks to explore and expose this currently observable femininity of artificial intelligence, particularly in chatbots and assistants. The project, therefore, intentionally highlights certain feminine traits conforming to gender stereotypes that become apparent in their anthropomorphization, the functions they perform and, particularly, the socially established feminine behavioral patterns this kind of entity can assume.

The project involves the development of chatbots with different personality traits, which are implemented on contexts in which the bots normally operate (such as Facebook messenger or Twitter). These are contextualized in a online platform, as a primary component of the project that seeks to briefly elucidate on what AI is, with another chatbot whose function is to explain its own creation process and including links to the other chatbots.

Each of the chatbots developed takes on a particular archetype, characteristic of the relation between AI and female stereotypes, which becomes evident through interaction with the user. In this manner, and as an exploratory project that motivated the ongoing research on which this paper is based,[13] *Conversations with ELIZA* seeks to incite reflection on the apparent predominance of the female gender in artificial intelligence, and how it can reinforce traditional and normative notions of gender and femininity.

13. *Conversations with ELIZA* was developed in the 1st year of the Masters degree in Communication Design and New Media at the Faculty of Fine-Arts, University of Lisbon as a preliminary approach to the topic which motivated this paper and the research we are now undertaking.

Fig. 1. Conversations with ELIZA website.

3.1. Concept

Seeking to question the relation between gender within AI systems the project focuses on designing the chatbots' dialogues, tasks and personality traits, whose femininity is gradually revealed through interaction. In terms of dialogue-based interaction, it proposes different types of experiences with chatbots, whose conversational subject matter develops around the ideas and concepts relating to AI and gender, in order to introduce the user to the topic. In terms of their role as assistants, each chatbot is designed to perform specific tasks in an attempt to simultaneously portray standard virtual assistant skills and functions associated with traditional female labour.

Femininity is also gradually revealed through their characteristic personality traits, that seek to emulate feminine archetypes, which are approached with a certain irony and in a somewhat caricatural manner, by portraying personalities specifically designed to meet their functions and by making evident the traits or attributes typically associated to them.

3.2. Implementation

Concerning the methods for implementing dialogues, tasks and personality traits, and in order to ensure successful interactions, we began by developing rules-based dialogues that allow retrieval-based responses. Focusing on common AI errors and how to avoid them, we aimed to eradicate off-track moments by presenting suggestions in a multiple-choice fashion.

Regarding their tasks, we first looked into common functions offered by chatbots, and then into traditional attributes associated with female labour as previously described. We then combined both, and came up with four different tasks that simultaneously referred to AI and feminine tasks: explaining how chatbots work and are made, sending to-do reminders, giving daily compliments and pep talks, and tweeting curious facts (in this case, regarding women).

These tasks also reflected upon the bots' personalities. But we also focused on particular archetypes, characteristic of AIs (such as Helper, Lover, Motherly Figure and Femme Fatale),[14] and combined them with traditional female stereotypes (namely Innocent, Orphan, Caregiver and Ruler)[15] in order to expose recognizable and expected social behavior, therefore drawing inspiration from popular culture and how it typically portrays femininity in AI (e.g.: Her, Ex Machina, Humans, Metropolis).[16]

Accordingly, we came up with a helpful, compliant assistant; a motherly, caregiving figure; a cheerful, understanding and intimate figure; and an irreverent, sarcastic figure, as described in the results. Adding to this, we also analyzed the dialogues,

14. These archetypes, retrieved from an article analyzing female robots and AIs, are mainly found in pieces of media that depict female AIs. The Helper archetype refers to helpful and compliant assistants, the Lover to figures that seek to satisfy lack of intimacy or emotional contact, the Motherly Figure to empathic, sympathetic figures who may also be worried or disappointed, and the Femme Fatale to a simultaneously attractive and dangerous figure that seeks power and conflict (Anders, 2015).

15. These stereotypes are also found in media depicting women, while also referring to Bem's stereotypes (1981 in Prentice and Carranza 2002, 269). The Innocent stereotype refers to naïve, optimistic women that try to follow the rules, the Orphan to women that try to please others and wish to be well seen as well as feel integrated, the Caregiver relates to maternal women that look after others and try to protect and ensure their well being, and the Ruler pertains to bold and competitive women that seek power and are not afraid to break the rules.

16. Most of these examples, despite portraying said archetypes, also include feminized bodies. In this sense, *Her* constitutes a particularly interesting example since Samantha (the AI) only takes form through its voice, revealing how stereotyped femininity (in this case, the Lover archetype) can still be portrayed without physical appearance.

17. We chose Siri because we had easy access to Apple products, making it easier and quicker to analyze this AI. We analyzed how it would react to certain commands and observe which stereotypes it displayed and noted, for example, that the AI often put itself down or reassured the user about being in control. Our ongoing research seeks to expand this analysis to other bots in order to further examine stereotypes in AIs.

18. This approach was inspired by an article, written in 2017, that showed how Siri, Alexa, Cortana and Google Home reacted to sexual harassment, noticing their reactions mostly reinforced traditional ideas about women. In this situation, the bots "most frequently evaded harassment, occasionally responded positively with either graciousness or flirtation, and rarely responded negatively" (Fessler, 2017). The article also concluded that "the fact that Apple writers selected "I'd blush if I could" as Siri's response to any verbal sexual harassment, [...] or coy, evasive responses like Alexa's "Let's change the topic" in response to "You are a slut" ends up reinforcing stereotypes of passive, subservient women (Fessler, 2017).

19. tinyurl.com/yaecumal

20. facebook.com/assistantcwe

Fig. 2. Section that provides context on AI.

tasks and personality traits displayed by Siri[17] to see how it would react to dominant or friendly acts,[18] borrowing from its answers in order to further develop the chatbots.

Finally, and although linked to the primary webpage of the project, the bots were integrated in platforms that resonated with their tasks and echoed common contexts in which chatbots typically operate, ranging from web-based chats to social pages like Twitter or Facebook.

3.3. Results

The project's website[19] provides context on what AI is and how it works, but also highlights gender and femininity within AI. Taking on a previous and broader research project on the relationship between gender and AI and on how to create perfect female cyborgs, bots and AIs, this site puts a particular focus on the creation of feminized chatbots.

Accordingly, the project presents four chatbots that engage with their users through different types of interactions such as dialogues, tweets and reminders, and thus explores gender in AI within and outside its website, while presenting the subject to possibly unaware users. It also allows for multiple interactions that range from texting bots, to reading tweets or even dragging them across the screen.

The main bot,[20] integrated on the project's webpage as well as on Facebook messenger, is an assistant whose function is to explain, through dialogue, the female AIs' creation processes, or how femininity emerges in these contexts. This bot borrows from female stereotypes associated with service contexts, such as being compliant, helpful, and gentle.
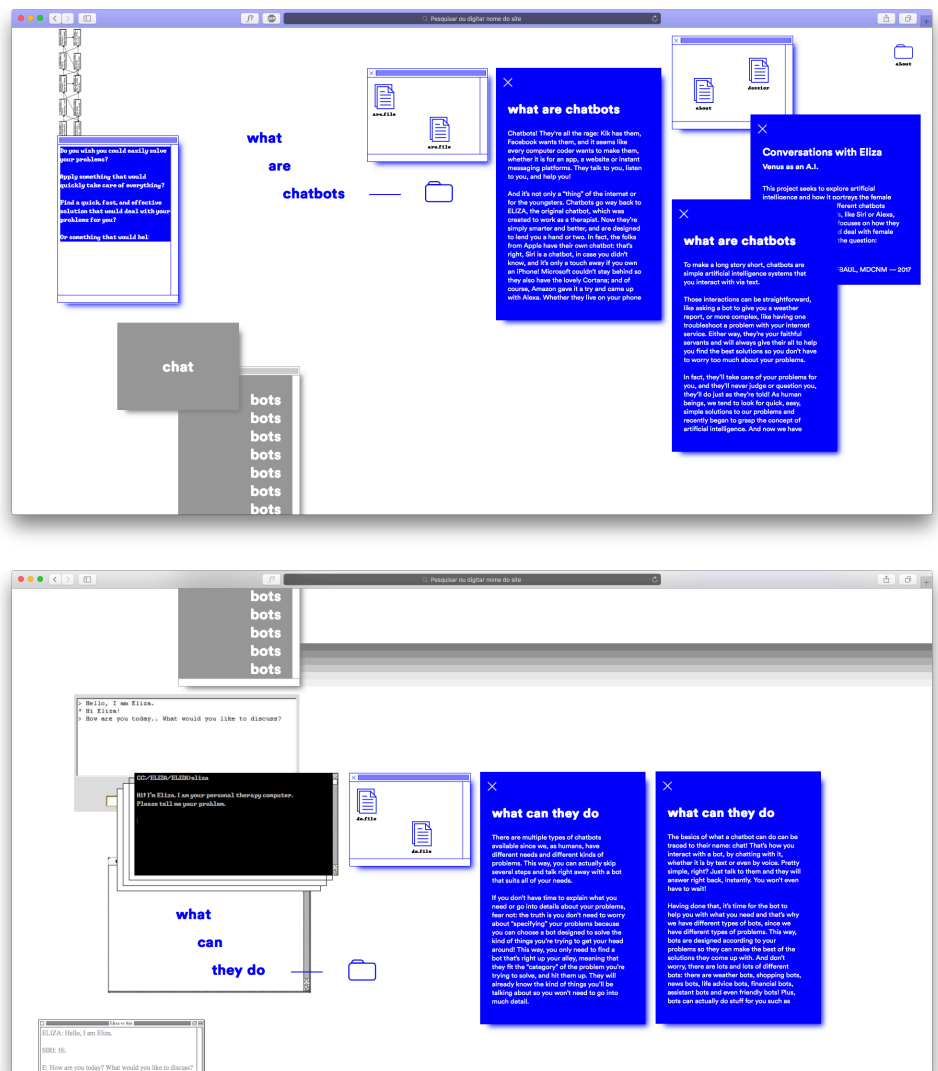
Fig. 2. (cont.) Section that provides context on AI.

Fig. 3. The Assistant bot, presenting itself.

21. twitter.com/cybelecwe

Subsequently, the other three bots are presented and named according to the different female stereotypes ascribed to them: Cybele, Iynx and Electra. The first, Cybele,[21] whose name is inspired on an Anatolian mother goddess, is a tweet scheduling bot that uses code lines to generate tweets and send daily "maternal" reminders, operating as a simultaneously caring, obsessive and disappointed motherly figure. It exhibits stereotypical behavior such as being compassionate, sensitive to the needs of others, and yielding.

Fig. 4. Assistant, accessible through its
Facebook page.
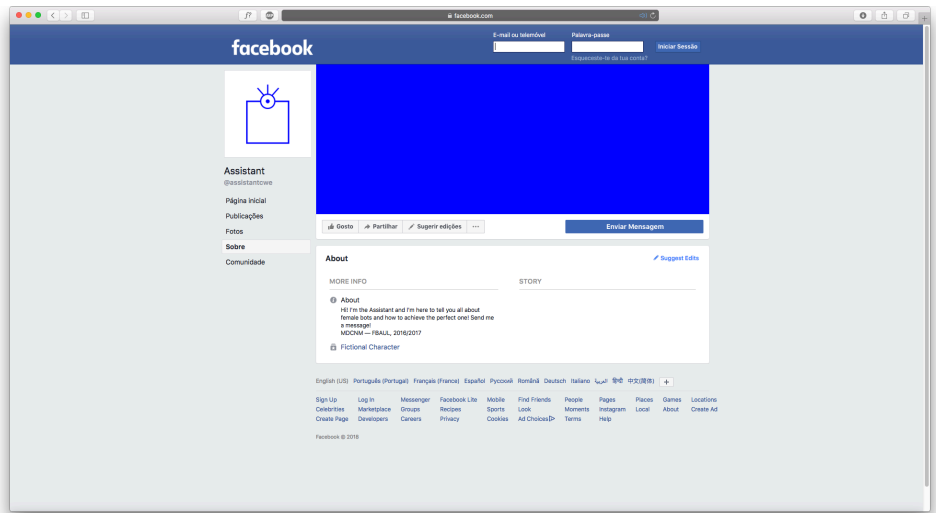
Fig. 5. The remaining bots Cybele, Iynx and
Electra.

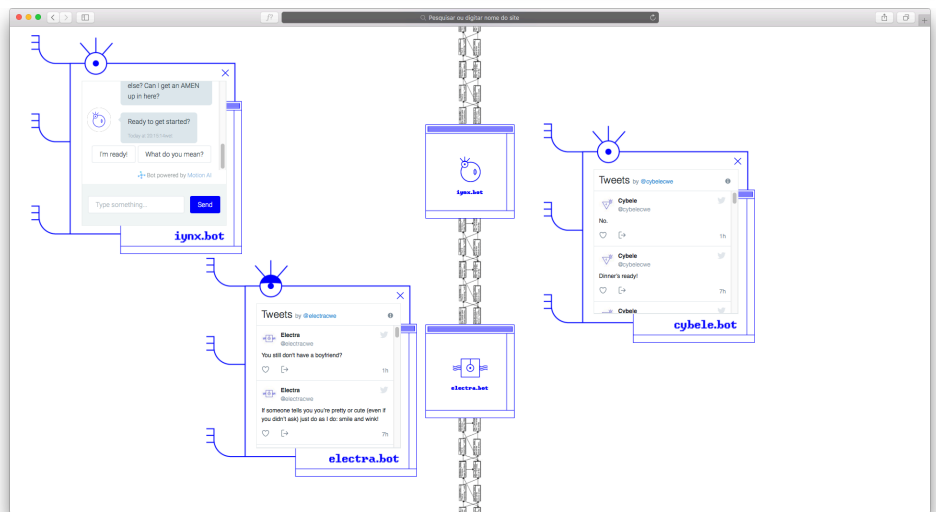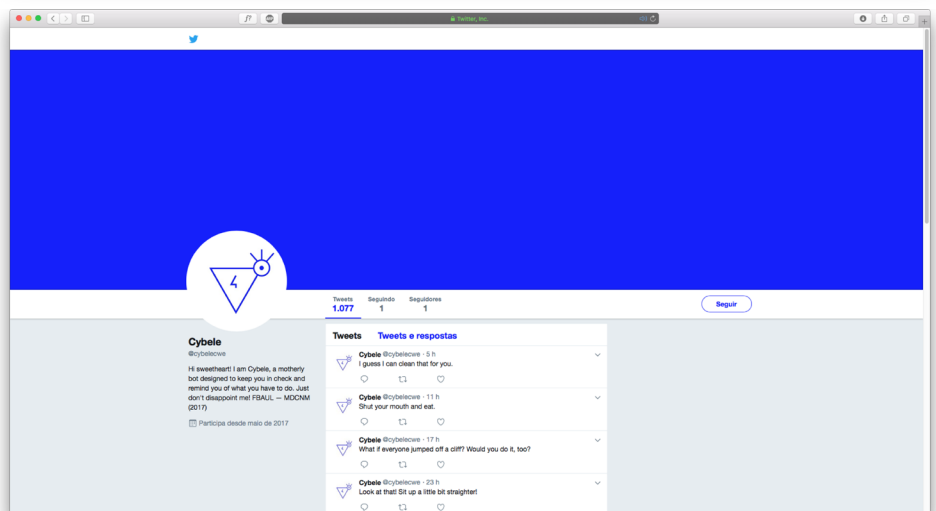

Fig. 6. Cybele's twitter profile.

22. facebook.com/iynxcwe

Inspired on a Greek nymph, Iynx[22] consists of a Facebook messenger based bot, which operates as a seductive, empathizing figure that tries to help its users with their self-esteem, by offering the feature of sending daily compliments and pep talks. Accordingly, it does not use harsh language and is eager to sooth hurt feelings, while being soft-spoken, childlike and understanding.

23. twitter.com/electracwe

Electra,[23] whose name is inspired on a Greek vengeful figure, follows a less conventional approach. By portraying a more defiant and bold attitude, this tweet scheduling talks about common assumptions regarding women, eventually twisting them or presenting them ironically. It tampers with female and male stereotypes, such as being assertive, self-sufficient and having a strong personality.

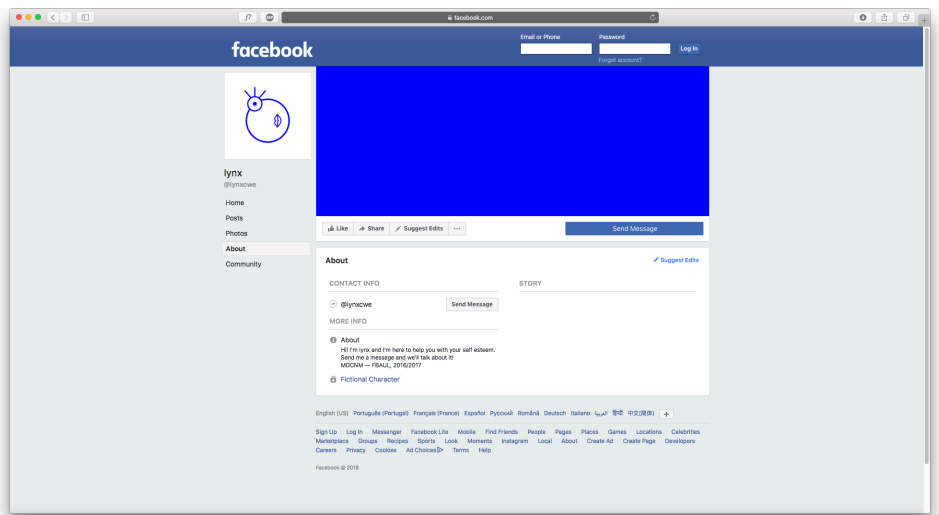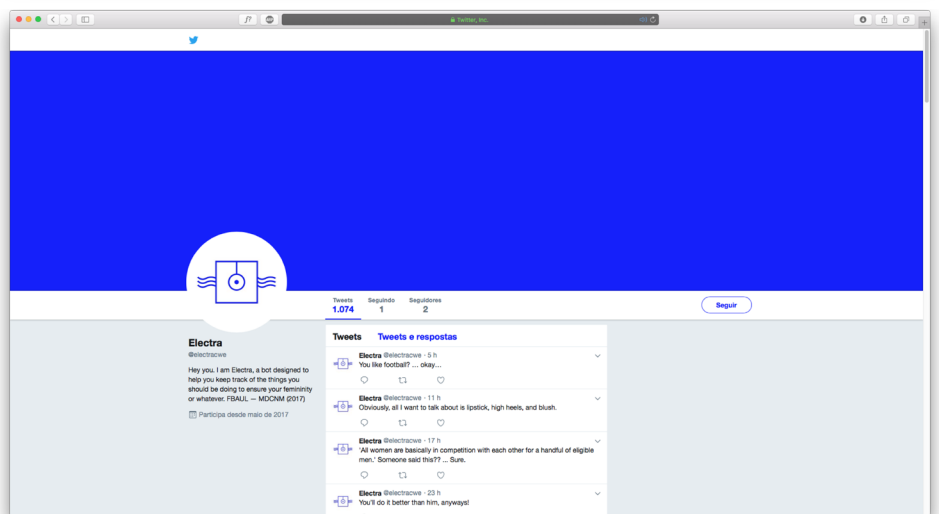Fig. 7. lynx, accessible through its Facebook page.

Fig. 8. Electra's twitter profile.

These different personalities and functions seek to incite reflection on feminized AI, its multiplicity and the questions that arise, particularly, when it reinforces traditional gender roles and accentuates cultural stereotypes. These bots seek to promote this kind of reflection by exposing, and intentionally accentuating, female stereotypes observed in AI. Accordingly, further developments of this exploratory project contemplate the expansion of these chatbots functions and femininity, aiming for more complex interactions and gender portrayals.

4. CONCLUSION

Artificial intelligence increasingly integrates our daily life and its development is pushing it towards a social, humanized realm. In this context, chatbots are no longer mere assistants given their ubiquity and their way of interacting that brings them closer to friendly companions. However, these ubiquitous companions often perform tasks that echo historically feminine roles and articulate these features with stereotypical behaviors. The attempt to bring them closer to human traits and interactions also reveals a biased view of gender through a feminized (often submissive) role, lacking a counterpart or gender neutral approach, or just mere cultural diversity.

This paper sought to examine and explore the relationship between gender and artificial intelligence and its significance as a field that, in its rapid development, often eludes awareness and critical stances on the social and cultural roots that inform its evolution. But rather than providing answers or guidelines to counter an observable tendency towards feminization of digital assistants, this paper sought to tackle into the questions that arise when the topic is subject to closer inspection.

According to this idea, *Conversations with ELIZA* sought to illustrate and comment on the phenomenon of feminized chatbots, ironically reinforcing some of the stereotypes we engage with. In this sense, it's partially inspired by portrayals of AI in popular culture but also, and more profoundly, by common assumptions about femininity often portrayed by chatbots such as Siri, Cortana or Alexa as an integral part of our daily lives. As abstract and neutral as these entities might want to be in their conception, they end up reflecting our common assumptions and views back to us.

References:

Anders, Charlie Jane. 2015. "From Metropolis to Ex Machina: Why Are So Many Robots Female?" In *io9, Gizmodo*. https://io9.gizmodo.com/from-maria-to-ava-why-are-so-many-artificial-inte-lige-1699274487

Butler, Judith. 1988. "Performative Acts and Gender Constitution: An Essay in Phenomenology and Feminist Theory." *Theatre Journal* 40 (4):519-531. The Johns Hopkins University Press.

Butler, Judith. 1990. *Gender Trouble: Feminism and the Subversion of Identity*. New York and London: Routledge Classics.

Dale, Robert. 2016. "Industry Watch: The return of the chatbots." *Natural Language Engineering* 22 (5):811-817. Cambridge University Press.

Deaux, Kay, and Brenda Major. 1987. "Putting Gender Into Context: An Interactive Model of Gender-Related Behavior." *Psychological Review* 94 (3):369-389. American Psychological Association Inc.

Fessler, Leah. 2017. "Siri, Define Patriarchy: We tested bots like Siri and Alexa to see who would stand up to sexual harassment." *Machines with Brains*. Quartz Media LLC. https://qz.com/911681/we-tested-apples-siri-amazon-echos-alexa-microsofts-cortana-and-googles-google-home-to-see-which-personal-assistant-bots-stand-up-for-themselves-in-the-face-of-sexual-harassment/

Halberstam, Judith. 1991. "Automating Gender: Postmodern Feminism in the Age of the Intelligent Machine." *Feminist Studies* 17 (3):439-460. Feminist Studies Inc.

Haraway, Donna. 1991. "A Cyborg Manifesto: Science, technology and socialist-feminism in the late twentieth century." In *Simians, Cyborgs, and Women: The Reinvention of Nature*. London: Free Association Books.

Hester, Helen. 2016. "Technology Becomes Her." *New Vistas* 3 (1):46-50.

Hofstadter, Douglas. 1995. "The Ineradicable Eliza Effect and Its Dangers." In *Fluid Concepts & Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*, 155-168. New York: Basic Books.

Morozov, Evgeny. 2013. *To Save Everything, click here: The Folly of Technological Solutionism.* New York: Public Affairs.

Morozov, Evgeny. 2013. "The Perils of Perfection." *The New York Times.* https://www.nytimes.com/2013/03/03/opinion/sunday/the-perils-of-perfection.html

Piccini, Gualtiero. 2004. "Functionalism, computationalism, and mental states." *Sutdies in History and Philosophy of Science* 35:811-833. Department of Philosophy, Washington University.

Prentice, Deborah A., and Erica Carranza. 2002. "What Women and Men Should Be, Shouldn't Be, Are Allowed to Be, and Don't Have to Be: The Contents of Prescriptive Gender Stereotypes." *Psychology of Women Quarterly*, 269-281. Blackwell Publishing.

Richardson, Kathleen. 2015. *An Anthropology of Robots and AI.* New York: Routledge.

Snyder, Mark. 1977. "On the Self-Fulfilling Nature of Social Stereotypes." Annual Meeting of the American Psychological Association, San Francisco, California.

Weber, Jutta. 2005. "Helpless machines and true loving care givers: a feminist critique of recent trends in human-robot interaction." *Info, Comm & Ethics in Society* 3:309-218. Troubador Publishing Ltd.

Weizenbaum, Joseph. 1966. "ELIZA - A Computer Program For the Study of Natural Language Communication Between Man And Machine." *Computational Linguistics* 9 (1):36-45. Communication of the ACM.

Weizenbaum, Joseph. 1976. Computer Power and Human Reason: From Judgment to Calculation. New York/San Francisco: W.H. Freeman and Company.

West, Candace, and Don H. Zimmerman. 1987. "Doing Gender." *Gender and Society* 1 (2):125-151. Sage Publications.

Zost, Mary. 2015. "Phantom of the Operator: Negotiating Female Gender Identity in Telephonic Technology from Operator to Apple iOS." Senior Thesis, BA, Faculty of College of Arts and Science of Georgetown University.