Inflo: A Revised System for Online News Sharing and Interaction

by

Nicholas Sukiennik

A Thesis Submitted to
The Hong Kong University of Science and Technology
in Partial Fulfillment of the Requirements for
the Degree of Master of Philosophy
in Technology Leadership and Entrepreneurship

December 2019, Hong Kong

Authorization

I hereby declare that I am the sole author of the thesis.

I authorize the Hong Kong University of Science and Technology to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize the Hong Kong University of Science and Technology to reproduce the thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Nicholas Sukiennik

Inflo: A Revised System for Online News Sharing and Interaction

by

Nicholas Sukiennik

This is to certify that I have examined the above MPhil thesis
and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by
the thesis examination committee have been made.

Prof. Pan Hui, Thesis Supervisor
Department of Computer Science and Engineering

Prof. Chi Ying Tsui, Division Head
Department of Integrative Systems and Design

Division of Integrative Systems and Design, HKUST

December 2019

Acknowledgements

I would like to thank Prof. Pan Hui for his supervision. Prof. Terrence Yee for his guidance and support. Dr. Tristan Braud for early feedback on the UI and concept. Ehsan for the original backend methods file which we are still building off of to this day. Labmates for their companionship and support. Michael for moral support and business advice from the beginning. Michelle for introducing this life-changing opportunity to me. Friends and family who have been supportive from far or near, with varying degrees of knowledge about my whereabouts or goings-on…

Finally, thanks to those who contributed their time and skills to building the platform. Your work will always be appreciated.

Table of Contents

List of Figures

List of Tables

Inflo: A Revised System for Online News Sharing and Interaction

by

Nicholas Sukiennik

Division of Integrative Systems and Design

The Hong Kong University of Science and Technology

Abstract

Inflo is a cross between a news aggregator and social media platform which proposes a comprehensive revision to the way in which news is shared and accessed online by addressing the root of the problem. The underlying hypothesis is that an over-reliance on ads for revenue has driven current platforms to a voracious quest for engagement, which has in-turn favored content with attention grabbing headlines over that which provides informational value. We argue that such platforms leverage the provocative nature of sensationalized content and polarized comments to achieve their business goals, while both users and publishers are left with the short end of the stick; users with a deficiency of quality information, and publishers with dwindling revenue streams and less incentive to maintain high journalistic standards. Inflo addresses these issues by proposing mechanisms which improve the dynamics of news propagation on social media, and leverages a "Netflix for news" model which is set to benefit publishers in an unprecedented way. It is the goal of the proposed platform to bring about a shift in incentives of the publishers and social media users in a way which would ultimately have a major effect in curbing the spread of misinformation. Also described in this thesis are three technical modules which serve as the core intellectual property of the platform: 1) a machine learning based news content classification system, 2) a crowdsourced iterative bias elicitation algorithm, and 3) a gamification system to reward user contributed insights.

Chapter 1: Introduction

There are currently more than 3 billion people using social media worldwide [1] and up to 60 percent of them use social media platforms as a source for news [2]. However, such platforms were not originally intended for the purpose of sharing and consumption of news content. As such, there have been several instances of these platforms contributing to the spread false news and misinformation in recent years, all of which have led to a slew of far reaching negative consequences off-platform. These include, most notably, controversies regarding the role of "fake-news" (misinformation and disinformation) to skew the outcomes of the US presidential election of 2016 [3] and the leading up to the Brexit referendum [4], as well as more recent cases of Facebook being used to spread extremist sentiment in Myanmar during the Rohingya Crisis [5], and in Sri Lanka following the 2019 Easter Sunday Bombings [6], among other instances [7]. Such cases show the potential of online platforms in creating and provoking real-world conflict and disorder. From these instances, it is clear that social media platforms play a pivotal role in political and civil discussion and awareness and can have far reaching social consequences.

It is the hypothesis of this thesis and platform described herein that the current social media platforms used for news (e.g. including but not limited to Facebook, the one with the largest market share [8] and Twitter, the one with the second largest proportion of users engaging with news content [9] are insufficient as tools for propagating news content and that it would be in the interest of general public dialogue that such use-case be delegated to a new platform, designed specifically for news content. Such a shift would serve to eliminate the aforementioned negative externalities caused by existing platforms having not been designed for the purpose of sharing and accessing news, including the problem of misinformation as one of the side effects. This thesis will delve deeply into each of the factors which make existing platforms insufficient, in both their business models and feature offerings, and contribute a framework for a revised platform which improves upon each of those factors to create a more effective social media platform for news. This framework corresponds with the web application of the same name which is intended to validate the hypothesis of the revisions proposed therein. The first section of this thesis will describe the background and introduction to such framework, followed by the technical components in section 2, including both core IP and non-IP components. Finally, the third part of this thesis will propose a business plan and monetization model intended to leverage the described application into a startup entity, complete with forecasted costs and revenue projections, in total serving as a comprehensive analysis feasibility of the startup in becoming a sustainable business.

1.1 Diagnosing the Problems with Existing Social Media Platforms

In order to create a revised social platform for news, we first have to diagnose the problems with the existing ones. That way those specific problems can be addressed, leading to a solution at the root of the issue, rather than just alleviating the symptoms.

There are a few major drivers of the problems with social platforms which have, in the long run, led to the online news ecosystem being diluted with unreliable content. One of the major and overarching problems is a reliance on advertisements as a primary means of revenue.

The reason relying on ads leads to a decrease in reliability of content is that they cause those platforms to prioritize user engagement above all else. This results in a cycle wherein reliability and content quality are ignored for the sake of profit:

1. In order to maximize ad revenue, platforms gather user data in order to present relevant ads;
2. in order to gather maximum data from users, they have to maximize engagement (i.e. user clicks: likes, comments, shares);
3. and finally, in order to maximize engagement, they present content which users are most likely to engage with, which tends to be those which are of low factual integrity: clickbait, or sensationalized, polarizing, and conspiratorial content.

It is thus clear that the reliance on ads drives low quality content, and as a result have brought about a lack of trust towards the news industry [10]. This same quest for engagement can be broken down in terms of the specific features that platforms use as tools to gather as much data from users as possible. These features and their role in propagating unreliable content are described below.

1.1.1 The Algorithmic Feed

Facebook's newsfeed recommendation algorithm (or equivalent in other platforms e.g. Twitter, Snapchat, Youtube, Pinterest; but we will use Facebook as the archetype of this trend for the purpose of this thesis) is tailored to each user, based on data of that user's interests, and promotes content it assumes the user will like, and then engage with, thus propagating the cycle of targeted advertising and revenue growth. The side effect of such recommendation algorithm is that it usually puts polarized or sensationalized content at the top of one's feed, which is valued for its provocative, as opposed to factually accurate, characteristics. The result of this is clearly good for Facebook, as it garners more insight into a user's interests, but at the same time, damages the integrity of the online news ecosystem, and dilutes it with low quality, unreliable content. This, in turn, provokes polarized views [11] and can create filter bubbles [12] which in turn can lead to an echo-chamber effect [13]. The result is that users who spend a lot of time on such social media platforms often end up being confirmed in their pre-existing biases, despite such biases often being provoked by misleading, inaccurate, or polarized information [14].

Prominent media researcher Pablo Boczkowski sums up the major flaw of social media platforms: "their reliance on algorithmic curation has endowed these procedures with a certain opacity that makes it even more difficult for the public to come up with strategies that successfully identify bias" [15].

1.1.2 The Like-driven Commenting System

Most online social platforms use a very naive system of user interaction: comments and likes. In the case of shared news, the comments thereupon show a tendency for people to simply like those which they most agree with, regardless of whether they provide any value to the community. Liked comments then feature more prominently to users to later view the respective posts. This creates a cycle wherein polarized comments are more quickly propagated than those comments which contribute meaningful information, including those which contain inaccurate information [16]

These two features, which are common to most current social platforms, have given rise to an ecosystem which thrives off of extremism and polarization. They have been optimized for the profit of corporations as opposed to the well-being (i.e. education and awareness) of the public. They have led to echo-chambers and filter bubbles which have had serious real-world consequences.

From this we conclude that the presence of ads sets off a race to get the most data, or engagement, and is a recipe for a toxic online information ecosystem, firstly because it has given rise to an echo-chamber inducing newsfeed, and secondly because it encourages polarizing comments among the user community. As such, a motivating factor for Inflo is the principle that a platform which does not host ads, does not use a recommendation algorithm to present content of interest, nor does it feature a like-driven commenting system, can create a better ecosystem for sharing accurate information and meaningful insights. If this hypothesis is correct, Inflo will be able to revise the social media paradigm and facilitate a healthy content news consumption and discussion environment.

The specific feature revisions are defined in the next section.

1.2 The Inflo Approach

Inflo is a news sharing and discussion platform which is designed for the informational benefit of its users, not the financial benefit of its advertisers. This foundational principle leads to a natural re-evaluation of the features we take for granted in contemporary social media platforms.

Inflo presents two major features in revision to existing platforms used for the propagation of news content: 1) a filterable feed powered by the latest machine learning technologies—giving more control to users over what kind of content they would like to see, and 2) a gamification system to incentivize quality discussion and interaction within the online community of users.

Additionally, Inflo addresses the need of more transparency of the credibility of sources and users on social media platforms to prevent the spread of misleading content and misinformation. To fulfill this need, we present a novel crowdsourced bias elicitation feature which will serve to place accountability on both the individuals and organizations which are responsible for conveying information online.

Each of these features is described in detail below.

1.2.1 A Personalized (Not Presumptuous) Content Feed

Inflo features a filterable (non-algorithmically driven) feed, which allows the user to curate the content in their feed according to that which is relevant to them, doing away with the recommendation system employed by most conventional social media. To make such feature possible, each piece of content shared within the platform will be analyzed and labeled using our proprietary news classification algorithm (see Section 2.1.1). The Inflo filtering system aims to revise the paradigm of social content discovery by labelling the content and allow users to choose what topics they want to see according to interests and mood at a given time, as opposed to making assumptions about what the user wants to see and feeding them content accordingly. Such recommendation-based feeds have led to existing platforms being exploited by third-parties attempting to spread false information, leading to an array of negative consequences (see Section 1.1.1). Thus, with the current approach, we intend to create a new paradigm of social content sharing which does not allow for such exploitation. Design decisions such as the filterable (non-algorithmic) feed, and a lack of advertisements are what could prevent such exploitation.

1.2.2 Impact System - a Non-like-driven System for User Interaction

The impact system is a framework for rewarding users for offering insights to the community, and has been designed in opposition to the concept of a liking system, which only reward users for "popularity" which often does not correlate with informational accuracy. This system will also quantify a user's expertise by keeping track of which topics they tend to post under most, thereby applying a weight to their contributions in that category. The goal of the system is to cultivate a self-regulating community which rewards quality of interaction, rather than virality and polarized comments. Ultimately, it will be the key factor in making Inflo a context hub, where people can go to get accurate and reliable information surrounding the articles they read, as opposed to a typical social media platform focused on superficial engagement.

Additionally, the impact system will serve to incentivize user retention by gamifying the activities of sharing and commenting, as well as drive the monetization mechanism which creates value from user insights (see "Revenue Model" below").

1.2.3 The Need for Quantifying Bias

One key method to eliminate low quality and misinformation is to detect and eliminate that which is extremely biased. Inflo will feature an innovative tool to quantify bias of news articles and that of users within the community. In line with the goal of aggregating context, Inflo's use of a bias evaluation tool can allow members of the community to inform their own opinion as to how credible an article may be, thereby mitigating spread of potentially unreliable information, or adoption of fringe views thereupon leading to a decrease in the overall informational value derived from the platform.

With the features described above, as well as a reconsidered user-interface specially designed to facilitate news (see Section 2.3), Inflo aims to bring about a shift in the status quo of online content sharing, away from the viral and unreliable, and instead toward informative insightful. As a whole, it serves as a major contribution to the ecosystem of online news through carefully designed features for responsible handling of news content, as well as a novel revenue model revenue model which supports those who have a role in the creation and publication of reliable content (see Section 3.1.1).

The framework described has been implemented as a web-application, which by the time of submission of this thesis (December 2019), is deployed as an “alpha” product and can be found at http://www.inflo.news. Note that the “alpha” release is yet an unfinished product and intended for demonstration purposes only. The intended release timeline can be found in Appendix V.

1.3 Intro to the Media Landscape

Inflo has the potential to make significant impact because it takes into consideration the news ecosystem as a whole, addressing it as a system rather than only focusing on one single component of the misinformation problem. In deciding the features and design of the platform, we have kept in consideration the root cause rather than the symptoms (i.e. prevent the spread of fake news by emphasizing informational value, as opposed to detecting and eliminating it). We believe that in order to curb the spread of misinformation, one has to have a full understanding of how *information* spreads, and the behaviors that define such spread. While this is a problem that could be approached from multiple academic fields, the current thesis will focus primarily on the role of human actors in countering the spread of misinformation (i.e. a crowdsourcing approach), as opposed to an automated, AI based approach, with the argument that, while it does have a role, AI is not a panacea, and should be limited only to certain narrow use-cases as a tool in a larger process. There are several research labs and papers which describe the usage of ML for detection of fake news in various contexts, but a majority of them are not universally applicable and only address a very narrow scope of application. While we do not discredit such efforts, it is the goal of the project described herein to develop a user-facing application which is so robust as to be resistant to the fallibility of approaches which rely too strongly on automated solutions, or those which are too theoretical in describing the dynamics at play in the spread of online information making them ineffective for the purpose of a real-world application. To do so, we believe that a user-focused approach is the most feasible.

One thing to note is that Inflo is not intended as a direct solution to the problem of “fake news”, but instead, a platform for news sharing and discussion, which rewards quality of insights over quantity of engagement: this does not entail eliminating “fake-news” which in itself is a complex and loaded term; the platform does not, and will never, bluntly remove content due to the alleged veracity or lack thereof. It is our goal to create an ecosystem which drives context and insights, in effect allowing the users to decide for themselves what is accurate and what is not. Truth, after all, is a subjective term, and we do not believe we have sufficient authority to make that distinction of our own accord. We believe that by rewarding people for a set of objectively constructive behaviors, we can cultivate a self-regulating community wherein accuracy and quality take precedence over virality, clickbait, sensationalism, or polarized content, and as such, will over the long-term, result in a more productive pursuit against misinformation other tools, whether human or AI driven.

Chapter 2: Technology

Inflo consists of some modules which are considered as “core intellectual property”, due to their novelty and difficulty of execution, as well as some modules which are considered just as important yet less technically advanced and are as such defined as “technical foundations”. This Section covers both types.

2.1 Core IP

The core IP and technological foundations for the Inflo platform are composed of three main parts: 1) the automatic classification algorithm, 2) a crowdsourced bias elicitation system for users and articles, and 3) the impact score reputation system.

2.1.1 The Automatic News Content Classification System (ANCCS)

The goal of the automatic classification system is to enable a content navigation experience which does not entail a recommendation algorithm (and therefore does not require a user’s personal data), but still provides a satisfactory level of personalization. To do so, we have developed a proprietary machine learning algorithm which analyzes the text of a news article and outputs respective categories and topics, otherwise referred to as keyphrases. While the current algorithm does successfully output both broad categories as well as more specific keyphrases for a given article, the implementation in the Inflo platform only makes use of the broad categories, due to deployment constraints.

Below, we describe the algorithm pipeline in-depth (this Section has been updated and adapted from a demo paper which has been published on arxiv and written in collaboration with Pranav Agrawal, under the title: “Inflo: News Categorization and Keyphrase Extraction for Implementation in an Aggregation System” [17]). A demo of the demo can be found at https://www.infloproject.com/cat-keyphrase-labeling-demo (the reason it is not a full demo is because the news articles available are not continuously updated, as they would be in the full demo, which can, at the time of writing, only be run on a local machine).

Preliminaries

The pipeline of the algorithm entails 1) analyzing article text to generate an appropriate discrete category for a given news article, and 2) extracting keyphrases with the help of category label and entity-based document frequencies. Results could be used directly in the platform for use in the filtering system (see Section 2.3.3), or for further processing to group similar articles together. First, we explain how and from which sources the datasets were created in order to train the model.

Dataset Collection

Data was collected via several news APIs including New York Times API [18], CNN News API [19] and News API [20]. In total, 500,024 news articles were collecting, published between the years 2000 to 2018. Then, 12 discrete news categories were decided upon, which each article would be grouped under; those categories are as follows: Regional Politics, Sports, Entertainment, International Relations, Science, Business, War and Conflicts, Law and Order, Technology, US, World, and Miscellaneous. There were approx. 20 to 50 thousand articles per category.

In order to train the model, 80% of the articles were apportioned for training, 10% for validation, and 10% for testing. Because different news publications discretize their articles according to a different category distribution, and also because many of the articles downloaded did not have categories associated with them, we manually applied one of the discrete categories defined above to each respective article. This was done by making use of the tags of the articles--which are found in their metadata, and reflect various topics covered in the given article (these tags are most often used for the purpose of being listed in search engine results). The tags were allocated under the most relevant broad category, and articles pertaining to each of those tags could then be migrated easily under the respective categories in turn.

Document Frequency Computation

The first step in classifying text data is to calculate the document frequency [21], which is simply a tabulation of the occurrence of each word. The frequency tabulations can then be used to algorithmically predict classes for given text documents. The two types of document frequencies we have used are 1) phrase based and 2) named-entity based, both of which are category-specific.

This means the frequency of given terms will be different according to its category, which will then influence the results for the following keyphrase extraction tasks.

Entity-based document frequency computation was done using SpaCy [22],an open source python package. First, entities were extracted from each article followed by document frequencies computed on the same set of extracted entities. The result is a separate corpus of entities for each category, enabling the models to incorporate common entities for each given news category.

News Classification

Text classification is one of the central problems in the field of Natural Language Processing (NLP), and Inflo has taken steps towards some significantly novel advancements in the field. Deep learning models, especially CNN (Convolutional Neural Networks) have been very effective for text classification for most general purposes [23] [24]. However, the challenges of these kinds of models is that they typically require massive datasets, computational power, and days to converge. Advances in transfer learning have greatly simplified the task by using a pre-trained model and fine-tuning for customized purposes [25]. A language model is a probabilistic distribution for a sequence of words, and AWD-LSTM (ASGD Weight-Dropped LSTM) has been used as a main architecture for this specific language model [26]. At the time of writing, ULMFiT (Universal Language Model Fine-tuning for Text Classification) is the current state-of-the-art language model for text classification with transfer learning [27].

Using ULMFiT eliminates the requirement of such a large dataset and reduces the required computational power to a reasonable level which could be done on a person computer, ultimately allowing for a significant increase in accuracy over a short period of time. That being said, however, ULMFiT does require fine-tuning of both the language model and classifier according to the target dataset.

The datasets in language models (PTB, WikiText) include “UNK” tokens which are replacements of low frequency based OOV (Out of Vocabulary) tokens. This replacement could hurt domain-specific datasets, if, for example in the case of scientific articles or news, the classification relies on such low frequency OOV entities. As to prevent this loss due to the UNK token, we instead use named entity-specific UNK tokens (e.g. instead of using UNK, we use PERSON-UNK, COUNTRY-UNK, PRODUCT-UNK and so on). Named entities are often a good indicator of trends and could contribute to better classification [28] [29]. For example, a news article pertaining to “lifestyle” could refer to products and an article belonging to “international relations” could refer to countries. As both “product” and “country” are entities, such entity-specific tokens could improve our results. The information regarding named entities is done by preprocessing through SpaCy [22].

We tested this hypothesis on our dataset, with the result as follows: a well-tuned CNN provides 62.4% accuracy; regular ULMFiT (without named-entities) provides 73.5%; and named-entities based ULMFiT provides 77.4% accuracy on the test dataset. The model was deployed using PyTorch [30].

With accuracy far from ideal, the usefulness of the applied categories will be subject to the discretion of users of the Inflo platform, and with feedback, results will be improved and made more appropriate given the use case.

It should be noted, however, that deciding news category for the article is often ambiguous and fuzzy (for example, an article about Apple could be categorized as “business” or “technology”). As such, one of the intents in category labeling is not only to achieve perfect accuracy, but also to facilitate better keyphrase extraction, as described in the next Section.

Keyphrase Extraction

The schematic diagram for keyphrase extraction module is shown in Figure 1. We have used PKE (Python Keyphrase Extraction) library for keyword extraction [31].

It should be noted that vocabulary styles are different across news categories. Thus, it is crucial to have category-specific corpora. Our extraction system is an aggregation of three different methods, each one being described below:

1. Statistically Extracted Keyphrases: This method is meant to extract keyphrases which are determined statistically. KP-Miner serves this purpose by making use of document frequencies, term frequencies, and positional occurrences [32]. Once the article category is defined by the classifier, we use the category-specific document frequencies for better retrieval. The resulting keyphrases were subjectively useful; however, most of the single-word keywords were redundant. Hence, this module will be most effectively used for multi-word keyphrases.
2. Statistically Extracted Entities: The method is meant to extract the entities (like persons, organizations, places) which are central to the article. To this end, we also use KP-Miner,

but on entity-based category-specific document frequencies. After extraction, we de-duplicate using singularization and "nounification". Nounification of a keyword is performed by retrieving a relevant noun in the synset of the corresponding keyword in WordNet [33]. This easily converts words like "elect" to "election" and "Italian" to "Italy".

3. Graphically Extracted Topics: Graphical-based methods for keyword extraction comprises topics as nodes and edges as semantic relations. This method is meant to extract the key-topics which influence other surrounding topics in an article.

The results from each of the above three methods are de-duplicated and can be further used within the Inflo filtering system as tags for content navigation.

Many other unsupervised approaches exist, such as SingleRank [34], TopicRank [35], TopicalPageRank [36], PositionRank [37], and MultipartiteRank [38], among which MultipartiteRank gives the best results for purpose of news. However, MultipartiteRank often returns redundant keyphrases which are eliminated by de-duplication and nounification (our contribution).

In order to be meaningful to the community, as a tool for navigation and content filtering, labels for a given piece of content will cover certain key pieces of information regarding the story in question; namely the location where it took place, the primary people involved, the name of a specific event the story surrounds, and/or any brands, organizations, or products the story may be discussing. Each of these key pieces of information can be extracted from an article via the use of entities, a topic modeling concept. Entities allocate one of the above "types" to each of the keyphrases which is extracted by conventional methods, if appropriate. We can then take the most common entity-associated keyphrase from each type, for each article, and set them as the tags used within the user interface for that article. This method will provide consistency across all articles shared within the platform, and will contribute to an effective user interface convention which makes it easy for the user to find articles of interest, specifically, based on their topics of interest, but not too specific to only cover a single story.

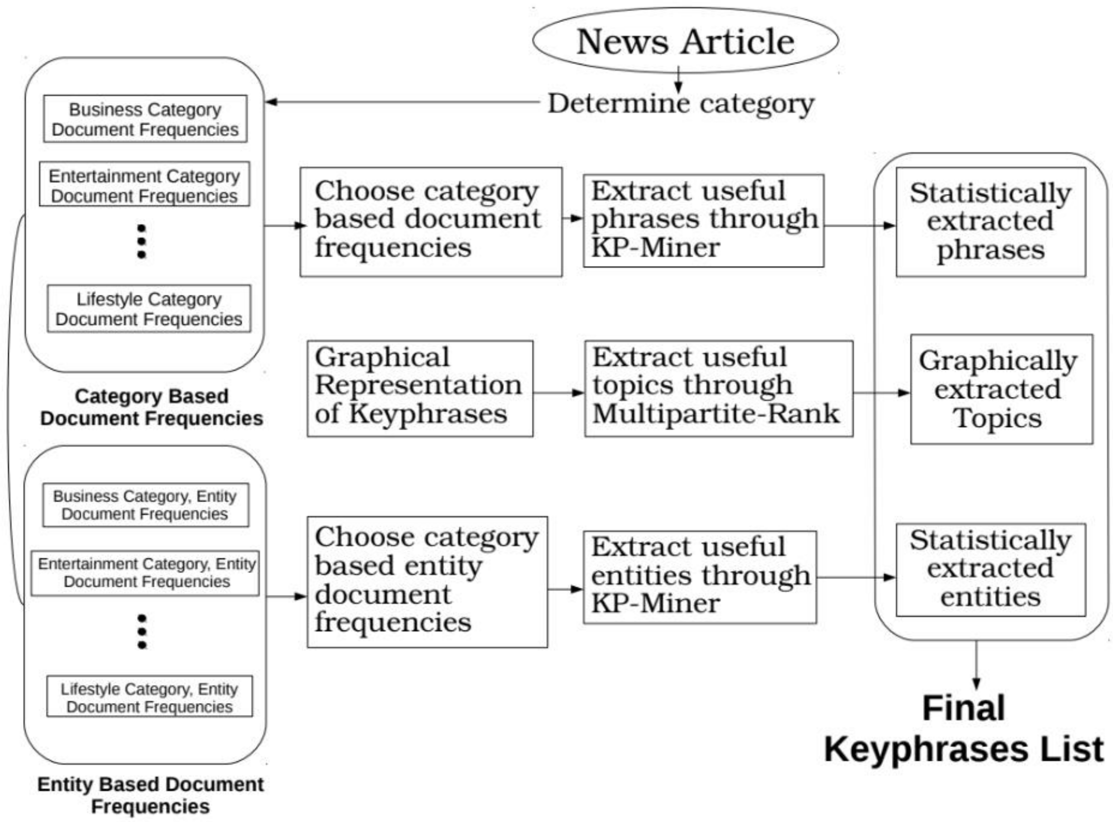


Figure 1: A schematic for comprehensive keyphrase extraction for news articles for implementation within the Inflo platform.

The Inflo automatic news content classification system contributes greatly to the Inflo platform by enabling instantaneous labelling of news articles from any source, which can then be used as part of content navigation tools such as topic and sub-topic filters, making for more efficient discovery of relevant content for each individual user.

This also eliminates the need for a single stream feed of recommended content often used in existing social media platforms, which have been shown to create filter bubbles and echo chamber effect as described in Section 1.1.1.

Hierarchy and Normalization

To process results of the ANCCS for use in the of the Inflo filtering system, there are two more steps that should be done: 1) ensuring that keyphrase results across related articles are exactly the same, and not just similar, allowing them to be displayed under a mutual tag, and 2) creating a hierarchy of tags each with an inductive or deductive relationship to each other. #1 will enable articles covering the same story (defined as a singular incident, event, or otherwise newsworthy phenomenon) but from different publications to be grouped together making it easy for users to take in a wide perspective of the news from several different angles (see Figure 2), while #2 will improve the content navigation experience by creating depth, allowing users to delve into more and more specific topics according to their interests.

Tag Normalization

In order to normalize tags across multiple articles, there are a few approaches we could choose to take.

The most naive (easiest) approach is to match the extracted keywords with a set of pre-set topic-associated keywords, and if the match is present, allocate the corresponding topic to the article. The pre-set topic associated keywords can be generated via latent Dirichlet allocation (LDA) [39], a topic modeling technique. By training the topic model on a set of labeled data (topic given), we can determine the most important keywords for the given topic. Thus, if our keyword extraction results in one or more of those *most important* keywords, there is a very high chance that it will be most closely associated with the topic of the given topic model. In order to be scalable for the purpose of the Inflo platform, we would have to constantly train new topic models as trends in global news and discussion come about. A set of dynamic trends can be generated via A) Twitter's trending tags/topics, B) Wikipedia. We would thus simply have to plug in our topic model with an API from one of these two sources [40] [41], train the LDA model given a corpus of data (e.g. from NewsAPI), and then save the resulting keywords and corresponding topic to the database to be cross checked each time a new article is shared to the platform. When a match is found between the ANCCS and the topic model results, the category will be automatically applied.

A more complex yet more unsupervised approach to the problem of dynamic topic generation is in the form of word vectors. In the field of NLP, a word vector is a numeric mapping to each word in a language model, which is, in turn, related to other, similar terms. While word vectors are often used in the case of machine translation to determine the appropriate word given context, we can also apply them to the given use case in the following way:

A. Compare keyword extraction results from each new article with those from all of the existing articles (assuming no topics are present). Then use a language model to obtain the word vectors for the new and old extracted keywords, and for those which are similar within a narrow margin, apply the most frequent extracted keyword across all similar documents as the topic.
B. For future articles, compare new word vectors which that of previous topics and allocate the topic with the most similar word vector to those of the current documents extracted keyword.
C. This second approach has the advantage of not requiring any labeling nor training to be done in advance, however it will most likely require quite a bit of trial and error before results are completely satisfactory to be deployed publicly.

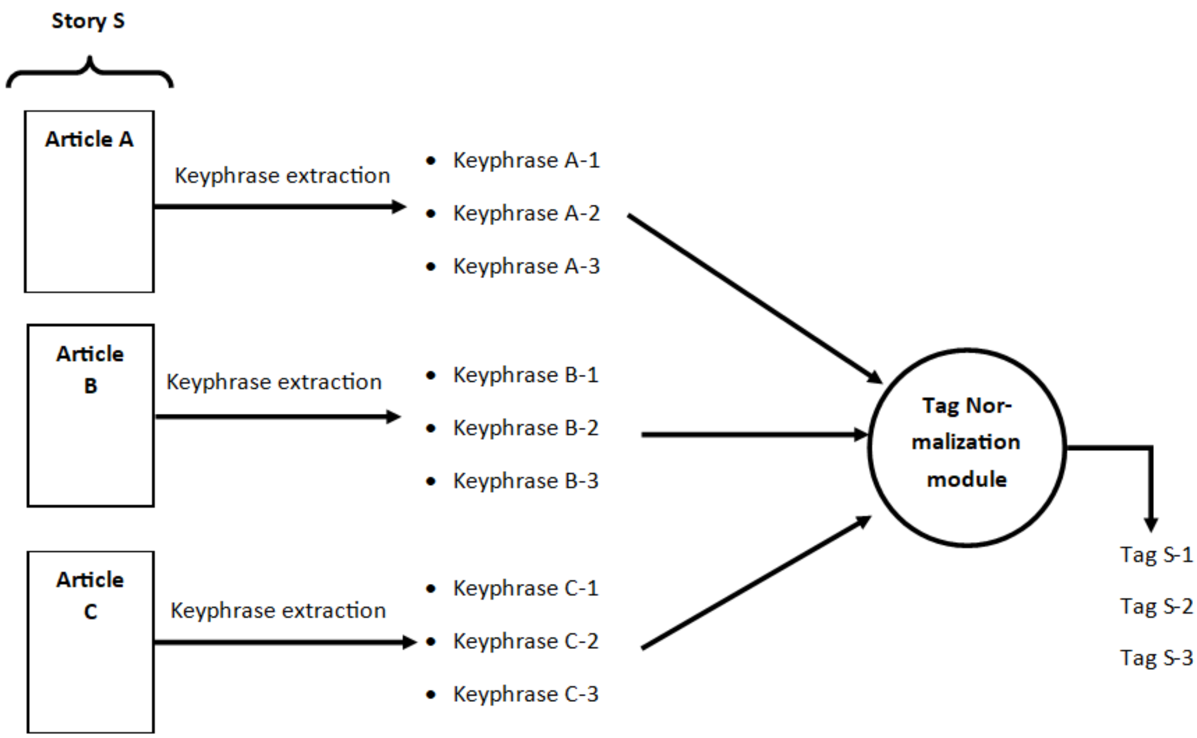


Figure 2: A schematic for normalizing tags across articles on the same story. This will allow them to be grouped together so that users can easily discover articles which are related

Deployment and API Implementation

The category classification component of the ANCCS described above is currently deployed and integrated into the Inflo platform for use within the filtering system (see Section 2.3.3). One minor thing to note is that, in the platform implementation, some categories have been consolidated and renamed, with the following mapping (only those which are different from those above are specified): Law & Order → Law; Regional Politics, International Relations → Politics; War & Conflicts → World. As such there are only a total of ten categories in the filter menu within the platform.

The key innovation we are contributing to the field the ANCCS is the deployment of a machine learning model for real time use in a social network.

The algorithm has been deployed using Render [42], a new web-app deployment platform which expedites the deployment process of deep learning models [43]. It is implemented as an API which accepts article text as a GET request, and returns the category as a JSON response. The response can then be parsed by the web application and implement the classification label into the application UI on the relevant news article preview.

The deployment of deep learning models, especially neural networks, has long been a challenge [44], especially when results are needed in real time to be of practical use, as with the current application. New technologies in transfer learning and cloud computing have made it recently more feasible [45] considering both computational cost and speed of obtaining results. The use of FastAI's ULMFiT language model [27] in combination with the Render deployment is a cutting-edge implementation which could start to revolutionize the way deep learning models are used in consumer-facing products online.

The automatic content classification system is one of the most important aspects of both the functionality and the IP of the Inflo platform. Thus, we plan to expand it with more features in the forthcoming iterations.

One such expansion will be the ability to generate unsupervised keywords/phrases for a given news article, which could then be used to group it with articles of the same topics. This allows for more discrete topic filtering and in turn, a more personalized user browsing experience. These keyphrases will also later be put into a hierarchical structure, so the user can effectively "zoom" in or out of a given topic.

Additionally, we will continue to analyze and extract different qualifying factors of shared articles to improve the filtering system, and allowed users to access more relevant and reliable content. Those factors are listed below along with a preliminary direction as to how such results can be achieved:

1. Bias labels - see Section 2.1.2 for the bias elicitation module.
2. Opinion vs Original Reporting – can be achieved using convolutional neural networks trained on a set of known editorial/opinion based articles [46].
3. Unreliable/Unknown source – can be achieved with a continuously updated database of trusted and non-trusted sources. Non trusted will be labeled as "unreliable" and anything from a source not in either list will be labeled as unknown. Similar lists have been compiled by third-party fact checking organizations, which could be leveraged as a starting point for this feature [47].
4. "Local" news - As described in the section regarding entities, each article will have associated with it a location "entity". Local news usually surrounds a certain city or metropolitan area. To recreate the "local news" feature which is found in existing news aggregators, we will use location services of the user's device to determine their nearest metropolitan area, and by searching the location entities of shared articles put those which match the location of the user in a special tab called "local".

The above features will be in addition to the current filtering options which only consist of categories, and can be expected to be completed by the release of the beta version of the platform (July 2020).

2.1.2 User-Article Bias Elicitation Module

The goal of this module is to use a crowdsourcing approach to accurately elicit the bias of articles from various sources and users' standpoints on those articles and the specific topics they concern.

The method we have devised is specifically made feasible by the Inflo platform in that it relies on: 1) a continuous input of user data, which can be easily collected through a social platform such as this one, and 2) and automatic method for classifying news articles by topic, which is also a core capability of Inflo as described in the previous Section.

The method described here is more generally applicable than any existing bias checking method in the ecosystem, as well as the least corruptible, given that it does not involve individual "checkers", but instead a community of up to several thousands of social media users. By crowdsourcing the stance detection of articles in our platform, we avoid places undue emphasis on individual fact-checkers or third-party agencies.

Via a crowdsourced approach, we hypothesize that we can determine an article's and user's bias through an iterative mechanism for collecting and interpreting user voting data dynamically.

Analysis Procedure and Benchmarking

In order to validate whether such module is feasible, we performed a preliminary analysis of an existing, the results of which could also be used to establish a benchmark for the type of results we can expect to achieve. The full report of such analysis is provided in Appendix I. Once implemented, we will compare the analysis data collected via the Inflo platform with the results of this analysis, and, if consistent, the conclusions of validity can be applied to our module, which includes both the method for collecting new data as well as the process for analysis and bias elicitation.

Implementation

In order to collect data within the platform to elicit the bias, we will present the user with a survey in the Inflo UI each time a user clicks on a news article of known topic (also referred to as "issue"). When the article link it clicked, it will open in a new tab, meaning that when the user finishes reading and comes back to Inflo, the survey will be displayed. Below we explain how the survey is to be presented.

For the purposes of demonstration, we will assume that the user is reading an article on the topic of "gun control".

Survey question 1:

For the given issue, where do *you* stand?

←------------|-------------|-------------|------------|-------------|-------------→

Against gun control **Pro-gun control**

Survey question 2:

For the given issue, where do you objectively think *the given article* stands?

←------------|-------------|-------------|------------|-------------|-------------→

Against gun control **Pro-gun control**

The user will respond by sliding a caret to the point on each spectrum which corresponds with their answer. In sequence, from left to right, each tick represents the following answer: extremely against, moderately against, neutral, strongly pro-, moderately pro-. The user can also select responses in between the labeled ticks; e.g. to the left of the extremely against tick, between the extremely against and moderately against ticks, between the moderately against and neutral ticks, and so forth. The responses will be interpreted numerically and normalized via outlier correction.

The responses should be stored in a database with each entry respective to each user and each article. The data can then be run through the algorithm for bias elicitation and the results displayed within the UI related to the respective user and article.

Significance

The reason current module is meaningful is that this is the first attempt on a large scale to rank the bias of individual news articles, along with their corresponding publication, on a spectrum *per topic, a*s opposed to placing sources or topics per source on a left/right political spectrum, as well as being the first dynamically updating crowdsourced bias classification method implemented into a social platform.

Overall with this module, we aim to allow users to evaluate their news content in a more holistic way, thus eliminating the effects of polarization from reading biased content, and which we expect will, in turn, translate to more insightful user commentary being contributed to the Inflo community.

2.1.3 Impact system

The Inflo impact system (IIS) is a gamification framework for rewarding users of a social network for objective contributions to the community, with the goal of facilitating for constructive interaction within the community, as opposed to the primarily sentiment-based contributions seen on conventional social media platforms [48]. The system will allocate "reputation points" to users for achieving various quantifiable tasks, as well as based input from the community. The rewards framework will be based on principles developed through existing literature in the area of gamification, psychology, and behavioral economics.

Many online communities utilize a gamification system to maintain user retention as well as to incentive quality of interactions [49]. Some such examples are Stack Overflow [50], Quora [51], and Reddit [52]. However, Inflo is the first such online platform which attempts to integrate gamification into a platform based primarily around news content.

To maintain a high rate of user retention, Inflo will create a rewards system which takes into account Flow theory [53], which defines four quadrants of user engagement (anxiety, apathy, boredom, and flow), of which the optimal one is "flow" which describes the state wherein the user is not experiencing either too much challenged (anxious), or too little challenge (boredom). By rewarding users in the flow state, they will be less likely to churn from (i.e. leave) the platform [54].

Furthermore, in order to maintain high quality in user interaction and prevent group manipulation, Inflo takes into account game theory, namely mechanism design and the revelation principle [55], which is a method of aligning a user's incentives with the incentives of the platform, ensuring that each user contributes in a way that is beneficial to the greater community [56].

The impact system, as a first step, will be built off of the voting system, which allows users to up- or down-vote commentaries in relation to given articles. Community guidelines dictate that votes should be given not purely for subjective agreement, but for objective insight that they provide.

Initially, Inflo will use a system of automated fact checks to allocate points to users. Specifically, Inflo will process and analyze commentaries for references, facts, and figures, use state of the art deep learning methods, and if the contents of a commentary are determined to be authentic and verifiable, the user will be rewarded. However, once a user has reached a given threshold, they we will able to propagate the points through endorsement of other commentaries, leading to a cascade effect, where expertise will be organically elicited via a cycle of automated fact-checking and user-driven fact checking.

The IIS is critical to the functioning of the Inflo platform as it effectively serves to quantify the credibility of the respective users. This is, in turn, necessary to apply different admin rights to users (see Appendix II for an example user rights distribution), based on their credibility, and to support the self-regulating nature of the platform. It will also be the starting point for an expertise system, designed to reward users who show a deep level of knowledge about a given topic.

Expertise

When users contribute insights repeatedly under a given theme or topic, they will earn expertise points, which will quantify not only general impact, but specific impact within a field of knowledge. After the user has achieved a certain expertise score in a given topic, they will be considered an "expert" in that topic, which will then unlock specific features only available to experts: for example, experts can endorse commentaries by other users within the given topic of expertise. Endorsements are different from upvotes as they contribute directly to the receiving users impact score, whereas typical upvotes do not. In this way, an expert has more of a role in regulating the community than other typical users (until the latter gain their own expertise). As such, experts will be pivotal in making Inflo into a self-sustaining platform, something similar to "moderators" on Reddit [57].

One thing to note is that the expertise system cannot be entirely user driven in the beginning, as there is no definition of a "base truth" within the platform. A baseline of base truth is necessary to prevent people from manipulating or "gaming" the system through coordinated action and trolling (i.e. give users who don't really deserve it, undue expertise). To avoid this, the first batch of experts will have to be manually chosen based on their off-platform qualifications, by undergoing a vetting process. Such process would take into account professional experience, education, research publications, degrees, or awards in a given field. Once the first batch of experts in a given field are determined, future experts can be created through a cascade effect [58] of expertise allocation, starting with the first one who was artificially selected (i.e. not via the impact system).

It is also worth mentioning the relationship between the impact system and the bias elicitation module (described in the previous section, Section 2.1.2). While the two will not be directly linked, in terms of allocating or detracting points for users who have a particular bias standpoint (e.g. add points for being neutral, detract for extreme views), one could also argue that there is a correlation between presence of bias, and overall reliability. Therefore, with bias standpoints publicly displayed along with a user's impact and expertise scores, other users can determine for themselves how reliable a given commentary is, and allow it to inform their own opinions as appropriate.

Intellectually Backed Currency

The impact system could also be described with the concept of an intellectually backed currency. A currency is built upon the concept of scarcity, and so with scarcity, people will be cautious as to not spend their currency unless they strongly believe it is deserved. Therefore, by making the commodity of an endorsement scarce, the mechanism will eliminate the negative externalities of the liking-system of traditional social media, which are unlimited, and therefore can be given frivolously, meaning that the user who gives them does not have any incentive to preserve them, as they have no value. The impact system will therefore reward users who provide the most value to the community, through the transfer of this currency. The difference between this currency, and say government currency or crypto currencies, is that this "intellectually backed currency" will be given in exchange for knowledge instead of goods and services. One could also consider using blockchain as the underlying technology to maintain the transactions conducted within the impact system.

2.2 Technical foundations

The technical foundations for the software aspect of the platform application itself (not including the core IP described above) consists of three modules: 1) a news article preview generator 2) a front-end user interface, made with React JS [59], and 3) a backend infrastructure for storing and querying data, made with Django [60], a Python web app framework, all of which is deployed using an Amazon Web Services [61] (AWS) E2 instance.

2.2.1 News article preview generator

Content preview are a common element in almost all contemporary social media platforms; that is, when a user chooses to share a link to another website, the social platform they are sharing it to will generate a preview of the content on the linked site. In accordance with this paradigm, as well as being a feature necessary for usability, we have built an in-house preview generator for use within the Inflo platform.

The Inflo news article preview generator is a process for scraping news article data via python, composed of a combination of open source tools as well as specially written python code as necessary, all combined to form a cohesive, robust method for generating article previews in real-time, i.e. whenever a user shares an article. The workflow for article preview generation is described below:

The preview generator is initiated when the user submits a news URL to share to the Inflo network. The URL is passed through a Python function which makes a request to that URL and parses the response to gather the relevant data. Those data are then saved to the database and will later be used to display a preview of the given article in the Inflo "Dashboard" (see Appendix III), and include the following: publication name, publication date, headline, body text, thumbnail image.

The response parsing is done primarily using Newspaper 3k, an open source Python package [62] (headline, source name, publication name, thumbnail); whereas the date is gathered using a tool called "Article Date Extractor". While these tools are not foolproof, they do work for a majority of conventional news websites. For those sites/articles with which these tools are not compatible, customized parsing functions have been developed.

Also included as part of this module is a validator which determines whether the URL submitted is a news article or not (from newspaper 3k package). If not, it will not be posted to the dashboard, with an error message displayed to the user saying: "this content is not currently supported". This is done so as to prevent spam and adult content from being shared on the platform, and to keep content focused around theme of news, as opposed to other forms of written web content.

In future iterations, we will add the capability to handle non-news content: for example, academic papers, programming tutorials, or blog posts, as well as content of different formats, including audio and video. While these are not news, their compatibility with the platform reflects the goal of Inflo to be a central place on the web for sharing content on the web which could be seen as having informational value to the community. As such, we will be undergoing continuous R&D efforts to effectively process those types of content for classification, and include the relevant classification labels in the filtering system. In this way, Inflo will eventually evolve into not only a hub for news, but a hub for any kind of web content which could be insightful and useful to the community.

2.2.2 Web application

The details on the frameworks used for development of the platform, along with their respective architectures are outside of the main theme of the thesis, and as such, are provided for reference in Appendix IV.

2.3 A Hybrid Social News Aggregation System (UI/UX Features)

The components in the design of the user-interface have been selected with the goal of creating an intuitive fusion between a news aggregator and social network.

In order to understand the design of the user interface of the Inflo platform, we investigate the defining features of the two types of applications below, along with examples of each.

1. ***News aggregator:*** examples of news aggregators include Flipboard, Google News, Apple News, and Yahoo News. The defining feature of each of these applications is a curated selection of news articles across various sources, displayed under a defined set of discrete categories. Content selection and category allocation is often done by human curators employed by the respective companies, and without affiliation to the publishers of the news.
2. ***Social network:*** examples of social network applications include: Facebook, Twitter, Instagram, and Reddit. The defining feature each of these apps is a feed of "posts" which are entirely user created (although they may contain links to articles or media published elsewhere). Posts are shared into a feed which is visible publicly, or only to a subset of users, according to the dynamics of the social network.

Inflo combines the defining features of the above two types of applications in the following way:

Users share news articles to the platform, but instead of being turned into “posts”, each article becomes a public entity which is not “owned” by any individual user, with a list of users who shared to the network, all listed in one place in the UI. The article then becomes part of the community, and visible to other users publicly. Each article is labeled by the machine learning driven classifier, which automatically puts the article under the relevant category (see Section 2.1.1), allowing the user to browse content in aggregator-like fashion using the Filtering Panel (see Appendix III).

2.3.1 Public Feed

In the first iteration of the Inflo platform, all shares (articles), and commentaries are public, as opposed to the feed only showing posts by a user’s connections, or only those accounts that they follow (this is the case in Facebook, Twitter, and Instagram, but not Reddit). While there is certainly value to be obtained from identifying connections between users, in the case of a hub for news insights, there is no added value from seeing only what one’s “friends” posts. An objective insight is an objective insight no matter which social circles one is a part of. Assuming that users come from different real-world communities, we also expect this design decision to results in a minimal occurrence of the echo chamber effect [63]. However, the downside to this is that there could be a sense of “information overload” or too many articles and commentaries that someone would not know how to most efficiently make use of the platform. As such, we have developed a “filtering system” which allows users to immediately discover the content that is most relevant to them, described below.

2.3.2 Filtering System

The filtering system is an implementation of the automatic classification system described in Section 2.1.1, using the classification results as filtering options, enabling users to hide or display articles according to their labels. The initial iteration of the platform only makes use of category filters. The categories can be toggled in the form of a checklist—more than one can be selected at a time. This offers the flexibility of the platform to be used in traditional aggregator style (only one category at a time), while also offering more personalization by allowing the users to select multiple categories according to their interests.

The filtering system is one of the pivotal features of Inflo as a content discovery tool, being a revision of existing social platforms, which present a single-stream recommendation feed (unfilterable), which has been shown to cause a several unintended negative externalities as described in Section 1.1. It is also a revision of the navigation tools of existing aggregators given that all the content is shared by other users and classified automatically, as opposed to being curated and classified by individuals which is costly, time consuming, and only includes content from a preselected list of sources. This feature has only recently become feasible due to recent advances in deep learning utilized in the ANCCS (Section 2.1.1), making it all the more valuable a component of the Inflo platform.

2.3.3 The Content Block

The hybrid aggregation/social network gives rise to a new UX concept which will serve to revise the status quo of content navigation and user interaction on the web: **the content block**.

A content block (CB) is defined as a preview of a news article within the platform which is not owned by any user (see Figure 3). No matter how many users share the same article, there will only be one CB, which will be an aggregate of all shares of a given article. In other words, all user-stories of a given story are represented by a single element in the UI.

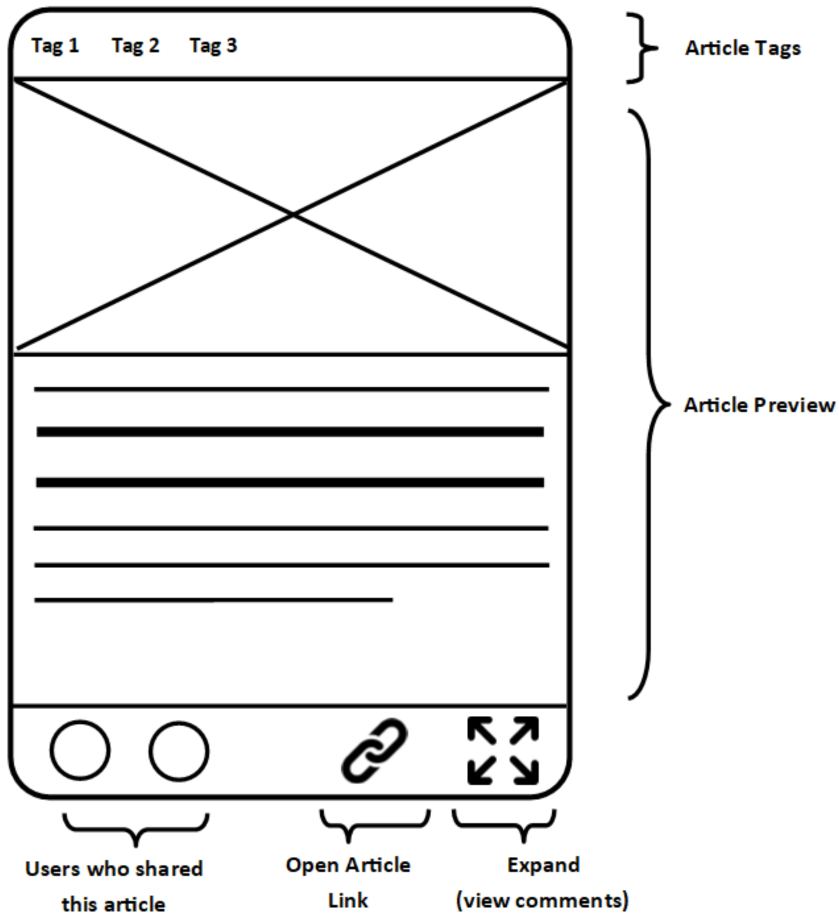


Figure 3: The content block, a new design paradigm for displayed news previews for use in the Inflo platform.

Rather than putting the content in the context of the user who shared it, e.g. as Facebook or Twitter do, we are flipping that paradigm on its head, and putting the shares and user contributed commentary in the context of the shared content (i.e. a content first approach, as opposed to a user first approach), with the goal of facilitating more responsible social media content interaction (i.e. there is no ability to "retweet a news article", which tends to be driven by impulse, leading to a spread of unverified or misleading information [16]).

Then, the comments will be displayed in an expanded version of the CB which opens when a user clicks on the CB (see Figure 4).

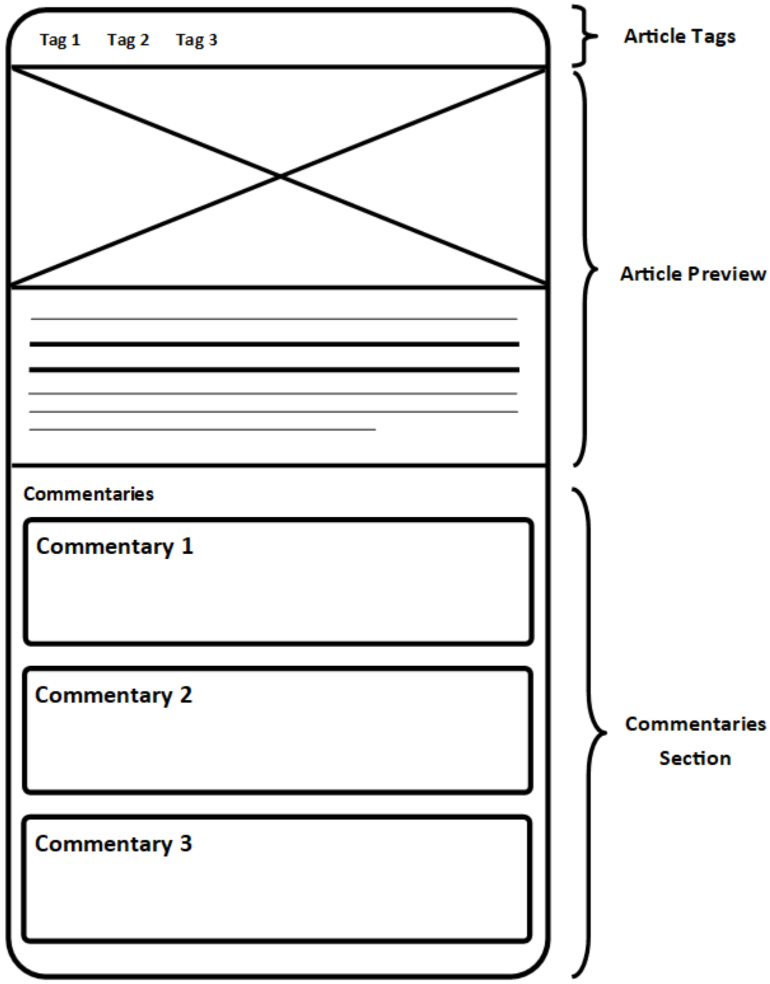


Figure 4: The expanded content block, which shows commentaries contributed from all users in the network who have shared the same article.

The CB design paradigm described above, although seemingly minor, leads to several improvements in the paradigm of online news content sharing:

A. All commentaries associated with a particular piece of content can be found in one place, not scattered around the network. This means that there is no filter bubble at play. Whereas in on Twitter, the same article might be received positively if Tweeted by one person, and negatively by another, we can see the reception of people across the network using the independent CB model.

B. Sharing of content to network will be realizable only when the content has not already been shared (by any users), removing duplication of effort. If a user shares an article to the network, and it has already been shared previously, their commentary contribution will be added to the existing expanded CB, allowing that user to then view what users have previously said about the same article.

C. All users who shared it will be visible on the CB itself. This let's everyone in the network know how many times a given article has been shared within the network in total. That could provide insight into the popularity or an article which could, albeit not necessarily, provide a hint as to the reliability of the article. It could also give more recognition to good quality articles from lesser known sources, seeing as there will be a sorting option in the dashboard to show articles shared most at the top of the feed.
D. Users will see the content in its original form before seeing commentaries (need to click to view commentariess). The dynamics of the CB being user independent means that users will first be exposed to the article as it was original published, before being influenced by other users' opinions. This could avoid the possibility of an echo-chamber effect where a person will have a pre-existing opinion of the topic prior to even reading an article, based on somebody's comment which they agree or disagree with. While the user can choose to read commentaries prior to reading the article, the design of the CB makes it easier to avoid that if the user so chooses.

Overall, the non-user-associated CB is where the merging of social network and news aggregator dynamics plays a pivotal role in eliminating low quality content (including misinformation). By putting content first, the user can get an uninfluenced perspective of the article. And by aggregating commentaries from across the community, it will be easy for users to gain a sense of the public's reception to—as well as informative value of—a given article, across the entire network.

2.3.4 Commentaries and Commentary Types

With Inflo, we have also reconsidered the way users react and interact to/with shared content with a system of commentaries and commentary types.

Commentaries

Firstly, a response to news article is considered a commentary, as opposed to a comment. This is to distinguish them from comments as they commonly exist in other social platforms--as concise, witty, or comical remarks intended to garner as many likes as possible, and which often contribute to polarization [11]. Inflo's commentaries, by contrast, are designed to encourage insights and informational value. The design of the commentary is intended to evoke a feeling of an expert contribution, as opposed to a trivial remark.

Commentaries are not limited by length. This decision is based on the hypothesis that the longer a comment, the more consideration a given user will put into writing it, and the more informational value it will resultingly contain, as opposed to comments which are deliberately short and whose contents are primarily emotion-based (i.e. not making any significant contribution to the community, yet still could garner many likes due to people feeling empathetic to a given comment). There is a body of research which investigates sentiment in social media posts, which could provide some insight into the veracity of the given hypothesis [64]. This is an area for future work which the Inflo team will pursue to validate the stated hypothesis.

Commentary Types

Furthermore, each commentary is designated by the user as a specific type, from a set of predefined option. Type selection serves two purposes: 1) to give others an idea of what a commentary will be about before they read it, and 2) to discretize the forms of context provided therein, and constrain them so as to avoid spamming/trolling--if it doesn't fall into one of the defined types, it most likely does not provide the type of contribution that is valued within the Inflo platform. A user then might consider reconsidering what kind of commentary to write, taking into account what would be valuable to the greater online community.

The available commentary types are based on those which will enhance the user's experience (sense of satisfaction in using the platform, stemming from intuitiveness, or designing features to correspond which the user's mental model of how it should work) [65], as well as based on which types of insights are most valuable in the context of our intended monetization model (see Section 3.3). The commentary type options are as listed in the following table (Table 1), along with a short explanation of a what given commentary should contain for each one.

Table 1: Commentary Type Options and Their Descriptions

Comment type	**Description**
Discussion	This type entails any general discussion or reaction to the content within the news article; not relating to the quality thereof (which should be contained under the "fact-check" or "bias-check" types below).
Highlight	A standout excerpt copied and pasted from the article. The user can then go on to provide a reaction to given excerpt.
Opinion	This corresponds to a personal reaction one would have to content of the article, such as a tragic event, natural disaster, or political controversy. This is different taking issue with the article's standpoint, in which case the "bias-check" type should be used.
Question	This type allows the user to ask for clarification on some part of the article; for example, some background information that may have been missed. It could also be used for rhetorical questions relating to the content of the article.
Bias Check	This type allows the users to present an alternative viewpoint on the given object of reporting, coming from perhaps an alternative political or cultural standpoint. It also provides the opportunity to pint of a flaw in an argument expressed in a given argument; e.g. a logical fallacy such as a "straw man" or "red herring" which are often employed in political opinion pieces. [66]
Context	This type allows the user to present additional background information relating to the subject of the article, such as statistics or prior developments relating to the same story (such information could be considered as a response to the questions posed using the "Question" type.

Fact Check	This type is where users can point out possible inaccuracies in reporting, for example, misstated statistics, misattributed quotations; or otherwise fabricated information. Users will be encouraged to post a citation to a fact checking website or a link with the “correct” information.
Related	This type allows the user to post a link to an article which is related in theme or topic to the posted article.

If we consider that each commentary type could also be used to address previous commentaries, as opposed to the article itself, there are plenty of further use-cases in which they could be employed. For example, “related” could also be used to elaborate on other commentaries with the “bias-check” type, to either confirm or refute them; bias check could be used to comment on another user’s “opinion” with evidence to back up their standpoint; “context” could be used to back up or refute an element of a previous “discussion” commentary, and so on.

Future iterations of the platform will include a feature for direct reactions to previously posted commentaries, whereas in the current iteration, users will be encouraged to do so by addressing the relevant username of the user whose commentary they are responding to.

2.3.5 Share Panel

The share panel is the interface through which the user will share content from around the web into the Inflo platform. The share panel is a defining feature of the platform, and as such, careful consideration has been put into all aspects of its design and execution, all with the goal of facilitating a flow of articles and insights in the most seamless way possible.

While the current share panel takes the form of a UI element within the platform itself, future iterations of it will take the form of a browser extension.

Such extension will bring about a major improvement in usability by allowing users to directly contribute to the body of content on the platform without having to navigate to the platform separately (e.g. sharing and commentaries can be done on the page of a news article rather than on the platform itself). See Figure 5 for details.

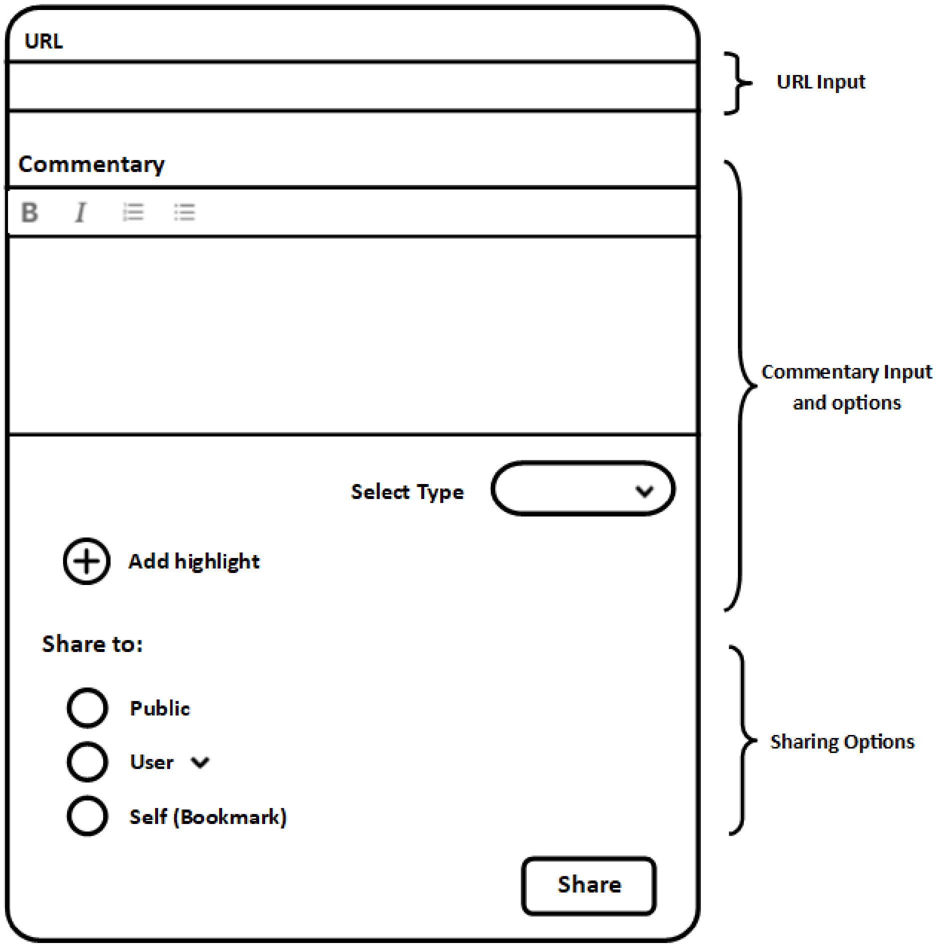


Figure 5: A wireframe depiction of the share panel, which is the portal with which users can share content from the web into the Inflo platform for users to view and interact with.

The browser extension would automatically input the URL from the user's currently opened tab/webpage, so that they would not have to copy and paste it into the sharing panel within the platform itself. This will greatly decrease the friction involved in sharing content and commentaries into the platform.

Additionally, future iterations of the share-panel browser extension will support the following features:

- pre-share article preview, and image selection (if there are multiple images on the webpage, the user can toggle to which one should be used in the preview)
- commentary text formatting

- direct article highlighting (rather than copying and pasting text into the commentary input area), done via highlighting the text and showing a highlight button next to the highlighted text.
- bookmarking (saving articles to one's personal account for later reading),
- and one-to-one sharing (sending an article to an individual user, as opposed to sharing it with the entire network).

With a user-friendly, intuitive share-panel, Inflo intends to become a central hub for content aggregation across the web, both in a social as well as a personal context.

2.3.6 Additional Features and Design Decisions

As with any application, one must continuously update and tweak its features according to 1) user feedback and 2) product-market fit. While this means that the future evolution of the app is unknown in a specific sense, we do plan to add certain UX features which have not yet been implemented due to time or funding constraints, although the individual features therein could take a more or less different form according to the aforementioned two factors.

See Appendix V for the complete list and timeline of expected rollout for new UX features over the next three years.

2.4 Twitter Integration - Solution to the Chicken and Egg Problem

New social platforms, unless they bring something incredibly new and fun to the table, are often neglected due to the network-effect, i.e. the fact that if all a user's friends are on a certain platform, they will stay there, and have little incentive to migrate to a new platform. In the case of Inflo, if there is no content being shared, users will find it uninteresting and quickly churn (leave) the platform. If there are no users, there will be nobody to post content, and so the cycle continues. This is what's known as the chicken or egg problem of consumer facing platforms [67]. In order to circumvent this, Inflo will leverage Twitter to populate the platform with content, which is easily done through Twitter's API. As such, the fact that there are few users posting to the platform initially does not mean that there will be no content on the platform, leading to an effective nullification of the chicken-egg problem.

The Twitter Integration will allow users to share content from Inflo into Twitter, and vice versa, essentially creating a two-way street between Inflo and Twitter. The workflow for this feature is described below, and visualized in Figure 6.

When a user signs in with their Twitter account, they give us permission to view their list of friends (i.e. accounts they follow). The Twitter API, accessed via Python's Tweepy package [68], is then used to go through the list of followers and extract and tweets containing supported content (i.e. news articles), and then display them in a separate feed. The Twitter feed is similar in appearance to the public feed (dashboard), except that it is private only to the given signed in user. CBs then, are the same as they would be if the news articles were shared directly to Inflo, except that in the "shared by" Section, the user icon will link to a Twitter profile as opposed to an Inflo profile. Any text that was shared along with the article in the tweet will also be adapted into the form of a comment, and appear in the expanded CB. A user can then choose to add a new commentary to the given article, upon which the article and the new commentary will be automatically visible in Inflo's public feed.

Then Inflo must motivate users not only to browse content within the platform, but also to interact with it, namely, contribute their opinions, input, and insights. The way we induce such motivation is threefold.

1. Users will be rewarded with recognition within the community, in the form of expertise points. Several platforms make use of a reputation system to incentivize user interaction in an online community, to powerful effect: Stackoverflow (reputation points), Reddit (Karma), Quora (upvotes, content views).
2. There is no current social media platform which is publication independent, and which aggregates all comments to a given article together in one place. In social media like Facebook, comments are grouped together under an individual user's post; in aggregators like Google News, one is led to the publication's website where the comments are all written by frequent visitors to that site (usually reflecting a one-sided group mentality).
3. In contrast to Twitter, which could be considered a major competitor, Inflo will not have a character limit with regard to user's responses, whereas a Tweet can only contain a maximum of 280 characters. This feature would allow for more in-depth discussion, and

decrease polarization and personal attacks, which are often provoked by the limited tweet structure.

Additionally, each commentary posted to an article shared within Inflo's network will have a share button, which, when clicked, will allow the user to share the given commentary to Twitter. The resulting tweet will be a URL leading back to the Inflo expanded CB with the relevant commentary highlighted, along with a preview of the given article.

In addition to allowing Twitter users to bypass the character limit imposed by Twitter itself, this features will also create a stream of organic marketing: users who have not heard of Inflo will be able to view a commentary via the shared Tweet, and if they feel inclined to respond, they can choose to create an account on Inflo itself and do so, bringing more active users to the platform.

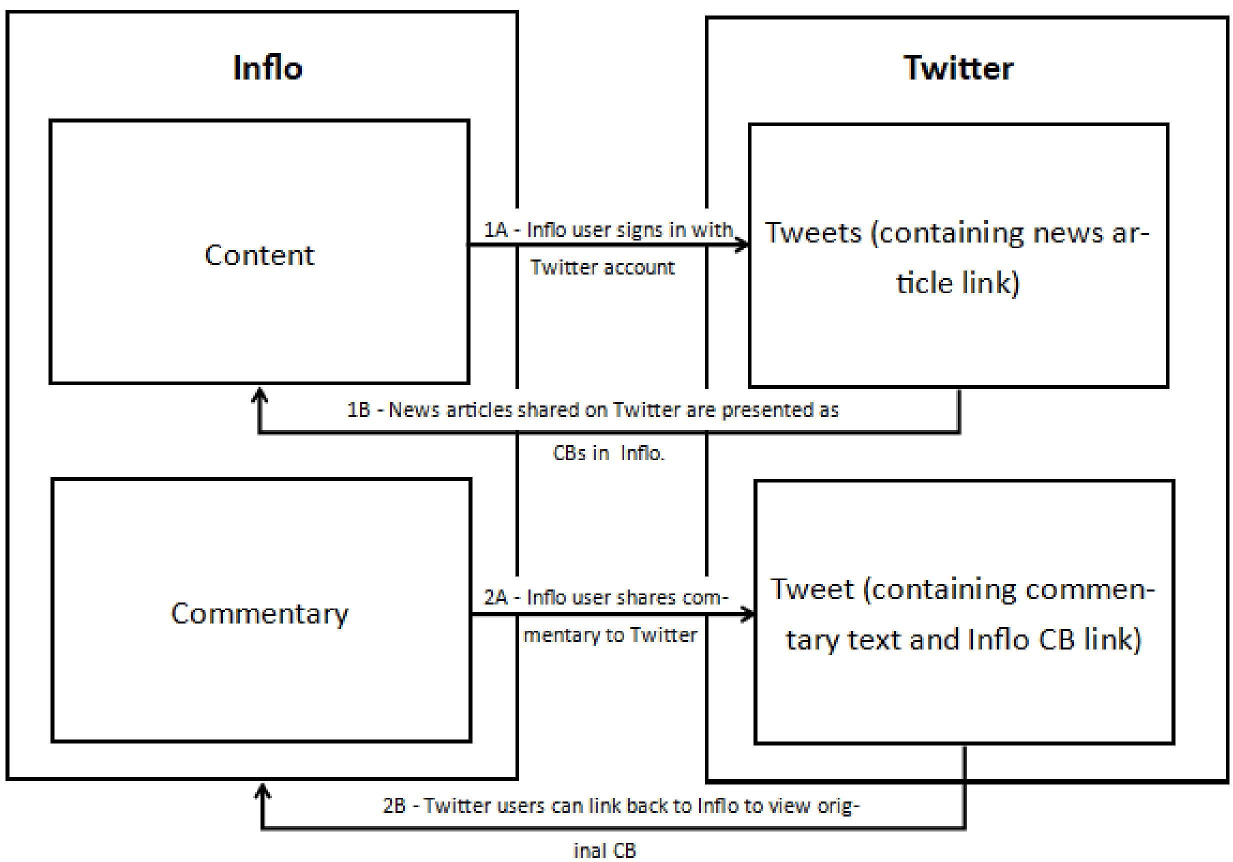


Figure 6: A schematic showing the planned interdynamics between Inflo and Twitter. Step 1 is the solution to the chicken-egg problem, bringing in content, while step 2 is a growth-hack mechanism, bringing in users.

Ultimately, the Twitter integration will be a pivotal feature of the platforms as it will serve as both a usability boost, and a growth-hack, as well as a means of avoiding the plaguing chicken-egg problem. The Twitter Integration is anticipated to be deployed in the Beta version of the platform (see Appendix V for full timeline).

It is important to note that there is no copyright conflict in using Tweet content in Inflo, because news articles shared to Twitter is also not owned by Twitter. Twitter's API allows anybody to develop on top of their network, and utilize the data generated by users [40]. The API is free for a certain number of Tweets, afterwards, one can pay for a broader range of access to Twitter data, which makes it a win-win for both Twitter and Inflo in the case we decide to pursue that route.

Chapter 3: Business Plan & Monetization

3.1 News Industry Landscape

With web 2.0 and ubiquitous internet, almost all news publishers now publish primarily, if not exclusively, online. In turn, the new dynamics of online publishing have caused the news industry to undergo two drastic disruptions: 1) a reduction in revenue due to lower digital ad prices (as opposed to print ads) [69], and 2) a competition for engagement which has led to a drop in quality of digital news content [70].

As a result of #1, some publishers have cut their budgets or closed down [71], while others have been sustained by VC funding to varying degrees of success [72]. For those that have taken a more adaptive approach, we have seen them implement a subscription model. This entails implementing a paywall, which can take different forms, but the most common one being a metered paywall, which blocks users from accessing full articles after they've read a certain number per month. The paywall can then be removed by with the purchase of a subscription. However, this model has worked for only a few of the most preeminent news publishers [73]. The subscription model, despite being seen as the path forward for a more sustainable digital news industry [74], has yet to find an effective manifestation. An effective subscription model would solve both the financial and trust problem that the digital news industry is now experiencing, due to the removal of the reliance on ads for either revenue or for driving engagement. However, the current problem with subscription models are that most people do not want to subscribe to individual news sources, given that they typically consume news from more than a single source per day [75].

Meanwhile, news promoted on social media tends to drive quality down, due to the fact that the business model of social media platforms are at odds with the goal of reporting accurate, informative information [76]. It is, therefore, no surprise that trust the news has decreased in recent years [9].

From the above it is clear that existing social media platforms have had a detrimental effect on both the quality of published content as well as the financial sustainability of news publication, generating an antagonistic relationship between the two. This in combination with the fact that a truly sustainable subscription model has yet to be achieved in scalable and replicable form are all indicators of a need for a new platform step in the arena to disrupt the disruption, and make publishing quality news more sustainable. The business model of Inflo has been designed to do exactly that: reconcile the financial incentives of publishing accurate information with those of maintaining a healthy online social community. Therefore, the antagonization will be replaced with a symbiotic relationship between the platform and the publication.

3.1.1 Revenue Models of News Aggregators

Ad-based Aggregators:

Despite the problem of quality within the news industry being in large part driven by platforms running ads, several news aggregators and social media apps have come about in recent years continuing to do just that [71]. They insist on leveraging news content as attention grabbers in order to gain user data and sell targeted ads. This is a one-sided relationship which benefits the aggregator/social platform much more than it does the publishers themselves. It enables users to access content for free, and only benefits publishers by bringing them traffic, instead of revenue [77].

Subscriptions based aggregators, on the other hand, provide an easy way to get across paywalls on articles from multiple publishers at one time.

One such example is INKL [78], a news aggregator and curation website/application which has a subscription offering which combines premium content across multiple publications into a single monthly package, or sells access to them on a per-article basis. This provides the benefit of cost efficiency and convenience, so that a user does not have to subscribe to multiple publications separately.

Another example is Blendle [79], a Dutch-language news aggregator which charges on a per article basis, has achieved a 20% paid user conversion rate, although it is still not profitable [80].

However, due to the fact that the above aggregators host the content on its own platform, rather than linking to the publisher, users can still find ways to access the content elsewhere for free. This has led to some dissatisfaction among users.

The above examples certainly show the promise of the subscription-based aggregator model, by bringing more audience revenue to publishers while creating a sustainable business in their own right.

Inflo takes this idea and expands upon it by adding the social network and interaction features, instead of purely aggregation as seen in INKL and Blendle, and minus the curation (which is effectively replaced by users sharing articles themselves).

3.2 The Inflo Model – Netflix for News

The Inflo monetization model is designed to provide value not to advertisers but those in the media industry—journalists and publishers—themselves. The goal of such a model is to both increase the quality of reporting through consumer intelligence, and to bring more audience revenue to publishers, thus representing a new sustainable model for the digital news industry.

3.2.1 Revenue Model

Inflo's revenue model is twofold. Firstly, Inflo will offer users a freemium subscription option to access locked (paywalled) content from multiple sources for a single monthly payment. Inflo will effectively purchase subscriptions in bulk, and serve them to as a middleman. Ideally, this would take advantage of economies of scale by purchasing them at lower cost than they would be if purchased individually, thereby creating a profit margin when they are resold. However, because multiple subscriptions will need to be purchased per user, the middleman/bulk analogy does not apply. To go around this obstacle, Inflo will purchase paywalled content for each user on a per article basis, as opposed to an unlimited content per time (e.g. month) basis. What this means is that publishers will receive less monthly revenue per subscribed Inflo user (as compared to a user who subscribed via their own subscription plan), but a greater quantity of paying users in total, effectively increasing conversion rate, and in turn, a greater share of revenue from subscriptions. The revenue model is presented as a flowchart in Figure 7 below.

The feasibility of this model takes into account the basic assumption that news consumers will be more willing to pay for a subscription which contains multiple sources than a single one supposing that they both cost approximately the same per month.

Because Inflo does not host original content, but only links to articles published elsewhere, the subscription will simply register Inflo users with each site individually when they purchase the Inflo subscription, so when they click on paywalled content within Inflo, they will not be blocked when they reach the publication's website. Technical considerations for such implementation are still in the pre-development stage.

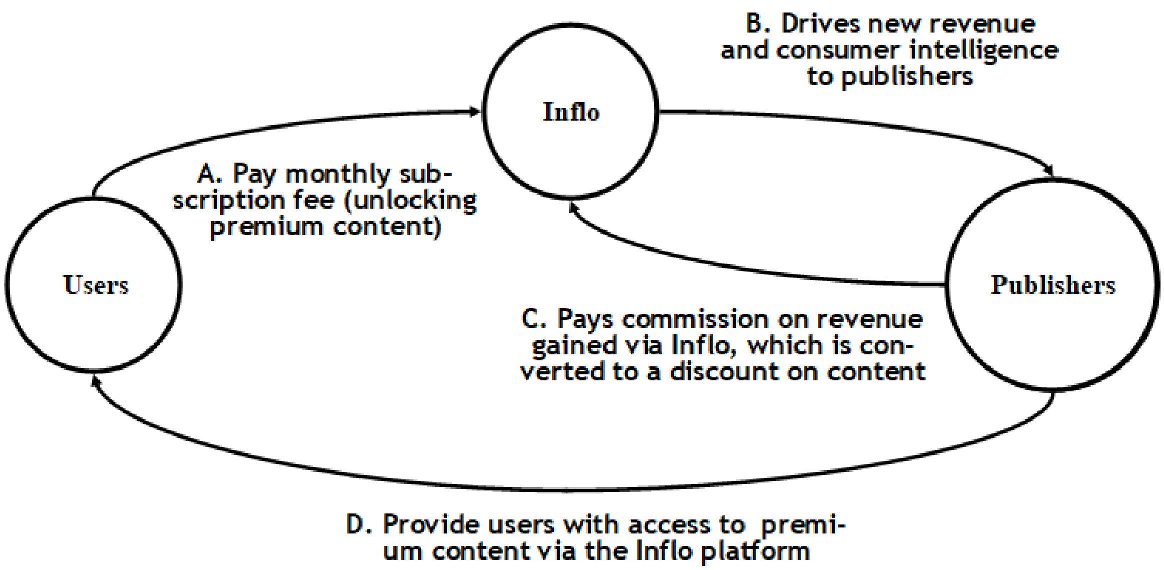


Figure 7: The Inflo revenue model, depicting the symbiotic relationship between users, publishers and the Inflo platform itself, allowing users access more diversified content for the same price as a single subscription to an individual publication. This is done by taking a commission rate and bringing publishers a new revenue stream, all resting on the principle of economies of scale.

Inflo's model differentiates itself from other startups which have attempted a similar "Netflix for news" model in one simple yet pivotal way. That is, the fact that Inflo's subscription unlocks paywalls on the publisher's site itself, as opposed to hosting the content on-platform. This results in a seamless experience from platform to publisher, wherein both parties are benefited: users benefit from a content consumption experience unimpeded by paywalls, while publishers gain more audience revenue and greater traffic overall. It also prevents users from finding the paywalled article for free elsewhere (because only paywalled articles on the original publication's website will need to be unlocked, rather than trying to monetize articles on-platform, which are free in their original form).

The second revenue stream takes the form of a consumer intelligence offering to partner publishers, which will be compiled from analysis of data[1] from different types of commentary the platform, and presented in terms of feedback on a selection of individual articles which have been shared. The forms of feedback will correspond to commentary types, and will be analyzed through a combination of machine learning and human input. The format and explanation of an intelligence analysis for a given article can be found in Appendix VI.

Each intelligence report will also include the following social data derived from community activity:

1. Popularity: how many times this article was shared within our platform versus others. Social data from other sites can be found via APIs (e.g. Twitter API, News API).
2. Engagement rate: percentage and number of users who viewed, read, commented on articles from a given source/topic.

The intelligence reports are designed to provide insights for media publishers, which, in turn, would allow them to improve the quality of their published content, and thus increase trust in their "name", so to speak, which will ultimately lead to higher subscription rates. This is an example of Inflo leveraging the value of *insights*—as opposed to engagement—driven by online news content, a distinction which is at the crux of the online misinformation/clickbait/fake news problem. The intelligence reports will be provided in return for A) a discount on premium articles from those publications who have a paywall, or B) a monthly subscription, as a B2B offering, for those publications who do not have a paywall. The revenue projections described in Section 3.3 take into account option A only.

3.2.2 Market sizing

[1] Data will be anonymized and thus user privacy will always be guaranteed when analyzing and presenting data to third-parties.

3.725 billion people actively use social media. That's 48% of the world's population [1]. Of that 48%, up to 36% see social media as an important way for them to view and access news content [2]. That's 1.3 billion people getting news from platforms like Facebook and Twitter. However, those platforms have been recently under fierce criticism for their poor handling of mis- and disinformation, leading many users to lose trust, if not leave the platforms completely. This leaves the market wide open for a new player to take on the role of a social media platform for news, leaving those 36% of social media users, or 1.3 billion people, as the total available market for our solution.

Of course, these people are spread across different regions in the world and speak different languages. Inflo will first target the English-speaking market, due to the technological constraints (see Section 2.1.1), and scale up to other markets as we deploy more language implementations). If we use those same statistics to estimate the number of total English speakers getting news content from social media, the total addressable market size is 65 million people.

Further, we will target as early adopters those who are most impacted by the poor handling of news on conventional social platforms: namely, journalists and media professionals. They have the largest stake in improving the media ecosystem, because it is their work which can be either exploited, or undermined by the spread of misinformation. Thus, they have the most incentive to contribute to a community in which quality and insights are rewarded. In the United States alone, there are 37,900 people working within the journalism industry as reporters, editors, photographers, or video editors, according to the Bureau of Labor Statistics [81].

3.3 Projections

In this section we lay out a 4-year (48 month) financial projection and the bases on which it is devised, in order to prove that the monetization model proposed above can be feasibly translated into a sustainable business. The projection includes one year for pre-monetization, which will be used for marketing to early adopters as well as product iteration, followed by a three-year monetization scenario which results in profitability at month 44 or the projection (or 32 post-profitability). The projections also include a corporate budgeting (inc. manpower) section, which will provide a more realistic figure for the amount of venture capital needed to be raised. Finally, based on the monetization scenario provided, we offer an estimation of how much a news publication can expect to earn in revenue by joining Inflo's premium content package. This final section does not factor into the company's profitability figures or timeline, but is nevertheless a critical aspect of the business model, as it could prove the benefit of Inflo as a partner as well as a pivotal factor in the digital news ecosystem as a whole.

3.3.1 Pre-monetization

Firstly, we expect to begin monetization one year after venture capital is raised, which in turn is expected to be six months after the platform is released. During that first year, we will focus our efforts on marketing the platform the platform to early adopters. To do so, we have designed a marketing strategy that appeals to our specific target groups, e.g. people in or related to the media industry, as well as those who are disenfranchised with the current slumped state of conventional social media.

As such we have chosen to take an inbound-first marketing approach by promoting the Inflo on other platforms where these people can be reached, through original content. Specifically, we aim to reach industry professionals via Quora and Medium, both of which support more than usual productive conversion as opposed to Facebook and Twitter. For those two platforms, content can be posted for free, and we expect that by posting frequently, with content related to improving the social media ecosystem, we can easily obtain several thousand users in the time period. Next, we will aim at the disenfranchised social media users via Twitter and DuckDuckGo. Twitter is often seen as a place of refuge for once-Facebook users, but is still in several ways insufficient in remediating what makes Facebook unsatisfactory. DuckDuckGo is a search engine which boasts privacy of personal data, which is exactly counter to Facebook's prime feature. Therefore, we feel that users on these two platforms will find Inflo appealing as a new, more responsible alternative to existing social media solutions. Finally, we include press and SEO as the final two marketing streams, which are convention for new startups, and from which we will be able to reach users who might already interested in such a product but didn't previously know it existed.

Below we present the timeline of user growth using the above-mentioned methods to achieve 60,000 active users by the end of the first year.

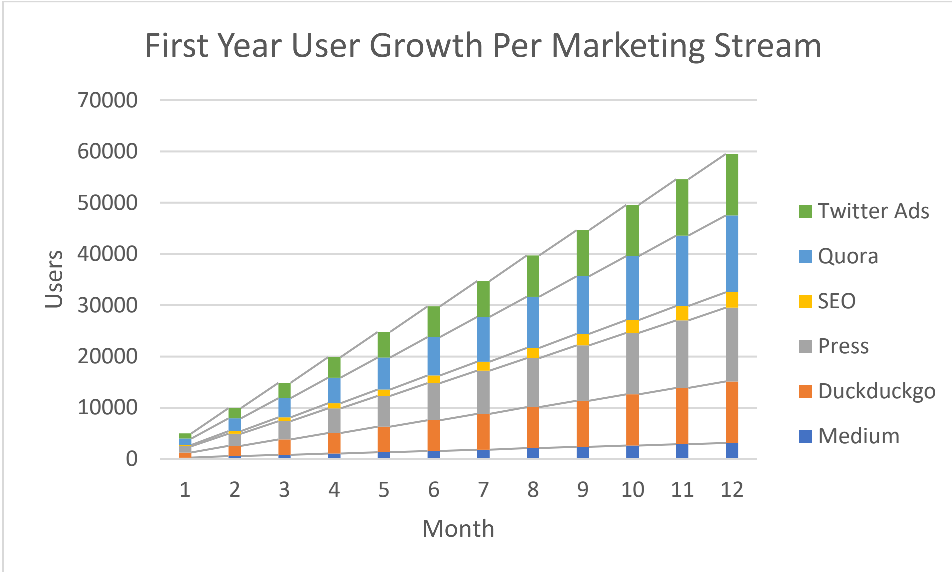


Figure 8: This chart depicts the marketing strategy used to gain the first 60,000 over the first year of release, pre-monetization. All financial projections are based on this number of users as a starting point.

We estimate that the above marketing strategy will cost a total of USD 9000 per month over the first year, after which we will increase the marketing budget by 10% per month, resulting in a total marketing expenditure of USD 3M over the 4 year period, and of USD 2.1M during the period prior to profitability (the latter number being useful for fundraising purposes). A detailed outline of the marketing strategy is provided in Appendix VII-A.

3.3.2 Monetization Scenario

Using appropriate assumptions and defined values, we have devised a monetization scenario which will result in achieving profitability before the end of three years post-monetization.

By the time monetization begins, we will assume that one percent of the total active users will start to pay for the monthly subscriptions, and that the number of paying users will increase by 5% month-over-month for three years. We also assume a rate of 20% month-over-month user adoption starting from the point of monetization, which results in a total of 35 million users by the end of the three year period (see Figure 9 below).

We intend to implement a comprehensive marketing strategy which will lead to the ramp up in user adoption (and in turn, revenue), according to the stated rate. The strategy will consist of:

1. Partnering with individual news outlets to allow users to share articles directly from their sites onto inflo. E.g. The Inflo icon will be available as a "sharing option" on the websites or apps of news publications, which will create a seamless transition from reading the article to sharing it to Inflo. We anticipate that user adoption will increase greatly will such an addition, and ramp up each time we collaborate with another publisher who agrees to implement the feature.
2. The ability to share Inflo commentaries to Twitter will bring in users to the platform.
3. A continuation of the marketing strategy outlined in section 3.3.1, during the phase of monetization will also contribute to increased user adoption by spreading the word of the platform beyond those who would be in the directly circles of the early adopters.

First, we present the user growth projection below, based on the aforementioned assumptions. The raw data can be found in appendix VII-B.

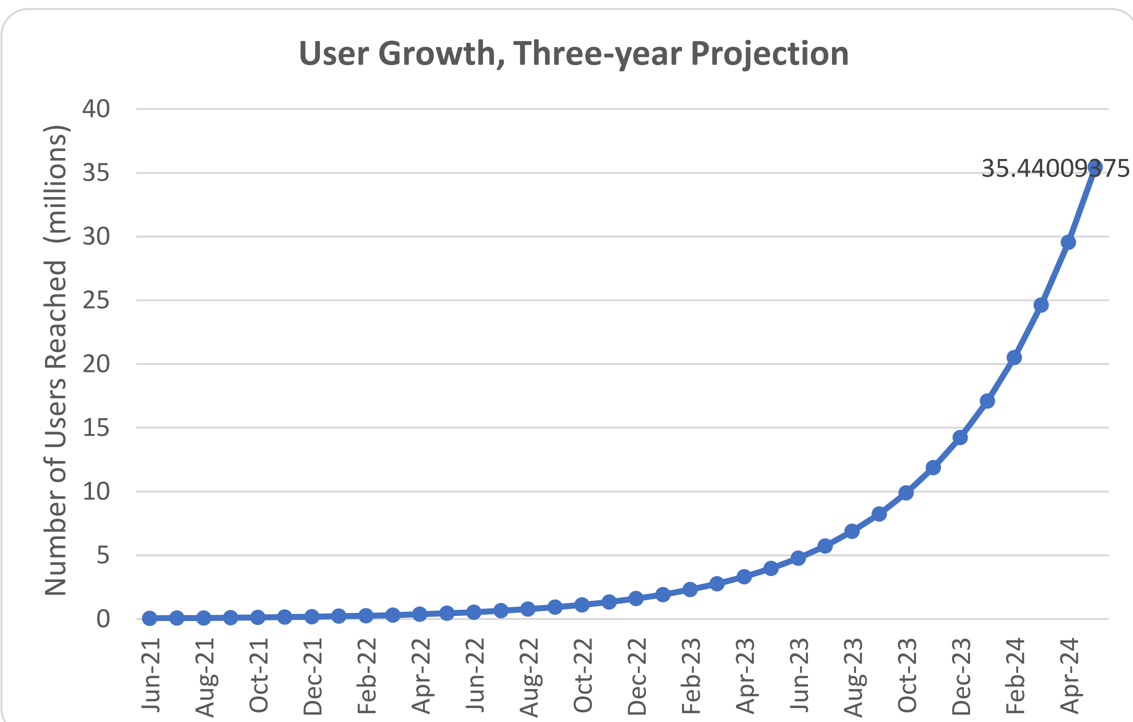


Figure 9: Three-year user growth (in millions) projection based on 20% month-over-month adoption, resulting in a total of 35 million users by the end of the third year.

We will set our subscription rate at USD 12 per month per user, and we will assume that, for the first three years, Inflo's subscription package will consist of ten publications.

According to the figures specified above, if we have 10 publications included in the package, and each user is paying a USD 12 subscription rate to Inflo, each publisher should only get USD 1.2. But instead, Inflo will have an initial expense rate of USD 30 per subscribed user per month, meaning that each publication will get USD 3 per subscribed Inflo user.

Inflo earns revenue by taking a commission from the amount of revenue publishers earn through our paying users. The commission rate used in the current projections is set at 10% of the revenue earned per month publication. The commission rate, although does not change in percentage, results in a month-over month increase in revenue to Inflo because is it based on the subscriptions of users, which is increasing by 5% each month. In other words, the commission rate could also be thought of as a discount on the content which we are purchasing from the publisher. The discount is based on the amount of revenue which a given publisher is earning via our subscribers. E.g. if a publisher earns $10,000 one month via Inflo, the publisher will give us a $800, which will be used to purchased the content for the next month (effectively, an $800 discount). Therefore, the commission rate does not need to change in order to the revenue for the platform to increase.

With the increase in revenue from commission as user adoption increases, the number of paying Inflo users will eventually be enough such that the discount on content will bring the expenditure per user lower than the rate at which users are paying *us*, resulting in a gross positive monthly profit margin. In other words, if the discount we get on premium content from each publisher (based on the commission off of the revenue they earn though our paying users) allows us to decrease our expenditure per user across all publications to USD 10, then if each user pays USD 12/month for a subscription, that's USD 2 of profit per paying user per month.

Below we present the increase in percentage of paying users out of all users, and the effective cost per user after the commission/discount is factored in. The raw data for these figures can be found in Appendix VII-B.

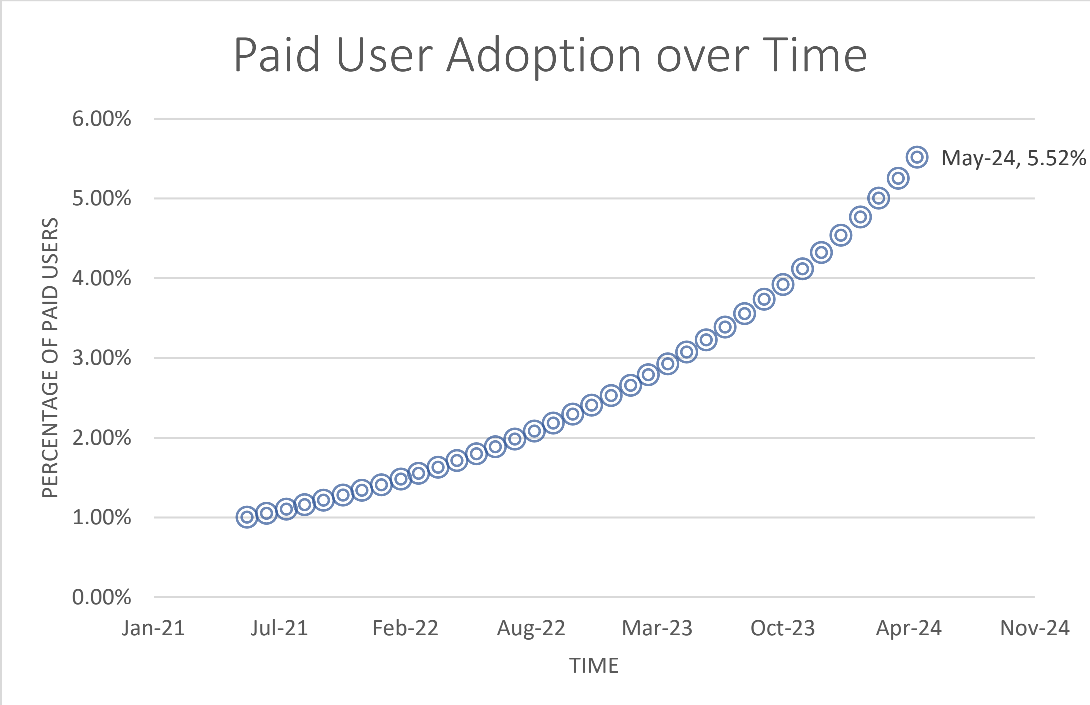


Figure 10: This figure shows the increase in paid user adoption over time, which results in a total of 5.5% of users paying by the end of the three-year monetization period.

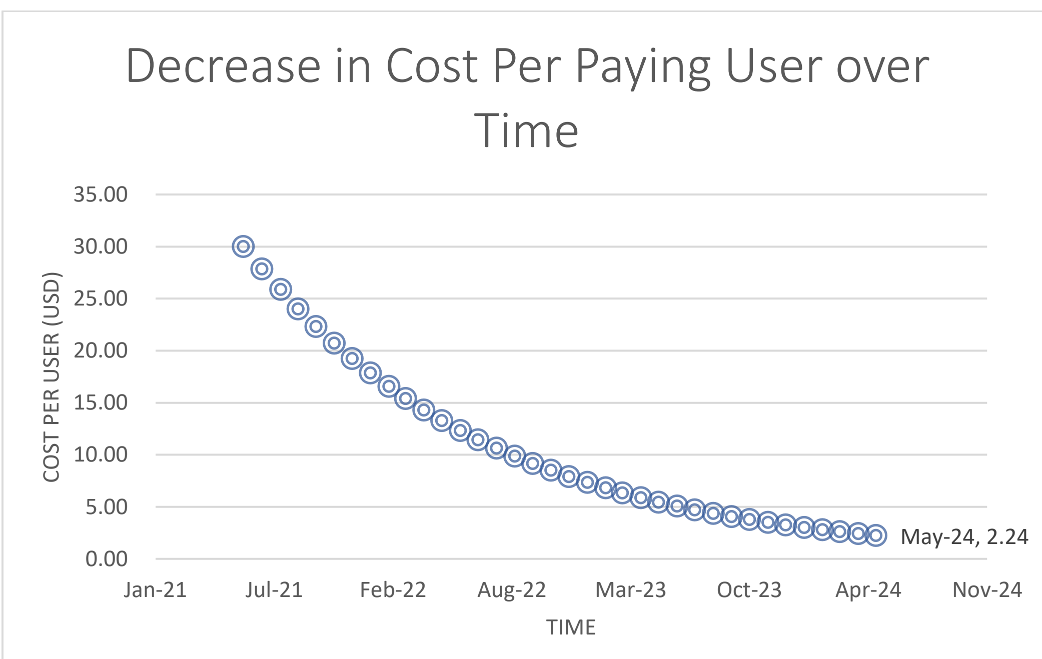


Figure 11: This figure shows the decrease in cost per user over time, which goes below $12 in July 2022, or the 14th month after monetization.

According to the calculations based on these figures, the expense per paid Inflo user will drop below USD 12 on month 14 after monetization, at USD 11.45 per subscribed user (a profit of USD 0.55 per paying user that month). Then, on month 19 after monetization, Inflo will achieve its first profitable month, with USD 4.10 in profit per paying user, and a total of USD 64K in profit. Cumulative profitability will be achieved six months later, on month 25 after monetization with a total of USD 787K in profit for that month. The full calculations for this scenario can be found in Appendix VII-B.

The time offset in profit from content and total monthly profitability is due to the inclusion of corporate expenses in the overall projection, which are outlined in the following section.

3.3.3 Corporate Budget

Inflo intends to adhere to a lean startup philosophy meaning that we intend to seek the minimum amount of venture capital needed to reach profitability in three years. Upon reaching profitability, Inflo as a corporation will decide whether to pursue further funding rounds of fund the company's operations entirely through earned revenue.

Keeping with this principle, we have devised a lean corporate budget to determine how much venture capital will be needed to cover corporate costs over the same period described above.

The manpower component of the corporate budget includes developers, data scientists, marketers, and interns. Over time, we will slowly hire more manpower for each of the given positions, resulting in a total manpower budget of USD 3.6M over the 4-year period (one-year pre-monetization, and USD 2.8M during the period prior to monetization.

The yearly salaries for the allocated positions are as follows: frontend developer, 65k; backend developer, 65k; data scientist, 80k; digital marketer, 50k; intern, 24k; CEO, 65k. Positions will be filled according to the following timeline.

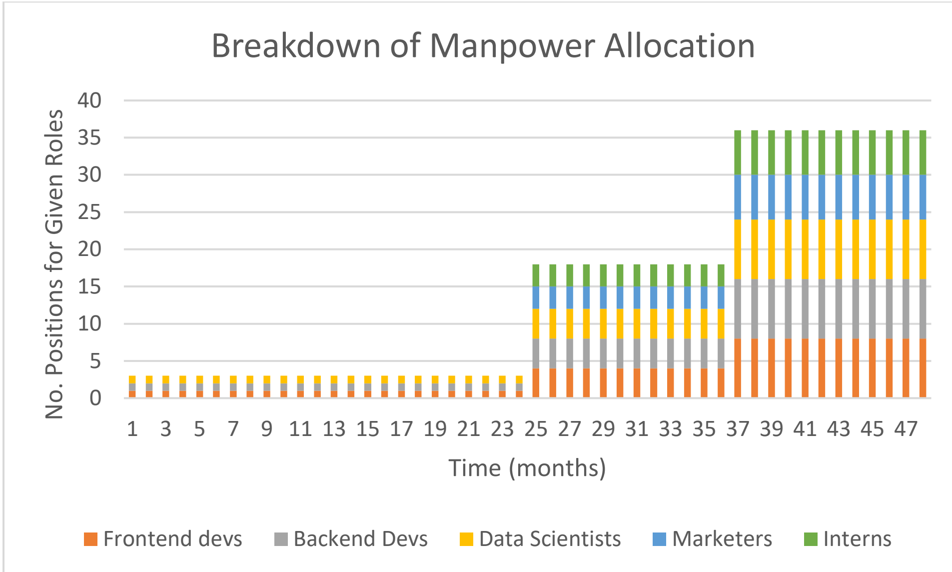


Figure 12: This figure shows the breakdown of manpower over time.

In addition to manpower, we include technical costs (cloud hosting), property costs (office space rental), corporate tax, and one-time fees including business registration costs and patents.

Rent is set at $3000 per month, and corporate tax is 21%. Cloud hosting is approximately $300 per month for an Amazon EC2 instance with the minimum storage requirements. The breakdown of corporate expenses over time is shown in the following figure. Rent and technical costs are included in "additional expenses".

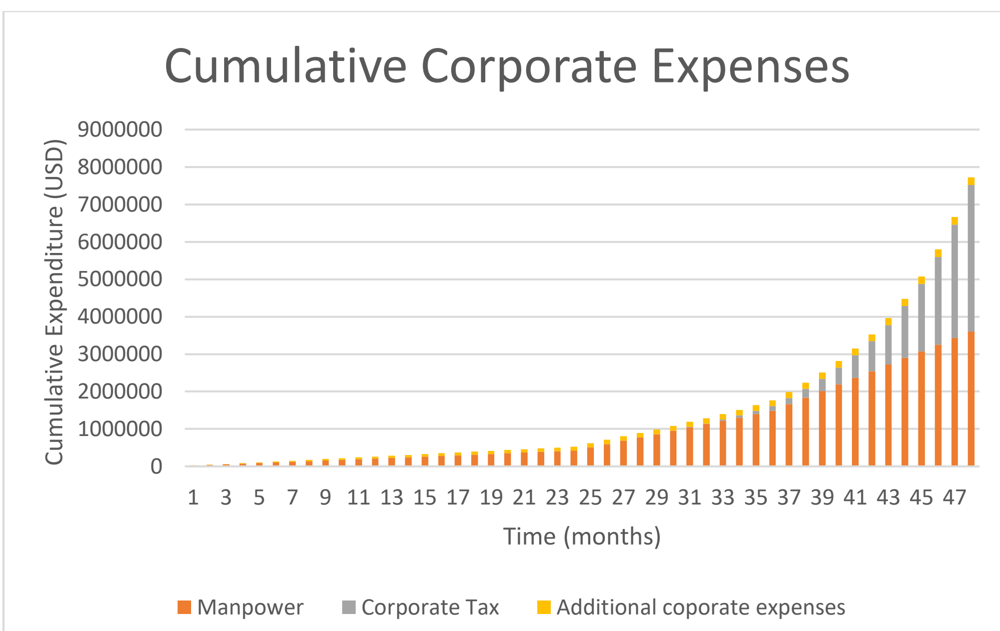


Figure 13: This figure shows the increase corporate expenses over time, broken down by manpower, corporate tax, and additional expenses.

The additional technical costs amount to a total of USD 90,000 for the period prior to monetization, and USD 100,000 over the full four years. The raw data for corporate expenses can be found in Appendix VII-C

3.3.4 Publisher Revenue Estimation

Although the rate at which Inflo pays a given publication per paying user is initially lower than the rate of a subscription a user would pay to a publisher directly, we assume that a majority of paying Inflo users are paying for a news subscription for the first time; i.e. they were not already subscribers of any individual publication. As such, a majority of the users paying via Inflo to each publisher is in addition to the revenue that they would already be making. In other words, if a publisher's individual subscription is USD 12, and Inflo is paying USD 5 per paying user to each publication, that does not mean we have to bring in more than twice the number of users the publication original had in order to break even. Depending on the proportion of actual users who switch from an individual publication to Inflo's subscription, the multiple can be much lower and still earn the publishers a net profit.

Accordingly, instead of using the gross 95% content revenue to estimate publisher gains, we have devised a scenario to calculate a more realistic estimation. To do so, we have assumed that 15% Inflo's paid users are those who have churned from an existing publisher's subscription. To calculate the amount of lost revenue per publication, we assume that each publication has 250K paying users. We then subtract the amount of revenue a publication would lose according to the number of churned users who migrated to Inflo as a percentage of the publication's paid user base, and recalculated the revenue earned by a publisher via Inflo per month.

According to this estimation, we were able to validate that the paid user adoption of Inflo in the above scenario will not occur so drastically as to result in a net loss being incurred by any publisher. On the contrary, a publisher who opts to be included as part of Inflo's premium content package will ultimately earn a net gain of USD 1M in additional revenue over the three-year monetization period via Inflo after losses from churned users, which is shown is the following figure.

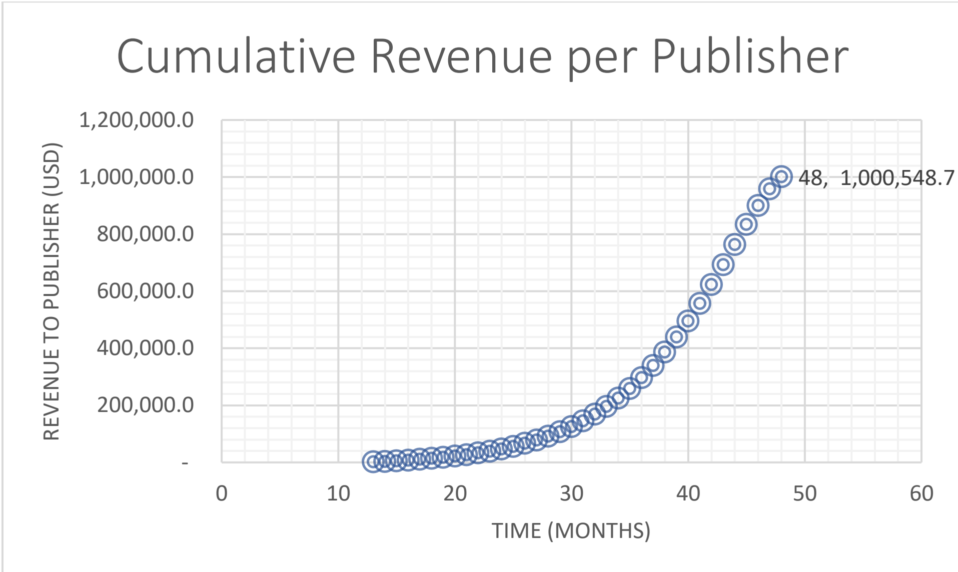


Figure 14: This figure shows the increase in paid user adoption over time, which results in a total of 5.5% of users paying by the end revenue earned by each publisher over the three-year monetization period, ending at a cumulative amount of USD 1 million.

The raw data for the publisher revenue scenario can be found in Appendix VII-D.

3.3.5 Projections Summary

Inflo's revenue model leverages economies of scale. By bringing in an increase in paying users, and resultingly, revenue, to individual news publishers, Inflo can pay less per user for the content from each publication and ultimately earn a profit.

According to the projections in the monetization scenario outlined above, Inflo will earn a positive monthly profit margin 27 months after monetization starts (see Figure 10), and break even in cumulative profitability 5 months later, making the business profitable before the end of 3 years post-monetization. At the end of the three years, the scenario projects a monthly profit of USD 11M and a cumulative profit of USD 31M (see Figure 11).

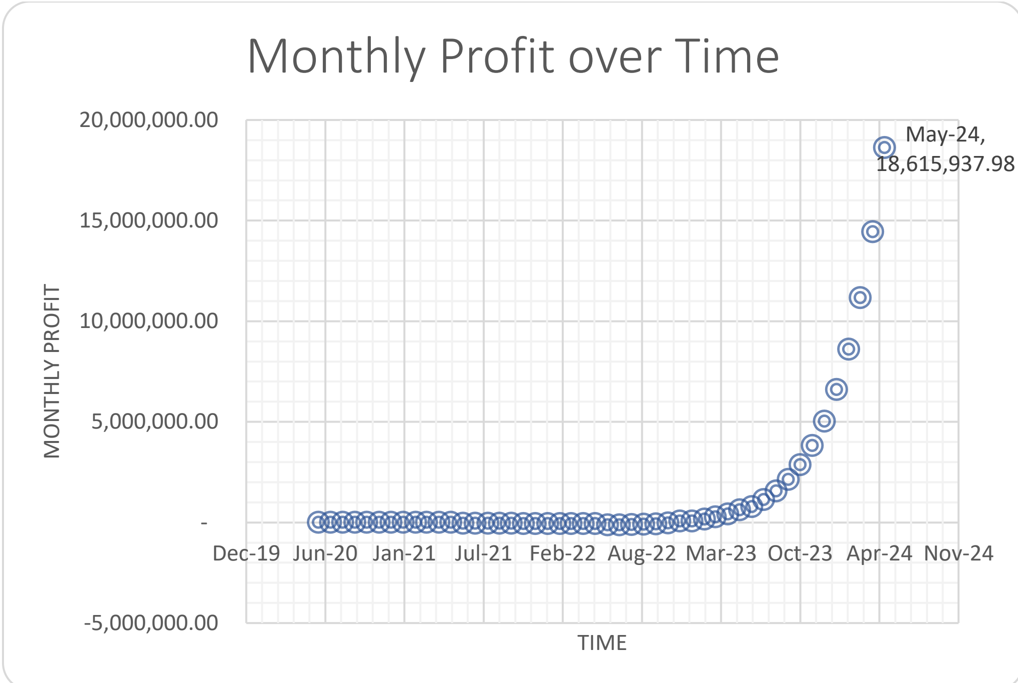


Figure 15: A projection of monthly profit (revenue – expenses) over a three-year period after monetization It is seen that monthly profit reaches a net positive just after December 2022.

It is projected that, at the point of profitability, Inflo will have nearly 2 million paying users, or approximately 5 percent of the total user base, which is projected to be 35 million users by then end of three years.

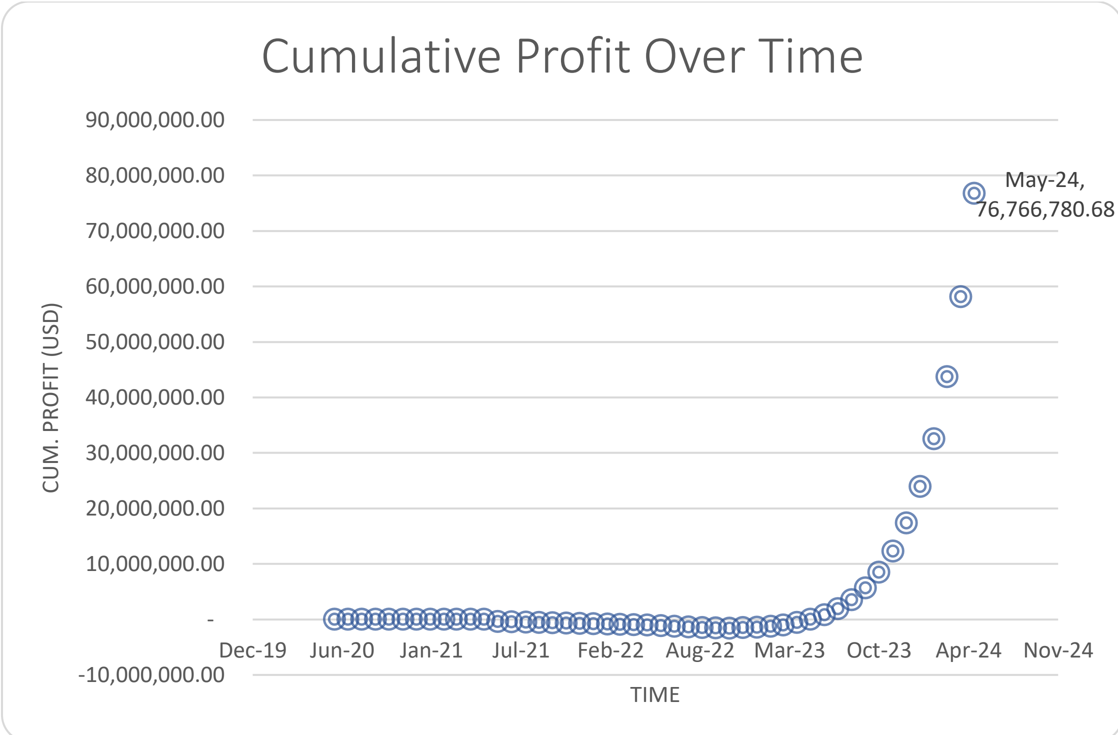


Figure 16: A projection of cumulative profit over a three-year period after monetization. It is seen that cumulative profit reaches a net positive just around May 2023, and at the end of three years post monetization, the overall profitability is approximately USD 76 million.

Budget and Required Funding

According to the revenue projections above, Inflo will need a sum of approximately US 7.4 million dollars in venture capital to cover all costs prior to achieving profitability. The respective expenses are broken down below.

Content: USD 2.5M

Manpower: USD 2.8M

Marketing: USD 2.1M

Operating costs: USD 90,000

Operating costs include office space, technology (cloud), patents and corporate registration costs. Full itemized expenses are provided in Appendix VII-D.

Chapter 4: Conclusion

The unique position of Inflo is that it is both a technology and market driven solution, using the latest machine learning and deployment tools in combination with a unique content sharing paradigm to create an engaging and insightful user experience. Inflo presents a new business model which feeds off of accurate and reliable content, creating an alternative to the exploitative ad-revenue model of conventional social media, while also creating a more sustainable revenue stream for publishers. This represents a shift in the digital news ecosystem towards a higher prioritization of information and quality.

On the technology side, Inflo's ongoing R&D efforts will continue to push the frontier of research in the field of online news propagation while integrating the findings thereof seamlessly into the commercial platform and in turn leveraging it into a sustainable and thriving business, and making an overall positive change in the social media ecosystem from both a user and publisher perspective.

The technical edge of Inflo is the deployment of fast deep learning models on the cloud and the results thereof being seamlessly integrated into the user experience for filtering and navigation. We are one of the (if not *the)* first companies to make use of ULMFiT and fast.ai's Python implementation thereof in a commercial, user facing platform. By keeping up with the latest technologies, and constantly improving speed and accuracy, we can expect to keep up with and even outpace any competitors.

From a consumer standpoint, Inflo aims to eventually be the primary source for news context on the web, as well as a complementary tool to other, non-news focused social platforms, as a hub for fact-checks, bias-checks, and expert insights, in addition to simply a means to share and discuss content of interest with a group or individual friends. In this way Inflo aims to assume a flexible role in the lives of social media users, allowing them to use it as casually or in-depth as they would like.

From a corporate perspective, Inflo aims to take the largest market share among social media platforms being used for the purpose of news, and as a result, to be the largest source of revenue for the top players in the digital news industry, much like Spotify has done for music, and Netflix for movies.

References

[1] S. Kemp, "The Global State of Digital in October 2019," We Are Social, 23 October 2019. [Online]. Available: https://wearesocial.com/blog/2019/10/the-global-state-of-digital-in-october-2019.

[2] E. Shearer, M. Barthel and J. Gottfried , "News Habits on Facebook and Twitter," Pew Research Center, 14 July 2018. [Online]. Available: https://www.journalism.org/2015/07/14/news-habits-on-facebook-and-twitter/.

[3] D. Folkenflik, "Facebook Scrutinized Over Its Role In 2016's Presidential Election", Npr.org, 2017. [Online]. Available: https://www.npr.org/2017/09/26/553661942/facebook-scrutinized-over-its-role-in-2016s-presidential-election.

[4] C. Cadwalladr, "Facebook's role in Brexit -- and the threat to democracy," TED, 2019. [Online]. Available: https://www.ted.com/talks/carole_cadwalladr_facebook_s_role_in_brexit_and_the_threat_to_democracy/details.

[5] M. Safi, "Revealed: Facebook hate speech exploded in Myanmar during Rohingya crisis", the Guardian, 2018. [Online]. Available: https://www.theguardian.com/world/2018/apr/03/revealed-facebook-hate-speech-exploded-in-myanmar-during-rohingya-crisis.

[6] A. Kiersz, "https://www.businessinsider.com/sri-lanka-blocked-access-facebook-instagram-social-media-easter-sunday-bombings-2019-4," Business Insider, Apr 21, 2019. [Online]. Available: https://www.businessinsider.com/sri-lanka-blocked-access-facebook-instagram-social-media-easter-sunday-bombings-2019-4.

[7] R. Kaser, "Facebook uncovers yet another fake news campaign", The Next Web, 2019. [Online]. Available: https://thenextweb.com/facebook/2018/07/31/facebook-uncovers-yet-another-fake-news-campaign/.

[8] "Global social media research summary 2019 | Smart Insights", Smart Insights, 2019. [Online]. Available: https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/.

[9] C. Schmidt, "Americans expect to get their news from social media, but they don’t expect it to be accurate," Nieman Lab, Sept. 10, 2018. [Online]. Available: https://www.niemanlab.org/2018/09/americans-expect-to-get-their-news-from-social-media-but-they-dont-expect-it-to-be-accurate/.

[10] M. Ingram, "Most Americans say they have lost trust in the media," Columbia Journalism Review, 12 September 2018. [Online]. Available: https://www.cjr.org/the_media_today/trust-in-media-down.php.

[11] D. Boyd and S. Yardi, "Dynamic Debates: An Analysis of Group Polarization Over Time on Twitter," *Bulletin of Science, Technology & Society*, 2010.

[12] "How Filter Bubbles Distort Reality: Everything You Need to Know," Farnam Street, 31 07 2017. [Online]. Available: https://fs.blog/2017/07/filter-bubbles/.

[13] K. Garimella, G. De Francisci Morales, A. Gionis and M. Mathioudakis, "Political Discourse on Social Media: Echo Chambers, Gatekeepers, and the Price of Bipartisanship," in *World Wide Web Conference*, 2018 World Wide Web Conference, 2018.

[14] S. B. Kleinman and S. Knobloch-Westerwick, "Preelection Selective Exposure: Confirmation Bias Versus Informational Utility," *Communication Research*, vol. 39, no. 2, pp. 170-193, 2011.

[15] P. Boczkowski, "Fake news and the future of journalism," Nieman Lab, 2017. [Online]. Available: https://www.niemanlab.org/2016/12/fake-news-and-the-future-of-journalism/.

[16] S. Vosoughi, D. Roy and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, pp. 1146-1151, 2018.

[17] P. Agrawal, N. Sukiennik and P. Hui, "Inflo: News Categorization and Keyphrase Extraction for Implementation in an Aggregation System," 2018. [Online]. Available: https://arxiv.org/abs/1812.03781.

[18] Miscellaneous: Ny times api, https://developer.nytimes.com/.

[19] Miscellaneous: Cnn news api, https://developer.cnn.com/.

[20] "Miscellaneous: News api - a json api for live news and blog articles," [Online]. Available: https://newsapi.org/ .

[21] J. M. Patel, "Natural language processing (NLP): term frequency - inverse document frequency (Tf-idf) based vectorization in Python," 02 July 2019. [Online]. Available: http://jaympatel.com/2019/02/natural-language-processing-nlp-term-frequency-inverse-document-frequency-tf-idf-based-vectorization-in-python/.

[22] Honnibal, M., Montani, I.: spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing. To appear (2017).

[23] Kim, Y.: Convolutional neural networks for sentence classification. arXiv preprint , arXiv:1408.5882 (2014) .

[24] Zhang, X., Zhao, J., LeCun, Y.: Character-level convolutional networks for text classification. In: Advances in neural information processing systems. (2015) 649–657.

[25] Bengio, Y.: Deep learning of representations for unsupervised and transfer learning. In: Proceedings of ICML Workshop on Unsupervised and Transfer Learning. (2012) 17–36 .

[26] Merity, S., Keskar, N.S., Socher, R.: Regularizing and optimizing lstm language models. arXiv preprint arXiv:1708.02182 (2017).

[27] Howard, J., Ruder, S.: Universal language model fine-tuning for text classification. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Volume 1. (2018) 328–339 .

[28] Dewdney, N.: Named entity trends originating from social media. In: 24th International Conference on Computational Linguistics. (2012) 1.

[29] Newman, D., Chemudugunta, C., Smyth, P., Steyvers, M.: Analyzing entities and topics in news articles using statistical topic models. In: International conference on intelligence and security informatics, Springer (2006) 93–104.

[30] Paszke, A., Gross, S., Chintala, S., Chanan, G.: Pytorch (2017) .

[31] Boudin, F.: pke: an open source python-based keyphrase extraction toolkit. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations. (2016) 69–73.

[32] El-Beltagy, S.R., Rafea, A.: Kp-miner: A keyphrase extraction system for english and arabic documents. Information Systems 34(1) (2009) 132–144 .

[33] Miller, G.A.: Wordnet: a lexical database for english. Communications of the ACM 38(11) (1995) 39–41.

[34] Wan, X., Xiao, J.: Single document keyphrase extraction using neighborhood knowledge. In: AAAI. Volume 8. (2008) 855–860 .

[35] Bougouin, A., Boudin, F., Daille, B.: Topicrank: Graph-based topic ranking for keyphrase extraction. In: International Joint Conference on Natural Language Processing (IJCNLP). 2013) 543–551.

[36] Sterckx, L., Demeester, T., Deleu, J., Develder, C.: Topical word importance for fast keyphrase extraction. In: Proceedings of the 24th International Conference on World Wide Web, ACM (2015) 121–122.

[37] Florescu, C., Caragea, C.: Positionrank: An unsupervised approach to keyphrase extraction from scholarly documents. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Volume 1. (2017), 1105–1115.

[38] Boudin, F.: Unsupervised keyphrase extraction with multipartite graphs. arXiv preprint arXiv:1803.08721 (2018) .

[39] D. M. Blei, A. Y. Ng and M. I. Jordan, "Latent Dirichlet Allocation," in *JMLR*, 2003.

[40] "Twitter API," [Online]. Available: https://developer.twitter.com/.

[41] "Wikipedia (Mediawiki) API," [Online]. Available: https://www.mediawiki.org/wiki/API.

[42] "Render," [Online]. Available: www.render.com.

[43] V. R. Martinez, "FastAi and Render: Quick and easy way to create and deploy Computer Vision Models," Data Driven Investor, 6 April 2019. [Online]. Available: https://medium.com/datadriveninvestor/fastai-and-render-quick-and-easy-way-to-create-and-deploy-computer-vision-models-14fd3f903cfc.

[44] R. Merritt, "Deploying Machine Learning Models is Hard, But It Doesn't Have to Be," Anaconda, 9 June 2018. [Online]. Available: https://www.anaconda.com/deploying-machine-learning-models-is-hard-but-it-doesnt-have-to-be/ .

[45] T. Ray, "Fast.ai's software could radically democratize AI," ZDNet, 2 October 2018. [Online]. Available: https://www.zdnet.com/article/fast-ais-new-software-could-radically-democratize-ai/.

[46] X. Zhang, J. Zhao and Y. LeCun , "Character-level Convolutional Networks for Text Classification," in *NIPS*, 2015.

[47] "Misinformation Directory," FactCheck.org, 12 November 2018. [Online]. Available: https://www.factcheck.org/2017/07/websites-post-fake-satirical-stories/.

[48] S. Volkova, T. Wilson and D. Yarowsky, "Exploring Sentiment in Social Media: Bootstrapping Subjectivity Clues from Multilingual Twitter Streams," in *Association for Computational Linguistics*, Sofia, Bulgaria, 2013.

[49] S. Bista, S. Nepal, N. Colineau and C. Paris, "Using Gamification in an Online Community," 2012.

[50] "StackOverflow," [Online]. Available: stackoverflow.com.

[51] "Quora," [Online]. Available: www.quora.com.

[52] "Reddit," [Online]. Available: www.reddit.com.

[53] Nakamura, J., Csikszentmihalyi, M.: Flow theory and research. Handbook of positive psychology pp. 195–206 (2009).

[54] H. M. Reza , G. Sujit , X. Ma and P. Hui, "HRCR: Hidden Markov-Based Reinforcement to Reduce Churn in Question Answering Forums," in *Pacific Rim International Conference on Artificial Intelligence*, 2019.

[55] L. N. Hoang, "Mechanism Design and the Revelation Principle," Science4All, 19 April 2012. [Online]. Available: http://www.science4all.org/article/mechanism-design/.

[56] P. Legros and E. Cantillon, "The Nobel Prize: What is mechanism design and why does it matter for policy-making?," VOX CEPR Policy Portal, 18 October 2007. [Online]. Available: https://voxeu.org/article/nobel-prize-what-mechanism-design-and-why-does-it-matter.

[57] J. N. Matias, "The Civic Labor of Online Moderators," The Platform Society, 2016.

[58] J. Borge-Holthoefer, R. A. Baños , S. González-Bailón and Y. Moreno, "Cascading behaviour in complex socio-technical networks," *Journal of Complex Networks*, vol. Volume 1, no. Issue 1, https://doi.org/10.1093/comnet/cnt006, p. Pages 3–24, June 2013.

[59] "React," [Online]. Available: reactjs.org/.

[60] Django, [Online]. Available: https://www.djangoproject.com/.

[61] "AWS," Amazon, [Online]. Available: https://aws.amazon.com/.

[62] L. Ou-Yang, "Newspaper3k: Article scraping & curation," [Online]. Available: https://newspaper.readthedocs.io/en/latest/.

[63] [Modeling Echo Chambers on Social Media. Sasahara, K.; Ciampaglia, G. L.; Flammini, A.; and Menczer, F. 2017. Presented at 4th Annual International Conference on Computational Social Science (IC2S2)].

[64] P. Nakov, "Semantic Sentiment Analysis of Twitter Data," in *Encyclopedia of Social Network Analysis and Mining*, 2017, pp. 1-12.

[65] A. L. Blackler, V. Popovic and D. Mahar, "Intuitive Interaction Applied to Interface Design," in *Intuitive Interaction Applied to Interface Design*, Douliou, Taiwan, 2005.

[66] J. Silva, "Logical fallacies and discourse," News Literacy Project, 15 August 2018. [Online]. Available: https://newslit.org/educators/civic-blog/logical-fallacies-and-discourse/.

[67] J. Wanner, C. Bauer and C. Janiesch, "Two-Sided Digital Markets: Disruptive Chance Meets Chicken or Egg Causality Dilemma," in *IEEE*, 2019.

[68] "Tweepy," [Online]. Available: http://docs.tweepy.org/en/v3.5.0/.

[69] M. Lindsay, "Digital Revenue Streams: What can newspapers teach other industries?," News Media Alliance, 12 December 2016. [Online]. Available: https://www.newsmediaalliance.org/newspapers-teach-other-industries/.

[70] F. Filloux, "Quality news shifts to a paid-for model. And that's good thing. But…," Monday Note, 6 Feb 2017. [Online]. Available: https://mondaynote.com/quality-news-shifts-to-a-paid-for-model-and-thats-good-thing-but-1b72a86a3b35.

[71] L. Moses, "The pivot to reality for digital media," Digiday, 20 November 2017. [Online]. Available: https://digiday.com/media/pivot-reality-digital-media/.

[72] N. Usher, "Digital Journalism Venture-backed News Startups and the Field of Journalism," *Digital Journalism*, p. DOI: 10. 1080/21670811.2016.1272064, 29 Nov 2017.

[73] M. Boland, "THE PUBLISHER PAYWALL REPORT: Variations, drivers for adoption, examples, and future of the pay-for-content model," Business Insider, 2016.

[74] "Publishers' key tools for driving subscriptions in 2019: Dynamic paywalls, newsletters, and sports subs," What's New in Publishing, May 2019. [Online]. Available: https://whatsnewinpublishing.com/publishers-key-tools-for-driving-subscriptions-in-2019-dynamic-paywalls-newsletters-and-sports-subscriptions.

[75] E. K. Thorpe, "The fatal flaw in subscription models," The Media Briefing, 9 Mar 2017. [Online]. Available: https://www.themediabriefing.com/analysis/the-fatal-flaw-in-subscription-models/.

[76] T. Rose-Stockwell, "This Is How Your Fear and Outrage Are Being Sold for Profit," Social Media Week, 9 October 2017. [Online]. Available: https://socialmediaweek.org/blog/2017/10/fear-outrage-sold-profit/.

[77] J. Marshall, "There's a Digital Media Crash. But No One Will Say It," Talking Points Memo, 17 November 2017 . [Online]. Available: https://talkingpointsmemo.com/edblog/theres-a-digital-media-crash-but-no-one-will-say-it.

[78] "inkl," [Online]. Available: http://www.inkl.com/newsletters/morning-edition.

[79] "Blendle," [Online]. Available: https://blendle.com/.

[80] L. Shekhtman, "Blendle: Can an 'iTunes for journalism' save the industry?," The Christian Science Monitor, [Online]. Available: https://www.csmonitor.com/Business/2016/0324/Blendle-Can-an-iTunes-for-journalism-save-the-industry.

[81] M. Barthel, "Newspapers Fact Sheet," Pew Research Center, 9 July 2019. [Online]. Available: https://www.journalism.org/fact-sheet/newspapers/.

Appendices

Appendix I – Bias Elicitation Module, In-depth Analysis

The dataset used for the hypothesis validation was supplied by Ad Fontes Media, a non-profit organization which has created a popular Media Bias Chart [2]. The data comes from a study which was conducted to create the most recent version of the chart, which was released in mid-2019. The study entailed gathering a group of individuals to rate a set of articles across a set of predefined sources over a period of several days[3].

Dataset Investigation

The dataset in total contains bias ratings of 1916 news articles across 108 sources. The bias ratings are provided on a points basis, with 5-point intervals each corresponding to a segment of the political spectrum: extreme left, moderate left, neutral, moderate right, and extreme right. The point values for each interval are specified as -36 to -18, -17 to -6, -5 to +5, +6 to +17, and +18 to +36, respectively.

Preliminary Dataset Validation

In order to determine the reliability of the dataset prior to performing our analysis, we compared the bias rankings with those of another well-known bias-checking resource: Media Bias Fact Check (MBFC). To do so, the average bias rankings from Ad Fontes source were calculated and converted to their corresponding interval (e.g. extreme left, moderate right, etc.) and cross-compared with the bias classification from MBFC for the same sources. In doing this, it's important to note here that MBFC does not provide the individual articles they used to classify the bias for each source.

---

[2] "Interactive Media Bias Chart 5.0," Ad Fontes Media, October 2019. [Online]. Available: https://www.adfontesmedia.com/interactive-media-bias-chart/.

[3] V. Otero, "Why Measuring Political Bias is So Hard, and How We Can Do It Anyway: The Media Bias Chart Horizontal Axis," Ad Fontes Media, 2019. [Online]. Available: https://www.adfontesmedia.com/why-measuring-political-bias-is-so-hard-and-how-we-can-do-it-anyway-the-media-bias-chart-horizontal-axis/.

The comparison using the 5 labels specified above resulted in a match of 34% for the sources which were present across both Ad Fontes and MBFC.

Next, we compared the classifications across Ad Fontes and MBFC solely based on side (e.g. moderate and extreme left would be grouped together as "left"). In this case the result was a 65% match.

Finally, we normalized all bias ratings that fell into the "moderate" part of the spectrum as neutral, and compared only the sources which were found to be "extreme" across the two sources. The match rate in this case was 80%. From this we can conclude that sources which are considered extremely biased are much easier to classify and thus the results are more likely to agree.

From the results of the comparison overall, we concluded that the dataset from Ad Fontes was worth looking into more deeply, so we proceeded with the in-depth analysis.

Analysis Motivation

The motivation of the current analysis is to contribute a more reliable framework for evaluating bias rankings in news media. The first step in doing so is to create more discrete bias rankings for each source. That means, instead of placing sources on a left/right political spectrum, we generate a bias value per topic within each news source. This is more insightful than the generalized bias per source, because different sources might take different stances on different topics. Additionally, topics concerning regions outside the US may not easily fall onto the conventional left-right political spectrum, so those topics should be evaluated individually and on an appropriate spectrum (e.g. for or against, rather than liberal or conservative).

The Ad Fontes dataset includes a list of articles for each source and the individual ratings thereof (as opposed to a generalized bias score per source), which gives us the opportunity to classify the articles by topic and generate average bias scores for each topic per source.

After classifying articles by topic, we proceeded to undergo a systematic analysis of the results in order to determine the significance thereof, and any broader conclusions that could be derived. That procedure is outlined in full below.

Analysis Procedure

The procedure of the in-depth analysis consisted of the following steps:

First, we primed the data for analysis and made some preliminary observations.

Data priming involved the following steps:

1. Classify each article by topic. The topics were predefined based on the major topics being discussion leading up to the coming 2020 US presidential election. Those topics are as follows: "reproductive rights', "gun control", "immigration", "marijuana", "2020 democratic primaries", "Russia investigation", "Michael Cohen scandal", "Brexit", "North Korea". Any topics which did not fall into at least one from this list were labeled as "no topic". The topics were defined by extracting article text and feeding it through the ANCCS module described in Section 2.1.1. In total, approximately 28% of the articles fell into one or more of the predefined topics (525 total).
2. Take the average of each source for each of the defined topics.
3. Split the averages into those which were calculated from A) sources containing any number of articles per topics, and B) sources contained three or more articles per topics. These two groups will be referred to as "all" and "3 plus" for the rest of this report.
4. Plot each source on a left-right spectrum for each given topic, as well as the topics for each source (for both groups). This allows us to easily make qualitative judgement of the biases across topics and sources, and any differences across groups. Examples for each of these charts are provided below:

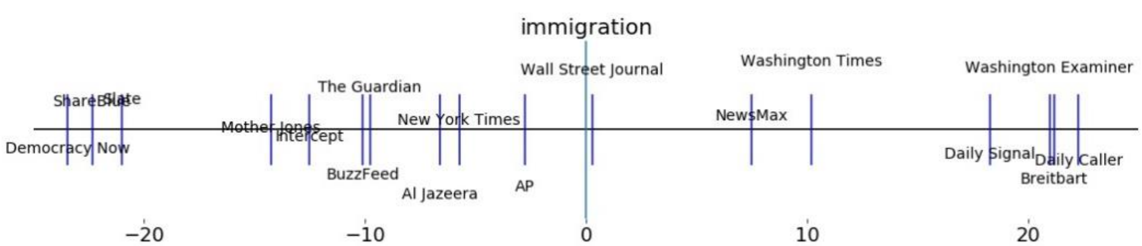


Figure A-1A - Sources per topic: "immigration", "3plus"

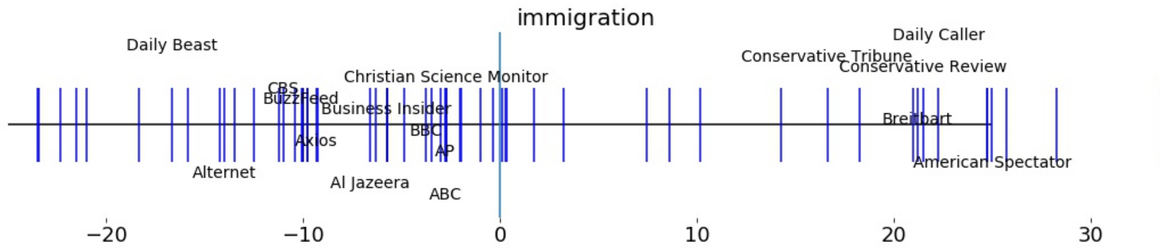


Figure A-1B - Sources per topic: "immigration", "all"

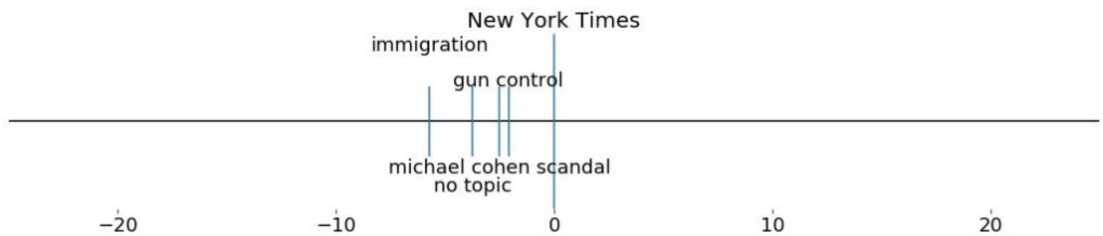


Figure A-2A - Topics per source: New York Times,"3plus"

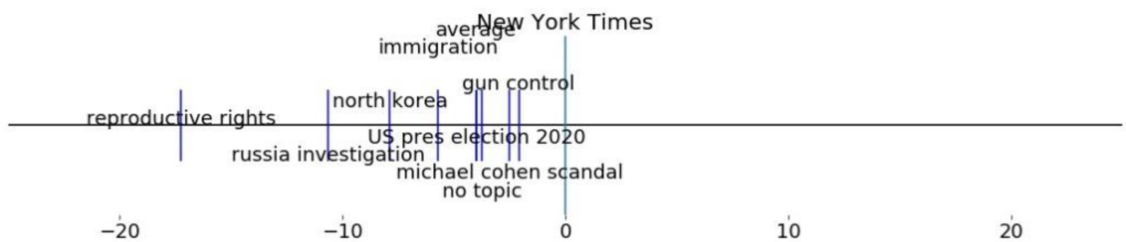


Figure A-2B - Topics per source: New York Times, "all″

Based on a qualitative comparison between the "all" and "3 plus" groups, it was observed that there was *much more side-agreement (all right, or all left) across topics when there were three or more articles present for a given topic (heretofore referred to as "original observation").* We then investigated why this was the case.

The three possible causes for the original observation are explained below:

1. One possible cause is that articles present in the "3 plus" group only included topics which have a clearly defined left/right distinction. This assumes that, out of the topics that were defined, some of them do not have clearly defined sides, leading to confusion of raters of the individual articles, causing them to rate those topics inconsistently.

2. Another possible cause is that articles present in the "3 plus" group only included sources which apply a fixed stance across all topics, meaning that some sources (those in the "all" group), have less control over the bias of their articles across different topics.
3. Finally, another possible cause is small sample bias, meaning that there are too few articles for a given topic in the "all" group for its bias rating to be valid.

In order to determine which reason was the actual cause of the original observation, the following plan was devised and executed.

1. Quantify the consistency of bias across sources. This meant comparing side-agreement for all topics under each source. If a given topic was opposite to the side of the overall source average, it was considered inconsistent (topic consistency). This stage is to validate the original observation.
2. Calculate the amount of variation across rankings within each topic given a source. This stage is to determine the presence of human error, and whether human error is present more often with inconsistent topics or not.
3. Finally, we conducted a source specific analysis to determine if agreement and consistency had any correlation with variation of ratings per topic for given sources. To do so, we first quantified agreement across articles under each topic/source (if a topic contains ratings on both sides for a given source, this is counted as one disagreement). Then we split the variation data into samples by separating the sources into two groups for a given topic, based on both A) consistency (e.g. those who rated that topic as consistent with the source average), and B) agreement (e.g. if a given topic contained and instances of disagreement. This analysis could help make conclusions on whether inconsistencies were a result of small sample bias or rather a deliberate effort by the publication. If the average standard deviation for the group with the consistent topics is higher, that could signify that the source has looser control of individual topics.

Topic Consistency:

After tabulating the number of inconsistencies per source, the topics which corresponded to each inconsistency were added up and divided by the total number of sources in which that topic was present, to generate a cumulative inconsistency rate per topic, for both the "all" and "3 plus" groups.

The results are provided below:

Table A-1: Total inconsistencies per topic for "all" group.

Topic	# sources in which topic is present	# inconsistencies per topic	inconsistency rate (%)
reproductive rights	16	1	6.3
gun control	22	3	13.6
immigration	59	4	6.8
marijuana	12	7	58.3
russia investigation	48	2	4.1
michael cohen scandal	45	5	11.1
Dem primaries 2020	53	1	1.9
Brexit	12	1	8.3
North Korea	26	7	26.9

From the table we can see that the topics with the highest inconsistency rate for the all group were "marijuana", "North Korea", "gun control", and "Michael Cohen scandal".

Next, we look at the results of the "3 plus" group.

Table A-2: Total inconsistencies per topic for "3 plus" group.

Topic	# sources in which topic is present	# inconsistencies per topic	inconsistency rate
reproductive rights	16	0	0
gun control	22	0	0
immigration	59	0	0
marijuana	12	0	0
russia investigation	48	0	0
michael cohen scandal	45	0	0
democratic primaries 2020	53	0	0
Brexit	12	0	0
North Korea	26	0	0

Thus, we can clearly see that the original observation is valid and there are no topic inconsistencies for the "3 plus" group.

The complete results for article consistency, featuring total number of inconsistencies, as well as total articles for each topic for each group, are provided in the figure below.

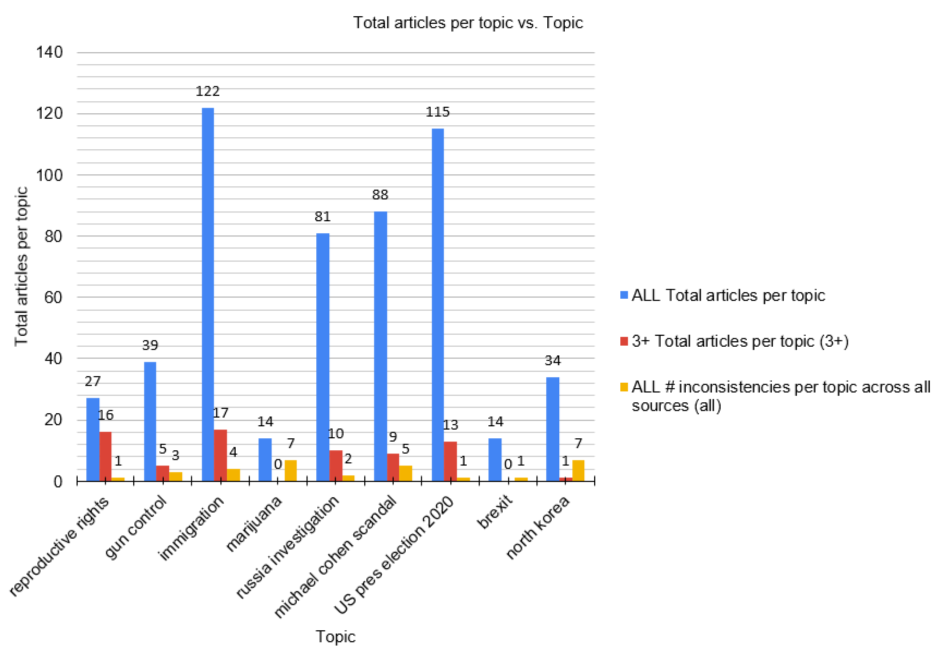


Figure A-1 - Topic Consistency Results and Article Count

From figure 3, we can see that some topics are more prone to inconsistency despite having a low number of total articles.

Variation Results

Next we provide the results of the variation in bias scores across topics per source. This is calculated by taking the standard deviation of all the articles under a given topic for each source, and then taking the average of the standard deviations across all sources for each topic.

First, we present the results for the "all" group.

Table A-3 - Average standard deviations across all sources for each topic, along with inconsistency rate, "all"

Topics	**Average stdev per topic**	**Inconsistency rate (%)**
reproductive rights	10.1	6.3
gun control	4.01	13.6
immigration	4.8	6.8
marijuana	1.94	58.3
US presidential election	5.21	4.1
russia investigation	4.81	11.1
michael cohen scandal	5.28	1.9
Brexit	N/A	8.3
North Korea	1.0	26.9

From the above results, we can conclude that there is no significant difference in the way articles for topics were rated across different consistencies. As such, we can rule out human error and conclude that inconsistencies are not caused by confusion of the raters.

Next, we present the results for variation within "3 plus" group.

Table A-4 - Average standard deviations across all sources for each topic, along with inconsistency rate, "3 plus"

Topics	Average stdev per topic	Inconsistency %
reproductive rights	6.22	6.3
gun control	3.12	13.6
immigration	5.55	6.8
marijuana	N/A	58.3
US presidential election	5.50	4.1
russia investigation	6.80	11.1
michael cohen scandal	6.77	1.9
Brexit	N/A	8.3
North Korea	1.0	26.9

We again note a similar non-correlation and reconfirm that there is no confusion among raters in evaluating the bias for different topics despite inconsistency.

Source Specific Analysis

The source specific analysis was conducted in the following scenarios:

a. Consistency split 1+, stdev averages 2+ (with agreement comparison 2+)
b. Consistency split 1+, stdev averages 2+ (with agreement comparison *3+*)
c. Consistency split *2+*, stdev averages 2+ (with agreement comparison 2+)
d. Consistency split 2+, stdev averages 2+ (with agreement comparison *3+*)
e. Agreement split 2+, stdev averages 2+ (with consistency comparison 1+)
f. Agreement split 2+, stdev averages *3+* (with consistency comparison 1+)
g. Agreement split 2+, stdev averages 2+ (with consistency comparison *2+*)
h. Agreement split 2+, stdev averages *3+* (with consistency comparison 2+)

For scenarios A through D, we created a sample split for each topic based on which sources had said topic as consistent or inconsistent. For each split, we compute the stdev of the source present for a particular topic, call it Stdev_source_topic. We then took a simple average of all the sources present for each topic, call it Stdev_topic. Then we take average over these topics using two statistics. One is simple average over Stdev_topic of all topics, the other is a weighted average, with the weight being #of sources per topic/#sources in the sample. The simple average of Stdev_topic gives higher weight to those Stdev_topic with fewer sources because it assumes the absent sources have the same stdev of the observed sources in that topic. The weighted average is intended to correct the overweighting. In other words, the weighted average is the simple average of the stdev of all sources across all topics, i.e. a simple average of Stdev_source_topic.

To come to a conclusion as to whether there is any correlation between variation and agreement, we compared both the simple and weighted resulting averages with the disagreement rate for each topic sample split (both consistent and inconsistent).

We then compared the stdev averages for each topic with the rate of inconsistency for the same topic. We added up the instances wherein the standard deviations were higher in the consistent sample vs in the inconsistent sample over all topic splits, and tabulated the totals for each scenario.

The outcomes were as follows:

A. Consistent sample tend to have higher standard deviation (6:3 for weighted and 7:2 unweighted)
B. Consistent sample tend to have higher standard deviation (5:2 for both weighted and unweighted)
C. Consistent sample tend to have lower standard deviation (2:3 for both weighted and unweighted)
D. Consistent sample tend to have lower standard deviation (2:3 for both weighted and unweighted)

Next, we looked at the disagreement rates for the same samples to determine whether the consistent or inconsistent samples for each topic basis had more disagreement:

A. All topics have a higher disagreement rate in the inconsistent sample
B. inconsistent sample has more disagreement (7:9)
C. inconsistent sample has more disagreement (2:3)
D. inconsistent sample has more disagreement (2:3)

Similarly, for groups E through H, we created a sample split for each topic based on which sources had said topic showing any instance of disagreement, or no disagreement, for each topic. The simple as weighted standard deviations were calculated for each topic per sample split as above. Then the resulting standard deviations (both simple and weighted), for each sample split were compared with the average inconsistency rate for the given split. We added up the instances wherein the standard deviations were higher in the agreement sample vs in the disagreement sample across all topic splits, and tabulated the totals for each scenario.

The outcomes for E through H are as follows:

E. Agreeable samples tend to have higher standard deviation (4:3 weighted, 5:2 unweighted)

F. Agreeable sample tend to have higher standard deviation (5:2 for both weighted and unweighted)
G. Agreeable sample tend to have higher standard deviation (4:3 for weighted and 5:2 unweighted)
H. Agreeable sample tend to have higher standard deviation (5:2 for weighted and 6:1 unweighted)

Next, we looked at the consistency rates for the same samples to determine whether the disagreement or no disagreement samples for each topic basis had more inconsistency:

E. Except for gun control, all topics which have a disagreement sample have higher inconsistency rates
F. Disagreement has more inconsistency (one is exactly the same) (1:5)
G. Disagreement has more inconsistency (3:4)
H. Disagreement has more inconsistency (3:4)

Conclusions

From the first consistency analysis (topic consistency), we determined that the topics that have the highest rate of inconsistency over all the sources were: "marijuana", "North Korea", "gun control", and "Michael Cohen scandal". We compare this with the "3 plus" data (which contains no inconsistencies) to find that two of these topics have a few or zero articles in that group:

"marijuana" (0), and "North Korea" (3). This points to small sample bias as a likely cause for the inconsistencies of those topics. So we proceed by investigating the cause of the inconsistencies for the other two topics, "gun control", and "Michael Cohen scandal", which arise when we look at the results for the "all" group. The remaining possibilities are: human error, sources having loose control of bias stance across topics, or, again, small sample bias.

We finally look at the results of variation analysis to rule out confirm human error. According to the results of the variation analysis, i.e. the averages of the standard deviations of ratings of articles under each topic for a given source, there is no significant difference in the variation of ratings whether the topic was inconsistent or not. Thus, we can rule out human error. Finally, in the source specific analysis, we can see that the standard deviations in almost all cases are higher for the consistent and no disagreement samples. From this we conclude that the inconsistent or disagreeing topics are most likely due to the news source's deliberately different position from the source average position rather than the small sample bias.

This conclusions above are in line with our hypothesis and as such, validate the use of this module in the Inflo platform, as a method to elicit article and user bias.

Appendix II – Preliminary User Rights Distribution

User privileges breakdown:

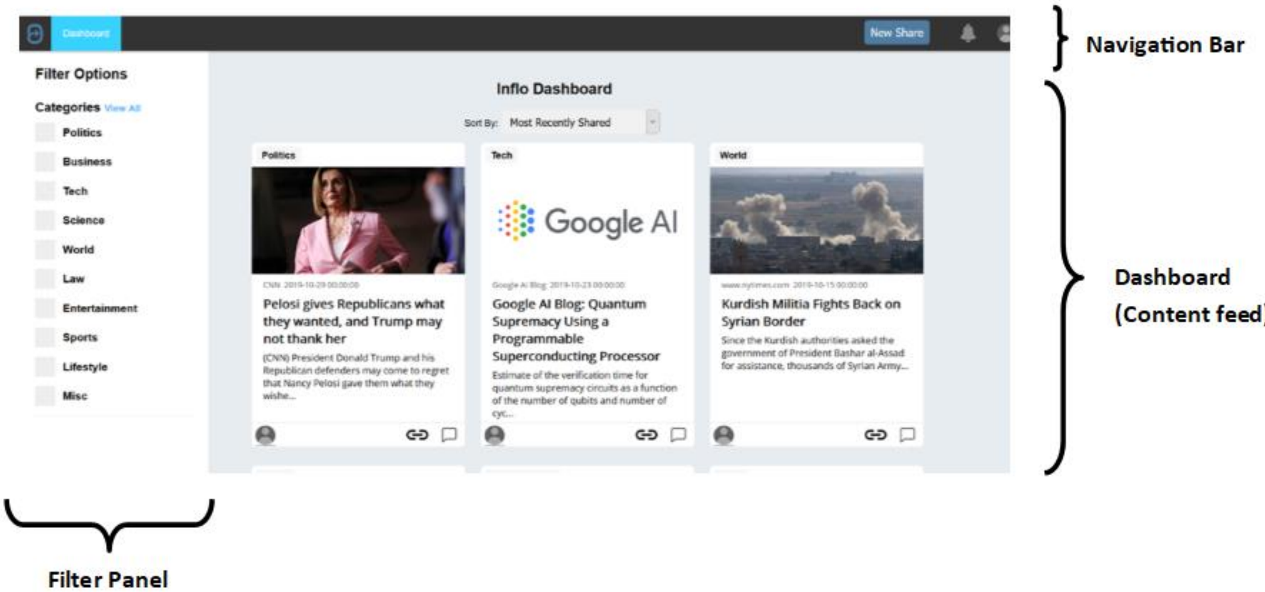
1. Admin
   - Can delete CBs, commentaries
   - Can ban users
   - Can promote users from common to expert
   - Can modify categories and add tags
2. Expert
   - Can upvote/downvote commentaries
   - Can modify categories and add tags
   - Can share CBs
   - Can write commentaries
3. Common/New User
   - Can write commentaries on existing CBs
   - After getting a certain number of upvotes, they can share new CBs

Privileges allocation:

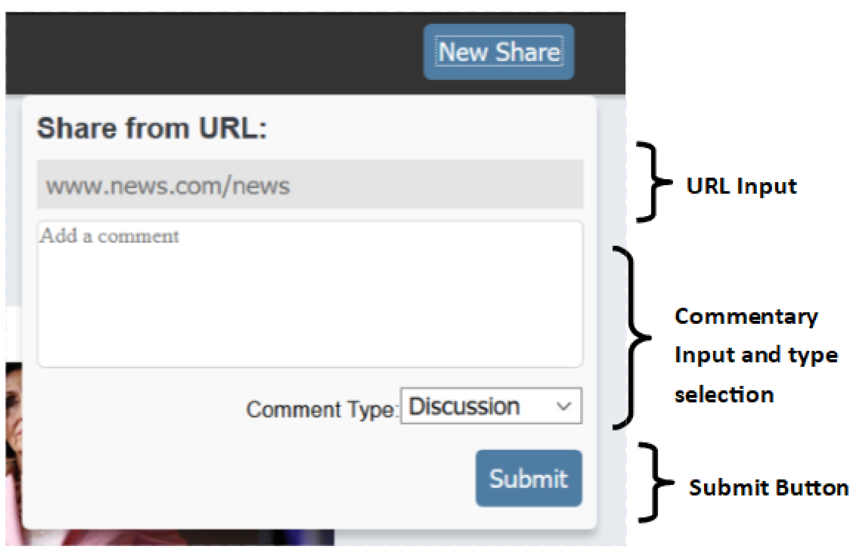
A common user can only be promoted to expert manually (by an admin) initially, and admin will initially be an employee of Inflo itself. This is to prevent exploitation from any potential loopholes in the system until it has been proven effective (and foolproof) over a long term.

Appendix III – Platform UI

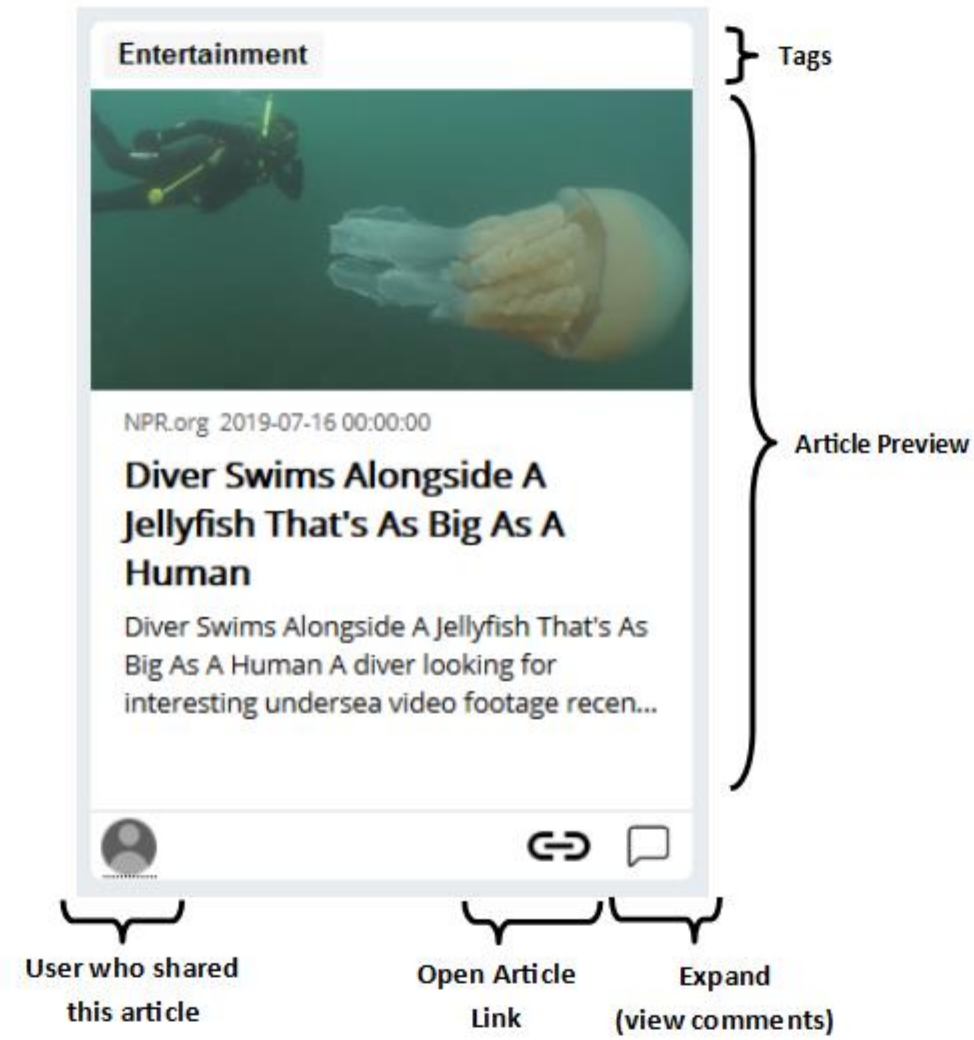
A. Main UI



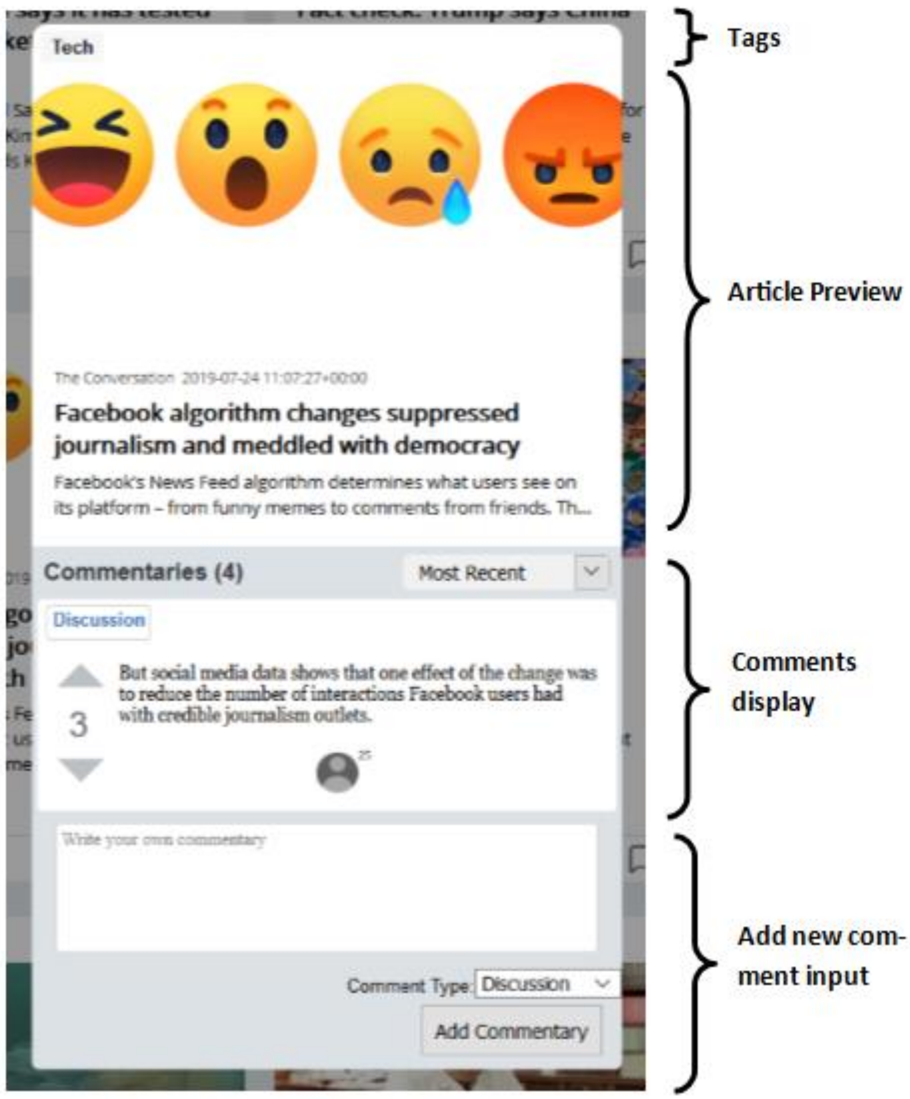
B. Share Panel



C. Content Block



D. Expanded Content Block



Appendix IV – Web-app infrastructure

A. Frontend:

React is used as the framework for the front-end application, which is a dynamic and responsive implementation of the user interface.

React is a framework which makes reuse of web elements trivial, by allowing components to be imported into others, and to be modified with new data according to their context (e.g. user, article). It also makes it easy to change or update layouts--when the layout is changed in a single Javascript component file, it will automatically update wherever that update is imported in any other components. React is ideal for the Inflo platform in that it allows for dynamic creation of UI elements without having to manually duplicate code.

The React frontend application is used in conjunction with the Django backend, by calling different function of the latter as API requests.

B. Backend:

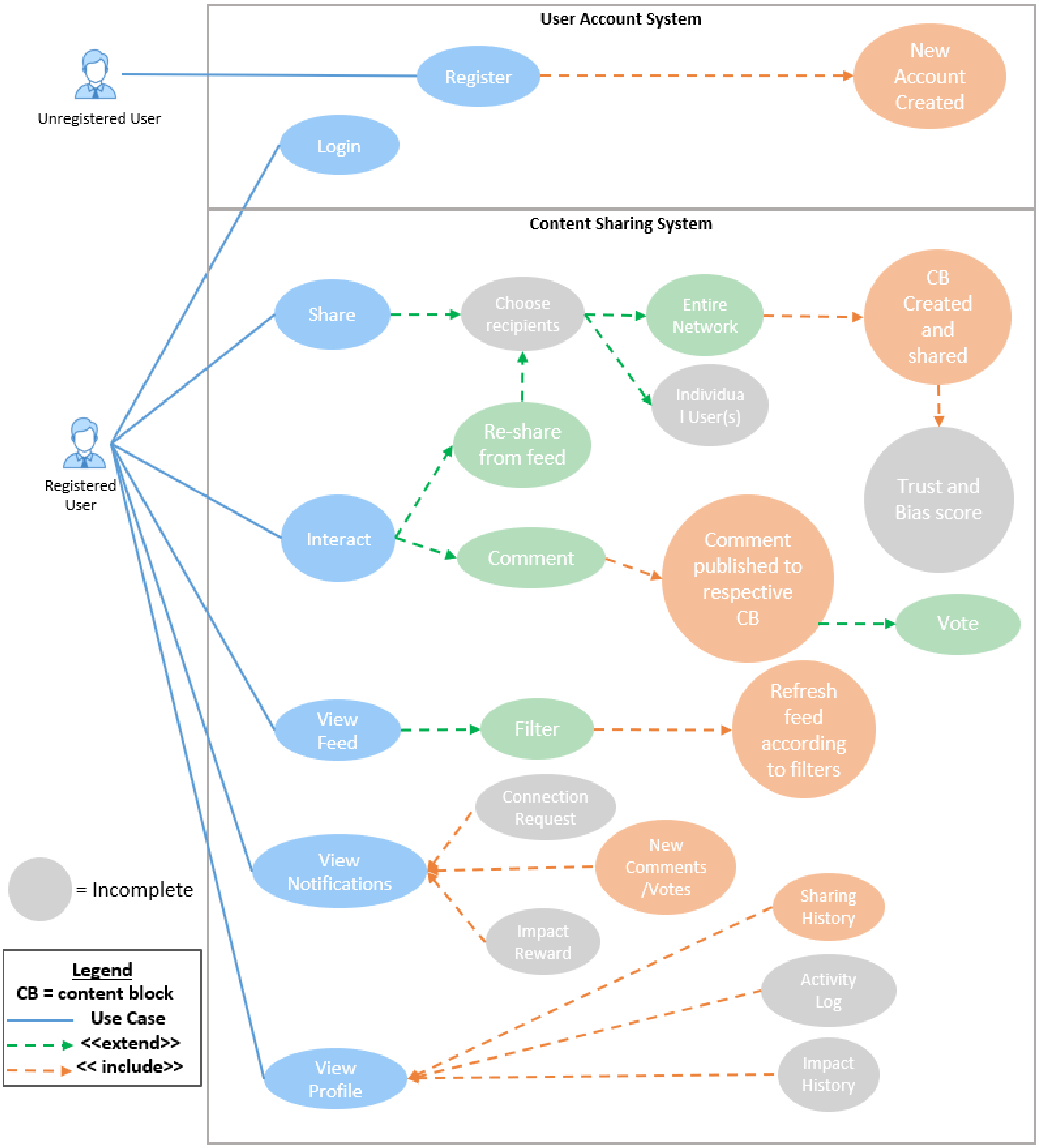
The backend uses Python's Django web framework in the form of a set of APIs that can be called from the frontend to save and query relevant data.

The database being used is a MySQL database.

C. Use-case diagram

The following is a depiction of the possible use-cases of the Inflo (alpha) platform, which is useful for software engineering purposes, but can also provide a broad view of the overall concept of the platform for laypersons. Thus, we provide it here:

Appendix IV-C: Inflo Use Case Diagram (Prototype)



Appendix V – Expected Features and Iterations Release Timeline

Iteration	Features				
	Content Classification/Navigation	*Social Network*	*Gamification*	*Supported Content*	*Other*
Alpha (January 2020)	• Filtering system, topics and categories	• Public Sharing • Commentaries and types	• voting	• English • Text • News	• Web-deployment
Beta (July 2020)	• Hierarchical tagging • Tag normalization articles • Search	• Personal connections • Twitter integration	• Impact Score • Expertise score • User Privileges	• Blogs • Tutorials • Academic Papers	• Bias Elicitation Module • Browser extension
Future Iterations (January 2021 Onwards)	• Article Summaries • Source analysis (reliable/unknown)	• Communities sharing • Request expert response	• User interest map and timeline	• Multi language Support • Support for image, video, audio	• Mobile deployment

Appendix VI - Intelligence analysis breakdown for a given article

Form of feedback	**Description of intelligence**	**Comment Type**	**Analysis method**
Reception of the article:	A general overview of public reception, whether the article was well/poorly received	Opinion	Sentiment analysis (machine learning) neg/pos [84][4]
Trust and accuracy:	whether people have pointed out corrections to reported facts/figures	Fact check	Human input
Bias perception:	how much/whether people felt the article represented a certain political leaning/bias.	Bias Check	Machine learning for bias detection [5]
Content completion:	Whether the article is lacking essential information to provide a complete picture	Context	Human input

[4] Smith, K.S., McCreadie, R., Macdonald, C. et al. Inf Syst Front (2018) 20: 1013. https://doi.org/10.1007/s10796-018-9827-x.

[5] A. Patankar and J. Bose, "Bias Discovery in News Articles Using Word Vectors," in *IEEE*, https://ieeexplore.ieee.org/abstract/document/8260730, 2017.

Appendix VII – Detailed Financial Projections

The full Inflo business model including revenue calculations over the three-year projection period are provided in excel format for convenience. They can be accessed at the following link:

https://docs.google.com/spreadsheets/d/1oEUxg8IWTKHYUaNaCnOKY5kuZLHg65Fliq4L_piApkw/edit?usp=sharing

The broken down calculations spreadsheets are copied below for convenience.

Appendix VII-A: Marketing Budget

Source	Cost/mo (USD)	Traffic	CTR	Conversion %	Total Users
Medium	0	26,000	20.00%	5.00%	260
Duckduckgo ads	3000	100000	10.00%	10.00%	1000
Press	0	120000	10.00%	10.00%	1200
SEO	0	50000	10.00%	5.00%	250
Quora	3000	100000	25.00%	5.00%	1250
Twitter Ads	3000	100000	10.00%	10.00%	1000
Total Cost / mo	9000				

Month	Medium	Duckduckgo	Press	SEO	Quora	Twitter Ads	Total
1	260	1000	1200	250	1250	1000	3960
2	520	2000	2400	500	2500	2000	7920
3	780	3000	3600	750	3750	3000	11880
4	1040	4000	4800	1000	5000	4000	15840
5	1300	5000	6000	1250	6250	5000	19800
6	1560	6000	7200	1500	7500	6000	23760
7	1820	7000	8400	1750	8750	7000	27720
8	2080	8000	9600	2000	10000	8000	31680
9	2340	9000	10800	2250	11250	9000	35640
10	2600	10000	12000	2500	12500	10000	39600
11	2860	11000	13200	2750	13750	11000	43560
12	3120	12000	14400	3000	15000	12000	59520

Appendix VII-B: Monetization Scenario (Detailed Projection)

User Growth Projection			
Month	**Date**	**Percentage of Available Market Reached**	Number of Users Reached
1	Jun-21	0.02%	60,000.00
2	Jul-21	0.02%	72,000.00
3	Aug-21	0.02%	86,400.00
4	Sep-21	0.03%	103,680.00
5	Oct-21	0.03%	124,416.00
6	Nov-21	0.04%	149,299.20
7	Dec-21	0.05%	179,159.04
8	Jan-22	0.06%	214,990.85
9	Feb-22	0.07%	257,989.02
10	Mar-22	0.08%	309,586.82
11	Apr-22	0.10%	371,504.19
12	May-22	0.12%	445,805.02
13	Jun-22	0.14%	534,966.03
14	Jul-22	0.17%	641,959.23
15	Aug-22	0.21%	770,351.08
16	Sep-22	0.25%	924,421.29
17	Oct-22	0.30%	1,109,305.55
18	Nov-22	0.35%	1,331,166.66
19	Dec-22	0.43%	1,597,400.00
20	Jan-23	0.51%	1,916,880.00
21	Feb-23	0.61%	2,300,256.00
22	Mar-23	0.74%	2,760,307.19
23	Apr-23	0.88%	3,312,368.63
24	May-23	1.06%	3,974,842.36
25	Jun-23	1.27%	4,769,810.83
26	Jul-23	1.53%	5,723,773.00
27	Aug-23	1.83%	6,868,527.60
28	Sep-23	2.20%	8,242,233.12
29	Oct-23	2.64%	9,890,679.74
30	Nov-23	3.17%	11,868,815.69
31	Dec-23	3.80%	14,242,578.83
32	Jan-24	4.56%	17,091,094.59
33	Feb-24	5.47%	20,509,313.51
34	Mar-24	6.56%	24,611,176.21
35	Apr-24	7.88%	29,533,411.46
36	May-24	9.45%	35,440,093.75

Cash Flow Estimate (USD)				
	Date	Cost per paying user	Percentage of users paying	Number of users paying
12	Jun-21	30.00	1.00%	600
13	Jul-21	27.86	1.05%	756
14	Aug-21	25.87	1.10%	953
15	Sep-21	24.02	1.16%	1200
16	Oct-21	22.30	1.22%	1512
17	Nov-21	20.71	1.28%	1905
18	Dec-21	19.23	1.34%	2401
19	Jan-22	17.86	1.41%	3025
20	Feb-22	16.58	1.48%	3812
21	Mar-22	15.40	1.55%	4803
22	Apr-22	14.30	1.63%	6051
23	May-22	13.28	1.71%	7625
24	Jun-22	12.33	1.80%	9607
25	Jul-22	11.45	1.89%	12105
26	Aug-22	10.63	1.98%	15252
27	Sep-22	9.87	2.08%	19218
28	Oct-22	9.17	2.18%	24215
29	Nov-22	8.51	2.29%	30511
30	Dec-22	7.90	2.41%	38443
31	Jan-23	7.34	2.53%	48439
32	Feb-23	6.81	2.65%	61033
33	Mar-23	6.33	2.79%	76901
34	Apr-23	5.88	2.93%	96895
35	May-23	5.46	3.07%	122088
36	Jun-23	5.07	3.23%	153831
37	Jul-23	4.70	3.39%	193827
38	Aug-23	4.37	3.56%	244222
39	Sep-23	4.06	3.73%	307720
40	Oct-23	3.77	3.92%	387727
41	Nov-23	3.50	4.12%	488537
42	Dec-23	3.25	4.32%	615556
43	Jan-24	3.02	4.54%	775601
44	Feb-24	2.80	4.76%	977257
45	Mar-24	2.60	5.00%	1231344

46	Apr-24	2.41	5.25%	1551493
47	May-24	2.24	5.52%	1954881

Cash Flow Estimate (USD)				
	Date	**Cum exp**	**Cumulative revenue**	**Accum Profit**
1	Jun-20	30,857.58	-	-
2	Jul-20	61,715.16	-	-
3	Aug-20	92,572.74	-	-
4	Sep-20	123,430.32	-	-
5	Oct-20	154,287.90	-	-
6	Nov-20	185,145.48	-	-
7	Dec-20	216,003.06	-	-
8	Jan-21	246,860.64	-	-
9	Feb-21	277,718.22	-	-
10	Mar-21	308,575.80	-	-
11	Apr-21	339,433.38	-	-
12	May-21	370,290.96	-	-
13	Jun-21	420,048.54	7,200.00	-412,848.54
14	Jul-21	473,856.12	16,272.00	-457,584.12
15	Aug-21	532,332.90	27,702.72	-504,630.18
16	Sep-21	596,196.41	42,105.43	-554,090.99
17	Oct-21	666,278.55	60,252.84	-606,025.72
18	Nov-21	743,544.25	83,118.58	-660,425.67
19	Dec-21	829,113.24	111,929.41	-717,183.83

20	Jan-22	924,285.47	148,231.05	-776,054.42
21	Feb-22	1,030,570.74	193,971.13	-836,599.62
22	Mar-22	1,149,723.21	251,603.62	-898,119.59
23	Apr-22	1,283,781.75	324,220.56	-959,561.19
24	May-22	1,435,116.99	415,717.90	-1,019,399.09
25	Jun-22	1,677,486.23	531,004.56	-1,146,481.67
26	Jul-22	1,943,097.51	676,265.74	-1,266,831.77
27	Aug-22	2,235,684.50	859,294.84	-1,376,389.66
28	Sep-22	2,559,593.82	1,089,911.49	-1,469,682.33
29	Oct-22	2,919,887.11	1,380,488.48	-1,539,398.63
30	Nov-22	3,322,460.15	1,746,615.49	-1,575,844.66
31	Dec-22	3,719,138.89	2,207,935.51	-1,511,203.38
32	Jan-23	4,228,014.73	2,789,198.75	-1,438,815.98
33	Feb-23	4,803,375.34	3,521,590.42	-1,281,784.92
34	Mar-23	5,456,099.31	4,444,403.93	-1,011,695.38
35	Apr-23	6,198,872.20	5,607,148.95	-591,723.25
36	May-23	7,046,489.48	7,072,207.68	25,718.21
37	Jun-23	8,104,710.57	8,918,181.68	813,471.11
38	Jul-23	9,305,172.60	11,244,108.92	1,938,936.32
39	Aug-23	10,671,373.93	14,174,777.24	3,503,403.30
40	Sep-23	12,230,739.41	17,867,419.32	5,636,679.91
41	Oct-23	14,015,281.00	22,520,148.34	8,504,867.34
42	Nov-23	16,062,370.12	28,382,586.91	12,320,216.79

43	Dec-23	18,415,640.48	35,769,259.50	17,353,619.02
44	Jan-24	21,126,043.58	45,076,466.98	23,950,423.39
45	Feb-24	24,253,082.74	56,803,548.39	32,550,465.65
46	Mar-24	27,866,255.92	71,579,670.97	43,713,415.05
47	Apr-24	32,046,742.72	90,197,585.42	58,150,842.70
48	May-24	36,889,376.96	113,656,157.63	76,766,780.68

Appendix VII-C: Corporate Budget

Manpower Costs	
Position	**Yearly Rate (USD)**
Frontend developer	65000
Backend developer	65000
Data Scientist	80000
Marketers	50000
Interns	24000
CEO	65000

Other Corporate Expenses			
Item	**Cost/mo.**	**Unit Cost**	**Total/yr.**
AWS Hosting	342.58		4110.96
Render Hosting	15		60
Patents			2000
Corporate Registration			2000
Tax			21%
Months	**Manpower**	**Additional corporate expenses**	Cum all corp + manpower + tax
1	17500	4357.58	21857.58

2	35000	8715.16	43715.16
3	52500	13072.74	65572.74
4	70000	17430.32	87430.32
5	87500	21787.9	109287.9
6	105000	26145.48	131145.48
7	122500	30503.06	153003.06
8	140000	34860.64	174860.64
9	157500	39218.22	196718.22
10	175000	43575.8	218575.8
11	192500	47933.38	240433.38
12	210000	52290.96	262290.96
13	227500	56648.54	284148.54
14	245000	61006.12	306006.12
15	262500	65363.7	327863.7
16	280000	69721.28	349721.28
17	297500	74078.86	371578.86
18	315000	78436.44	393436.44
19	332500	82794.02	415294.02
20	350000	87151.6	437151.6
21	367500	91509.18	459009.18
22	385000	95866.76	480866.76
23	402500	100224.34	502724.34
24	420000	104581.92	524581.92
25	508500	108939.5	617439.5
26	597000	113297.08	710297.08
27	685500	117654.66	803154.66
28	774000	122012.24	896012.24
29	862500	126369.82	988869.82
30	951000	130727.4	1081727.4
31	1039500	135084.98	1188159.65
32	1128000	139442.56	1282643.914
33	1216500	143800.14	1393276.663
34	1305000	148157.72	1509876.523
35	1393500	152515.3	1634209.449
36	1482000	156872.88	1768535.585
37	1659000	161230.46	1985658.57
38	1836000	165588.04	2237935.734
39	2013000	169945.62	2511483.687
40	2190000	174303.2	2812291.287
41	2367000	178660.78	3147980.141
42	2544000	183018.36	3528241.743

43	2721000	187375.94	3965390.409
44	2898000	191733.52	4475062.438
45	3075000	196091.1	5077099.973
46	3252000	200448.68	5796668.054
47	3429000	204806.26	6665666.067
48	3606000	209163.84	7724510.815

Appendix VII-D: Publisher Revenue (Detailed Projection)

Publisher Revenue Projection					
Month	**% pub users users being lost to IF / pub**	**# pub subs leaving pub / sub**	**Revenue gain /pub from inflo**	**Revenue from IF after assumption of loss of users to IF / pub**	**Cumulative Rev**
13	0.00%	9.00000	1,620	1,512.0	1,512.0
14	0.00%	11.34000	1,895	1,759.3	3,271.3
15	0.01%	14.28840	2,218	2,046.2	5,317.5
16	0.01%	18.00338	2,595	2,378.6	7,696.0
17	0.01%	22.68426	3,036	2,763.5	10,459.5
18	0.01%	28.58217	3,552	3,208.8	13,668.3
19	0.01%	36.01354	4,156	3,723.4	17,391.7
20	0.02%	45.37706	4,862	4,317.5	21,709.2
21	0.02%	57.17509	5,689	5,002.5	26,711.7
22	0.03%	72.04062	6,656	5,791.1	32,502.8
23	0.04%	90.77118	7,787	6,697.8	39,200.6
24	0.05%	114.37168	9,111	7,738.4	46,939.0

25	0.06%	144.10832	10,660	8,930.4	55,869.4
26	0.07%	181.57648	12,472	10,292.9	66,162.3
27	0.09%	228.78637	14,592	11,846.6	78,009.0
28	0.12%	288.27082	17,073	13,613.5	91,622.5
29	0.15%	363.22124	19,975	15,616.4	107,238.9
30	0.18%	457.65876	23,371	17,879.0	125,117.9
31	0.23%	576.65003	27,344	20,424.1	145,542.0
32	0.29%	726.57904	31,992	23,273.4	168,815.4
33	0.37%	915.48959	37,431	26,445.2	195,260.6
34	0.46%	1,153.51689	43,794	29,952.2	225,212.7
35	0.58%	1,453.43128	51,239	33,798.2	259,010.9
36	0.73%	1,831.32341	59,950	37,974.2	296,985.1
37	0.92%	2,307.46750	70,142	42,452.0	339,437.1
38	1.16%	2,907.40905	82,066	47,176.8	386,613.9
39	1.47%	3,663.33540	96,017	52,056.8	438,670.7
40	1.85%	4,615.80260	112,340	56,950.1	495,620.8
41	2.33%	5,815.91128	131,437	61,646.5	557,267.3
42	2.93%	7,328.04821	153,782	65,845.2	623,112.6
43	3.69%	9,233.34075	179,925	69,124.6	692,237.2
44	4.65%	11,634.00934	210,512	70,903.8	763,141.0
45	5.86%	14,658.85177	246,299	70,392.7	833,533.8
46	7.39%	18,470.15323	288,170	66,528.0	900,061.7
47	9.31%	23,272.39307	337,159	57,889.9	957,951.7

Appendix VIII – Additional Company/Corporate information

IP/Copyright

We are currently (Dec. 2019) in the process of registering a software patent with the HKUST Technology Transfer Office for the bias elicitation algorithm.

Corporate Information

Inflo was incorporated as a limited private corporation in Hong Kong on October 2018. Nicholas Sukiennik is holder of 100% of the equity of the company. Corporate registration will not be renewed in Hong Kong.

Start-up Fundraising Activities

Inflo has been primarily supported at the expense of its sole founder. We have also received funding from the HKUST Alumni Endowment Fund at a total of 2500 HKD, as well as been offered 8000HKD in reimbursement from the HKUST Entrepreneurship Acceleration Fund.

Additionally, Inflo has actively applied and been rejected by the following programs/funds:

A. TSSSU (2018),
B. Cyberport Creative Micro-fund,
C. The STILE Initiative,

and has failed in qualifying as an entrant in the following competition: HKUST-Sino Million Dollar competition 2018.

Market Positioning:

The following is a comprehensive marketing positioning chart laying out the significant competitors in the field of social media and news aggregation.

Feature	f / Twitter	reddit	Medium	头条	The New York Times / WSJ / wp / FT	Flipboard	Google news / YAHOO! NEWS	inkl	inflo
Third-party content sharing	✓	✓							✓
Topic/Category Navigation			✓	✓	✓	✓	✓	✓	✓
Social network	✓	✓	✓			✓			✓
Premium content Subscription			✓		✓			✓	✓
Gamification System		✓							✓

Team & Roles

The following people were responsible in developing significant elements of the platform:

1. Nicholas Sukiennik: prototype, UI/UX.
2. Pranav Agrawal: training and deployment of the ANCCS classification algorithm.
3. Shiven Kapur: backend integration, user authentication system, migration to React framework.