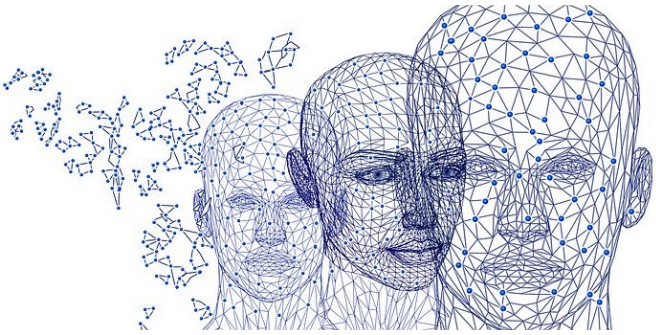
NOVEMBER 2020, ISSUE CODE NL-2021-2

DDD (Digital Data Deception) Technology Watch Newsletter

Table of Contents



"All warfare is based on deception. Hence, when we are able to attack, we must seem unable; when using our forces, we must appear inactive; when we are near, we must make the enemy believe we are far away; when far away, we must make him believe we are near."

— Sun Tzu, *The Art of War*

Editors: Enes Altuncu, Virginia Franqueira, Sanjay Bhattacherjee and Shujun Li
Affiliation: Kent Interdisciplinary Research Centre in Cyber Security (KirCCS), University of Kent, UK
Contact Us: `ddd-newsletter@kent.ac.uk`

© 2020 University of Kent, UK

Editorial

In this second issue of the Digital Data Deception (DDD) Technology Watch Newsletter, we decided to mainly focus on an important topic for DDD: *conversational agents* or *chatbots*. Although different definitions exist for these two terms, for the purpose of this newsletter we decided to use these terms interchangeably to avoid unnecessarily complicating related concepts.

In recent years, conversational agents have become more pervasive in our lives. They can be text-based customer service agents, technical support agents or even voice-enabled cyber humanoid with face, gestures, personality traits and background. They are used for a wide range of tasks (e.g., to conduct interviews, make payments, and advise on purchases) but, most importantly, for social purposes as well (e.g., counselling, chatting, and coaching). Despite great benefits, conversational agents can bring a number of failures such as racist and partisan rhetoric conveyed by chatbots which have been observed (`https://www.bbc.co.uk/news/technology-54718671`). The training dataset used to drive a conversational agent may be directly or indirectly deceiving, e.g., Google announced that their Meena generalist chatbot has been trained with data "filtered from public domain social media conversations" (`https://ai.googleblog.com/2020/01/towards-conversational-agent-that-can.html`). Moreover, leveraging their potential of persuasion and the easy access by anyone to the technology (`https://www.clearvoice.com/blog/build-facebook-chatbot-10-minutes/`), rogue chatbots can become the next "great" tool for criminals to launch more sophisticated social engineering attacks (`https://www.cmswire.com/digital-workplace/is-your-enterprise-ready-to-fight-off-rogue-chatbots/`).

The malicious applications of conversational agents for DDD are the main reason why we decided to choose the topic as this issue's main focus. This topic also nicely connects three important areas of DDD: AI, Psychology, and Human-machine Teaming. Such connections will be reflected from the papers covered in this issue.

In addition to research work directly related to conversational agents, in this issue we also include a number of papers broadly related to conversational agents and DDD, covering topics such as deceptive social bots, human detection of deception and machine-generated content, and human and automatic credibility assessment.

This issue is based on 29 research papers published since 2019, which were identified and selected following a venue-driven systematic literature review (SLR) approach: (1) for a number of target topics related to conversational agents and DDD a number of highly relevant scientific journals, conferences and workshops were identified; (2) all papers published at the selected venues since 2019 were manually inspected to identify a number of candidate papers; (3) at least two independent members of the DDD editorial team checked each candidate paper for their relevance; (4) the whole editorial team met and discussed to agree on the final list of papers for inclusion in the newsletter. Since we decided to focus on conversational agents in this issue, some selected papers (e.g., a number of papers on recommendation systems) are reserved for inclusion in a future issue. In addition to papers identified from the above venue-driven approach, we also included a relevant book chapter [18] from a 2019 book *The Palgrave Handbook of Deceptive Communication* (`https://doi.org/10.1007/978-3-319-96334-1`), which was discovered in our ad hoc searches for related work. More chapters of this book will be considered in future issues of the newsletter.

We hope you enjoy reading this issue. Feedback is always welcome, and should be directed to `ddd-newsletter@kent.ac.uk`.

List of Acronyms

We list all acronyms used in this newsletter as a useful index for reference purposes.

- Adjusted Rand Index (ARI)
- Artificial Intelligence (AI)
- Clustering Coverage (CC)
- Conditional Random Fields (CRF)
- Conversational Topic Suggestion (CTS)
- Convolutional Neural Network (CNN)
- End-User Development (EUD)
- Fuzzy-set Qualitative Comparative Analysis (fsQCA)
- Generative Adversarial Network (GAN)
- Generative Pre-trained Transformer (GPT)
- Hierarchical Attention Networks (HAN)
- High Self-Disclosure (HD)
- Intelligent Personal Agent (IPA)
- Interactive Emotional Dyadic Motion Capture (IEMOCAP)
- Linguistic Inquiry and Word Count (LIWC)
- Long Short-Term Memory (LSTM)
- Low Self-Disclosure (LD)
- Miami University Deception Detection Database (MU3D)
- Multi-Layer Perceptron (MLP)
- Natural Language Generation (NLG)
- Natural Language Processing (NLP)
- Natural Language Understanding (NLU)
- Non Self-Disclosure (ND)
- Random Forest (RF)
- Recurrent Neural Network (RNN)
- Reinforcement Learning (RL)
- Spoken Dialog System (SDS)
- Status Quo Bias Perspective (SQBP)
- Text-To-Speech (TTS)

© 2020 University of Kent, UK

Three Papers with Broad Relevance

Introduction

We start with three papers with broad relevance for the focus theme of this issue: conversational agents. The first paper discusses how intelligent systems that are capable of socially interacting with humans (e.g., conversational agents) can use their power to pursue different types of deception by exploiting humans' intrinsic vulnerabilities. The other two papers look at recent research developments on how humans detect deception, which can provide useful insights on how to design more deceptive chatbots and how humans can detect deception more effectively.

"Parasitic" (Deceptive) Social Bots

Sætra [25] focused on social bots, i.e., intelligent systems designed to be capable of socially interacting with humans. A major part of the article focuses on the so-called "parasitic nature" of social bots, i.e., social bots can misuse their power to deceive humans. The authors described such deception at two different levels. *Full deception* happens when, in a human-machine interaction, the machine (e.g., a robot or chatbot) "manages to fool a person into believing it is actually real". *Partial deception* happens when the machine starts being treated as "subjects" and manages to trigger irrational responses (social and emotional) although, at the rational level, the human understands they are interacting with a machine. The author further elaborated on three forms of robot deception:

1. *external state deception* which happens when the machine "lies" to the human about something external (e.g., issues false statements unrelated to its purpose);
2. *superficial state deception* which happens when the machine "suggests it has some capacity or internal state that it actually lacks" (e.g., pretends/fakes emotions such as empathy and sadness);
3. *hidden state deception* which happens when the machine provides or inhibits cues and signals "to conceal abilities or capacities that it does have" (e.g., conceals data sharing).

The author argued that human characteristics, such as willingness to help, need to bond, the fact we are social by nature, easiness to be fooled, and assumption that others are like us, become our intrinsic vulnerabilities that can be exploited by "social robots" powered by artificial intelligence (AI).

Human Detection of Deception

Markowitz [20] integrated theories and empirical evidence to provide a nuanced understanding of human deception and its detection. They point out that a lie is not holistically fabricated and liars often used a genuine prior experience to construct their false story. Deception uses truthful details and are consistent with principles of communication efficiency and the way discourse is produced incrementally. The author integrated the crucial components of deception – lie-truth base-rates, deception expectations, and goals – through a *deception faucet* metaphor. The metaphor describes deceptive discourse production as violations of conversational maxims (Quantity, Quality, Manner, Relation). As illustrated in Figure 1, the author represented *Quantity* as water flow volume (e.g., a full stream or drips of water), *Quality* as water temperature (deceptive discourse as hot water and honest discourse as cold water), *Manner* as the style of water flow, *Relation* as water not being "off-colour" and *deception goals* as sink type. As an example, the author explained that while deception was represented as hard water calcification that can accumulate and leave residue in the sink, deception detection was represented as realising that a clog due to the calcification prevents water from exiting the drain. In this case, people can check flow volume, temperature, the style of flow and discolouration to identify the clog, according to the author. This study can serve as a starting point to understand theoretical research and empirical studies about human deception production and detection.

Levine [18] reviewed studies on human deception (lie) detection, and discussed how findings have changed across the time. The author argued that 2006 was a milestone as some new deception detection approaches were introduced after 2006. Based on this, the author compared deception detection methodologies studied via experiments and resulting accuracy levels for pre-2006 and post-2006 deception detection research. The comparison showed

© 2020 University of Kent, UK

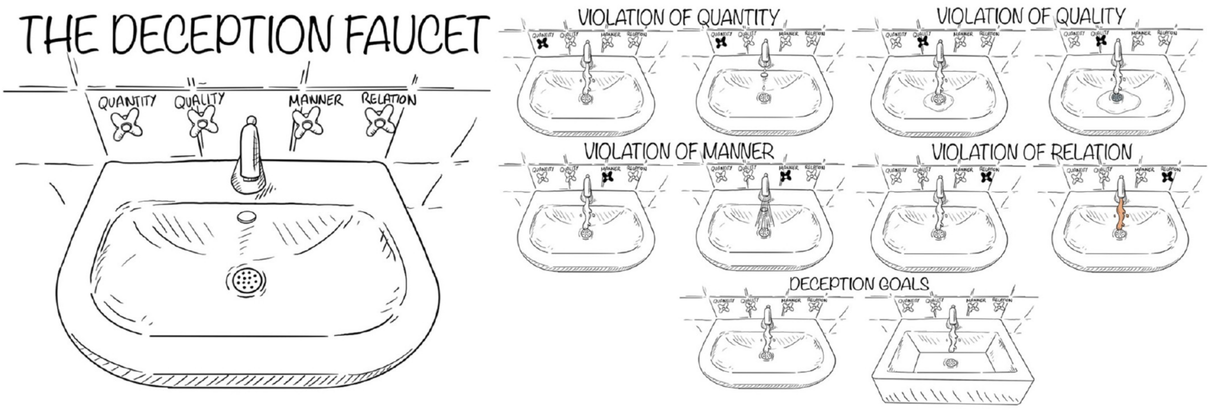


Figure 1: An overview of the deception faucet proposed by Markowitz [20] (left panel) and violations of maxims reflected in the deception faucet metaphor (top four right panels).

that deception detection studies started to focus on communication content rather than cues after 2006. This significant change in detection approach considerably increased the average accuracy of the studies. The average accuracy for pre-2006 studies, which were mostly following cue-based approach, was 54%. While post-2006 studies following different approaches, including strategic use of evidence, content in context or situational familiarity, and persuasion-based approaches, have all produced accuracy levels over 70%.

© 2020 University of Kent, UK

Conversational Agents

Introduction

At the centre of human-machine interactions are conversational agents. Powered by AI technologies, especially natural language processing (NLP) and speech recognition, they are able to engage humans in a dialogue or a group conversation for different purposes. The deceptive power of social bots, such as conversational agents, has been elaborated generically in one paper covered in the previous section.

In this section, we cover a number of more specific DDD-related aspects of conversational agents, including making conversational agents seem more natural to humans (including to humans of different cultures), equipping conversational agents with the ability to suggest better topics, increasing conversational agents' power of persuasion, improving conversational agents' ability to learn and adapt, turning the development of conversational agents easier to achieve, and increasing the conversational power for a wider range of tasks.

As highlighted in the editorial, different definitions exist for the terms "conversational agents" and "chatbots". One could argue that all chatbots are conversational agents, but not all conversational agents are chatbots. Nevertheless, for the purpose of this section, we consider these terms interchangeably to simplify the related concepts.

Evaluating Human-like Features & Behaviour for Conversational Agents

Zepf et al. [30] proposed the EmphaticSDS (Emphatic Speech Dialogue System) prototype to evaluate the use of *lexical empathy* and *acoustic empathy* in human-machine speech interactions in two driving scenarios: set car temperature and start navigator. The former empathy relates to the ability of the system to re-use the user's command words in the response provided, while the later empathy relates to the ability of the system to match the perceived user's voice emotion (when giving commands) in the response provided. The architecture of the prototype had four modules, as illustrated in Figure 2. A user study with 33 participants (16 females and 17 males) was undertaken to evaluate the prototype and subjective user-ratings regarding perceived "empathy", "personalisation", "naturalness" and "efficiency" (of the system in providing information to the user). Results indicated that lexical mimicry was particularly appreciated by participants as improving empathy and personalisation without negative impact on efficiency. This result potentially informs new design of conversational agents, including deceptive ones.

Dubiel et al. [6] studied the effect of synthetic voice in conversational systems (i.e., Text-To-Speech (TTS) systems). They used a methodology consisting of two stages: (Stage 1) Voice Selection, and (Stage 2) Interactive Evaluation. Stage 1 included the selection of datasets representatives of

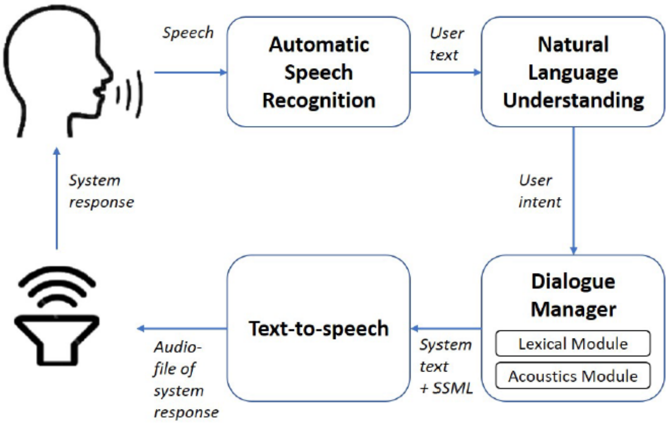


Figure 2: Architecture of the EmphaticSDS prototype proposed by Zepf et al. [30]; SSML refers to "Synthesis Markup Language".

© 2020 University of Kent, UK

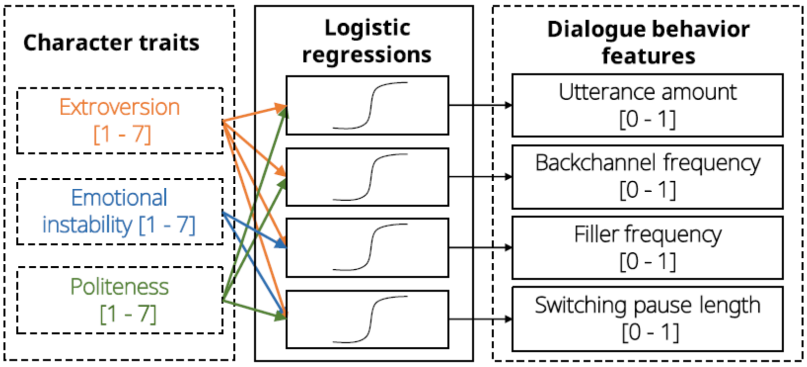


Figure 3: Character traits, obtained in Step 1, were used to train a character expression model in Step 2 (Yamamoto et al. [29])

.

both classes of "persuasive synthetic voice" and "expressive synthetic voice". For the former class, they made a selection of speakers from the IBM Debater dataset (debating speeches) and, for the latter, they made a selection of speakers from the LibriTTS dataset (with audiobooks). An online listening test was then carried out with 30 crowd-sourced participants (equally balanced between males and females) with the goal of selecting the most persuasive (male) voice and the most expressive (male) voice from each pre-selection. Test results showed that the debate speaker with the *lower speech rate and lower mean pitch* was considered as the most persuasive, and the audiobook speaker with the *fastest and broadest pitch range* was selected as the most expressive. Stage 2 included the creation of a conversational agent prototype, and the setup of an experiment with four interactive tasks related to flight booking. 26 participants (14 males and 12 females) took part. The main findings were that the debater speaker's voice scored better among participants in terms of perceived personal qualities "truthful" and "involved", but the study did not find any significant difference to suggest that persuasive voice affects user behaviour in terms of following recommendations. Such preliminary results are encouraging to indicate that voice alone is not enough to convince people to follow recommendations potentially provided by an adversarial system.

Yamamoto et al. [29] addressed the aspect of *character personality expression* of conversational agents, recognising that the personality conveyed by an agent should align with the "social role" it is meant to fulfil. Rather than focusing on agents' utterance content, the authors focused on agents' *utterance behaviour* (i.e., spoken dialogue behaviour) and their relationship with the character traits extroversion, emotional instability, and politeness. They reported on a study composed of 4 main steps. Step 1 ("impression evaluation") was an user experiment conducted with 46 university students (18 females and 28 males) to determine their impression of character trait for different dialogue conditions, exercising the following attributes: utterance amount, backchannel (interjections) frequency, filler frequency, and switching pause length. In Step 2, as shown in Figure 3, results from Step 1 were used to train a "character expression model" to control the dialogue behaviours. Step 3 evaluated the validity of the model. For that, dialogues between the authors' Android chatbot ERICA (using the trained model) and 4 human operators performing "participant roles" (i.e., a first-time dating partner, a job interviewer, and an attentive listener) for 3 types of "dialogue tasks" (i.e., speed dating, job interview, and attentive listening) were recorded. Results indicated that the model was able to represent appropriate character traits according to each dialogue task. In Step 4, a user evaluation of the model was conducted (with 13 participants) to establish whether humans perceived the emulated character traits from listening to sample dialogues. Results showed they did with an average accuracy ratio of 0.770.

Shi et al. [27] investigated the effect of identities of chatbots and the strategies adopted by chatbots for inquiry in human-centred conversations with persuasive goals. The authors conducted an online study involving 790 participants, and their goal was to study how humans can be persuaded by a chatbot to donate to charity. They designed a two by four facto-

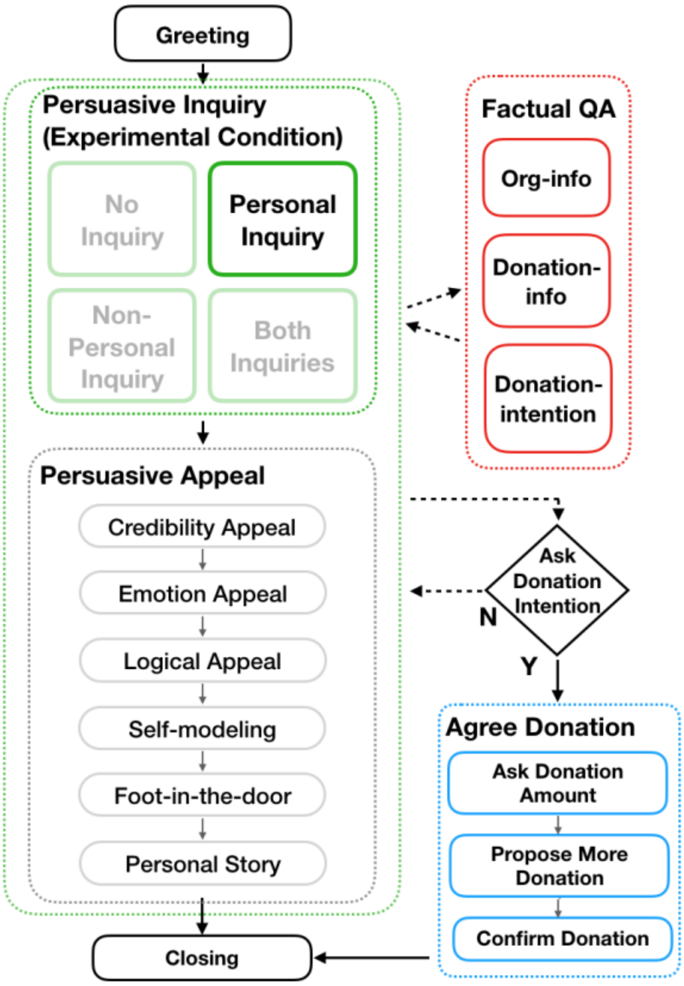


Figure 4: The System Agenda by Shi et al. [27].

rial experiment: a chatbot that could pose with two displayed identities (human or chatbot) and four inquiry strategies leading to perceived identities. The inquiry strategies explored are: (1) no inquiry and proposing a donation directly, (2) non-personal inquiry only, (3) personal inquiry only, and (4) both inquiries. The overall structure of the conversation led by the chatbot followed the pattern described in Figure 4. The participants were randomly assigned to different interaction settings. The findings of the experiment showed that the *perceived identity* of the chatbot had significant effects on the persuasion outcome (i.e., donation) and interpersonal perceptions (i.e., competence, confidence, warmth, and sincerity). Their findings had further details on the effects of interaction with perceived identities and inquiry strategies. They concluded that it is not the displayed identity but the perceived identity by the user that matters. They discussed these findings for theoretical and practical implications towards developing ethical and effective persuasive chatbots. The ethical angle comes from the need for a bot to disclose its artificial nature, i.e., to not deceive the human user, which has been required by law in some countries, e.g., in the US state California via the 2019 Autobot Law [17].

Laban and Araujo [15] studied customers' perception of conversational agents as having a less or more *cooperative behaviour* in the context of a service-oriented task. In order to evaluate that, the authors created four agents (using the Conversational Agent Research Toolkit) focused on simple or complex service, and cooperative or non-cooperative agent using a between-subjects experiment design. A *mediator model* was adopted to observe the relationship between the independent variable "perceived cooperation" and the dependent variable "perceived service performance" using the following variables

© 2020 University of Kent, UK

as mediators: (1) “Perceived anthropomorphism” (human-like features/behaviour of the agent); (2) “Perceived social presence”; and (3) “Perceived information quality”. A pool of 91 crowd-sourced participants were asked to perform a given task, followed by a questionnaire to gather perceptions. Key findings indicated that the most influential mediator between *perceived cooperation* and *service performance* was (3), and the second most influential mediator was (1). A lack of association was observed in relation to mediator (2). This study indicates that information quality, rather than anthropomorphism, has more influence on customers, which has implications on development of deceptive chatbots.

Beyond Two-party Conversation & Task-oriented Agents

Ahmadvand et al. [1] defined the Conversational Topic Suggestion (CTS) problem for an *open-domain conversational agent*, i.e., an agent which engages in conversations with users on a number of topics rather than a task-oriented specific topic. In such a context, the agent aims to propose topics that maximise the probability of acceptance by the user. The authors evaluated a mix of different types of recommendation strategies, and implementation models (i.e., Conditional Random Fields (CRF) model, Convolutional Neural Network (CNN) model with 3 layers, and Recurrent Neural Network (RNN) model with 256 hidden layers). Results were obtained using the Amazon Alexa Prize 2018 competition dataset composed of 14,707 conversations collected over a period of 15 days. Comparison of metrics *Micro-averaged Accuracy* (individual topic suggestions) and *Macro-averaged Accuracy* (suggestions across classes of topic) indicated that the CRF model outperformed the deep learning-based models CNN and RNN for both metrics (81.9% and 79.9%, respectively). One explanation, pointed out by the authors, is that CRF models are not so negatively affected by small volume of training data, compared to CNN and RNN models. Generalising the results of this study, the ability of open-domain conversational agents to suggest topics to users may be valuable for adversarial agents to influence the flow of conversations and gather sensitive information from users.

Martinez and Kennedy [21] proposed a multi-party conversational system based on the observation that research and development focused, until now, on two-party interactions consisting of an automated agent and a user. In such multi-party interactions, the agent should be able to track multiple conversations involving different topics over different channels. Therefore, this new setting brings challenges such as the ability of an agent to address a group as a whole and participants individually, to manage turns across conversations and topics, to detect the current speaker, and remember past interactions with recurring participants. Their solution relies on three main ingredients: *agent managers*, *memory component*, and *sensor fusion component*. The prototype system was evaluated with a case study drawing from a pool of 6 participants (equally represented by males and females) paired randomly. Each pair engaged in three interactions with an agent which had a name, a background (made explicit in replies), and was able to visually show human-like movements (e.g., blinking and breathing). Lessons learned from the case study included issues when participants addressed the agent simultaneously, issues with the current manual approach to test the logic of the system, and issues for the human participants when too many interrupt events happened making it difficult for them to keep track of the thread. Despite being at early stages, research in multi-party social-chatbot systems aims at a more natural experience to participants – potentially amplifying the likelihood of an adversarial agent to convincingly engage with them.

Conversational Agents Inducing Self-Disclosure & Other Actions

Lee et al. [16] explored the influence of *chatting styles* on self-disclosure to inform a two-party conversational systems’ design. They used off-the-shelf components (i.e., ManyChat and Google Dialogflow) to build and monitor three types of chat sessions:“Small talk”, “Journaling” (non-sensitive questions) and “Sensitive Questions”. Students were recruited for a period of three weeks, and were assigned to three different groups of 15-16 students: ND control group (non self-disclosure), LD group (low self-disclosure) and HD group (high self-disclosure). What differed between groups was the amount of self-disclosure in the replies supplied *by the chatbot agent* (e.g., answers including feelings, thoughts and detailed information against general and neutral answers). The study design is illustrated

© 2020 University of Kent, UK

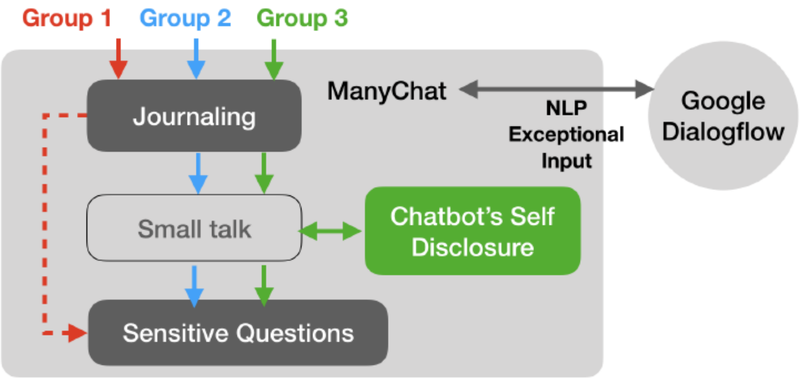


Figure 5: Design of the study undertaken by Lee et al. [16].

in Figure 5. The experiment was followed by an interview with participants to gather perceptions and a survey to gather feelings about "trust", "intimacy" and "enjoyment". Compared to the other groups, the following were observed among the HD group: an increased perception of rapport and easiness to answer sensitive questions, and a feeling of genuine exchange of information with the agent. This observation, however, was not replicated for the journaling questions. The effect of time had a positive impact on the HD group with an increased level of enjoyment and willingness to keep interacting for longer, compared to the other groups. Findings of this study suggested how the design of conversational agents can influence the behaviour of humans interacting with them, which again may be exploited by adversarial chatbots.

Ischen et al. [10] investigated the increasing use of chatbots in a commercial context to make product- or service-related recommendations. Therefore, these bots need to collect personal information of the user, similar to other online services, leading to potential privacy issues. This study investigated the extent to which privacy concerns in chatbot interactions are related to users' attitudes and recommendation adherence. It also investigated the extent to which users feel comfortable sharing personal information with a human-like chatbot in comparison with a machine-like chatbot, or a website. There were 231 individuals participating in the study, all recruited through the Dutch online panel PanelClix with age ranged from 18 to 73 (mean = 41.83 and standard deviation = 14.01), 48.5% were female (51.5% male) and 51.6% indicated to have a high educational level (38.9% middle, 9.5% low). The findings showed that a human-like chatbot leads to more information disclosure, and recommendation adherence mediated by higher perceived anthropomorphism (human-like characteristics) and subsequently, lower privacy concerns in comparison to a machine-like chatbot. However, no statistically significant difference was observed on the perceived anthropomorphism or the affect on information disclosure when a human-like chatbot is compared with a website. These results suggest that more evidence is required about the usefulness of human-like chatbots in personal recommendations for commercial products. That in turn implies that deceptive behaviour of bots may not be more effective than simpler methods like fake websites for encouraging users to disclose more information.

Følstad and Taylor [8] presented findings from a study where a solution for expressing uncertainty and suggesting likely alternatives was implemented in a live chatbot for customer service. Seven hundred chatbot dialogues were sampled at two points in time (immediately before and after implementation) and compared by conversational quality, which contains response adequacy, dialogue directness, dialogue conclusiveness and dialogue helpfulness. The authors showed that their solution for conversational repair reduced the proportion of false positives in chatbot dialogues from 30% to 11%. At the same time, expressing uncertainty and suggesting likely alternatives did not seem to strongly affect the dialogue process and the likelihood of reaching a successful outcome. The study is in its early phase and includes samples for only one implementation. Moreover, the sampled data does not provide insights in possible long-term changes in the effects of expressing uncertainty since the dialogues to be compared were collected in the first week after the implementa-

© 2020 University of Kent, UK

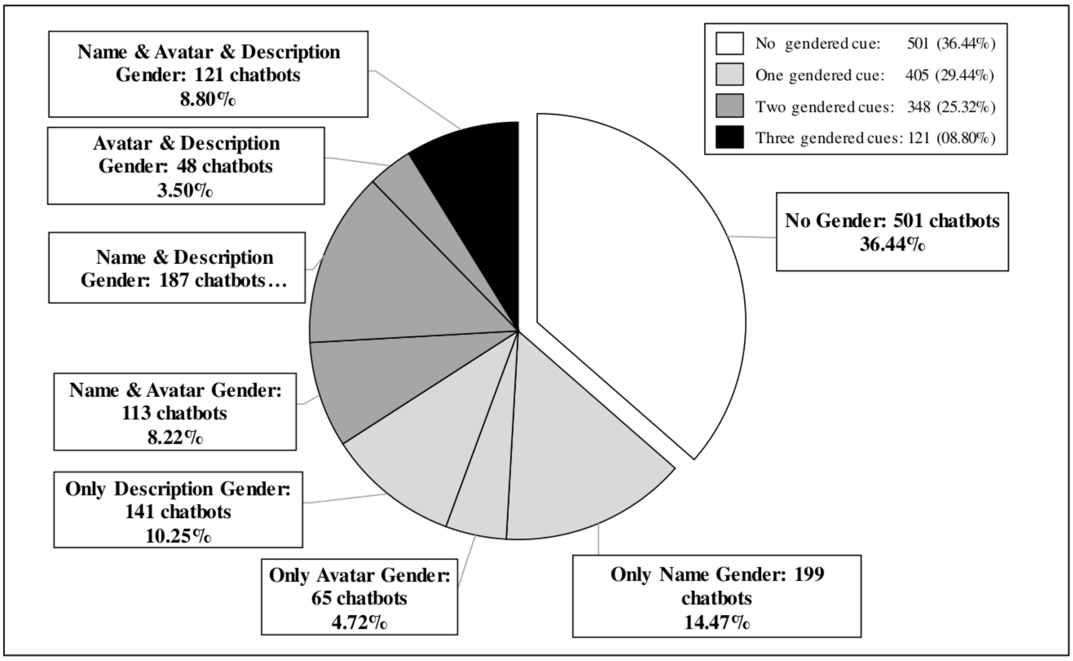


Figure 6: The distribution of gender-specific names, avatars, and descriptions from the investigated list of chatbots in the work by Feine et al. [7].

tion. On the other hand, with the experiments with larger samples and multiple implementations across longer periods of use, the results could be improved to inform development of more deceptive chatbots.

The Role of Conversational Agents' Gender and Culture

Feine et al. [7] leveraged from a report published by the UNESCO which revealed that most popular voice-based chatbots are designed to be female, and the potentially harmful effects this can have on society. The UNESCO report focused primarily on voice-based chatbots, while Feine et al.'s work considered text-based chatbots. Since chatbots can be gendered in their design, the authors used an automated gender analysis approach to investigate three gender-specific cues in the design of 1,375 chatbots listed on the platform `https://chatbots.org/`. They used: two gender APIs to identify the gender of the name, a face recognition API to identify the gender of the avatar, and a text mining approach to analyse gender-specific pronouns in the chatbot's description. Their results suggested that gender-specific cues are commonly used in the design of chatbots and that most chatbots are – explicitly or implicitly – designed to convey a specific gender. Most of the chatbots have female names, female-looking avatars, and are described as female chatbots. This is particularly evident in three application domains (i.e., branded conversations, customer service, and sales). Therefore, they found evidence that there is a tendency to prefer one gender (i.e., female) over another (i.e., male). Thus, they argued that there is a gender bias in the design of chatbots in the wild. A pictorial view of there findings can be found in Figure 6. Based on these findings, they formulated propositions as a starting point for future discussions and research to mitigate the gender bias in the design of chatbots. Such results are helpful for the design of malicious chatbots.

Miehle et al. [22] have investigated whether culture-specific parameters can be trained by following a supervised learning approach so that the system can response to user actions according to the user's culture. The authors used a dialogue management framework based on the concept of probabilistic rules and a data set including 258 spoken dialogues on healthcare topics in the languages of four different cultures, German, Polish, Spanish and Turkish. They implemented a rule for each user actions, Accept, Declare, Goodbye, Greet, Reject, Request and Thank, by using *if...then...else* structure. For training the system shown in Figure 7, the authors used Wizard-of-Oz learning in which they used dialogue transcripts to estimate the culture-specific

© 2020 University of Kent, UK

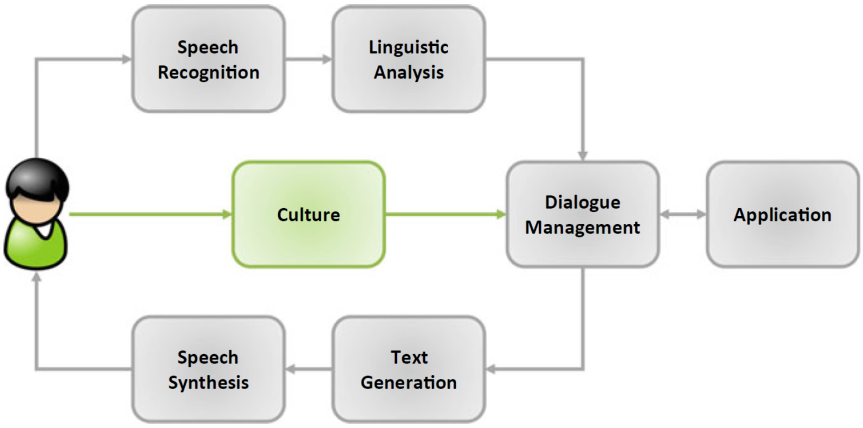


Figure 7: In the system proposed by Miehle et al. [22], the user's culture was used in the dialogue management to adapt the system behaviour to the user.

parameter. For evaluation, the authors compared the mean values of probability distributions of the parameters which are responsible for the selection of the next system action. The results showed that the different characteristics of the cultures result in different parameters with highest mean values, meaning that the system response to a user action varies depending on the culture. Although this study only covered four European languages in a healthcare context, the authors stated that they are interested in extending the proposed approach to other conversational topics and further cultures. This study can also be useful for the assessment of cultural differences related to deception.

Lloyd et al. [19] introduced a new database, Miami University Deception Detection Database (MU3D), which contains 320 videos of 80 individuals telling truths and lies. Out of 80 individuals, blacks-whites and females-males were equally represented with 20 individuals from each combination of these race and gender. Each individual recorded four different videos (i.e., positive truth, negative truth, positive lie, negative lie), yielding 320 videos fully crossing target race, target gender, statement valence, and statement veracity. The authors provided descriptive analyses of the video characteristics (e.g., length) and subjective ratings (e.g., target attractiveness). The stimuli and an information codebook, providing additional information about each video (e.g., trustworthiness ratings, anxiety ratings, length of video, transcriptions of videos), as well as information about the targets featured in the videos (e.g., attractiveness ratings, self-reported age, and self-reported race), are available for free for academic research purposes at `http://hdl.handle.net/2374.MIA/6067`. This database could be useful for understanding cultural and demographic differences for deception detection research.

Empowering Conversational Agents with the Ability to Learn and Adapt

Jacovi et al. [11] highlighted that task-oriented conversational systems (e.g., customer support agents) typically adopt a rule-based architecture and are unable to automatically learn and improve their support in production. As such, improvements to the system require experts' input and manual changes to dialogue execution graphs; inability to handle users' query or explicit requests for support from humans are recorded on *escalation logs*. The authors leveraged from such logs and proposed a 5-stepped architecture to recommend and integrate new nodes to deployed execution graphs. A prototype using IBM Watson Assistant was evaluated using a non-public dataset with 7,605 banking customer support conversations, and the public MultiWOZ dataset with 10,000 conversations from a variety of customer support domains. Escalation logs were simulated, and execution graphs for the datasets were either created or "destructively modified" (e.g., by removing nodes and edges reflected on the corresponding escalation log). The evaluation focused on the classifier used by the solution (i.e., Decision Tree) compared to Random Forest (RF) and XGBoost applied to the banking dataset. The evaluation metrics used were *Adjusted Rand Index (ARI)*, *Clustering Coverage (CC)*, *Number of Child Nodes*, and *Condition*

© 2020 University of Kent, UK

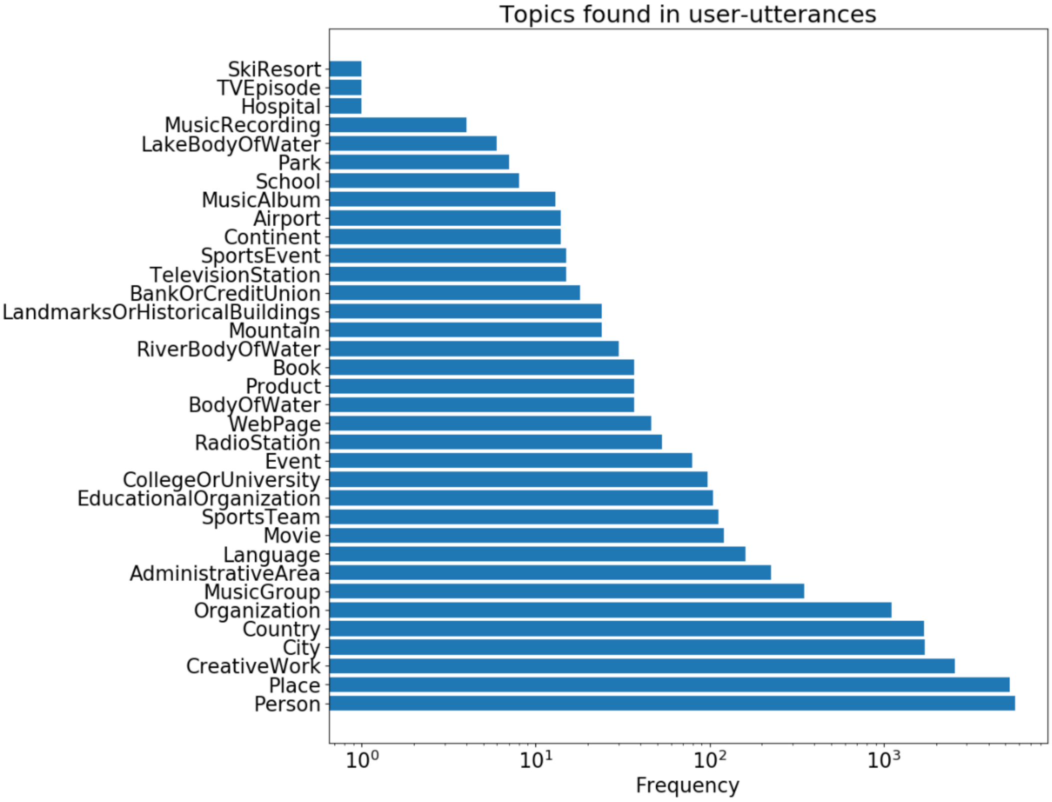


Figure 8: The topics found in user utterances as identified by DBpedia Spotlight in the work by Jalota et al. [12].

Length. The results indicated that the proposed solution outperformed the others in terms of ARI but were slightly below in terms of CC achieved by Random Forest. An evaluation of human-to-human logs generated was also undertaken using the MultiWOZ dataset (which contained agents' action labels), and indicated that conditions deriving from multiple, often disjoint paths, reaching a node were long and difficult to interpret. This prompted addition of variables, which the authors aimed to detect automatically in the future.

Jalota et al. [12] presented an approach to performing retrospective analysis of the logs of knowledge base-driven dialogue systems in order to examine whether these systems are serving their intended purpose and catering to the needs of their users. In particular, they tried to understand (1) how users interact with knowledge-driven chatbots, (2) whether the chatbots can sufficiently satisfy the expectations of the users, and (3) the overall flow of conversations and hence possible avenues for improving chatbot quality and subsequently the user experience. They identified common user topics that users ask a knowledge-driven chatbot and the use of anaphora (repeated words) in their conversations. These topics have been listed in Figure 8. They suggested three general analytical streams for investigating knowledge-driven chatbots. Using the DBpedia Chatbot as a case study, the authors inspected three aspects of the interactions: user queries and feedback, the bot's response to these queries, and the overall flow of the conversations. They discussed key implications based on their findings. All the source code used for the analysis have been made available at `https://github.com/dice-group/DBpedia-Chatlog-Analysis`.

Biancardi et al. [4] presented a computational model aiming at detecting users' impression of the agent and producing appropriate agent's verbal and nonverbal behaviours to maintain a positive impression of warmth and competence. User's impressions were recognised using a machine learning approach through facial expressions. The agent could *adapt in real-time* its verbal and nonverbal behaviour, with a reinforcement learning algorithm that takes user's impressions as reward to select the most appropri-

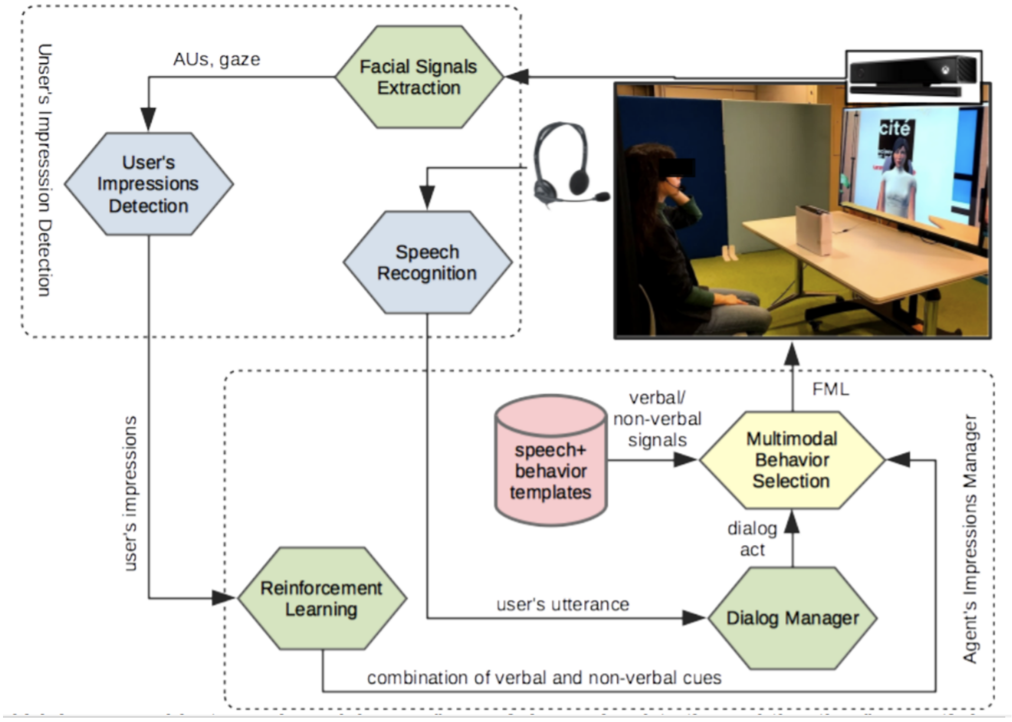

Unser's Impresssion Detection
AUs, gaze
Facial Signals Extraction
User's Impressions Detection
Speech Recognition
user's impressions
verbal/ non-verbal signals
FML
speech+ behavior templates
Multimodal Behavior Selection
dialog act
Agent's Impressions Manager
Reinforcement Learning
user's utterance
Dialog Manager
combination of verbal and non-verbal cues

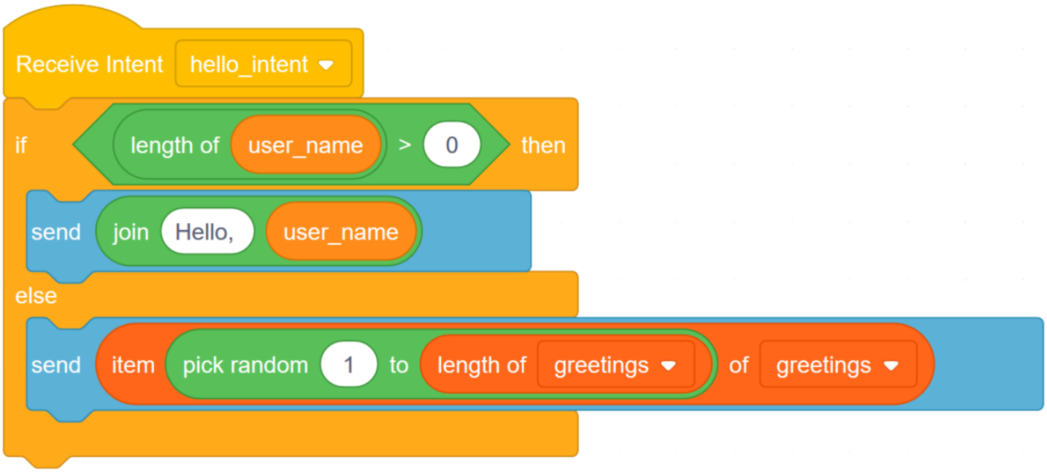
Figure 9: The system architecture proposed by Biancardi et al. [4] showing the set up of their user study.

ate combination of verbal and non-verbal behaviour to perform. The authors also performed a user study with 71 participants to test the model in a contextualised interaction with humans, as shown in Figure 9. The results of ANOVA indicated that the agent performs significantly better when using the proposed model than in the random condition, in which the agent randomly chose its behaviour, without considering user's reactions. Moreover, results showed that users' ratings about an agent's warmth are influenced by their a priori experience and perceptions about virtual characters in general, and that users judged the agent as more competent when it adapted its behaviour compared to the random condition.

Hussain et al. [9] presented a method for training a robot for generating backchannels, which mean reactions in a conversation like non-verbal gestures (nods and smiles), non-verbal vocalisations (mm, uh-huh, laughs) and verbal expressions (yes, right), during a human-robot interaction to maintain high engagement level. The authors considered the problem as a Batch Reinforcement Learning problem. Since online learning by interaction with a human is highly time-consuming and impractical, they used the IEMOCAP (Interactive Emotional Dyadic Motion Capture) dataset, that consists of dyadic human-to-human conversations on a range of scenarios. A total of 151 dialogues from the dataset were performed by 10 professional actors in pairs on scripted and improvised scenes. In order to treat this as a batch data for training, the authors assumed that of the two actors, one represents a behaviour policy which takes the actions and the second actor behaves as an environment that generates states and rewards. The authors performed some experiments with Multi-Layer Perceptron (MLP) and fully-connected Long Short Term Memory (LSTM), and evaluated their relative performance by using Bellman residual and Off-policy Policy Evaluation (OPE) techniques to understand the effectiveness of the resultant policy. The results showed that the fully-connected LSTM outperforms MLP. This study contains more theoretical experiments, and the authors stated that they will evaluate the effectiveness of their method with experiments in a human-robot interaction setting.

Towards More Accessible Development of Conversational Agents

Rough et al. [24] raised attention to the gap between the amount of research devoted to the technical development of the components that make up Intelligent Personal Agents (IPAs) and Spoken



```
Receive Intent  hello_intent ▾
if  length of user_name > 0  then
    send  join Hello, user_name
else
    send  item  pick random 1 to length of greetings ▾  of  greetings ▾
```

Figure 10: A simple greeting program written in a Blockly-derived language designed by Klopfenstein [13]. The program uses blocks to express conditional logic and chatbot behaviour. The program relies on basic Blockly features such as variables and lists.

Dialog Systems (SDSs, i.e., voice-based conversational agents), and the number of trained programmers with skills to develop nuanced systems that can learn and adapt to individual users. Although IPAs like Apple's Siri and Amazon's Alexa, have upgraded speech-based interfaces to become a key feature of smartphones and personal in-home devices, they can only do predefined tasks. This paper is primarily from the field of end-user development (EUD) with the goal to make building and prototyping customised speech interface experiences more accessible to end-users. This is to allow users to shape and personalise their experience of such speech-based systems. They proposed providing visual programming tools for novice end-users to customise and extend the functionality of their IPAs. They suggested that these tools can lower the barriers to the adoption, personalisation and everyday use of IPAs. Their proposal has been explored through an initiative called the B-SPOKE project with the aim to empower users to personalise their experience with IPAs. Their goal was to enable speech and dialogue researchers to rapidly innovate and prototype new ideas, potentially encouraging new forms of interactions and applications. They associated the idea of "democratising development" of the speech interface through the project and listed the salient challenges in building such enabling tools. While this initiative may be empowering users, it may also be providing easily accessible deceptive capabilities to malicious actors, similar to what happened with the generative adversarial network (GAN) technology (which has been used for creating DeepFake media for deception purposes).

Klopfenstein [13] pointed out that Google Blockly has been adopted over the years by a variety of software development tools – primarily designed for children. Blockly provides a visual block-based paradigm as an intuitive and easy-to-use approach to programming. In this paper, the design of a Blockly-derived language for chatbots was proposed. It portrayed common conversational interface concepts within the framework for block-based programming elements. The technical goal was to delegate the user input handling to an external NLP system, which transforms raw user utterances into user intents and parameters. This allows abstract intents to be easily processed by a block-based dialog manager. Figure 10 shows an example block program that is handling a user intent (without parameters) and sending back a greeting based on simple logic and internal chatbot state. The incoming user input has been represented as an "event" block (i.e., the "Receive intent" block in yellow). Other special blocks can be added for common messaging features (like the "Send" block in blue, taking a simple string as input).

Adopting Conversational Approach for Tasks not Conversational by Nature

Xiao et al. [28] reported on an empirical study to evaluate the effectiveness and limitations of

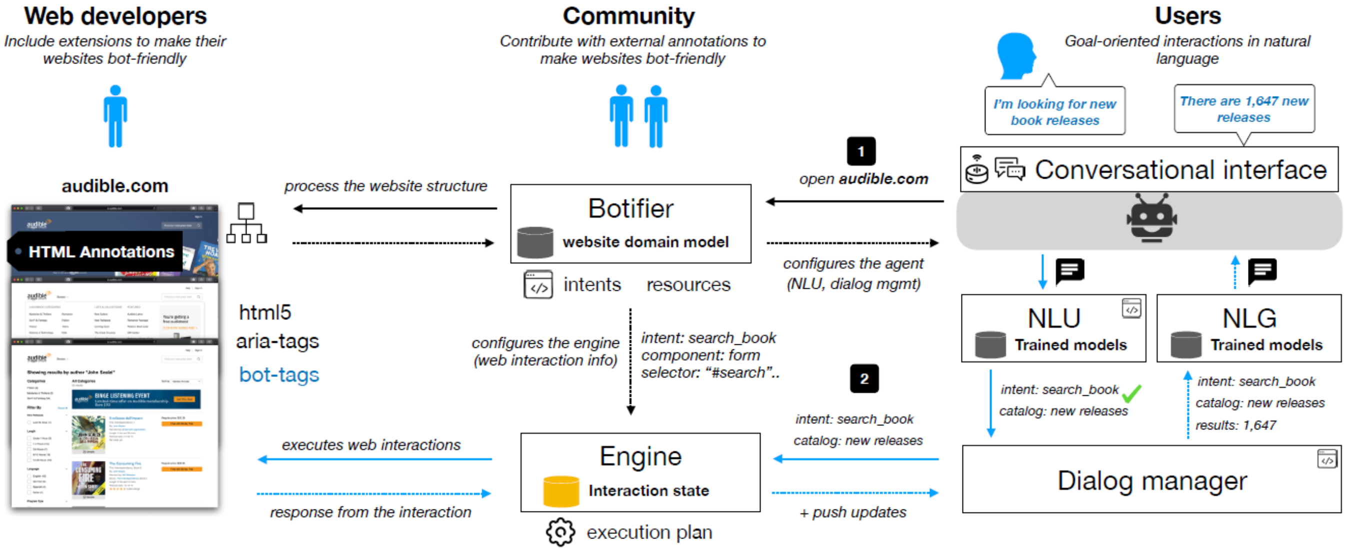


Figure 11: Reference architecture for a conversational website proposed by Baez et al. [2]
.

an AI-powered chatbot survey (implemented with `https://juji.io/`) compared to a traditional web-based survey (implemented with `https://www.qualtrics.com/`). The off-the-shelf chatbot Juji had a set of "conversational" skills, being able to ask for more information, provide response feedback, probe answers, verbalise emotions and convey understanding, control and maintain conversation flow, and handle user excuses and questions. Six-hundred participants were involved in the study, and they were equally invited for each survey format, provided a URL. The survey content, consisting of choice and free-text type of questions, targeted US-based video gamers. The survey formats were compared in terms of (1) *information quality* of free-text questions using metrics "informativeness", "specificity", "relevance" and "clarity"; (2) *level of engagement* using metrics "engagement duration", "response length" and "self-disclosure"; (3) and *response and completion rates.* The authors' main findings indicated that the chatbot survey outperformed the traditional survey in several aspects, such as it collected almost 40% more and richer information, answers were more relevant, participants provided lengthier answers (therefore engaged for a longer period of time), the completion rate was 2.2 times higher (although the response rate was 6% lower), and the majority indicated a positive reaction to the human-like *Juji.* An important finding was that 32% of the chatbot survey participants disclosed personal information, compared to 16% using the traditional format. This indicates that the potential for a malicious chatbot to induce users to disclose sensitive information is real. The authors mentioned the possibility of obfuscating such information assuming a non-deceptive chatbot.

Schneeberger et al. [26] presented the results of a study that examined the obedience of human users towards an embodied virtual agent in the role of an instructor, and a human in the role of an instructor. Under a cover-story of a creativity test, the authors asked 60 women participants to fulfil 18 stressful and shameful tasks. The results indicated that participants obeyed the virtual agent at the same level as the human instructor. On average, around 14 tasks were fulfilled in both conditions. In addition, the agent was able to elicit the same level of the negative feelings such as stress and shame. On the other hand, this study had two significant limitations. Firstly, the experiments have been performed via video-chat, meaning that a human instructor in a video-chat may not have the same authority compared to a human in a face-to-face interaction. Secondly, the authors relied on self-reported stress and shame levels rather than an objective measurement.

Baez et al. [2] discussed the concept of *conversational web interactions* without the need to implement tailored conversation logic for existing websites and to train Natural Language Understanding (NLU) and Natural Language Generation (NLG) models. The authors' idea leveraged from two ingredients: "chatbot" technology to establish the interface between user and website (vocally or textually), and "annotation" of websites' content and functionalities to empower the chatbot with application domain knowledge. The proposed reference architecture, illustrated in Figure 11, refines the typical *input-intent-action-response* paradigm of conversational agents applied to an example audiobook website. The "botifier" component parses and processes a website, given its URL, to extract domain knowledge. The authors discussed implementation options and challenges involved in implementing such conversational websites. *Privacy and security* was identified as a challenge requiring further research. The potential for deception, e.g., via malicious "actions" (with misleading "response") triggered by user "input" with legitimate "intent" may become reality, specially with security as an after-thought.

© 2020 University of Kent, UK

Miscellaneous

Introduction

This section covers four papers each on a different topic related to DDD. The first two papers both report empirical studies involving human participants, looking at how human credibility assessment on people can be influenced by stereotypical deception cues and how humans detect machine-generated content, respectively. The third paper looks at an AI model for automatic credibility assessment of Twitter events. Finally, the last paper studies the human mental model of AI's error boundary in "AI-advised human decision making", a case of human-machine teaming, whose results could find applications in joint human-machine detection of DDD.

Stereotypical Deception Cues and Human Credibility Assessment

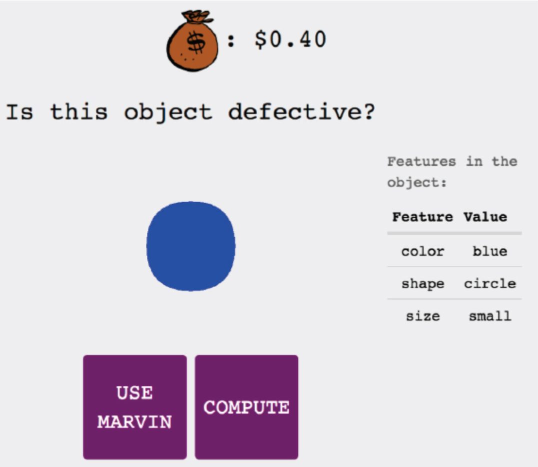
Bogaard and Meijer [5] investigated to what extent nonverbal stereotypical deceptive behaviours, such as gaze aversion and fidgeting, affect people's credibility assessment. In this manner, they showed four truthful videos, in which stereotypical deception cues were manipulated, to each of 75 participants and asked participants to assess the credibility of the person in each video. They also investigated if the timing of cues influenced credibility assessment due to "the primary effect", which implies that people form an opinion early in the decision process and this opinion will have the largest influence on how subsequent information will be interpreted. The results showed that stereotypical deception cues, such as gaze aversion and fidgeting, significantly decrease the observed credibility. On the other hand, the authors found that the timing of the cues has no effect on credibility assessment. The authors used ANOVA in their statistical analyses. This study shows that stereotypical cues could negatively affect the process of human deception detection.

Human Detection of Machine-generated Content

Köbis and Mossink [14] conducted two experiments to evaluate behavioural reactions to the Natural Language Generation (NLG) algorithm Generative Pre-trained Transformer (GPT)-2 with 830 participants. The authors generated poem samples with GPT-2 by using the identical starting lines of human poems. From these samples, either a random poem was chosen (human-out-of-the-loop) or the best one was selected (human-in-the-loop), and the selected one has been matched with a human-written poem. While the authors used poems written by novice users in the first study, they used poems of professional poets in the second study. In a new incentivised version of the Turing test, participants failed to reliably detect the machine-generated poems in the human-in-the-loop treatment, yet succeeded in the human-out-of-the-loop treatment. Further, participants revealed a slight aversion to machine-generated poetry, independent of whether participants were informed about the algorithmic origin of the poem (transparency) or not (opacity). This study showed that humans have difficulties in identifying machine-generated content, implying that NLG algorithms may be useful for automatically creating deceptive content.

Automatic Credibility Assessment

Patro and Rathore [23] proposed a credibility analysis approach based on the linguistic structure of tweets. They made use of a novel deep learning architecture to characterise Twitter events and predict their perceived credibility. The authors used the CREDBANK dataset, which contains 66M tweets and 1,377 events reported in Twitter, to conduct their experiments. They also used two linguistic analysis tools (for categorising Twitter events): Linguistic Inquiry and Word Count (LIWC) and Empath. The analyses with these tools showed that standard LIWC categories like "negate", "discrep", "cogmech", "swear", and Empath categories like "hate", "poor", "government", "worship" and "swearing-terms" correlate negatively with the credibility of events. For the second step of the study, the authors used the derived categories to predict the credibility of a Twitter event by using each of the categories as a feature in the deep learning model. The authors used Hierarchical Attention Networks (HAN) for the classification of events. Using their deep learning architecture, the authors achieved an accuracy of 54% which is 26% better than the best-known state-of-the-art.



: $0.40

Is this object defective?

Features in the object:

Feature	Value
color	blue
shape	circle
size	small

Figure 12: Interface of the CAJA game platform used in experiments by Bansal et al. [3].

Human-Machine Teaming

Bansal et al. [3] studied the human mental model of AI's *error boundary* (i.e., when AI makes mistakes) in the context of a case of human-machine teaming called "AI-advised human decision making", namely, an AI system provides recommendations while a human makes the final decision. According to the authors, the error boundary of an AI model h is a function f that describes for each input x whether model output $h(x)$ is the correct action for that input, i.e., $f : (x, h(x)) \rightarrow \{T, F\}$, where T is true and F is false. The article described a controlled experiment using CAJA, a game-based platform previously developed by the same authors for studying AI-advised human decision making. As illustrated in Figure 12, the game used an AI model called Marvin and allowed participants to accept a recommendation made by Marvin (option "USE MARVIN") or to reject it (option "COMPUTE") for a number of different scenarios. A scheme of points rewarded participants' ability to learn the error boundary. A total of 25 crowd-sourced participants engaged in the experiment. In order to evaluate how human participants formed a mental model of Marvin's error boundary and how the mental model evolved, roles of the following properties were examined: the *parsimony* and the *stochasticity* of the error boundary, and the *task dimensionality*. Based on the results, the authors concluded that parsimony positively affects the team performance, while the stochasticity and the task dimensionality have a negative affect. They therefore recommended that AI models used in human-AI collaboration should have parsimonious error boundaries and minimize the stochasticity of system errors, and the task dimensionality should be reduced as much as possible. The results reported in this work could be useful for DDD-related applications, e.g., for improving the performance of human-machine teaming efforts for detecting DDD. The work is based on a relatively small number of participants and the tasks involved are more abstract, so more research is needed to confirm the findings and produce more generalisable insights.

References

[1] Ali Ahmadvand, Harshita Sahijwani, and Eugene Agichtein. 2020. Would you Like to Talk about Sports Now? Towards Contextual Topic Suggestion for Open-Domain Conversational Agents. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*. ACM, 83–92. https://doi.org/10.1145/3343413.3377974

[2] Marcos Baez, Florian Daniel, and Fabio Casati. 2020. Conversational Web Interaction: Proposal of a Dialog-Based Natural Language Interaction Paradigm for the Web. In *Chatbot Research and Design: Third International Workshop, CONVERSATIONS 2019, Amsterdam, The Netherlands, November 19–20, 2019, Revised Selected Papers (Lecture Notes in Computer Science, Vol. 11970)*. Springer, 94–110. https://doi.org/10.1007/978-3-030-39540-7_7

[3] Gagan Bansal, Besmira Nushi, Ece Kamar, Walter S. Lasecki, Daniel S. Weld, and Eric Horvitz. 2019. Beyond Accuracy: The Role of Mental Models in Human-AI Team Performance. In *Proceedings of the 7th AAAI Conference on Human Computation and Crowdsourcing*. AAAI, 2–11. https://aaai.org/ojs/index.php/HCOMP/article/view/5285

[4] Beatrice Biancardi, Chen Wang, Maurizio Mancini, Angelo Cafaro, Guillaume Chanel, and Catherine Pelachaud. 2019. A Computational Model for Managing Impressions of an Embodied Conversational Agent in Real-Time. In *Proceedings of 2019 8th International Conference on Affective Computing and Intelligent Interaction*. IEEE, 234–240. https://doi.org/10.1109/ACII.2019.8925495

[5] Glynis Bogaard and Ewout H. Meijer. 2020. Stereotypical Behavioural Cues – But Not Their Order – Influence Credibility Judgements. *Journal of Investigative Psychology and Offender Profiling* 17, 2 (2020), 131–141. https://doi.org/10.1002/jip.1543

[6] Mateusz Dubiel, Martin Halvey, Pilar Oplustil Gallegos, and Simon King. 2020. Persuasive Synthetic Speech: Voice Perception and User Behaviour. In *Proceedings of the 2nd Conference on Conversational User Interfaces*. ACM, Article 6, 9 pages. https://doi.org/10.1145/3405755.3406120

[7] Jasper Feine, Ulrich Gnewuch, Stefan Morana, and Alexander Maedche. 2020. Gender Bias in Chatbot Design. In *Chatbot Research and Design: Third International Workshop, CONVERSATIONS 2019, Amsterdam, The Netherlands, November 19–20, 2019, Revised Selected Papers (Lecture Notes in Computer Science, Vol. 11970)*. Springer, 79–93. https://doi.org/10.1007/978-3-030-39540-7_6

[8] Asbjørn Følstad and Cameron Taylor. 2020. Conversational Repair in Chatbots for Customer Service: The Effect of Expressing Uncertainty and Suggesting Alternatives. In *Chatbot Research and Design: Third International Workshop, CONVERSATIONS 2019, Amsterdam, The Netherlands, November 19–20, 2019, Revised Selected Papers (Lecture Notes in Computer Science, Vol. 11970)*. Springer, 201–214. https://doi.org/10.1007/978-3-030-39540-7_14

[9] Nusrah Hussain, Engin Erzin, T. Metin Sezgin, and Yücel Yemez. 2019. Batch Recurrent Q-Learning for Backchannel Generation Towards Engaging Agents. In *Proceedings of 2019 8th International Conference on Affective Computing and Intelligent Interaction*. IEEE, 392–398. https://doi.org/10.1109/ACII.2019.8925443

[10] Carolin Ischen, Theo Araujo, Hilde Voorveld, Guda van Noort, and Edith Smit. 2020. Privacy Concerns in Chatbot Interactions. In *Chatbot Research and Design: Third International Workshop, CONVERSATIONS 2019, Amsterdam, The Netherlands, November 19–20, 2019, Revised Selected Papers (Lecture Notes in Computer Science, Vol. 11970)*. Springer, 34–48. https://doi.org/10.1007/978-3-030-39540-7_3

© 2020 University of Kent, UK

[11] Alon Jacovi, Ori Bar El, Ofer Lavi, David Boaz, David Amid, Inbal Ronen, and Ateret Anaby-Tavor. 2020. Improving Task-Oriented Dialogue Systems in Production with Conversation Logs. In *Proceedings of the KDD 2020 Workshop on Conversational Systems Towards Mainstream Adoption*. CEUR Workshop Proceedings (CEUR-WS.org), Article 3, 10 pages. http://ceur-ws.org/Vol-2666/KDD_Converse20_paper_3.pdf

[12] Rricha Jalota, Priyansh Trivedi, Gaurav Maheshwari, Axel-Cyrille Ngonga Ngomo, and Ricardo Usbeck. 2020. An Approach for Ex-Post-Facto Analysis of Knowledge Graph-Driven Chatbots – The DBpedia Chatbot. In *Chatbot Research and Design: Third International Workshop, CONVERSATIONS 2019, Amsterdam, The Netherlands, November 19–20, 2019, Revised Selected Papers (Lecture Notes in Computer Science, Vol. 11970)*. Springer, 19–33. https://doi.org/10.1007/978-3-030-39540-7_2

[13] Lorenz Cuno Klopfenstein. 2019. Loquacious Blockly: Designing Building Blocks for Conversational Interfaces. In *Online Proceedings of CONVERSATIONS 2019 (3rd International Workshop on Chatbot Research)*. Article 24, 7 pages. https://conversations2019.files.wordpress.com/2019/11/conversations_2019_position-paper_24_web.pdf

[14] Nils Köbis and Luca D. Mossink. 2021. Artificial Intelligence Versus Maya Angelou: Experimental Evidence that People Cannot Differentiate AI-Generated from Human-Written Poetry. *Computers in Human Behavior* 114, Article 106553 (2021), 13 pages. https://doi.org/10.1016/j.chb.2020.106553

[15] Guy Laban and Theo Araujo. 2020. Working Together with Conversational Agents: The Relationship of Perceived Cooperation with Service Performance Evaluations. In *Chatbot Research and Design: Third International Workshop, CONVERSATIONS 2019, Amsterdam, The Netherlands, November 19–20, 2019, Revised Selected Papers (Lecture Notes in Computer Science, Vol. 11970)*. Springer, 215–228. https://doi.org/10.1007/978-3-030-39540-7_15

[16] Yi-Chieh Lee, Naomi Yamashita, Yun Huang, and Wai Fu. 2020. "I Hear You, I Feel You": Encouraging Deep Self-Disclosure through a Chatbot. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Article 48, 12 pages. https://doi.org/10.1145/3313831.3376175

[17] Dominique Shelton Leipzig, Bo W. Kim, Nicola Menaldo, and Laura Mujenda. 2019. I Am Robot: California's New Law Requires Disclosure of Use of Bots. https://www.privacyquicktipsblog.com/2019/05/i-am-robot-californias-new-law-requires-disclosure-of-use-of-bots/ (last accessed on 28th October 2020).

[18] Timothy R. Levine. 2019. An Overview of Detecting Deceptive Communication. In *The Palgrave Handbook of Deceptive Communication*. Springer, 289–301. https://doi.org/10.1007/978-3-319-96334-1_15

[19] E. Paige Lloyd, Jason C. Deska, Kurt Hugenberg, Allen R. McConnell, Brandon T. Humphrey, and Jonathan W. Kunstman. 2019. Miami University Deception Detection Database. *Behavior Research Methods* 51, 1 (2019), 429–439. https://doi.org/10.3758/s13428-018-1061-4

[20] David M. Markowitz. 2020. The Deception Faucet: A Metaphor to Conceptualize Deception and Its Detection. *New Ideas in Psychology* 59, Article 100816 (2020), 9 pages. https://doi.org/10.1016/j.newideapsych.2020.100816

[21] Victor R. Martinez and James Kennedy. 2020. A Multiparty Chat-Based Dialogue System with Concurrent Conversation Tracking and Memory. In *Proceedings of the 2nd Conference on Conversational User Interfaces*. ACM, Article 12, 9 pages. https://doi.org/10.1145/3405755.3406121

[22] Juliana Miehle, Nicolas Wagner, Wolfgang Minker, and Stefan Ultes. 2020. Culture-Aware Dialogue Management for Conversational Assistants. In *Conversational Dialogue Systems for the Next Decade*. Lecture Notes in Electrical Engineering, Vol. 704. Springer, 103–115. https://doi.org/10.1007/978-981-15-8395-7_8

[23] Jasabanta Patro and Pushpendra Singh Rathore. 2020. A Sociolinguistic Route to the Characterization and Detection of the Credibility of Events on Twitter. In *Proceedings of the 31st ACM Conference on Hypertext and Social Media*. ACM, 241–250. https://doi.org/10.1145/3372923.3404795

[24] Daniel Rough, Vincent Wade, and Benjamin R. Cowan. 2019. Democratising Speech Interfaces through Visual Programming: The B-SPOKE Project. In *Online Proceedings of CONVERSATIONS 2019 (3rd International Workshop on Chatbot Research)*. Article 12, 6 pages. https://conversations2019.files.wordpress.com/2019/11/conversations_2019_position-paper_12_web.pdf

[25] Henrik Skaug Sætra. 2020. The Parasitic Nature of Social AI: Sharing Minds with the Mindless. *Integrative Psychological and Behavioral Science* 54 (2020), 308–326. https://doi.org/10.1007/s12124-020-09523-6

[26] Tanja Schneeberger, Sofie Ehrhardt, Manuel S Anglet, and Patrick Gebhard. 2019. Would you Follow My Instructions If I was not Human? Examining Obedience towards Virtual Agents. In *Proceedings of 2019 8th International Conference on Affective Computing and Intelligent Interaction*. IEEE, 227–233. https://doi.org/10.1109/ACII.2019.8925501

[27] Weiyan Shi, Xuewei Wang, Yoo Jung Oh, Jingwen Zhang, Saurav Sahay, and Zhou Yu. 2020. Effects of Persuasive Dialogues: Testing Bot Identities and Inquiry Strategies. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Article 714, 13 pages. https://doi.org/10.1145/3313831.3376843

[28] Ziang Xiao, Michelle X Zhou, Q Vera Liao, Gloria Mark, Changyan Chi, Wenxi Chen, and Huahai Yang. 2020. Tell Me About Yourself: Using an AI-Powered Chatbot to Conduct Conversational Surveys with Open-Ended Questions. *ACM Transactions on Computer-Human Interaction* 27, 3 (2020), 1–37. https://doi.org/10.1145/3381804

[29] Kenta Yamamoto, Koji Inoue, Shizuka Nakamura, Katsuya Takanashi, and Tatsuya Kawahara. 2020. A Character Expression Model Affecting Spoken Dialogue Behaviors. In *Conversational Dialogue Systems for the Next Decade*. Lecture Notes in Electrical Engineering, Vol. 704. Springer, 3–13. https://doi.org/10.1007/978-981-15-8395-7_1

[30] Sebastian Zepf, Arijit Gupta, Jan-Peter Krämer, and Wolfgang Minker. 2020. EmpathicSDS: Investigating Lexical and Acoustic Mimicry to Improve Perceived Empathy in Speech Dialogue Systems. In *Proceedings of the 2nd Conference on Conversational User Interfaces*. ACM, Article 2, 9 pages. https://doi.org/10.1145/3405755.3406125