



# Utilization of Speculative Design for Designing Human-AI Interactions

Sukwoo Jang<sup>1</sup>, Ki-Young Nam<sup>2\*</sup>

<sup>1</sup>Department of Industrial Design, Ph.D. Student, KAIST, Daejeon, Korea

<sup>2</sup>Department of Industrial Design, Associate Professor, KAIST, Daejeon, Korea

---

## Abstract

**Background** Understanding artificial intelligence (AI) and contemplating its effects present substantial challenges. A powerful approach to tackling this issue is speculative design (SD), which greatly involves constructing narratives that mobilize discussion on the design and social adoption of technologies. However, research on SD narratives in the context of AI has been scarce. Therefore, this study aims to identify narrative themes in SD that concern human-AI interaction.

**Methods** To begin with, 22 related research cases were collected from the Association for Computing Machinery (ACM) digital library based on selection criteria. Subsequently, a constant comparative method was employed to analyze the selected research cases, which resulted in 16 narratives. Thereafter, affinity diagramming was conducted to form higher-order categories, resulting in identifying five narrative themes.

**Results** The analysis yielded five narrative themes: 1) AI Revealing its Ways of Learning, 2) Exposing the Creator of AI, 3) Staging Conflict among Users, 4) Situating Users as Hackers, and 5) Betrayal of AI. All five narrative themes were found to create a discursive space about human-AI interaction and to generate design insights that concern the socio-technical issues of AI.

**Conclusions** The findings of this study add understanding to the growing field of critical thinking in human-computer interaction (HCI) research. They provide insights into developing more ready-to-use methodological devices that can stimulate discourse around human experience of AI. It is expected that scholars and practitioners alike may use the findings of this study to apply an SD approach for investigating human-AI interaction.

**Keywords** Artificial Intelligence, Human-AI Interaction, Speculative Design, Narrative

---

This research was supported by the 4th BK21 through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (MOE) (NO.4120200913638).

\*Corresponding author: Ki-Young Nam (knam@kaist.ac.kr)

*Citation:* Jang, S., & Nam, K. -Y. (2022). Utilization of Speculative Design for Designing Human-AI Interactions. *Archives of Design Research*, 35(2), 57-71.

<http://dx.doi.org/10.15187/adr.2022.05.35.2.57>

**Received :** Jul. 27. 2021 ; **Reviewed :** Mar. 08. 2022 ; **Accepted :** Mar. 08. 2022

**pISSN** 1226-8046 **eISSN** 2288-2987

**Copyright :** This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted educational and non-commercial use, provided the original work is properly cited.

---

## 1. Introduction

Artificial intelligence (AI) has become increasingly prevalent in our lives. Everyday objects from speakers to automobiles are embedded with AI; all of which are capable of carrying out more complex tasks with greater autonomy. This kind of machine autonomy brings about new user experiences (UXs) and social implications that need careful understanding and preparation. To account for such potential changes, scholars of HCI and design have directed noticeable interest towards critical practice: re-imagining the use of technology and opening a discursive space to reflect upon the relationship between humans and technology (Bardzell & Bardzell, 2013; Pierce et al., 2015).

One approach for achieving criticality in HCI research is speculative design (SD), which involves constructing narratives of alternative futures to challenge the conventional assumptions about technology design (Dunne & Raby, 2013; Lindley, Coulton, & Sturdee, 2017). Narratives play an important role in SD because they situate SD artifacts in a use context that fosters imagination and discussion about human-technology interaction (Malpass, 2013). With this in mind, how should SD narratives be constructed to open up discussion about human-AI interaction? The literature presents conceptual discussion on the value and roles of SD but does not provide knowledge on creating SD narratives in the context of AI. As such, three research aims were set as follows:

- To build an understanding on the significance of narratives in SD for investigating the issues of human-AI interaction.
- To identify the narrative themes in SD that mobilize discussion on the design of AI products and services.
- To propose methodological implications of SD for AI.

The term “theme” in narrative themes is intentionally used in this research to describe the types of narratives utilized in SD artifacts that foster discussion on human-technology interaction. This choice of term is based on the fact that, in the field of literature and literary theory, a theme is defined as the semantic value of a narrative that encourages discussion on human experience (De Beaugrande, 1982; Smarr, 1979).

In order to identify the narrative themes, a qualitative content analysis was conducted on 22 research cases that had employed SD to gain design insight on human-AI interaction. As such, this research is positioned as “research into” SD by studying cases of “research through” (Frayling, 1993) SD. The analysis yielded five narrative themes, which carried socio-technical issues associated specifically with AI, supported the imagination of alternative human-AI interaction, and stimulated discussions on the design of AI and its social implications.

This paper begins with a theoretical background section on human-AI interaction, the roles of SD, and the importance of narratives in SD. This is followed by a research method section that illustrates the approach to collecting and analyzing the research data. Finally, a findings and discussion section is presented, which details the five narrative themes and their methodological implications.

---

## 2. Theoretical Background

In this section, the issues of human-AI interaction that require attention in designing intelligent agents responsive to human values are outlined. Subsequently, the strengths and current limitations of SD with respect to its operational basis are illustrated. Thereafter, the role of narratives in SD are described and the impetus for identifying the narrative themes concerning the interaction between humans and intelligent agents is established.

### 2. 1. Issues of Human-AI Interaction

#### 2. 1. 1. Human and AI Agency

Barandiaran, Di Paolo, and Rohde (2009) define three conditions for an object to be considered as having agency:

- **Individuality:** An object must be a distinguishable entity that differs from its environment.
- **Interactional Asymmetry:** An object must be the active source of activity in its environment.
- **Normativity:** An object must regulate this activity in relation to certain norms.

In the field of socio-technical systems and social sciences, there are varied theoretical positions toward human and machine agency. In actor-network theory (ANT) (Law, 1992), the human and machine are considered to have symmetrical (equal) effects in a network. In contrast, the double dance of agency model (Rose & Jones, 2005) and activity theory (Kaptelinin & Nardi, 2006) view the human and machine agency as asymmetrical, with only humans having intentionality and machine agency being shaped through human intent. Despite their differences in understanding the symmetry between human and machine agency, all three theories reject the anthropocentric perspective of limiting agency to humans. They acknowledge that machines are key players in human-machine interaction. They maintain human and machine agency to be understood as an assemblage and suggest examining socio-technical issues from the perspective of both human values and technology attributes.

The aforementioned constructs can be noted as especially important in the context of human-AI interaction. AI, according to a general definition, is technology that performs context detection and provides information or services to a user (Groce et al., 2013). As an intelligent agent undergoes the cycle of learning and adapting, it has the potential to act as an independent agent (Yearsley, 2017) and shift the dynamic between human and AI (Legaspi, He, & Toyozumi, 2019). In order to avoid negative or unwanted UX due to the learning ability of AI, AI systems need to be designed to update and adapt with caution, inform the user on the consequences of user actions that the AI learns, and notify the user about changes (Amershi et al., 2019).

#### 2. 1. 2. User's Data Sharing

User data (i.e., data related to or generated by the user) needs to be fed to the AI system for the system to produce maximal benefits for the user. The system not receiving adequate user data can risk unwanted outcomes such as unfairness and data bias (Leese, 2014; Trewin et al., 2019). Therefore, the issue of user's data sharing needs to be handled with care to prevent

such happenings (Saffarizadeh, Boodraj, & Alashoor, 2017). As user data can become a target to AI-related crimes such as data theft/fraud/forgery and impersonation (King et., 2020), AI systems need to be implemented with safeguards that monitor and prevent such misconduct for the user's safety (Genpact, 2017). At the same time, the user need to be notified of the benefits he/she will receive from interacting with AI-embedded products (Ostrom, Fotheringham, & Bitner, 2019). It is essential to identify the types of data that the user is willing to share and understand his/her expectations when sharing data with intelligent agents.

### **2. 1. 3. AI's Transparency and Explainability**

The transparency and explainability of AI relates to how people understand the inner-workings of AI systems. It affects how the user works with/control an AI system and weigh its level of accessibility, trustworthiness, fairness, and privacy protection (Arrieta et al., 2020). Intelligent agents have now reached a stage wherein they are capable of making decisions for or instead of the user. On the one hand, such ability has potential to add convenient value to AI-related UX. On the other hand, the user may also take this as frustrating and meddlesome, especially when he/she cannot understand why or how the decisions were made for them (Castelli et al., 2017). To ensure that such negative UX does not occur, AI products and systems should be designed in a way that allows the user to easily understand their functioning.

### **2. 1. 4. AI's Sociality**

AI operates differently from human intelligence as it takes a mathematical approach to compute analytical results. Intelligent agents may not follow social norms and cause the user to perceive AI as intrusive and unsettling (Ostrom et al., 2019). To prevent such problems, intelligent agents are commonly designed to have sociality – the ability of following human rules of social interaction (Purington, Taft, Sannon, Bazarova, & Taylor, 2017). The degree or type of sociality the user prefers from intelligent agents may vary, so the sociality of intelligent agents needs to be carefully managed. Hence, researchers need to question when and how such sociality affects the user's experience of interacting with intelligent agents (Liao, Davis, Geyer, Muller, & Shami, 2016).

### **2. 1. 5. User's Autonomy**

It has been reported that machine automation can hinder the user's desires of self-expression, confirming to self-identity, and maintaining self-autonomy (Ostrom et al., 2019). In such situations, the user not only needs to be supported by intelligent agents but also needs to perform their desired goals with their own capabilities (Leung, Paolacci, & Puntoni, 2018). In other words, the user will benefit from intelligent agents when intelligent agents augment the user's desired goals and do not substitute the user's identity (Pew Research Center, 2018). The user should be able to control intelligent agents so that he/she feels a sense of independence within his/her daily life (Fong, Indulska, & Robinson, 2011). It is necessary to understand the user's needs of increasing, maintaining, or lessening their control of intelligent agents and design AI systems in such a way that they respond to the user needs.

## 2. 2. Speculative Design for Creating a Discursive Space

Design itself is a “fundamentally imaginative act that involves picturing the world as other than it is” (Blythe, 2017). If design is already an activity of envisioning new possibilities, it would be necessary to define what is meant by “speculative” in SD.

Lindley and Potts (2014) maintain that most designs and prototyping explore an optimal solution, whereas SD involves envisioning a plurality of futures that portray human experience with technology ranging from preferable to undesirable. Such envisioning allows us to examine the various possibilities, and at the same time, question the current reality and its inherent biases of technology design (DiSalvo, 2012; Lindley, 2015; Oogjes & Wakkary, 2017; Wakkary, Odom, Hauser, Hertz, & Lin, 2015). The use of SD essentially opens up discussion about the values and politics entangled in human-technology interaction (Wong & Khovanskaya, 2018). Thereby, it is possible to generate design insights related to the requirements for preferable futures (Dunne & Raby, 2013) and technology adoption (Lindley et al., 2017).

Despite the consensus among scholars on how design can be speculative, there is a need to establish an operational basis of SD that allows it to be more accessible to researchers (Pierce et al., 2015). To address this need for knowledge-building, narratives in SDs have been highlighted as crucial for making sense of SD, thereby making it more accessible.

## 2. 3. Narrative Themes in Speculative Design

Narrative is the prime device that helps the user comprehend what is proposed by speculative artifacts. The narratives in SD are created through the use of literary devices such as transgression of norms, provocation, satire, and staging of dilemmas (Bardzell & Bardzell, 2013). These devices evoke suspense (Vogler, 2007), contempt, shock, and righteous indignation (Malpass, 2013) and therefore stimulate discussion on the social, cultural, and political implications of technology. As such, researchers should consider narratives with utmost importance when taking a SD approach.

Prior HCI and design studies have attempted to build knowledge on narratives in SD by analyzing SD exemplars. The meta-analyses present SD plots that encourage discussion on human-technology interaction (Blythe, 2017), techniques that help create believable and engaging SD artifacts (Auger, 2013), strategies that evoke imagination on technology (Knutz, Lenskjold, & Markussen, 2016), and semantic themes/tactics for the practice of SD (Ferri, Bardzell, Bardzell, & Louraine, 2014). These scholarly works establish the importance of narratives in SD and clarify approaches to developing narratives that generate design insights related to the social implications of emerging technology. However, an extensive literature search yielded no research cases that explored narrative issues targeted toward AI. Human-AI interaction faces difficult questions of socio-technical issues, and using SD is a viable pathway to exploring these questions. To address this gap in HCI and design research, the research aimed at clarifying narrative themes in SD that problematize socio-technical issues of human-AI interaction. Hence, it was chosen to collect and analyze HCI research cases that examine human-AI interaction through the use of SD.

---

### 3. Research Method

This section illustrates our approach to identifying the narrative themes in SD for human-AI interaction, which consists of two stages: (1) case collection and (2) analysis. The SD cases collected for analysis were HCI research cases that empirically examined human-AI interaction through the use of SD. Research through SD cases were chosen as the material for analysis because they detailed the type of human-AI interaction issues targeted for investigation, the method of using SD, and the provoked discussions from the use of SD, all of which were information needed for identifying the narrative themes in SD for AI. Since there exist multiple ways of utilizing SD for Human-AI interaction, it would be difficult to find different narrative themes of SD through a single empirical study. Examining multiple existing empirical research cases would be better suited to identifying various types of SD narrative themes for AI. Although this method has a limitation of lacking affluent empirical data of grounded theory or phenomenological research, it can be used as a starting point for building an effective SD for human-AI interaction.

#### 3. 1. Case Collection

The ACM digital library was used to search for the research cases. The search range for publication year was set to post-2010 to obtain the latest research cases. Both SD-related (3) and AI-related (15) keywords were used for search: “speculative design,” “design fiction,” “critical design” AND “artificial intelligence,” “machine learning,” “AI,” “chatbot,” “robot,” “autonomous,” “smart home,” “smart devices,” “smart object,” “smart city,” “smart technology,” “analytics,” “voice assistant,” “monitoring technology,” “sensing technology.” The definition of AI by Kaplan and Haenlein (2019) was used to select the AI-related keywords: AI is a system’s ability to acquire, interpret, learn, and use external data. Therefore, the selection criteria were expanded to include the keywords related to these system abilities.

45 keyword pairs were created by combining one SD with one AI keyword (3 x 15). The title and keyword sections of the first 100 search results per keyword pair were scanned (approximately 4,500 cases in total), most of which were found not to be cases of research through SD. Therefore, it was necessary to introduce tighter selection criteria to ensure the relevance of the keyword pairs to the main topic of the research cases: the title or keyword section of the research articles would include at least one SD AND one AI relevant keyword. By using this filter, 22 research cases were selected. Table 1 details the published year and title of the 22 research cases.

Table 1 Description of Research Cases

No.	Title
1	Rudiments 1, 2 & 3: Design Speculations on Autonomy
2	Chatbots of the Gods: Imaginary Abstracts for Techno-spirituality Research
3	A Machine Learning: An Example of HCI Prototyping with Design Fiction
4	NewSchool: Studying the Effects of Design Fiction through Personalized Learning Scenarios
5	Operationalising Design Fiction for Ethical Computing
6	Homes for Life: A Design Fiction Probe
7	Using Design Fiction to Reflect on Autonomy in Smart Technology for People Living with Dementia
8	Infrastructures of the Imagination: Community Design for Speculative Urban Technologies
9	Real-fictional Entanglements: Using Science Fiction and Design Fiction in Interrogate Sensing Technologies
10	Futuristic Autobiographies: Weaving Participant Narratives to Elicit Values around Robots
11	Near Future Cities of Things: Addressing Dilemmas through Design Fiction
12	Eyespy: Designing Counterfunctional Smart Surveillance Cameras
13	Ad Empathy: A Design Fiction
14	Intimate Futures: Staying with the Trouble of Digital Personal Assistants through Design Fiction
15	The Adventures of Older Authors: Exploring Futures through Co-design Fictions
16	Judgment Call the Game: Using Value Sensitive Design and Design Fiction to Surface Ethical Concerns Related to Technology
17	A World Following Farmer Almanac: Speculation on Lifestyle Interweaving Folk Religion and Smart Home
18	Understanding Parents' Perspectives on Mealtime Technology
19	I Beg to Differ: Soft Conflicts in Collaborative Design Using Design Fictions
20	Designing an Escape Room in the City for Public Engagement with AI-enhanced Surveillance
21	Hawkeye – Deploying a Design Fiction Probe
22	Our Friends Electric: Reflections on Advocacy and Design Research for the Voice Enabled Internet

### 3. 2. Analysis

For analysis, the constant comparative method – a qualitative data analysis process that incorporates data collection, inductive coding, categorization, and comparison (Glaser, 1965) – was employed to conceptualize the narrative themes of SD. Three coders experienced in qualitative research, with backgrounds including UX for AI and probes/generative toolkits, carried out the analysis. The three coders were selected to analyze the data from a HCI/design methods view and increase impartiality. The coders followed a four-step process to analyze the selected cases. Figure 1 illustrates the process.

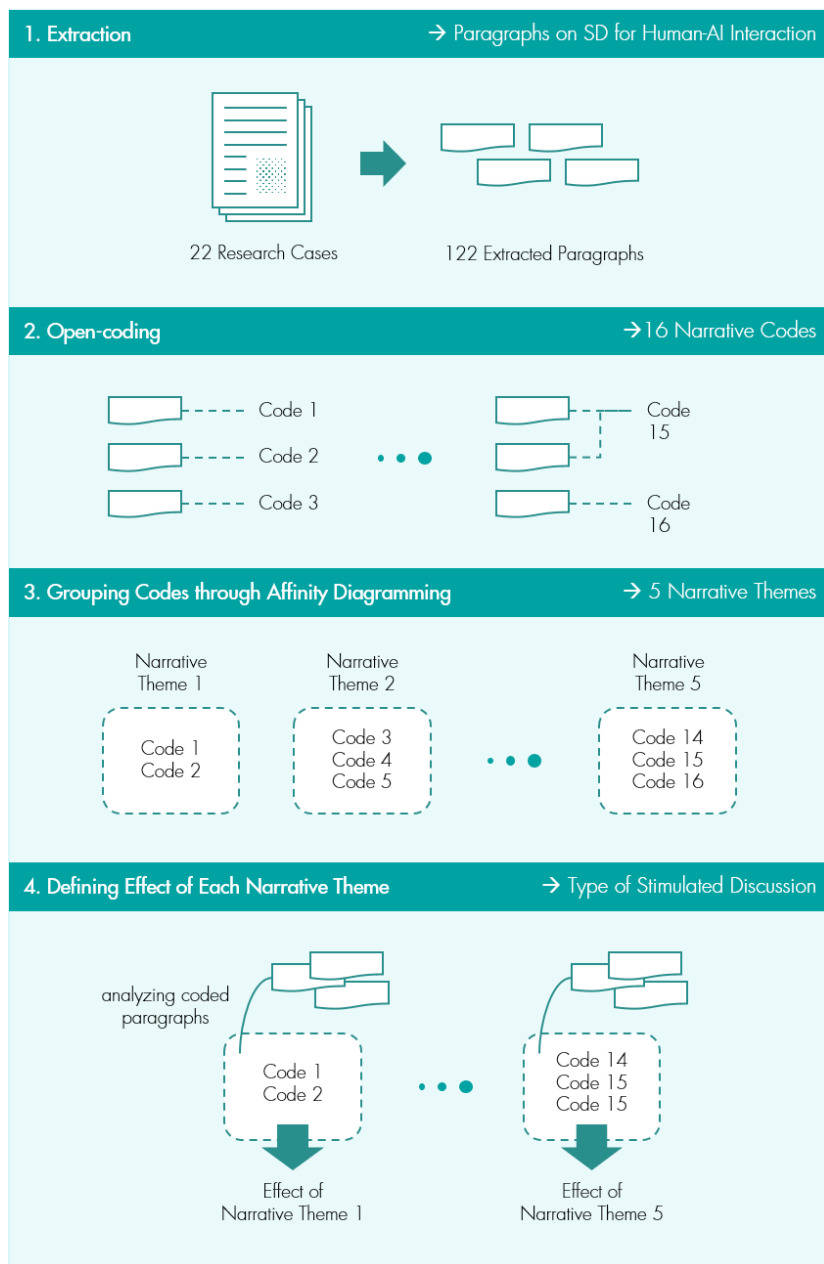


Figure 1 Analysis Process

First, the coders extracted paragraphs from the selected cases that related to the narratives used for envisioning human-AI interaction. This was to ensure a more efficient and comprehensive understanding of the cases. A total of 122 paragraphs were extracted. The extracted paragraphs were analyzed in the context of their corresponding cases, and therefore, the unit of analysis was the individual research case.

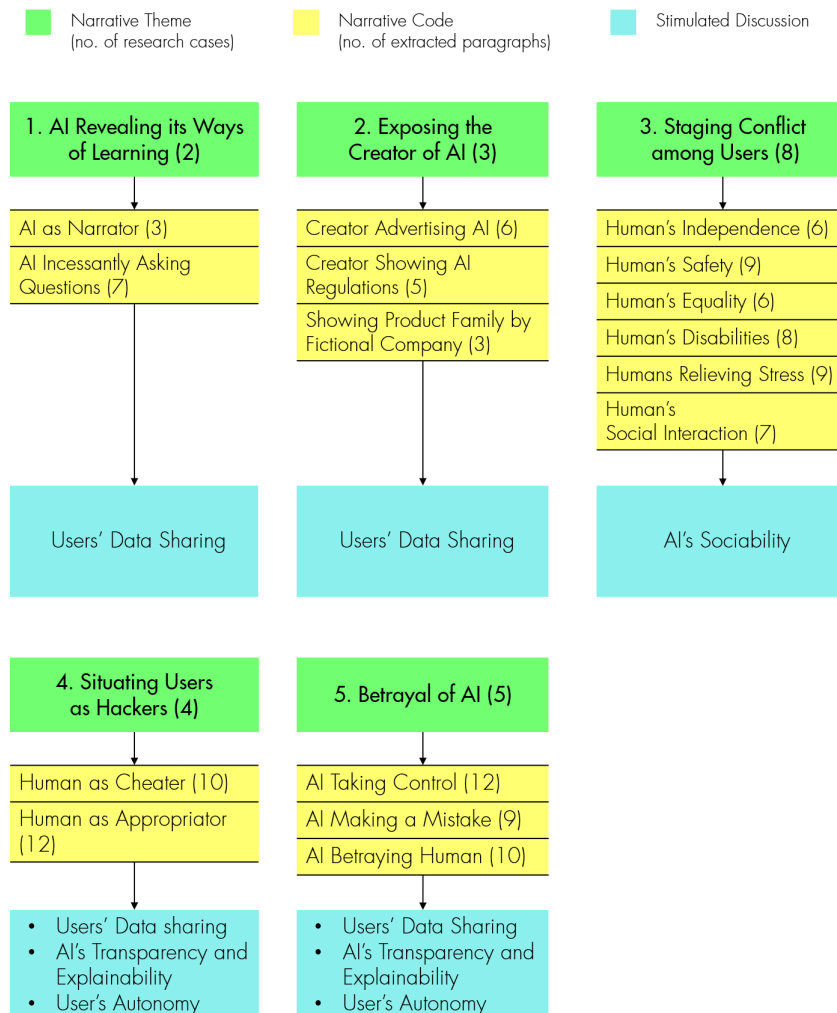
Second, the coders carried out thematic coding on the extracted paragraphs based on the narratives used for describing human-AI interaction (e.g., AI as a narrator, AI betraying the user). Consequently, 16 narrative codes were coded and labeled.

Third, the coders used affinity diagramming to group the 16 narrative codes and form



higher-order categories, namely, narrative themes. The coders grouped the 16 narrative codes based on their similarity of how the human or AI agency was illustrated. As a result, five narrative themes were found.

Fourth, the coders re-examined the extracted contents that referred to each narrative theme to define its main effect of use. This resulted in identifying the effects for each narrative theme, which all related to stimulating discussion on specific issues of human-AI interaction. Figure 2 shows the hierarchy among the identified five narrative themes and 16 narrative codes, the effect (stimulated discussion) produced through the use of the narrative themes, the number of research cases that used each narrative theme, and the number of paragraphs coded to each narrative code.



**Figure 2** Summary of narrative themes

---

## 4. Findings and Discussion

By conducting a constant comparative analysis, five narrative themes in SD were identified. The narrative themes were found to provoke discussion on the issues of human-AI interaction that are presented in the theoretical background section of this paper. This section illustrates the five narrative themes, provides discussions on the merits of each narrative theme, and highlights the scope for future studies on making SD more accessible in AI context. Figure 2 shows the narrative themes, the narratives belonging to each theme, and each theme's following effects-of-use (stimulated discussion).

### 4. 1. AI Revealing its Ways of Learning

This narrative theme involves creating SD artifacts that showcase the contextual learning process of intelligent agents. The theme was found to provoke discussion and help generate design insights related to user's data sharing with intelligent agents. For example, Lindley and Potts (2014) created a speculative video, positioning an artificially intelligent device as the narrator in order to "invite the audience into the world of a computer that is capable of contextual learning with subtlety, and thus allowing space for heuristic interpretation." This setting highlighted the intelligent agent's needs to learn and be offered more situation-specific data and stimulated discussion about showing contextually rich data to AI.

Knutz et al. (2016) mention narrative anthropomorphism, a SD strategy, which is to give voice to technology in order to understand complex ecologies. AI Revealing its Ways of Learning aligns with narrative anthropomorphism in the sense that the narrative in SD is crafted so that the technology's point of view is made apparent. Such approach to SD has much value in the context of AI because, users' data sharing relates to intelligent agents using their learning abilities to offer more personalized and beneficial events to users and users' reluctance of sharing data (Saffarizadeh et al., 2017). Creating narratives that shift the focus from user to technology can be a powerful approach when aiming to generate design insights related to the issue of user's data sharing.

### 4. 2. Exposing the Creator of AI

Exposing the Creator of AI is a narrative theme that presents the creator of intelligent agents and the ideology that actuates AI decision-making. Using this theme was identified to motivate conversation about the responsibilities that the creators of intelligent agents should have for user's data protection. Skirpan and Fiesler (2018) created a speculative artifact in the form of an advertisement, which shows a fictitious company (the creator) promoting their AI-based marketing solution. The intelligent agent's creator is directly exhibited and therefore AI products are framed as entities that operate for the profit of companies and not just for user needs, and therefore grounded "debates around fair use of data, and the boundaries of ethical design" (Ibid).

Preexisting meta-analyses of SD such as Blythe (2017) share findings on narratives that concern the user's journey of utilizing interactive products and services. However, Exposing the Creator of AI suggests that there is value in breaking away from centering the user in narratives when studying human-AI interaction. The use of Exposing the Creator of AI provokes a thought-process on the ulterior motives and hidden ideologies sown by

the creators of AI products. The SD artifacts that embody this narrative theme drives conversation on the ethical use of AI and the design of safeguards for user data protection. Therefore, HCI and design researchers could signal to Exposing the Creator of AI to spark more in-depth conversations and to make an inquiry into the issues surrounding users' data sharing.

#### **4. 3. Staging Conflict among Users**

Staging Conflict among Users represents narratives of which intelligent agents cause or aggravate conflict among users. The use of this theme was found to give voice to the issues on AI's sociality, i.e., the ability of intelligent agents to assimilate the varying user desires, beliefs, opinions, and values. In the case of Schulte, Marshall, and Cox (2016), the researchers created a speculative scenario depicting a situation wherein senior citizens desire independence and their children want real-time notifications about their parents from AI care-givers. The scenario was constructed to "define a group of actors of the story between which potential conflicts play out: to articulate what the world should look like in which these technologies are expected to live" (Ibid). The scenario stimulated discussion about the conflicting values of stakeholders and how the values should be addressed through the design of AI's sociality.

The work of Ferri et al. (2014) was the only pre-existing SD meta-analysis found to touch on the concept of conflict. It reports a speculative design tactic named social transgression, which is to change the use context of everyday objects and to stage a conflict between the changed state and status quo. Staging Conflict among Users makes further use of conflict by not only portraying new possibilities but also representing the different values of users clashing due to AI. As such, Staging Conflict among Users reinforces the acknowledgment of multitude of values about specific situations. These conflicts and their following dialogs contain and represent the varying values and perspective of users. As such, they can be rich resources for generating design insights related to the social norms that AI need to follow and inform the design of sociable AI.

#### **4. 4. Situating Users as Hackers**

This narrative theme refers to narratives in SD that position users as hackers of intelligent agents, which in turn encourages discussions on user's data sharing, AI's transparency and explainability, and users' autonomy. To prevent confusion, the concept of hacking in the research cases related to Situating Users as Hackers did not concern the "cultures of making" (Bardzell, Bardzell, & Toombs, 2014) and was specific to the activities of overriding computer systems. Derboven and Vandenberghe (2016) exemplify the use of Situating Users as Hackers and created a speculative scenario in which students hacked into an AI-based learning analytics system. In the scenario, the students find a way to feed the AI system with false data so that they can receive higher school records. The scenarios provoked discussion on the data biases of AI systems and promote design that "take into account the messy reality of unmotivated students, platform misuse, and discontinuous data gathering" (Ibid).

Auger (2013) states that it is only possible to achieve design criticality through SD when the narratives in SD are understandable and not far-fetched from the audience's perception of their world. Based on his arguments, Auger proposes SD techniques to building

understandable SD narratives. However, AI remain as black-boxes to users (Knight, 2016), so in itself, AI is hard to understand let alone it being portrayed through an imaginative narrative. Situating Users as Hackers adds understanding to the practice of constructing understandable SD artifacts concerning human-AI interaction. Situating Users as Hackers allows users to gain a heuristic understanding on the inner-workings of AI and the consequences of AI misuse. HCI and design researchers could build upon the concept of Situating Users as Hackers and build SD artifacts that display narratives wherein AI needs to be overcome by users. SD artifacts that reflect such narratives could increase the awareness of what is happening inside intelligent agents and deepen the understanding of supporting user autonomy and AI fairness.

#### 4. 5. Betrayal of AI

The theme of Betrayal of AI involves rendering situations wherein intelligent agents seem to help users at first, but misguides them at the end. Staging such narrative stimulated discussions on user's data sharing, AI's transparency and explainability, and users' autonomy. The research case of Søndergaard and Hansen (2018) made use of Betrayal of AI by creating a speculative video that depicts a smart toilet assistant. The smart assistant reliably helps the user track her menstrual cycle until it makes a mistake with the user's birth control. This speculative video engaged the audience to question the inner-workings of intelligent agents and promoted discussions on "which biases and conflicts these intimate algorithmic conversations [between user and AI] might foster" (Ibid).

It is mentioned that a narrative polarized to one perspective has the potential problem of oversimplifying matters of concern and ignoring many intermediate viewpoints (Vogler, 2007). Betrayal of AI can lessen such problems. Speculative artifacts that embedded Betrayal of AI showcased both optimistic and dismal imaginaries of intelligent agents. The narrative theme displays the duality of intelligent agents to the audience. Therefore, it gives the audience the opportunity to think about the plurality of possible futures in human-AI interaction. These imaginations can encompass a variety of beliefs, values, ideals, and fears and "open new perspectives on the challenges facing us" (Dunne & Raby, 2013). The narrative theme can bring about discussions on how the contrasting aspects of intelligent agents should be mediated (e.g., user's autonomy vs. benefits of machine autonomy). A narrative portrayal of intelligent agents' duality can support the understanding of the possible and preferable states of AI.

---

### 5. Conclusion

This study established the narrative themes in SD concerning AI, which has its own unique characteristics. We have attempted to accomplish this by content-analyzing 22 empirical research cases that employed SD to gain an understanding of human-AI interactions.

The five narrative themes identified through this study tend to exaggerate the issue of human-AI interaction. Such portrayals of human-AI interactions showed interactions beyond the existing social norms regarding the relationship between human and machine, as shown by AI agency acting on its own, going beyond the social norm of machine being controlled

by humans. The narrative themes can help craft the speculation on various human-AI interactions wherein humans are knowingly or unknowingly affected by AI due to overt AI agency, lack of AI transparency and explainability, and disregard of user's autonomy. They can also be used to envision speculations that involve users willingly sharing their data with AI and AI assimilating to user needs and values. These envisions can help define desirable states of human-AI interactions of which humans and AI organically communicate with each other and recognize what the other needs in order for the AI to deliver utility and value to humans.

The findings of this study add understanding to the growing field of critical thinking in HCI research and provide insights into developing more ready-to-use methodological devices that can stimulate the discourse around the relationships between human experience and AI. However, the findings and discussions may not be fit for contextualizing other areas of technology (e.g., augmented/virtual reality). Future studies will be needed to further clarify the ways of utilizing SD and building narratives in the context of various emerging technologies.

## References

1. Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for Human-AI Interaction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. doi:<https://doi.org/10.1145/3290605.3300233>
2. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., & Benjamins, R. (2020). Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges Toward Responsible AI. *Information Fusion*, 58, 82–115.
3. Auger, J. (2013). Speculative Design: Crafting the Speculation. *Digital Creativity*, 24(1), 11–35.
4. Barandiaran, X. E., Di Paolo, E., & Rohde, M. (2009). Defining Agency: Individuality, Normativity, Asymmetry, and Spatio-temporality in Action. *Adaptive Behavior*, 17(5), 367–386. doi:[10.1177/1059712309343819](https://doi.org/10.1177/1059712309343819)
5. Bardzell, J., & Bardzell, S. (2013). What is Critical About Critical Design? *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. doi:<https://doi.org/10.1145/2470654.2466451>
6. Bardzell, J., Bardzell, S., & Toombs, A. (2014). "Now That's Definitely a Proper Hack": Self-made Tools in Hackerspaces. *Proceedings of the 2014 CHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/2556288.2557221>
7. Blythe, M. (2017). Research Fiction: Storytelling, Plot and Design. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery. doi:<https://doi.org/10.1145/3025453.3026023>
8. Castelli, N., Ogonowski, C., Jakobi, T., Stein, M., Stevens, G., & Wulf, V. (2017). What Happened in my Home? An End-User Development Approach for Smart Home Data Visualization. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery. doi:<https://doi.org/10.1145/3025453.3025485>
9. De Beaugrande, R. (1982). The Story of Grammars and the Grammar of Stories. *Journal of Pragmatics*, 6(5–6), 383–422.
10. Derboven, J., & Vandenberghe, B. (2016). NewSchool: Studying the Effects of Design Fiction through Personalized Learning Scenarios. *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI '16)*. doi:<https://doi.org/10.1145/2971485.2993926>
11. DiSalvo, C. (2012). FCJ-142 Spectacles and tropes: Speculative Design and Contemporary Food Cultures. *The Fibreculture Journal* (20 2012: Networked Utopias and Speculative Futures).

12. Dunne, A., & Raby, F. (2013). *Speculative Everything: Design, Fiction, and Social Dreaming*. The MIT Press.
13. Ferri, G., Bardzell, J., Bardzell, S., & Louraine, S. (2014). Analyzing Critical Designs: Categories, Distinctions, and Canons of Exemplars. *Proceedings of the 2014 Conference on Designing Interactive Systems (DIS '14)*. doi:<https://doi.org/10.1145/2598510.2598588>
14. Fong, J., Indulska, J., & Robinson, R. (2011). A Preference Modelling Approach to Support Intelligibility in Pervasive applications. *Proceedings of the 2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*.
15. Frayling, C. (1993). *Research in art and design*.
16. Genpact. (2017). The Consumer: Sees AI Benefits But Still Prefers the Human Touch. Retrieved from <https://www.genpact.com/downloadable-content/the-consumer-sees-ai-benefits-but-still-prefers-the-human-touch.pdf>
17. Glaser, B. G. (1965). The Constant Comparative Method of Qualitative Analysis. *Social problems*, 12(4), 436-445.
18. Groce, A., Kulesza, T., Zhang, C., Shamasunder, S., Burnett, M., Wong, W.-K., & Bice, F. (2013). You Are the Only Possible Oracle: Effective Test Selection for End Users of Interactive Machine Learning Systems. *IEEE Transactions on Software Engineering*, 40(3), 307-323.
19. Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in My Hand: Who's the Fairest in the Land? On the Interpretations, Illustrations, and Implications of Artificial Intelligence. *Business Horizons*, 62(1), 15-25.
20. Kaptelinin, V., & Nardi, B. A. (2006). *Acting with Technology: Activity Theory and Interaction Design*. The MIT press.
21. King, T. C., Aggarwal, N., Taddeo, M., & Floridi, L. (2020). Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions. *Science and Engineering Ethics* 26, 89-120. doi: <https://doi.org/10.1007/s11948-018-00081-0>
22. Knight, W. (2016). The Dark Secret at the Heart of AI. Retrieved from <https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/>.
23. Knutz, E., Lenskjold, T. U., & Markussen, T. (2016). Fiction as a Resource in Participatory Design. *Proceedings of DRS 2016 International Conference: Future-Focused Thinking*.
24. Law, J. (1992). Notes on the Theory of the Actor-Network: Ordering, Strategy, and Heterogeneity. *Systems Practice*, 5(4), 379-393.
25. Leese, M. (2014). The New Profiling: Algorithms, Black Boxes, and the Failure of Anti-Discriminatory Safeguards in the European Union. *Security Dialogue*, 45(5), 494-511.
26. Legaspi, R., He, Z., & Toyoizumi, T. (2019). Synthetic Agency: Sense of Agency in Artificial Intelligence. *Current Opinion in Behavioral Sciences*, 29, 84-90.
27. Leung, E., Paolacci, G., & Puntoni, S. (2018). Man Versus Machine: Resisting Automation in Identity-based Consumer Behavior. *Journal of Marketing Research*, 55(6), 818-831.
28. Liao, Q. V., Davis, M., Geyer, W., Muller, M., & Shami, N. S. (2016). What Can You Do? Studying Social-Agent Orientation and Agent Proactive Interactions with an Agent for Employees. *Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS '16)*. doi:<https://doi.org/10.1145/2901790.2901842>
29. Lindley, J. (2015). A Pragmatics Framework for Design Fiction. *Proceedings of the 11th European Academy of Design Conference*.
30. Lindley, J., Coulton, P., & Sturdee, M. (2017). Implications for Adoption. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. *Association for Computing Machinery*. doi:<https://doi.org/10.1145/3025453.3025742>
31. Lindley, J., & Potts, R. (2014). A Machine Learning: An Example of HCI Prototyping with Design Fiction. *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational (NordiCHI '14)*. doi:<https://doi.org/10.1145/2639189.2670281>
32. Malpass, M. (2013). Between Wit and Reason: Defining Associative, Speculative, and Critical Design in Practice. *Design and Culture*, 5(3), 333-356.

33. Oogjes, D., & Wakkary, R. (2017). Videos of Things: Speculating on, Anticipating and Synthesizing Technological Mediations. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. doi:<https://doi.org/10.1145/3025453.3025748>
34. Ostrom, A. L., Fotheringham, D., & Bitner, M. J. (2019). Customer Acceptance of AI in Service Encounters: Understanding Antecedents and Consequences. *Handbook of Service Science*, 2, 77–103.
35. Pew Research Center. (2018). AI and the Future of Humans: Experts Express Concerns and Suggest Solutions. Retrieved from <https://www.pewresearch.org/internet/chart/ai-and-the-future-of-humans-experts-express-concerns-and-suggest-solutions/>
36. Pierce, J., Sengers, P., Hirsch, T., Jenkins, T., Gaver, W., & DiSalvo, C. (2015). Expanding and Refining Design and Criticality in HCI. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. doi:<https://doi.org/10.1145/2702123.2702438>
37. Purington, A., Taft, J. G., Sannon, S., Bazarova, N. N., & Taylor, S. H. (2017). "Alexa is My New BFF": Social Roles, User Satisfaction, and Personification of the Amazon Echo. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*. doi:<https://doi.org/10.1145/3027063.3053246>
38. Rose, J., & Jones, M. (2005). The Double Dance of Agency: A Socio-theoretic Account of How Machines and Humans Interact. *Systems, Signs & Actions*, 1(1), 19–37.
39. Saffarizadeh, K., Boodraj, M., & Alashoor, T. M. (2017). Conversational Assistants: Investigating Privacy Concerns, Trust, and Self-disclosure. *Proceedings of the International Conference on Information Systems*.
40. Schulte, B. F., Marshall, P., & Cox, A. L. (2016). Homes for Life: A Design Fiction Probe. *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI '16)*. doi:<https://doi.org/10.1145/2971485.2993925>
41. Skirpan, M., & Fiesler, C. (2018). Ad Empathy: A Design Fiction. *Proceedings of the 2018 ACM Conference on Supporting Groupwork*. doi:<https://doi.org/10.1145/3148330.3149407>
42. Smarr, J. L. (1979). Some Considerations on the Nature of Plot. *Poetics*, 8(3), 339–349.
43. Søndergaard, M. L. J., & Hansen, L. K. (2018). Intimate Futures: Staying with the Trouble of Digital Personal Assistants through Design Fiction. *Proceedings of the 2018 Designing Interactive Systems Conference*. doi: <https://doi.org/10.1145/3196709.3196766>
44. Trewin, S., Basson, S., Muller, M., Branham, S., Treviranus, J., Gruen, D., & Manser, E. (2019). *Considerations for AI Fairness for People with Disabilities*, 5. Association for Computing Machinery.
45. Vogler, C. (2007). *The Writer's Journey*. Michael Wiese Productions.
46. Wakkary, R., Odom, W., Hauser, S., Hertz, G., & Lin, H. (2015). Material speculation: Actual Artifacts for Critical Inquiry. *Proceedings of The Fifth Decennial Aarhus Conference on Critical Alternatives*.
47. Wong, R. Y., & Khovanskaya, V. (2018). Speculative Design in HCI: From Corporate Imaginations to Critical Orientations. *New Directions in Third Wave Human-Computer Interaction*, 2, 175–202.
48. Yearsley, L. (2017). *We Need to Talk about the Power of AI to Manipulate Humans*. MIT Technology Review.