Persuasive Content Generator

The design, development and validation of persuasive content generator based on social media profiles

By

Amirsam Khataei (MASc)

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Ph.D.)

in

Electrical and Computer Engineering

Carleton University
Ottawa, Ontario

January 2019

ABSTRACT

Providing content and experiences to users in order to influence their thoughts and opinions is essential in many fields such as marketing, education, and politics. With advances and the growing availability of the Internet and Web technologies, posting documents online has become an effective means of communicating to large audiences. However, given individuals' different respective interests, characteristics, and abilities, shared documents will not likely be equally persuasive to all users. To communicate persuasively, the categorization of users according to factors like age, marital status, education, and occupation is becoming increasingly prevalent, as is providing users with content specific to their categories. This approach assumes individuals within a category are similar, which is not necessarily true. Persuasive communication should recognize differences at the individual level and personalize—rather than categorize—the content. However, a systematic software solution that tailors and prepares separate persuasive content for individuals does not yet exist. Given the increasing popularity and usage of social media, social-network extractable data can potentially provide a tremendous source of insight and background about individuals. Inspired by the Yale Attitude Change approach, this thesis proposes a multi-layer model called Pyramid of Individualization and the related software framework to generate persuasive content based on initial author input and audience social-media data. Preliminary results show the proposed system can create personalized information that (a) matches reader interests (attention), (b) is tailored to reader ability to understand the information (comprehension), and (c) is supported by a trustable source (acceptance).

ACKNOWLEDGMENT

First and foremost, I would like to express my sincere appreciation to my supervisors Dr. Ali Arya and Dr. Michael Hine without whom this study would not have been possible. Their guidance, patience, and encouragements made all the difference.

I deeply would love to thank my dear wife, Yasamin Jan, for her unconditional love, understanding and support that she has had for me. I was amazed by her willingness to proof read countless pages of technical documents. Special thanks to my Ameh Manijeh and Amoo Farhad for all their incredible support since the moment I moved to Canada and started my studies. Thanks to my bigger brother for all the help and for always being on my side and never leaving me alone.

Finally, I would like to dedicate this dissertation to Maman and Baba. I am extremely grateful to them for their love, prayers, caring, and encouragement.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1.

INTRODUCTION

1.1 BACKGROUND AND MOTIVATION

Aesop, the ancient Greek fabulist (620–564 BCE) said, "Persuasion is often more effective than force" [1]. Persuasion has been interpreted as "human communication designed to influence the autonomous judgment and actions of others" [2]. *The American Heritage Dictionary* defines it as the act of causing "[someone] to do something by means of argument, reasoning, or entreaty; to win over [someone] to a course of action by reasoning or inducement; to make [someone] believe something; convince" [3]. Persuasion is an essential part of education, marketing, sales, and politics when people should be convinced that a certain product, skill, or approach is suitable for them. Persuasive content is meant to influence opinions, beliefs, attitudes, or behaviours of intended audiences. Given the development of the Internet and Web technologies and their growing availability and popularity, generating a digital document and sharing it with others has become an efficient means of communicating with large audiences. However, since individuals possess different respective interests, characteristics, and abilities, these shared documents will not necessarily be equally persuasive to all users.

To provide persuasive communication, users are often categorized according to characteristics like age, marital status, education, and occupation and provided with content specific to their characteristics. For example, television/online advertisements will target a program/website's viewers, or financial institutions provide different promotional material specifically targeting different demographics based on their perceived common needs. Although

this approach can improve the persuasiveness of the content, it assumes individuals within a category are similar, which is not necessarily true. Individuals may have different points of view on a topic because of respective personal experiences, epistemic beliefs [4][5], or even dispositions toward knowledge [6]. For instance, a generic persuasive text targeted at students may not persuade them to act similarly in relation to a common issue since they may have different respective reading comprehension levels, or they may not all trust the author equally. Examples of more personalized content include customized online ads based on browsing history, and social-media newsfeeds based on existing connections and recent activities. Persuasion can focus on either behaviour or attitude which is a long-term pattern of behaviours based on opinions and beliefs. Previous research on persuasive software systems has focused mainly on changing behaviour leaving attitude an unexplored construct in systems research [7]. To our knowledge there is currently no systematic software solution for tailoring individualized persuasive content to influence attitude.

Persuasive Content Generation (PCG) is the process of algorithmically generating and controlling persuasive content. Although much has been said about motivating users to perform a target behaviour or to meet a target goal [7]–[9], research into influence on user attitudes and automated tools for creating persuasive content is far less common. Difficulties in designing such tools include the lack of personal data and associated computational models capable of manipulating content to facilitate persuasion. Given the increasing popularity and use of social media, extractable data from users' social network accounts constitutes a tremendous source of personal data about users. This thesis presents the Pyramid of Individualization (PoI) model and its related software framework for generating personalized persuasive content based on initial author input and target audiences' social media information. PoI is a multilayer model inspired by

the Yale Attitude Change (YAC) approach that provides a mechanism for building user profiles which is used for persuasive applications.

1.2 PROBLEM STATEMENT

Nearly twenty years ago, David Weinberger, co-author of the *Cluetrain Manifesto* [10], defined *personalization* as "the automatic tailoring of sites and messages to the individuals viewing so we can feel that somewhere there is a piece of software that loves us for who we are" [11]. This concept has been widely interpreted as the process of preparing content for a specific person based on stated or implied preference [12][13][14] and has resulted in the breaking of groups of people into microsegments [15]. Persuasion is more than segmentation, however. For example, a persuasive system should not try to convince "single mothers" to choose a product; it should persuade "Mary" or "Jasmin" that its product meets their specific needs. Thus, segmentation—based on personal characteristics like interest or a readability score—becomes only the first step in generating persuasive content.

Although a presentation or document should target an audience and address their needs, an author cannot generate content that fits all target audience's member personalities and backgrounds. This thesis addresses the problem of generating personalized persuasive content by designing, implementing, and evaluating:

- A theoretical model that defines user profiles suitable for generating persuasive content
- The related software framework that enables automatic generation of persuasive content based on user profile and original author's objectives

In order to help solve this problem, this thesis will answer five particular research questions relating to the automatic generation of persuasive text:

1. What parameters should be considered in a personalized model?
2. How can we build a personalized model?
3. Can such a personalized model improve persuasion?
4. What software framework can create the persuasive content based on personalized model and original author's objective?
5. What are the essential processes required by this model and software framework?

While persuasive content can include more than just text, the main focus of the thesis is textual content since unstructured text is the largest source of data and information [16]. Future research can include multimedia data which will add more complexity and potentially strength to the personalization.

1.3 PROPOSED SYSTEM

To address the problem of the lack of personal data and related computational models in generating persuasive content, we propose the multilayer model called Pyramid of Individualization (PoI) and the corresponding Persuasive Content Generation (PCG) software framework. PoI defines the personalization process as a series of layers each narrowing down the population until we reach an individual. This is done by segmentation (user categories), comprehension (personal readability level), and individualization (adding personal characteristics, and in particular, trusted evidence). The corresponding software framework consists of a pipeline that processes the author and end user information as input and generates the persuasive content. This is done by using information from the end users' social network accounts and, through a series of author (persuader)–provided rules [17].

In designing the PoI model, we relied on the four persuasive factors defined by Yale Attitude Change (YAC) approach [18]: attention, comprehension, acceptance, and retention. These factors

directly influenced our PoI model and are reflected in different aspects of the software-framework process: segmentation specifies content that can draw the user's attention, comprehension makes the content easily readable, and individualization adds trusted sources so that the content is more acceptable. Together, these layers will create a personalized and persuasive content that we believe will be more memorable (so addressing the retention factor in YAC). While there are many other personal characteristics that can be added in the Individualization layer, in this research we have focused on providing evidence and supporting arguments from trusted sources. This is due to the direct role of this factor in decision making and the importance of trust in acceptance [19] [18].

1.4 RESEARCH APPROACH

Our research approach included developing a model, prototyping the PCG software, and using a mixed method analysis for evaluation. Our software system design followed a reusable component-based approach (creating, using, and modifying reusable components) and a web-based structure to allow maximum flexibility in software design and deployment. To evaluate the generated content's persuasiveness, we adopted the YAC factors (attention, comprehension, acceptance, and retention) and group them into personalized and persuasive criteria:

1. Readers perceive content generated by the proposed system as personalized: evaluation criteria include whether content generated by the proposed system (a) gains reader attention and (b) is comprehensible; and
2. Readers perceive content generated by the proposed system as persuasive: evaluation criteria include whether content generated by the proposed system (a) is acceptable and (b) retainable.

1.5 CONTRIBUTION

Our major contributions are:

- The PoI as a new model and the basis for creating persuasive content. The new model extends the YAC factors into the context of persuasive text by defining different layers of personalization.
- A pluggable software framework that integrates newly designed re-usable software components with existing ones. The components designed and developed through this research project are:
    - A rule-based component to create and control personalized content by the author (persuader).
    - User (persuadee) profile builder based on extracting information from social media data.
    - User trust graph and related algorithms for selecting trustworthy supportive information.

1.6 THESIS STRUCTURE

This document provides an overview of the PoI, our proposed persuasive personalization model, and reports on PCG components based on the model and the software framework. Chapter 2 reviews the related work in persuasive content and details the tools and methods we used. Chapter 3 includes an overview of the proposed system and highlights our PoI model and software framework. Chapter 4 contains a general overview of our implementation notes, and a report on PCG prototyping. Chapter 5 defines our research approach and evaluation methodology. Chapter

6 covers results and discusses the final experiment. Finally, Chapter 7 describes how this system can be optimized, limitations, and future work.

Chapter 2.

RELATED WORK

2.1 PERSUASION

Persuasion has been interpreted as "human communication designed to influence the autonomous judgment and actions of others" [2]. *The American Heritage Dictionary* defines it as the act of causing "[someone] to do something by means of argument, reasoning, or entreaty; to win over [someone] to a course of action by reasoning or inducement; to make [someone] believe something; convince" [3]. One of the early works in the field of persuasion is Yale Attitude Change (YAC) [18]. After WWII, Carl Hovland and his colleagues [19] initiated a study that would, through propaganda, boost US soldier morale. Their work resulted in a theoretical structure that, in late 1960s, William McGuire [18] developed into the YAC model. According to YAC [18], effective persuasion of an audience first requires gaining that audience's attention, adjusting the message's comprehension level to a level the intended user can consume, ensuring argument acceptance, and ensuring the message will be remembered (see Figure 2.1 - YAC overview

YAC has shaped marketing and advertising [20] strategies and is influential because of its simplicity. Most well-known theories of persuasion [9][21][22][23] are either inspired by YAC or share common elements with it. Furthermore, YAC Information Processing Model (its dissection of how a system persuades) is used in notable Persuasive Systems Design (PSD) Model [24].

This section summarizes four other well-known theories of persuasion. It explains our rationale behind choosing YAC as our primary source for the development of our own persuasive content model.

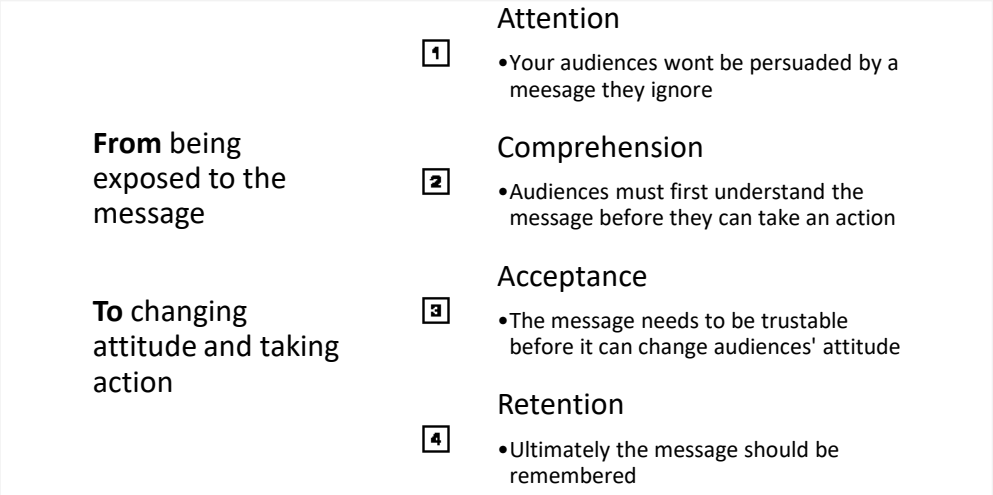


Figure 2.1 - YAC overview

In the 1990s, Icek Aizen studied persuasion by means of perceived behavioural control. His research resulted in the theory of planned behaviour (TPB), which links beliefs and behaviour [9], stating that one needs to know people's intentions to understand their volitional behaviour. An individual's intentions reflect their thinking outcomes about who they are and what their objectives are. According to TPB, three psychological variables (attitude, norms, and control) best predict intention:

- Attitude expresses what is good or bad about an action,
- Norms express which actions are approved, and which ones are not, and
- Control expresses what makes it easier or harder to do the action.

Unlike YAC, a major part of TPB focuses on using direct controls to affect behaviour, as opposed to influencing opinions and attitudes. Controls (factors that may facilitate or prevent

performance of the behavior) directly impact changing behaviour and do not rely on textual content to change attitude.

Lawrence Kohlberg's stages of moral development highlights people's justifications for their own behaviour [25]. Kohlberg defines three levels of moral development:

- At the preconventional level, authority is outside the individual and reasoning is based on the physical consequences of actions. In this level, the subject has not yet internalized societal moral conventions;
- At the conventional level, the individual internalizes moral standards from valued adult role models; and
- At the postconventional level, the individual bases their judgment on self-chosen beliefs and their moral reasoning on individual rights and justice.

Although Kohlberg's theory does not explain what triggers behaviour, the theory could be a good reference for analyzing post-behaviour evaluation. While YAC is a suitable persuasion theory to change one's attitude, Kohlberg's stages of moral development are suitable to be repurposed and then used to help foresee behaviour or predict decision-making processes.

The Elaboration Likelihood Model (ELM) describes the nature of changes in attitude using two different processes (central and peripheral) [21]. The central route includes thoughtful consideration of the message, while the peripheral route consists of decisions based on cues other than message content's strength, including source credibility. In this model, an author describes two separate routes leading to a change of attitude, but both processes may operate in parallel to persuade an individual. Both YAC and ELM highlight trust audience for a message as the main criteria in audience persuasion. YAC, however, is more complete than ELM since it also listed other criteria that can increase the likelihood of persuasion.

Fazio et al.'s model links attitudes and behaviour: it explains that people's attitudes guide their perception of behaviour [8], meaning:

- Explicit attitudes are measured at the conscious level and are deliberately formed and easy to self-report, and
- implicit attitudes are measured at an unconscious level and are involuntarily formed and are typically unknown to individuals.

Both attitudes can shape individuals' behaviours, but only explicit attitudes are based on cognitive processes. According to Fazio et al., behaviours resulting from implicit attitudes are likely to be caused when the individual feels stressed or distracted. As with the ELM, this model can be interpreted as a subsection of YAC. Data extracted using Fazio's theory needs to be reshaped if that data is to become a software input.

After completing the literature review, we decided not to strictly follow any of the above persuasion theories, though we use YAC's persuasive factors (i.e., attention, comprehension, acceptance, and retention) as a primary source for our persuasive content model.

2.2 PERSUASIVE TECHNOLOGIES

The study of persuasive technologies first emerged in the late 1990s [26]. Fogg defined *persuasive technologies* as technology designed to change attitudes or behaviors of users through persuasion and social influence but not through coercion [27]. How researchers individuate persuasive strategies varies. Cialdini developed six principles [28][29], Fogg described forty approaches under a general definition of persuasion [27], and others listed over 100 distinct tactics [30]. Scholarly writing on these technologies began increasing in 2005 [7]. More recently, Maurits and colleagues studied how persuasive technologies can dynamically adapt to how users are persuaded [31]. While these technologies all aim to directly change user behaviour, altering

behaviour by first changing attitude has not received enough attention: according to Kristian and Harri [7], five out of six persuasion-related studies address behavioural change and not attitude change.

While behaviour involves the expression of feelings or action, attitude involves the mind's predisposition to ideas, values, or people. Attitude structures can be described in terms of three components [7]:

- The affective component, involving a person's emotions about the attitude object (e.g., "I am scared of spiders");
- The behavioural (or conative) component, involving how attitude influences how individuals act or behave (e.g., "I will avoid spiders and scream if I see one");
- The cognitive component, involving a person's belief and knowledge about an attitude object (e.g., "I believe spiders are dangerous")

The ABC attitude model uses the above three components [32] and assumes the link between attitudes and behaviour: that people are rational, that they always attempt to behave rationally, and that a person's behaviour should be consistent with their attitude(s). While this principle may be sound, people do not always follow it and sometimes behave in seemingly illogical ways, such as by smoking cigarettes while knowing that smoking causes lung cancer and heart disease. LaPiere shows that the cognitive and affective components of attitude do not always match behaviour [33]. YAC is designed explicitly for influencing attitude. As for behavior, it aligns to match attitude. Thus, designing a system such as PCG to influence user attitude to change the user's behaviour would be useful.

Torning and Oinas-Kukkonen [7] and, later, Juho Hamari et al. [34] provide overviews of the history of applications in persuasive-systems design. According to Juho Hamari's

comprehensive literature review, the two most prevalent foci of persuasion-related studies are health and/or exercise (47.4%) and ecology (21.1%), which included, for example, technologies aimed at conserving energy.

Most of the above studies, some of which have resulted in commercial products or services, base behaviour persuasion on control sensors. Fit4Life [35], for example, combines the input of number of sensors to help users achieve their weight-loss goals [35]. The model tracks individual progress and generates actionable items by tracking sensor data and comparing that data against user goals. While these applications are intended to persuade users by controlling user behaviour, they are not designed to influence user attitude.

Busch and Patil used personalized content generated through surveys and gamification to train people to follow cyber-security best practices [36]. Gamification uses game concepts and mechanisms, such as competition and leader boards, in nonentertainment applications [37][38]. Once Busch and Patil's users read the personalized content they were given, they were presented with a quiz, a challenge, and a score. To encourage individuals to compete, the system posted scores on a board. While gamification can be used to increase engagement and potentially persuasion, the personalization in existing work is primarily in the form of categorization of users based on age, gender, or personality/gamer type rather than unique individual features [39][40][36].

Similar to Busch and Patil's research, our research uses personalized content, but we focus on changing attitude rather than specific behaviour and create user profiles based on social media data.

2.2.1 Generic Summarization

Generic text summarization approaches are generally divided into four categories [41]: (1) heuristic approaches that rely on sentence position within a document; (2) corpus-based methods like TF-IDF [42] that are based on numerical statistics that measure word weight in a document or corpus; (3) lexical-chains methods that account for discourse structure and search for chains of context words in text; and (4) knowledge-rich approaches that use Latent Semantic Analysis (LSA). Gong and Liu [43] were one of the first researchers to use LSA for text summarization. This method can find salient topics or concepts in a document and measure their relative importance within the document. The method represents concepts in the summary with sentences that best capture those respective concepts. Steinberger and Ježek [44], however, showed this approach fails to be a generic solution and can only be used in particular domains. Although none of the four categories is designed for personalization, they could be used as the foundation of personalized-summary-generator methods.

2.2.2 Personalized Summarization

A personalized summary provides information assumed to be the most important or interesting in a document for a user. Since a generator's personalized-summary outcome aligns with user attention (YAC's first criteria), our software framework and PCG system uses a summary generator as an essential module.

Sun et al. [45] analyzed click-through data. Click-through data records how users find information through queries (i.e., clicking on a search-result link to a web page indicates that terms from the search query describe the page and can be weighted more when summarizing the page).

Park et al. [46] summarized comments (i.e., descriptions) and tags that users added when creating bookmarks through social bookmarking services like Delicious (they did not summarize content). This approach can also summarize documents with no or minimal text, but which have other multimedia content. On the other hand, this approach depends on the number of added bookmarks and cannot summarize documents with no bookmarks [47].

Móro and Bieliková introduced a linear method based on a combination of raters that allowed researchers to consider parameters like user characteristics [48]. To build user-characteristic profiles, their approach relied on a user-modeling process designed for open corpus web-based systems, wherein content can be dynamically added or changed [49]. Diaz et al. developed a similar approach to personalized summarization that used a vector of weighted keywords [50]. Móro, Bieliková, and Diaz's model's main disadvantage is that users must manually insert keywords and weights into their model. In addition, many existing personalized-interest approaches isolate users from what they disagree with or are unaware of—a problem known as the filter bubble [51] and which prevents learning about and being exposed to new concepts and cultures. Further, as Frans and colleagues suggest, users may lose interest in content that lacks sufficient complexity [52]. For these reasons, our PCG system generates personalized summaries through an enhanced module that uses Diaz et al.'s and Frans et al.'s respective approaches [52].

2.2.3 Complexity and Comprehension

Adjusting content readability plays an important role in creating persuasive content. Classical (manually tuned) readability indices like the Flesch-Kincaid Index [53], the Gunning Fog Index [54], and the Coleman-Liau Index [55] have been proposed. These indices provide a reliable way to tell what level of education someone will need to be able to read a piece of text easily.

Recent research relies on statistical methods to measure readability. Callan and colleagues [56] predicted reading difficulty by using statistical language modeling. All these methods measure and adjust content readability to maximize the likelihood of persuasion.

Measuring and adjusting content processing difficulty is another way to improve content persuasion. Processing difficulty is commonly described as the measurable effort required to process a new token of information [57]. By measuring the textual features that make reading easier or harder to process, a textual-complexity analysis predicts the difficulty respective users will have with a given text. Processing difficulty, according to verbal efficiency theory [58], propagates from the bottom up: when lower-level textual interpretations (i.e., words) fail, higher-level interpretations (i.e., sentences) are also incorrect. However, less-than-optimal word processing may be sufficient to understand a text [59], indicating a difference between processing difficulty and comprehension. This means word-level difficulty may contribute to processing difficulty but not hurt comprehensibility.

Frans van der Sluis and coauthors [52] sorted methods for computing textual comprehensibility into four classes: traditional, lexical familiarity, priming, and dependency-locality. These methods are appropriate for large-scale applications because of their low-level computational complexity—they either use simple word-based representational models or highly optimized parsers.

Given the importance of comprehension and readability scores, their inclusion is crucial to personalized summarization and recommender systems. A system that can, with minimal supervision, access text-document repositories and generate persuasive content that matches user comprehension might be the next logical step in this field.

2.2.4 Opinion Mining

Evidence suggests that social media is causing people to become increasingly reliant on others' opinions. Online peers, via social media's information stream and social dynamics, influence each other's knowledge, opinions, and behaviours [60][61][62][63]. Given the rapid growth of user-generated content published on social networks, a tool for Web-mining that captures sentiments and opinions while evaluating the large-scale content credibility has become essential. Design-system research serving product vendors' and policy makers' analyze and aggregate customer attitudes to products, services, and organization features has mainly aimed to enable organizations to do so along dimensions like time, location, and personal/professional experience [64][65][66]. This research aims to boost trust between persuadee and author (content provider). Pang and Lee [67] and Ravi [66] published surveys of methods and approaches that focused on techniques to address new challenges raised by sentiment-aware applications (as opposed to challenges already addressed by more traditional fact-based analyses). While existing applications for opinion mining, sentiment analysis, and classification are suitable for entities like product reviews [67], [68], they are not optimal for online discussions in which comments convey argument and sentiment toward the discussion [69].

Preliminary literature in the field of trust-based-social-recommender systems demonstrates the advantage of recommending trustable content in marketing [70][71]. These works focus on recommender systems that, before recommending content to users, consider trustworthiness. Predicting the trust rating between content publishers and receivers is critical. For instance, in their work on inference of binary-trust relationships in web-based social networks, Goldbeck and Hendler [72] considered social networking sites on which users explicitly provide trust ratings to other members. Assigning trust ratings to each member is unfeasible on large social networks, so

Goldbeck and Hendler propose an inferring mechanism to assign trustworthy and nontrustworthy ratings to those who have not been assigned one. Their research did not create, test, or evaluate this inferring mechanism.

Although mining personal opinions and experiences consists of a series of challenging procedures, evaluating the credibility of collected data is also complex. Microblogs can both provide truthful news and spread misinformation and false rumors. Castillo and Mendoza analyzed the credibility of news propagated through Twitter using a classifier [73]. Their model uses features from user tweets, users' retweeting behaviours, posts' text, and citations to external sources. With precision and recall ranging from 70% to 80%, Castillo and Mendoza classified messages as credible or noncredible. On the other hand, instead of using classifiers, Soo Cho [74] proposed applying a profiling-and-text-analysis program called LIWC [75] to predict individual expertise. The system assigns a credibility level, based on the subject matter in relation to collected background knowledge about individuals, to content.

The individualization layer is meant to prove the potential value of opinionative data in elevating trust (i.e., persuasion) for content semiauto-generated by our system. Predicting content credibility is essential to this task.

2.3 Summary and Gap Analysis

As discussed, personalization has been widely interpreted as the process of preparing content for a specific person based on stated or implied preference [12][13][14]. This approach typically breaks groups of people into segments. Some scholars refer to this process as micro-segmentation [15];

In our literature review, we did not encounter a model that translated persuasion theories to a software that automatically generate persuasive, custom-made, personalized content. Existing

content generation methods are ad-hoc and do not follow comprehensive models. As such, we introduced our PoI model (inspired by YAC) and an associated software framework.

While there is no comprehensive model for PCG in the context of persuasive documents, and no integrated software framework to implement that model, there are modules, as reviewed above, that can be used and/or improved in such an integrated system. We identified appropriate modules during the literature review that could be refined and reused in our research. For example, in the segmentation layer, we decided to enhance Diaz and Frans' approach [52] as the module to generate personalized summaries.

Predicting credibility and trust between two users has been the main missing module in building persuasive systems. Calculating an information-trustworthiness score is the most important part of the individualization layer. Kawak studied a conventional follow-followee social graph [10] to calculate information-trustworthiness scores. He demonstrated that the graph's flaws are a means to investigate Twitter's role as a social or information network [76]. Most Twitter users follow back their followers in accordance with informal courtesy, but few users in follow relationships communicate among each other [6]; therefore, a Twitter-social-graph algorithm that focuses only on follow relationships is insufficient to measure trust with. Current research has not yet taken into proper consideration that people's relations consist of more than totalled retweets, likes, and follows. The role of feelings or sentiments needs to be considered more.

Intuitively, creating persuasive content has three basic tasks. First, we need to understand the intended audience, then find existing content, and finally tailor the content so it becomes persuasive. Figure 2.2 shows various components required for creating persuasive content. The gray boxes identify the ones missing in current research.

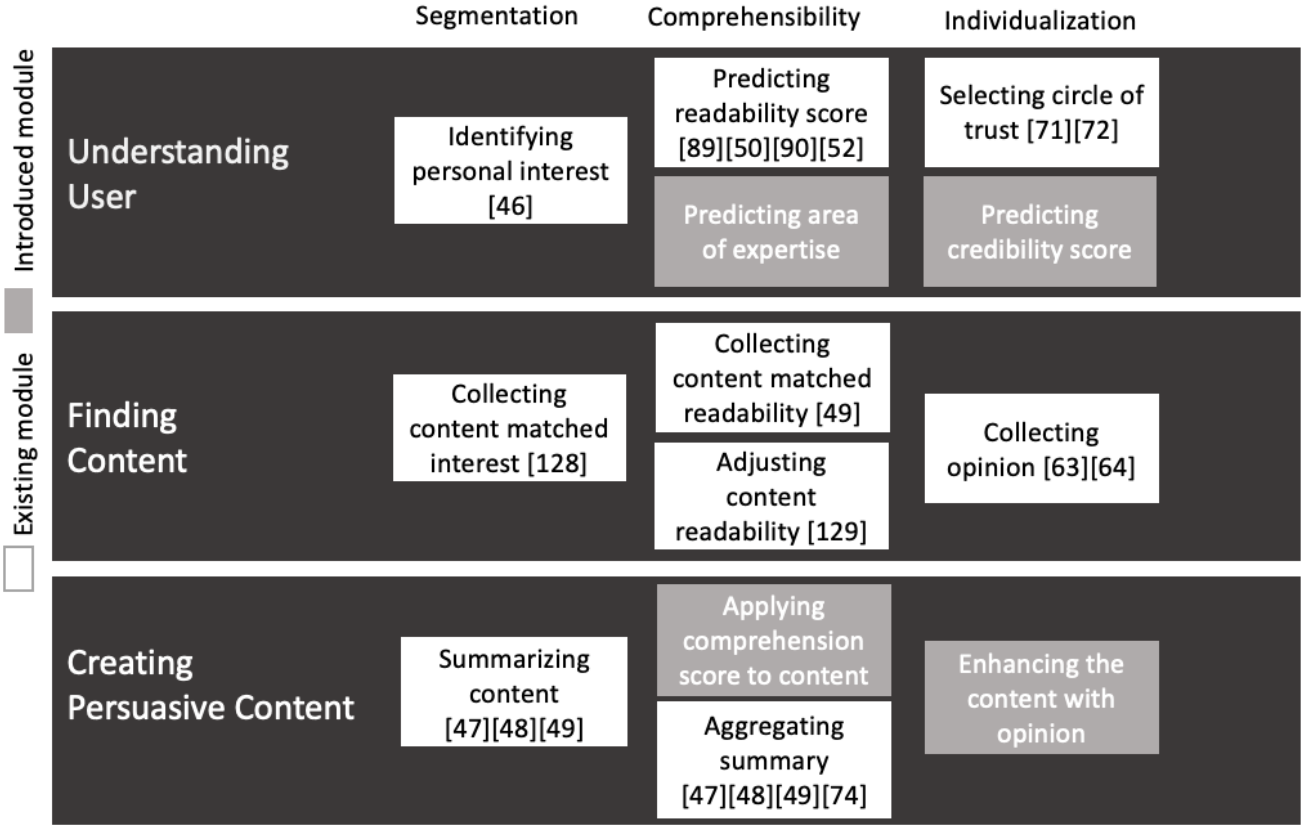


Figure 2.2 – Gap analysis in respect for the modules to create persuasive content

Chapter 3.

PROPOSED SYSTEM

3.1 PYRAMID OF INDIVIDUALIZATION

A large population can be divided into smaller population segments. Such segments enable groupings based on common interests and needs. The ability to understand or comprehend text is another distinguishing factor in the context of persuasive text and the reason why educational material is written to accord with vocabulary and difficulty levels appropriate to its target readers. Personality traits are another example of such distinguishing characteristics. While these factors allow people to be grouped, within each group individuals can be differentiated by individualized characteristics like personal experiences, friends and family, and opinions. We modeled these distinguishing factors as a multilayer pyramid that uses the general population for its large base and narrows to a specific individual through layers of distinguishing characteristics.

Figure 3.1 shows how we divided an individual's basic characteristics into four main layers: segmentation (demographic characteristics, such as age and gender), comprehension, personality traits, and individualization. The first three layers are usually shared amongst a large group of people, and their order may not be important. In contrast, the individualization layer includes personal features—like family history or personal opinions—that are shared within smaller crowds and can be truly individual (mutually exclusive). We represent these intimate features and final individualization at the pyramid's tip. It should be noted that each layer progressively narrows down the population, but the order of the first three layers does not have to

be as proposed. Particularly, Comprehension and Segmentation can be applied in any order or side-by-side.

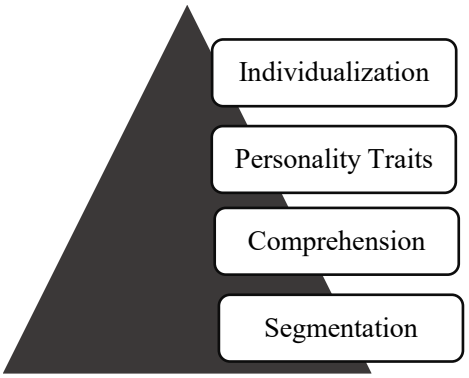


Figure 3.1 – Pyramid of individualization

Our primary research goal was to design and develop a system based on our PoI model and then evaluate whether it can improve the persuasiveness of content, as compared to a generic persuasive text written by an expert. Our first step was to build user profiles by gathering information from users' social-network accounts and other digital traces. Working from user profiles, we applied the PoI model to construct user profiles.

Although we named personality traits as a PoI layer, that layer is out of the scope of the software framework and the PCG system.

3.1.1 Segmentation Layer

YAC identifies recipient-attention attainment as the first step towards persuasion success. "Interesting," "relevant," and "enjoyable" are parameters that increase the likelihood of gaining reader attention [77][14]. The above parameters are associated with user preferences, and the segmentation layer is used to draw user attention based on user-preferences. Basic segmentation, according to age, gender, or marital status, is the easiest means to initial classification of individuals. Gender has always been a main segment for comparison. Rachel Croson and Uri Gneezy [66] reviewed the literature on gender differences and identified robust variations in three

main sections: risk preferences, social preferences, and competitive preferences. Furthermore, Andrews suggests relationships between gender differences and persuasive communication [78]. For instance, findings revealed that although men are more inclined to criterion-based argumentation, women are more likely to invent their own [78]. Segregating people according to age also effectively both determines people's content preferences and gains their initial attention, as has been studied in domains such as the entertainment and movie industry [79].

3.1.2 Comprehensibility Layer

The pyramid's second layer is comprehensibility adjustment. YAC suggests that audience acceptance and retention first require that an audience understand the presented message. For example, through three experiments, Eagly demonstrated how lowering comprehensibility lessened message acceptance [80]. This directly relates to Classical (manually tuned) readability indices like the Flesch-Kincaid Index [53], the Gunning Fog Index [54], and the Coleman-Liau Index [55]. Moreover, the criteria "easy to follow," "less complex," and "easy to understand" have been identified by researchers as a means to assess content comprehensibility [81][52].

3.1.3 Personality-Traits Layer

The third layer matches presentation wording with recipient personality traits. Various personality models categorize people according to long-term characteristics instead of short-term moods and emotions [82][83]. A common personality model is the big five personality traits. It is based on the five major traits of human personalities: openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism [84]. In a sample of 324 survey respondents it was found that persuasion (i.e., improving the chances of an audience accepting an argument and changing that audience's opinions) can be increased by aligning message with audience persona

[84]. Although the personality-traits layer is the third PoI layer, it is out of the scope of this thesis. Due to its complexity, personality can be the subject of further research.

3.1.4 Individualization Layer

The individualization layer's design is premised on the idea that personalized content can be enriched with personal data like family history or personal opinion. For example, a restaurant's online flyer that encourages people to reserve the restaurant for their future events could be enriched by incorporating a persuadee's personal photo from a past festive event instead of using a generic restaurant image. This could reinforce positive emotion and advertisement persuasiveness. However, using personal images in personalized content is only a simple example of individualization. Images of friends in the restaurant or a quote from them describing their happy experience there could increase the likelihood that an intended audience will trust the place more than another establishment. Creating trust and credibility is a persuasiveness strategy that directly relates to the YAC acceptance stage. To be more granular, parameters like "trustworthy," "accurate," "authentic," and "believable" have been commonly accepted as a means to increase the likelihood of reader acceptance of content [85][86].

3.2 System Overview

The general pipeline of our PCG system, implementing three layers of the PoI model, can be seen in Figure 3.2.

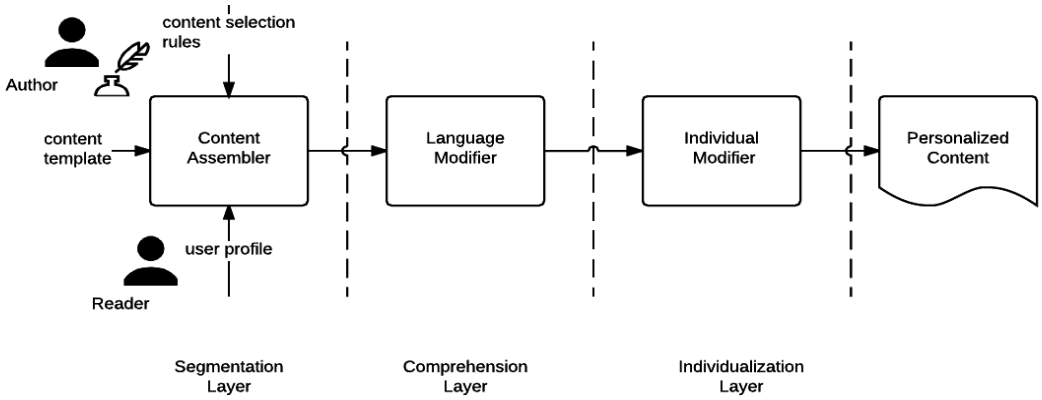


Figure 3.2 – PCG-software-framework-process overview

Figure 3.3 shows a detailed version of the segmentation, comprehension, and individualization layers within the PCG system. It consists of tasks like user-profile construction, content selection, and manual customization.

Our proposed system is a collaboration between multiple actors with assigned roles. Each actor is described below.

- System: the personalization system that assembles and prepares the content
- Author (persuader): the content designer who creates the document templates and rule files
- Reader (persuadee): the intended receiver of personalized content who logs into the system with proper authentication through their social-network account
- Expert: a close friend or family member or a public figure or organization whom the reader trusts and who has published material the system can use as supporting opinion

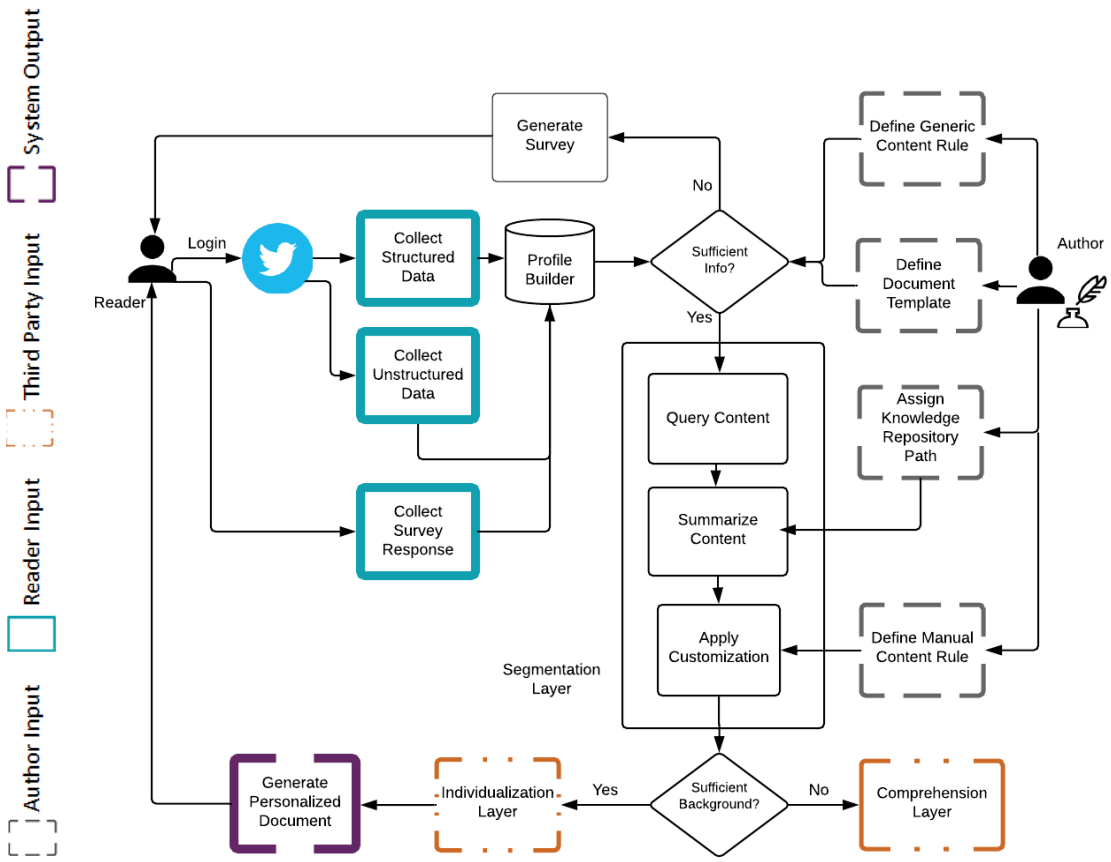


Figure 3.3 – PCG detailed overview

The following list overviews the system's main components:

- Document template: The document contains the template for the persuasive content. The places that require personalization by the system are identified with a unique ID in this template. It allows the aggregation of content from multiple sources.
- Content rules: the essential core logic fed to the system and used by the persuader to collect and summarize content. Rules consist of:
  - Segments: persuadee data from user profiles
  - Goals: topics the persuadee might potentially be interested in

- Knowledge repository: a collection of author-provided articles and documents[1]
- Profile builder: Using the collected data from the persuadee's social network account, the system builds the user profile. The data from the user profile is later used to predict the user's comprehension and interest score.
- Segmentation layer: the component where the system selects and assembles the initial content
- Comprehension layer: the component that allows the system to simplify wording to improve readability
- Individualization layer: the component where the system collects and appends related content (opinion) from users' credible sources as supporting information. Credibility may consist of trustworthiness, expertise, third-party endorsements, or verifiability and could also include images, posts, or other types of media from the users' content.

3.3 RULES DEFINITION

Content rules are the essential core logic prepared by the persuader and fed to the personalization engine to collect, analyze, and personalize content. There are two types of rules: generic and manual. Table 3.1 and Table 3.2 show samples for both types. Sections 3.4 and 3.6 further detail the rules' XML structure.

3.3.1 Generic Rule

A Generic rule consists of five main parts: segment, weight, content, topic ID, and polarity. In the XML context, segment and weight are mandatory content-element attributes, and content is

[1] In the current implementation, the knowledge repository is a Solr Apache Lucene [132] collection that contains articles for a specific topic.

the XML value for the content element. Topic and polarity are mandatory rule-element attributes. Except polarity, the above fields can accept string values. Polarity can only accept positive or negative values. Figure 3.4 shows their relationship. All attributes that are defined in rule schema are mandatory.

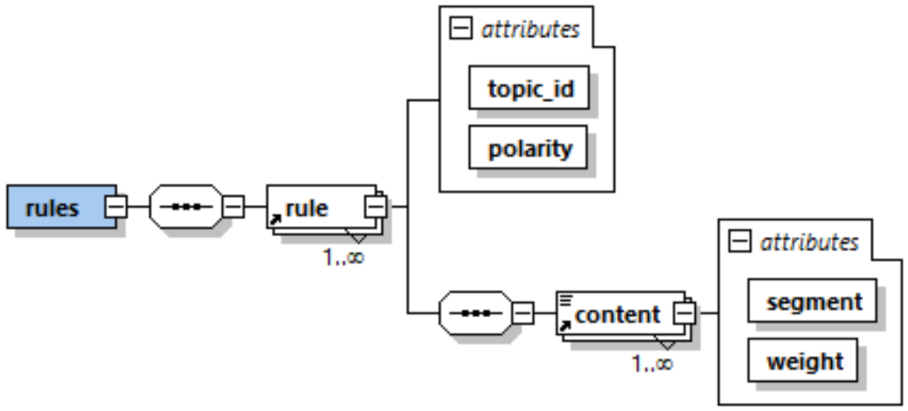


Figure 3.4 – Rule schema

Segments are persuadee data, like preferred location. Contents are topics an author has selected in association with a segment to display to the persuadee. To activate a rule, the system obtains all the necessary criteria (persuadee data) for the rule. Once the rule is activated, the system applies the segment and contents as a search query to retrieve and recommend the most relevant information from the knowledge repository. If the system selects more than one segment for a persuadee, it uses weight values to assimilate each segment's content into one item. A weight value is a number from 0 to 100. To provide more flexibility to the persuader, each segment may have one or more contents. For instance, in Table 3.1, the "New York City" segment has two content elements (smart cities and urban lifestyle). On the other hand, a rule with a blank segment becomes a global rule that applies to all persuadees.

Generic rule

```
<rule topic_id="city_generic" polarity="negative">
 <content segment="New York City" weight="20">smart cities;urban life style</content>
 <content segment ="Ottawa" weight="20">IoT</content>
 <content segment ="Vancouver" weight="20">Amazon</content>
</rule>
```

Table 3.1 - Generic rule example

Manual rule

```
<rule topic_id="city_manual" polarity="negative">
 <content segment ="New York City" weight="20">NYC is one of a couple of dozen cities
   around the world edging ever closer to becoming what is known as a "smart city,"
   an urban location tightly connected with advanced forms of technology involving
   not only mobile devices and ads but sophisticated forms of healthcare,
   energy, transportation, property management, and waste and water systems.
   A city must be advanced in several of those areas in order to become a smart city,
   according to consulting firm Frost & Sullivan. </content>
 <content segment ="Ottawa" weight="20">IoT (The actual content) </content>
 <content segment ="Vancouver" weight="20">Amazon (the actual content) </content>
</rule>
```

Table 3.2 – Manual rule example

Each rule element contains a topic ID and polarity attributes. The topic ID uniquely identifies a rule, while polarity indicates an author's negative or positive perspective toward content. While the system searches for information in the knowledge repository, it conducts sentiment analysis to ensure polarity matches author perspective (Section 3.5).

The system also uses the polarity attribute when collecting supporting opinions (Section 3.6). Polarity alignment between supporting opinions and retrieved information from the knowledge repository is essential. To avoid a contradiction between a supporting opinion and the

information, the system needs to conduct a sentiment analysis on the content to identify content polarity.

3.3.2 Manual Rule

Besides creating generic rules, the author can add manual rules to overwrite the contents that have been retrieved by generic rules. Doing so allows the author to prepare content and directly insert it into the final document. Table 3.2 shows an example for manual rule.

3.4 USER PROFILING

Before collecting content for the personalized document, the system needs to identify the target reader. The primary data source for user-profile creation is user social-media accounts. We used Twitter accounts because of their public-access flexibility using REST application programming interface (API). Our system creates user profiles from three information sources:

- Structured data: Twitter-profile information;
- Unstructured data: user tweets; and
- Persuadee direct input: additional information directly provided by the persuadee.

The Twitter API provides access to default profile information (structured data) such as user names, locations, and bios. It also provides access to other basic user information, such as numbers of followees, followers, and tweets.

User tweets are unstructured data that need to be managed and organized. We built our platforms for converting unstructured data into structured records through keyword extraction, sentiment analysis, domain classification, complexity assessment, and statistical analysis.

Despite using different machine-learning techniques for classifying tweets, the system may not be able to identify reader awareness of, and interest in, certain topics. By generating a short survey, the system collects the reader data needed for personalization.

Figure 3.5 overviews how the system, using the above-stated data sources, generates and updates user profiles. We discuss all the above-stated data sources in more detail in the next sections.

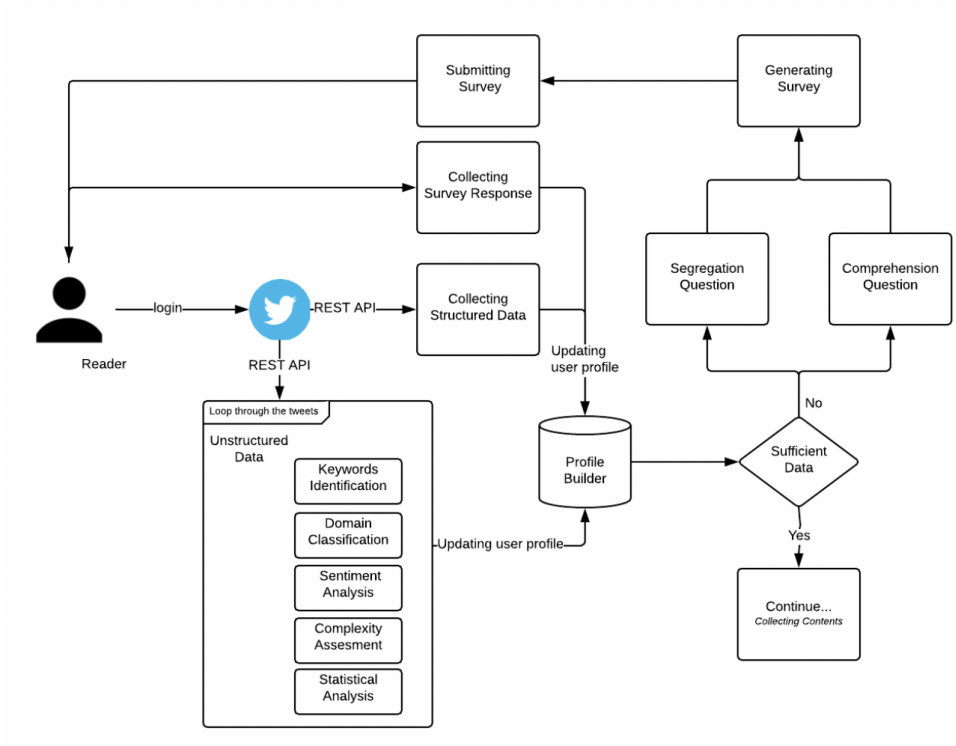


Figure 3.5 – User-profiling overview

3.4.1 Structured Data

Personal user information, such as name, age, location, and a summary of interests, is available in most social network and microblog services for researchers. Through open API, this information is easy to collect, and since the data is preorganized, it can be stored as is; however, data can be incomplete (a user may choose not to post bio details) or misleading. Furthermore,

other relevant user attributes, such as explicit and implicit interests or political preferences, are usually missing. Cheng [72] estimated that only 26% of users report a specific location, while the rest provide either general locations (e.g., states/provinces, countries) or nonexistent places. Pennacchiotti and Popescu [87] conducted a pilot study of a similar nature to assess direct use of public profile information, such as gender and ethnicity, from Twitter. In a corpus of 14M active users in April 2010, they found 48% of users provided a short bio and 80% a location. Therefore, we decided not to rely solely on extracted data from user profiles.

3.4.2 Unstructured Data

Users' tweets are a good example of unstructured data and cover insights on personal attitudes toward different topics (e.g., political orientation or ethnicity). To extract the information, the system relies on features like n-grams models [88], simple sociolinguistic features (e.g., presence of emotions), and communication behaviour (e.g., frequently retweeted content).

Figure 3.5 shows that the process of collecting and organizing knowledge from tweets by running the data through 5-stage pipeline. After extracting a keyword from the tweet (first module), the system runs the domain-classification (categorization), sentiment-analysis (polarization), and complexity-assessment modules. The following sections discuss these modules in more detail. Data retrieved from these modules is recorded in user profiles and updated by the statistical analysis module.

3.4.2.1 Keyword Extraction

Term frequency–inverse document frequency (TF-IDF) [89] and TextRank [90] are two of the most practical techniques for keyword extraction. Both techniques can extract user interests from a collection of tweets with great precision [91]. They are not, however, appropriate for a

single tweet since most terms in a tweet are used only once. We, therefore, used the default MAUI [92] toolkit as a baseline for automated keyword collection. MAUI enables the extraction of a list of potential keywords from a document and trains a decision tree using features like TF-IDF.

3.4.2.2 Domain Classification

Keywords on their own can be vague and ambiguous. As Figure 3.6 illustrates, a word like *bond* can have a variety of meanings in various contexts. Classifying keywords by domains helps identify keyword hyponymies and hypernyms: word X is a hyponym of word Y if X is a subtype or instance of Y. For example, *bond* in the finance domain is a hyponym of debt; at the same time, it is a hypernym of a broad range of asset types, such as government bonds, municipal bonds, and corporate bonds. Defining such a hierarchy for keywords would help predict user preferences. For instance, if a user prefers content about Java, C++, and Java Script, the person would likely be interested in other programming languages.

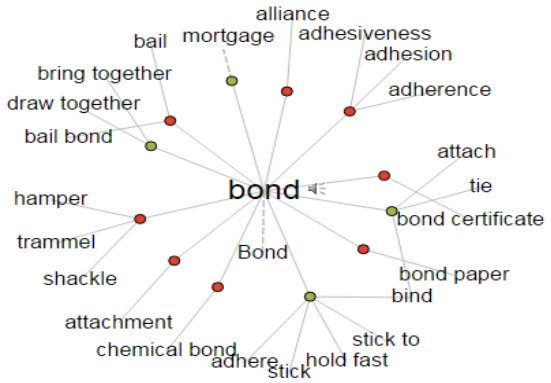


Figure 3.6 – Word map for bond created by Visual Thesaurus

A Word-Class Lattice (WCL) is a supervised lattice-based model used to identify textual definitions and extract hypernyms from those definitions [93]. After evaluating WCL with a dataset of 1000 first-paragraph sentences randomly sampled from Wikipedia, we found the tool

adequately rates precision, recall, and accuracy of definition and hypernym extraction. Thus, we used the tool in PCG system.

3.4.2.3 Sentiment Analysis

Negative user preferences should be segregated from positive user preferences. The polarity of a sentiment is the point on the evaluation scale that corresponds to how positive or negative a sentiment is. By comparing the polarity of user sentiment toward the extracted keywords and the polarity of the presented content, we ensured user sentiment aligns with author-selected polarity. To conduct the analysis, we relied on the uClassify sentiment-analysis toolkit [94]. uClassify treats sentiments as non-binary and in this study, sentiments were classified using common categories like positive, negative and neutral. This thesis's information-retrieval and opinion-mining sections further discuss how this module is used.

3.4.2.4 Complexity Assessment

Our personalization process assesses user comprehension via a user-document-keyword-complexity evaluation that ensures that the complexity of generated content matches user preference. The simplest way to perform a comprehensibility-based ranking for a given topic is to build a classifier that assigns a comprehensibility score. Inspired by Tan et al.'s approach [95], we trained a logistic-regression classifier by using comprehensibility indices to extract pages from Simple English Wikipedia and English Wikipedia [53][96][55][54]. The maximum number of characters per tweet is 280. Since the above method relies on word length and sentence length and since 280 characters is insufficient for calculating comprehension score, we only assessed the comprehension score for tweets that contained URLs to other documents.

3.4.2.5 Statistical Analysis

Given that user opinions, preferences, and attitudes change, user profiles must be updated to reflect changes. Updating profiles includes updating keyword frequency, comprehensibility scores, and variance within hypernyms. A large keyword (topic) variance within a hypernym means a user probably has in-depth knowledge of the topic. Variance is also a sign that a user is comfortable being exposed to new topics. For example, if users are familiar with C++, Java, and JS, they are not likely to be computer-programming novices. Thus, they may be interested in higher complexity content within this field.

3.4.3 Persuadee Direct Input

Ambiguity and a lack of consistency in collected tweets is a major challenge for automated systems and can introduce misrepresentative data to user profiles. To alleviate data sparseness, we built a survey-generator module. The system uses the survey-generator module when it fails to identify a user preference on a topic. We designed the survey to enable sufficient data collection to tune a predicted user-interest-and-comprehension score. The system generates two question types: segregation preference and comprehension preference:

- Segregation-preference questions are generated based on the segment attribute in the generic-rule file (e.g., in Table 3.3, the questionnaire asks for the user's income level) and help the system segregate users and assemble persuasive content;
- Comprehension-preference questions: Comprehensibility is user-specific because users need adequate background knowledge to comprehend the persuasive content. These questions are meant to be used in the collection of users' comprehensibility preferences for a specific topic. For instance, a well-known professor might possess advanced knowledge

of computer science but be an intermediate mechanic; thus, he prefers easy-to-understand mechanical repair instructions. As Table 3.3 shows, the system asked readers to rank their expertise level (from 1 to 5) of a given topic. This information is necessary for the system to adjust the level of presented background knowledge in the final personalized document.

Document Template

```
<div id= "RRSP_benefits">
 Default text(content) given by the author. The text will be used as the
 fallback in case the system fails to identify a reliable personalized content.
</div>
```

Rule Template

```
<rule topic="RRSP_benefits" polarity="positive">
 <content segment="upper class income" weight="20">tax benefit</content>
 <content segment ="middle class income" weight="20">retirement fund</content>
 <content segment ="lower class income" weight="20">first time home buyer</content>
</rule>
```

Persuadee Direct Input

On scale of 1 to 5, where "1" means "Not Familiar" and "5" means "Very Familiar", how familiar you are with the topic of "RRSP Benefits"? - Purpose: Determining comprehensibility

○ 1 (Not comfortable) ○ 2 ◉ 3 ○ 4 ○ 5 (Very comfortable)

Which category would you put yourself in? - Purpose: Assigning segregation

○ Lower class income ○ Middle class income ◉ Upper class income

Table 3.3 – RRSP Benefit document template, rule template and survey layout

To improve user experience, we limited the number of questions we asked users. One way to achieve this is by enhancing the system's ability to extract more accurate data from tweets. Since this may not always be possible, we required authors to assign a weight factor (importance level) to each content. The weight factor is a number from 0 to 100 where 0 means not important and

100 means extremely important. If content has a high-weight factor, the system generates a question for that content. A high-assigned weight means, the weight factor is greater than a defined threshold. The default value for the threshold is 25. Otherwise, the system discards the content and uses the generic content (the fallback option). This process is illustrated in Figure 3.7.

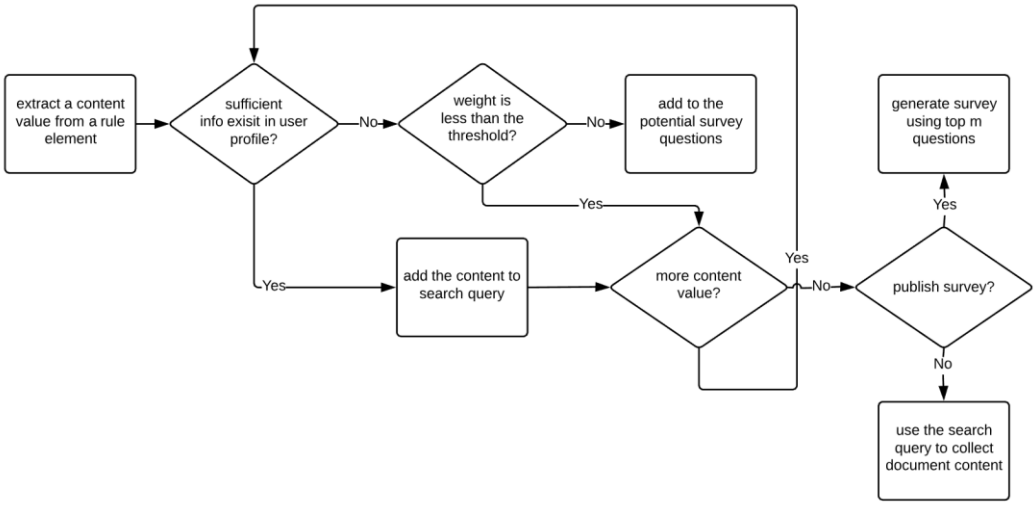


Figure 3.7 – Generating segmentation and comprehension-preference process

3.5 INFORMATION ASSEMBLER

The segmentation layer acts as our system's primary content assembler. Therefore, retrieving and assembling persuasive information is the main subcomponent of the segmentation layer. The segmentation layer initiates by building a query to retrieve documents from the given local-knowledge repository. As a part of our technical contribution, besides building a query to retrieve documents, the system analyzes sentiment to identify documents matching author polarity.

The current implementation's knowledge repository is a Solr Apache Lucene collection that contains articles for a specific topic (articles are indexed according to title and author-assigned keywords). The author gives the system the knowledge-repository URL. The Solr query to retrieve documents consists of two rule factors that are ordered according to priority:

- Rule content specialization within the content's domain (e.g., retirement fund in Table 3.3)
- Rule segment User criteria required for content segregation (e.g., middle-class income in Table 3.3)

When the system executes a given query by an author, it identifies documents with relevant information on an anticipated topic that are likely to contain content within the specialization. The system reuses the sentiment-analysis module from the polarization of user-profile keywords to compare the extracted content sentiment with the assigned sentiment value given in the rule. In addition, building the query means a user does not necessarily fall into a single category. For example, because a user may relate to multiple domains (rule topic), the query will expand to include all categories.

Solr next returns a list of relevant documents. Then, the system summarizes the documents using an unsupervised technique inspired by the TextRank algorithm [90]. TextRank is easy to adapt and ranks all sentences in a text. It consists of similar tasks like building graphs for texts, where the graph vertices represent the units to be ranked.

Our system includes the classic TF-IDF weighting system (a common numerical statistic used in TextRank) [97] in a slightly modified form, wherein TF-ISF weights are computed instead. TF-ISF is a more suitable weighting system for our system, since it ranks sentences instead of words. $TF_{i,j}$ is term frequency of ith index term in the jth sentence, and ISF_i is inverse-sentence frequency of ith index term [98][99]. Like the TF-IDF model, the corresponding weight for a sentence is computed as

$$ws_{ij}=tf_{ij}.\ isf_i \tag{3.1}$$

Due to effectiveness of cosine similarity, we have used it to measure the similarity weight between the two sentences (s_a and s_b).

$$w_{sim}(s_a, s_b) = \frac{\sum ws_{i,a} . ws_{i,b}}{\sqrt{\sum ws_{i,a}^2} . \sqrt{\sum ws_{i,b}^2}} \tag{3.2}$$

Furthermore, the system sorts sentences based on ranks of node using the PageRank [100] algorithm. To preserve content coherence, the system retrieves and reorders the top n sentences according to the source document's original order. If the system fails to retrieve relevant content from the knowledge repository, it relies on author-given default content in the presentation template.

3.6 OPINION MINING

As Section 3.1.4 discusses, highly trustworthy communicators inspire positive attitudes toward the positions they advocate better than less trustworthy communicators [101]. Collecting supporting opinion from a trustable source requires multiple subprocesses (as Figure 3.8 illustrates).

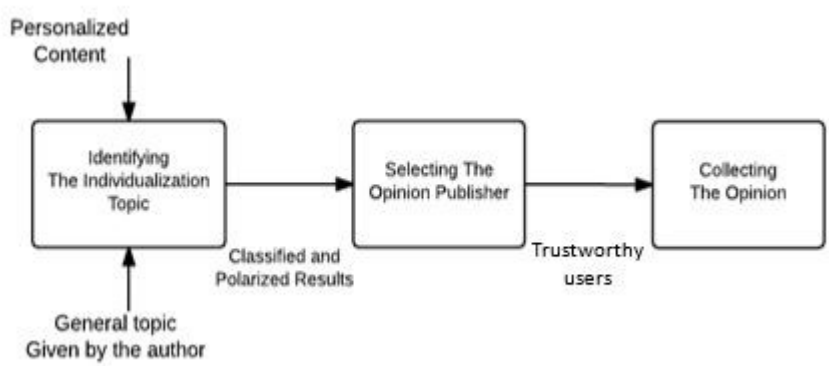


Figure 3.8 – General overview of the individualization layer

3.6.1 Identifying the Individualization Topic

Domain classification of content is the prerequisite for mining supporting arguments (opinions). As the segmentation-layer section discusses, the system collects content via user profiling and by applying author-driven content rules. The content rules are the logic that guides the recommender system's collection, analysis, and personalization of content. Table 3.2 shows that each rule element contains a polarity attribute. This attribute indicates an author's perspective toward a content element. To avoid contradiction between a supporting opinion and the content, the system first conducts a sentiment analysis on the content to identify content polarity. For instance, the following sentence implies a negative opinion about investing: *While stocks have historically performed well over the long term, there's no guarantee you'll make money on a stock at any given time, and you could lose all your principal*. Thus, the following would not make a good supporting opinion because it expresses the opposite perspective: *Just made a $3,000 investment in stock last year, and it turned into a $25,000 fortune within a couple years*.

A sentiment's polarity is the point on the evaluation scale that corresponds to the system's positive or negative evaluation of the sentiment. Our system verifies whether the supporting opinion has sentiment information that matches the intended personalized content.

3.6.2 Searching for the Trustable Source

Acquiring expertise in any domain means going beyond ordinary learning and progressing from rule- and fact-based knowing to experience-based knowing [102]. Unfortunately, data generated in social-network accounts does not have the semantic structure of product reviews on typical electronic-commerce websites. Such a data structure allows publishers to build reputation and expertise in a community. In Twitter, data is limited to the number of likes and tweets of posts

and our polarity analysis of the comment. These tools are essential to building trust graphs that evaluate trust between two people.

Neither a social graph nor structured data (such as total likes or retweets) are sufficient to measure user trustworthiness in a given social Twitter group. The sentiment associated with tweets shared between two parties is important. Existing approaches only use structured data to consider relationships among users. To measure user trustworthiness in a given social Twitter group, we introduced the user-trust graph to estimate relationship-trust strength from interactions (e.g., communication, tagging) and common interests. This graph is based on trust-score calculation as a function of time and tweet sentiment.

The graph consists of nodes (corresponding to user accounts and tweets) and edges (corresponding to follow and retweet relationships). Unlike the Twitter social graph, which is relatively static, the user-trust graph is dynamic and reconstructed when: one user mentions the other in a tweet, both users share a retweet or hashtag, one user likes a tweet by the other.

We modelled interactions between two Twitter users through a trust graph with multiple links. In this user-trust graph, each node represents a user. There are two types of edges in this graph: directed and undirected. A directed edge between two nodes—A and B—exists if User A mentions User B in at least one tweet or likes a tweet from User B or retweets at least one of User B's tweets. When two users share a common hashtag or retweet, there is an undirected edge between two nodes (as Figure 3.9 illustrates).

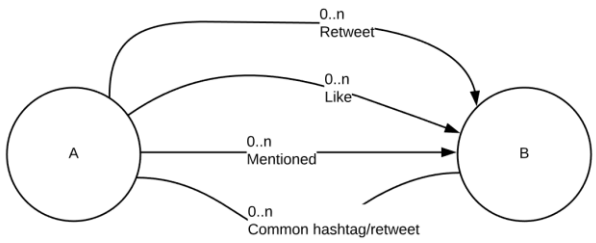


Figure 3.9 – User-trust graph

The system performs sentiment analysis on interactions between users and assigns respective positive, neutral, or negative scores. Despite the sentiment of a tweet's content, when a user retweets another person's tweet, the interaction is considered to have a positive score. But a retweet with a negative comment is considered to have a negative score. Intuitively, a greater positive score is interpreted as representing stronger trust strength between the two users.

Human relations may change over time; correspondingly, trust strength varies between two people over time. As Dai and Davison discuss [103], information should be ranked according to recency. Further, recent interaction with a positive sentiment is interpreted as a sign that the users currently share interest in that subject. To apply recency during weight calculation, the system labels edges (user interactions) with event timestamps. This allows the system to filter edges based on temporal parameters and thereby measure trend within a timeframe.

The system then tries to calculate interaction weight. Equation 3.3 is based on defining trust as a function of time. If the edge (interaction) from Node A to Node B exists with sentiment value e, the interaction's trust weight is $w = e / \mu_t$, where μ_t is the weight coefficient for the given timestamp. The weight coefficient depends on two factors: ΔTE_{ab} (elapsed time from the interaction time (e_t) between Users A and B to the current time) and ΔTI_{ab} (elapsed time from the initial interaction time between Users A and B to the current time).

$$\mu_t = {\Delta TE_{ab}}/{\Delta TI_{ab}} \tag{3.3}$$

In Equation 3.4, DT is the direct trust experience from A to B within a given domain, and W is the sentiment (weight) associated with direct interaction between the users. In this equation, W_{DL}, W_{DM}, and W_{DR} are weights associated to likes, mentions and retweets respectively.

$$DT_{ab} = (W_{DL}, W_{DM}, W_{DR}) \tag{3.4}$$

$$DTW_{ab} = {\sum w_{ab}}/{|W_{ab}|} \tag{3.5}$$

In Equation 3.5, $W_{ab} = \{w_{ab1}, w_{ab2}, \dots, w_{abn}\}$ captures the set of all sentiment score in the form of likes, retweets, and mentions that User A gives User B's content.

While *DT* is the direct trust strength, common preferences and content (in the form of retweets and hashtags) exist between users. This data can also be used as a second input source to improve trust-strength prediction, even though there is no direct interaction or trust path from the User A to User B. This value is demonstrated as *CT.* Equation 3.6 shows that *DT* and *CT* become the foundation for Twitter trust weight (T_{ab}).

$$T_{ab} = \alpha DT_{ab} + (1 - \alpha)CT_{ab} \tag{3.6}$$

$$where\ CT_{ab} = (W_{CR}, W_{CH}) \tag{3.7}$$

$$CTW_{ab} = {\sum w_{ab}}/{|W_{ab}|} \quad 0 < W_{ab} < 1 \tag{3.8}$$

Parameter α is meant to establish a minimum criteria that produces trustable scores calculated from direct communication. Full confidence in a direct trust-experience value requires a certain number of direct interactions between two users. As the number of direct interactions increases, confidence (reliability) increases until it reaches a point signifying a close relationship between two users. On the other hand, if user *a* has a small number of direct interactions with user

b, the value for DT_{ab} cannot reliably quantify trust value. Thus, we proposed that the combination mechanism (Equation 3.6) could reduce the weight for DT and increase the importance of the common-interaction trust value. CT_{ab} helps complement the uncertainty from the lack of direct feedback. To calculate the reliability of DT_{ab}, we rely on Equation 3.9, which Xu et al. [104] and Sabater [105] proposed.

$$\alpha_{ab} = \begin{cases} \sin(\pi/2 \cdot |W_{ab}|/N_{min} & |W_{ab}| \in [0, N_{min}] \\ 1 & Otherwise \end{cases} \tag{3.9}$$

In the above formula, N_{min} is the minimum number of sentiment score in the form of likes, retweets, and mentions that User A gives on User B's content.

Chapter 4.

SOFTWARE PROTOTYPING

4.1 IMPLEMENTATION OVERVIEW

This section highlights the principles that shaped our prototype design and development process. We implemented the first half of the PCG system (PCG v1) to cover components associated with the segmentation and comprehension layers. PCG v2 added the individualization layer's components and revisions to the PCG v1 components.

PCG v1 allowed us to build user profiles from Twitter accounts. By parsing and reading rule files, the system built relevant queries and retrieved information. The retrieved information matched users' respective interests and comprehensibility preferences. Finally, the system (PCG v1) trimmed the retrieved documents and assembled the content according to the document template. PCG v2 addressed issues discovered during PCG v1 prototyping and evaluation. It also covered individualization-layer components like the trust-graph and the collection of opinionative information.

During implementation, we factored in the following strategies:

- Component Reusability: We employed a reusable component-based approach for our system design using and improving existing modules, developing new components, and creating a framework. For summarization or sentiment analysis, we relied on TextRank [90] and the uClassify-sentiment-analysis toolkit [94]. As a separate component for generating UI presentation, we used impress.js, an open-source presentation toolset comprised of a JS library, and CSS3.

- Web Portability: Unlike desktop applications, web applications can be easily ported to virtually any platform with a web browser and can be easily accessed without a complex setup process. There are many existing packages, such as a variety of JS libraries, to help us with development.
- Adaptability: Adaptability to other components, such as other social networks, was one of our main design goals. Although we implemented the current system to integrate with Twitter to generate user profiles, the design can support expansion to other types of social media accounts, such as Facebook or LinkedIn. Aggregating different sources of information could potentially result in more accurate user profiles, although that is beyond the scope of this thesis.

Opinion-mining [67] involves various methods and techniques, including information retrieval, artificial intelligence, and natural language processing. This confluence of different approaches is explained by the nature of the data being processed (free-form text) and the application requirements (scalability, online operation) [106]. For this research, we selected a typical opinion-mining process that involves identifying, classifying, and analyzing sentiment polarity. Besides collecting the content, it conducts binary-sentiment classification of positive and negative emotions. The system handles sentiment classification with the same module we used to build user profiles. To improve opinion quality, the system connects to BrainyQuote.com's repository of quotes.[1]

[1] A resource that provides free access to an extensive database of quotes from prominent historical figures and contemporary newsmakers—celebrities, athletes, politicians, authors, and other public figures

4.2 PROTOTYPING OVERVIEW

Prior to designing and launching the final experiment, we built three prototypes to improve accuracy and overall quality of the delivered PCG-system result. Our aim in conducting multiple pilot studies was to identify major system issues that could cause the final experiment to fail. Through our pilot studies, we developed and tested research-instrument adequacy, assessed research feasibility, designed a research protocol, and assessed whether our protocol is workable and effective. The pilot studies also enabled us to collect valuable information that helped us shape the final experiment's structure. The pilot study:

- Established the effectiveness of the sampling frame and technique,
- Assessed the likely success of proposed recruitment approaches,
- Identified logistical problems in performing the experiment,
- Estimated outcome variability to help determine sample size,
- Collected preliminary data,
- Determined resources (e.g., finance, staff, etc.) needed for a planned study,
- Assessed proposed data-analysis techniques to uncover potential problems, and

4.3 PCG v1

With PCG v1, we evaluated the system using heuristic analysis. To do so, we picked three Twitter profiles as our target audience (Table 4.1), built user profiles using those Twitter accounts, and generated the persuasive content.

Once, the system successfully generated the content, we took a holistic view to identify problems. We found this approach to be a suitable means of quickly generating results with minimal effort, especially as compared to conventional usability-inspection methods.

Walter Knitl	@praxiem
•A product manager who loves all technology, with special interest in IoT.	
Peter Stewart	@peterstew
•An entrepreneur, small business enthusiast, and practical business advisor with an emphasis on planning, operations and execution.	
Kirk Denininger	@kirkdeininger
•A tech entrepreneur with a specialization in business process automation and branding.	

Table 4.1 – User-persona overview

4.3.1 Testing Preparation

All three profiles belonged to people with business backgrounds. We assumed the authors' shared goal was to put together educational documents to promote using a new marketing strategy. We downloaded fifty articles related to marketing strategy from the Harvard School of Business portal. Each article contained an average of 3000 words. The author's task involved creating a rule file and a document template. Once the user logged in, the system generated the persuasive content and appended it to the document template. The document template consisted of three sections (as Figure 4.1 demonstrates):

- Introducing the problem,
- Offering a solution, and
- Elaborating an example.

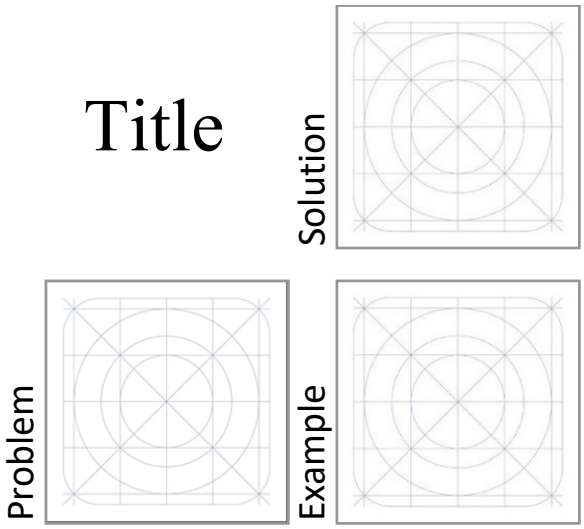


Figure 4.1 – Document template

This exercise's problem and example sections are associated with a rule. As Section 3.3 discusses, a rule is a set of segments and contents. Figure 4.2 and Table 4.2 show three segments (start up, social media, and brand) for the problem section. Each segment is directly associated with a content query. For example, for people related to start ups, the author recommends content related to managing cost. Since people may be associated with all given segments, the system applies a weight factor to balance coverage for each content query.

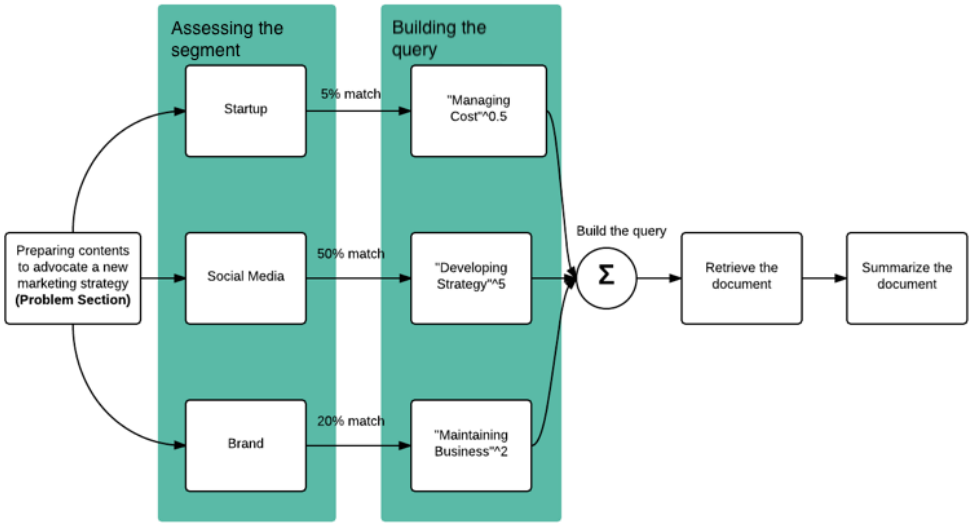


Figure 4.2 – Rule overview

Generic Rule

```
<?xml version="1.0" encoding="UTF-8"?>
<rules>
<rule topic_id="problem" polarity="positive">
 <content segment="startups" weight="20">cost</content>
 <content segment ="social media" weight="20">strategy</content>
 <content segment ="brand" weight="20">business</content>
</rule>
<rule topic_id="solution" polarity="positive">
 <content segment ="product manager" weight="20">sales</content>
 <content segment ="designer" weight="20">user experience</content>
 <content segment ="entrepreneur" weight="20">client focus</content>
</rule>
<rule topic_id="example" polarity="negative">
 <content segment ="new york city" weight="20">smart cities</content>
 <content segment ="ottawa" weight="20">iot</content>
 <content segment ="vancover" weight="20">Amazon</content>
</rule>
</rules>
```

Document Template

```
<html lang="en">
 <head>
 <meta http-equiv="Content-Type" content="text/html; charset=utf-8"/>
 <title>Personalized Document</title>
<link rel="stylesheet" href="http://maxcdn.bootstrapcdn.com/.../css/bootstrap.min.css">
 <link rel="stylesheet" href="style.css" media="screen">
 </head>
 <body>
   <div id="header_1" class="header"></div>
   <div id="problem" class="topic"></div>
   <div id="header_2" class="header"></div>
   <div id="solution" class="topic"></div>
   <div id="header_3" class="header"></div>
   <div id="example" class="topic"></div>
</body>
</html>
```

Table 4.2 – Generic-rule-and-document-template detailed overview

4.3.2 Prototype Results

Once a user logged into the system using their Twitter account, we started building their user profile by collecting their tweets. If a tweet contained a URL, we captured the URL and extracted the top five keywords from the text linked to by the URL using TextRank. Since tweets are usually short, we also collected the timeline's top ten keywords - a limitation defined by API restrictions. Then, using uClassify [94], we identified the content's hyponymy (domain). As Section 5.1.2 discusses, we used the hierarchical structure to adjust the complexity level of the retrieved content. After we identified the content's hyponymy, we calculated content-readability indices [53][96][54] to estimate a given domain's comprehension level.

The keywords given in a user profile demonstrated that user's topic-relation factor. When building the query, the system looked for matching rule segments in the user profile's keyword collection. The system picked the segment with the highest ratio match for the query. If the segment failed to match any of the user's keywords, the system looked for matching rule goals in the user profile as a fallback. Once the query was completed, the system passed the query to Solr so that the most relevant documents from the knowledge repository are retrieved. The summarized results from the retrieved documents of all three users follow below in Figure 4.3, Figure 4.4, and Figure 4.5 respectively.

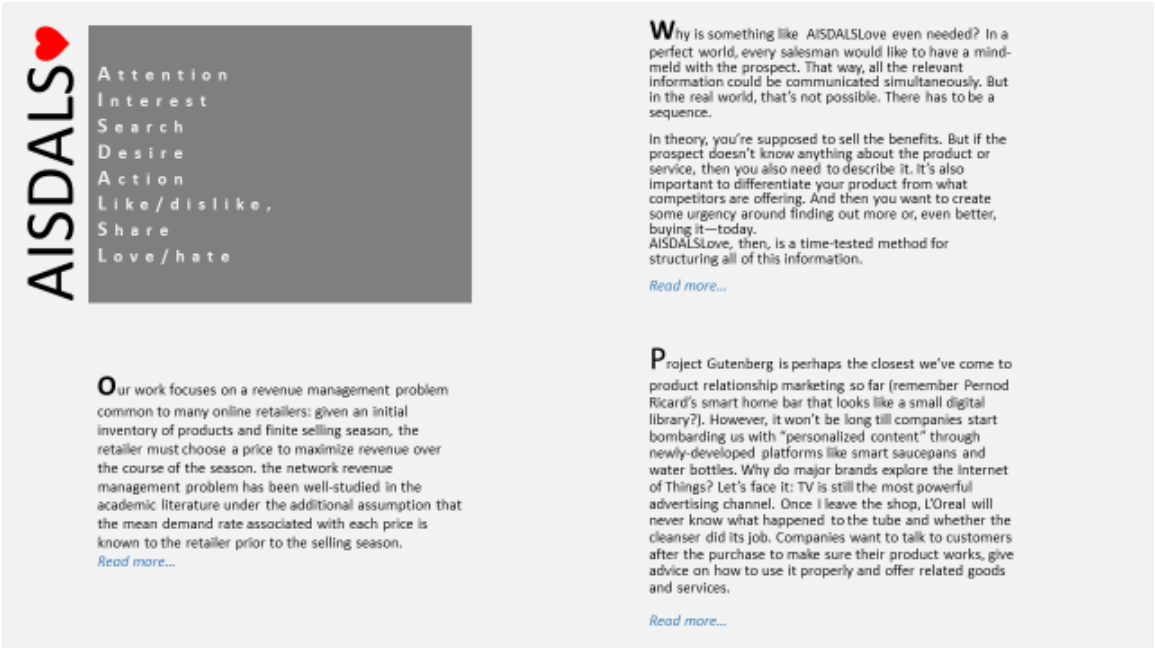


Figure 4.3 – Kirk presentation

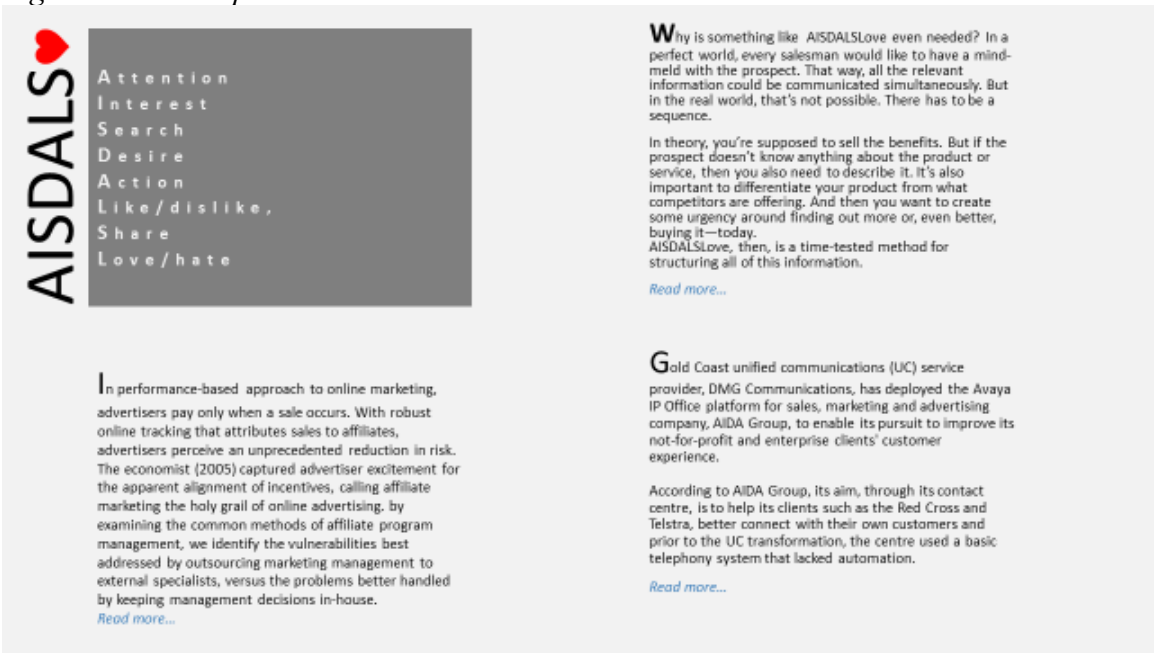


Figure 4.4 – Peter presentation

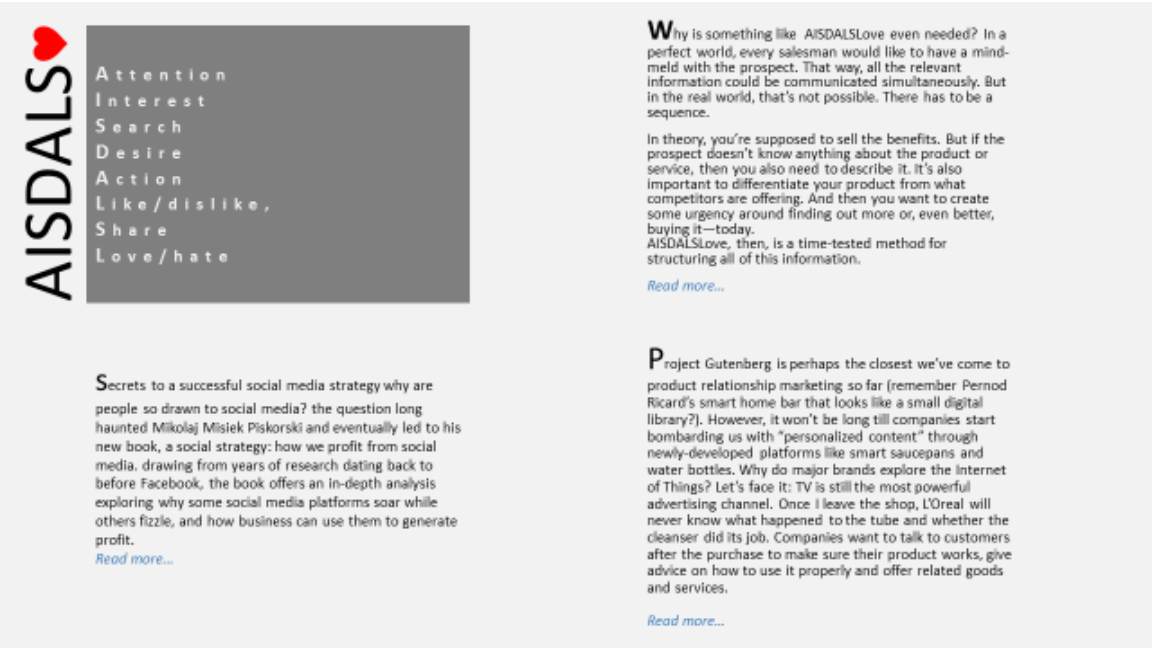
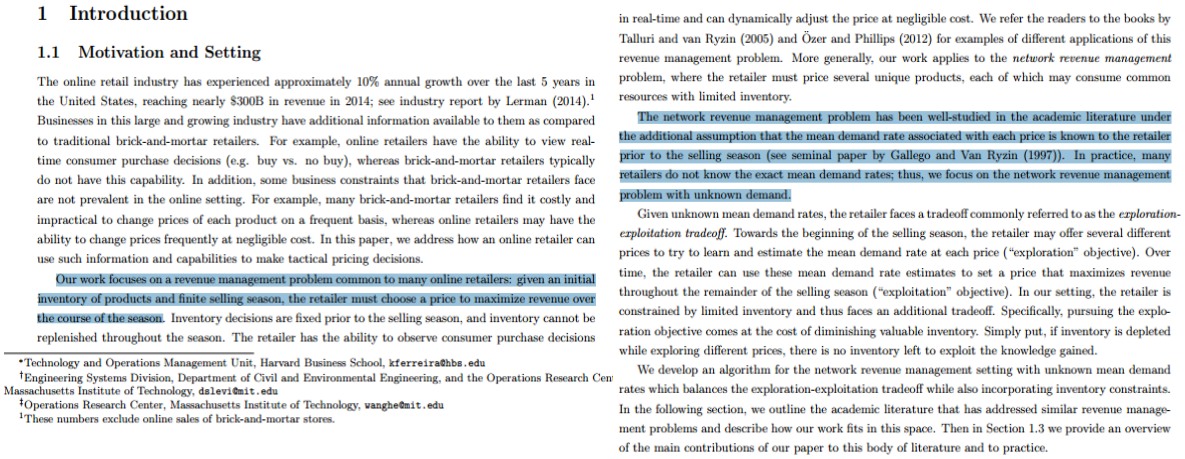


Figure 4.5 – Walter presentation

Kirk's (Figure 4.3) problem section contains a personalized summary of "Online Network Revenue Management using Thompson Sampling" by Kris Johnson Ferreira et al. [107], an article publicly available on the Harvard School of Business website. For performance reasons, we decided to use only the abstract, introduction, and conclusion sections of the article for summarization (roughly 1800 words). The system structured the graph for the document and sorted the sentences based on relevancy to cost management. As illustrated below, the first two sentences are selected from two different parts of the article.

Since the solution section is not associated to a rule, the system used the author-provided text to fill the section; thus, all three users saw the exact same content. For the next step to evaluate the system, we assigned a rule for the solution section such as one in Figure 4.7.



1 Introduction

1.1 Motivation and Setting

The online retail industry has experienced approximately 10% annual growth over the last 5 years in the United States, reaching nearly $300B in revenue in 2014; see industry report by Lerman (2014).[1] Businesses in this large and growing industry have additional information available to them as compared to traditional brick-and-mortar retailers. For example, online retailers have the ability to view real-time consumer purchase decisions (e.g. buy vs. no buy), whereas brick-and-mortar retailers typically do not have this capability. In addition, some business constraints that brick-and-mortar retailers face are not prevalent in the online setting. For example, many brick-and-mortar retailers find it costly and impractical to change prices of each product on a frequent basis, whereas online retailers may have the ability to change prices frequently at negligible cost. In this paper, we address how an online retailer can use such information and capabilities to make tactical pricing decisions.

Our work focuses on a revenue management problem common to many online retailers: given an initial inventory of products and finite selling season, the retailer must choose a price to maximize revenue over the course of the season. Inventory decisions are fixed prior to the selling season, and inventory cannot be replenished throughout the season. The retailer has the ability to observe consumer purchase decisions in real-time and can dynamically adjust the price at negligible cost. We refer the readers to the books by Talluri and van Ryzin (2005) and Özer and Phillips (2012) for examples of different applications of this revenue management problem. More generally, our work applies to the *network revenue management* problem, where the retailer must price several unique products, each of which may consume common resources with limited inventory.

The network revenue management problem has been well-studied in the academic literature under the additional assumption that the mean demand rate associated with each price is known to the retailer prior to the selling season (see seminal paper by Gallego and Van Ryzin (1997)). In practice, many retailers do not know the exact mean demand rates; thus, we focus on the network revenue management problem with unknown demand.

Given unknown mean demand rates, the retailer faces a tradeoff commonly referred to as the *exploration-exploitation tradeoff*. Towards the beginning of the selling season, the retailer may offer several different prices to try to learn and estimate the mean demand rate at each price ("exploration" objective). Over time, the retailer can use these mean demand rate estimates to set a price that maximizes revenue throughout the remainder of the selling season ("exploitation" objective). In our setting, the retailer is constrained by limited inventory and thus faces an additional tradeoff. Specifically, pursuing the exploration objective comes at the cost of diminishing valuable inventory. Simply put, if inventory is depleted while exploring different prices, there is no inventory left to exploit the knowledge gained.

We develop an algorithm for the network revenue management setting with unknown mean demand rates which balances the exploration-exploitation tradeoff while also incorporating inventory constraints. In the following section, we outline the academic literature that has addressed similar revenue management problems and describe how our work fits in this space. Then in Section 1.3 we provide an overview of the main contributions of our paper to this body of literature and to practice.

*Technology and Operations Management Unit, Harvard Business School, kferreira@hbs.edu

†Engineering Systems Division, Department of Civil and Environmental Engineering, and the Operations Research Cen Massachusetts Institute of Technology, dslevi@mit.edu

‡Operations Research Center, Massachusetts Institute of Technology, wanghe@mit.edu

‡These numbers exclude online sales of brick-and-mortar stores.

Figure 4.6 – Sentence selection

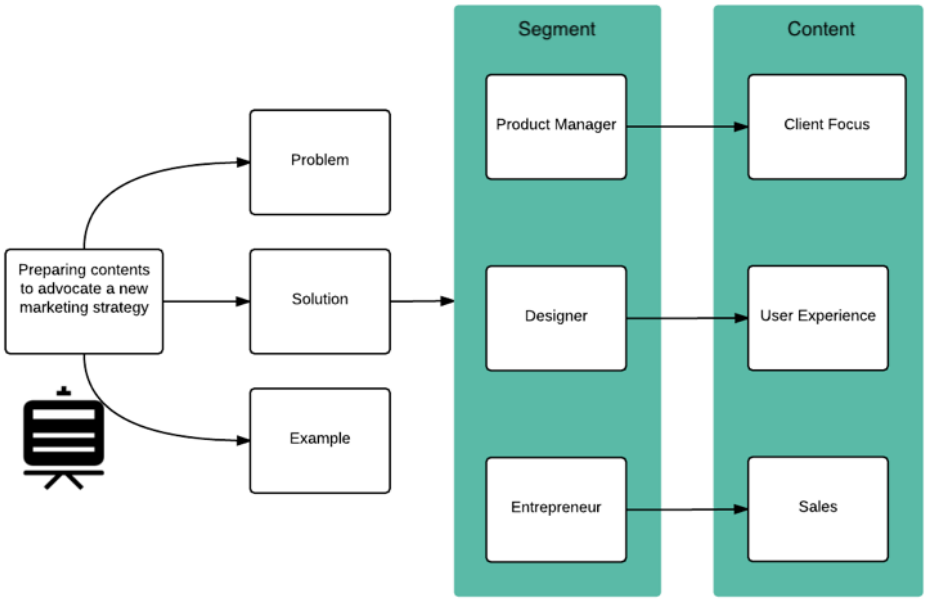


Figure 4.7 – Adding a rule for the solution section

Thus, we adjusted the document template and changed the content to the one illustrated in Figure 4.8. The new solution section contained content related to design and user experience,

because Kirk's user profile reflects his familiarity with the design domain. Since the rule segment is based on user expertise, the solution section content is different for each user.

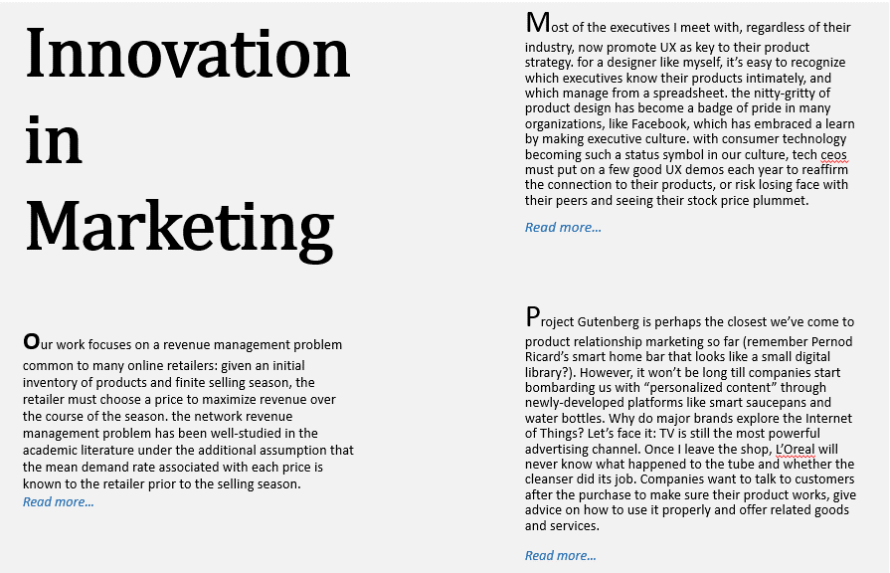


Figure 4.8 – A different example for PCG v1 presentation

4.3.3 PCG v1 Discussion

We designed the first prototype to capture the essentials for picking up content based on user profiles. Implementation of the first prototype covered some PCG core components like the user-profile builder (Section 3.4) and the information assembler (Section 3.5). During the implementation phase, we observed issues related to user experience and database performance that could have potentially impacted participant engagement in the final experiment.

Heuristic evolution is an ideal approach to identify usability problems [108]. Since it does not require the involvement of real users, it became a rational choice to inspect and evaluate PCGv1 design by the researcher. We also used Cognitive walkthrough [109] by asking a small number of individuals to sign in to PCGv1 and share their thoughts and experience. Below are what we identified using Heuristic evaluation and Cognitive evaluation walkthrough:

- The content's text should be easy on users' eyes so they can scan the content quickly. Using short paragraphs, strong titles/headers, bold text, bullet points, and quotes can make the content scannable.
- The content should be cohesive and flow smoothly from one part to the next. Adopting a well-designed timeline to present the content might keep participants engaged.
- The content should be prepared and presented in a timely fashion to keep participants engaged. Refactoring code, caching the result, optimizing database queries, and using background tasks will improve performance.

4.4 PCG v2

The PCG v1 scope included capturing the essential user data that was needed to query content based on user profiles and to assemble the presentation. For PCG v2, we implemented the individualization layer components and improved the system to enhance participant engagement.

4.4.1 Improvement Overview

To implement the individualization components and to improve participant engagement, we altered the PCG system in the following ways:

- We made the content easier to read by decreasing the number of sentences for each section in the PCG content from six to three and adjusted contrast between characters and background.[1]

[1] A contrast ratio of 3:1 is the minimum level recommended by ISO-9241-3 and ANSI-HFES-100-1988 for standard text and vision. According to WCAG 2.0, this means the relative luminance of text colour differs from the surrounding text's by a contrast ratio of at least 3:1.

- We structured the content more clearly, adopting an inverted pyramid writing style [110] wherein content started with an overview of main points so participants could better relate to subsidiary points. Our intent for the PCG content was to convey subsequent narrative points. Brehmer et al. [94], using a survey of 263 timelines, identified a rich set of timeline designs that can be used for narrative storytelling [111]; we adopted design elements from this set to make the persuasive content more coherent.
- We improved performance by doing the following:
    - We reduced the number of HTTP requests to improve system performance when loading PCG content.[1] We followed a variety of best practices to reduce the number of requests, including CSS Sprites, refactoring code, reducing the number of places in the code that make large numbers requests, and combining CSS and JS files.
    - We ran the PCG system as multiple lightweight processes instead of executing it as a monolithic entity. We designated processes for tasks like creating user profiles (Section 3.4), summarizing content (Section 3.5), and building trust-graphs (Section 3.6) that require intensive computation. This helped keep process times short. This approach allowed us to background long-running work and provide indication of work in progress to end users.
    - Using MySQL Tuner script, we produced an overview of the PCG database's performance, enabling us to clean out old, unused tables and to create indexes for faster access. We also reviewed and refactored SQL queries that had been used to handle operations like selecting user profiles and collecting opinionative information from the database.

[1] Browsers rely on HTTP to fetch data from the server that hosts the PCG system. The more HTTP requests browsers make, the slower PCG content will load.

4.4.2 Content Domain

Although, we could have reused the marketing domain for PCG v2, we decided to select a different domain that could be reused for the final experiment as well. We did this because:

- Selecting a different domain allowed us to evaluate PCG under a different context; and
- From a participant-engagement point of view, a theoretical marketing subject may not be the most suitable for the final experiment as interest in the topic may be low.

Climate change was chosen as a suitable domain for the final experiment because many studies have attempted persuade people to take action against climate change, and research has demonstrated that carefully crafted persuasive text can promote both conceptual and attitudinal change [4][112][113]. According to Murphy, "a persuasive text should be structured to promote understanding and the view that the alternative perspective is worth considering" [114]. In addition, changing people's attitudes about socio-scientific issues like climate change has unique challenges [112][5][115]: according to Sakamoto and Goldstone, "it is complex, multidimensional, and requires systems thinking, that is, the ability to think and reason abstractly about systems to appreciate their interactive nature" [116].

Furthermore, climate change presents a challenge in that people have limited knowledge or misconceptions about the topic. For example, individuals may believe the planet is not warming. Misconceptions may also include conflation of weather and climate [115]. For instance, when people are asked for their opinion on climate change, they often refer to their experience and use their recollection of weather temperature fluctuations to assess whether the planet is warming. Thus, we decided that climate change is a suited domain for the final experiment. Despite the effort to choose a proper subject, it is logical to assume that no topic could be equally interesting to all

participants and this could affect the persuasion. We hoped that our choice was suitable overall for the randomly selected participants.

4.4.3 Rule-File Design

Before preparing the content's HTML document template, we needed to design the author XML rule files needed for the PCG v2 prototype and the final experiment. As discussed in Section 3.3, the PCG system relies on a set of XML rule files to assemble persuasive content. To assist us with capturing the basic requirement needed for the rule files, we followed Norwegian psychologist Espen Stoknes's advice on how to prepare persuasive contents. Stoknes [100] lays out a psychological approach for moving society to climate action in *What We Think About When We Try Not To Think About Global Warming* [117]. He argues that if inductive reasoning is not effective to fight climate change, perhaps we need to embrace the irrational human mind:

> "Unthinkingly, the same social experiment has been repeated over and over: Simply give people the information, and then wait and see if the facts trickling into people will persuade them to change their behaviour. The outcome has been consistently underwhelming. But that hasn't held rational people like climate scientists, public servants, and environmentalists back from trying the same experiment on the public again and again–each time with yet more facts and, each time, for some weird reason, expecting a different outcome." [100]

Moreover, Stoknes named localized content with positive sentiment that shows a way to take visible and consistent action to fight climate change as an appropriate strategy. He also argued that convincing an individual in a group to act requires finding another member of that group to have an aligning opinion. The next section covers Stoknes's criteria for persuading people to fight climate change.

4.4.3.1 Personalizing and Localizing the Content

According to Stoknes, climate change is seen as an abstract problem and speaking exclusively about increased carbon emissions only makes it more abstract. People will downplay problems that will manifest in the distant future. Stoknes argues against talking about global effects and says targeting messages to local perceptions will more effectively communicate the dangers of climate change. For instance, after Hurricane Sandy, New Yorkers understood sea-level rise; similarly, Californians now understand long-term drought. As Section 3.1 discusses, the first step of the PoI is attracting audience by presenting relevant content.

4.4.3.2 Presenting Content with a Positive Sentiment

People generally talk about climate change in terms of doomsday scenarios. Though apocalyptic concern is justified by fact, such scenarios discourage people from listening. Instilling feelings of helplessness only overwhelms audiences. Alternatively, discussion of opportunity and solution positively impact persuasion. Stoknes says, "while a message of doom doesn't inspire people to act, a focus on solutions could" [117].

4.4.4 Rule-File Preparation

To maximize PCG content's persuasiveness, we applied the following criteria from Section 4.4.3 to prepare our XML rule file and presentation template:

- Content should be personalized as well as localized, and
- Content should convey a positive sentiment message.

For PCG v2 and the final experiment, we downloaded 100 articles related to climate-change problems and solutions from Yale Environment 360, an online magazine published by *Yale*

School of Forestry & Environmental Studies and Yale University.[1] Each article contained 1000 to 2000 words. In addition, we connected the PCG to *The Guardian* API so that the PCG could collect articles without having to download and host them separately.

As discussed, authors had to create a rule file and document template. Once a user logged in, the system generated persuasive content using the rule file and appended that to the document template. The document template for the second prototype and the final experiment consisted of following sections (as Table 4.3 demonstrates).

Section Title	**Definition**
A common definition	A shared chapter among all participants containing a general definition for climate change
What's on NEWS?	Latest and most relevant news on climate change to the participant
Effect of climate change on our lives	A personalized chapter describing a problem associated with climate change and containing content from a scientific article
A successful example for how to fight climate change	A positive solution on how to fight back the climate change and containing content from a scientific article

Table 4.3 – Presentation-template-chapter definition

Table 4.4 – PCG v2 HTML template includes four segments (refugee crisis, natural-resource contamination, agricultural development, and lifestyle behaviour) assigned to the

[1] It provides opinion, analysis, reporting, and debate on global environmental issues and features original articles by scientists, journalists, environmentalists, and academics.

problem section. Each segment is directly associated with multiple contents. For example, to people that fit the “lifestyle behaviour” segment, the author recommends content related to “consumption,” “car,” and “population.” Since people may be associated with all given segments, the system applies a weight factor to balance the coverage for each content query (weight = 25).

As Stoknes recommends [117], by using generic rule polarity attribute we aimed to display a message with positive sentiment and so, for the solution section, filtered out any articles with negative sentiment.

Chapter 1: A common definition

```
<div class="cd-timeline-block">
  <div class="cd-timeline-img cd-picture">
    <img src="img/cd-icon-picture.svg" alt="Picture">
  </div> <!-- cd-timeline-img -->

  <div class="cd-timeline-content">
    <div id="common_def">
      <h2>OUR GOALS AND OBJECTIVES</h2>
      <p>There's a lot of information floating around about climate change.
        Most people know it has something to do with industrial pollution.
        In a nutshell, climate change occurs when long-term weather patterns are altered - for example,
        through human activity. Global warming is one measure of climate change, and is a rise in the average
        global temperature. Here we would like to tailer a personalized message approach for persuading people to climate action.</p>
    </div>
    <span class="cd-date">
      Chapter 1</br>A common definition
      <div class="cd-timeline-quote">
        <div class="cd-quote-img cd-quote">
          <img src="img/cd-icon-quote.svg" alt="Quote">
        </div> <!-- cd-timeline-img -->
        <!-- will be updated by npg -->
        <div id="quote_section">
          <!-- will be updated by npg -->
        </div>
      </div>
    </span>
  </div> <!-- cd-timeline-content -->
</div> <!-- cd-timeline-block -->
```

Chapter 2: What’s on news

```
<div class="cd-timeline-block">
  <div class="cd-timeline-img cd-picture">
    <img src="img/cd-icon-picture.svg" alt="Picture">
  </div> <!-- cd-timeline-img -->

  <div class="cd-timeline-content">
    <div id="news_section">
      <!-- will be updated by npg -->
    </div>
    <span class="cd-date">Chapter 2</br>What's on news?</span>
  </div> <!-- cd-timeline-content -->
</div> <!-- cd-timeline-block -->
```

Chapter 3: Effect of climate change on our life

```html
<div class="cd-timeline-block">
  <div class="cd-timeline-img cd-picture">
    <img src="img/cd-icon-picture.svg" alt="Picture">
  </div> <!-- cd-timeline-img -->

  <div class="cd-timeline-content">
    <div id="problem_example">
    <!-- will be updated by npg -->
    </div>
    <span class="cd-date">Chapter 3</br>Effect of climate change on our life</span>
  </div> <!-- cd-timeline-content -->
</div> <!-- cd-timeline-block -->
```

Chapter 4: A successful example to fight climate change

```html
<div class="cd-timeline-block">
  <div class="cd-timeline-img cd-picture">
    <img src="img/cd-icon-picture.svg" alt="Picture">
  </div> <!-- cd-timeline-img -->

  <div class="cd-timeline-content">
    <div id="solution_example">
      <!-- will be updated by npg -->
    </div>
    <span class="cd-date">Chapter 4</br>A successful example to fight climate change</span>
  </div> <!-- cd-timeline-content -->
</div> <!-- cd-timeline-block -->
```

Table 4.4 – PCG v2 HTML template

4.4.5 PCG v2 Processing Pipeline

This version's first process was identical to PCG v1's. Once a user logged into the system using their Twitter account, we started building their user profile by collecting their tweets and processing the data. As mentioned in section 4.3.2, the data processing task consists of extracting keywords and identifying the keywords' hyponymy (domain). As Section 3.4.2 discusses, we used a hierarchical structure to adjust the retrieved content's complexity level. We calculated the content's readability indices [53][96][54] to estimate comprehension level in a given domain.

Keywords given in user profiles demonstrated users' respective topic-relation factors. When the system was building the query, it looked for matching rule segments in the user profile`s keyword collection. The segment with the highest ratio match was picked for the query. If the segment failed to match any of the user's keywords, the system looked for matching rule goals in the user profile as a fallback. Once the query was completed, the system passed the query to Solr so that the most relevant documents from the knowledge repository could be retrieved.

To identity a quote on climate change from a trustable source, the system first generated a user-trust graph. As Section 3.6 discusses, we relied on users' Twitter interactions (likes, mentions, and retweets) to build trust graphs. The system sorted graph results using a descending trust value; the system used the first fifty trustable sources to generate the list. As a fallback method, we created a second list from all the profiles that the user followed. The system passed both lists to Solr so Solr could retrieve the most relevant quotes from the knowledge repository.

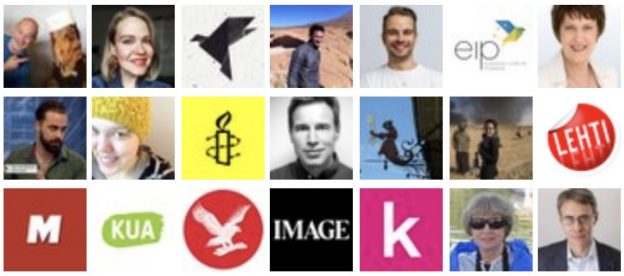
As Table 4.4 demonstrates, the presentation template was designed to follow a narrative-timeline design meant to improve user focus on content. Intuitively, lack of focus negatively impacts participant comprehension and engagement (Section 4.4.1). A narrative timeline enables users to only focus on one area at a time. The presentation started with the common definition and a trustable quote. As the user scrolled down into the content, new chapters appeared.

We used couple of users to test PCG v2 before the final experiment. To demonstrate the differences between generated content, we list the results for both test users below:

Popular Keywords

crisis syrian nyt social miten society article wars says conference rauhanprosessinytimessolepolitical actors iltaa booming mik planning role dialogue media syyrian possibilities israel juttuni oltiin maai asioiden response kundays eilen people problem build siit syria cohesion mutta kyttjlt donors conflict help peace security support kiitos empower local funding interesting keskustelem assa

Trustable List



Popular Keywords

design work discussion time future amendment automated privacycars project recognized connected highly webinar governmentnews austin look scrutiny city texas regulation vehicles ditch ready lawsuit quickly vehiclelearn public reached data 9th member proprietary car changing ive drivers transportation statessays 2018 colleagues google detection european help interesting 4th years behaviorlyft law tesla

Trustable List

Table 4.5 – PCG v2–Test-Users Description

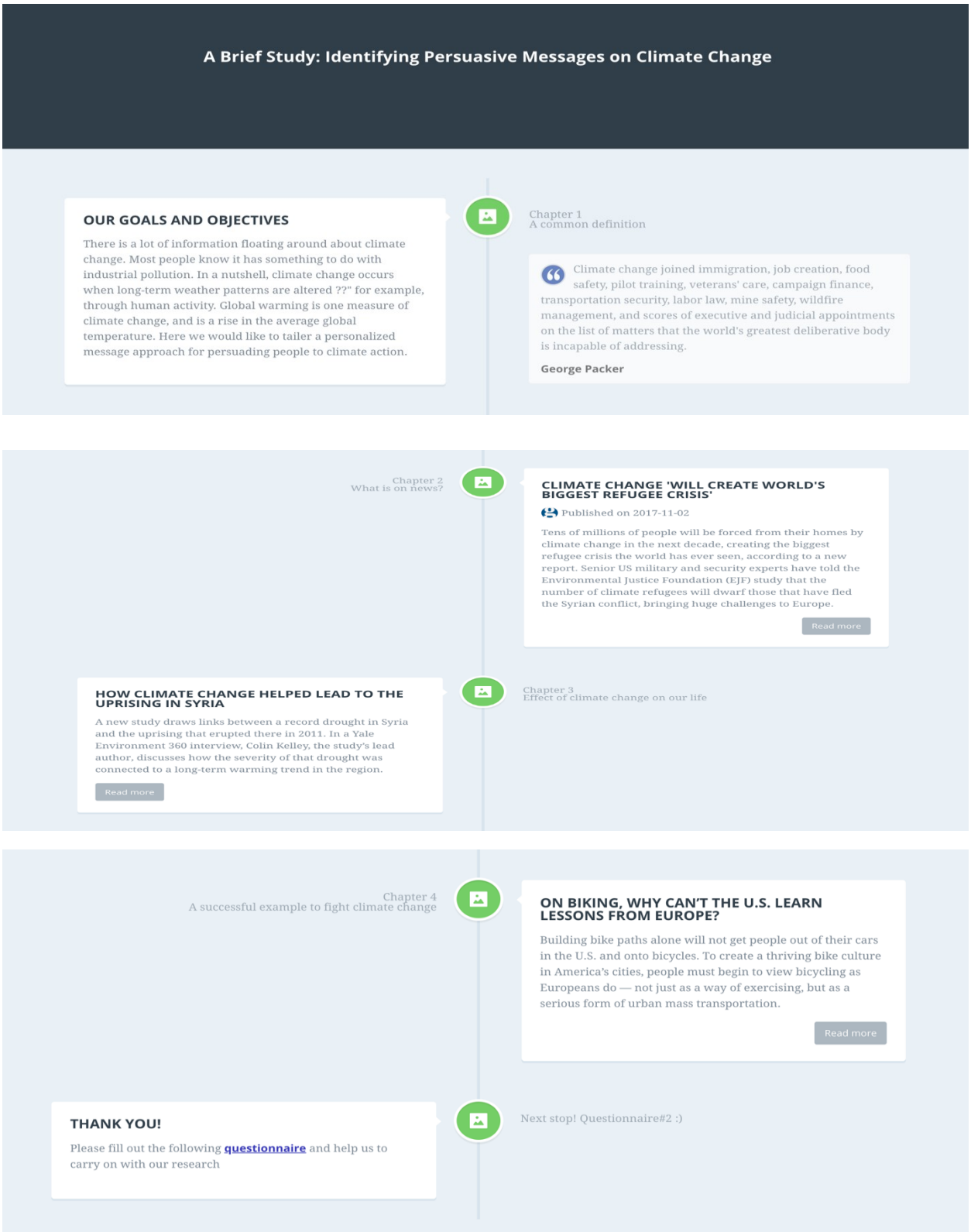


Figure 4.9 – @LeaPakkanen presentation

Figure 4.10 – @SLGALLUN presentation

4.5 User-Trust-Graph Pilot Study

As Section 3.6.2 discusses, we introduced the user-trust graph to help collect supporting opinions from readers' trustworthy sources as part of the individualization layer. Since we designed this component from scratch, we decided to measure its design effectiveness separately. To evaluate the graph's effectiveness, we performed a pilot study. In this pilot study, the system built a trust graph by collecting and monitoring Twitter activities of ten users who might also have followed and interacted with each other on Twitter. To simplify the sentiment analysis, we decided to only use English tweets. Due to the Twitter API limitation, we could only collect a maximum of 500 tweets per user. For each user, the model returned a top-ten list of trustworthy users. Users were asked to evaluate the generated result by adjusting the list. We repeated the same experiment using TwitterRank [118], which correlates trust with tweet popularity. For both models, we applied the Bobadilla formula [119] to calculate the error measurement of the predicted trust strength for each user. Figure 4.11 shows the differences between the two approaches.

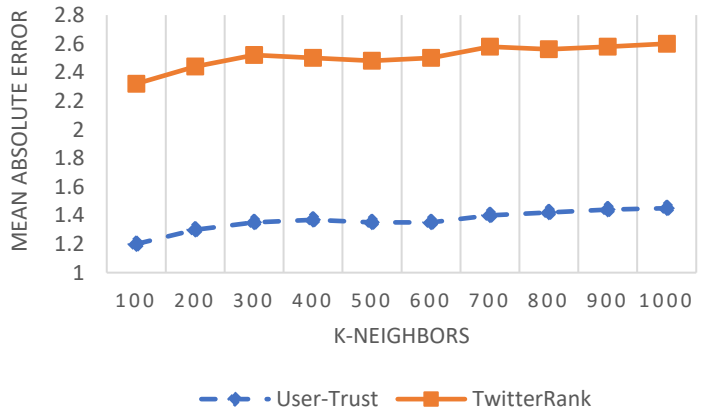


Figure 4.11 – Error measurement of the predicted trust strength

The proposed model (user-trust graph) has a lower error rate than TwitterRank. In some places, we noticed our model's predicted value was 40% to 45% more accurate than TwitterRank. Although results are promising, the evaluated sample is small and only adequate for a pilot study.

Chapter 5.

EVALUATION METHODOLOGY

5.1 METHODOLOGY OVERVIEW

We evaluated our system to determine whether the generated content was personalized and persuasive. We used YAC parameters as our evaluation criteria. According to YAC [18], to maximize the chance of persuading a user to take action or change opinion, first, their attention must be gained and the content's comprehension level must be adjusted so the user can read and understand the message. This will make the content appropriate for that user. As such, we considered an artifact (the content) personalized if:

- The artifact attracted reader's attention, and
- The artifact was comprehensible to the reader.

YAC defines the persuasion cycle as complete when the persuadee accepts and retains the presented information. Thus, to determine whether the content was persuasive, we evaluated if:

- The reader accepts the persuasive text, and
- The reader remembers the persuasive text.

The study used a pre-test, post-test experimental design with two treatments. Recruited subjects were assigned to receive either content generated by the PCG system or generic expert-generated content. Subjects responded to questionnaires prior to the application of the treatment, immediately after the treatment, and two weeks after the treatment. Details of the study are presented in the following sections.

5.2 HYPOTHESES

Our study operationalizes personalization and persuasion as follows:

- That content is personalized if it attracts reader attention and is comprehensible to the reader.
- Content is persuasive if the reader accepts the content and retains it.

This operationalization maps directly to layers in the PoI model. Specifically, 'attention' corresponds to the 'segmentation' layer; 'comprehensible' corresponds to the 'comprehension' layer; and 'acceptance' corresponds to the 'individualization' layer. For each construct (personalization and persuasion), we defined the following series of hypotheses:

1. Personalization:

<u>The generated content attracts reader's attention</u>

As Vesanen and, later, Zhou et al. state, "Interesting," "relevant," and "enjoyable" are parameters that increase the likelihood of gaining reader attention [77][14]; we, therefore, proposed the following hypotheses:

H1.1 The readers find the PCG content significantly more interesting than the generic content.

H1.2 The readers find the PCG content significantly more enjoyable than the generic content.

H1.3 The readers find the PCG content significantly more relevant than the generic content.

<u>The generated content is comprehensible to the reader</u>

The criteria “easy to follow,” “less complex,” and “easy to understand” have been identified by researchers as a means to assess content comprehensibility [81][52]; therefore, we proposed the following hypotheses:

H1.4 The readers find the PCG content significantly easier to follow than the generic content.

H1.5 The readers find the PCG content significantly less complex than the generic content.

H1.6 The readers find the PCG content significantly easier to understand than the generic content.

2. Persuasion:

<u>The reader accepts the persuasive text</u>

Parameters like “trustworthy,” “accurate,” “authentic,” and “believable” have been commonly accepted as a means to increase the likelihood of reader acceptance of a content [85][86]; we, therefore, defined the following hypotheses:

H2.1 The readers find the PCG content significantly more trustworthy than the generic content.

H2.2 The readers find the PCG content significantly more accurate than the generic content.

H2.3 The readers find the PCG content significantly more authentic than the generic content.

H2.4 The readers find the PCG content significantly more believable than the generic content.

Furthermore, as part of persuasion, we assess the impact of the treatment on influencing participants' attitude towards climate change.

H2.5 There is a positive significant change in PCG-group-participant attitudes toward the study's subject (climate change) from the pre-treatment to post-treatment responses.

<u>The reader retains the persuasive text</u>

In the context of our study, retention refers to remembering the effect of the treatment. Some scholars define it as the persistence to perform a learned knowledge or behavior [120][121]. While individuals may remember the content due to factors other than the persuasiveness of content (such as their personality traits and memory), the overall statistical result can be a measure of persuasiveness as it includes a random set of participants. If the PCG content had no effect on changing participants' attitudes towards climate change immediately after the treatment, then retention was not assessed. Prior research has used a two week delay [122][123][124] to assess retention of a message.

H2.6 There is no significant change in PCG-group-participant attitudes toward the study's subject (climate change) between their post-treatment responses to their two-weeks post-treatment responses.

5.3 DATA COLLECTION

We set the minimum requirements for participation as:

- Having an age of 18+,
- Having a valid Twitter account, and
- Being a fluent English speaker.

We did not want to recruit from youth since, intuitively, they exhibit different behaviours than adults. We required users to have a Twitter account since the PCG builds user profiles using participant tweets. Speaking fluent English is also important because PCG contents were to be in English, and, according to YAC, comprehension is directly related to persuasion.

We designed the final evaluation to have two phases (short term and long term) and two participant groups (one exposed to the proposed system and one using generic persuasive content written by a professional writer). We designed three questionnaires to measure participants' opinions and feedback during the experiment at three different points in time.

1. Pre-treatment: assessed participants' current views on the experiment topic before reading the persuasive (PCG or generic) content
2. Post-treatment: assessed participants' opinions on content personalization and persuasion immediately after reading the persuasive (PCG or generic) content
3. Two-weeks post-treatment: assessed participants' respective opinions on the experiment topic two weeks after the experiment. Two week intervals have been used in other similar projects [121][123][124] to assess retention of a message.

The questionnaires are shown in Appendices I, II, and III, respectively. To help us investigate the evaluation objectives, the questionnaires covered the following types of questions:

- Attitude towards Climate Change (ACC) questions, in which participants shared their attitude toward the generic and PCG content topic through three different timeframes, and
- Direct-Assessment (DA) questions, in which participants were asked to share their respective opinions on content personalization and persuasiveness.

The ACC questions were:

1. I believe that there is evidence of global climate change.
2. I believe that things I do, have no effect on the quality of the environment.
3. I believe most of the concerns about environmental problems have been exaggerated.

The DA questions were:

1. After reading the provided text, I am persuaded to take action against climate change.
2. I found the content interesting.
3. I found the content enjoyable.
4. I found the content relevant to me.
5. I found the content clear and easy to follow.
6. I found the content's word and sentence structure complex.
7. I found the content easy to understand and comprehend.
8. I found the content trustworthy.
9. I found the content accurate.
10. I found the content authentic.
11. I found the content believable.

Respondents assessed questions using a 5-point Likert scale. Using a symmetric agree-to-disagree scale for a series of statements, respondents had the option to express the intensity of their attitude toward the given topic or question. Table 5.1 maps the above questions to their related hypotheses.

Question	Hypotheses
ACC Question 1	H2.5 and H2.6
ACC Question 2	H2.5 and H2.6
ACC Question 3	H2.5 and H2.6
DA Question 1	H2.5 and H2.6
DA Question 2	H1.1
DA Question 3	H1.2
DA Question 4	H1.3
DA Question 5	H1.4
DA Question 6	H1.5
DA Question 7	H1.6
DA Question 8	H2.1
DA Question 9	H2.2
DA Question 10	H2.3
DA Question 11	H2.4

Table 5.1 - ACC and DA questions and their related hypotheses

5.4 DATA ANALYSIS

We started the analyses by calculating the mean responses for Attitude toward Climate Change (ACC) questions during different timeframes (pre-treatment, post-treatment, and two-weeks post-treatment). This was the first step toward analyzing participant changes of attitude toward climate change. To conduct general analysis on distribution and reliability of the calculated mean value, we needed to calculate standard-deviation error bars for the dataset. A small standard-deviation bar indicates that data is clumped around the mean, and a larger standard deviation indicates that the distribution of the data is more variable from the mean. A larger standard-deviation bar indicates the mean value is less reliable and that the smaller bar can be interpreted as a sign of accuracy in the dataset [125].

Calculating skewness and kurtosis for the dataset is also important. Skewness helps provide a sense of data distribution. For instance, a negative skewness value indicates data skews left—that is, the left tail is relatively longer than right tail. Kurtosis helps identify the likelihood of outliers and extreme values.

As we obtained results from Likert-scale data (ranges 1-5), with a limited number of responses, we could not assume our data was parametric, leading us to analyze them via nonparametric statistical methods. We used the nonparametric Mann-Whitney test (alpha = 0.05) to assess whether two sets of data were significantly different from each other for ACC and DA questions' responses. This was a reliable way to examine differences between participants' ACC questions responses within pre-treatment, post-treatment, two weeks post-treatment timeframes. It also let us determine whether persuasion was resultant from either the generic text prepared by an expert or our system's text.

5.5 EXPERIMENT SETUP

5.5.1 Recruitment

Since the experiment involved human participants, the research team applied and received Carleton University Research Ethics Board-B (CUREB-B) ethics clearance for the experiment (Appendix VI).

To optimize the data-collection process, we hosted the experiment on a public server. By visiting the experiment website, participants were randomly assigned either generic or PCG text. The recruitment script (call for research participation) was posted on bulletin boards and designated areas for flyers and ads at Carleton University in accordance with Carleton's posting policy. The recruitment script was also posted on affiliated Carleton University social media channels, such as the Carleton Research Participants Facebook group. An URL to the experiment landing page was included in the recruitment script. By clicking the link, candidates could read the consent form and register for the study.

5.5.2 Experiment-Material Design

As mentioned before, we designed three questionnaires to measure participants' respective opinions and feedback during the experiment during three different stages (timeframes):

- Stage 1: Pre-treatment feedback,
- Stage 2: Immediate post-treatment feedback, and
- Stage 3: Two weeks post-treatment feedback

We followed two main rules to maximize questionnaire effectiveness:

- As DeVellis recommends, we attempted to create questionnaire items that were strongly worded, unambiguous, and declarative statements without jargon [126].

- The participants rated each item on a Likert scale from one (strongly disagree) to five (strongly agree). We intentionally unified the scale across all three questionnaires to simplify data analysis.

Questionnaire 1 (ACC questions in Section 5.3, also listed in Appendix I) measured initial attitudes about climate change. It provided a benchmark of participants' attitudes before being exposed to the treatment. It covered how much participants believed in climate change and whether they thought human activity has been the driving force behind climate change.

To measure and assess the system's impact from the PoI standpoint of personalizing and individualizing content, we included a set of questions in Questionnaire 2 (DA questions in Section 5.3, also listed in Appendix II)

Questionnaire 3 (Appendix III) used a similar structure as Questionnaire 1 and was designed to collect participant attitudes about climate change two weeks after the experiment. This questionnaire was used in assessing the retention of the persuasive text. It used the ACC questions in Section 5.3 plus the first DA question to evaluate persuasion.

In addition to the questionnaires, we selected a generic persuasive content from *National Geography*[1] to compare with the PCG content. The generic content was 598 words in length and had a Flesch-Kincaid readability score of 10.5. A Flesch-Kincaid readability score of 10.5 means the content is easy to read and easily understood by an average eleven-year-old student. Based on the definition of persuasive text by Chambliss and Gardner [127], we concluded the text was persuasive. According to Chambliss and Gardner, ''A message is persuasive, if it is structured to counter the current beliefs of a typical reader as well as to present new ones by capitalizing on a reader's existing knowledge and beliefs." The text was structured to demonstrate solid evidence

[1] https://www.nationalgeographic.com/environment/global-warming/global-warming-solutions/

for global climate change and to support that humans are contributing to the trend. Figure 5.1 is a screenshot of how the generic content was incorporated into the experiment's web application and of the actual content.

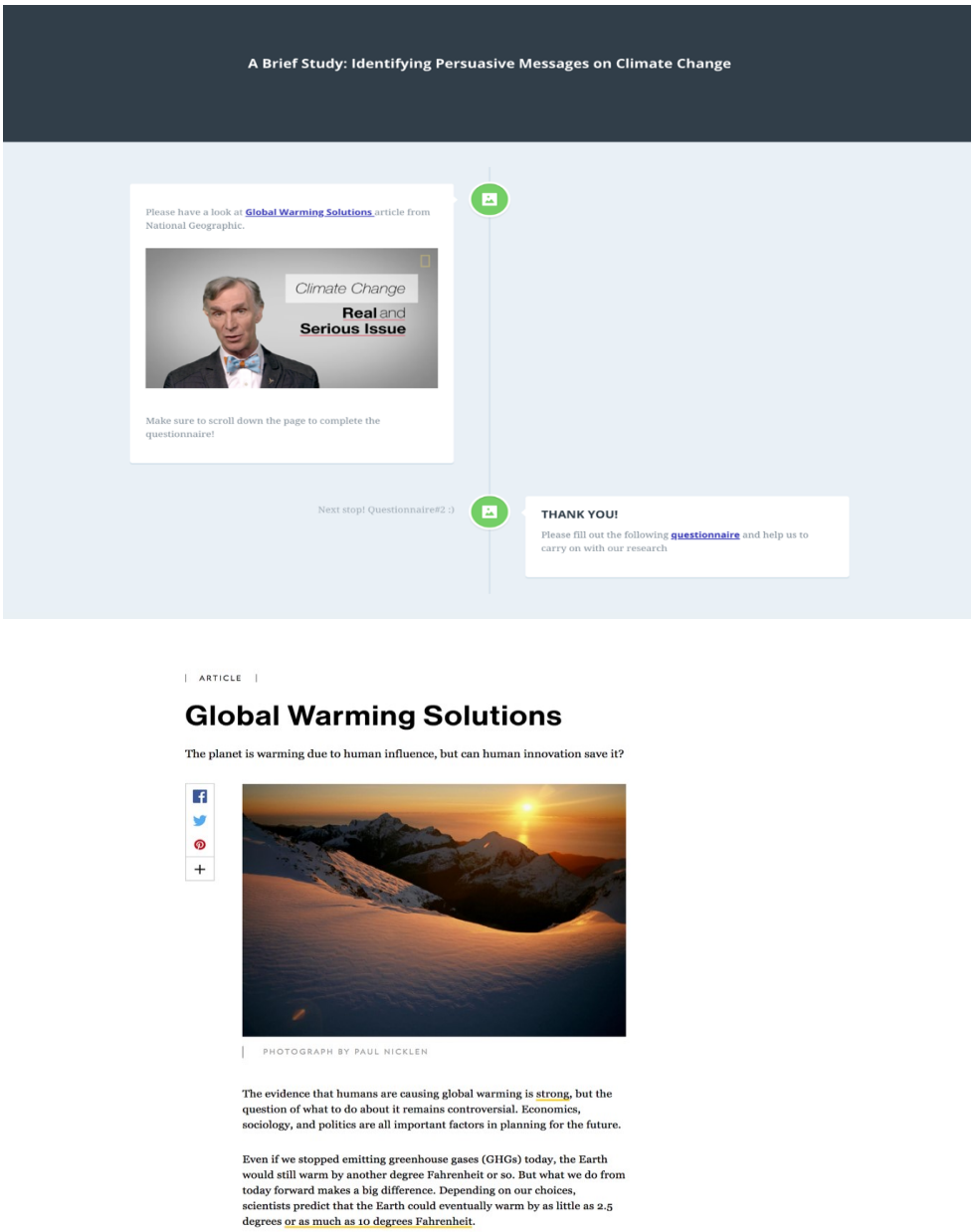


Figure 5.1 – Generic content on climate change

5.5.3 Experiment Flow

To simplify the data-collection process, we built a web application to launch the experiment on a public server. The experiment landing page contained a consent form and a short online questionnaire (Questionnaire 1) regarding participant opinions and attitudes toward climate change. Participants were asked first to read the consent form (Appendix VI) before completing Questionnaire 1. Figure 5.2 shows this part of the experiment.

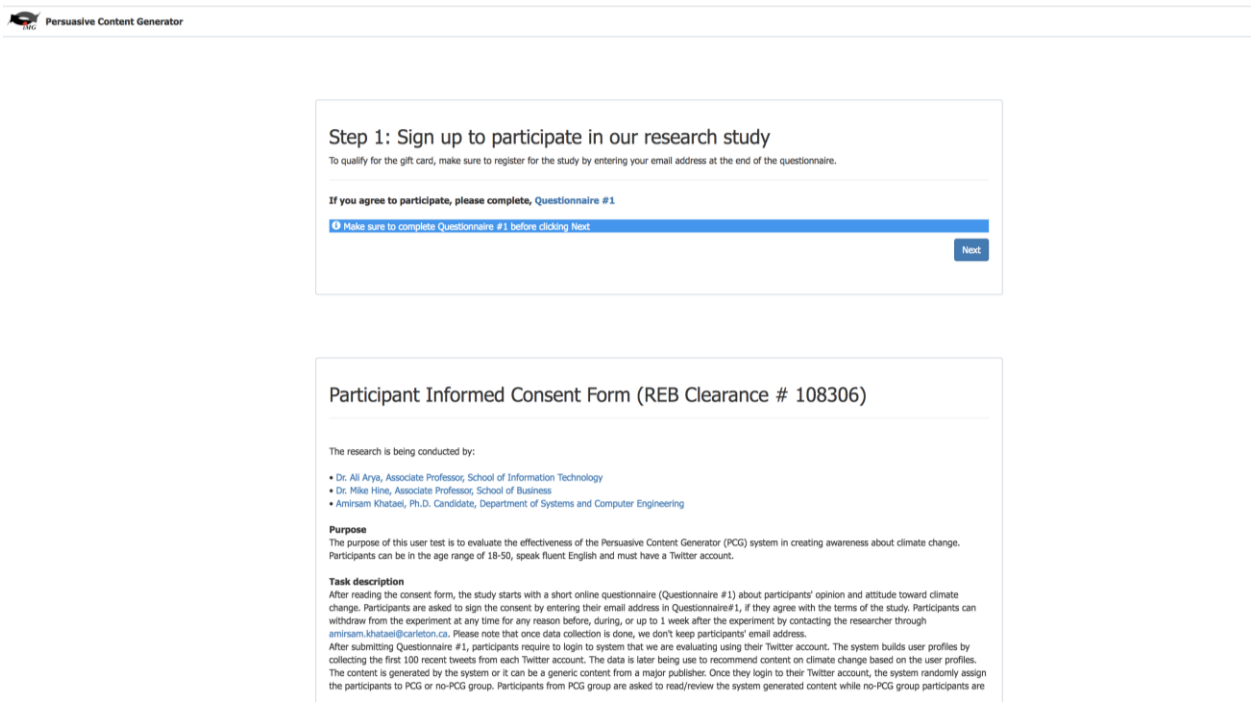


Figure 5.2 – Experiment-web-application landing page

After submitting the questionnaire, participants were asked to log into the PCG system using their Twitter account. The system built user profiles by collecting the first 500 recent tweets from each Twitter account. The data was later used to recommend and personalize content on climate change based on the user profiles. The climate-change content was either generated by the system or generic content from a major publisher. After reading the content, the system asked

participants to complete a second questionnaire (Questionnaire 2). Questionnaire 2 focused on content structure, user opinions, and attitudes toward climate change. In addition, the second questionnaire asked participants to rate content attractiveness, comprehensibility, and creditability.

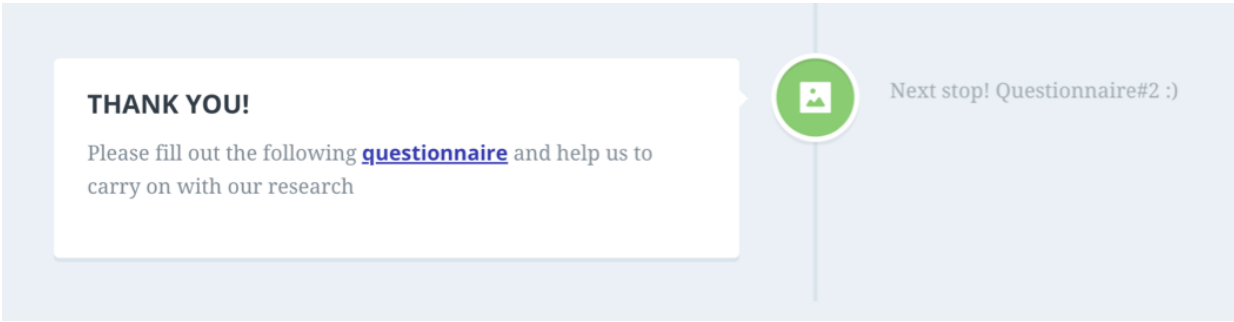


Figure 5.3 – Questionnaire 2 incorporated into the experiment flow

The first and second questionnaires covered the same three questions related to climate change (ACC questions). By comparing an individual's responses, we intended to measure the content's immediate persuasive impact (i.e., how it changed participant attitude).

The process of completing Questionnaire 1, reading the climate change content, and answering Questionnaire 2 was designed to take approximately fifteen minutes. However, we enforced no hard limit on how long participants spent on the experiment.

Participants were also asked to respond to a follow-up questionnaire two weeks from the date of the study. The follow-up questionnaire (Questionnaire 3) was sent to participants via email.

Chapter 6.

EXPERIMENT RESULT

6.1 RESULT OVERVIEW

This chapter reports on results that have been obtained from forty-eight participants who completed the entire experiment. Twenty-five participants were given the generic content, and twenty-three were given the PCG content. In our analyses, we refer to these groups as the generic and PCG groups. The upcoming subsections demonstrate our analyses on the following:

- General analysis: reports on participants' general opinions toward climate change during different phases of the experiment by reporting on ACC question responses.
- PoI analysis: reports on participants' assessments regarding content from a PoI standpoint (segmentation, comprehension, and individualization layers) by covering DA questions 2-11.
- Attitude Change analysis: Reports on amount of change in participants' attitude toward climate change before and after each timeframe within each treatment level using ACC question responses. It also reports on difference in change in participants' attitude toward climate change between the two participant groups using ACC question responses.
- Persuasion analysis: reports on participants' self-assessments on content persuasion effectiveness by covering DA persuasion question responses,

We defined an outlier as a data point 1.5 times the interquartile range above the upper quartile and bellow the lower quartile. Although we display the outliers in boxplots, all outliers

have been manually removed from our calculations. This dropped the sample size for each group by one (22 PCG group and 24 generic group). We performed all the calculations with and without the outliers and saw no significant difference in findings.

6.2 General Analysis

We started the analyses of participants' changes of attitude toward climate change by calculating the mean response to Attitude toward Climate Change (ACC) questions within different timeframes (pre-treatment, post-treatment, and two-weeks after treatment) for the two treatments. The calculation results are demonstrated in box plots (Figure 6.1 - Figure 6.6). In all the figures, × represents the mean value and the dashed line represents the median. The graphs show that both groups (generic and PCG) expressed similar responses before being exposed to the persuasive contents but did not share attitudes after being exposed to the contents.

To conduct a general analysis on the distribution of data and the calculated mean value, we included standard-deviation error bars in the bar charts. The small standard-deviation bar is interpreted as data clumped around the mean, and the larger standard deviation is a sign of data spread far from the mean.

According to responses to the first two ACC questions across both the PCG and generic treatment groups (Figure 6.1 -Figure 6.4), there was no significant change in participant opinions toward climate change. This is not the case for the third ACC question where the calculated post-treatment mean value shows a 32.5% drop from the calculated pre-treatment mean value for PCG-group participants. Alternatively, for the same question, the calculated post-treatment mean value shows only a negligible increase in the calculated pre-treatment mean value for generic-group participants. Table 6.1 and Table 6.2 show statistical analysis for three ACC questions

across PCG and generic treatment groups. The purpose of both tables is to describe the sample data.

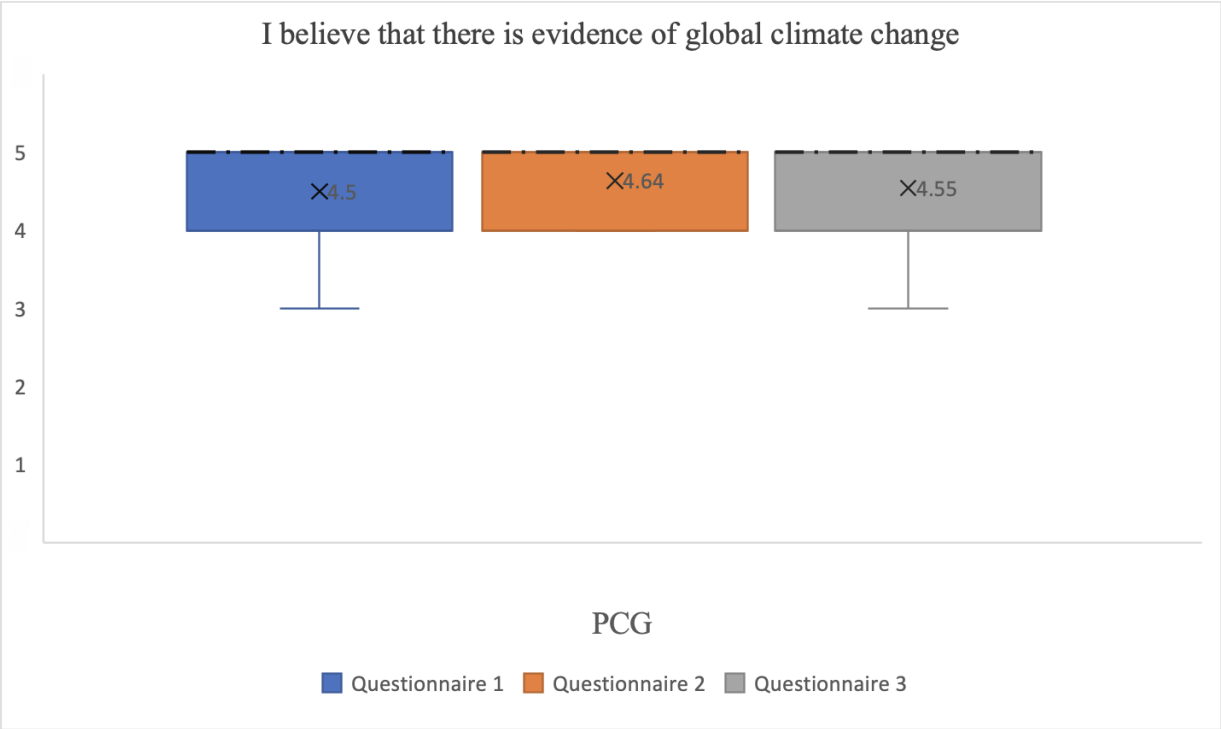


Figure 6.1 – Distribution of ACC Question 1 responses for the PCG group

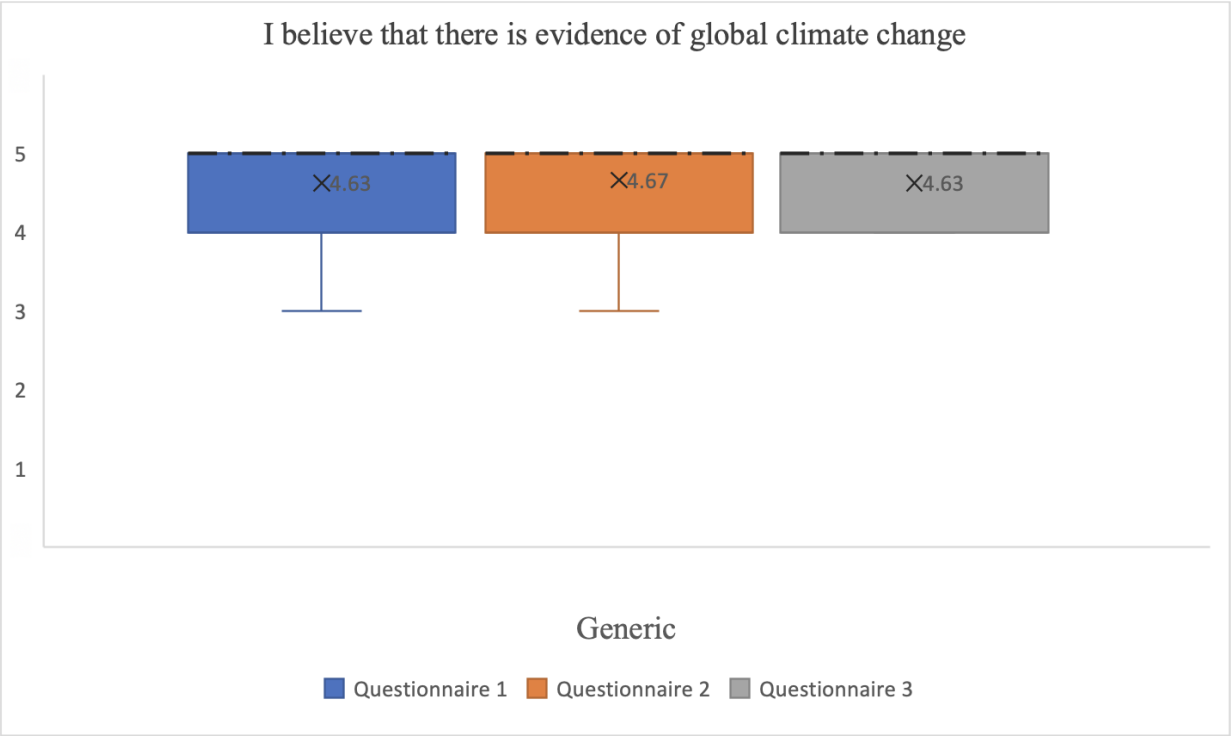


Figure 6.2 – Distribution of ACC Question 1 responses for the generic group

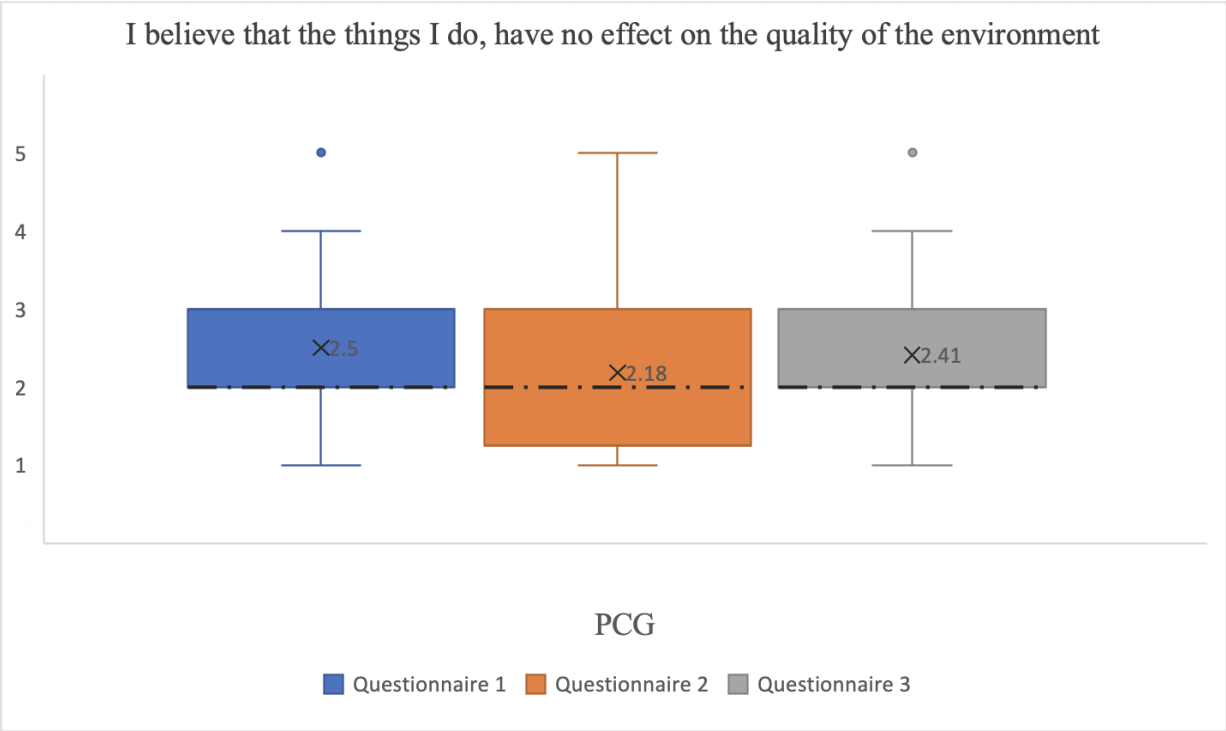


Figure 6.3 – Distribution of ACC Question 2 responses for the PCG group

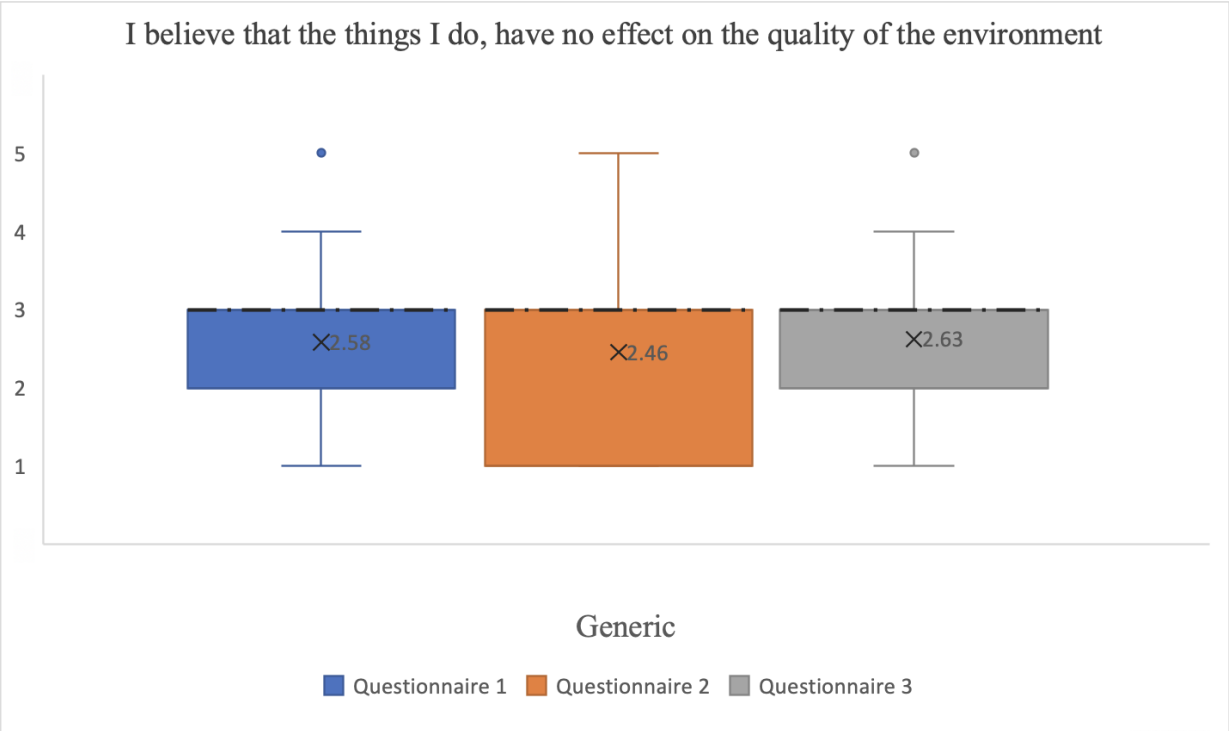


Figure 6.4 – Distribution of ACC Question 2 responses for the generic group

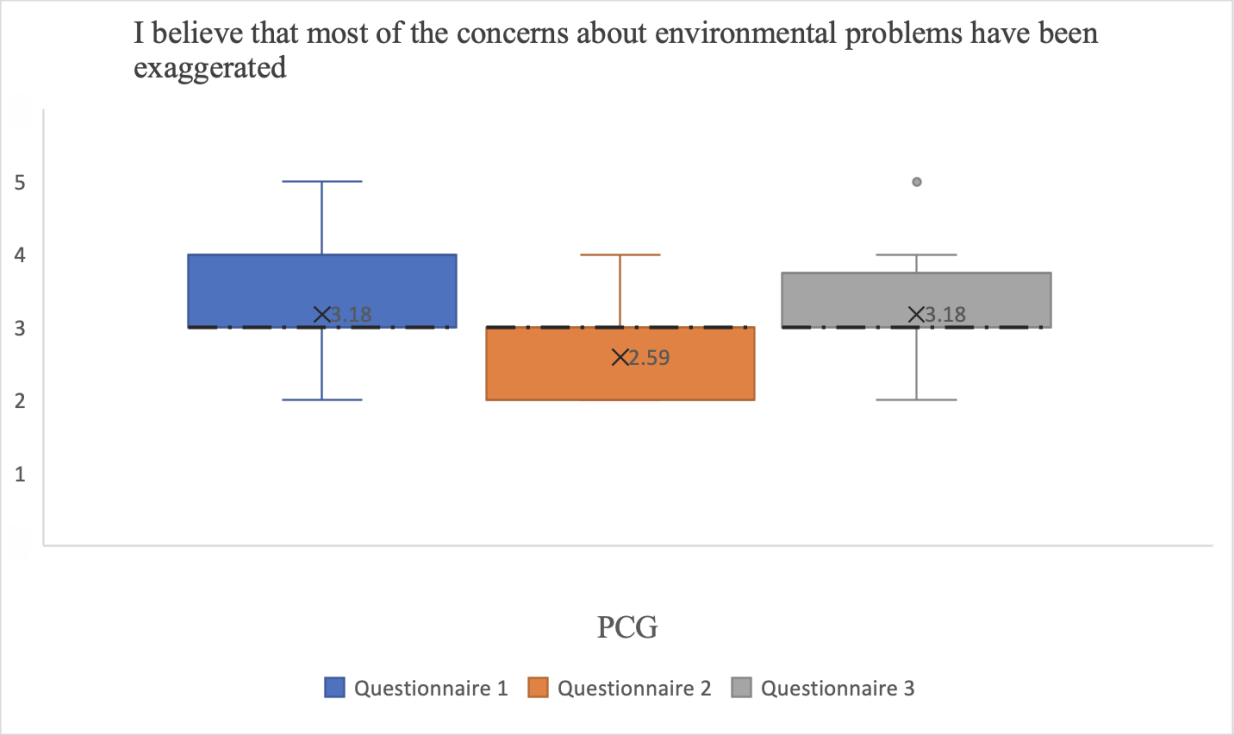


Figure 6.5 – Distribution of ACC Question 3 responses for the PCG group

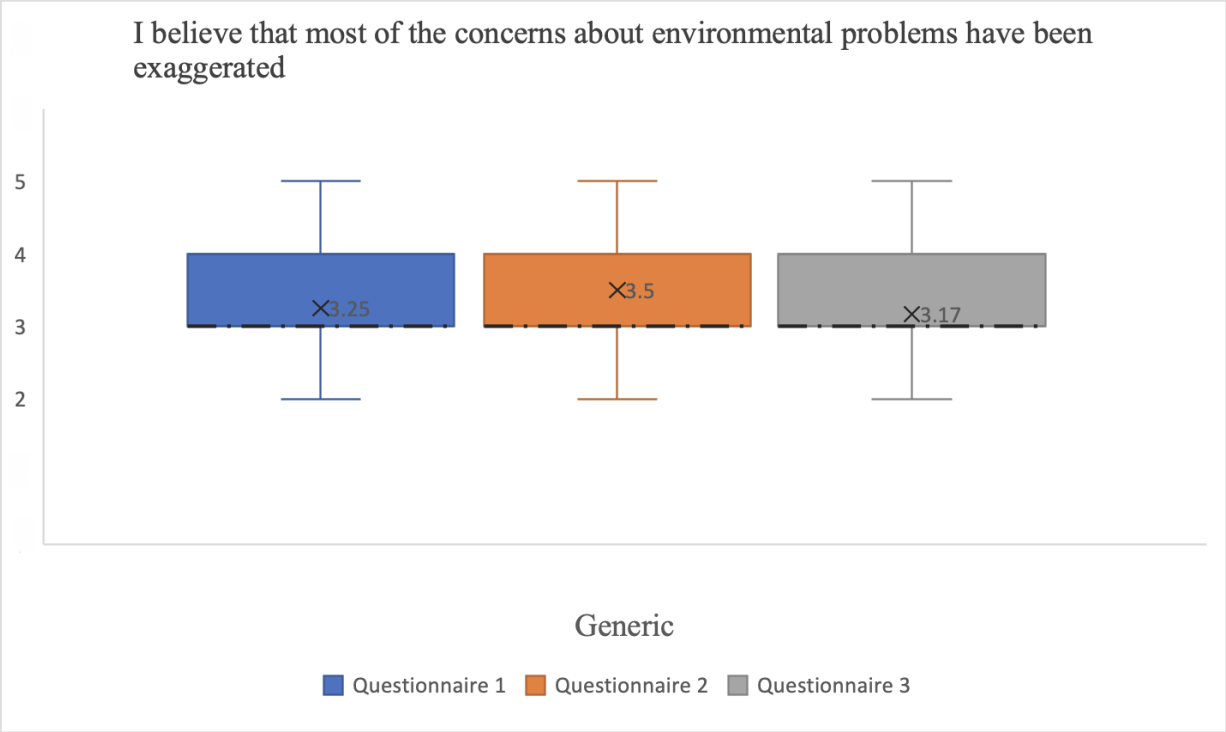


Figure 6.6 – Distribution of ACC Question 3 responses for the generic group

Category	Mean	Variance	Skewness	Kurtosis	Confidence Level (95%)
Pre-treatment					
ACC Question 1	4.500	0.357	-0.736	-0.312	0.250
ACC Question 2	2.500	1.119	1.195	1.297	0.442
ACC Question 3	3.182	0.537	0.495	0.707	0.306
Post-treatment					
ACC Question 1	4.636	0.242	-0.609	-1.802	0.206
ACC Question 2	2.182	1.013	0.836	1.334	0.421
CC Question 3	2.591	0.348	0.379	-0.626	0.247
Two weeks post-treatment					
ACC Question 1	4.545	0.355	-0.933	0.025	0.249
ACC Question 2	2.409	1.206	0.729	0.044	0.459
ACC Question 3	3.182	0.537	0.495	0.707	0.306

Table 6.1 - ACC questions: statistical analysis for PCG-group participants

Category	Mean	Variance	Skewness	Kurtosis	Confidence Level (95%)
Pre-treatment					
ACC Question 1	4.625	0.332	-1.281	0.859	0.230
ACC Question 2	2.583	1.384	0.394	-0.232	0.471
ACC Question 3	3.250	0.804	0.247	-0.536	0.359
Post-treatment					
ACC Question 1	4.667	0.319	-1.522	1.626	0.226
ACC Question 2	2.458	1.389	0.195	-0.765	0.472
ACC Question 3	3.500	0.696	0.736	-0.343	0.334
Two weeks post-treatment					
ACC Question 1	4.625	0.245	-0.551	-1.859	0.198
ACC Question 2	2.625	1.462	0.325	-0.521	0.484
ACC Question 3	3.167	0.667	0.194	-0.393	0.327

Table 6.2 - ACC questions: statistical analysis for generic-group participants

6.3 PYRAMID OF INDIVIDUALIZATION ANALYSIS

As discussed, our PoI goal was to improve the likelihood of persuading users through the PCG process model (i.e., generating content using segmentation, comprehension, and individualization layers). The following subsections report mean value, standard deviations, skewness, kurtosis, and confidence level (95%) for all outcome variables associated with the three layers for the generic and PCG content.

6.3.1 Segmentation-Layer Analysis

This section covers our analyses of participant responses to the following Direct Assessment (DA) questions:

- Did you find the content interesting?
- Did you find the content enjoyable?

- Did you find the content relevant?

As Table 6.3 and Table 6.4 demonstrate, the result from PCG- and generic-group responses show negative skewness and kurtosis. Negative values indicate the data skews left with light-tailed distribution—that is, the left tail is longer than the right tail, and the sample does not contain extreme values.

Category	Mean	Variance	Skewness	Kurtosis	Confidence Level (95%)
Interesting	4.045	0.426	-0.042	-0.367	0.273
Enjoyable	3.545	1.117	0.268	-1.194	0.442
Relevant	4.364	0.433	-0.547	-0.528	0.275

Table 6.3 – Segmentation layer: statistical analysis for PCG-group participants

Category	Mean	Variance	Skewness	Kurtosis	Confidence Level (95%)
Interesting	4.167	0.667	-0.852	0.623	0.327
Enjoyable	4.000	0.870	-0.702	-0.149	0.373
Relevant	2.917	1.036	-0.091	-0.368	0.407

Table 6.4 – Segmentation layer: statistical analysis for generic-group participants

To compare responses to the questions across the treatment levels, we used the Mann-Whitney U test. As Table 6.5 highlights, in the "Interesting" and the "Relevant" categories, the computed p-values are lower than the significance level when alpha is equal to 0.05. In the "relevant" category, the PCG mean value is higher than generic mean value. However, for the "interesting" question, the mean value for the Generic group is higher than that of the PCG group which is the opposite of what was hypothesized.

	PCG		Generic		Mann-Whitney U test	
Category	**Mean**	**Variance**	**Mean**	**Variance**	**U**	**p-value**
Interesting	4.045	0.426	4.167	0.667	297	< 0.0001
Enjoyable	3.545	1.117	4.000	0.870	332	0.133
Relevant	4.364	0.433	2.917	1.036	68	< 0.0001

Table 6.5 – Segmentation layer: comparison of samples from PCG- and generic-group participants

Based on these results, we accept H1.3 and reject H1.1 and H1.2. Overall, there was partial support that the PCG content attracts more reader's attention than the generic persuasion content.

6.3.2 Comprehension-Layer Analysis

This section covers our analyses of participant responses to the following questions:

- Did you find the content clear and easy to follow?
- Did you find the content's word and sentence structure complex?
- Did you find the content easy to understand and comprehend?

As demonstrated in Table 6.6 and Table 6.7, both groups' "Flow" (i.e., easy to follow) and "Comprehensible" categories have negative skewness and positive kurtosis. This shows the dataset skews left with heavy tail.

Category	**Mean**	**Variance**	**Skewness**	**Kurtosis**	**Confidence Level (95%)**
Flow	4.545	0.450	-1.221	-0.141	0.280
Complexity	2.455	1.212	0.598	-0.123	0.460
Comprehensible	4.227	0.755	-0.963	0.408	0.363

Table 6.6 – Comprehension layer: statistical analysis for PCG-group participants

Table 6.7 – Comprehension layer: statistical analysis for generic-group participants

Category	**Mean**	**Variance**	**Skewness**	**Kurtosis**	**Confidence Level (95%)**
Flow	3.792	1.129	-0.734	0.498	0.425
Complexity	2.958	1.259	0.894	-0.545	0.449
Comprehensible	3.792	1.303	-0.895	0.152	0.457

To compare responses across the two treatment groups, we used the Mann-Whitney U test. As highlighted in Table 6.8, the computed p-value is lower than the significance level when alpha is equal to 0.05 in the "Flow" category, and the PCG mean is higher than the generic mean. We, therefore, accept H1.4 and reject H1.5 and H1.6. Overall, there was partial support that the PCG generated content is more comprehensible than the generic persuasion content.

	PCG		Generic		Mann-Whitney U test	
Category	**Mean**	**Variance**	**Mean**	**Variance**	**U**	**p-value**
Flow	4.545	0.450	3.792	1.129	152	0.002
Complexity	2.455	1.212	2.958	1.259	327	0.146
Comprehensible	4.227	0.755	3.792	1.303	208	0.124

Table 6.8 – Comprehension layer: comparison of samples from PCG- and generic-group participants

6.3.3 Individualization-Layer Analysis

This section covers our analyses on participants' responses to the following questions:

- Did you find the content trustworthy?
- Did you find the content accurate?
- Did you find the content authentic?
- Did you find the content believable?

As shown in Table 6.9 and Table 6.10, skewness and kurtosis are negative for all categories related to the individualization layer. The negative value indicates the data skews left with light-tailed distribution. The purpose of reporting skewness and kurtosis is solely to describe the sample data.

Category	Mean	Variance	Skewness	Kurtosis	Confidence Level (95%)
Trustworthy	4.136	0.790	-0.734	-0.179	0.371
Accurate	4.273	0.684	-0.574	-1.282	0.346
Authentic	3.955	0.807	-0.338	-0.764	0.375
Believable	4.545	0.355	-0.933	-0.025	0.249

Table 6.9 – Individualization layer: statistical analysis for PCG participants

Category	Mean	Variance	Skewness	Kurtosis	Confidence Level (95%)
Trustworthy	4.208	0.694	-0.426	-1.426	0.496
Accurate	4.042	0.650	-0.621	-0.301	0.323
Authentic	3.875	0.984	-0.023	-1.531	0.397
Believable	3.625	1.201	-0.024	-1.312	0.238

Table 6.10 – Individualization layer: statistical analysis for generic-group participants

To compare responses across the treatment groups, we used the Mann-Whitney U test. As Table 6.11 highlights, the computed p-value is lower than the significance level alpha=0.05 for two questions. Mean responses for both the "accurate" and "believable" questions are significantly higher for PCG than the generic group. Thus, by accepting H2.2 and H2.4 and rejecting H2.1 and H2.3, we found partial support for PCG content as being more acceptable than generic content. Overall, there was partial support for PCG content being more 'individualized' than generic persuasion content.

	PCG		Generic		Mann-Whitney U test	
Category	**Mean**	**Variance**	**Mean**	**Variance**	**U**	**p-value**
Trustworthy	4.136	0.790	4.208	0.694	273.5	0.075
Accurate	4.273	0.684	4.042	0.650	221	0.002
Authentic	3.955	0.807	3.875	0.984	251	0.893
Believable	4.545	0.355	3.625	1.201	137.5	0.001

Table 6.11 – Individualization layer: comparison of samples from PCG- and generic-group participants

6.4 ATTITUDE CHANGE ANALYSIS

In this section, we start with examining hypothesis H2.5 for all three ACC questions to assess the impact of the treatment on influencing participants' attitude towards climate change as well as retaining the attitude:

> H2.5 There is a positive significant change in PCG-group-participant attitudes toward the study's subject (climate change) from the pre-treatment to post-treatment responses.

If the PCG content has an impact on changing participants' attitudes towards climate change immediately after the treatment, we also assess whether participants retain the attitude.

> H2.6 There is no significant change in PCG-group-participant attitudes toward the study's subject (climate change) between their post-treatment responses to their two-weeks post-treatment responses.

We also tested the above hypotheses against the generic group dataset so that we could look for significant changes across different timeframes within generic group separately. We used the non-parametric Mann-Whitney test with alpha equals to 0.05 to assess whether two sets of data are significantly different from each other. Table 6.12, Table 6.13, and Table 6.14 show the p-value (two-tailed) for each ACC question for the separate PCG and generic groups. In the following tables, T1, T2, and T3 correspond to the pre-treatment, post-treatment, and two-weeks post-treatment timeframes respectively.

	PCG				Generic			
Timeframe	**Mean**	**Standard deviation**	**U**	**p-value**	**Mean**	**Standard deviation**	**U**	**p-value**
T1 vs T2	T1 = 4.500	T1 = 0.598	216	0.717	T1 = 4.625	T1 = 0.576	276	0.994
	T2 = 4.636	T2 = 0.492			T2 = 4.667	T2 = 0.565		
T2 vs T3	T2 = 4.636	T2 = 0.492	257	0.998	T2 = 4.667	T2 = 0.565	307	0.933
	T3 = 4.545	T3 = 0.596			T3 = 4.625	T3 = 0.495		

Table 6.12 – Comparison of samples from ACC Question 1 during different timeframes

	PCG				Generic			
Timeframe	**Mean**	**Standard deviation**	**U**	**p-value**	**Mean**	**Standard deviation**	**U**	**p-value**
T1 vs T2	T1 = 2.318	T1 = 1.211	249.5	0.871	T1 = 2.583	T1 = 1.176	308	0.670
	T2 = 2.182	T2 = 1.006			T2 = 2.375	T2 = 1.176		
T2 vs T3	T2 = 2.182	T2 = 1.006	217	0.548	T2 = 2.375	T2 = 1.176	262	0.601
	T3 = 2.455	T3 = 1.184			T3 = 2.583	T3 = 1.139		

Table 6.13 – Comparison of samples from ACC Question 2 during different timeframes

	PCG				Generic			
Timeframe	**Mean**	**Standard deviation**	**U**	**p-value**	**Mean**	**Standard deviation**	**U**	**p-value**
T1 vs T2	T1 = 3.182	T1 = 0.733	346	0.009	T1 = 3.250	T1 = 0.897	249	0.403
	T2 = 2.591	T2 = 0.590			T2 = 3.500	T2 = 0.834		
T2 vs T3	T2 = 2.591	T2 = 0.590	150	0.056	T2 = 3.500	T2 = 0.834	341	0.242
	T3 = 3.091	T3 = 0.684			T3 = 0.684	T3 = 0.816		

Table 6.14 – Comparison of samples from ACC Question 3 during different timeframes

According to the calculated p-values, there were no significant attitude changes expressed by the generic group. The same statement is true for ACC Questions 1 and 2 for the PCG-group dataset. As Table 6.14 highlights, the computed p-value is lower than the significance level alpha=0.05 for a single question. We can only accept Hypothesis H2.5 for ACC Question 3 for the PCG group and reject both hypotheses for ACC Questions 1 and 2. Though, for T2 vs T3 timeframe, the calculated p-value for ACC Question 3 is higher than the significance level alpha=0.05, so we reject H2.6. Overall, there was partial support for PCG content to change participants' attitudes toward the study's subject (climate change) from the pre-treatment to post-treatment assessment. However, that significant change was not retained when the ACC questions were re-assessed two weeks later.

To examine whether there were any significant changes between the groups, we calculated the mean-value differences between two timeframes. To calculate the differences, we subtracted pre-treatment from post-treatment or post-treatment from two-weeks post-treatment for PCG- and generic-group responses. To compare the PCG- and generic-group responses with each other, we applied the nonparametric Mann-Whitney test (alpha = 0.05) to their mean-value differences. Table 6.15, Table 6.16, and Table 6.17 show the calculated p-value (two-tailed). In the following tables, T1, T2, and T3 correspond to the pre-treatment, post-treatment, and two-weeks post-treatment timeframes respectively.

	PCG		Generic		Mann-Whitney U test	
Timeframe	**Delta mean**	**Delta SD**	**Delta mean**	**Delta SD**	**U**	**p-value**
T1 vs T2	-0.136	0.468	-0.042	0.204	277.5	0.690
T2 vs T3	0.091	0.526	0.042	0.464	262	0.751

Table 6.15 – Comparison of PCG and generic differences for ACC Question 1

	PCG		Generic		Mann-Whitney U test	
Timeframe	**Delta mean**	**Delta SD**	**Delta mean**	**Delta SD**	**U**	**p-value**
T1 vs T2	0.136	0.468	0.208	0.204	274	0.997
T2 vs T3	-0.273	0.550	-0.208	0.658	220	0.084

Table 6.16 – Comparison of PCG and generic differences for ACC Question 2

	PCG		Generic		Mann-Whitney U test	
Timeframe	**Delta mean**	**Delta SD**	**Delta mean**	**Delta SD**	**U**	**p-value**
T1 vs T2	0.591	0.796	-0.250	-0.250	126	< 0.0001
T2 vs T3	-0.500	0.673	0.333	0.761	195.5	0.065

Table 6.17 – Comparison of PCG and generic differences for ACC Question 3

Although there were no significant changes between the first two ACC questions, there is a significant difference in the mean deltas across the PCG and generic treatment groups for the third ACC question. The result supports the previous section's result. We can only accept Hypothesis H2.5 for ACC Question 3 for the PCG group and reject H2.6. Overall, there was partial support for PCG content to change participants' attitudes toward the study's subject (climate change) from

the pre-treatment to post-treatment assessment. However, that significant change was not retained when the ACC questions were re-assessed two weeks later.

6.5 PERSUASION QUESTION ANALYSIS

This section, focusing on participants' self-assessments on content persuasion by examining participants' response on DA persuasion question ("After reading the provided text, I am persuaded to take action against climate change") during post-treatment and 2 weeks post-treatment.

As Table 6.18 demonstrates, we observed a higher mean in content persuasiveness responses from the PCG group than for the generic-content group immediately after the treatment was provided. This is based on the question: "Did you find the content about climate change persuasive?" Skewness and kurtosis are negative for both groups. The negative value indicates data skews left with light-tailed distribution.

Group	Mean	Variance	Kurtosis	Skewness	Confidence level (95%)
PCG	4.227	0.660	-1.310	-0.460	0.340
Generic	4.042	0.660	-0.318	-0.534	0.344

Table 6.18 – Content quality: statistical analysis for PCG- and generic-group participants

Figure 6.7 and Figure 6.8 illustrate a similar statement using the DA persuasion question. By comparing the mean value for the PCG and generic groups within the post-treatment

timeframe, we noticed approximately 10% improvement in persuasion when using PCG content. Similar standard-deviation values signify an equal level of reliability for both groups' responses.

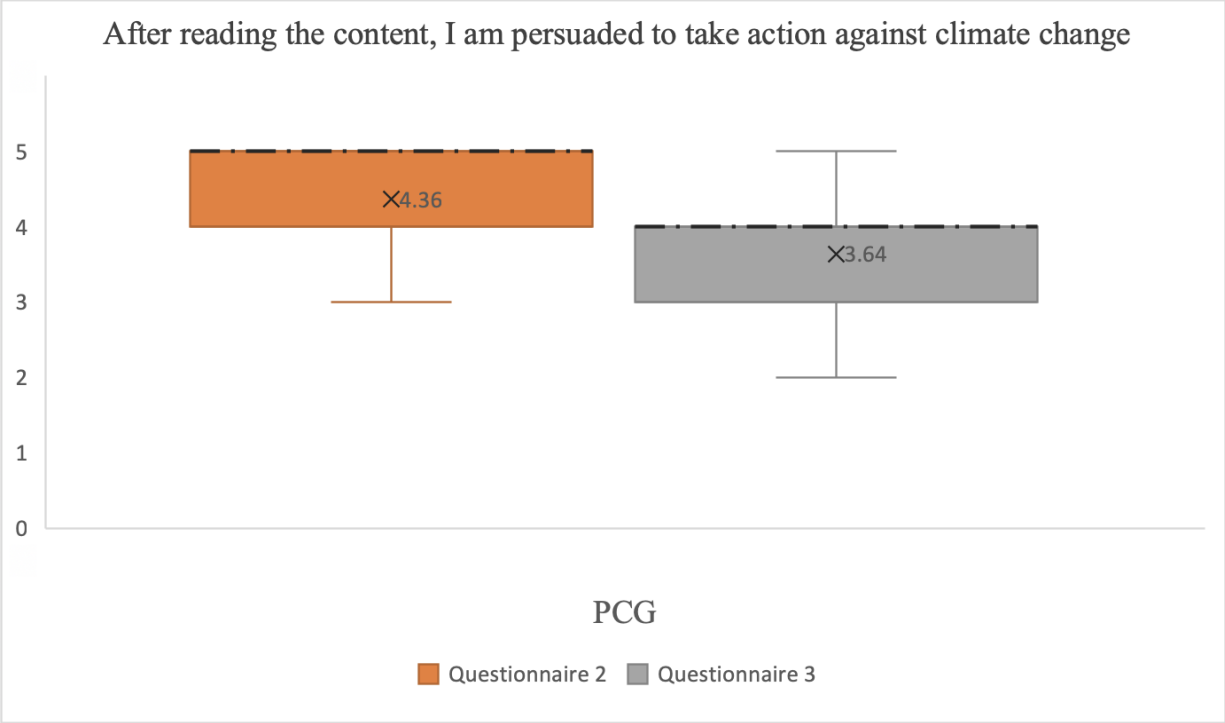


Figure 6.7 – Distribution of DA-persuasion responses for the PCG group

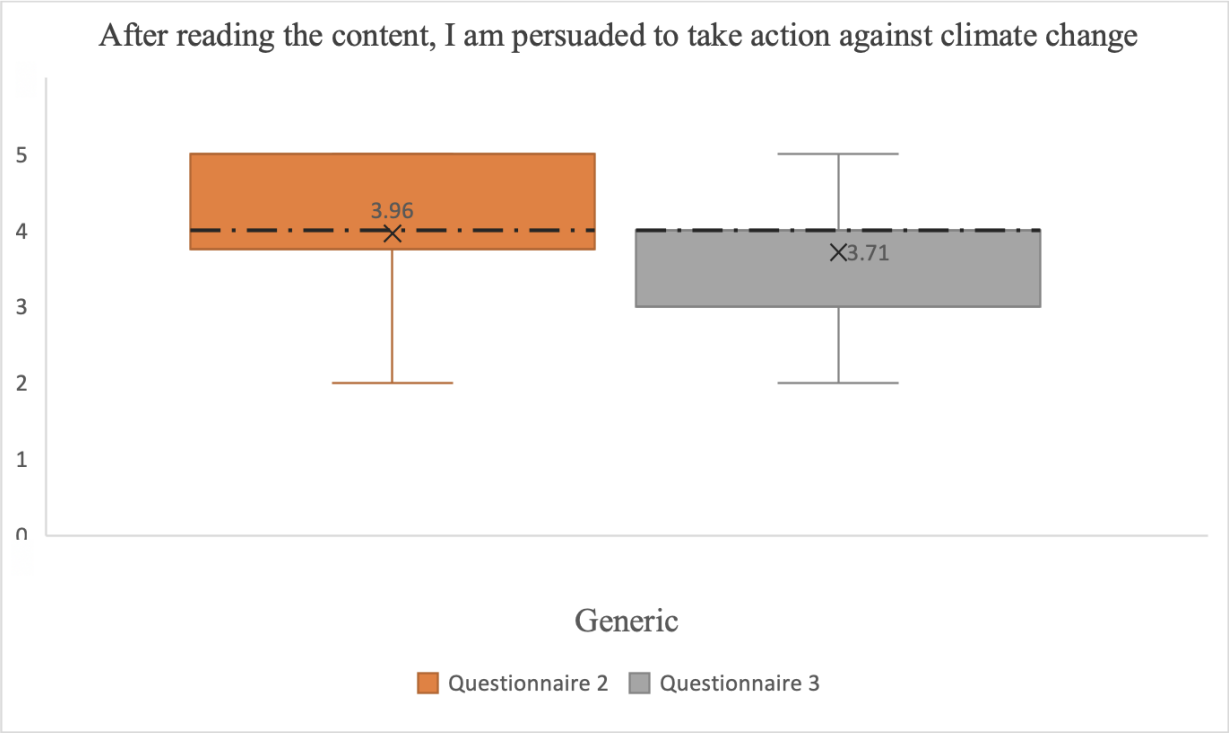


Figure 6.8 – Distribution of DA-persuasion responses for the generic group

To determine any significant changes between the post- and two-week responses, we applied the nonparametric Mann-Whitney test (alpha = 0.05) to each data group. Table 6.19 shows the calculated p-value (Two-tailed). Furthermore, Table 6.20 captures, DA persuasion question calculated p-value for PCG and generic differences.

	PCG				Generic			
Timeframe	**Mean**	**Standard deviation**	**U**	**p-value**	**Mean**	**Standard deviation**	**U**	**p-value**
T2 vs T3	T2 = 4.364	T2 = 0.790	320	0.052	T2 = 3.958	T2 = 0.908	336	0.259
	T3 = 3.727	T3 = 1.120			T3 = 3.708	T3 = 0.908		

Table 6.19 – DA-persuasion-question comparison during different timeframes

	PCG		Generic		Mann-Whitney U test	
Timeframe	**Delta mean**	**Delta SD**	**Delta mean**	**Delta SD**	**U**	**p-value**
T2 vs T3	0.727	0.767	0.250	0.794	181	0.05

Table 6.20 - Comparison of PCG and generic differences for DA persuasion question

As the above tables show, since the computed p-value is just equal to the significance level alpha=0.05, we do not see any significant changes in either group. Thus, the result is aligned with the Section 6.4 and Section 6.5 results on retention and we reject H2.6.

6.6 DISCUSSION

From personalizing content to making content persuasive, we observed different results during the analysis. In this section, we discuss our findings in terms of the following:

- General discussion: a short summary of the hypotheses.
- ACC questions discussion: discussion of changed participant attitudes toward climate change.
- Effectiveness of persuasion discussion: discussion of results from the YAC acceptance-and-retention-factor standpoint.

6.6.1 General Discussion

As Table 6.21 demonstrates, in some areas, PCG content did not show any advantages over the generic content. This can be interpreted as a sign of multiple factors:

- The PCG failed to be effective in those categories,
- PCG and generic content are both equally effective in those categories, and
- Participants responded inaccurately or falsely to questions (response bias).

We should also take into account the Hawthorne effect, which typically results in demand characteristics, a type of response bias with which participants change their responses simply because they are part of an experiment [128]. Martine Orne says, “This arises because participants are actively engaged in the experiment, and may try and figure out the purpose, or adopt

certain behaviours they believe belong in an experimental setting" [129]. Due to the nature of this experiment, this effect was one of the hardest inbuilt biases to eliminate or factor into the design.

Hypotheses	**Status**
H1.1 The readers find the PCG content significantly more interesting than the generic content.	Rejected
H1.2 The readers find the PCG content significantly more enjoyable than the generic content.	Rejected
H1.3 The readers find the PCG content significantly more relevant than the generic content.	Accepted
H1.4 The readers find the PCG content significantly easier to follow than the generic content.	Accepted
H1.5 The readers find the PCG content significantly less complex than the generic content.	Rejected
H1.6 The readers find the PCG content significantly easier to understand than the generic content.	Rejected
H2.1 The readers find the PCG content significantly more trustworthy than the generic content.	Rejected
H2.2 The readers find the PCG content significantly more accurate than the generic content.	Accepted
H2.3 The readers find the PCG content significantly more authentic than the generic content.	Rejected

H2.4 The readers find the PCG content significantly more believable than the generic content.	Accepted
H2.5 There is a positive significant change in PCG-group-participant attitudes toward the study's subject (climate change) from the pre-treatment to post-treatment responses.	Partially Accepted
H2.6 There is no significant change in PCG-group-participant attitudes toward the study's subject (climate change) between their post-treatment responses to their two-weeks post-treatment responses.	Rejected

Table 6.21 – Summary of hypotheses verification

6.6.2 ACC Questions Discussion

In the General Analysis

- Acquiescence bias, a response category in which participants tend to be agree with all statements [130];
- Social-desirability bias, a response-bias category in which participants deny undesirable traits that are not socially desirable [131]; or
- The possibility that climate change was already well-known among participants, which is likely since our sample is from people affiliated with Carleton University (i.e., participants were like to belong to a moderate-to-highly educated group of people).
- The 5-point Likert scale could have caused some data skew.

Although we did not record any significant response changes for ACC Question 2 ("I believe that things I do have no effect on the quality of the environment") for different timeframes, the mean value here was lower than for ACC Question 1 (2.5). We expected this result for generic content since it focused only on problem awareness by backing up that awareness with scientific facts. On the other hand, the PCG content included a designated solution section. The section showed how others have changed their behaviour to fight climate change, yet it did not significantly impact participant responses. Most examples here were of large-scale behavioural changes, like switching to alternative-energy sources or investing in innovative transportation solutions. It may be that these solutions are more applicable to governments and large organizations than people.

We observed a different pattern in ACC Question 3 ("I believe most of the concerns about environmental problems have been exaggerated") than ACC Questions 1 and 2. According to the calculated mean value, participants attitudes changed by nearly 20% immediately after reading the PCG content. This indicates a significant positive change in participant attitudes toward climate change.

In the Attitude Change Analysis section, we observed a similar pattern to the General Analysis. The calculated p-value for the pre-treatment and post-treatment timeframes for ACC Question 3 were lower than the alpha value. This could support the pattern we documented in General Analysis for ACC Question 3. It could also signify the PCG's effectiveness to make content more realistic and believable than the generic content. After all, one of the PCG-content sections was dedicated to the latest personalized news related to climate change. Typically, during this phase of personalization, the PCG system looked for news articles within participants'

respective geolocations. Thus, participants were likely personally connected to the news-article content rather than scientific facts in the generic content.

The PoI analysis also led us to a similar conclusion: the content generated from our system was significantly more relevant to users than the generic content. Due to the PCG-segmentation components, PCG content had a higher relevancy mean value than the generic content. But for the other two categories (finding the content interesting and enjoyable), the mean values or the PCG and generic content were not significantly different. We think the embedded video in the generic content gave an extra boost to the mean values for interesting and enjoyable categories.

6.6.3 Effectiveness-of-Persuasion Discussion

YAC defines the persuasion cycle as complete when the persuadee accepts and retains the information. Thus, to verify the persuasion effectiveness, we measured whether users accepted the persuasive text and whether the users remembered the persuasive text.

For all three sections (6.4, 6.5 and 6.6), we conclude that the PCG content has a higher likelihood of being more persuasive than the generic. As observed in the analysis section and the results from the three sections, the immediate persuasion impact for the PCG content is significant and reliable, a statement we can make with high confidence. This statement also aligns with results from the General Analysis which demonstrate a similar statement, though we cannot say the same for the retention impact, which indicates that the PCG content was not retainable. We also documented that enhancing the PCG content with news articles from reputable news agencies improved the content's believability. These two factors may have directly increased overall persuasiveness.

Chapter 7.

CONCLUSION

This thesis addresses the problem of creating persuasive content through a personalized model. Our primary research goal was to design and develop a system based on our proposed PoI model and perform a preliminary evaluation to determine whether the system can improve the persuasiveness of content compared to generic persuasive text written by an expert.

7.1 CONTRIBUTIONS

We explored the research goal by completing the design and implementation of the PoI system based on YAC [18] criteria. To evaluate our proposed system's effectiveness, we defined and attempted to verify the below criteria by comparing PCG content with a generic persuasive content:

- Content generated by the proposed system is perceived by readers to be personalized, and
- Content generated by the proposed system is perceived by readers to be persuasive.

From a personalization standpoint, the PCG content was found to be personalized and, as expected, more relevant to users. From a persuasion standpoint, according to our analyses, we conclude that PCG content has greater likelihood of being persuasive than the generic text. As observed from the analysis sections from Chapter 6, the post-treatment persuasion impact for PCG content is significant. We did not find enough evidence to support the same in terms of the persuasion-retention impact.

While further research is required to fine-tune the model, current findings demonstrate potential uses in education and informative applications, such as customer briefing, marketing, etc.

The primary contributions of the presented research include:

- The PoI as a new model and the basis for creating persuasive content. The new model extends the YAC factors into the context of persuasive text by defining different layers of personalization;
- A pluggable software framework that integrates newly designed re-usable software components with existing ones. The components designed and developed through this research project are a:
  - rule-based component to create and control personalized content by the author (persuader);
  - user (persuadee) profile builder based on extracting information from social media data
  - user trust graph and related algorithms for selecting trustworthy supportive information

7.2 LIMITATIONS OF STUDY

Our findings are limited for the following reasons:

- As mentioned, the sample we used to evaluate PCG was insufficient to generalize the results. We realized this limitation when we calculated a large standard-deviation error bar for both PCG and generic groups' mean values for the post-treatment timeframe. We interpreted this to mean that the mean value is not reliable to represent the dataset. Using a small sample size could be the cause of this issue. Intuitively, future

improvement could be made by recruiting more participants for the evaluation. In this study, we were limited to a sample size of 48 because of resource constraints.[1]

- Our recruitment script (Call for Research Participation) was posted on bulletin boards, on designated areas for flyers, on the Carleton University campus, and on university-affiliated social-media channels. This means participants were likely to be moderately to highly educated, and so the sample set does not represent a diverse group of people.
- A two-week period to measure message memorability may not be sufficient since some scholars mention four or more weeks as an appropriate waiting time to measuring retention of persuasion content [121][123][124]. However, since we rejected H2.6, a longer waiting period would not likely change our current result.
- Due to limitation that we had with hosting the surveys and PCG software on a public server, we cannot comment on how much it took for participants to complete the experiment. The experiment was designed in such a way that participants had to go through the persuasive contents before completing the surveys. Therefore, we have high confidence that participants at least viewed the contents.

7.3 FUTURE RESEARCH

In addition to the ways this study could be improved by addressing the above limitations, this research can be further developed in multiple directions. The privacy concerns with respect to such system that relies on user's social media data were out of the scope of this project but need to be investigated properly.

[1] A larger sample size could also minimize the Hawthorne effect.

Comparing the PCG content to a set of generic contents instead of one single comparator would increase researcher confidence to evaluate the research goals. Furthermore, the PCG content could also be compared to a content selected by a conventional recommender system. Both PCG and conventional recommender systems rely on building user profiles, even though one assembles contents, and the other recommends contents.

Although we implemented the current system to integrate with Twitter, the system can be expanded to other types of social media accounts, such as Facebook or LinkedIn. Aggregating different sources of information would result in more accurate user profiles.

To increase the likelihood of choosing the best possible supporting opinion, the system may generate a list of top experts for the given individualization topic instead of generating a list of top trustworthy people. This could be achieved by initiating a semantic approach to organizing each candidate in terms of their expertise level of a given topic. Based on Dreyfus and Dreyfus terminology [102], this task would have two phases:

- Confirming the rule- and fact-based-knowledge factor, which can be summarized as the processing and classification of candidate activities and the finding of the total number of activities[1] associated with an individualization topic. The posts may contain links to websites, which is also reflected in the classification. The goal would be to define how well a candidate knows the topic.
- Estimating the experience-based-knowledge factor: Well-known readability indexes like Flesch-Kincaid [53], Gunning Fog [54] or Coleman-Liau [55] would be formulated to evaluate and indicate comprehension difficulty of texts. The calculated readability score for the candidate's published work or for the recommended content would be the

[1] Activity is a candidate's posts and reposts from other users.

primary source from which to determine the expertise level. Although there is no ranking mechanism on social networks to build up reputation, the number of followers, likes, and reposts can help estimate reputation level.

Once the above steps are taken, it would be possible to select the most relevant and appropriate supporting opinions. However, the personalization system would likely fail to find supporting opinions from the legitimate-experts list; it is also likely that the accuracy level for the closest match would be too low. To address a lack of supporting opinions, future research should study the effectiveness of mining opinion through second- or third-tier experts (trustworthy people) recommended by readers' preliminary experts (first-tier experts). In the context of social networks, *transitivity* means the close friends of one's friends are also one's friends. In other words, a reader would trust people who are referenced as credible sources by that reader's experts.

To improve retention, future researchers might consider incorporating relevant personal tweets and opinions into the PCG content rather than relying on BrainyQuote or similar repositories. Researchers could also further enhance the PCG system so that upon reviewing the persuasive content, users would be sent follow-up notices by the PCG system to remind them about the persuasive content. In this way, researchers could examine whether periodic reminders about the persuasive content have an impact on improving persuasion retention.

The PCG system could also be expanded to include the personality traits layer in the PoI model. The goal for this layer would be to match content wording with user persona. A common method in classifying individuals is based on the five major trait domains of the human personality. In a sample of 324 survey respondents suggests that persuasive effectiveness can be increased by aligning the message with audience persona, which would improve the likelihood argument acceptance by users and changing user opinion [84].

Overall there are many opportunities to improve and explore on both the design and validation of the PCG system. In continually improving the system design future researchers could incorporate other complementary persuasion theories. Besides, given its pluggable framework, we anticipate this initial work will provide a foundation for a sustainable stream of research for many years.

REFERENCES

[1] Aesop, *The North Wind and the Sun*. .

[2] H. W. Simons, J. Morreale, and B. Bronbeck, "Persuasion in Society. Thousand Oaks: Sage Publications, Inc.," 2001.

[3] *The American Heritage Dictionary of the English Language*. Boston: Houghton Mifflin, 2000.

[4] C. M. Kardash and R. J. Scholes, "Effects of preexisiting beliefs, epistemological beliefs, and need for cognition on interpretation of controversial issues.," *J. Educ. Psychol.*, vol. 88, no. 2, pp. 260–271, 1996.

[5] L. Mason, F. Scirica, and L. Salvi, "Effects of beliefs about meaning construction and task instructions on interpretation of narrative text," *Contemp. Educ. Psychol.*, vol. 31, no. 4, pp. 411–437, 2006.

[6] H. M. Crowson and T. K. DeBacker, "Belief, Motivational, and Ideological Correlates of Human Rights Attitudes," *J. Soc. Psychol.*, vol. 148, no. 3, pp. 293–310, Jun. 2008.

[7] K. Torning and H. Oinas-Kukkonen, "Persuasive System Design: State of the Art and Future Directions," 2009.

[8] R. H. Fazio, "Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework," *Adv. Exp. Soc. Psychol.*, vol. 23, pp. 75–109, 1990.

[9] I. Ajzen, "The theory of planned behavior," *Organ. Behav. Hum. Decis. Process.*, vol. 50, no. 2, pp. 179–211, 1991.

[10] D. Weinberger, R. Levine, C. Locke, and D. "Doc" Searls, *The Cluetrain Manifesto*. Basic Books, 2000.

[11] David Weinberger, “Journal of the Hyperlinked Organization,” 1999. [Online]. Available: http://www.hyperorg.com/backissues/joho-sept30-99.html. [Accessed: 05-Dec-2018].

[12] A. Cöner, “Personalization and Customization in Financial Portals,” *J. Am. Acad. Bus.*, vol. 2, no. 2, pp. 498–504, 2003.

[13] M. Lou Roberts and D. Zahay, *Internet Marketing: Integrating Online and Offline Strategies: Mary Lou Roberts, Debra Zahay: 9781133625902: Amazon.com: Books*, 3rd ed. Cengage Learning, 2003.

[14] J. Vesanen, “What is personalization? A conceptual framework,” *Eur. J. Mark.*, vol. 41, no. 5/6, pp. 409–418, Jun. 2007.

[15] Tianyi Jiang and A. Tuzhilin, “Improving Personalization Solutions through Optimal Segmentation of Customer Bases,” *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 3, pp. 305–320, Mar. 2009.

[16] V. Gupta and G. Lehal, “A Survey of Text Mining Techniques and Applications,” *JETWI*, vol. 1, no. 1, pp. 60–76, 2009.

[17] A. Khataei and A. Arya, “Personalized Presentation Builder for Persuasive Communication,” *Commun. Des. Q. Rev.*, vol. 3, no. 3, pp. 25–32, 2015.

[18] W. J. McGuire, “Personality and attitude change: An information-processing theory,” *A. G. Greenwood, T. C.*, 1968.

[19] C. I. Hovland, I. L. Janis, and H. H. Kelley, “Communication and persuasion; psychological studies of opinion change.” Yale University Press, 1953.

[20] E. Aronson, T. D. Wilson, and R. M. Akert, *Social Psychology*, 7th ed. Pearson, 2009.

[21] R. E. Petty and J. T. Cacioppo, *Communication and Persuasion*. New York, NY: Springer New York, 1986.

[22] H. Oinas-Kukkonen and M. Harjumaa, “A systematic framework for designing and evaluating persuasive systems,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5033 LNCS, pp. 164–176, 2008.

[23] J. T. Cacioppo, R. E. Petty, C. F. Kao, and R. Rodriguez, “Central and peripheral routes to persuasion: An individual difference perspective.,” *J. Pers. Soc. Psychol.*, vol. 51, no. 5, pp. 1032–1043, 1986.

[24] H. Oinas-Kukkonen and M. Harjumaa, “Persuasive Systems Design: Key Issues, Process Model, and System Features,” *Commun. Assoc. Inf. Syst.*, vol. 24, no. 28, pp. 485–500, 2009.

[25] L. Kohlberg, “The Claim to Moral Adequacy of a Highest Stage of Moral Judgment,” *J. Philos.*, vol. 70, no. 18, p. 630, Oct. 1973.

[26] B. Fogg, “Persuasive computers,” in *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI ’98*, 1998, pp. 225–232.

[27] B. J. Fogg and B. J., “Persuasive technology: Using Computers to Change What We Thinks and Do.,” *Ubiquity*, vol. 2002, no. December, p. 2, Dec. 2002.

[28] R. Cialdini, “The Science of Persuasion,” *Scientific American*, 2004.

[29] R. B. Cialdini, *Influence : science and practice*. Pearson Education, 2009.

[30] K. Rhoads, “How Many Influence, Persuasion, Compliance Tactics & Strategies Are There?,” 2007. [Online]. Available: http://www.workingpsychology.com/numbertactics.html. [Accessed: 03-Jul-2017].

[31] M. Kaptein, P. Markopoulos, B. de Ruyter, and E. Aarts, “Personalizing persuasive technologies: Explicit and implicit personalization using persuasion profiles,” *Int. J. Hum. Comput. Stud.*, vol. 77, pp. 38–51, May 2015.

[32] A. Ellis, “Reason and emotion in psychotherapy.” Lyle Stuart, 1962.

[33] R. T. LaPiere, “Attitudes vs. Actions,” *Soc. Forces*, vol. 13, no. 2, pp. 230–237, Dec. 1934.

[34] J. Hamari, J. Koivisto, and T. Pakkanen, “Do Persuasive Technologies Persuade? - A Review of Empirical Studies,” Springer International Publishing, 2014, pp. 118–136.

[35] S. Purpura, V. Schwanda, K. Williams, W. Stubler, and P. Sengers, “Fit4Life: The Design of a Persuasive Technology Promoting Healthy Behavior and Ideal Weight,” in *SIGCHI Conference on Human Factors in Computing Systems*, 2011, pp. 423–432.

[36] M. Busch, S. Patil, G. Regal, C. Hochleitner, and M. Tscheligi, “Persuasive Information Security: Techniques to Help Employees Protect Organizational Information Security,” *Persuas. Technol.*, vol. 9638, pp. 339–351, Apr. 2016.

[37] R. Hunter, *The gamification handbook : everything you need to know about gamification.* Lightning Source, 2011.

[38] J. McGonigal, *Reality is broken : why games make us better and how they can change the world*. Random House, 2012.

[39] R. Orji, J. Vassileva, and R. L. Mandryk, “Modeling the efficacy of persuasive strategies for different gamer types in serious games for health,” *User Model. User-adapt. Interact.*, vol. 24, no. 5, pp. 453–498, Dec. 2014.

[40] G. F. Tondello, R. Orji, and L. E. Nacke, “Recommender Systems for Personalized Gamification,” in *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization - UMAP ’17*, 2017, pp. 425–430.

[41] A. Nenkova and K. McKeown, “Automatic Summarization,” *Found. Trends Inf. Retrievals*, vol. 5, no. 2–3, pp. 103–233, 2011.

[42] K. Sparck Jones, “A statistical interpretation of term specificity and its application in retrieval,” *J. Doc.*, vol. 28, pp. 11–21, 1972.

[43] Y. Gong and X. Liu, “Generic text summarization using relevance measure and latent semantic analysis,” in *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR ’01*, 2001, pp. 19–25.

[44] K. J. Josef Steinberger, *Text Summarization and Singular Value Decomposition*, vol. 3261. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005.

[45] J.-T. Sun, D. Shen, H.-J. Zeng, Q. Yang, Y. Lu, and Z. Chen, “Web-page summarization using clickthrough data,” in *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval - SIGIR ’05*, 2005, p. 194.

[46] J. Park, T. Fukuhara, I. Ohmukai, H. Takeda, and S. Lee, “Web content summarization using social bookmarks,” in *Proceeding of the 10th ACM workshop on Web information and data management - WIDM ’08*, 2008, p. 103.

[47] H. Zhang, Z. C. W. Ma, and Q. Cai, “A study for documents summarization based on personal annotation,” in *Proceedings of the HLT-NAACL 03 on Text summarization workshop -*, 2003, vol. 5, pp. 41–48.

[48] R. Móro and M. Bieliková, “Personalized text summarization based on important terms identification,” *Proc. - Int. Work. Database Expert Syst. Appl. DEXA*, pp. 131–135, 2012.

[49] M. Barla, “Towards Social-Based User Modeling and Personalization,” *Inf. Sci. Technol. Bull. ACM Slovakia*, vol. 3, no. 1, pp. 52–60, 2011.

[50] A. Díaz and P. Gervás, “User-model based personalized summarization,” *Inf. Process.*

Manag., vol. 43, no. 6, pp. 1715–1734, 2007.

[51] E. Pariser, *The Filter Bubble: What the Internet Is Hiding from You*. The Penguin Group, 2011.

[52] F. Van der Sluis, E. L. Van den Broek, R. J. Glassey, E. M. A. G. Van Dijk, and F. M. G. De Jong, "When complexity becomes interesting," *J. Assoc. Inf. Sci. Technol.*, vol. 65, no. 7, pp. 1478–1500, Jul. 2014.

[53] J. P. Kincaid, J. R. P. Fishburne, R. L. Rogers, and B. S. Chissom, "Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel," Feb. 1975.

[54] R. Gunning, *The Technique of Clear Writing*. McGraw-Hill, 1952.

[55] M. Coleman and T. L. Liau, "A computer readability formula designed for machine scoring."

[56] K. Collins-Thompson and J. Callan, "Predicting reading difficulty with statistical language models," *J. Am. Soc. Inf. Sci. Technol.*, vol. 56, no. 13, pp. 1448–1462, Nov. 2005.

[57] T. F. Jaegar and H. Tily, "On language utility:Processing complexity and communicative efficiency," *Wiley Interdiscip. Rev. Cogn. Sci.*, vol. 2, no. 3, pp. 323–335, 2011.

[58] C. A. Perfetti, "Verbal efficiency in reading ability," *Read. Res. Adv. theory Pract.*, vol. 6, pp. 109–143, 1998.

[59] D. L. Long, J. Wilson, R. Hurley, and C. S. Prat, "Assessing text representations with recognition: The interaction of domain knowledge and text coherence," *J. Exp. Psychol. Learn. Mem. Cogn.*, vol. 23, no. 4, pp. 816–827, 2006.

[60] L. Bode, "Political News in the News Feed: Learning Politics from Social Media," *Mass*

Commun. Soc., vol. 19, no. 1, pp. 24–48, Jan. 2016.

[61] R. M. Bond *et al.*, "A 61-million-person experiment in social influence and political mobilization," *Nature*, vol. 489, no. 7415, pp. 295–298, Sep. 2012.

[62] S. Messing and S. J. Westwood, "Selective Exposure in the Age of Social Media," *Communic. Res.*, vol. 41, no. 8, pp. 1042–1063, Dec. 2014.

[63] J. Turcotte, C. York, J. Irving, R. M. Scholl, and R. J. Pingree, "News Recommendations from Social Media Opinion Leaders: Effects on Media Trust and Information Seeking," *J. Comput. Commun.*, vol. 20, no. 5, pp. 520–535, Sep. 2015.

[64] J. Liu, Y. Cao, C. Lin, Y. Huang, and M. Zhou, "Low-Quality Product Review Detection in Opinion Summarization," in *Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 2007.

[65] A. Ghose and P. G. Ipeirotis, "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 10, pp. 1498–1512, 2011.

[66] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications," *Knowledge-Based Syst.*, vol. 89, pp. 14–46, 2015.

[67] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," *Found. Trends Inf. Retr.*, vol. 2, no. 1–2, pp. 1–135, 2008.

[68] B. Pang and L. Lee, "Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales," in *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics - ACL '05*, 2005, pp. 115–124.

[69] T. Wilson and J. Wiebe, "Annotating attributions and private states," *Proceedings of the Workshop on Frontiers in Corpus Annotations II: Pie in the Sky*. Association for

Computational Linguistics, pp. 53–60, 2005.

[70] P. Massa and P. Avesani, “Trust-aware Recommender Systems,” in *RecSys*, 2007.

[71] H. Ma, I. King, and M. R. Lyu, “Learning to recommend with social trust ensemble,” in *SIGIR ’09 Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, 2009.

[72] J. Golbeck and J. Hendler, “Inferring binary trust relationships in Web-based social networks,” *ACM Trans. Internet Technol.*, vol. 6, no. 4, pp. 497–529, Nov. 2006.

[73] C. Castillo, M. Mendoza, and B. Poblete, “Information Credibility on Twitter,” in *20th international conference on World wide web*, 2011.

[74] Kyung Soo Cho, J.-S. Ryu, J.-H. Jeong, Y.-H. Kim, and U.-M. Kim, “Credibility Evaluation and Results with Leader- Weight in Opinion Mining,” in *Cyber-Enabled Distributed Computing and Knowledge Discovery*, 2010.

[75] J. W. Pennebaker, M. E. Francis, and R. J. Booth, “Linguistic Inquiry and Word Count,” Austin, Texas, 2007.

[76] H. Kwak, C. Lee, H. Park, and S. Moon, “What is Twitter, a social network or a news media?,” in *Proceedings of the 19th international conference on World wide web - WWW ’10*, 2010, p. 591.

[77] X. Zhou, Y. Xu, Y. Li, A. Josang, and C. Cox, “The state-of-the-art in personalized recommender systems for social networking,” *Artif. Intell. Rev.*, vol. 37, no. 2, pp. 119–132, 2012.

[78] P. H. ANDREWS, “Gender Differences in Persuasive Communication and Attribution of Success and Failure,” *Hum. Commun. Res.*, vol. 13, no. 3, pp. 372–385, Mar. 1987.

[79] M.-L. Mares and Y. Sun, “The Multiple Meanings of Age for Television Content

Preferences," *Hum. Commun. Res.*, vol. 36, no. 3, pp. 372–396, Jun. 2010.

[80] A. H. Eagly, "Comprehensibility of persuasive arguments as a determinant of opinion change.," *J. Pers. Soc. Psychol.*, vol. 29, no. 6, pp. 758–773, 1974.

[81] K. Collins-Thompson, "Computational Assessment of Text Readability: A Survey of Past, Present, and Future Research," *Www-Personal.Umich.Edu*, 2014.

[82] L. R. Goldberg, "The structure of phenotypic personality traits," *Am. Psychol.*, vol. 1, no. 48, pp. 26–34, 1993.

[83] E. C. Tupes and R. E. Christal, "Recurrent Personality Factors Based on Trait Ratings," 1961.

[84] J. B. Hirsh, S. K. Kang, and G. V. Bodenhausen, "Personalized Persuasion," *Psychol. Sci.*, vol. 23, no. 6, pp. 578–581, Jun. 2012.

[85] S. Ravikumar, R. Balakrishnan, and S. Kambhampati, "Ranking tweets considering trust and relevance," in *Proceedings of the Ninth International Workshop on Information Integration on the Web - IIWeb '12*, 2012, pp. 1–4.

[86] A. Appelman and S. S. Sundar, "Measuring Message Credibility," *Journal. Mass Commun. Q.*, vol. 93, no. 1, pp. 59–79, Mar. 2016.

[87] M. Pennacchiotti and A.-M. Popescu, "A Machine Learning Approach to Twitter User Classification.," *Icwsm*, pp. 281–288, 2011.

[88] A. Z. Broder, M. S. Glassman, Steven C. Manasse, and G. Zweig, "Syntactic clustering of the web," *Comput. Networks ISDN Syst.*, vol. 8, no. 29, pp. 1157–1166, 1997.

[89] G. Salton, A. Wong, and C. S. Yang, "A vector space model for automatic indexing," *Commun. ACM*, vol. 18, no. 11, pp. 613–620, Nov. 1975.

[90] R. Mihalcea and P. Tarau, "TextRank: Bringing order into texts," *Proc. EMNLP*, vol. 4,

no. 4, pp. 404–411, 2004.

[91] T. Vu and V. Perez, "Interest mining from user tweets," in *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management - CIKM '13*, 2013, pp. 1869–1872.

[92] O. Medelyan, V. Perrone, and I. H. Witten, "Subject metadata support powered by Maui," in *Proceedings of the 10th annual joint conference on Digital libraries - JCDL '10*, 2010, p. 407.

[93] R. N. S. U. D. R. Stefano Faralli, "A Java Framework for Multilingual Definition and Hypernym Extraction."

[94] "uClassify." .

[95] C. Tan, E. Gabrilovich, and B. Pang, "To Each His Own : Personalized Content Selection based on Text Comprehensibility," in *WSDM*, 2012.

[96] G. H. M. Laughlin, "SMOG Grading-a New Readability Formula," *Wiley*. [Online]. Available: http://www.jstor.org/stable/40011226?seq=1#page_scan_tab_contents. [Accessed: 13-Feb-2016].

[97] G. Salton and M. Mcgill, *Introduction to Modern Information Retrieval*. New York: McGraw-Hill, 1986.

[98] A. G. Alberto Díaz, Pablo Gervás, "Evaluation of a System for Personalized Summarization of Web Contents," *User Model. 2005*, vol. 3538, pp. 453–462, 2005.

[99] R. A. García-hernández, R. Montiel, and Y. Ledeneva, "Text Summarization by Sentence Extraction Using Unsupervised Learning *," *Work*, pp. 133–143, 2008.

[100] S. B. and L. Page, "The Anatomy of a Search Engine," *Comput. Networks ISDN Syst.*, no. 30, pp. 1–7, 1998.

[101] B. Sternthal, D. Ruby, and C. Leavit, “The Persuasive Effect of Source Credibility: Tests of Cognitive Response,” *J. Consum. Res.*, vol. 4, no. 4, pp. 252–260, 1978.

[102] H. L. Dreyfus and S. E. Dreyfus, *Mind over Machine: the power of human intuition and expertise in the age of the computer*. Basil: Blackwell, 1986.

[103] N. Dai and B. D. Davison, “Freshness matters: in flowers, food, and web authority,” in *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval - SIGIR ’10*, 2010, p. 114.

[104] M.-L. Germain and D. McGuire, “The Role of Swift Trust in Virtual Teams and Implications for Human Resource Development,” *Adv. Dev. Hum. Resour.*, vol. 16, no. 3, pp. 356–370, Aug. 2014.

[105] J. Sabater and C. Sierra, “REGRET: reputation in gregarious societies,” in *Proceedings of the fifth international conference on Autonomous agents - AGENTS ’01*, 2001, pp. 194–195.

[106] M. Tsytsarau and T. Palpanas, “Survey on mining subjective data on the web,” *Data Min. Knowl. Discov.*, vol. 24, no. 3, pp. 478–514, May 2012.

[107] K. J. Ferreira, D. Simchi-Levi, and H. Wang, “Online Network Revenue Management using Thompson Sampling.”

[108] J. Nielsen and R. Molich, “Heuristic evaluation of user interfaces,” in *Proceedings of the SIGCHI conference on Human factors in computing systems Empowering people - CHI ’90*, 1990, pp. 249–256.

[109] J. Nielsen, R. L. Mack, C. Lewis, and P. Polson, *Usability inspection methods*. Wiley, 1994.

[110] “The Inverted Pyramid Structure,” *Purdue Online Writing Lab*. [Online]. Available:

https://owl.purdue.edu/owl/purdue_owl.html.

[111] M. Brehmer, B. Lee, B. Bach, N. H. Riche, and T. Munzner, "Timelines Revisited: A Design Space and Considerations for Expressive Storytelling," *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 9, pp. 2151–2164, Sep. 2017.

[112] G. M. Sinatra, C. M. Kardash, G. Taasoobshirazi, and D. Lombardi, "Promoting attitude change and expressed willingness to take action toward climate change in college students," *Instr. Sci.*, vol. 40, no. 1, pp. 1–17, 2012.

[113] C. Hynd, "Conceptual change in response to persuasive messages," in *Intentional Conceptual Change*, 2003, pp. 291–315.

[114] P. K. Murphy, "Teaching as Persuasion: A New Metaphor for a New Decade," *Theory Pract.*, vol. 40, no. 4, pp. 224–227, Nov. 2001.

[115] D. Lombardi and G. M. Sinatra, "College Students' Perceptions About the Plausibility of Human-Induced Climate Change," *Res. Sci. Educ.*, vol. 42, no. 2, pp. 201–217, Apr. 2012.

[116] R. L. Goldstone and Y. Sakamoto, "The transfer of abstract principles governing complex adaptive systems," *Cogn. Psychol.*, no. 46, pp. 414–466, 2003.

[117] P. E. Stoknes, *What we think about when we try not to think about global warming : toward a new psychology of climate action*. Chelsea Green Publishing, 2015.

[118] J. Weng, E.-P. Lim, J. Jiang, and Q. He, "TwitterRank: Finding Topic-sensitive Influential Twitterers," in *WSDM*, 2010.

[119] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," *Knowledge-Based Syst.*, vol. 46, pp. 109–132, Jul. 2013.

[120] D. Ausubel, "The psychology of meaningful verbal learning," Oxford, England, 1963.

[121] D. C. Rubin and A. E. Wenzel, "One hundred years of forgetting: A quantitative

description of retention.," *Psychol. Rev.*, vol. 103, no. 4, pp. 734–760, 1996.

[122] G. Cameron, "Does Publicity Outperform Advertising? An Experimental Test of the Third-Party Endorsement," *J. Public Relations Res.*, vol. 6, no. 3, p. 185, 1994.

[123] J. Liebowitz and M. Frank, *Knowledge Management and E-Learning*. CRC Press, 2010.

[124] A. Shaddock and M. Carroll, "Influences on metamemory judgements," *Aust. J. Psychol.*, vol. 49, no. 1, pp. 21–27, Apr. 1997.

[125] A. G. Bluman, *Elementary statistics : a step by step approach*, 10th ed. McGraw-Hill Education, 2017.

[126] R. DeVellis, *Scale Development Theory and Application*, Fourth Edition. 2016.

[127] M. J. CHAMBLISS and R. GARNER, "Do Adults Change their Minds after Reading Persuasive Text?," *Writ. Commun.*, vol. 13, no. 3, pp. 291–313, Jul. 1996.

[128] R. McCarney, J. Warner, S. Iliffe, R. van Haselen, M. Griffin, and P. Fisher, "The Hawthorne Effect: a randomised, controlled trial.," *BMC Med. Res. Methodol.*, vol. 7, p. 30, Jul. 2007.

[129] M. T. Orne, "On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications," *Am. Psychol.*, vol. 1, no. 145, pp. 42–73, 1962.

[130] D. WATSON, "Correcting for Acquiescent Response Bias in the Absence of a Balanced Scale," *Sociol. Methods Res.*, vol. 21, no. 1, pp. 52–88, Aug. 1992.

[131] A. J. Nederhof, "Methods of coping with social desirability bias: A review," *Eur. J. Soc. Psychol.*, vol. 15, no. 3, pp. 263–280, Jul. 1985.

[132] Apache Software Foundation, "Apache Solr." Forest Hill, Maryland, USA, 1999.

APPENDICES

APPENDIX I: QUESTIONNAIRE 1 (A PRETEST ON CLIMATE CHANGE)

On scale of 1 to 5, where "1" means "Strongly Disagree" and "5" means "Strongly Agree"

	Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree
1. I believe that there is evidence of global climate change.	1	2	3	4	5
2. I believe that things I do, have no effect on the quality of the environment.	1	2	3	4	5
3. I believe most of the concerns about environmental problems have been exaggerated.	1	2	3	4	5

APPENDIX II: QUESTIONNAIRE 2 (MEASURING IMMEDIATE PERSUASION EFFECT)

On scale of 1 to 5, where "1" means "Strongly Disagree" and "5" means "Strongly Agree"

	Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree
1. I believe that there is evidence of global climate change.	1	2	3	4	5
2. I believe that things I do, have no effect on the quality of the environment.	1	2	3	4	5
3. I believe most of the concerns about environmental problems have been exaggerated.	1	2	3	4	5
4. After reading the provided text, I am persuaded to take action against climate change.	1	2	3	4	5
5. I found the content interesting.	1	2	3	4	5
6. I found the content enjoyable.	1	2	3	4	5
7. I found the content relevant to me.	1	2	3	4	5
8. I found the content clear and easy to follow.	1	2	3	4	5
9. I found the content's word and sentence structure complex.	1	2	3	4	5
10. I found the content easy to understand and comprehend.	1	2	3	4	5
11. I found the content trustworthy.	1	2	3	4	5
12. I found the content accurate.	1	2	3	4	5
13. I found the content authentic.	1	2	3	4	5
14. I found the content believable.	1	2	3	4	5
15. Did the system recommend Climate Change 101 with Bill Nye | National Geographic?				Yes	No

APPENDIX III: QUESTIONNAIRE 3 (MEASURING LONG-TERM PERSUASION EFFECT)

On scale of 1 to 5, where "1" means "Strongly Disagree" and "5" means "Strongly Agree"

	Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree
1. I believe that there is evidence of global climate change.	1	2	3	4	5
2. I believe that things I do, have no effect on the quality of the environment.	1	2	3	4	5
3. I believe most of the concerns about environmental problems have been exaggerated.	1	2	3	4	5
4. After reading the provided text from the experiment (two weeks ago), I am persuaded to take action against climate change.	1	2	3	4	5

APPENDIX VI: ETHIC CLEARANCE FORM

Carleton UNIVERSITY
Canada's Capital University

Office of Research Ethics and Compliance
5110 Human Computer Interaction Bldg | 1125 Colonel By Drive
| Ottawa, Ontario K1S 5B6
613-520-2600 Ext: 4085
ethics@carleton.ca

CERTIFICATION OF INSTITUTIONAL ETHICS CLEARANCE

The Carleton University Research Ethics Board-B (CUREB-B) has granted ethics clearance for the research project described below and research may now proceed. CUREB-B is constituted and operates in compliance with the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans* (TCPS2).

Ethics Protocol Clearance ID: Project # 108306

Research Team: Amirsam Khataei (Primary Investigator)
Ali Arya (Research Supervisor)
Dr. Michael J. Hine (Research Support)

Project Title: Generating Persuasive Content through a Software Framework

Funding Source (If applicable):

Effective: **January 31, 2018** Expires: **January 31, 2019**.

Restrictions:

This certification is subject to the following conditions:

1. Clearance is granted only for the research and purposes described in the application.
2. Any modification to the approved research must be submitted to CUREB-B via a Change to Protocol Form. All changes must be cleared prior to the continuance of the research.
3. An Annual Status Report for the renewal of ethics clearance must be submitted and cleared by the renewal date listed above. Failure to submit the Annual Status Report will result in the closure of the file.If funding is associated, funds will be frozen.
4. A closure request must be sent to CUREB-B when the research is complete or terminated.
5. Should any participant suffer adversely from their participation in the project you are required to report the matter to CUREB-B.

Failure to conduct the research in accordance with the principles of the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans 2ndedition* and the *Carleton University Policies and*

Procedures for the Ethical Conduct of Research may result in the suspension or termination of the research project.

Upon reasonable request, it is the policy of CUREB, for cleared protocols, to release the name of the PI, the title of the project, and the date of clearance and any renewal(s).

Please contact the Research Compliance Coordinators, at ethics@carleton.ca, if you have any questions or require a clearance certificate with a signature.

CLEARED BY: **Date: January 31, 2018**

[redacted]

Andy Adler, PhD, Chair, CUREB-B

[redacted]

Bernadette Campbell, PhD, Vice-Chair, CUREB-B

APPENDIX V: CALL FOR PARTICIPANT SCRIPT

Carleton UNIVERSITY
Canada's Capital University

Call for Research Participation(REB Clearance # 108306)

Participants are required for a research project entitled:

Generating persuasive content through a software framework

The research is being conducted by:

- Dr. Ali Arya, Associate Professor, School of Information Technology
- Dr. Mike Hine, Associate Professor, Sprott School of Business
- Amirsam Khataei, Ph.D. Candidate, Department of Systems and Computer Engineering

The purpose of this user test is to evaluate the effectiveness of Persuasive Content Generator (PCG) system in creating awareness about climate change

Participants can be in the age range of 18-50, speak fluent English and must have a Twitter account. If you're interested in taking part in the study, please visit: https://pcg-app.herokuapp.com. Participants will receive a $10 Amazon gift card after completing three questionnaires.

The study starts with the consent form and a short online questionnaire (Questionnaire #1) about participants' opinion and attitude toward climate change. Participants are asked first to read the consent form before completing Questionnaire#1.

After submitting the questionnaire, participants are asked to login to the PCG system using their Twitter account. The system (PCG) builds user profiles by collecting the first 100 recent tweets from each Twitter account. The data is later used to recommend/personalize content on climate change based on the user profiles. The climate change content is generated by the system or may be generic content from a major publisher. After reading the content the participants will be provided a second questionnaire (Questionnaire #2) by the system. Questionnaire #2 is about structure of the content as well as user opinion and attitude toward climate change. Please note that participant data are only cached temporarily and will not be stored in the system.

The process of completing Questionnaire #1, reading the content on climate change, and answering Questionnaire #2 should take approximately 15 minutes. However, there is no hard limit on how much time participants have to spend on the experiment.

Participants are asked to respond a follow-up questionnaire in two weeks from the date of the study. The follow-up questionnaire (Questionnaire #3) will be sent to them through email.

If you have questions or concerns, please contact Amirsam Khataei via email: amirsam.khataei@carleton.ca.

Participants can request to destroy their responses and withdraw from the experiment at any time for any reason before, during or up to 1 week after the study by contacting the researcher through the above email. Gift card will not be given in cases of withdrawal.

APPENDIX VI: CONSENT FORM

Carleton UNIVERSITY
Canada's Capital University

Participant Informed Consent Form (REB Clearance # 108306)
You have been solicited as a research participant for our project entitled:
Generating persuasive content through a software framework
The research is being conducted by:

- Dr. Ali Arya, Associate Professor, School of Information Technology
- Dr. Mike Hine, Associate Professor, School of Business
- Amirsam Khataei, Ph.D. Candidate, Department of Systems and Computer Engineering

Purpose
The purpose of this user test is to evaluate the effectiveness of the Persuasive Content Generator (PCG) system in creating awareness about climate change.
Participants can be in the age range of 18-50, speak fluent English and must have a Twitter account.
Task description
After reading the consent form, the study starts with a short online questionnaire (Questionnaire #1) about participants' opinion and attitude toward climate change. Participants are asked to sign the consent by entering their email address in Questionnaire#1, if they agree with the terms of the study. Participants can withdraw from the experiment at any time for any reason before, during, or up to 1 week after the experiment by contacting the researcher through amirsam.khataei@carleton.ca. Please note that once data collection is done, we don't keep participants' email address.
After completing Questionnaire #1, participants are asked to login to the PCG system using their Twitter account. The system builds user profiles by collecting the first 100 recent tweets from each Twitter account. The data is later used to recommend/personalize content on climate change based on the user profiles. The content is generated by the PCG system or it can be generic content from a major publisher. Once participants logon to their Twitter account, the system randomly assign the participants to PCG or no-PCG group. Participants from PCG group are asked to read/review the system generated content while no-PCG group participants are asked to read/review the generic content.
Participants are asked to complete an online questionnaire (Questionnaire #2) on climate change after reading an article about fighting climate change. The process of completing Questionnaire #1, reading the content on climate change, and answering Questionnaire #2 should take approximately 15 minutes. However, there is no hard limit on how much time participants have to spend on the experiment.
While the system does not record further information from the no-PCG group participants, the system creates use profiles for PCG group participants by capturing the essential information needed by the system from their Twitter accounts. Participants' name and contact information will not be stored in the system and will not appear in any final research project artefacts. The data from Twitter accounts is only cached temporarily and will not be stored in the system. Surveys responses will also be destroyed, after completing data collection. Reports and presentations will include anonymized data only. At the end of the second questionnaire,

participants are notified that they are asked to complete a follow-up questionnaire in two weeks. The follow-up questionnaire (Questionnaire #3) will be sent to the participant via email.

Dissemination

The result of this research will be used in Amirsam Khataei's thesis and may be published and presented in conferences, journals, as well as grant applications.

Anonymity/Confidentiality

Participants' name and contact information will not appear in the final research project. The collected data from Twitter accounts are only cached temporarily on Salesforce Heroku server located in the United States and will not be stored. After generating the personalized content, the system will clean the cache. Your responses to questionnaires will be stored and protected by Microsoft Forms which stores data on servers located in the United States; the data may be disclosed via a court order or data breach. Survey responses will also be destroyed, after completing data collection. Reports and presentations will include anonymized data only. Please note that, only aggregate/not attributable data will be disseminated in publications/ conferences.

For information on Microsoft Forms security policy:https://support.office.com/en-us/article/security-and-privacy-in-microsoft-forms-7e57f9ba-4aeb-4b1b-9e21-b75318532cd9

For information on Salesforce Heroku security policy: https://www.heroku.com/policy/security

Risk

There are no physical or psychological risks that are associated with this experiment.

Right to Withdraw

As a participant, you can withdraw at any time for any reason before, during or up to 1 week after the experiment by contacting the researcher through email. The researcher's email address is given in the recruitment advertisement and all three questionnaires. You can send an email to the researcher to submit their formal withdraw request. If you withdraw, your data is destroyed and not used. The gift card will not be given in cases of withdrawal.

Compensation

You receive our gratitude and will have the opportunity to interact with PCG prototype. By completing all three questionnaires you will receive a $10 Amazon gift card. We will use the provided email address from the questionnaire to send you the gift card. The gift card will not be given in cases of withdrawal.

Research Ethics Board:

This research has been reviewed and cleared by the Carleton University Research Ethics Board (REB) and questions and concerns can be addressed to the REB chair.

Should you have any ethical concerns with the study, please contact Dr. Andy Adler, Chair, Carleton University Research Ethics Board-B (by phone: 613-520-2600 ext. 4085 or by email: ethics@carleton.ca).

If you agree to participate, please complete, Pretest on Climate Change questionnaire (Questionnaire #1) by opening the following link:
https://forms.office.com/Pages/ResponsePage.aspx?id=lRjZagbeXki8UfzhJsyFMN6z3ZnOHNRMogfuyktXQqNURjFLQzdWRTVSMEdFMDFNOUYwNVRYQkk2Ti4u

Make sure to register for the evaluation by entering your email address at the end of the questionnaire so you will qualify for the gift card.

APPENDIX VII: FOLLOW-UP EMAIL TEMPLATE

Carleton UNIVERSITY
Canada's Capital University

REB Clearance # 108306

Automating the process of creating persuasive content
The research is being conducted by:

- Dr. Ali Arya, Associate Professor, School of Information Technology
- Dr. Mike Hine, Associate Professor, School of Business
- Amirsam Khataei, Ph.D. Candidate, Department of Systems and Computer Engineering

Thanks for participating in our research. As the final step, please complete a short questionnaire about your opinion/attitude toward climate change. The process of completing Questionnaire #3 should take less than 5 minutes approximately. Make sure to use the exact same email address that you used for the previous questionnaires. The questionnaire needs to be completed to receive the $10 Amazon gift card.

Follow up questionnaire about climate change:
https://forms.office.com/Pages/ResponsePage.aspx?id=lRjZagbeXki8UfzhJsyFMN6z3ZnOHNRMogfuyktXQqNUMjUyQk1CSFJRTkFJN0xHWERLM0lINjZZWC4u

Cheers
Amirsam Khataei