

Factors Affecting Conversation Design for Digital Political Personae

Abstract

Digital political personae are common on social media, representing a potential avenue to inform and engage citizens in political conversation. While personae that function as a digital extension of politicians and commentators are clearly identifiable, large numbers of automated personae of limited sophistication also engage in political exchanges in online spaces. Despite their prevalence, little prior work has addressed conversation design approaches to maximise the effectiveness of digital political personae in their interactions with users.

A review of the literature highlighted a range of approaches for effective use of digital tools in political contexts, including strategies for conveying information, sustaining engagement and employing and responding to emotive language on polarising topics. Examination of interactions with existing political personae on social media revealed that many of these approaches were limited or absent from the current conversational paradigm.

Chatbot software was used to explore methods that would address the issues identified with existing digital political personae through an iterative design process. A layered interaction scenario was developed that supports branching political conversation on a central topic, with a base of secondary topics to enhance the utility of the persona. The conversation design developed incorporates lessons from the literature on use of effective digital tools for political conversation, and has the potential to engage and inform large numbers of participants, as well as gather information from them for analysis.

Acknowledgements

Walter Langelaar

Table of Contents

Abstract.....	1
Acknowledgements.....	1
Table of Contents.....	2
1. Introduction.....	3
Design factors for digital personae.....	3
2. Literature Review.....	7
3. Methodology.....	14
3.1 Analysis of existing digital personae.....	14
3.1.1 Sentiment misclassification.....	15
3.1.2 Failure to engage.....	15
3.1.3 Off-topic replies.....	17
3.1.4 Media tweets.....	18
3.1.5 Third-party conversation.....	20
3.2 Interaction scenario development.....	21
4. Design findings.....	23
4.1 Derived design criteria.....	23
4.1.1 Sustained conversation.....	23
4.1.2 Closed conversation.....	23
4.1.3 Triage, translation and layering.....	24
4.1.4 Tone and language.....	24
4.2 Conversation design iteration.....	24
4.2.1 Approaches for sustaining conversation.....	25
4.2.2 Approaches for closing conversation.....	30
4.2.3 Incorporation of triage, translation and layering.....	31
4.2.4 Modifying tone and language to suit audience.....	34
4.2.5 Final design output.....	35
5. Discussion.....	37
5.1 Limitations.....	37
5.1.1 Social media analytics.....	37
5.1.2 Chatbot software.....	37
5.1.3 Conversation design.....	38
5.2 Applications of this research.....	39
6. Conclusion.....	41
7. References.....	42
Appendix A.....	46

1. Introduction

This research project will determine the factors affecting engagement with a chatbot persona, specifically in the area of political discourse. We hope to derive a list of conversation design criteria affecting conversation and engagement for digital political personae, and develop a platform that allows the effects of these criteria on user interactions to be further investigated. The project builds on the prior work ‘SAM: Virtual Politician’ (Langelaar, Gerritsen, & Smith, 2017), which used a chatbot to engage users in conversation from the perspective of a simulated digital politician.

The research project seeks to answer the question:

How can we approach conversation design to inform and engage people interacting with digital political personae?

Governments and non-governmental organisations (NGOs) are increasingly investing in the research and development of technological solutions to address issues around trust and engagement in political systems, while citizens are using existing digital platforms to organically solve the same problems.

One avenue being explored is the creation of digital political personae, either to promote a specific position or to act as an impartial actor with the aim of raising trust and engagement across the system as a whole, as seen in the SAM project (Langelaar et al., 2017).

Artificial personae governed by programming, machine learning or artificial intelligence are increasingly common (López, Quesada, & Guerrero, 2017), though there are presently few sophisticated examples, and none of them are openly operating in the political sphere. As machine learning and AI-driven personae become increasingly common, we may expect their incorporation into e-government initiatives as part of a broader push to expand voter engagement and participation.

Understanding the conversation design factors that affect user interactions with such personae will be crucial to their success or failure.

Design factors for digital personae

Digital personae require the combination of several technologies to function effectively, as well as a brand presence. We will focus on the technological and conversational aspects of digital personae, while keeping “branding” factors such as colour and graphical presentation constant. For example, some colours have particularly strong political associations, which may trigger strong emotional responses depending on the political identity of the user (Sawer, 2007), impacting interactions with the persona. Accordingly, neutral graphical presentation will be assumed, to focus on the factors affecting conversation design.

AI & Machine Learning

The artificial digital personae discussed in this paper rely on natural language processing, which may be powered by several different forms of artificial intelligence. An artificially intelligent system may be defined as any system that mimics intelligent behaviour, including responding to its environment and learning from it (Russell & Norvig, 2010). AI for natural language processing may

employ human defined rules and responses, data-driven statistical and probabilistic methods and artificial neural networks (Otter, Medina, & Kalita, 2016).

Human-defined responses allow for predetermined responses to textual or speech input, with the interaction governed by predefined rules. This approach does not employ statistical analysis or learning systems, and the impression of intelligence is limited by what can be foreseen when responses are defined.

Statistical and probabilistic methods use analysis of large datasets to better recognise system input and determine appropriate responses. Machine learning approaches allow for input recognition and outputs to evolve over time within defined rules.

Artificial neural networks employed in natural processing transform an input through processing in a number of hidden layers before generating an output. The “neurons” in these layers can perform weighted transformations, allowing for simultaneous analysis and interpretation of different aspects of the input. Neural networks require training using large datasets to develop a series of transformations that will produce the desired output. Recurrent neural networks are heavily used in natural language processing because they have the ability to “remember” previous inputs. A network that does not remember the previous input will respond to each new input in isolation, limiting the potential for human-like dialogue (Otter et al., 2016).

Commercial and open source natural language processing tools based on statistical methods or artificial neural networks are increasingly common, allowing for easy creation of a chatbot-driven digital persona. However, using neural networks to create purely generative dialogue has proven challenging even for well-resourced, high-profile artificial personae like Microsoft’s Tay, which was swiftly shut down after learning racist and offensive dialogue from users (Buiten, 2019). A persona based on predefined intents, rules and conversation design allows for different design factors to be explored without complex programming.

Persona control and coding

Commercially available and open-source chatbot software allows chatbot designers to control the flow of conversation between users and the chatbot. Several concepts are commonly found in most chatbot design:

- *Intents* allow the chatbot to recognise user intent based on their input.
- *Entities* allow the chatbot to store and retrieve information based on prior input
- *Dialogues* are the responses delivered by a chatbot to a given input or combination of inputs and entities
- *Links and structure* control how the conversation moves from dialogue to dialogue.

Control of these factors may be accomplished through coding directly in the chatbot interface, through uploading pre-prepared files in common file formats such as .json, or through a graphic user interface (Mattka, 2019).

A common feature of chatbots is the ability to “train” them based on user inputs. This is not equivalent to statistical training using artificial neural networks or AI; rather it is linking an input that the chatbot does not recognise to a predefined intent. This allows the chatbot to trigger dialogue based on this intent when the unrecognised phrase - or similar phrases as determined by the NLP software - are encountered in future.

Platform

A functioning chatbot requires a platform to interact with users and other entities. A platform provides an interface for users to interact with the chatbot, and exchanges inputs and responses with natural language processing functions. Some chatbot software allows for simplified deployment to specific platforms, while other platforms require the creation of middleware to exchange code between platforms and back-end processing functions.

Typically, chatbots are deployed via one of three types of platform:

- Social media websites
- Messaging applications
- Direct deployment on a website

The lines between these platforms may be blurred. A chatbot deployed via Facebook Messenger is effectively accessible via Facebook itself (as a social media platform), via the standalone Facebook Messenger application, and may also be deployed directly on a website by embedding a Facebook Messenger chat widget into the page.

Chatbots are generally employed for a specific purpose, such as providing information or taking orders or reservations. The mix of platforms they are deployed on will depend on the marketing or engagement strategy of the parent organisation.

Visual design

User sentiment towards a digital persona or chatbot will be affected by the degree of alignment between a user's own preferences and the perceived "brand personality" of the persona or chatbot. Aaker (1997) defines five core dimensions of brand personality:

- Sincerity
- Excitement
- Competence
- Sophistication
- Ruggedness

Visual design factors, such as the presentation of the chatbot's interface, choice of fonts, layout and graphics will contribute to perception of these dimensions, and therefore to user perception of the chatbot. The effect of visual design factors on the perception of digital political personae is a broad topic that merits more detailed investigation than can be covered here. Employing neutral visual design approaches allows the brand personality of the chatbot to be defined and driven by conversation design factors.

Conversation design

Conversation design also contributes to brand personality as described by Aaker (1997). The structure and content of conversation during interaction with digital personae is the primary mechanism by which information is conveyed to users. Understanding the factors that affect conversation design for chatbots is essential to constructing effective interaction scenarios with potential users.

This project will attempt to determine the conversation design factors affecting human interaction with existing digital personae. Once identified, these factors will be explored using an iterative design approach to develop an understanding of how decisions surrounding interaction design can impact digitally mediated conversation in a chatbot scenario.

2. Literature Review

A review of the literature on digitally mediated political conversation

Modern political institutions have been criticised as being ill-equipped to legislate effectively in the face of rapid technological and social change. Although roughly half of the world's peoples live in democratic societies, many are characterised by declining engagement, ineffective representation, increasing partisanship and a lack of political transparency. Against this backdrop, it is perhaps unsurprising that public confidence in political institutions is falling around the world, and democracy is experiencing a steady erosion as a result (The Economist Intelligence Unit, 2017).

Governments and non-governmental organisations (NGOs) are increasingly investing in the research and development of technological solutions to address these issues, while some citizens are using existing digital platforms to organically solve the same problems. This review will examine the literature to determine the criteria that will drive the creation of novel digital government or e-government systems such as chatbots and digitally mediated conversations, looking specifically at four key areas: engagement and participation; representation; polarisation and populism and information and transparency.

Engagement and participation

One of the more obvious signs of modern political dysfunction is falling engagement and participation in democratic processes such as voting, and many e-government solutions set out to address this issue. Meneses et al. (2017) describe a typical approach to an e-participation Platform “México Participa,” designed to foster engagement by connecting citizens directly with candidates during the 2015 Mexico midterm election. The platform was deliberately kept simple and the mechanism for voicing political opinions was contextualised in a way that made intuitive sense to participants by using a similar grading scale to that found in the Mexican school system. The system gave users a direct channel to interact with politicians and give a “grade” reflecting their current perception of key issues, as well as provide free-form responses on the most important issue to them. Like many such systems, México Participa operated for a limited time with a fixed purpose due to lack of ongoing funding, resourcing and support, despite the success of the program in driving engagement among its users.

Lukensmeyer (2017) observes that engagement and participation are among the most common aims for digital political tools, but suggests that this evinces a worrying underinvestment in tools that would drive higher-order engagement in public decision-making and deliberative democracy. However, Janowski (2015) describes a model of digital government that passes through several

stages, with Engagement as the penultimate stage before full digital maturity. In this case, the investment in participatory systems might be a positive sign that governments are following the journey towards digital maturity, with more complex systems likely to follow.

Low engagement and participation are interrelated with the other issues identified in the literature. Inadequate representation leads to dissatisfaction, which increases partisan dysfunction and lowers participation, further widening the gap between citizens and representatives and leading to greater dissatisfaction and further reduced participation.

Representation

Lukensmeyer (2017) describes the driving force behind political dysfunction in America as “the fact that the collective citizen voice is no longer a primary driving force in the direction of the country.” Wealthy political donors, partisan redistricting and a lack of accurate news reporting have resulted in a political establishment that does a poor job of representing the average citizen. The rise of government that does not accurately reflect the will of the average citizen is a key driving factor behind low participation and engagement. However, digital media, particularly the Internet, provides the potential to create a deliberative public space in which meaningful political communication and opinion formation can occur. From such a space, citizens can hold decision-makers accountable, improving the representative fit between citizens and politicians (Dahlberg, 2011).

The Obama administration attempted to create open online deliberative spaces with Change.gov, Ideascale and We The People, providing a common environment from which anyone could petition the government on issues that they felt passionately about, including calling out unfulfilled promises. However, the ability of citizens to engage in meaningful deliberation on these platforms was limited by their universalist, populist approach (Farina, Epstein, Heidt, & Newhart, 2014). The low barrier to use and simplicity of some actions – such as voting a particular petition or idea up or down – required minimal investment and knowledge from participants, making it difficult to draw meaningful information from these actions. The Web 2.0 assumption was that, left to their own devices, citizens would curate and organise their own system (Bland & Reisdorf, 2012), from which information could be harvested for government purposes.

However, few online civic spaces have had an observable impact on public policy and representation, as they have not been designed with robust mechanisms to translate public opinion into public policy or communicate the resulting changes to participating citizens. Gastil & Richards (2017) consider that “online engagement processes have had no visible impact on governmental legitimacy,” and that the often disconnected nature of these spaces serves to diffuse and lessen the impact of participation. The dissipation of civic energy required by engagement in disconnected

platforms, in addition to the lack of visible results, may further reduce enthusiasm and lower participation and engagement.

In the case of Change.gov and Ideascale, a further hurdle to meaningful outcomes was that citizens were not well enough informed about the workings of government to be able to participate. In some cases, the digital system created was simply the conventional process made available online. Although putting the process online makes it easier for users to access rulemaking materials and processes and to submit comments, citizens still needed to understand the original process to be able to make meaningful contributions (Solivan & Farina 2013). After several months' exposure to Ideascale some agencies abandoned the platform, as they were unable to justify the cost of analysing and responding to user participation against the limited value gained.

Information and transparency

Digital government systems are often seen as a tool for educating citizens and making political processes more transparent. However, government-led approaches often fail in one of two ways – they are either too open, producing little meaningful contribution due to a lack citizen knowledge (Solivan & Farina, 2013; Farina et al., 2014), or they are too closed, requiring participants to go to a specific online environment at a particular time, and interact with content or moderators in a prescribed way (Collin, 2015).

Solivan and Farina (2013) analysed the information behaviour of participants on the political e-rulemaking site RegulationRoom.org, and found the following informational barriers to participation:

1. A lack of awareness that relevant political action is occurring, and that participation is possible.
2. Information overload due to high quantities of complex or technical information.
3. A lack of familiarity with the process, and a lack of understanding of how to participate effectively.

They successfully employed several interventions to circumvent these issues, including:

1. Triage – limiting the information displayed to the most relevant.
2. Translation – simplifying the information to make it more understandable.
3. Layering – making more complex information available through hyperlinks so those who would benefit from it were able to learn more.
4. Indexing – similar to layering, the designers of RegulationRoom made all of the rules formerly and currently under consideration available in an index, so users could broaden their understanding if desired.

However, these interventions did not come without a cost, requiring heavier moderation of the website, its content and participants. Moderators were required to guide users towards information they might miss, encourage them to dig deeper into topics and point out connections to others' work.

Greater internet availability is associated with greater political knowledge and higher political engagement. This effect is most pronounced among citizens with low socio-economic status. This effect is likely due to the pervasive coverage of political stories across social media and digital news media. However, engagement overall is still low, and additional knowledge and engagement are largely the result of incidental learning during high-profile political events such as elections (Morris & Morris, 2013).

Epstein & Leshed (2016) cautioned that based on the experience of RegulationRoom “effective online civic engagement in policymaking is a resource-intensive endeavour, particularly when it involves human facilitative moderation.” They suggest that future systems should be designed to accommodate participation by less knowledgeable citizens by employing more plain language and simpler presentation. Failure to do will inevitably result in the system designers acting as mediators between politicians and the public.

Citizens have not waited for the government or NGOs to set up digital political resources. Collin (2015) notes that many non-government systems such as blogs, content creation platforms and peer-to-peer communication are used for civic and political engagement. However, these systems are reliant upon self-moderation and curation for informed and open-minded discussion to occur (Farina et al., 2014). These processes again rely on an educated, rational citizenry and are easily abused. Partisan participants can easily skew the direction and content of self-moderated platforms, resulting in rapid dissemination of inaccurate information.

Polarisation and populism

“The capacity to disseminate misinformation, wild conspiracy theories, to paint the opposition in wildly negative light without any rebuttal—that has accelerated in ways that much more sharply polarize the electorate and make it very difficult to have a common conversation.” – US President Barak Obama, in Remnick (2016).

Polarisation is the last challenge for digital government, and it is enabled by each of the other three. Lack of perceived representation fuels polarised thinking, as citizens become more extreme in their views to differentiate themselves (Westfall, Van Boven, Chambers, & Judd, 2015). Low participation makes it easier for more polarising politicians to attain power (Lukensmeyer, 2017), while a lack of good information ill-equips citizens to resist the lure of demagogues (Mendes, 2016).

Ironically, higher polarisation appears to drive higher engagement; politicians with more extreme positions have more Twitter followers, even when controlling for traditional media exposure (Hong & Kim, 2016). In the US, Donald Trump's successful candidacy for president relied in part on his more extreme positions, which strongly appealed to a segment of the American electorate (Mendes, 2016). When considering the rules for successful engagement derived from RegulationRoom (Solivan & Farina, 2013), it is apparent that many Trump supporters were resistant to complex information and lacked familiarity with the political process. Then-candidate Trump demonstrated that only triage and translation are necessary to keep supporters engaged – by limiting the amount of information available and translating it into simple terms, more complex information is unnecessary to secure substantial political support. We might hesitate to design an e-government system in this way – though we should be prepared for the potential of a counter-system operating in this fashion to push a point of view.

Further exploration of emotive language was conducted by Tan, Danescu-Niculescu-Mizil & Lee, (2016), in an examination of the online forum “Change My View” on reddit.com, which allows users to submit a position and invite counterargument from other users. The authors found that most participants were unsuccessful in changing the view of the original user, but that more successful arguments were likely to exhibit common characteristics.

Characteristics of arguments include:

- Arousal (words which are more likely to reflect intense emotion)
- Concreteness (words which denote something perceptible rather than concepts)
- Dominance (the decree of control expressed by words)
- Valance (how “pleasant” words are)

More successful arguments exhibited progressively more arousal, dominance and valance through continued discussion, but began and ended with less concreteness. Opening up discussion and drawing conclusions with a degree of hedging may allow participants to create space for discussion and later invite support of a conclusion, rather than commanding it.

The ability to precisely target individuals is already in evidence in political campaigns around the world, and is both potentially beneficial to a digital government system, and open to abuse. Chester & Montgomery (2017) urge the creation of enforceable best practices to “ensure that digital technology enhances democratic institutions, without undermining their fundamental goals.”

Despite the common perception that digital and social media is largely behind growing partisan polarisation, Boxell, Gentzkow and Shapiro (2017) found that polarisation has increased the most in populations that are least likely to use the Internet and social media.

Political chatbots

Openly political chatbots are uncommon, but chatbots that are not openly declared as either politically oriented or as chatbots are numerous on some platforms, such as Twitter, where they are heavily engaged in sharing links to external media (Wojcik, Messing, Smith, Raine, & Hitlin, 2018).

Undeclared politically oriented chatbots have the potential to significantly affect public opinion by controlling public discourse in online spaces. Twitter has closed millions of accounts suspected to be controlled by undeclared chatbots (Neudert, 2018).

The undeclared nature of political chatbots may be a factor in allowing them to influence public discourse, by posing as ordinary commenters rather than agents with an agenda. Some commentators regard chatbots as a potential threat to political discourse because of their lack of transparency and regulation, and potential to engage with and influence large numbers of people (Susskind, 2018). A small number of openly political chatbots have received popular attention (Prakash, 2018), but do not have the reach of the much larger number of undeclared chatbots.

Chatbots and natural language interfaces presently depend on large volumes of interaction data, but misinterpretation is common despite the availability of huge data sets. Even the most sophisticated chatbots tend to break down in longer conversation threads, as they are unable to appropriately relate to prior conversation. Accordingly, chatbot design tends to focus on driving users towards specific goals and conversation pathways, while providing acceptable responses to recover the conversation in the event of misunderstanding (Følstad & Brandtzæg, 2017).

A conceptual design that abstracts frameworks and goal may address some of these challenges (Telang, Kalia, Vukovic, Pandita, & Singh 2018). Dividing the functions of a chatbot into sub-functions that work together in order to achieve their goals has the potential to allow for longer conversation threads to be maintained coherently, and for the chatbot to continue to shepherd users towards desired goals in the event of misunderstanding.

Conclusions

The existing literature suggests the use of chatbots and digital personae for political purposes has several challenges to contend with. To sustain engagement, these systems must be meaningfully coupled to political activity, and must deliver visible results from citizen participation. Participation can be facilitated through employing techniques developed for prior systems, including triage, translation, layering and indexing of information. The potential use of chatbots or artificially intelligent systems to communicate information through a conversational approach has not been well explored in the literature and may present an avenue that circumvents the challenges of moderation experienced in previous explored systems.

If such a political chatbot were to produce a large volume of meaningful output, we would then have other considerations: who will determine how citizen's contributions are interpreted, analysed and weighted; how can they be incorporated into decision-making, and how will technology be used to sort, analyse and summarise contributions? These questions represent an important avenue for future research as more complex systems for e-government are developed.

3. Methodology

3.1 Analysis of existing digital personae

The literature review highlighted that few examples of political chatbots are presently operational. However, many politicians and political parties are active on multiple digital platforms and are essentially operating digital political personae as extensions of their real-world identities. This provides an opportunity to examine digital political discourse to determine potential design factors.

Several platforms were considered for analysis. While Facebook is a common platform used by most politicians and political parties, an examination of the Facebook pages of several major politicians and parties did not yield large amounts of conversation data, with use focused around campaigning, policies, legislation and events. YouTube and Instagram accounts examined were similarly focused on promotion and had low levels of conversational engagement. Twitter was selected for more detailed examination of digital political conversation, because political users engage in back-and-forth conversation much more often than on the other platforms examined, and because of the large number of public exchanges available for analysis from a range of political figures and parties.

Twitter's REST api (Twitter Inc., 2020a) allows for large volumes of tweets to be retrieved in formats that can be readily analysed in open-source applications such as RStudio (Mishra, Khanna, Kumar, & Sinha, 2017). A data corpus was assembled by sampling tweets from multiple political Twitter accounts. Accounts were specifically selected to cover broad political territory, including politicians from each of New Zealand and America's major political parties. Both "high-profile" (>1m followers) and "low-profile" (<1m followers) accounts were included in the sample.

Detailed examination of the data corpus generated through bulk sampling of tweets revealed several issues that affect the utility of using this methodology to inform conversation design. The initial intent of the study was to use a combination of sentiment analysis (Cresswell, 2009) and content analysis (Leavy & Prior, 2014) to inform the development of an interaction scenario. The limitations of available tools in conducting a robust sentiment analysis led to greater focus on content and thematic analysis (Braun & Clarke, 2012).

The sampled tweets were coded and grouped according to common themes that were identified. This approach yielded clear themes from analysis of just a few hundred tweets from the personae selected to compile the data corpus. While strong themes were readily identifiable through manual coding of responses, we do not consider that the sample is sufficiently robust to make quantitative claims about the relative prevalence of these themes across political conversation on Twitter. Variation in thematic prevalence was observed between personae which could be considered to have similar political leanings. Coupled with the subjective nature of assigning conservative-leaning versus progressive-leaning labels to sampled accounts, we consider that this observation would make attempts to support conclusions such as "conservative leaning Twitter accounts attract more "praise/insult" type responses to posts" unsafe.

Accordingly, quantitative methods were not employed.

3.1.1 Sentiment misclassification

A sentiment analysis package employing a dictionary approach was used to analyse tweets containing specific hashtags and @usernames, as well as tweets and replies associated with the accounts of prominent politicians. The dictionary approach to sentiment classification employed aggregates the pre-defined sentiment scores of non-common words in a conversation fragment such as a tweet in order to generate a positive or negative overall sentiment classification (Goyal & Kakkar, 2019).

A detailed examination of the resulting classifications showed that the sentiment package consistently misclassified tweets. For example, the following tweet was classified as neutral to positive, whereas in context it would be understandable to a human as negative:

"@BernieSanders my Oakland landlord showed up one day and doubled the rent. "If you can pay it you can stay. If not I have people waiting."

Bakliwal et al. (2013) describe classification of political tweets as challenging because of the frequent use of sarcasm, comparative expressions, domain-specific terminology and inferred sentiment from otherwise neutral conversation. Responses that combined two or more of the common forms of exchange described in Figure 1 proved particularly challenging for accurate sentiment analysis.

The weightings ascribed to words vary depending on the sentiment lexicon employed. However, the misclassification described by Bakliwal et al. (2013) was expected to similarly apply to different lexicons, because they are primarily caused by the effects of context and inference that are not tied to the raw sentiment score of words. Testing using a different sentiment lexicon resulted in similar misclassification, confirming this hypothesis. Use of automated sentiment analysis was therefore disregarded in favour of manual coding and classification of conversation, as described in Cresswell (2009). A random purposeful approach was taken, analysing batches of 100 tweets or replies chosen at random from the selected accounts.

3.1.2 Failure to engage

The most common issue identified with using Twitter exchanges to inform conversation design is that sustained conversation on the platform appears rare, despite the greater prevalence of back-and-forth exchanges compared to other social media platforms hosting political personae. Most exchanges between political personae and other users of the platform comprise a single statement and response.

Shegloff and Sacks (1973) described adjacency pairs as a model for a unit of conversation, in which the second part of an adjacency pair responds to the first and effectively closes that part of the conversation. Littlejohn, Foss and Oetzel (2017) highlight a number of adjacency pairs, including assertion-assent/dissent, insult-response, and question-answer.

Comparing sampled Twitter exchanges against common adjacency pairs revealed a high prevalence of "open" or "unclosed" exchanges between participants using the platform. Four common forms were identified through content analysis, shown in Figure 1.

Of the common forms of exchange identified, the only “closed” adjacency pair was an assertion followed by assent / dissent. This exchange naturally terminates – further engagement was rare in the responses sampled. Praise/insult is typically considered to be the opening of an adjacency pair, but in the tweets sampled this opening was typically left without a response, leaving the conversation effectively open. Questions that would be expected to elicit an answer to complete an adjacency pair were also overwhelmingly left without a response. Comments that invite the continuation of a conversation, but do not act as part of an adjacency pair, were also rarely responded to.

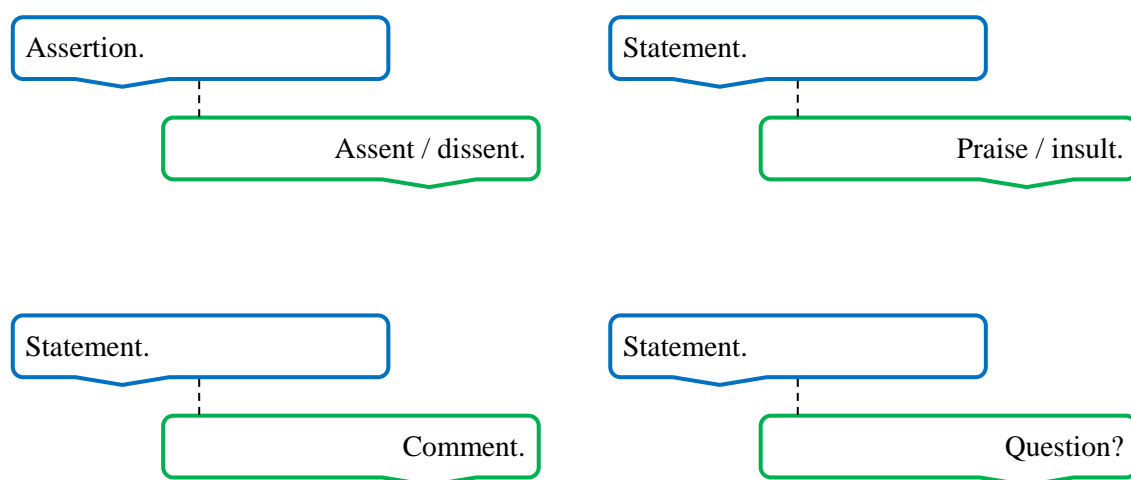


Figure 1. Common forms observed in political exchanges on Twitter.

The prevalence of different response types varied between political personae and between tweets from those personae, and also with overall engagement with the initiating tweet (as measured by number of responses). The most frequently observed pattern was that the proportion of simple “agreement / disagreement” or “praise / insult” responses increased as the total engagement with the initiating statement increased. This finding was not unexpected, as these types of responses are simpler to compose – hence initial statements that engender many responses are more likely to be ones that can be responded to with a simple statement of agreement or disagreement.

Combinations of response types were frequently observed. Common instances of combined forms were agreement combined with praise or disagreement with insult, and rhetorical questions indicating disagreement or insults.

Responses were classified as a “comment” when they expanded on the information contained in the original statement, and did not solely state agreement / disagreement or praise / insult for the originating persona (although comments were often combined with these response types). Non-rhetorical questions were the least frequently observed interaction, and responses to either comments or questions were extremely uncommon (see Figure 2).

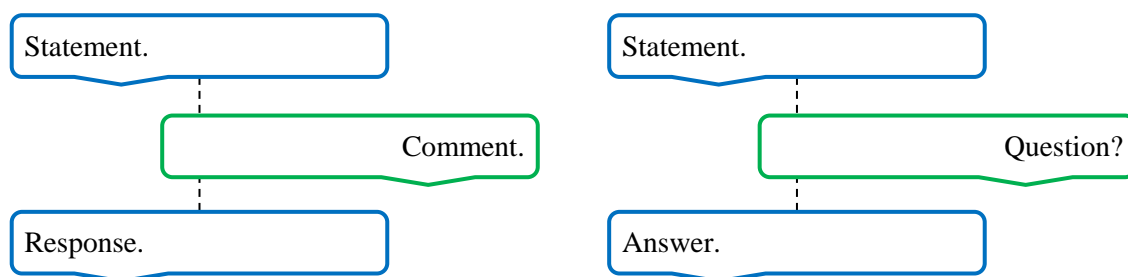


Figure 2. Uncommon forms observed in political exchanges on Twitter.

A second response from the originating persona was often followed by a subsequent comment or question when it was observed, but longer exchanges were not observed in the data corpus collected. This suggests that while digital personae on Twitter may have millions of followers and tens of thousands of replies to individual tweets, sustained conversation on the platform is rare.

3.1.3 Off-topic replies

Any response to an initial tweet that did not directly or tangentially engage with the subject matter of the tweet was considered to be an off-topic reply (see Figure 3). Purely off-topic replies were excluded from analysis when considering common forms of exchange (see Figure 1), though off-topic replies that also contained praise or insults for the originating persona were included.



Figure 3. An off-topic reply to a political tweet

For each of the individual accounts of political personae identified and examined, off-topic replies constituted a substantial fraction of all responses. While off-topic replies might be considered

to reduce the utility of data corpus analysis, they are an important metric of engagement in their own right.

In the corpus examined, the proportion of off-topic replies varied between tweets and between personae. The number of off-topic replies varied in inverse proportion to the total engagement with the tweet; likely driven by the similarly observed proportion of simple responses increased as engagement with the initial tweet increased.

3.1.4 Media tweets

Tweets that either consist solely of an image or video, or of an image or video with accompanying text, were categorised as media tweets. Media tweets present a challenge to bulk analysis, as the media content cannot reliably be parsed by automated tools. The intent of media may be determinable from accompanying text, but the unreliability of sentiment analysis for political tweets described by Bakliwal et al. (2013) makes any inference as to the purpose of accompanying media unreliable. Media tweets also frequently occur without any accompanying text, which presents challenges to this type of analysis.

Inclusion of media in tweets can improve the likelihood of the message being shared (or retweeted), particularly if the content is “news” that can be classified as negatively affecting (Hansen, Arvidsson, Neilsen, Colleoni, & Etter, 2011). This would appear to align with the goal of increased engagement and participation highlighted by Lukensmeyer (2017) for digital political tools. The previously identified failure to engage in real conversation would suggest that Lukensmeyer’s concern that broad engagement may come at the expense of higher-order participation is broadly correct – at least on Twitter.

While images and some videos can be scraped in bulk using Twitter’s API, associating them with original responses and classifying the intent behind them was done manually to ensure accuracy.

Media tweets were categorised into one of three types: informative, emotive and memetic.

Informative media

Informative media tweets use embedded media to convey additional information not contained within the text of the tweet itself. Informative media was found to commonly consist of videos or infographics produced by third parties, with fewer instances of this form of media post being created and shared directly by the accounts analysed. Informative media was found to be more commonly employed by the accounts sampled than by respondents.

The use of informative media is not unexpected, as digital political personae act as a platform to communicate political positions and campaign information, which frequently requires more detail than can fit into a single tweet.



Figure 4. Examples of “informative” media tweets

Emotive media

Emotive media tweets use media to enhance or accentuate the primary message of the tweet. This media is distinct from informative media in that its primary purpose is not to provide information, but to elicit positive or negative feelings through the use of emotive imagery.

Emotive media was found to be commonly employed by both the accounts sampled and by users responding to tweets.



Figure 5. Examples of positive and negative “emotive” media tweets

Memetic media

Memetic media content may also be informative or emotive, but its primary composition is the presentation of a simple idea, concept or piece of information in an impactful readily sharable format. Memetic content often follows a standard format known as an image macro, consisting of an image on

a black background, overlaid with white text. Image macros in particular are commonly referred to as memes, though any form of information or behaviour spreading through a population potentially qualifies as a meme. Short videos (typically between 10 and 30 seconds) and simple charts were other commonly observed forms of memetic media content.

Classification of memetic media content is challenging. To some extent, any form of message that can act as a vehicle to convey as a unit of information may be considered a meme. For the purposes of this study memetic media was considered to comprise images following the popular image macro format, as well as images or short video focused on conveying an idea rather than solely conveying information or eliciting an emotional response. Significant overlap was observed with the other types of media identified.

Memetic content was much more commonly observed in replies to political personae rather than originating with them.



Figure 6. Example of memetic media content

3.1.5 Third-party conversation

Exchanges between users responding to a political persona are common on Twitter. These exchanges were not captured by bulk sampling tools used to assemble the data corpus, and so were not included in the analysis.

3.2 Interaction scenario development

Long conversational exchanges may generally be more common on Twitter than in the observed sample of tweets on political accounts. Honey and Herring (2009) found that the median number of exchanges in conversations in randomly sampled tweets was 3, and the mean was 4.62. From the sampled exchanges between users and political personae *no* conversations longer than three messages were observed, suggesting again that the current paradigm for conversation with digital political personae is limited when compared to general conversation on the same platform.

No prior work was identified that explicitly explored the difference in conversational factors between digital political conversation and more general conversation. Conversation design that explores this space is therefore breaking new ground. However, we note that some of the observed features of conversation with digital political personae are not necessarily caused by the type of conversation or the political nature of the personae. Limited conversational exchanges with any given respondent are to be expected when initiating comments attract thousands of responses, because a human-driven account would not have the time to exchange messages with more than a fraction of respondents. The existing conversational paradigm does not conform to the strategies identified for effective political communication identified in the literature review, highlighting the need for novel approaches to enhance conversation with commonly encountered digital political personae.

A conversation design approach was required that would allow exploration of the factors identified through the analysis of political discourse on Twitter, and the prior work highlighted by the literature review. Commercially available chatbot tools were investigated for this purpose, because they allow for rapid prototyping, iteration and testing of conversational structures and language.

Initial prototyping was undertaken in a simple custom environment that was able to select responses based on keywords in messages. While this prototyping confirmed the viability of using chatbot software to explore conversation design, it highlighted several critical requirements that would support effective use of such software.

IBM Watson Assistant was adopted as a more comprehensive platform that would support complex conversation design. Watson Assistant's primary advantage as a conversation design platform is its mature natural language processing capability, and associated ability to perform fuzzy matching of intents and entities. Watson Assistant was linked to a Facebook Messenger bot to enable testing of the conversation design developed, although the chatbot can be linked to and accessed from multiple platforms.

Chatbot solutions allow for database queries or business processes to be activated by users through a conversational interface, with the chatbot software – in this case Watson Assistant – interpreting the input and responding with the appropriate information from a database. An example architecture design is shown in Figure 7.

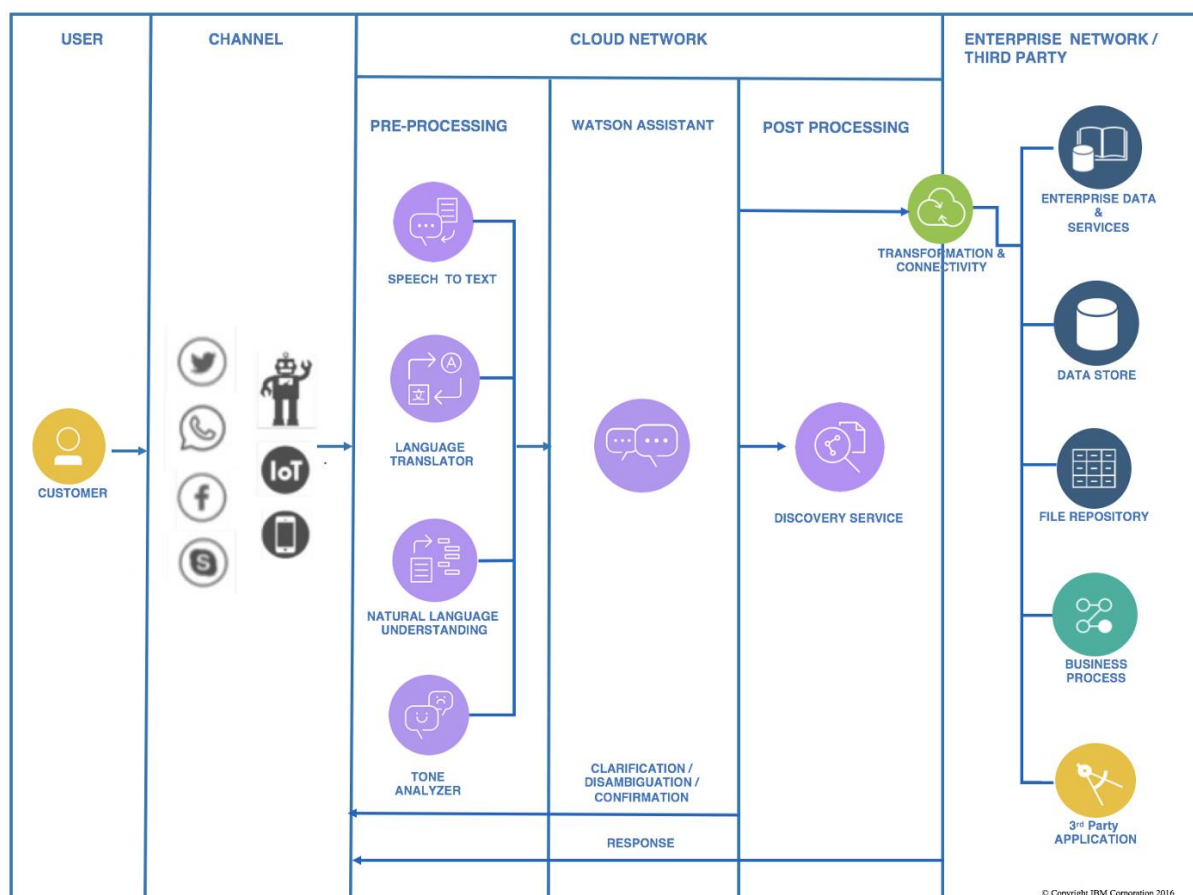


Figure 7. Sample architecture diagram for integrating Watson Assistant with other services (IBM, 2016).

For the purpose of conversation design, post processing and integration with file repository or data services was not required. Design investigation was therefore confined to the pre-processing and Watson Assistant sections of Figure 7. The natural language capabilities of the Watson Assistant service enable understanding user intent from a broad range of utterances, while only requiring limited examples. The tone analyser was also explored as a potential approach to generate sustained conversation.

Interface effects, such as variation in the ability of digital political personae to employ colour and branding, were deemed out of scope for this study. Language translation and speech to text functionality of Watson Assistant were similarly excluded from detailed investigation, as they primarily affect the user's ability to interact with the persona, rather than the features of the persona itself.

The communication channel employed by users may influence conversation, due to both interface effects and differences between user populations. Population effects are beyond the scope of this digital conversation design study, though the conversation design concepts explored can potentially provide a vehicle to investigate differences between user populations of different channels.

4. Design findings

4.1 Derived design criteria

This thesis set out to determine novel methods for conversation design for digital political personae based on a review of the existing literature and analysis of existing examples of digital personae. Analysis of political discourse on Twitter was considered in the context of the information gathered during the literature review to derive criteria to explore during conversation design. In order to maintain a focus on conversation design, other factors that might affect user interactions such as branding and platform effects were held neutral and constant where possible.

Content analysis of existing digital political personae on Twitter revealed that while engagement with personae appeared to be high, sustained conversation was extremely rare. Several apparent limitations of the prevailing paradigm for communication employed by digital political personae were identified. Other design factors, such as the impact of channel, format and branding on user engagement and perception, were excluded from consideration for the purposes of this research.

4.1.1 Sustained conversation

The most significant limitation to political discourse on Twitter under the present paradigm is the lack of sustained conversation; most exchanges are limited to a single statement and reply. A significant volume of user responses to political personae comprise low-effort positive or negative affirmations, praise, or insults. While conversation may terminate naturally following these responses, each of these rapid terminations represents failed potential to continue the conversation and generate additional engagement and participation that are typically a common aim for digital political tools (Lukensmeyer, 2017).

Consequently, conversation design should meet the design criteria of generating sustained conversation that supports deeper engagement and participation.

4.1.2 Closed conversation

Almost all exchanges sampled between political personae and other users are short, and many do not experience “closure” – the conversational exchange does not terminate at a satisfactory point, with more dialogue implied by the state of the conversation but never realised. The existing personae sampled on Twitter almost never respond to public messages or questions, whether or not they originate with another user or are asked in response to an initiating message from the persona. This leaves a significant number of exchanges without closure.

Ending a conversation part-way through an exchange might reasonably be expected to be unsatisfying for participants. Where possible, conversation should be structured so that conversations on a given topic reach an appropriate endpoint.

4.1.3 Triage, translation and layering

Despite the limitations observed in conversational exchange on Twitter, the structuring of information presented in tweets by political personae closely matched several of the interventions identified by Solivan and Farina (2013) to reduce barriers to participation in political discussion. Twitter's low character limit of 280 characters promotes triage and translation mechanisms that limit dialogue to the most relevant information and encourage the use of simple, direct language. Political personae on Twitter were observed to frequently make use of external links and videos, creating layered access to information that allows users to pursue additional information if desired.

Although Watson Assistant does not have the 280 character limit of Twitter, conversation exchanges should make use of simple language addressing only the most essential information, and should make use of hyperlinks to provide access to deeper information where possible. Replies should be limited to 280 characters to make the resulting conversation platform-agnostic; able to be published on Twitter or any other online platform.

4.1.4 Tone and language

Tan et al. (2016) identified emotive language as a key component of online discussion. Analysis of political discourse on Twitter has confirmed the use of emotive language dominates the most common forms of exchange (assent / dissent and praise / insult).

The conversation design will endeavour to create a structure that allows for investigation of the impact of tone and language on participation and engagement.

4.2 Conversation design iteration

A framing scenario was chosen to inform conversation design iteration. A digital political persona was created with the intent of being able to hold a sustained conversation on climate change and the environment, while also being able to respond appropriately (but briefly) to other questions that may be asked outside of this primary conversation. Climate change was chosen as the primary topic because many examples of dialogue on the topic were observed in the exchanges sampled from Twitter, and there were a range of commonalities and disagreements in expressed views that must be accounted for.

Conversation design approaches were developed through iterations of this scenario to address the design criteria developed through examination of the literature and existing digital political personae. During each iteration, the conversation design was tested using sample user dialogue similar to the comments observed on Twitter. Comments were also selected at random from the complete set of all scraped Twitter comments to test the response of the persona to different topics. Persona responses were limited to 280 characters in order to make the resulting conversation "platform agnostic" – the chatbot software underpinning the conversation could potentially be linked to any digital platform.

4.2.1 Approaches for sustaining conversation

Establishing focused user intents and dialogues

Initial conversation design covered a small number of potential exchanges relating to climate change. These exchanges typically involved the digital persona responding to questions about climate change. User utterances were based on common queries and statements observed on Twitter when scraping political tweets about climate change. These utterances were grouped into intents, each of which triggered a dialogue response when detected.

Watson Assistant's natural language processing was found to link many climate change related queries or statements to these intents, even where user statements were only loosely related to the existing utterances used as examples. Initial design iterations sought to take advantage of this capability, linking many user utterances to a limited number of responses (see Figure 9). Responses in this design iteration were necessarily simplistic, to ensure that they were loosely appropriate replies to multiple potential user utterances. Initial testing showed that this approach created a passable facsimile of existing conversations with digital personae sampled from Twitter, with the same limitations; exchanges were shallow, did not conclude cleanly, and in most cases would feel like participants were not directly engaging with one another.

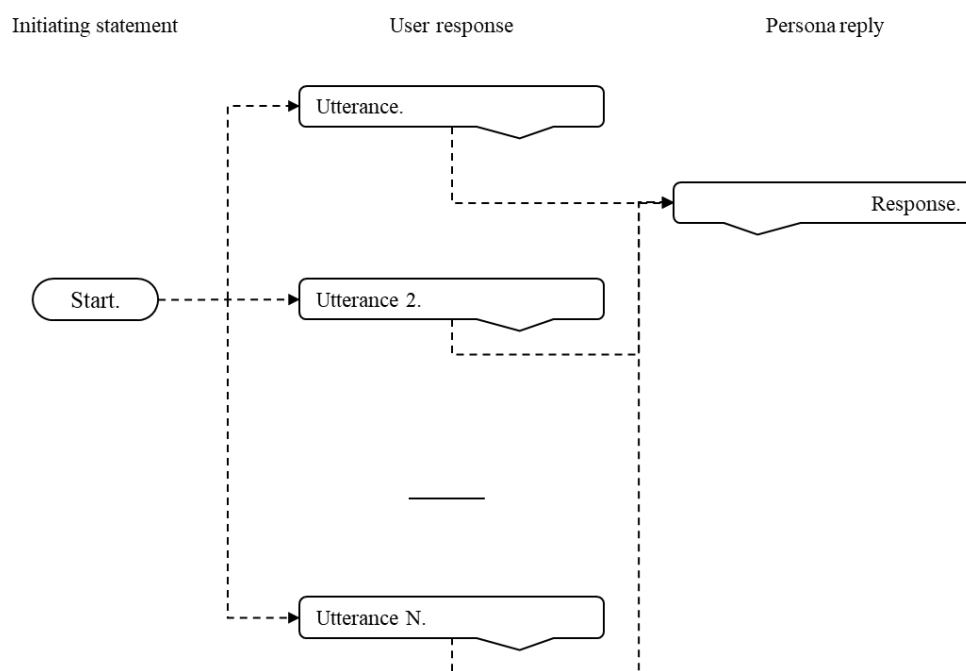


Figure 8. Initial prototype conversation linking multiple utterances to each response.

A broader range of intents and responses was created to capture more potential queries, statements and responses. Intent segmentation within topics creates a significantly larger pool of required dialogues, but also enables more granular engagement with users, expanding the potential for longer or more detailed exchanges. In early iterations, the utterances shown in Figure 8 were all linked to the same intent, and so would trigger the same response. However, the utterances in group A are statements with a negative sentiment, while the utterances in group B and C are queries with a neutral sentiment, with narrow topic focus in group B and broad topic focus in group C. Associating

these utterances with three separate intents allows for conversation that more closely addresses user intent, in addition to providing more options for sustained conversation.

User examples (7) ↑	
emissions targets cost too much	A
emissions targets will hurt business	
emissions targets won't work	
should government set emissions targets	B
should government set greenhouse gas targets	
should we limit greenhouse gas emissions	
what should government do about global warming	C

Figure 9. Intent segmentation within topics.

Establishing a broad range of focused user intents and response enabled specific responses to many queries in a specific topic. However, the persona gave a generic response to any questions outside this topic.

Extra-topic responses

In order to explore the effort required to create a more “rounded” persona, intents and responses were defined outside the primary topic. Additional responses enhance the overall utility of the persona, as well as making it more human-like which may further increase the perceived utility (Reitz, Benke, & Maedche, 2019). Intents and responses were informed by the dataset scraped from Twitter, as well as user queries collected during the ‘Sam: Virtual Politician’ project (Langelaar et al., 2017).

Creation of responses outside the primary topic have the effect of making the persona more human-like by preventing repetition of a standard response when extra-topic intents are detected. This approach also enables sustained (though limited) conversation outside of the primary topic. Intents were bulk-created by generating a series of standard utterances and using variable substitution on the key words in each utterance. The resulting set of intents could be directly uploaded to the chatbot software.

Combined with topic-specific dialogues, the effect observed in iteration was to create a persona that was able to respond to a large number of user utterances with an appropriate response, as shown in Figure 10. This design iteration was able to sustain longer simulated conversation – provided users continued to ask questions that would trigger new responses. The “failure to engage” problem identified in analysis of existing personae was only partially addressed by this approach: while the designed persona actively responds in a way that most existing digital personae do not, it is still limited to a single statement and response in any on exchange. Longer exchanges on the same broad topic are effectively a sequence of two-part exchanges triggered in sequence by the user.

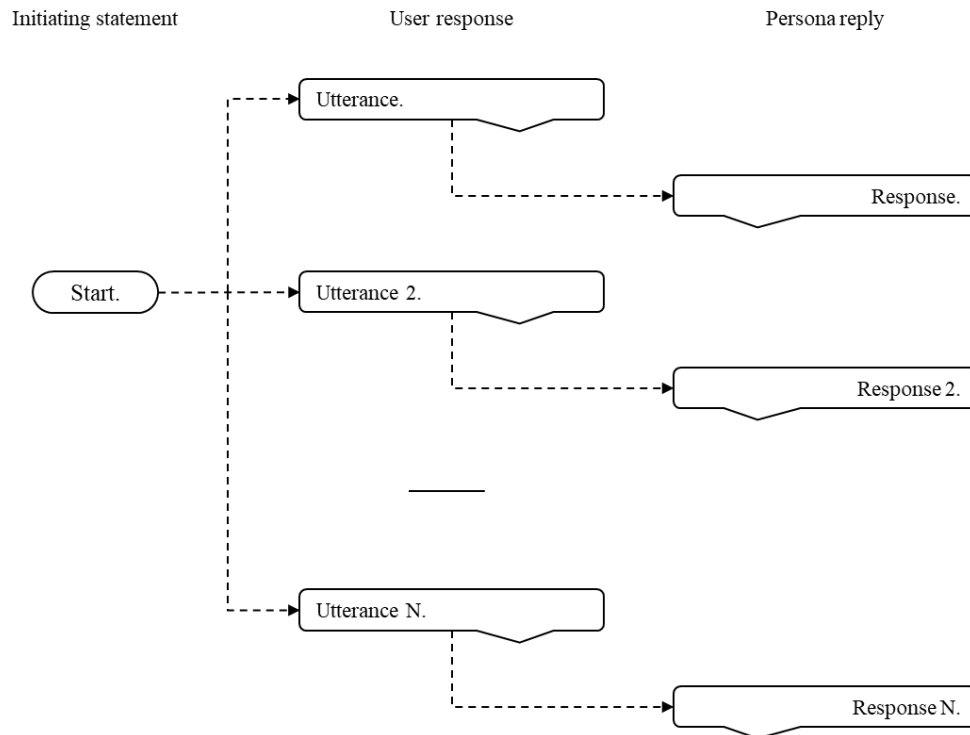


Figure 10. Individual responses to a large number of user utterances.

Dialogue chains

The limitations identified in the iteration shown in Figure 10 informed the development of longer dialogue chains. Dialogue chains were created using queries or linking statements to prompt additional responses on the same topic, enabling more detailed conversation to occur. When questions could be answered in multiple ways (for example Yes / No), multiple branching responses could be created where appropriate. An example of a query-based dialogue chain is shown in Figure 11.

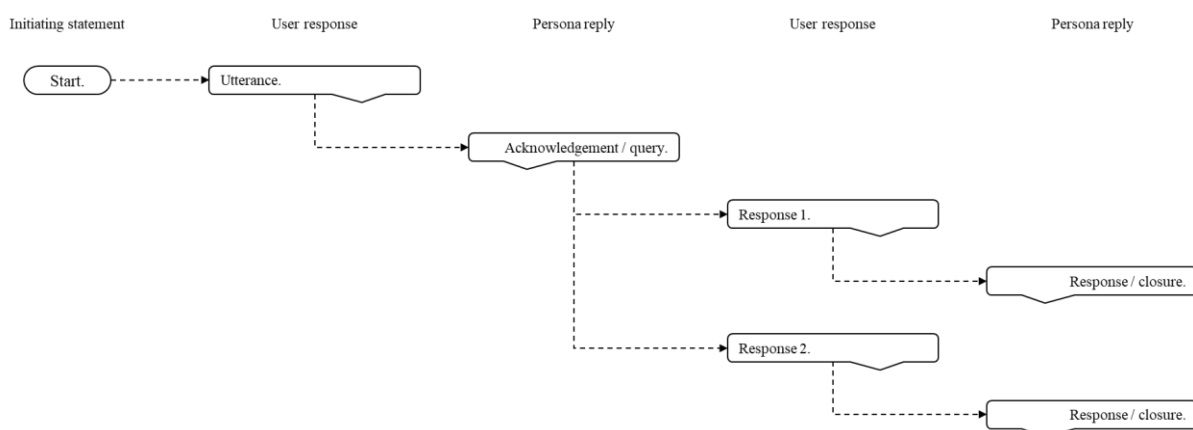


Figure 11. Example branching dialogue chain following a query.

Dialogue chains were primarily extended using the “question-answer” adjacency pair identified by Littlejohn, Foss and Oetzel (2017), but use of the “challenge-response” adjacency pair was also explored. After asking for a user’s opinion on the previous statement, the conversation was extended

in one of three ways: if the user agreed with the statement, an acknowledgement was given and the conversation moved onto the next topic; if the user disagreed with the statement, the persona challenged the user to provide justification; if the intent of the user could not be determined, the persona used a non-committal statement to prompt elaboration. The effectiveness of this approach was strongly dependent on the ability of the natural language processing capability of the chatbot software to correctly determine the intent of the user.

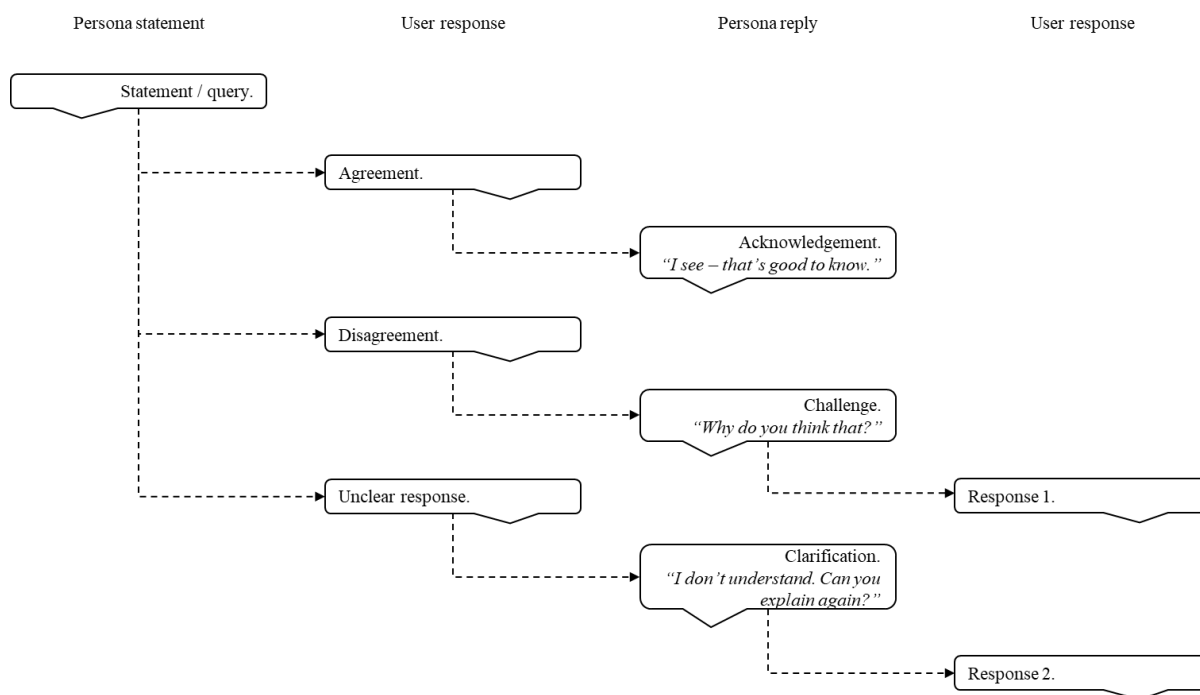


Figure 12. Example dialogue tree incorporating a “challenge-response” adjacency pair.

Digressions

Extra-topic or off-topic responses were a commonly observed feature of exchanges with existing digital political personae. During dialogue chains, the chatbot searches for specific intents to trigger the next dialogue in the chain – off-topic responses were set to trigger a standard unknown response. Testing with this configuration revealed that it was impossible to effectively break out of dialogue chains without a mechanism to effectively address off-topic responses.

Dialogue nodes in Watson Assistant allow for both inbound and outbound digressions. This enables conversation design that allows digressions out of a dialogue chain (by allowing outbound digressions), while preventing users from jumping into the middle of a dialogue chain (by forbidding inbound digressions). Digressions were found to present a natural approach for effectively layering information, as described by Solivan and Farina (2013).

For example, a user might be in the middle of a dialogue chain about emissions targets for greenhouse gases, but might ask the question “What are the effects if we do nothing about global warming?” The chatbot software can digress to an appropriate dialogue to answer this question, before returning to the original dialogue chain, or closing the conversation if a return was not enabled. Additional dialogue was added to inform users that they were returning to the original dialogue chain when this occurred. An example dialogue tree showing a digression structure is shown in Figure 13.

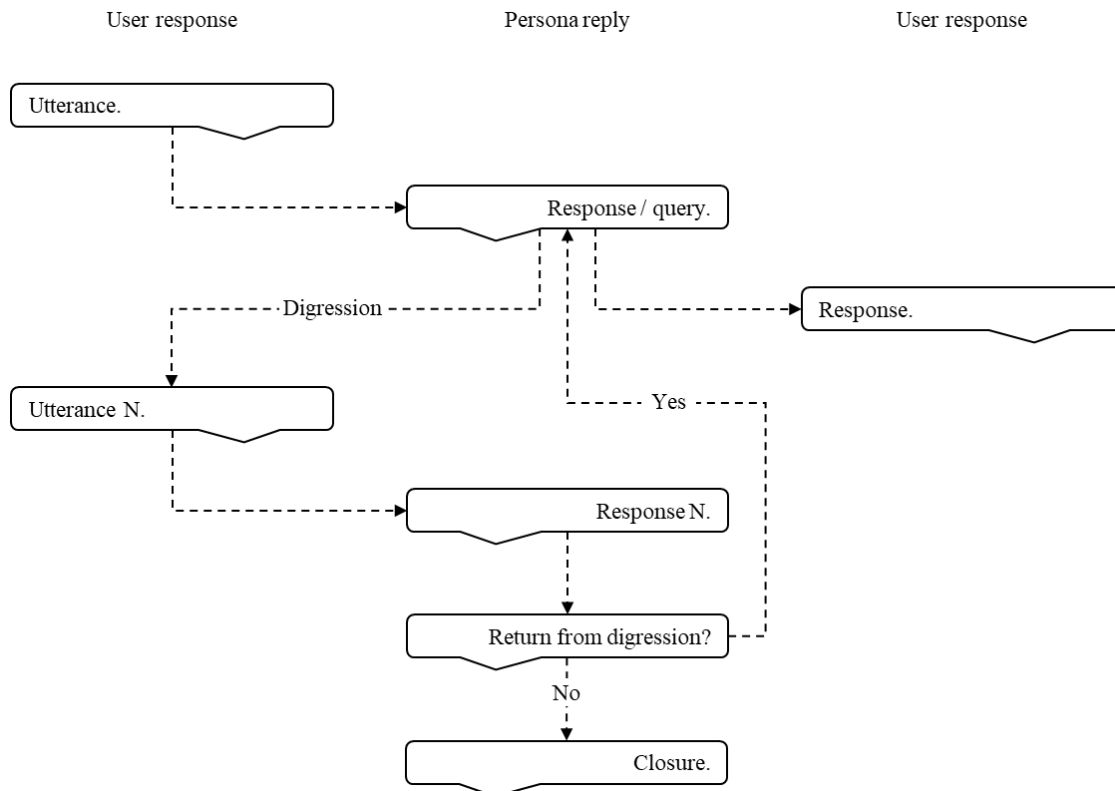


Figure 13. Example dialogue tree incorporating a digression.

The ability of the chatbot software to handle digressions had a notable limitation: conversation could continue to digress away from the primary dialogue chain but would always eventually return to the initial conversation if permitted. Return could be controlled based on user choice within the digression (“Do you want to go back to the previous topic?”), but this choice would need to be created and triggered for each potential dialogue with inbound digression. Consequently, this facility was not incorporated into the conversation design.

Alternate responses

Reitz et al. (2019) found that anthropomorphic features positively influence user perception of chatbots, especially in terms of perceived utility. When considering conversation design for digital political personae, user perception of the persona is an important consideration. Repetitious responses may heighten the feeling that a user is talking to low-value entity that has little to offer. While the artificial nature of the persona would not be hidden from the user, the ability to vary responses in a humanlike way is likely to improve user perception of the persona and promote engagement.

Great to hear! What would you like to talk about?	Thanks for your input! I'll store your response and use it to generate deeper conversations in the future.
That's good. What do you feel like talking about today?	Thanks for your point of view. I'll use your response to generate a better policy position on this issue.
Excellent! What do you want to talk about today?	That's an interesting perspective - thank you for sharing. I'll use it to refine my understanding of this issue.

Figure 14. Alternate responses for commonly used dialogue.

In situations where responses are public, such as on Twitter or other social media, variation in responses would be particularly important in preventing the persona from displaying large numbers of identical messages. Alternate responses were therefore developed and included for frequently encountered dialogues (see Figure 14). Alternate responses were considered for more complex dialogue, but variation of vocabulary and sentence structure in complex dialogue was deemed to have potentially unexpected impacts on tone and meaning, so this was not attempted.

4.2.2 Approaches for closing conversation

Many political exchanges observed in the dataset scraped from Twitter did not experience closure – questions were frequently unanswered, and statements were not addressed. Design iterations endeavoured to close any dialogue that did not lead to further conversation to improve user experience interacting with the persona.

Closing dialogue chains

Dialogue chains were closed when there were no other responses included in the chain. The persona was programmed to thank the user for the comments, perspective or question, effectively closing the chain by not inviting any more responses from the user on the topic.

While the goal was to terminate the dialogue chain, it was not necessary to also close the whole conversation. Asking “Was there anything else you wanted to talk about?” enables the overall conversation to be sustained, even after the user has closed a dialogue chain.

Extra-topic responses

The extra-topic responses developed during iteration enable more exchanges to close satisfactorily, in addition to supporting digressions and presenting the appearance of a more well-rounded persona. Extra topic responses enable the persona to more effectively close many of the types of dead-end exchanges observed during sampling of political conversations on Twitter, by responding to questions or opinions that might otherwise go unanswered.

The “digital psychiatrist” approach

Extra-topic responses can be relied upon to close many common exchanges, but the nearly limitless range of potential exchanges users may engage in makes pre-empting all of them with scripted dialogues impossible. The default approach employed by the Watson Assistant software when handling unknown utterances is to respond that the input is not recognised, and attempt to steer the user towards a topic or dialogue where responses have been defined. This approach was primarily employed, with variations in this default response reducing the potential instances of repetition.

A secondary approach was also developed to see if a more meaningful conversation could be simulated without the digital persona understanding the topic. When encountering an unrecognised utterance, the persona would indicate that the topic was not known to it, but would then ask the user to explain and elaborate. The user response could be parsed for emotive language, allowing the persona

to recognise a strong or neutral response and reply accordingly. An example dialogue chain is shown in Figure 15.

The dialogue chain shown in Figure 15 includes variation in the persona responses, reducing the likelihood of dialogue repetition if the persona encountered multiple sequential unknown topics. Repeated testing with example dialogue found that this approach was of limited utility, because the structure of the predefined replies often did not match the variation possible in user utterances, leading to a discordant exchange that would be transparently artificial to the user.

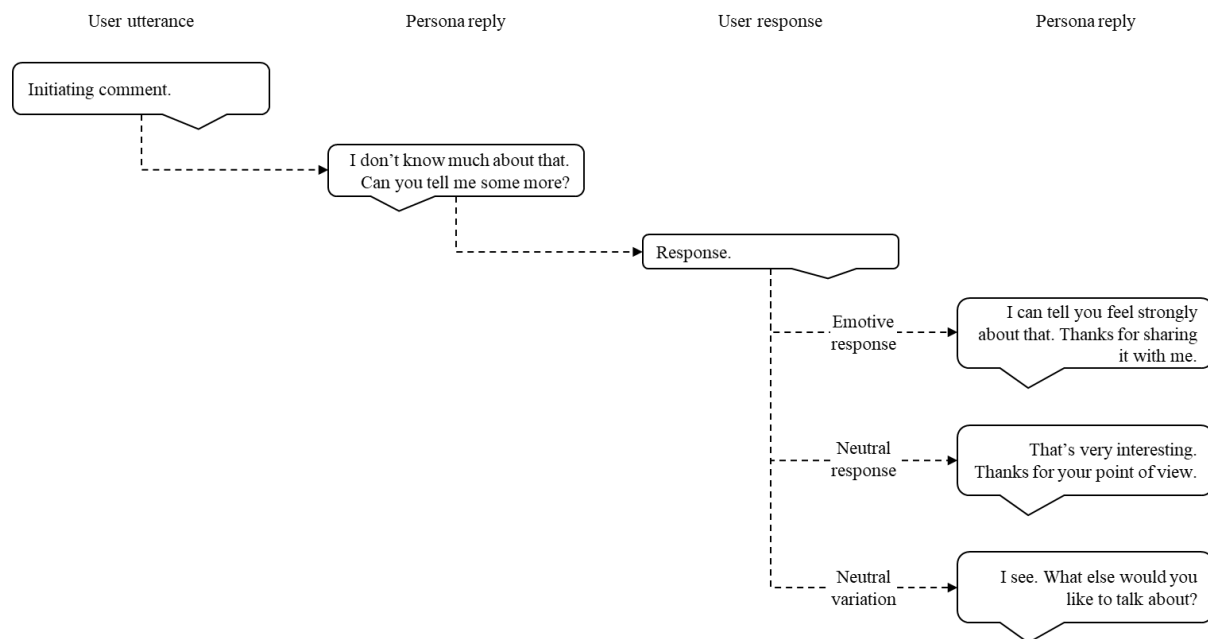


Figure 15. Alternate responses for commonly used dialogue.

4.2.3 Incorporation of triage, translation and layering

When developing a persona to sustain conversation, attention was paid to the cooperative principle of conversation as defined by the four Gricean maxims (Grice, 1975).

1. Quantity: provide as much information as is needed and no more
2. Quality: provide only true and accurate information that is supported by evidence
3. Relation: remain relevant to the discussion
4. Manner: be clear, brief, and unambiguous

These maxims are aligned with the concepts of triage, translation and layering identified by Solivan and Farina (2013), and formed the basis of dialogue structuring.

Triage

Triage is related to the maxims of quantity and relation: limiting the information provided to the most relevant to the conversation. Early design iterations sought to link many different user

utterances on the same topic to a small number of responses, each of which contained enough information to at least partially address the initiating comment. The example shown in Figure 16 shows the effect of this approach: the initial question is not directly addressed, and the user is overwhelmed by a large amount of information that is unlikely to encourage sustained engagement.

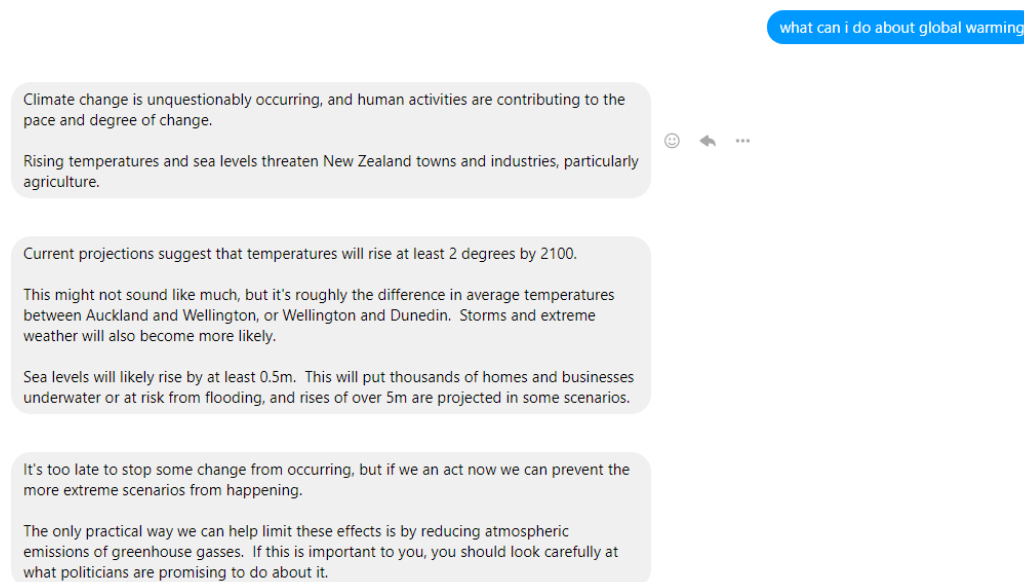


Figure 16. Example of a lengthy response that was later separated into multiple responses.

As previously discussed in section 4.2.1, multiple intents and dialogues were established to encourage sustain conversation and increase the relevance of each response to the initiating user utterance. Responses were restricted to 280 characters in most instances, to enable the resulting persona to be deployed to all common platforms including Twitter, which at the time of development maintains 280 character limit for each comment.

Chained comments as shown in Figure 16 enable this limit to be circumvented, but this form of exchange was rarely observed on Twitter, and is often frowned upon as a form of “conversational assault” on the user, known colloquially as a “Tweetstorm” (Merriam-Webster, 2019). Accordingly, this was avoided wherever possible, and only employed where all relevant information could not reasonably be contained within 280 characters.

Translation

Translation as defined by Solivan and Farina (2013) involves simplifying information to make it as accessible and understandable as possible, relating to the Gricean maxim of manner – being clear, brief and unambiguous.

Constraining responses to 280 characters had an unintended side-effect: the language of persona responses became increasingly complex in order to convey more information in a smaller number of words. A web-based tool was used to analyse persona responses against the Common European Framework of Reference for Languages (CEFR) (University of Cambridge Local Examinations Syndicate, 2020).

Analysis found that the CEFR level of the prototype persona responses was generally in the highest category, which had the potential to limit the accessibility of the conversation to users with lower language proficiency. A loose analysis of a selection of the tweets sampled in the analysis of existing digital political personae found that they generally exhibited a middling CEFR score, though there were marked differences in the CEFR score between personae in the sample of tweets examined.

Attempts were made to limit the use of complex language in order to reduce the CEFR level of the dialogue, though this was only partially successful. Ultimately, it was found to be difficult to separate the communication style of the conversation designer from the persona itself. Simple language may or may not be desirable – as discovered in the examination of existing tweets, some digital political personae consistently employ more complex language than others. We chose to accept a level of increased complexity in the dialogue as a feature of the persona, while still aiming to reduce complexity as far as possible.

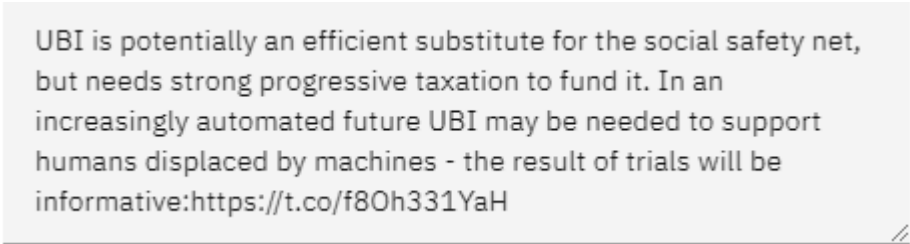
Layering

Layering makes more complex information available beyond the initial dialogue. Layering is essential to fully meet the Gricean maxim of quality, because self-imposed character limits make providing comprehensive information and supporting evidence challenging. Several mechanisms were examined that enable layering within digital conversation.

The conversation chains developed to sustain conversation present an opportunity to layer information at a high level: an initial response to a topic can lead to multiple follow-up dialogues that address other aspects of the subject. However, if further responses are limited to the same 280 characters this limits the amount of layered information that can be conveyed. While multiple responses can be presented together to provide more information, this approach can break the flow of the conversation, as seen when considering triage of information in Figure 16. As a result, this approach was not considered effective.

As seen in the examination of media tweets in section 3.1.4, images and video are frequently embedded into responses by existing digital political personae to provide additional information. Watson Assistant supports embedding of images and video into responses, as do most common social media platforms that might be linked to the chatbot persona. During conversation design iteration, images and video were embedded into responses where appropriate to the conversation topic.

Inclusion of hyperlinks to web pages and articles represents another common form of layering. Unlike embedded images and video, hyperlinks may count against character limits on platforms like Twitter. Twitter uses a link shortening service to automatically reduce the length of hyperlinks (Twitter Inc., 2020b); this supports the creation of dialogue that includes hyperlinks because it provides a consistent mechanism for reducing the length of links. This was the most commonly used approach for layering information that was employed during conversation design iteration, because it allowed for direct linking to detailed sources for information discussed in dialogue (see Figure 17).



UBI is potentially an efficient substitute for the social safety net, but needs strong progressive taxation to fund it. In an increasingly automated future UBI may be needed to support humans displaced by machines - the result of trials will be informative:<https://t.co/f80h331YaH>

Figure 17. Dialogue including a hyperlink providing layered access to information

4.2.4 Modifying tone and language to suit audience

Structuring the conversation design to respond individual characteristics of the user was the most challenging concept investigated. To accomplish this, the user was asked their political leaning during the initial exchange with the persona, and was given three options to choose from: “lean conservative,” “lean liberal,” and “somewhere in the middle.” This information was captured and stored in an entity during in an entity during this exchange.

When the user triggered subsequent dialogues with the persona, the reply could be set to vary based on the user’s self-identified political leaning. This effectively enables a tailored response to be provided to different types of users, provided they can be identified. Tailored responses were limited to the topic of global warming. Users who identified as “lean conservative” received responses that described the impact on business, while users who identified as “lean liberal” received responses that described the impact on the environment.

The utility of this capability warrants more detailed investigation. A digital political persona that panders to its audience may not be desirable, but the ability to frame information in language that is accessible to the audience certainly is. Feinberg & Willer (2015) found that arguments are more effective in political debate when they are reframed to match the moral framework of the audience. While the variations employed in this design research were crude proxies that reflect the personal biases of the researcher, investigation into framing responses to generate desired outcomes from target populations presents an avenue to use these design outputs to reach, engage and inform people with differing political perspectives.

A user’s political leaning was not the only information that was captured: the chatbot software also has the capability to ask and store a user’s name, allowing the persona to address them personally in later dialogues. However, many dialogues already employed the maximum number of characters implemented in earlier iterations, limiting the number of dialogues where the user’s name could be employed without breaking dialogue into multiple responses.

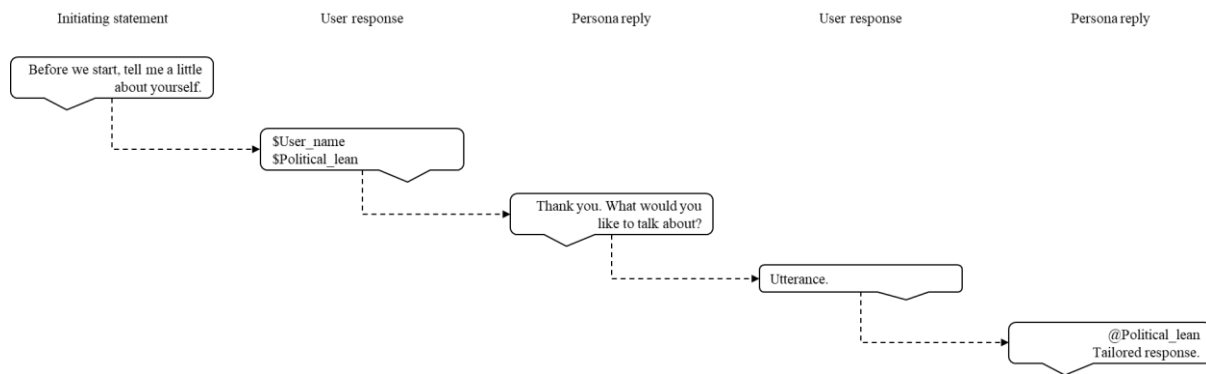


Figure 18. A dialogue chain that generates a tailored response to a user utterance based on information collected about the user during an initial exchange.

4.2.5 Final design output

The final design output builds on the conceptual framework described by Telang et al. (2018) to specifically address the design criteria identified for conversation design with digital political personae. A simplified representation of the final conversation design is shown in Figure 19. An example conversation is presented in Appendix A. The conversation design employs a base of short responses to multiple types of exchange to support a smaller number of rich dialogue chains that allow for longer conversation on one or more specific topics.

Dialogues both embed and employ links to external media and websites, providing access to more detailed information while maintaining conversation flow. Digression between dialogue chains and shorter responses allows discussion of the main topic to pause and resume naturally. An initiating exchange gathers limited information about the user's political views, enabling subsequent responses to be tailored to address these views. Responses may also be tailored based on the use of emotive language by the user, enabling the chatbot persona to modify conversational tone in response to user sentiment. When a given exchange is complete conversation terminates naturally: a user's final utterance is not left unanswered.

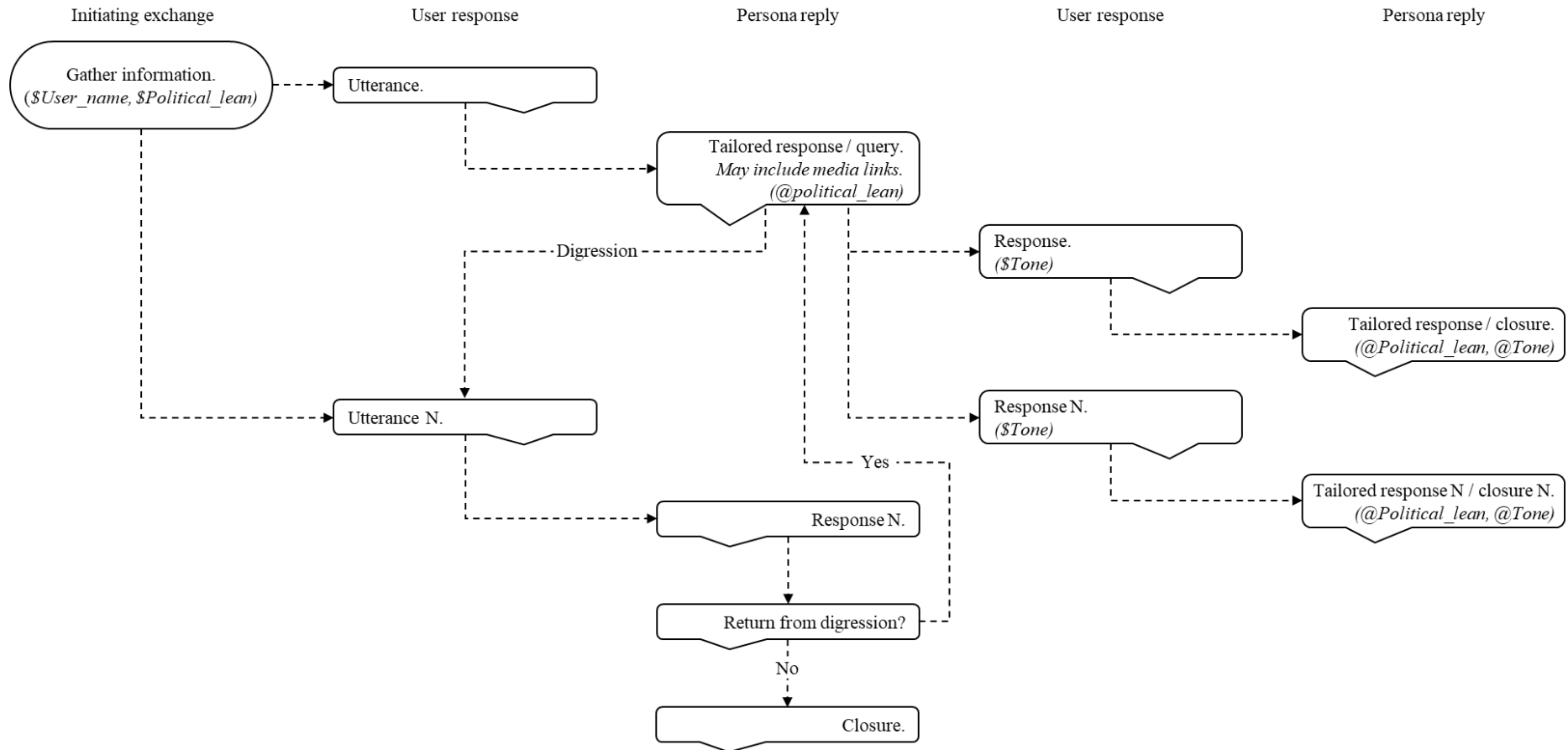


Figure 19. Conversation design for a chatbot-based digital political persona. Variables prefaced by \$ symbols indicate information collection; those prefaced by @ symbols indicate where previously stored information may be used to vary responses. Conversation chains may be longer than those shown, and may employ some, all or none of the indicated variables depending on the topic.

5. Discussion

5.1 Limitations

5.1.1 Social media analytics

The analysis of existing digital political personae was affected by limitations placed on the APIs used to sample tweets (Twitter Inc., 2020a). Twitter only supports the return of 3,200 of the most recent tweets from a user's timeline, which limits the bulk collection of tweets for analysis from a single account. Datasets could be scraped from multiple accounts to yield a large overall dataset, but this restriction effectively prevents bulk sampling of historical tweets.

Limitations to bulk sampling were rendered irrelevant by further limitations uncovered in the use of bulk analysis tools on the collected dataset. The complexity and nuance inherent in political exchanges caused automated tools to struggle to accurately parse and categorise tweets. Sentiment analysis tools in particular were found to consistently misclassify tweets, leading to the decision to manually analyse and classify a smaller number of tweets in order to inform the development of conversation design principles.

Manual examination of tweets uncovered further limitations in analysing conversations on social media. Scraping large numbers of tweets from a single user account does not accurately capture longer dialogue chains. While replies to specific tweets can be scraped by making adjustments to common API calls, capturing longer dialogue chains in bulk was not achieved during this research. While the subset of tweets that was manually coded was examined to identify longer dialogue chains, we cannot reliably comment on their prevalence in the wider dataset.

A further limitation of examining open social media fora like Twitter to inform conversation design is that only public-facing comments can be scraped and analysed. Conversations that take place via private or direct message between participants are not available to the researcher. Private exchanges may have a very different character to public exchanges, but there are obvious privacy implications around gaining access to such data.

These limitations are common to anyone using social media to do research into digital conversation design. Ultimately, ownership or control of the channel would be required to enable access to complete datasets.

5.1.2 Chatbot software

Design iteration was constrained by the capabilities of publicly available chatbot software. Of the software considered for this research, the most common issues encountered derived from the limited interface available to the user. The chatbot platforms examined were found to be targeted at users with minimal coding experience, and so relied heavily on user-friendly graphic user interfaces (GUIs).

However, allowing users to create intents and dialogues “with the touch of a button” often precludes development of more complex features or bulk actions. To create multiple intents and

dialogues requires significant manual effort if done directly through the interface of most chatbot platforms examined. IBM Watson Assistant has the capability to download, edit and upload intent, entity and dialogue files, which allowed for bulk editing and creation using variable substitution, but several alternative chatbot platforms explored, including the initial prototyping platform, did not have this capability.

Linking dialogues into chains was particularly challenging, requiring manual creation of parent and child dialogues and linking rules at each step. Creating and editing dialogues in code rather than in the GUI was significantly faster, but the lack of an effective visual map of the conversation made it more challenging to keep track of parent and child dialogues in dialogue chains. No chatbot software examined featured an editor that showed both the full detail of the underlying code and a visual representation of the conversation at the same time. An example of the visual display of dialogue chains in IBM Watson Assistant is included in Appendix A.

The pattern developed as the final design output of this research provide a framework that enables these limitations to be circumvented. By substituting dialogues and intents into the pattern developed, large number of complex exchanges may be quickly built. Control of the chatbot provides access to user dialogues that enable ongoing sampling of all conversations encountered by the persona, providing a potential mechanism for gathering large datasets of both public and semi-private conversation with a customisable digital persona.

5.1.3 Conversation design

Although the chatbot software employed in this research has many features that enable human-like conversation, such as the ability to vary repeated responses and support digression from topics, its limits would still be apparent to most users engaged in a longer conversation. The natural language processing capabilities employed to determine user intent were found to be generally accurate, but confusion between similar intents on different topics was repeatedly observed during design iteration. ‘Training’ the chatbot software by linking confused intents to the desired dialogue would be an ongoing requirement when employing a chatbot-driven persona for public conversation.

Responses that employed generic dialogue to respond to unknown utterance also revealed the limits of conversation design using this system. Pre-defined response variations often did a poor job of responding appropriately to user dialogue, failing to differentiate reliably between types of statements and queries. Unknown dialogue about a city or country would typically require a different structure of response than dialogue about a person or concept, which was not possible with the approach employed. A potential avenue for improving the ability of the system to respond to unknown utterances would be to train the system to recognise types of object (such as a location, person or concept) and respond appropriately to the unknown object type.

The most significant conversation design challenges concern the injection of the implicit tone and biases of the researcher into chatbot responses. As noted in section 4.2.3, the language employed during conversation design was at a higher level than may be desirable for accessibility by a broad audience. Repeated passes to simplify language were partially successful, but long-term use of a chatbot system would likely require continuous adjustment of the language employed to better reflect the requirements of its user base.

The ability to detect and respond to emotive language is a potentially powerful tool to engage users or change their views (Tan et al., 2016). However, this is another area where the emotional perspective of the researcher may be expected to affect the dialogue employed by the chatbot persona. The tone, language and emotional range of the persona might together be considered its “personality.” The impact of user perception of this personality on the overall reliability of the chatbot represents a potential topic for future research.

The impact of researcher bias is particularly important when considering the approaches employed to tailor responses based on a user’s self-identified political leanings. While Feinberg & Willer (2015) found that re-framing information to match the moral framework of the recipient made it more likely that it would be accepted, such re-framing would be implicitly affected by the perspective of the conversation designer on those same moral frameworks. It may be challenging for the conversation designer to craft responses aligned with a moral framework that they do not share.

Re-framing dialogue based on the political leanings of the user has potentially significant implications for a publicly deployed chatbot system because these variations would be swiftly discovered. Would a system that frames topics differently to conservatives and liberals be praised for savvy outreach, or criticised for insincere pandering? Given the rise in polarisation observed in some political climates (Boxell et al., 2017) it may be reasonable to assume the worst.

5.2 Applications of this research

The literature has much to say about the use of digital tools to both inform and drive political engagement, but thus far limited research has been devoted to the design of chatbot-driven digital personae to accomplish these activities. Investigation of existing conversation with digital political personae showed that the existing paradigm involves little genuine engagement between participants and does not make use of methodologies for robust political engagement identified in the literature.

General wariness to openly develop automated systems that mimic human conversation for political purposes may be driven by the prevalence and perception of “bad bots” in online fora. Automated accounts on Twitter have been estimated to share over 40% of political links to external websites (Wojcik et al., 2018), and are key vectors for the spread of misinformation. These bots act within the existing paradigm, using mass link sharing to drive users to external content (Shao et al., 2018). Using a conversational approach for a political chatbot-driven persona represents a novel approach for engaging users and sharing political information.

The conversation design developed through this research has the capacity to sustain longer and more complex conversations than are typically observed in interactions with existing digital personae. Despite the decision to impose character limits on response length, the resulting conversation was able to incorporate media and external links that enabled layering of more complex information. As a result, the persona developed would be suitable for situations where conveying information on political topics requires application of the concepts of triage, translation and layering identified by Solivan and Farina (2013). The branching nature of the conversation, coupled with tightly constrained character limits in responses, further serves to prevent information overload for participants.

The platform-agnostic design enables the chatbot persona to be deployed to multiple digital platforms, including direct messaging applications, social media, and other websites. This prevents limited accessibility to the chatbot acting as a barrier to participation in political conversation, identified as a challenge for digital political systems in the literature (Collin, 2015). A further advantage of utilising existing platforms is that their use is familiar – understanding of new systems and rules are not required for citizens to engage with the persona. The automated nature of the system limits human involvement in maintaining conversations, potentially making it an effective tool for countering misinformation-spreading bots by disseminating reliable information to large numbers of people.

The conversation design developed through this research specifically addresses the challenges evident in engaging in informed political conversation in digital spaces. Because the good practice approaches employed in this research are rooted in examination of the literature on tools for e-governance, the conversation design developed may form a useful tool for informing and engaging citizens where a conversational approach is required. Future research could use the conversation approach developed here to test the applicability of these approaches in different situations and with different types of participant.

The chatbot platform employed records users exchanges with the persona, allowing for large datasets of user interactions on political topics to be collected. This capability enables a wide range of future research opportunities, allowing for investigation of the political opinions of targeted demographics on specific topics. Utilisation of targeted chatbot-driven conversation has significant advantages over conventional approaches that sample social media, as utilised in this paper, because these approaches do not capture direct or private messages, and are limited by the capabilities of platform APIs in terms of the type and number of exchanges they can sample. With direct control over the channel, all responses can potentially be captured. A potentially valuable area of future research is the investigation of variation between “public” and “private” opinions in digital political spaces.

This research did not explore the intersection between other factors that may affect user perception of the “brand personality” of the developed persona. These factors may include colour, branding, interface, and the platform that houses the persona. Modification of these factors is likely to influence user perception of the persona and may therefore influence conversation and behaviour (Aaker, 1997). Further investigation of these factors may inform their effect on user behaviour and allow this to be incorporated into the overall persona design.

The utility of the conversation design developed here may not be limited to the political arena – future research might examine employing the approaches developed here in other areas, such as promoting or soliciting feedback on products and services.

6. Conclusion

Effective political discourse employs approaches that engage and inform participants, while making political topics accessible at varied levels of understanding. Presentation of clear, relevant information coupled with the ability to explore layered information in more detail can inform participants without overwhelming them. Responding to and utilising emotive language can translate topics into accessible language in polarised environments. Employing these approaches may serve to insulate participants against polarising rhetoric and improve participation in the democratic process. Despite studies showing the effectiveness of these approaches, few tools have been developed that utilise them to positive effect.

Political discourse in online spaces exhibits many of the negative consequences of failing to take account of these principles. Despite making and receiving large volumes of messages, extant political personae exhibit little true engagement with participants, with many exchanges shallow, limited, or receiving no response at all.

Commercially available chatbot software can be used to create conversation-driven personae that incorporate good practice approaches for political discourse. The design derived through this research enables sustained, detailed conversation that flows naturally between topics and provides access to layered information through embedded hyperlinks and other media. Large volumes of conversation data may be collected through the application, giving it potential utility in surveying and polling participants as well as engaging and informing them.

It is difficult to avoid the contention that the issues identified in the literature and examination of political exchanges on Twitter have become increasingly apparent during the period that this research was performed. The links between political polarisation in online and offline spaces seem no less strong today than they do than when this research was begun, amid continued erosion of political discourse on national and international stages. The grounding of this research in approaches for meaningful dialogue means that the conversation-driven system described here may offer a potential mechanism for addressing these challenges.

The design outputs of this research address the limitations to engagement with digital political personae observed in existing online fora. While the resultant conversation design is geared towards this use case, it has not escaped the researcher's notice that this design pattern may have similar utility in other scenarios where engaging and informing users via a conversational interface is desirable.

7. References

- Aaker, J. (1997). Dimensions of Brand Personality. *Journal of Marketing Research*, 34(3), 347–356. <https://doi.org/10.2307/3151897>
- Bakliwal, A., Foster, J., Puil, J.V., OBrien, R., Tounsi, L., & Hughes, M. (2013). *Sentiment Analysis of Political Tweets: Towards an Accurate Classifier*.
- Blank, G., & Reisdorf, B. (2012). The Participatory Web: A user perspective on Web 2.0. *Information, Communication & Society*, 15(4), 537–554. <https://doi.org/10.1080/1369118X.2012.665935>
- Boxell, L., Gentzkow, M., & Shapiro, J. (2017). Greater Internet use is not associated with faster growth in political polarization among US demographic groups. *Proceedings of the National Academy of Sciences of the United States of America*, 114(40), 10612-10617. doi:<https://doi.org/helicon.vuw.ac.nz/10.1073/pnas.1706588114>
- Braun, V., & Clarke, V. (2012). *Thematic analysis*. (pp. 57-71). American Psychological Association.
- Buiten, M. (2019). Towards Intelligent Regulation of Artificial Intelligence. *European Journal of Risk Regulation : EJRR*, 10(1), 41–59. <https://doi.org/10.1017/err.2019.8>
- Chester, J., & Montgomery, K. (2017). The role of digital marketing in political campaigns. *Internet Policy Review*, 6(4).
- Collin P. (2015) Young citizens and political participation in a digital society: Addressing the democratic disconnect. Palgrave Macmillan, London
doi:https://doi.org/10.1057/9781137348838_6
- Cresswell, J. (2009) *Research design: Qualitative, Quantitative and Mixed Methods Approaches* (3rd ed). London, Sage Publications.
- Dahlberg, L. (2011). Re-constructing digital democracy: An outline of four 'positions'. *New Media and Society*, 13(6), 855-872.
- Epstein, D., & Leshed, G. (2016). The magic sauce: Practices of facilitation in online policy deliberation. *Journal of Public Deliberation*, 12(1) Retrieved from <http://www.publicdeliberation.net/jpd/vol12/iss1/art4>
- Følstad, A., & Brandtzæg, P. (2017). Chatbots and the new world of HCI. *Interactions*, 24(4), 38–42. <https://doi.org/10.1145/3085558>
- Farina, C. R., Epstein, D., Heidt, J., & Newhart, M. J. (2014). Designing an online civic engagement platform: Balancing "more" vs. "better" participation in complex public policymaking. *International Journal of E-Politics*, 5(1), 16-40.
- Feinberg, M., & Willer, R. (2015). From Gulf to Bridge. *Personality and Social Psychology Bulletin*, 41(12), 1665-1681. doi: 10.1177/0146167215607842 Frankel, L., & Racine, M. (2010, July).

- The complex field of research: For design, through design, and about design. In *Proceedings of the Design Research Society (DRS) International Conference* (No. 043).
- Gastil, J., & Richards, R. (2017). Embracing Digital Democracy: A Call for Building an Online Civic Commons. *50*(3), 758-763. doi:10.1017/S1049096517000555
- Goyal, R., & Kakkar, M. (2019). Sentiment Analysis of Twitter Data Using Statistical Analysing Tool R Studio. *Indian Journal of Computer Science*, 4(5), 29-35.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics, Vol. 3: Speech Acts* (pp. 4158). New York: Academic Press.
- Hansen, L., Arvidsson, A., Nielsen, F., Colleoni, E., & Etter, M. (2011). Good Friends, Bad News - Affect and Virality in Twitter. arXiv.org. Retrieved from <http://search.proquest.com/docview/2081359714/>
- Honey, C., & Herring, S. (2009). Beyond Microblogging: Conversation and Collaboration via Twitter. 2009 42nd Hawaii International Conference on System Sciences, 1–10. <https://doi.org/10.1109/HICSS.2009.89>
- Hong, S., & Kim, S. H. (2016). (2016). Political polarization on Twitter: Implications for the use of social media in digital governments. *Government Information Quarterly*, 33(4), 777-782. doi: <https://doi.org/10.1016/j.giq.2016.04.007>
- IBM. (2016). Sample Architecture Diagram for Integrations. Retrieved from: <https://medium.com/ibm-watson/integrate-watson-assistant-with-just-about-anything-695bc1d29875>
- Janowski, T. (2015). Digital government evolution: From transformation to contextualization. *Government Information Quarterly*, 32(3). doi:<https://doi.org/10.1016/j.giq.2015.07.001>
- Langelaar, W., Gerritsen, N. And Smith, A. (2017) *Sam: Virtual Politician*. Retrieved from <http://business.scoop.co.nz/2017/11/15/meet-sam-the-virtual-politician-for-the-digital-age/>
- Leavy, P., & Prior, L. (2014). Content Analysis. In *The Oxford Handbook of Qualitative Research*. Oxford University Press.
- Littlejohn, S., Foss, K., & Oetzel, J. (2017). Theories of Human Communication (Eleventh edition.). Long Grove, Illinois: Waveland Press, Inc.
- López, G., Quesada, L., Guerrero, L.A.: Alexa vs. Siri vs. Cortana vs. Google assistant: a comparison of speech-based natural user interfaces. In: Nunes, I. (ed.) *AHFE 2017. AISC, vol. 592* (pp. 241–250). Springer, Cham (2018). https://doi.org/10.1007/978-3-319-60366-7_23
- Lukensmeyer, C. J. (2017). Civic tech and public policy decision making. *PS, Political Science & Politics*, 50(3), 764-771. doi:<http://dx.doi.org/helicon.vuw.ac.nz/10.1017/S1049096517000567>
- Mattka, R. (2019). Artificial Intelligence for the Web. *Tech & Learning*, 39(6), 28-33.

- Mendes, A. E. (2016). Digital Demagogue: The Critical Candidacy of Donald J. Trump. *Journal of Contemporary Rhetoric*, 6(3/4), 62-73.
- Meneses, M. E., Nonnecke, B., del Campo, A. M., Krishnan, S., Patel, J., Kim, M., Crittenden, C., & Goldberg, K. (2017). Overcoming citizen mistrust and enhancing democratic practices: Results from the e-participation platform México Participa. *Information Technologies & International Development*, 13, 138-154.
- Merriam-Webster. (2019). What's a 'Tweetstorm'? Retrieved from: <https://www.merriam-webster.com/words-at-play/tweeting-up-a-tweetstorm>
- Mishra, S., Khanna, P., Kumar, S., & Sinha, A. (2017). Sentiment analysis: an approach to opinion mining from Twitter data using R. *International Journal of Advanced Research in Computer Science*, 8(8). Retrieved from <http://search.proquest.com/docview/1953784071/>
- Morris, D. S., & Morris, J. S. (2013). Digital inequality and participation in the political process: Real or imagined? *Social Science Computer Review*, 31(5), 589-600.
- Neudert, L. (2018). Future elections may be swayed by intelligent, weaponized chatbots. *MIT Technology Review*. Retrieved from <https://www.technologyreview.com/s/611832/future-elections-may-be-swayed-by-intelligent-weaponized-chatbots/>
- Otter, D. W., Medina, J. R., & Kalita, J. K. (2018). *A Survey of the Usages of Deep Learning in Natural Language Processing*.
- Prakash, A. (2018, August 9). AI-Politicians: A Revolution in Politics. Retrieved from: <https://medium.com/politics-ai/ai-politicians-a-revolution-in-politics-11a7e4ce90b0>
- Rietz, T., Benke, I., & Maedche, A. (2019). The Impact of Anthropomorphic and Functional Chatbot Design Features in Enterprise Collaboration Systems on User Acceptance, presented at 14th International Conference on Wirtschaftsinformatik, 2019, At Seigen.
- Remnick, D. (2016). Obama reckons with a Trump presidency. *The New Yorker*, 28. Retrieved from <https://www.newyorker.com/magazine/2016/11/28/obama-reckons-with-a-trump-presidency>
- Russell, S., & Norvig, P. (2010). *Artificial intelligence : a modern approach* (3rd ed.). Upper Saddle River, N.J: Prentice Hall.
- Sawer, M. (2007). Wearing your Politics on your Sleeve: The Role of Political Colours in Social Movements. *Social Movement Studies*, 6(1), 39–56. <https://doi.org/10.1080/14742830701251294>
- Schegloff, E. A., & Sacks, H. (1973). Opening up closings. *Semiotica*, 8(4), 289-327.
- Shao, C., Ciampaglia, G., Varol, O., Yang, K., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. *arXiv.org*, 9(1), 1–9. <https://doi.org/10.1038/s41467-018-06930-7>

- Solivan, J., & Farina, C. R. (2013). Regulation Room: How the Internet Improves Public Participation in Rulemaking. *Cornell e-Rulemaking Initiative Publications*. Paper 13.
- Susskind, J. (2018). Chatbots Are a Danger to Democracy. *The New York Times*. Retrieved from <https://www.nytimes.com/2018/12/04/opinion/chatbots-ai-democracy-free-speech.html>
- Tan, C., Nicolae, V., Danescu-Niculescu-Mizil, C., & Lee, L. (2016). Winning Arguments: Interaction Dynamics and Persuasion Strategies in Good-faith Online Discussions. In J. Bourdeau, J. A. Hendeler, R. N. Nkambou, I. Horrocks, & B. Y. Zhao (Eds.), *Proceedings of the 25th International Conference on World Wide Web* (pp. 613-624). Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee. doi: 10.1145/2872427.2883081
- Telang, P., Kalia, A., Vukovic, M., Pandita, R., & Singh, M. (2018). A Conceptual Framework for Engineering Chatbots. *IEEE Internet Computing*, 22(6), 54–59. <https://doi.org/10.1109/MIC.2018.2877827>
- The Economist Intelligence Unit (2017). *Democracy Index 2017*. Retrieved from http://pages.eiu.com/rs/753-RIQ-438/images/Democracy_Index_2017.pdf
- Twitter Inc. (2020a). Get Tweet timelines. Retrieved from: https://developer.twitter.com/en/docs/tweets/timelines/api-reference/get-statuses-user_timeline
- Twitter Inc. (2020b). About Twitter's link service (<http://t.co>). Retrieved from: <https://help.twitter.com/en/using-twitter/url-shortener>
- University of Cambridge Local Examinations Syndicate. 2020. International Language Standards. Retrieved from: <https://www.cambridgeenglish.org/exams-and-tests/cefr/>
- Westfall, J., Van Boven, L., Chambers, J. R., & Judd, C. M. (2015). Perceiving political polarization in the United States: Party identity strength and attitude extremity exacerbate the perceived partisan divide. *Perspectives on Psychological Science*, 10(2), 145-158.
- Wojcik, S., Messing, S., Smith, A., Raine, L., & Hitlin, P. (2008). Bots in the Twittersphere. *Pew Research Center*.

Appendix A

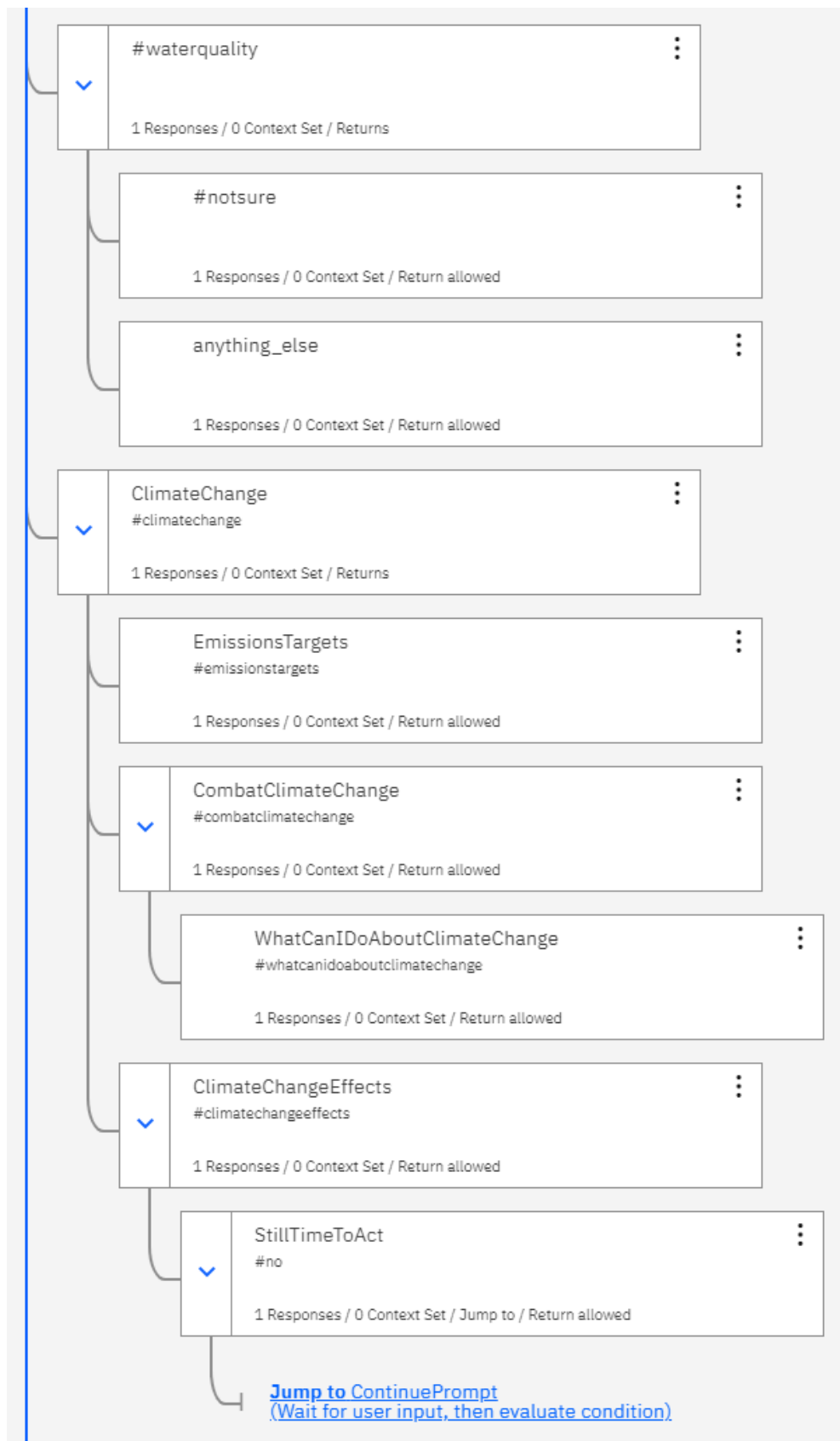


Figure 20. Dialogue chains as show in the IBM Watson Assistant interface.

Hello. I'm an automated chatbot designed to talk about political topics. What would you like to talk about today?

Is climate change real? What happened to global warming?

#climatechange

Climate change is happening now, and human activities are contributing to the pace and degree of change. Rising temperatures and sea levels threaten towns and industries, particularly agriculture, and will have significant effects all over the world. See: <https://climate.nasa.gov/effects/>

Doesn't the climate change all the time? What can we do about it?

#combatclimatechange

Research suggests that the only way to limit the effects of climate change is to reduce our use of fossil fuels. We should reduce our consumption of coal and oil products and shift to renewable sources of energy generation including wind and solar energy. What do you think?

Shouldn't we be more concerned about pollution?

#waterquality

Maintaining the health of lakes and rivers is important for several industries, and ensures that our environment is preserved for future generations. What is your biggest concern around water quality?

I was talking about pollution in general, but I guess water is part of that. Everything ends up in the sea and we end up eating it.

#waterquality

That's an interesting perspective - thank you for sharing. I'll use it to refine my understanding of this issue.

So we were talking about climate change...

Research suggests that the only way to limit the effects of climate change is to reduce our use of fossil fuels. We should reduce our consumption of coal and oil products and shift to renewable sources of energy generation including wind and solar energy. What do you think?

Wind and solar is fine I guess, but I don't see what I can do about climate change.

#whatcanidoaboutclimatechange

It can be hard to feel like individual action is meaningful, but if we use our cars less, properly insulate our homes, consume fewer disposable products, and purchase sustainably generated electricity (or use our own solar panels), collectively we can make a difference.

What happens if we don't do anything?

#climatechangeeffects

Even with preventative action temperatures will likely rise 2C by 2100, and sea levels by at least 0.5m. Many homes and businesses will be underwater or at risk from flooding, and agriculture and water security will be threatened. Extreme weather events will also be more likely.

That sounds bad.

#bad

I know it sounds bad, but there is still time to act. We will need local, national and international cooperation, but we can mitigate the worst of the effects. You can learn more at: <https://www.nature.com/articles/d41586-019-02734-x>

Do you want to talk about anything else?

No, I'm good.

#no

Ok! It was nice talking with you.

Figure 21. A simulated user conversation involving the dialogue chains in Figure 20. This conversation includes a digression into the “water quality” sub-topic, which was the determined to be the closest match to a user utterance about pollution by the natural language processing capability of IBM Watson Assistant. Layering of information is accomplished with hyperlinks. These links are automatically shortened if the dialogue is posted on the Twitter platform. User responses are in white; detected intents are indicated by #. The chatbot can be “trained” if incorrect intents are detected by selecting the correct intent using the visible dropdown arrow to the right of each intent.