Modeling Public Perceptions of Science in Media

Jiaxin Pei[1], Dustin Wright[2], Isabelle Augenstin[2], and David Jurgens[3]

[1]Stanford University
[2]University of Copenhagen
[3]University of Michigan

pedropei@stanford.edu, dw@di.ku.dk, augenstein@di.ku.dk,
jurgens@umich.edu

Abstract

Effectively engaging the public with science is vital for fostering trust and understanding in our scientific community. Yet, with an ever-growing volume of information, science communicators struggle to anticipate how audiences will perceive and interact with scientific news. In this paper, we introduce a computational framework that models public perception across twelve dimensions, such as newsworthiness, understandability, importance, and surprisingness. Using this framework, we create a large-scale science news perception dataset with 10,489 annotations from 2,101 participants who rated 1,506 science news stories. This dataset, sourced from diverse US and UK populations, provides valuable insights into public responses to scientific information across multiple domains. We further develop natural language processing models that predict public perception scores with a strong performance. Leveraging the dataset and model, we examine public perception of science from two perspectives: (1) Perception as an *outcome*: What factors affect the public perception of scientific information? (2) Perception as a *predictor*: Can we use the estimated perceptions to predict public engagement with science? We find that individuals' frequency of science news consumption is the strongest driver of perception, whereas demographic factors exert minimal influence. More importantly, through a large-scale analysis and carefully designed natural experiment on Reddit, we demonstrate that the estimated public perception of scientific information has direct connections with the final engagement pattern. Posts with more positive perception scores receive significantly more comments and upvotes, which is consistent across different scientific information and for the same science, but are framed differently. Overall, this research underscores the importance of nuanced perception modeling in science communication, offering new pathways to predict public interest and engagement with scientific content.

1   Introduction

Science communication plays a key role in disseminating scientific information and promoting public engagement with science [1]. Effectively communicating science to the public holds significance for both scientists and the broader audience outside the science community. On the one

hand, individuals living in modern society rely on scientific knowledge to make important life deci-sions [2], and in general, the public shows great interest in scientific information [3]. On the other hand, public engagement and discussions about scientific information could also help to promote societal understanding of science and inform new scientific developments [4].

Despite its crucial role in our society and the considerable efforts devoted to it, effectively communicating science to the public remains challenging and encounters numerous obstacles [5, 6]. One major challenge for science communicators is understanding the public [7]. Assumptions about the public represent an important latent dimension that distinguishes between the two classic models of science communication. In the deficit model, the public is simply viewed as a group of individuals that are passive and ignorant, and science communication is to simply "sell science" to the public [8]. Such a model has long been criticized because it overlooks the diversity of the audience [8]. The engagement model calls for science communicators to better contextualize the communication of science to promote science engagement with diverse groups of people [4]. However, in practice, understanding the public is challenging, even for professional journalists [9]. Additionally, existing research also suggests that many current science communication practices may only attract people who are already well-informed about certain science issues but fail to engage with a broader and more diverse set of audiences [10, 11]. Therefore, understanding the broad audience of science communication could provide further insights into facilitating effective science communication.

Existing studies have explored the interactions between science and diverse audiences from two primary perspectives. Firstly, researchers have investigated how various social factors influ-ence people's attitudes and attention toward scientific information [12, 13, 14]. These studies often employ surveys, where participants rate specific scientific issues such as gene technology [15] and climate change [16]. Another relevant area of research examines how different framing and lan-guage usage impact people's responses to scientific information [17]. However, within this line of research, while some studies have explored how pre-existing beliefs or political orientations shape individuals' consumption of scientific information [18, 19], there remains a gap in understanding the nuanced interactions between content and individual backgrounds. In short, we identify three major gaps in existing studies: (i) a tendency to narrowly focus on very specific topics (e.g. climate change), overlooking the nuanced variations across and within domains; (ii) the nuanced interac-tions between people's backgrounds and individual scientific information presented in media are underexplored; (iii) a lack of datasets and computational tools to understand public perception of scientific information in media at scale. In this study, we aim to help promote public engagement with science by addressing two questions: (1) What factors affect the public's perception of scien-tific information in the media? (2) Do scientific information with more positive perceptions receive more public engagement?

We introduce a computational pipeline designed to assess the public perception of scientific information in the media. Firstly, we propose a novel comprehensive framework comprising 12 di-mensions: NEWSWORTHINESS, UNDERSTANDABILITY, EXPERTISE, INTERESTINGNESS, FUN, IMPORTANCE, BENEFIT, SHARING, READING WILLINGNESS, EXAGGERATION, SURPRISING-NESS, and CONTROVERSY. Following this framework, we curate a large-scale science news per-ception dataset containing 10,489 annotations from 2,101 participants across 1,506 science news stories. Our dataset draws from representative samples of the US and UK populations, ensuring

diversity and representativeness in measuring public perceptions. Each participant evaluated a variety of science news stories based on our proposed framework. Finally, we develop NLP models to automatically predict public perception of science news.

This computational framework allows us to systematically examine the public perception of science from two perspectives: (1) Perception as an outcome: What factors affect people's perception of scientific information in media? (2) Perception as a predictor: Can estimated perceptions predict public engagement with scientific information? We find that individuals' science news consumption frequency is the strongest factor affecting their perceptions of science news stories, while demographic factors only have a minimal effect. Furthermore, through a large-scale analysis and a natural experiment on Reddit, we examine the connection between the estimated perceptions and the engagement patterns with scientific information. We find that the potential perceptions of a post correlate with public engagement with it. For example, posts with higher estimated SURPRIS-INGNESS, IMPORTANCE, and FUN tend to receive more upvotes and comments on Reddit. Such a pattern persists even when we compare different framings of the same scientific information, suggesting a strong causal relationship between the perception and engagement. Overall, our analysis highlights the significance of modeling public perception of scientific information in the media. It not only helps to better understand the public interest in science but also holds the potential to predict engagement patterns with scientific information.

2   A Multi-dimensional Framework for Modeling the Public Perception of Scientific Information in Media

Modeling public perception of scientific information is a key step towards effective science communication [20]. However, public perception is a subjective and multi-dimensional construct. Here, drawing on existing literature on news values, public understanding of science, and science communication, we develop a comprehensive framework to analyze the public perception of scientific information in the media.

NEWSWORTHINESS In science journalism practices, newsworthiness is regarded as a core dimension that guides the selection and coverage of information [21]. Similarly, scholars have studied the news values of scientific information in news stories [22] or research articles [23]. While many existing studies focus on modeling journalists' understanding of newsworthiness, some also emphasize the importance of studying perceived newsworthiness in various settings [24]. Here, we approach the newsworthiness of scientific information from the perspective of the general public. Specifically, our goal is to measure whether the public views a news article as "should be published in news or not". NEWSWORTHINESS represents the holistic evaluation of the perceived value of the presented information.

UNDERSTANDABILITY In science communication, presenting content that is accessible and comprehensible to the audience is crucial, as the value of scientific information in news is contingent on its understandability [25]. Without clear comprehension, the lay public is unlikely to benefit from the scientific news presented and further participate in the public discussion about it [26]. However, despite existing efforts to make science more accessible to the public, existing studies

suggest that understandability remains a challenge in current science news reporting [27]. Therefore, it is important for us to measure how the public perceives the understandability of science presented in the media.

EXPERTISE Scientific information usually contains jargon or specialized knowledge that the general public may not easily understand [28, 29]. Here, we aim to quantify whether a given science news story requires specialized knowledge for the public to understand. UNDERSTANDABILITY models an individuals' capability to understand the content, which could also depend on the individual's background in the relevant topic. Unlike understandability, EXPERTISE focuses on the relatively more objective "difficulty" of the content.

IMPORTANCE While scientific findings may be highly valued within specific scientific communities, their significance to the broader society can vary. Thus, assessing whether the science covered is of sufficient importance to the general audience is crucial in understanding the overall perception of science news [30]. Importance has been considered as a key dimension in evaluating science news values [22] and here we focus on the public perception of scientific importance. By doing so, we aim to gain insights into how the public perceives the significance of research outputs presented in news media.

FUN Entertainment is recognized as an essential news factor in journalism [21]. Despite science often being seen as a serious topic, making scientific information engaging is essential for science communicators to foster public engagement with science [1, 4]. Therefore, assessing the entertainment value or the fun aspect of reading science news stories is important in evaluating the overall perceptions.

SURPRINGNESS Surprisingness, characterized by content that contrasts with people's existing knowledge or perceptions, is widely recognized as a core dimension of news value [31]. Here, we aim to measure how the public perceives the surprisingness of scientific information presented in media. An existing study suggests that findings that surprise the audience may be deemed to have a higher news value [32], as they not only inform but also engage by introducing novel perspectives or unexpected information.

CONTROVERSY Scientific information in the media can be controversial due to the natural uncertainty of scientific research and the significance of the topic it tackles [33, 34]. Furthermore, existing studies also suggest that controversies discussed in the media can also potentially affect scientific publications [35]. Therefore, it is important to measure how the public perceives the scientific information presented in news media to be controversial.

EXAGGERATION Science reporting has long grappled with the issue of exaggeration or hype [36]. While exaggeration may initially attract more audience attention, existing studies suggest that perceived exaggeration could erode public trust in science [37, 38]. Hence, measuring how the public perceives exaggerations in science news stories is crucial for a comprehensive evaluation of scientific information in the media.

INTERESTINGNESS INTERESTINGNESS represents a subjective but essential dimension in assessing science news, tailored to the preferences of specific audiences. In journalism research, perceived interestingness is considered a core factor affecting people's perceptions of overall newsworthiness [24]. One of the primary goals of science communication is to engage the public with

information that captures their interest; therefore, it is essential for us to measure the perceived interestingness of scientific information presented in news articles.

BENEFIT Scientific knowledge plays a key role in informing individuals' decision-making and facilitating societal transformations [39]. Consequently, the direct applicability or usability of science news stories becomes a critical dimension of their value [22]. This dimension underscores the importance of presenting scientific findings in a way that is not only informative but also practically relevant to the audience's daily lives and decisions.

SHARING WILLINGNESS Social media platforms have gradually become one of the primary sources of news information for the general public [40]. Therefore, studying the sharability of news stories and users' sharing behaviors has attracted attention from journalism scholars [21, 41]. In this context, we consider the likelihood of readers sharing scientific information as an important dimension to measure the overall perception of science news, reflecting the impact and reach of science communication in the digital age. We use SHARING to represent this category in the rest of the paper.

READING WILLINGNESS Reading represents one of the most basic types of engagement with scientific information. READING WILLINGNESS measures whether the public would potentially be willing to read about the content in different venues. Good science news should attract the audience and prompt them to read further about the topic. Articles that the audience does not even want to read are less likely to facilitate further discussions and engagements with the corresponding topic. Therefore, we consider reading willingness as an important dimension in our framework. We use READING to represent this category in the rest of the paper.

Collectively, the 12 dimensions provide a comprehensive evaluation of public perceptions of scientific information and can be used to measure science presented in diverse types of media, including news stories, social media posts, and research articles. Appendix A details the descriptions of each dimension and the statements used to measure them.

3   Results

How the public perceives scientific information has long been a core question for the science community [42]. In this study, we conduct a series of experiments to investigate how factors like demographics and trust in science affect public perception of scientific information and the connection between perception and downstream engagement on social media. More specifically, we focus on two research questions: **RQ1 Perception as an outcome:** What factor has the largest influence on people's perception of science news? **RQ2 Perception as a predictor:** Are there any connections between the estimated perceptions and the final public engagement with scientific information?
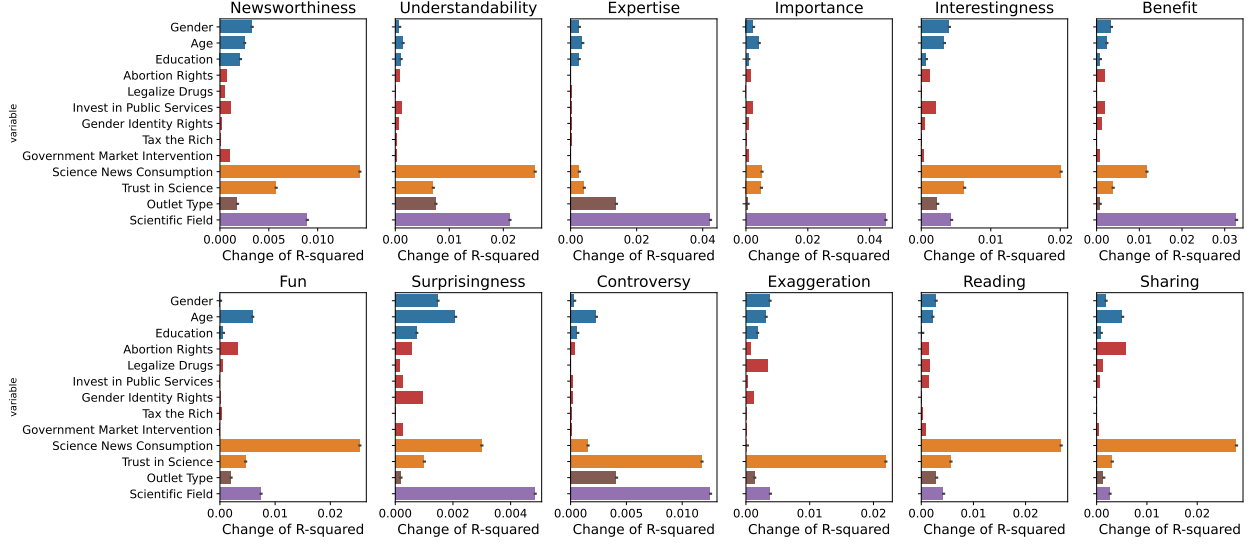
Figure 1: Individual's science news consumption frequency, trust in science, and the domain of scientific information are the strongest predictors of people's perception of scientific information in the media. Surprisingly, the widely studied demographic factors and political orientation factors have minimal effects.

3.1 Analyzing Factors Affecting the Public Perceptions of Science

Public attitudes toward science are subjective and can be affected by individuals' backgrounds such as gender [12] and education [43]. However, existing studies mainly focus on people's broad attitudes toward science, while science comes with large differences across fields and communities [39]. Therefore, it is important to study how different background factors affect people's perceptions of scientific information presented in media. In this study, we examine the connection between background factors and public perceptions of scientific information in news articles using the nuanced background information as well as individual ratings for science news stories. To understand how background factors affect people's perceptions of individual science news stories, we construct separate mixed-effect regression models leveraging the background factors to predict each dimension of public perceptions. The background factors include three major dimensions: (1) demographic information, including gender, age, and educational level; (2) experience with science, including science news consumption frequency and trust in science; (3) political orientations, including the attitudes towards women's reproductive health rights, legalizing drugs, government investments in public services, gender identity rights, taxing the rich and government interfering with the market. To account for the effect of article-level variations, we control the article ID as the random effect. Additionally, we also control the scientific field and the outlet type as fixed effects. Figure 2 presents the aggregated results from the regressions.

Science News Consumption Frequency has the largest effect on people's perception of scientific information As shown in Figure 2, science news consumption frequency is the driver of individuals' rating for scientific information. For example, people who consume science news on a daily basis tend to rate the same scientific information 0.47 higher than people who never consume science news stories on a 5-point scale. The effect size becomes even larger for FUN and SHARING,

6

	Newsworthiness	Understandability	Expertise	Importance	Interestingness	Benefit	Fun	Surprisingness	Controversy	Exaggeration	Reading	Sharing
Gender: Non-binary	-0.15	0.18	-0.38*	0.09	0.11	0.09	0.05	-0.34*	-0.00	-0.15	0.02	-0.12
Gender: Woman	0.12***	0.05	-0.11***	0.12***	0.13***	0.10***	0.00	0.04	0.04	-0.12***	0.09***	0.09**
Age: 25-29	-0.04	-0.10	0.01	-0.04	-0.00	-0.01	-0.02	0.04	-0.05	0.00	-0.00	-0.04
Age: 30-34	0.01	0.00	-0.16	-0.07	-0.06	-0.07	-0.15	-0.07	-0.02	-0.07	-0.01	-0.13
Age: 35-39	0.04	-0.06	-0.05	0.03	-0.01	-0.02	-0.14	-0.06	-0.05	-0.09	0.02	-0.10
Age: 40-44	0.05	-0.06	-0.10	0.00	-0.03	-0.05	-0.15	-0.05	-0.02	-0.05	-0.01	-0.15
Age: 45-49	0.07	-0.08	-0.13	0.07	0.03	-0.02	-0.25***	-0.12	-0.03	-0.12	-0.03	-0.03
Age: 50-54	0.14	0.00	-0.20**	0.09	0.12	0.08	-0.11	-0.05	0.03	-0.18**	0.10	0.01
Age: 54-59	0.01	-0.07	-0.20*	0.09	0.03	-0.00	-0.30***	-0.14	0.04	-0.14*	-0.05	-0.27***
Age: 60-64	0.14	-0.04	-0.18*	0.18*	0.12	0.01	-0.19*	-0.05	0.10	-0.06	0.02	-0.08
Age: >65	0.07	-0.08	-0.16	0.14	0.07	-0.03	-0.25***	-0.10	0.11	-0.07	-0.03	-0.09
Education: Graduate degree (e.g. MSc, MA, PhD)	-0.12***	-0.02	0.08	-0.08	-0.07	-0.05	0.04	-0.02	-0.04	0.03	-0.02	-0.05
Education: High school diploma or equivalent	0.01	-0.07	-0.09	-0.03	-0.01	-0.01	-0.02	0.02	-0.04	-0.06	0.01	0.02
Education: Less than a high school diploma	0.03	0.03	0.12	0.00	-0.04	-0.05	0.12	0.07	0.18	0.13	0.04	0.02
Attitude: Abortion rights	-0.04	-0.03	-0.01	-0.05**	-0.04**	-0.05***	-0.08***	-0.03	-0.03	0.03	-0.04**	-0.10***
Attitude: Legalize drugs	0.02	0.01	-0.02	0.00	-0.00	0.01	0.02	-0.01	-0.01	-0.05***	0.03**	0.03**
Attitude: Public services	0.06*	0.05*	-0.03	0.08***	0.07***	0.06***	0.01	0.03	0.03	-0.03	0.05**	0.04
Attitude: Gender identity rights	0.02	0.03	-0.03	0.03	0.02	0.03*	0.01	-0.04	-0.02	-0.04*	0.00	-0.01
Attitude: Tax the rich	0.01	-0.02	0.03	-0.02	0.01	0.00	0.03	-0.00	0.01	-0.01	0.02	0.01
Attitude: Government market intervention	0.04*	-0.02	0.02	0.04	0.02	0.03	0.00	0.02	0.01	-0.01	0.03	0.03
Science news consumption: Weekly	-0.06	-0.15***	-0.08	-0.06	-0.08	-0.04	-0.20***	-0.02	0.05	-0.02	-0.06	-0.21***
Science news consumption: Monthly	-0.21***	-0.28***	-0.03	-0.15**	-0.20***	-0.12***	-0.37***	0.01	-0.05	-0.02	-0.22***	-0.38***
Science news consumption: Yearly	-0.33***	-0.48***	0.08	-0.20**	-0.32***	-0.21***	-0.55***	-0.05	-0.06	-0.02	-0.40***	-0.52***
Science news consumption: Never	-0.47***	-0.63***	0.10	-0.33***	-0.53***	-0.33***	-0.74***	-0.21**	-0.05	0.01	-0.47***	-0.75***
Trust in science: High trust	-0.20***	-0.21***	0.14	-0.11	-0.11*	-0.08	-0.11	-0.04	0.23***	0.26***	-0.07	-0.07
Trust in science: Moderate trust	-0.24***	-0.32***	0.28***	-0.20***	-0.18***	-0.13***	-0.18***	-0.01	0.42***	0.45***	-0.13***	-0.08
Trust in science: Some trust	-0.30***	-0.34***	0.29***	-0.30***	-0.22***	-0.16**	-0.18	-0.02	0.53***	0.61***	-0.20***	-0.10
Trust in science: Little trust	-0.55***	-0.44***	0.24	-0.46***	-0.53***	-0.32***	-0.57***	-0.18	0.51***	0.70***	-0.42***	-0.46***
Trust in science: No trust	-0.56**	-0.36	0.29	-0.55*	-0.50*	-0.38*	-0.44	-0.32	0.13	0.47*	-0.43**	-0.08
Outlet type: PressRelease	-0.08*	-0.20***	0.34***	0.01	-0.10***	-0.01	-0.12***	0.02	-0.16***	-0.08	-0.10***	-0.09**
Outlet type: SciTech	-0.11**	-0.15***	0.24***	-0.05	-0.10**	-0.05	-0.06	-0.01	-0.16***	-0.07	-0.09***	-0.07
Scientific field: Humanities	-0.45***	-0.02	-0.17	-1.18***	-0.18	-0.78***	0.31**	0.01	-0.15	0.10	-0.17	-0.28**
Scientific field: Engineering	-0.19***	-0.21***	0.25***	-0.38***	-0.10*	-0.11***	0.15***	0.10	-0.16***	0.04	-0.09*	-0.09
Scientific field: Science	-0.22***	-0.19***	0.27***	-0.49***	-0.10***	-0.26***	0.20***	0.01	-0.28***	-0.06	-0.13***	-0.11**
Scientific field: Social Science	-0.18***	0.14***	-0.35***	-0.49***	-0.01	-0.25***	0.23***	-0.14***	0.02	0.09*	-0.05	-0.07

Figure 2: Experience with science and individuals' backgrounds could both affect people's perception of scientific information. People's science news consumption frequency and trust in science have the largest effect sizes.

suggesting that people's science news consumption frequency is significantly associated with their tendency to enjoy and share science. Similarly, trust in science also has a large effect on people's perceptions of scientific information regarding newsworthiness and interestingness. Additionally, trust in science significantly affects individuals' ratings for controversy and exaggeration, while science news consumption does not. This result suggests that prior trust in science could potentially affect individuals' trust in scientific information. Overall, our results suggest that individuals' science news consumption frequency and trust in science play a significant role in shaping people's perceptions of scientific information compared with other background factors.

Gender, age, education, and political attitudes have relatively small effects on public perception of science Existing studies have widely studied how demographic factors affect people's perception of science and scientific information [12, 44, 45, 46]. Here, we present the most comprehensive analysis on demographic factors, with the control of many other factors, including content and media type. Surprisingly, we found that age, education level and most of the political
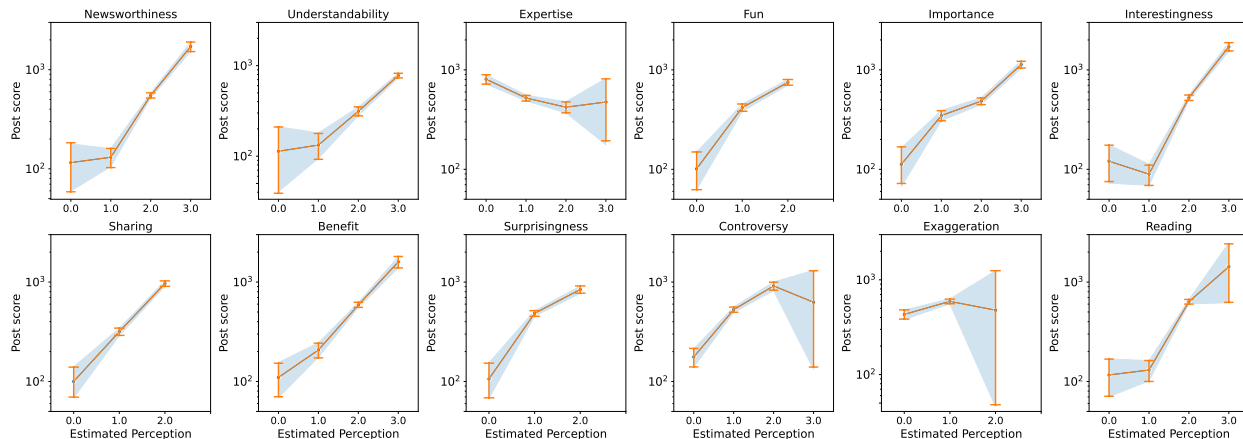
Figure 3: Posts with more positive perception values receive higher scores on Reddit.

orientation factors do not significantly affect individual ratings for science news stories on most of the dimensions. Furthermore, while some existing studies suggest that men hold more positive attitudes towards science [12, 45], our results suggest that women may actually show greater interest in scientific information presented in media. Moreover, while existing studies generally suggest a positive correlation between people's education level and their attitudes towards science [44], we observe that people with a graduate degree may actually consider the same scientific information with lower newsworthiness compared with individuals with a college degree. One potential explanation is that people with a higher degree may have a more stringent perception of news values due to their graduate-level training.

Science domain and media type strongly affect public perception of science In comparison to general news outlets, science news stories presented in Press Releases and SciTech blogs are consistently perceived as less newsworthy. This may be attributed to the more technical nature of the content covered by these outlets, which may be less understandable to the general public. Larger differences are observed in the scientific domain of the content. Although the Humanities are perceived as more fun and understandable, they are considered significantly less important and less newsworthy than other subjects. In contrast, science news stories about medicine are perceived as more important and thus considered more newsworthy compared to other subjects. The effect size of domains is significantly larger than that of other demographic information, indicating that content plays a more substantial role in influencing how the public perceives scientific information.

3.2 Predicting and Promoting Public Engagement with Science

One of the fundamental goals of science communication is to promote public engagement with scientific information [4]. Social media platforms offer unique opportunities for the public to engage with scientific information. However, predicting public coverage and engagement with science has been notoriously hard [47] because many types of information are competing for the public's attention in the current media environment [6]. Here, we study the connection between estimated perceptions and public engagement with scientific information on Reddit, a popular
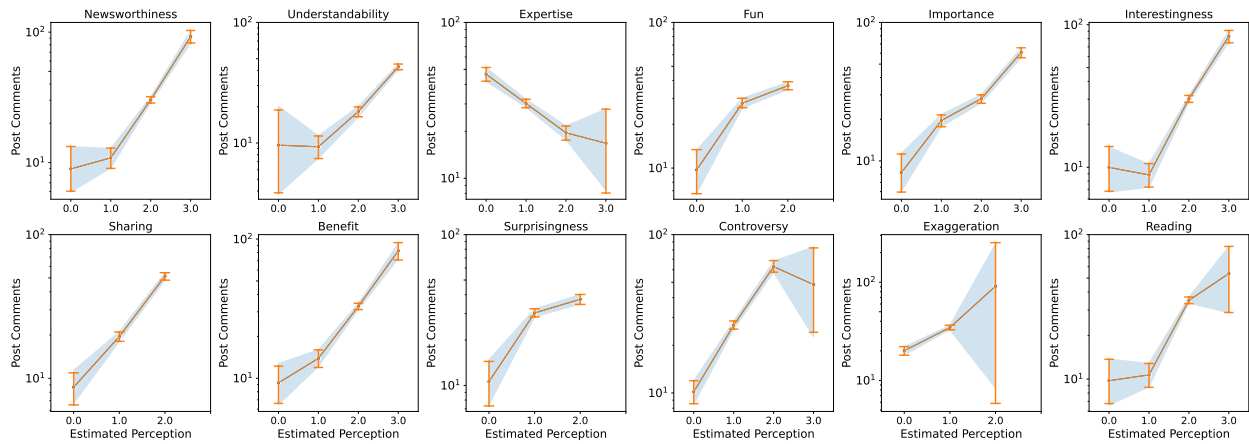
Figure 4: Posts with more positive perception values receive more comments on Reddit.
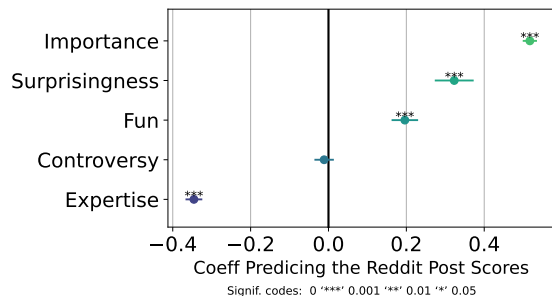
Figure 5: When sharing the same science news story, posts that are framed to be more important, surprising, and fun receive higher scores on Reddit. Posts that require more specialized knowledge to understand receive lower scores.
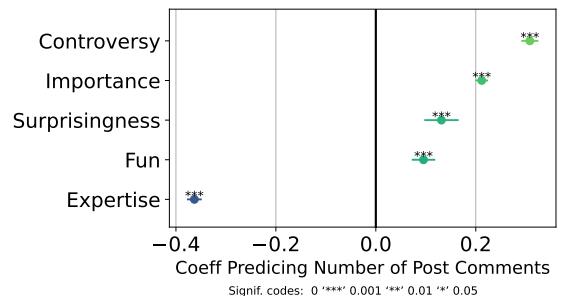

Figure 6: When sharing the same science news story, posts with higher estimated importance, surprisingness, fun, and controversy receive more comments on Reddit. Posts that require more specialized knowledge to understand receive fewer comments.

social media platform that contains many active science-sharing communities.

To study the connections between estimated perceptions and public engagement with social media posts about science, we construct a dataset with 95,465 Reddit posts mentioning 29,099 science-related content r/science subreddit. We conduct two analyses to understand the connection between the perceived values and engagements: (1) a descriptive analysis showing the overall trends across different sciences and (2) a natural experiment where the same science is framed differently in different posts.

Posts with more positive perception scores receive more scores and comments Figure 3 and Figure 4 shows the overall trend of estimated perceptions and public engagement patterns on Reddit. In general, public engagement with science-related posts is positively correlated with the estimated perceptions on most of the dimensions. For example, posts with higher estimated newsworthiness are significantly more likely to receive more upvotes and comments. More specifically, one unit increase of perceived importance could lead to a 68% increase of post scores. On the

9

contrary, posts that require specialized knowledge to understand receive lower scores and fewer comments. Such a result suggests that the estimated perception values could be used to predict public engagement with science-related posts on social media.

When describing the same science, posts with more positive perceptions attract more comments and upvotes Figure 5 and Figure 6 show the results of the natural experiment. We find that posts with higher IMPORTANCE, SURPRISINGNESS and FUN are more likely to receive higher scores and more comments, reflecting the public's preferences for the presentations of the scientific information. On the contrary, posts that require specialized knowledge to understand receive lower scores and fewer comments, suggesting that science jargon or specialized topics can reduce public engagement with science. For controversy, we observe different effects for comments and scores: posts with higher controversy values receive significantly more comments, but not scores. Such a result suggests that controversial content attracts more engagement through comments. While the number of comments is usually highly correlated with the post scores, in this setting, it is possible that controversial posts also receive more downvotes, which removes the effect of increased engagement. Our result also relates to the literature on framing [48, 49]. Existing research has found that framing the same issue in different ways could lead to different communication outcomes and audience perceptions [50]. While we do not study the linguistic features that contribute to different potential perceptions, our model could allow future researchers or journalists to directly estimate the potential public perceptions of the same information, but with different frames.

4   Discussions

Effectively engaging the public with science is important but challenging, partly because of the gap between the audience and the science communicators – science communicators may not be able to fully grasp what the audience is interested in and how they might perceive the scientific information they are receiving. To address this problem, in this study, we design a computational framework that helps to understand and predict public perception of scientific information in the media. Through a series of analyses, we find that the science news consumption frequency is the key factor that drives people's perception of scientific information: people who consume scientific information frequently are more likely to rate science more positively. Such a result reflects a key challenge in science communication that only people who trust and consume science are more likely to positively react to scientific information, but attracting people who do not frequently consume science is challenging. However, our follow-up analysis on Reddit shows a potential hope: even for the same science, framing it in a different way could change the final engagement pattern. Therefore, our computational tool for perception estimation provides science communicators a valuable approach to "preview" the potential public perception when drafting the scientific information, helping them craft messages that better engage diverse audiences.

5 Methods

To enable large-scale analysis of public perceptions of science, we develop the following computational pipeline combining large-scale crowdsourcing and NLP modeling.

Raw dataset Our raw dataset is obtained from `altmetric.com`, which contains 2.4 million media mentions of research papers. We follow the data processing pipeline from existing research [51, 52] and further extract a subsample of 128,942 news/article pairs where we are able to collect the full text of the news articles. Our data spans 273 different news outlets and 57,807 different scientific articles in over 100 scientific domains. Appendix B details more information about the dataset.

Crowdsourcing annotation We recruit participants from `Prolific.com` as it allows us to collect annotations from representative samples for the US and UK population. More specifically, we recruited 1,057 participants from the US and 1,044 participants from the UK. Both are drawn from a representative sample balanced for age, sex, and race. We require each participant to be fluent in English to participate in our study, as the science news articles are written in English. For each news article, we set out to collect 4 labels from the US and the UK samples separately. We use POTATO to set up all the annotation interface [53]. Appendix B presents all the details of our annotation experiment.

NLP modeling To allow automatic modeling of perceived news values for each news story, we fine-tune a RoBERTa-Large model [54] on the annotated dataset. We split the annotated data into train, validation, and test sets following the ratio of 7:1:2. We follow a multi-task regression setting and on the output layer, the model predicts a score ranging from 1-5 for each statement so that for each science news article, we could directly obtain the predictions for all the dimensions. Appendix C details the modeling process and results.

Reddit science posting dataset We first collect all the posts from 2017 to 2021 shared in r/science, a subreddit where users actively share science news stories. We then extract all the shared links in these posts and further extract all the posts on Reddit mentioning the links. We further removed all the duplicated posts and URLs that are only shared once. The final dataset contains 95,465 Reddit posts mentioning 29,099 science-related content.

Regression We construct mixed effect models using the estimated perceptions to predict two engagement metrics: the number of comments and the post scores, both log-transformed. Post scores are calculated by subtracting the number of downvotes from the total upvotes. Additionally, comments about scientific information represent an important engagement with scientific information as they mark open discussions about science. We control the URL as a random effect and also control the subreddit, the domain of the URL, and whether a URL was shared in a subreddit for the first time as a fixed effect. Furthermore, given that public perception correlates with each other, we followed the step-wise regression procedure to gradually remove the variables with the highest variance inflation factor score. The final regression contains five perception dimensions (i.e. IMPORTANCE, SURPRISINGNESS, FUN, CONTROVERSY and EXPERTISE) and other control variables.

References

[1] Baruch Fischhoff. The sciences of science communication. *Proceedings of the National Academy of Sciences*, 110(supplement_3):14033–14039, 2013.

[2] Randy L Bell and Norman G Lederman. Understandings of the nature of science and decision making on science and technology based issues. *Science education*, 87(3):352–377, 2003.

[3] John R Durant, Geoffrey A Evans, and Geoffrey P Thomas. The public understanding of science. *Nature*, 340(6228):11–14, 1989.

[4] Jack Stilgoe, Simon J Lock, and James Wilsdon. Why should we promote public engagement with science? *Public understanding of science*, 23(1):4–15, 2014.

[5] Donghong Cheng, Michel Claessens, Toss Gascoigne, Jenni Metcalfe, Bernard Schiele, and Shunke Shi. Communicating science in social contexts. *New Models*, 2008.

[6] Dietram A Scheufele. Communicating science in social settings. *Proceedings of the National Academy of Sciences*, 110(supplement_3):14040–14047, 2013.

[7] Mike S Schäfer, Tobias Füchslin, Julia Metag, Silje Kristiansen, and Adrian Rauchfleisch. The different audiences of science communication: A segmentation analysis of the swiss population's perceptions of science and their information and media use patterns. *Public understanding of science*, 27(7):836–856, 2018.

[8] Patrick Sturgis and Nick Allum. Science in society: Re-evaluating the deficit model of public attitudes. *Public understanding of science*, 13(1):55–74, 2004.

[9] Charles K Atkin, Judee K Burgoon, and Michael Burgoon. How journalists perceive the reading audience. *Newspaper research journal*, 4(2):51–63, 1983.

[10] Kirby Goidel and Matthew Nisbet. Exploring the roots of public participation in the controversy over embryonic stem cell research and cloning. *Political Behavior*, 28:175–192, 2006.

[11] Matthew C Nisbet and Dietram A Scheufele. What's next for science communication? promising directions and lingering distractions. *American journal of botany*, 96(10):1767–1778, 2009.

[12] Molly Weinburgh. Gender differences in student attitudes toward science: A meta-analysis of the literature from 1970 to 1991. *Journal of Research in science Teaching*, 32(4):387–398, 1995.

[13] Peter Achterberg, Willem De Koster, and Jeroen Van der Waal. A science confidence gap: Education, trust in scientific methods, and trust in scientific institutions in the united states, 2014. *Public Understanding of Science*, 26(6):704–720, 2017.

[14] Aaron M McCright, Katherine Dentzman, Meghan Charters, and Thomas Dietz. The influence of political ideology on trust in science. *Environmental Research Letters*, 8(4):044029, 2013.

[15] Anna Olofsson, Susanna Öhman, and Saman Rashid. Attitudes to gene technology: the significance of trust in institutions. *European Societies*, 8(4):601–624, 2006.

[16] Jing Shi, Vivianne HM Visschers, and Michael Siegrist. Public perception of climate change: The importance of knowledge and cultural worldviews. *Risk Analysis*, 35(12):2183–2201, 2015.

[17] Lauren Feldman and P Sol Hart. Upping the ante? the effects of "emergency" and "crisis" framing in climate change news. *Climatic Change*, 169(1-2):10, 2021.

[18] Sara K Yeo, Michael A Xenos, Dominique Brossard, and Dietram A Scheufele. Selecting our own science: How communication contexts and individual traits shape information seeking. *The ANNALS of the American Academy of Political and Social Science*, 658(1):172–191, 2015.

[19] P Sol Hart and Erik C Nisbet. Boomerang effects in science communication: How motivated reasoning and identity cues amplify opinion polarization about climate mitigation policies. *Communication research*, 39(6):701–723, 2012.

[20] Nancy Longnecker. Good science communication considers the audience. In *Teaching science students to communicate: a practical guide*, pages 21–30. Springer, 2023.

[21] Tony Harcup and Deirdre O'neill. What is news? news values revisited (again). *Journalism studies*, 18(12):1470–1488, 2017.

[22] Franziska Badenschier and Holger Wormer. Issue selection in science journalism: Towards a special theory of news values for science news? In *The sciences' media connection–public communication and its repercussions*, pages 59–85. Springer, 2011.

[23] Sachita Nishal and Nicholas Diakopoulos. From crowd ratings to predictive models of newsworthiness to support science journalism. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2):1–28, 2022.

[24] Angela M Lee and Hsiang Iris Chyi. When newsworthy is not noteworthy: Examining the value of news from the audience's perspective. *Journalism studies*, 15(6):807–820, 2014.

[25] Baruch Fischhoff. Communicating uncertainty fulfilling the duty to inform. *Issues in Science and Technology*, 28(4):63–70, 2012.

[26] Johanna Varner. Scientific outreach: toward effective public engagement with biological science. *BioScience*, 64(4):333–340, 2014.

[27] José J Morosoli, Lucía Colodro-Conde, Fiona Kate Barlow, and Sarah E Medland. Scientific clickbait: Examining media coverage and readability in genome-wide association research. *Plos one*, 19(1):e0296323, 2024.

[28] Russel Hirst. Scientific jargon, good and bad. *Journal of technical writing and communication*, 33(3):201–229, 2003.

[29] Aviv J Sharon and Ayelet Baram-Tsabari. Measuring mumbo jumbo: A preliminary quantification of the use of jargon in science communication. *Public Understanding of Science*, 23(5):528–546, 2014.

[30] Risa B Burns, Mark A Moskowitz, Michael A Osband, and Lewis E Kazis. Newspaper reporting of the medical literature. *Journal of general internal medicine*, 10:19–24, 1995.

[31] Johan Galtung and Mari Holmboe Ruge. The structure of foreign news: The presentation of the congo, cuba and cyprus crises in four norwegian newspapers. *Journal of peace research*, 2(1):64–90, 1965.

[32] Tint Hla Hla Htoo, Na Jin-Cheon, and Michael Thelwall. Why are medical research articles tweeted? the news value perspective. *Scientometrics*, 128(1):207–226, 2023.

[33] Daniel Sarewitz. How science makes environmental controversies worse. *Environmental science & policy*, 7(5):385–403, 2004.

[34] Sharon M Friedman, Sharon Dunwoody, and Carol L Rogers. *Communicating uncertainty: Media coverage of new and controversial science*. Routledge, 1999.

[35] Craig O Stewart. How a media controversy can influence a scientific publication. *Discourse Approaches to Politics, Society and Culture (DAPSAC)*, 2008.

[36] Timothy Caulfield and Celeste Condit. Science and the sources of hype. *Public Health Genomics*, 15(3-4):209–217, 2012.

[37] Kristen Intemann. Understanding the problem of "hype": Exaggeration, values, and trust in science. *Canadian Journal of Philosophy*, 52(3):279–294, 2022.

[38] Ariadne Neureiter, Marlis Stubenvoll, Ruta Kaskeleviciute, and Jörg Matthes. Trust in science, perceived media exaggeration about covid-19, and social distancing behavior. *Frontiers in Public Health*, 9:670485, 2021.

[39] Yian Yin, Yuxiao Dong, Kuansan Wang, Dashun Wang, and Benjamin F Jones. Public use and public funding of science. *Nature human behaviour*, 6(10):1344–1350, 2022.

[40] Elisa Shearer and Elizabeth Grieco. Americans are wary of the role social media sites play in delivering the news. *Pew Research Center*, 2, 2019.

[41] Monika Bednarek. Investigating evaluation and news values in news items that are shared through social media. *Corpora*, 11(2):227–257, 2016.

[42] National Academies of Sciences, Engineering, and Medicine. Communicating science effectively: A research agenda. 2017.

[43] Hee-Je Bak. Education and public attitudes toward science: Implications for the "deficit model" of education and support for science and technology. *Social Science Quarterly*, 82(4):779–795, 2001.

[44] Jonathan Osborne, Shirley Simon, and Sue Collins. Attitudes towards science: A review of the literature and its implications. *International journal of science education*, 25(9):1049–1079, 2003.

[45] M Gail Jones, Ann Howe, and Melissa J Rua. Gender differences in students' experiences, interests, and attitudes toward science and scientists. *Science education*, 84(2):180–192, 2000.

[46] Joshua M Blank and Daron Shaw. Does partisanship shape attitudes toward science and public policy? the case for ideology and religion. *The ANNALS of the American Academy of Political and Social Science*, 658(1):18–35, 2015.

[47] Ansel MacLaughlin, John Wihbey, and David Smith. Predicting news coverage of scientific articles. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 12, 2018.

[48] Zhongdang Pan and Gerald M Kosicki. Framing analysis: An approach to news discourse. *Political communication*, 10(1):55–75, 1993.

[49] Dietram A Scheufele. Framing as a theory of media effects. *Journal of communication*, 49(1):103–122, 1999.

[50] Paul D'Angelo and Matthew Lombard. The power of the press: The effects of press frames in political campaign news on media perceptions. *Atlantic Journal of Communication*, 16(1):1–32, 2008.

[51] Dustin Wright, Jiaxin Pei, David Jurgens, and Isabelle Augenstein. Modeling information change in science communication with semantically matched paraphrases. *arXiv preprint arXiv:2210.13001*, 2022.

[52] Jiaxin Pei and David Jurgens. Measuring sentence-level and aspect-level (un) certainty in science communications. *arXiv preprint arXiv:2109.14776*, 2021.

[53] Jiaxin Pei, Aparna Ananthasubramaniam, Xingyao Wang, Naitian Zhou, Apostolos Dedeloudis, Jackson Sargent, and David Jurgens. Potato: The portable text annotation tool. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 2022.

[54] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

[55] Helen Caple. News values and newsworthiness. In *Oxford research encyclopedia of communication*. 2018.

[56] Linden Dalecki, Dominic L Lasorsa, and Seth C Lewis. The news readability problem. *Journalism Practice*, 3(1):1–12, 2009.

[57] Tara Agarwal. *An analysis of the readability of science and technology reporting in four national newspapers*. PhD thesis, Loughborough University of Technology, 1992.

[58] Michael Bourk, Bienvenido León, and Lloyd S Davis. Entertainment in science: Useful in small doses. In *Communicating science and technology through online video*, pages 90–106. Routledge, 2018.

[59] Marehalli G Narasimhan. Controversy in science. *Journal of biosciences*, 26(3):299–304, 2001.

[60] Andrew F Hayes and Klaus Krippendorff. Answering the call for a standard reliability measure for coding data. *Communication methods and measures*, 1(1):77–89, 2007.

[61] Mohsen Tavakol and Reg Dennick. Making sense of cronbach's alpha. *International journal of medical education*, 2:53, 2011.

[62] Regina Dittrich, Brian Francis, Reinhold Hatzinger, and Walter Katzenbeisser. A paired comparison approach for the analysis of sets of likert-scale responses. *Statistical Modelling*, 7(1):3–28, 2007.

[63] Dennis Reidsma and Jean Carletta. Reliability measurement without limits. *Computational Linguistics*, 34(3):319–326, 2008.

[64] Ron Artstein and Massimo Poesio. Inter-coder agreement for computational linguistics. *Computational linguistics*, 34(4):555–596, 2008.

A Twelve-dimension framework for public perception of science

Table 1 shows the statements and descriptions for the twelve-dimensional framework for measuring public perception of science.

B Building a New Dataset to Model Public Perception of Scientific Information

Data resources are key to scientific research. To support long-term research on public perception of scientific information and promote public engagement with science, we create a new large-scale dataset containing nuanced ratings for individual science news stories as well as detailed background information of the participants. In this section, I describe the creation process of this dataset.

Category	Statement	Description
NEWSWORTHINESS	The science news story should be published in the news	In journalism, newsworthiness is considered as the core dimension of news stories and is usually operationalized as "worthy of being published as news" [55]. Here we measure the perceived newsworthiness from the perspectives of the public.
UNDERSTANDABILITY	I can understand the science news story	Understandability measures whether a given science news story can be understood by the public. Existing studies consider understandability as an important dimension for news stories [56, 57].
EXPERTISE	Understanding the science news story requires specialized knowledge	Scientific information may involve specialized knowledge and the EXPERTISE dimension measures whether a given news story requires specialized knowledge to understand.
IMPORTANCE	The science news story tackles an important issue	Importance has been considered as one of the core value factors of science news stories [22]. Here we measure how the public perceives the importance of the scientific information presented in a news article.
FUN	The science news story is fun to read	Entertainment has long been considered as one of the core news values in journalism [21] and the public also actively perceives the entertainment value of science [58]. Here we measure whether a given scientific information is perceived as fun and entertaining to the public.
SURPRISINGNESS	The scientific finding seems surprising to me	Science sometimes provides surprising results that contradict people's existing beliefs and surprisingness has been considered as an important factor for science news values [22]. Here we measure whether the public perceives a given science news story as surprising.
CONTROVERSY	The scientific finding could be controversial	Science tackles many controversial topics [59] and here we measure whether a given science is perceived as controversial by the public.
EXAGGERATION	This science news story is overstated or exaggerated	Exaggeration is a common issue in science reporting [36] and can potentially undermine the public trust in science [37]. Here, we measure the subjective perceptions of exaggeration in scientific information.
INTERESTINGNESS	The science news story sounds interesting to me The science news story could be interesting to the general public	Perceived interestingness is considered as a key factor affecting
BENEFIT	Who will benefit from learning about this finding? The general public A segment of the public Policy makers Companies in the related industries I learned something useful from the science news story Knowing about this science could benefit a lot of people	
SHARING	I would share this science news story with someone I know directly I would share this science news story with a wider forum like a mailing list, Twitter, Reddit I would be unlikely to share this science news story with anyone	The sharability of a news article has been considered as an important dimension of the news values [21]. Here we
READING	If I'm browsing news articles from __, I would be likely to open and read the science news story above General news outlets (e.g. BBC, New York Times, Fox News) Science and Technology media (e.g. Scientific American, National Geographic) Other popular printing media (e.g. Vogue, GQ, Elle) Other popular media (e.g. TV and radio) It should not be published in public media outside the science community	Reading willingness measures the audience's willingness to read the presented information in different venues (e.g. general news outlets, science and technology media and

Table 1: The twelve-dimension framework to measure public perceptions of scientific information in media.

B.1 Raw Dataset

Our raw dataset is obtained from `altmetric.com` which contains 2.4 million media mentions of research papers. We follow the processing pipeline from previous research [51, 52] and further extract a subsample of 128,942 news/article pairs where we are able to collect the full text of the news articles. Our data spans 273 different news outlets and 57,807 different scientific articles in over 100 scientific domains.

B.2 Crowdsourcing with Representative Samples

To measure the public perception of science news stories, we designed the following pipeline to create a large-scale crowdsourcing annotation task.

Data Cleaning For our annotation task, we aim to present the full news articles while excluding information that might bias the annotators' perception of scientific information. To this end, we designed a data-cleaning pipeline that removes the following types of information: (1) Outlet: people's perception of news articles might be biased by their pre-existing attitudes towards the outlet (2) Other meta information like author, date, city, and URL, as such information might create an impression that the presented news article has already been published, which may further bias the ratings.

Data Sampling for Annotation The overall goal of our data sampling is to provide a generally balanced sample for annotation across different settings. We design a sampling procedure that balances the type of outlet, the science domain, and the popularity of the paper. Below we describe our sampling process for a batch of around 1000 news articles:

1. Sampling based on coverage types. We first categorize each outlet into three categories: General, Press Release, and SciTech blogs. Furthermore, we balancedly sample the same amount of papers (e.g. 80 in this case) from three settings: (1) mentioned in all three types of outlets (2) mentioned in both PR and SciTech outlets (3) mentioned in both PR and General outlets. For each selected paper, we randomly sample 1 article from the corresponding category. Such a process will lead to 560 sampled news articles.

2. Sampling based on outlet types. We further expand our sample by randomly sampling 50 news articles from each type of outlet separately. This step leads to 810 news stories.

3. Upsampling based on domains. To balance the stories in different science domains, we further upsample another 50 news stories in Social Science, 50 in Humanities, and 100 in engineering, which leads to 1,010 stories.

4. Upsampling based on popularity. Previous steps may lead to a sample of papers with relatively low popularity, which might have lower news values than papers that are covered in many news stories. In this step, we further upsample another 30 news articles that cover popular research articles (coverage count larger than 30).

Category	Statement	K's α	C's α
NEWSWORTHINESS	Should be published in news	0.130	
IMPORTANCE	Tackles an important issue	0.311	
EXPERTISE	Needs specialized knowledge	0.248	
UNDERSTANDABILITY	I understand the news story	0.185	
FUN	News story is fun to read	0.122	
CONTROVERSY	Finding could be controversial	0.147	
SURPRISINGNESS	Finding is surprising to me	0.096	
EXAGGERATION	This science news story is overstated or exaggerated	0.045	
INTERESTINGNESS	Could interest the general public	0.137	0.77
INTERESTINGNESS	The science news story sounds interesting to me	0.092	
BENEFIT	Could benefit policy makers	0.233	
BENEFIT	Could benefit many people	0.220	
BENEFIT	Could benefit related industry companies	0.156	
BENEFIT	Could benefit the general public	0.166	0.84
BENEFIT	Could benefit a segment of the public	0.113	
BENEFIT	I learned something useful	0.082	
SHARING	I would share this science news story with someone I know directly	0.050	
SHARING	I would be unlikely to share this science news story with anyone	0.024	0.81
SHARING	I would share this science news story with a wider forum (e.g., Twitter)	0.022	
READING	Would be willing to read it from general news outlets (e.g., BBC, NYT)	0.083	
READING	Would be willing to read it from other popular media (TV, radio)	0.074	
READING	Would be willing to read it from other print outlets (Vogue, Time, etc.)	0.082	
READING	Would be willing to read it in my social-media feed	0.047	0.83
READING	Would be willing to read it from Sci&Tech outlets (NatGeo, SciAm)	0.070	
READING	It should not be published outside the science community	0.012	

Table 2: Krippendorff's α and Cronbach's α for rating the public perceptions of science news stories.

After this process, we are able to obtain a sample of news stories that are balanced for the type of coverage, type of outlet, domain, and popularity of the paper.

Particpants We recruit participants from Prolific.com as it allows us to collect annotations from representative samples for the US and UK population. More specifically, we recruit 1,057 participants from the US and 1,044 participants from the UK. Both are drawn from a representative sample balanced for age, sex, and race. We require each participant to be fluent in English to participate in our study as the science news articles are written in English. For each news article, we set to collect 4 labels from the US and the UK sample separately.

Annotation Procedure We first ask each annotator to read the introduction of our study and then consent to participate in the annotation. Once they agree to participate, we display 5 randomly sampled science news stories in a sequence. To measure the public perception of science news values, we follow the framework introduced in Sec 2 and ask each annotator to rate the given science news stories based on the 25 statements in Table 1. Each annotator is asked to rate the 25 statements for each science news story following a 1-5 Likert scale, where 1 means strongly disagree and 5 means strongly agree. After the annotation phase, each annotator is asked to answer a list of questions regarding their demographic information, educational background, experience/interest in science, and political orientation.

B.3 Quality Check and Post Processing

Our final dataset contains 10,489 annotations for 1,506 news articles. To evaluate the quality of the created dataset, we calculate Krippendorrf's α and Cronbach's α to evaluate the reliability of the inter-annotator agreement and the internal consistency of the scales.

Inter-annotator Agreement Table 2 shows the inter-annotator agreement as calculated by Krippendorff's α [60] for each dimension. The average IAA across all the statements is 0.11, suggesting that raters' perceptions of science news stories is subjective. Despite its subjectivity, we observe that the IAA for rating certain statements is generally higher, for example, the IAA for "The finding needs specialized knowledge", "Could benefit many people" and "Tackles an important issue" are generally above 0.2, suggesting that the annotators have more agreement when rating relatively more objective dimensions. While the overall IAA seems low in our data, it is worth noting that a similar level of IAA was observed in a similar annotation task conducted by **(author?)** [23]. **(author?)** [23] collected newsworthiness ratings from both expert raters and crowd workers. The IAA between expert raters is 0.16 and ranges from 0.09 to 0.18 for crowd workers, suggesting that the low IAA in our task is potentially caused by natural subjectivity instead of low-quality annotations.

Internal Consistency To evaluate the internal consistency of the instruments used for measuring public perceptions, We calculate Cronbach's α [61] for each group of statements including INTERESTINGNESS, BENEFIT, SHARING, and READING and the C's α is 0.77, 0.84, 0.81, 0.83 respectively, suggesting that the scales used to measure public perceptions are generally reliable.

Post Processing For dimensions measured with multiple latent statements (e.g. READING and SHARING), we average all the relevant ratings from an annotator to create a unified score for the dimension. After this step, each news article receives 12 scores from a single annotator. We then average the ratings of all the annotators to obtain the overall perception value. As a result, the final dataset contains perception scores for 1,506 news stories, and each story has 12 scores representing the public perceptions of it on the 12 dimensions.

Usage for Modeling Training One of the final goals of our study is to build NLP models that are able to automatically estimate public perceptions of scientific information. Therefore, it is worth discussing the suitability of using this data for building machine learning models. First of all, our post-processing steps ensure that we are modeling the mean relative perceptions instead of the numerical ratings. More specifically, when a news article receives a high average newsworthiness rating, it is more likely to be more newsworthy than another article with a low average score. To quantify the reliability of our final average scores as a relative measurement of news perceptions, we follow a similar procedure as [62] to convert the Likert ratings into ranking scores. We then calculate the correlation between the average rating and the ranking score on all the dimensions. The average correlation across all the dimensions is 0.8, suggesting that the average rating could reliably measure the relative perception orders across articles. Secondly, while we observe relatively low inter-annotator agreement, existing studies have suggested that low-IAA data can still be used to train reliable machine learning models [63, 64]. For example, in a similar task modeling the newsworthiness of research articles [23], while the IAAs between crowd workers are around 0.1, a model trained on the annotated data could still precisely predict the expert ratings

for newsworthiness. As we are following a similar post-processing pipeline as [23], we argue that our annotations, while with relatively low IAA (from 0.01 to 0.3), could still be used to train NLP models for automatically predicting public perceptions of scientific information.

C   Computational Modeling of Public Perceptions

To model the public perceptions of scientific information at scale, we develop a data-driven NLP pipeline that is able to estimate the public perception of each science news article.

Model To allow automatic modeling of perceived news values for each news story, we fine-tune a RoBERTa-Large model [54] on the annotated dataset. We split the annotated data into train, validation, and test set following the ratio of 7:1:2. We concatenate the title and the body of each news article as the input document. We follow a multi-task regression setting and on the output layer, the model predicts a score ranging from 1-5 for each statement, so that for each science news article, we could directly obtain the predictions for all the dimensions. We fine-tune each model for 10 epochs and then choose the best-performing model on the validation set as the final model and further evaluate the model's performance on the hold-out test set.

Result Table 3 shows the model performances on the test set. We found that the model is generally able to predict the perception of science news stories on most dimensions. The correlations for 7 out of the 12 dimensions are above 0.5. However, the model performs poorly on dimensions that are relatively more subjective. For example, the prediction performances for exaggerating and sharing are only 0.31 on news stories. Overall, our experiment suggests that automatically measuring the public perceptions of scientific information in media is doable but also challenging.

Dimension	News	Reddit	Twitter	Paper
NEWSWORTHINESS	0.54	0.69	0.42	0.52
READING	0.44	0.52	0.49	
UNDERSTANDABILITY	0.71	0.75	0.79	0.65
EXPERTISE	0.74	0.40	0.39	0.66*
FUN	0.60	0.76	0.72	
IMPORTANCE	0.77	0.62	0.20	
INTERESTINGNESS	0.52	0.34	0.14	
SHARING	0.31	0.69	0.48	
BENEFIT	0.67	0.35	0.24	
SURPRISINGNESS	0.41	0.56	0.57	
CONTROVERSY	0.49	0.16	0.05	0.44
EXAGGERATION	0.31	0.38	0.16	
Overall	0.54	0.52	0.39	

Table 3: Model performances as measured by Pearson's *r* over different test sets. The model generally performs well on News and Reddit posts but performs poorly on tweets. "Paper" refers to the Arxiv abstract newsworthiness dataset [23]. We calculated the correlation between our model's predictions and crowd ratings for the corresponding category. Overall, our model also shows good performance on the Arxiv abstract newsworthiness dataset. * means that an absolute correlation score is used because the two measured dimensions are conceptually opposite.

Model Generalizability on Social Media Text To evaluate the model performance in other domains, we created another test set for Reddit posts and tweets mentioning research articles or news stories. A similar annotation pipeline was adopted to annotate 50 Reddit posts and 50 tweets about scientific information. Each post is annotated by 8 annotators and Table 4 shows the internal consistency and inter-annotator agreement for the two datasets. We further run the trained model over this dataset and the results are shown in Table 3. Overall we observe that the model performs well on Reddit posts but performs poorer on tweets. This is potentially because the tweets are usually shorter and may contain more noise.

	K's α Reddit	K's α Twitter	C's α Reddit	C's α Twitter
Newsworthiness	0.41	0.32		
Understandability	0.36	0.45		
Expertise	0.33	0.53		
Fun	0.31	0.06		
Importance	0.60	0.46		
Surprisingness	0.26	0.13		
Controversy	0.31	0.28		
Exaggeration	0.08	0.23		
Interestingness	0.40	0.33	0.79	0.75
Benefit	0.43	0.41	0.86	0.84
Sharing	0.15	0.23	0.73	0.81
Reading	0.30	0.24	0.81	0.77

Table 4: Krippendorff's α and Cronbach's α for rating the public perceptions of Reddit and Twitter posts. We observe similar or higher levels of reliability and internal consistency compared with news annotations.

Model Generalizability on Paper Abstracts To test the generalizability of our model on research articles, we also evaluate model performance on the Arxiv newsworthiness dataset [23], which contains readability, newsworthiness, and controversy labels for computer science paper abstracts. We calculate the correlation between our model predictions and the ground truth crowd labels for overlapped dimensions. More specifically, for NEWSWORTHINESS and CONTROVERSY, we directly calculate the correlations. For UNDERSTANDABILITY and EXPERTISE, we calculate the correlation with the readability labels in the Arxiv newsworthiness dataset. As shown in Table 3, we found moderate to high correlations between our model predictions and the ground truth labels. For example, the Pearson's r for NEWSWORTHINESS is 0.52 and is 0.65 for UNDERSTANDABILITY. As a comparison, the correlation between the ground truth expert ratings is 0.33. Since the Arxiv abstracts were all sampled from computer science papers, such a result suggests that our model is able to capture nuanced differences for content in the same science domain despite the low IAA of the annotations.