Steering AI-Driven Personalization of Scientific Text for General Audiences

TAEWOOK KIM*, Northwestern University, USA
DHRUV AGARWAL*, Cornell University, USA
JORDAN ACKERMAN†, Center for Advanced AI, Accenture, USA
MANASWI SAHA, Accenture Labs, USA

Digital media platforms (*e.g.,* science blogs) offer opportunities to communicate scientific content to general audiences at scale. However, these audiences vary in their scientific expertise, literacy levels, and personal backgrounds, making effective science communication challenging. To address this challenge, we designed *TranSlider*, an AI-powered tool that generates personalized translations of scientific text based on individual user profiles (*e.g.,* hobbies, location, and education). Our tool features an interactive slider that allows users to steer the degree of personalization from 0 (weakly relatable) to 100 (strongly relatable), leveraging LLMs to generate the translations with chosen degrees. Through an exploratory study with 15 participants, we investigated both the utility of these AI-personalized translations and how interactive reading features influenced users' understanding and reading experiences. We found that participants who preferred higher degrees of personalization appreciated the relatable and contextual translations, while those who preferred lower degrees valued concise translations with subtle contextualization. Furthermore, participants reported the compounding effect of multiple translations on their understanding of scientific content. Drawing on these findings, we discuss several implications for facilitating science communication and designing steerable interfaces to support human-AI alignment.

CCS Concepts: • **Human-centered computing → Interactive systems and tools**; **Human computer interaction (HCI)**; **Empirical studies in HCI**; • **Computing methodologies → *Natural language generation.***

Additional Key Words and Phrases: Human-AI alignment, Scalable personalization, Science communication, Steering AI, Translation

1 Introduction

Science communication refers to the process of "communicating complex scientific information with *general audiences*[1] to improve public awareness, interest, and understanding of science" [22]. Both academic and industry professionals strive to make their ideas more accessible. As a result,

---

*Work done during an internship at Accenture Labs, San Francisco
†Work done during their tenure at Accenture Labs
[1]We refer to "general audiences" as individuals who are non-scientists, coming from diverse educational backgrounds and domains such as arts, business, education, healthcare, and product design.

Authors' Contact Information: Taewook Kim, Northwestern University, Evanston, IL, USA, taewook@u.northwestern.edu; Dhruv Agarwal, Cornell University, Ithaca, NY, USA, da399@cornell.edu; Jordan Ackerman, Center for Advanced AI, Accenture, San Francisco, CA, USA, jordan.ackerman@accenture.com; Manaswi Saha, Accenture Labs, San Francisco, CA, USA, manaswi.saha@accenture.com.

science communication platforms, such as science blogs [42] and science magazines [90], have been translating scientific information into more accessible formats for decades. However, a single version of a text often falls short of meeting the diverse needs of a general audience [10]. For instance, a journalist with an interest in quantum physics would have vastly different context and comprehension needs than a college student majoring in physics.

Scientific research articles often employ domain-specific language, including jargon and complex sentence structures, which can pose significant barriers to comprehension for general audiences. Language models offer the capability to personalize content to various user contexts, transforming both the style (*e.g.,* casual to formal [27] or romantic [51]) and content (*e.g.,* simplifying complex ideas [10]). This adaptive personalization creates opportunities for enhancing audience engagement and understanding through more effective content dissemination [50]. Notably, large language models (LLMs) can achieve this personalization at scale (*i.e.,* AI-scalable personalization), generating multiple tailored variations of content for diverse audiences [10, 50, 82]. This scalability positions LLMs as powerful tools in advancing inclusive and effective science communication [80].

Among the many strategies in science communication from prior HCI/CSCW literature, analogies have proven effective in translating technical content into more accessible forms [9]. For instance, the structure of the solar system is often used as an analogy to explain the structure of an atom. While such analogies can broaden understanding of scientific information, they are limited in number, can be generic, and may not resonate equally well with all audiences as individual comprehension is shaped by societal, cultural, educational, and personal backgrounds. For example, the space analogy assumes familiarity with the solar system [50]. LLMs have demonstrated the ability to generate personalized analogies that can help people understand complex concepts and ideas within their own context [30]. Building on this capability, we leverage LLMs to generate multiple personalized analogies tailored to individual readers' contexts.

We designed and implemented *TranSlider* (*Trans*late through *Slider*), an interactive reading interface that enables users to steer the degree of personalization in scientific text through an adjustable slider. *TranSlider* allows the user to specify a degree of personalization from 0 to 100 and utilizes their background information (*e.g.,* education, hobbies, location) to present relevant analogies. The slider enables intuitive exploration of various personalization degrees, allowing users to quickly review multiple translations. We employed *TranSlider* as a research probe to ask the following research questions:

RQ1: What is the utility of AI-driven personalized translations of scientific text?
RQ2: What is the impact of interactive reading features on user experience? Specifically:
    2.1: How does exploring multiple translations influence readers' comprehension and engagement with scientific text?
    2.2: How does the slider interaction to steer the degree of personalization influence readers' comprehension and engagement with scientific text?

To answer these questions, we conducted a user study with 15 non-expert participants who used the tool to understand two scientific texts by exploring multiple personalized translations. We conducted post-session semi-structured interviews to elicit feedback on their experience. We used thematic analysis on their qualitative responses and conducted descriptive quantitative analysis on tool usage logs to understand their exploration behavior (*e.g.,* number of generated translations. range of explored personalization degrees).

Participants found the analogies useful in understanding the content. Some favored general analogies (*e.g.,* a participant liked a construction analogy describing the body's cells as building sites—low personalization), while others preferred detailed, personalized analogies (*e.g.,* a participant appreciated a baking analogy to explain harmful emissions from lithium-ion batteries—high

personalization). Participants noted that reading multiple translations with varied analogies allowed them to piece together a fuller understanding of the scientific text, correcting misunderstandings along the way. Participants found the tool beneficial for learning about unfamiliar topics but were cautious about the reliability of AI-generated translations. Based on these findings, we discuss implications for science communication and the need for interactive techniques to steer models towards human pluralistic preferences, rather than solely relying on machine learning approaches.

Overall, our contributions to the CSCW community include,

(1) A novel slider-based LLM interaction to enhance understanding of scientific content through the exploration of analogy-driven personalized translations
(2) An investigation to examine the utility, benefits, and limitations of AI-driven personalized translations in science communication
(3) Implications for science communication and more broadly for cross-disciplinary communication, and designing interfaces for human-AI alignment.

2 Related Work

We review prior work covering science communication on digital media platforms, personalization in science communication, and existing interactive reading interfaces for scientific articles.

2.1 Science Communication on Digital Media Platforms

Science communication takes many forms in modern media for communicating complex scientific information to general audiences for improving public awareness, interest, and understanding of science [22]. Traditional channels include museums [13], exhibitions [66], and TV series [84], while web-based digital platforms have emerged as a powerful new stream. For example, online videos have become an influential source for science learning, especially during the COVID-19 pandemic, when many people turned to digital resources for scientific information [18]. Social media platforms such as Reddit (*e.g.,* r/science), Twitter (now X), Bluesky, and Mastodon offer interactive spaces for scientists to share research and engage with the public [35, 48, 69]. Finally, science blogs have been valuable channels for explaining complex scientific content to general audiences [46]. For example, more than half of college students actively use blogs as a learning resource [41]. A key advantage of science communication via digital media, a.k.a. online science communication [92], is its ability to reach a wider audience (*e.g.,* [48, 74]).

Prior HCI/CSCW literature has uncovered several key challenges scientists face in online science communication: engagement, translation, and dissemination [92]. First, scientists rarely participate in digital media platforms [99]. This limited engagement is due to the conventional academic nature that has put scientists as neutral observers rather than active communicators [70, 79], and practical constraints such as limited time and lack of communication training [14]. While recent research shows a gradual shift towards more active public engagement by scientists [19], the extent and impact of this change remain to be fully understood. Secondly, even for scientists who engage with digital communication, translating complex scientific concepts into accessible language is challenging [81]. These challenges may be compounded by the variety of readers with different backgrounds, levels of scientific literacy, and preferences [81, 87, 89]. Creating multiple versions tailored to different expertise levels is infeasible for individual scientists [20, 32]. Lastly, translation challenges can entail broader concerns such as misinformation on social media platforms [25]. The dissemination of inaccurate interpretations or misrepresentations of their work could have unwanted consequences for public understanding and their professional reputation [67, 93].

Given these challenges, this paper focuses on addressing the translation barrier in personalized science communication on digital media platforms. Specifically, we explore how to facilitate creating

multiple translations of scientific content to effectively accommodate general audiences with varying levels of scientific background at scale.

2.2 Personalization in Science Communication: Tools and Approaches

In science communication, personalization is an approach that translates scientific content into contexts familiar to each individual, making it more interesting, accessible, and engaging for non-scientists [6]. For example, when explaining biological concepts to individuals without a background in biology, saying "try not to drink soda and eat candy" may be more effective than "reduce glucose consumption", as the former uses more familiar contexts while conveying the same information to the intended audience [62]. Given its potential effectiveness for science communication, researchers have explored various strategies for personalizing information across contexts.

Prior approaches in HCI and CSCW have personalized scientific content using structured templates, often yielding limited benefits to the reader. For instance, Persalog personalizes news articles by substituting specific segments—such as city, state, or country names—based on geographical relevance [5]. Other studies have employed template-based rephrasings. For instance, Hullman *et al.* and Kim *et al.* generate alternative expressions of measurements following a template: "{number} lb is about (X.X) times the weight of {familiar object}" [44] or "The distance between locations is (number) times your distance to a {familiar landmark}" [54]). These methods rely on structured databases, such as geographical information [5], landmark references [54], and patients corpora [29], to compute contextual relevance. While some work has attempted more flexible language generation (*e.g.,* personalized medical brochures [29]), these efforts result in unnatural and ill-formed sentences, limiting their effectiveness.

These approaches predate the emergence of LLMs that could address these limitations: reliance on structured templates and databases. LLMs can transform text at scale without any databases, making it feasible to generate multiple translations on behalf of scientists. Recent HCI studies have explored the feasibility of LLMs in simplifying complex ideas (summarization) [11, 37, 53], retargeting concepts into another domain (contextualization) [12, 30, 64], and explaining content through user-familiar contexts (personalization) [10, 50]. Albeit in other domains, these studies show that LLMs can reduce the reliance on databases to compute contextual relevance and generate natural-sounding language expressions tailored to various user contexts. Building on these works, our study explores integrating contextual personalization with LLMs' capability to generate analogies [30]. We aim to design an LLM-powered tool that leverages readers' contextual information to generate personalized analogies for scientific texts.

2.3 AI-Powered Reading Interfaces for Scientific Articles

Recent research has proposed AI-powered interactive reading interfaces for science articles. These systems summarize sections of the paper and help navigate within the article to efficiently find relevant information [11, 33]. For example, ScholarPhi assists users in quickly looking up definitions of nonce words and symbols defined elsewhere in the paper [40]. Researchers have also explored more free-form question-answering systems to help users identify relevant sections of a paper based on their queries [97]. However, these primarily cater to users already motivated to engage with scientific literature, such as healthcare stakeholders (*e.g.,* patients [49]), students, and scientists.

While general-purpose tools like ChatGPT or AI-enabled PDF readers (*e.g.,* Adobe Acrobat with built-in chatbots) can assist with reading scientific texts, they are not designed for scientific content, and require users to know what to ask. In contrast, our tool focuses on everyday readers with a casual interest in science—individuals who may lack the time, background, or motivation to actively engage with science articles. We position our tool as an AI-powered intermediary [50] that delivers accessible, engaging science communication through analogy-driven, personalized translations.

Rather than expecting readers to extract meaning themselves, our tool meets them where they are by adapting content to their context.

3 Interaction Paradigms

3.1 Existing Design: Model-Initiative

Existing personalization systems predominantly employ a model-initiative design, where AI models automatically collect and process user data to deliver personalized content without explicit user intervention [26]. While efficient, this interaction design presents two primary human-AI alignment issues, namely personality alignment and user privacy.

3.1.1 Personality Alignment Issues. Model-initiative approach faces the *personality alignment* issue [98], *i.e.,* the personalized outcomes are not always well aligned with users' actual interests or needs [31, 100]. For example, consider an AI system that automatically personalizes results based on a user's current location (*e.g.,* Tokyo, Japan) when the user is actually interested in contextualized results of a different location (*e.g.,* Bergen, Norway). In this case, the AI personalization is misaligned with the user's actual interests. This alignment problem is compounded by the contextual nature of personalization preferences. Individuals may desire different levels of personalization depending on the context and motivation of their reading [10, 31, 36, 59]. For example, someone might prefer highly personalized content when reading medical articles relevant to their health conditions [63], but relatively less or lower personalization when casually exploring scientific topics of general interest. To accommodate varying contextual needs for each individual at scale, the model-initiative approach would naturally require invasive data collection and inference about users' contexts.

3.1.2 Privacy Issues. Model-initiative design assumes that AI models have access to users' private data [26]. However, from a user's perspective, it is unclear how much information is gathered, what types of information is gathered, and when [31]. Furthermore, users typically have minimal control over their personal data [60]. These privacy concerns become more problematic in the context of science communication, which aims to reach diverse, mass audiences [22]. The model-initiative paradigm would require collecting and processing private data from the entire population of potential readers to provide personalized translations of scientific texts. This approach not only raises ethical concerns but also creates significant barriers to adoption, as many people would be reluctant to share their personal information.

3.2 Our Design: Controlled User-Initiative

We explore an alternative approach: user-initiative design. Existing LLM interfaces, such as Chat-GPT [76] and Claude [7], provide prompt interfaces where users can specify all personalization parameters and their intentions. While a prompt-based approach maximizes user control, it presents challenges specifically for science communication: allowing unlimited user modification risks compromising the scientific accuracy crucial for effective information dissemination. High degrees of freedom could lead to the distortion of key ideas or inadvertently promote misinformation, even when users have no intention to mislead [50, 96].

We propose a controlled user-initiative design where readers can explicitly adjust specific aspects for personalization. To minimize the risk of misinformation or hallucination, rather than allowing free-form prompt interactions that could introduce unexpected results, we constrained user inputs to two structured elements: i) steerable **degrees** of personalization using sliders, addressing the alignment issue, and ii) editable structured **type** of contextual information via user profiles with predefined fields to alleviate the privacy issue. This design philosophy aims to preserve privacy by
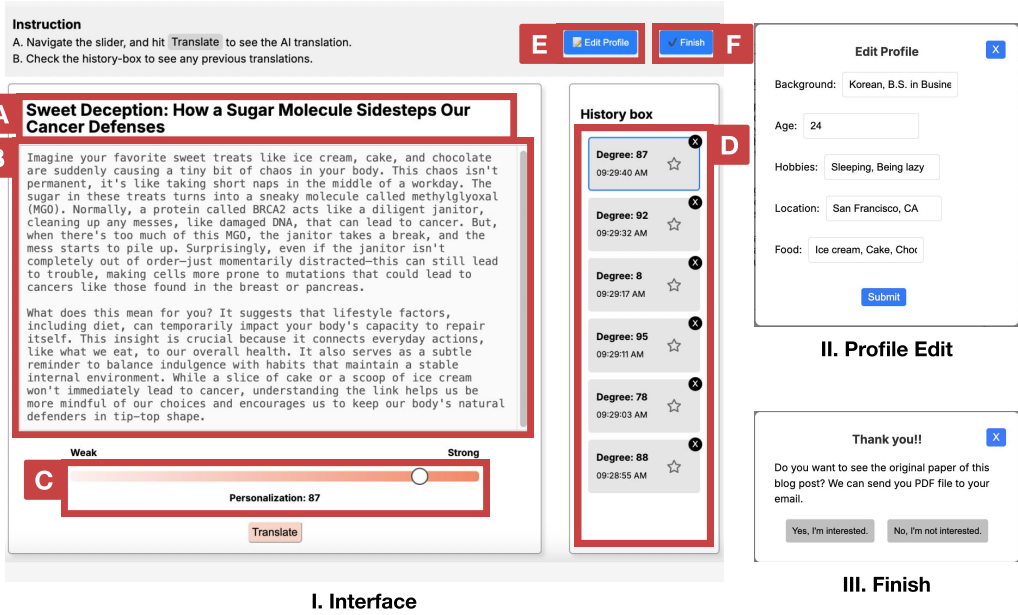
Fig. 1. The interface design of *TranSlider*. [A] The AI-personalized title of the science blog post appears here. [B] AI-personalized translations of science articles will appear here. [C] Users can navigate the slider to adjust their preferred degree of AI personalization for a science article intro. [D] For each personalized translation, a history item is stacked in the history box. Users can revisit any previous history items to see the translations. [E] Users can click the edit profile button to see the Profile Edit window [II] to adjust their profile information. [F] The Finish button triggers a pop-up [III] to collect the user's interest in reading the original paper. We use the response as a proxy measure for their interest in the content after reading the personalized translations.

minimizing detailed personal data collection and granting users control over input, while enabling contextually relevant personalization that aligns with their needs and preferences.

4 TranSlider: Design and Implementation

With this design philosophy in mind, we designed and implemented *TranSlider* (*Trans*late through *Slider*), an LLM-powered reading interface to help general audiences better understand scientific texts by generating personalized translations through a slider interaction and user profiles.

4.1 Interface Design

We designed our tool with three main components: the translation panel, the history box, and the edit profile button (Figure 1). The header provides instructions on using the interface; these instructions were brief as participants received a detailed tutorial at the start of the study.

4.1.1 Translation Panel. The translation panel has the main personalization features (Figure 1 [A], [B], [C]). The personalized title of the scientific article being viewed is displayed at the top with the personalized translation underneath it. A slider, labeled from weak to strong personalization, allowed participants to adjust the degree of personalization. The current personalization degree, a numeric value between 0 and 100, is shown just beneath the slider. Participants were instructed to adjust the slider to their preferred degree value before translating using the Translate button. This action triggers an LLM call, which generates and displays a personalized translation.
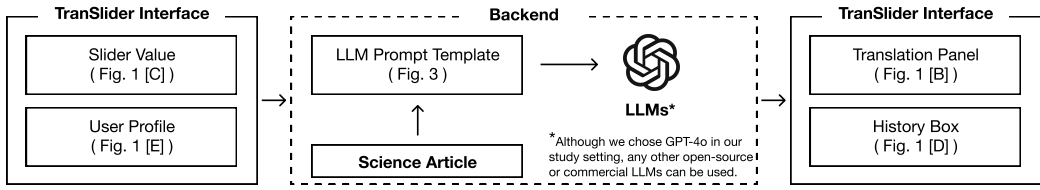
Fig. 2. The workflow of *TranSlider*.

4.1.2 History Box. The history box on the right allows participants to revisit, review, and compare translations they had seen during the session. Each new translation is automatically added to the stack, labeled with its personalization degree and the time of generation. Users can click on any past translation to view it, which would update the translation panel accordingly. They could also mark a translation as a favorite by clicking on the star icon beside it. Additionally, users can delete translations from the history by clicking the cross button, though we found that participants rarely used this feature in our study.

4.1.3 Edit Profile. The edit profile dialog box is accessible by clicking a button on the tool header (Figure 1 E). This displays a short form (Figure 1 [II]) containing information about the participant including their background, age, hobbies, location, and favorite food. This profile information is used to tailor the translation to the user's context.

4.1.4 Finishing a Session. The finish session button (Figure 1 F) served as a proxy measure for their interest in the content after reading the personalized translations. The button triggers a pop-up (Figure 1 [III]) that collects user response on whether they would like to receive the original research paper PDF via email. This button was used at the end of each article's translation session.

4.2 System Workflow and Implementation

The workflow of *TranSlider* is illustrated in Figure 2. Once a user adjusts the personalization slider and submits a translation request on the frontend, the backend integrates both the selected personalization degree and the original scientific text into the prompt template (detailed in Figure 3). Inspired by existing LLM-powered tools' prompt designs (*e.g.,* [23]), we used the chain-of-thought style instruction [91] to guide the model to reason how the slider input value (personalization degree) should be reflected in the output (personalized translation). With this combined prompt input, the backend makes an API call to the LLM to generate the personalized translation. The API response with the AI-personalized translation is then transmitted back to the translation panel in the frontend. We implemented *TranSlider* using React.js for the frontend interface and Node.js for the backend. We used OpenAI's GPT-4o model for the LLM.

Note on LLM Model Choice. While our implementation utilized GPT-4o, this model could be readily replaced with alternative commercial or open-source models. For instance, specialized models like OpenScholar-8B—specifically designed for scientific contexts through retrieval augmentation—outperform GPT-4o for scientific content [8] and could potentially generate higher-quality personalized translations. Conversely, substituting with older or less capable models such as GPT-Neo [16] would likely result in lower quality and unreliable translations. Our preliminary testing confirms that alternative models like Meta's Llama-3 (open source) [71], Anthropic's Claude 3.5 (commercial) [7] also generate comparable quality of translations, suggesting that our tool could incorporate different model architectures. Whether there would be any noticeable distinctions in the overall translation quality depending on the model choice is left for future investigation.

Task: " Generate an introduction of a science blog by generating translation with *analogies* of the following <scientific article>. I want it to be more easily understandable and accessible for audiences who do not have a strong science background. "

1. Profile Setup

<profile>
My background is **${ profile.background }**.
I am **${ profile.age }** years old.
I live in **${ profile.location }**.
My hobby is **${ profile.hobby }**.
My favorite food is **${ profile.food }**.
</profile>

2. The Degree Range Setup

We define the spectrum gradients of personalization for the translation ranging from ${0} to ${100}.

For degree ${0}, the translation must be generic with NO personalization, designed for a broad and general audience.

For degree ${100}, the translation must be fully personalized and tailored to the context of **${ background }**

For degree ${50}, the translation should contain a half of generic context with another half of some personalization with some similar or high-level context of **<profile>**.

3. Examples of Generic/Personalized Context

In generating more *generic* translation, first infer the "high-level context" from the **<profile>**. Then tailor the translation based on the inferred high-level context.

<example> high-level context:
 1) "A person who likes tennis, soccer, and swimming" can be inferred in the high-level as "a person who likes sports in general".
 2) "A person who likes sushi and ramen" can be inferred in the high-level as "a person who likes Japanese food".
 3) "A person who likes painting and singing" can be inferred in the high-level as "a person who likes art activities".
</example>

In generating more *personalized* translation, consider the specific user context in the **<profile>**, and also "similar context" to tailor the translation.

<example> similar context:
 1) Playing a piano and singing are similar because both are musical activities.
 2) Swimming and jogging are similar because both are cardio exercises.
 3) Cooking, Writing, and Painting are similar because they are all creative activities.
</example>

Chain-of-Thought Style Prompt

Try to carefully consider the above-mentioned **<example>** when reasoning how the translation with the degree of **${ sliderValue }** should be much more or less personalized to the profile.

4. Generation Instructions

Given the instructions above, generate translations in *two paragraphs*:

1: A translation with an *analogy* of the research, given the degree of **${ sliderValue }**.
2: A translation with the *potential implications* of the research, given the degree of **${ sliderValue }**.

Use simple sentences. No complex sentences. Answer immediately! No preambles! No titles! No labels! No quotes!
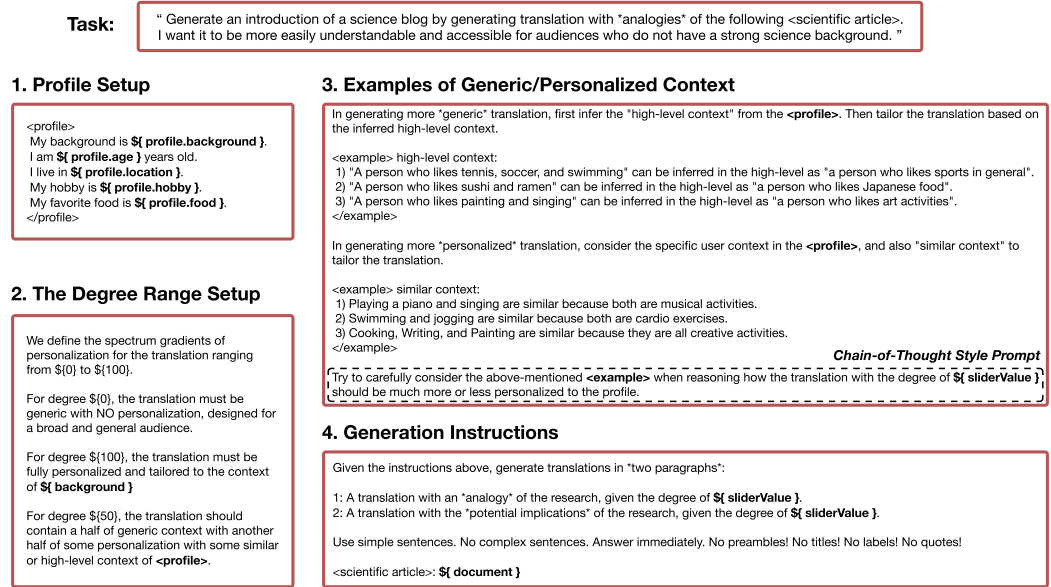
<scientific article>: **${ document }**

Fig. 3. Prompt Template Structure. The prompt begins with a **Task** overview, instructing the model to use analogies to explain a science article to general audiences. The **prompt context** consists of four parts: 1. The user profile; 2. Personalization spectrum with descriptions for degrees 0, 50, and 100; 3. In-context examples of both generic and personalized translations; 4. Original scientific article content with the slider value.

4.3 Pilot Study

Once the research team was satisfied with the LLM's output, we conducted a pilot study to determine whether the translations were understandable and free from hallucinations or misinformation. We recruited three research scientists outside of our research team. They all had published their work in the past three years. We used their authored publications in Computer Science, Biomedical Engineering, and Electrical Engineering, respectively, for the verification process.

4.3.1 Procedure. The first author conducted this study in a one-on-one in-person setting. Upon participants' arrival, we introduced *TranSlider* as a tool designed to help general audiences understand complex scientific texts. We explicitly stated that the purpose of this pilot study was to verify whether the AI-personalized translations were comprehensible and did not hallucinate any misinformation. We then asked participants to provide a link to their recent publication that they would like to share with the general audience. After a brief tutorial of our tool, we had them explore and interact with *TranSlider* at their own pace without any time constraints. We proceeded to the debriefing interview — the overall impression of the personalized translations and whether they had seen any misinformation or hallucinations — only when participants indicated they were ready to share their thoughts. Each of the three pilot studies took about 30 minutes to complete.

4.3.2 Results. All three scientists found the translations surprisingly understandable. They also confirmed that they could not detect any hallucinations or misinformation in the translations they reviewed. The electrical engineering researcher provided additional feedback on two aspects: the topic selection and the structure of translations. First, they noted that a scientific article's subject matter could influence its perceived benefits to general audiences. Some topics, like health and nutrition, might naturally interest the general audience, while they doubted whether some

Original Science Article	Personalized Translation
Title: A glycolytic metabolite bypasses ``two-hit'' tumor suppression by BRCA2	**Title**: Sugar Deception: How a Sugar Molecule Sidesteps our Cancer Defenses
Abstract: Knudson's ``two-hit'' paradigm posits that carcinogenesis requires inactivation of both copies of an autosomal tumor suppressor gene. Here, we report that the glycolytic metabolite methylglyoxal (MGO) transiently bypasses Knudson's paradigm by inactivating the breast cancer suppressor protein BRCA2 to elicit a cancer-associated, mutational single-base substitution (SBS) signature in nonmalignant mammary cells or patient-derived organoids. Germline monoallelic BRCA2 mutations predispose to these changes. An analogous SBS signature, again without biallelic BRCA2 inactivation, accompanies MGO accumulation and DNA damage in Kras-driven, Brca2-mutant murine pancreatic cancers and human breast cancers. MGO triggers BRCA2 proteolysis, temporarily disabling BRCA2's tumor suppressive functions in DNA repair and replication, causing functional haploinsufficiency. Intermittent MGO exposure incites episodic SBS mutations without permanent BRCA2 inactivation. Thus, a metabolic mechanism wherein MGO-induced BRCA2 haploinsufficiency transiently bypasses Knudson's two-hit requirement could link glycolysis activation by oncogenes, metabolic disorders, or dietary challenges to mutational signatures implicated in cancer evolution.	**Translation**: Imagine **basketball**-friendly streets in **Berkeley**. Imagine there are always two security guards on watch who ensure that no one interrupts your game. Now, suppose one of those guards gets distracted by something shiny (let's call this shiny thing MGO), leaving the field unprotected momentarily. Smash! Someone throws a rock, ruining your perfect game. This is like what happens in cells with the protein BRCA2. Usually, both guards (or both copies of BRCA2) have to be taken out for cancer to intrude. But MGO shortcuts this by temporarily distracting one guard, letting small DNA changes sneak in without permanently disabling BRCA2. Just as losing focus during a basketball game can lead to a ruined play, this temporary distraction caused by MGO can lead to cancer. The implications are significant because understanding this mechanism means we can explore new ways to prevent cancer, especially for those who enjoy a diet that influences glucose metabolism, much like loving **Korean BBQ** or **beef noodle soup** which are high in sugars. This research could lead to new dietary recommendations or treatments targeting MGO to help keep your body's game strong and uninterrupted.

Fig. 4. An example translation (degree = 100) of P14, whose hobby is basketball and rock climbing, and lives in San Francisco, CA. Their favorite foods are Korean barbeque and beef noodle soup. The first paragraph gives a brief explanation of the scientific content based on their personal context: location (San Francisco) and hobby (Basketball). The second paragraph presents the implications of this scientific knowledge to the reader's diet and health, saying that both Korean BBQ and beef noodle soup have a lot of sugar.

other topics, like environmental or geological sciences, would draw similar interest. Secondly, they suggested that highlighting some practical implications of studies could make translations more directly beneficial to general audiences.

4.3.3 Implications. The results of this pilot study influenced our prompt and study design. We refined the instruction prompt (see 4. in Figure 3) to generate two distinct paragraphs: the first presenting an analogy explaining the study, and the second discussing implications to the user (example in Figure 4). Previously, the translation only explained the study without offering personalized implications. We also decided to include two different types of scientific articles for our study: one with directly relatable content and another with less relatable content (Section 5.1).

5 User Study

Our study aimed to obtain readers' feedback and usage data regarding two aspects of the tool: i) the utility of AI-personalized translation of scientific text (RQ1), and ii) the impact of interactive reading features on user experience (RQ2). Below, we describe the details of our study design, including the choice of science articles, participants, study procedure, and data analysis methods.

5.1 Science Articles

We selected scientific articles from two distinct domains: *health* and *environment*. While health-related topics can generate strong interest among broad audiences due to their direct personal relevance (*e.g.,* [28, 45, 88]), environmental topics might feel less immediately relevant to individuals' personal contexts. We deliberately chose these contrasting domains to demonstrate that the utility of AI-personalized translations may not be dependent on a particular topic; if participants could perceive the value of AI-personalized translations across both more personal (health) and less personal (environmental) topics.

To select science articles that the general audience would be interested in, we searched articles in the Reddit sub-channel r/science[2]. The first author compiled a list of 10 candidate articles for

---

[2]https://www.reddit.com/r/science/

		Gender	Age	Education	Experience		Science Articles	
					w/ Sci$^\alpha$	w/ LLMs$^\beta$	Session I	Session II
A	P1	Female	43	UX Design†	3	Yes | No	health	environment
	P2	Male	23	Economics	2	Yes | No	health	environment
	P3	Male	32	Industrial Eng.	4	Yes | Yes	health	environment
	P4	Female	26	Industrial Eng.	2	Yes | No	health	environment
	P5	Female	24	Statistics	2	Yes | No	health	environment
	P6	Male	34	Business	1	Yes | Yes	health	environment
	P7	Female	24	Business	4	Yes | No	health	environment
B	P8	Male	25	Economics	3	Yes | No	environment	health
	P9	Female	25	Economics	2	Yes | No	environment	health
	P10	Female	23	Biochemistry	4	Yes | Yes	environment	health
	P11	Female	24	Computer Sci.	2	Yes | No	environment	health
	P12	Female	23	Information Sys.	3	Yes | No	environment	health
	P13	Female	25	Business	5	Yes | No	environment	health
	P14	Male	24	Cognitive Sci.	3	Yes | Yes	environment	health
	P15	Female	25	Business	2	Yes | No	environment	health

Table 1. Demographic and background of participants. †P1 does not have a degree, but has a certificate in UX Design. $^\alpha$Experience reading scientific text in the last three years (1: Never, 2: A few times per year, 3: A few times per month, 4: A few times per week, 5: Daily). $^\beta$Experience with LLMs (first value represents general LLMs experience, second value represents LLM use specifically for reading).

each domain from top-rated posts within the past year. Through iterative discussions among collaborators, we narrowed our selection to two articles: one in health [55] and one in environment [38]. To mitigate potential ordering effects, we counterbalanced the presentation sequence across participants, with half viewing the health article first followed by the environment article, and vice versa for the remaining participants (last two columns in Table 1).

5.2 Participants

We recruited 15 participants (10 female and five male) through mailing lists and word-of-mouth referrals. Our recruitment criteria were designed to represent the general public with an interest in science, but not scientists. We recruited people above 18 years of age and different educational domains—a mix of science-based (*e.g.,* engineering) and non-science based fields (*e.g.,* business). We excluded people who: (a) had a doctorate degree or a research position, and (b) had a degree in *health* and *environment* fields—the topics of the scientific articles used in our study. This approach allowed us to secure participants with various domain backgrounds with a baseline level of interest in science and unfamiliarity with the chosen articles' fields. We also asked participants how often they read scientific articles (*e.g.,* science blogs) to ensure representation across a spectrum of regular readers and those who rarely engaged with scientific content.

Table 1 illustrates the details of participants' demographics. All participants were fluent in English and aged between 23-43 years (*M*=26.7, *SD*=5.3). Most participants held bachelor's degrees in various fields, while one participant (P1) had no degree. In the past three years, eight participants read scientific articles (*e.g.,* science blogs) at least monthly, while the others did so rarely. Regarding LLM experience, all participants were familiar with tools like ChatGPT; four actively used them to understand complex texts (*e.g.,* technical reports or academic papers), while the others did not.

5.3 Procedure

The study was conducted one-on-one and in-person. We audio and screen recorded study sessions that took an hour on average (*M*=60.5 min, *SD*=10.9 min). All participants received a 75 USD virtual gift card as compensation for their time.

5.3.1 Pre-study Survey (5 min). Before conducting the exploratory study, we collected participant background information through a pre-study survey. This included their experience with LLMs and reading scientific texts. We also gathered personal information such as their professional background, hobbies, familiar locations, and favorite foods, which was later used for implicit AI personalization in Session I.

5.3.2 Tutorial (15 min). After the pre-study survey, we showed the tool interface through a shared monitor in the same room. We explained the tool's purpose and provided a tutorial using a structured script. The tutorial covered tool features such as sliders, buttons, history box items, and included a demonstration of an example interaction. Throughout the tutorial, we encouraged participants to ask questions to ensure their complete understanding of the tool. After the guided tutorial, participants freely explored the tool until they felt comfortable to begin the main study sessions.

5.3.3 Session I (Implicit) & II (Explicit) (10 min × 2). As described earlier, all participants conducted the two study sessions in the same order to ensure gradual exposure to the full-featured tool interface. In the first session, participants interacted with the tool version that *implicitly* employs their profile to tailor the scientific articles. We explained to participants that the tool used the profile data they provided in the pre-study survey for personalization. To illustrate this concept, we gave them familiar examples of personalization, such as how Google Chrome leverages users' account data or how social media platforms provide personalized content based on their accounts [47]. After completing the first session, they began the second session with our fully featured tool that enabled users to *explicitly* edit their profiles.

In both sessions, we asked participants to freely explore varying degrees of personalization through the slider control. We ensured they reviewed at least five different degrees of personalized translations, and then selected three favorite translations by marking a star to the associated history item. After completing each session, participants responded to a post-session survey, where they rated their familiarity with the scientific articles and wrote the key takeaways about the translated article in 1-2 sentences. We then conducted a post-session interview focusing on: (a) their thought process while interacting with the personalization slider, (b) their rationale for selecting favorite translations, and (c) examples of translations they disliked and their reasons.

5.3.4 Debrief Interview (20 min). After completing both sessions (implicit and explicit), we conducted a semi-structured interview aligned with our research questions. The questions included i) the participants' perceived quality and utility of AI-personalized translations (RQ1), ii) their experience with interactive reading features such as multiple translations, history items, and the slider (RQ2), iii) any particular aspects they liked or disliked, and iv) overall concerns and suggestions for improvement. At the end, participants were free to ask any questions or share additional comments.

5.4 Analysis

5.4.1 User Behavioral Logs. We recorded user logs to analyze participants' tool exploration behaviors. The collected data includes user profile information (implicit), profile edits (explicit), explored degrees of personalization, all generated translation texts, favorite translations marked by users, deleted translations, and precise timestamps for all these actions. We analyzed them by plotting and computing descriptive statistics (*e.g.,* mean, std, max, and min). We did not conduct a statistical
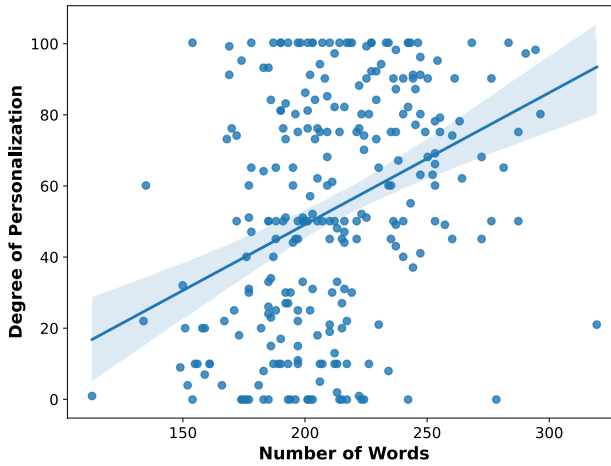
Fig. 5. The correlation between the personalization degree and personalized translation length (#words).

analysis of our data between conditions or groups, because we had a small sample size (*N*=15) [34]. More importantly, our study design was intended to enable qualitative analysis.

5.4.2 Interview Logs. We used automated services to transcribe all interview recordings. Then we conducted thematic analysis [17] on the quotes organized by interview question for each participant. For the open-coding process, the first author read quotes and labeled them with initial codes (*i.e.,* low-level codes) using Miro board cards[3]. Then the first and the second authors collaboratively reviewed the low-level codes to discuss and group codes into potential themes (*i.e.,* high-level themes). The first author structured a thematic map showing themes for each corresponding research question. Next, the research team collaborated to refine the thematic map more clearly, build a coherent narrative, and reach a consensus. During this process, we excluded themes that were not directly relevant to our research questions. Ultimately, we finalized the themes for each research question: the utility of AI-personalized translations (RQ1), the impact of interactive reading features on user experience (RQ2), and proposed use cases and associated concerns.

6 Findings

We first summarize findings from user behavior log data. We then synthesize participants' interview responses supported by user log data to substantiate additional insights and suggest potential opportunities for future research directions. Our findings cover three key aspects: i) the utility of AI-personalized translation for scientific text, ii) the impact of interactive reading features on user experiences, and iii) the potential use cases and concerns.

6.1 Quantitative Findings from User Logs and Pre-study Surveys

6.1.1 Users' Exploration Behaviors. Participants generated a total of 268 AI-personalized translations using *TranSlider*, with an average of 8.93 translations explored per session (min: 5, max: 19). We observed a positive correlation (Pearson's *r* = 0.36) between the degree of personalization and length of the translation (Figure 5), showing that more personalized translations tended to be slightly longer. This subtle increase in length could be attributed to the need for integrating more

---

[3]http://miro.com/

(a) Participants tend to explore extreme personalization degrees (0 or 100) more often, yet their preferred translations typically fell within moderate ranges, despite less exploration of middle ones.

(b) Those with extrinsic motivation for reading scientific articles tend to prefer personalization lower degree, while those with intrinsic motivation like translations with higher degrees.

Fig. 6. Illustration (a) shows the histogram of two distributions: all translations vs. user-selected translations. Illustration (b) displays a box plot of favorite translations by motivations for reading scientific articles.

detailed personal context in highly personalized translations, whereas low personalization required little to no incorporation of personal information.

Participants demonstrated various exploration trajectories, including: i) incremental—starting from low degrees like 2 and moving toward higher degrees, ii) decremental—starting from high degrees like 98 or 100 and moving toward lower degrees, and iii) edge degrees—exploring the extremes (0 or 100) before exploring moderate degrees. Regardless of their starting points, participants often revisited and compared previous translations, moving back and forth between different personalization degrees. While participants largely explored edge degrees more (0 and 100), many of them eventually favored translations with moderate degrees (Figure 6a). The mean value of participants' favorite translations was 53.14 (SD=33.24, $Median$=51.0), indicating an average tendency towards moderate personalization rather than extremes such as hyper-personalization or none at all. However, the high SD shows that participants prefer translations across a wide range of degrees, rather than clustering around a central tendency. This supports our original motivation that singular translations of science articles may not be enough to capture diverse audiences.

6.1.2 Motivations for Reading Scientific Articles: Extrinsic and Intrinsic. We categorize participants' motivation for scientific articles as extrinsic and intrinsic based on their pre-study interview responses. Those with extrinsic motivations (N=7) read scientific articles primarily when required for work-related purposes, while intrinsically motivated individuals (N=8) read them voluntarily, even in their spare time and regardless of work relevance. Figure 6b illustrates the distribution of participants' preferred translation degrees within each motivation category. Descriptively, we see that participants with extrinsic motivation tend to prefer a lower degree of personalization ($Mean$=40.52, SD=29.84), while those with intrinsic motivation favor higher degrees ($Mean$=62.75, SD=29.35). While our study was not designed to compare statistical differences between these

two groups [73], this descriptive result[4] shows that motivation could be an important factor in designing personalization interfaces.

6.2 Utility of AI-Personalized Translation for Scientific Text (RQ1)

All participants reported the benefits of personalized translations for scientific articles. However, individuals preferred varying degrees of personalization for different reasons. We asked participants to explain their preferences and rationale with concrete examples they had explored. Based on these examples, we categorized their responses into two groups: those preferring higher degrees of personalization (ranging from 51 to 100) and those preferring lower degrees (ranging from 0 to 45).

6.2.1 Participants Who Preferred Higher Degree Liked Relatable and Informative Translations. Participants who preferred translations with higher degrees of personalization appreciated that these highly relatable explanations facilitated their understanding of complex ideas (*e.g.,* P7, P8, P9, P10, P14). They further stated that these personalized translations, which extensively incorporated participants' personal contexts, resonated with them and were funny. For instance, P10 appreciated the translation began with 'running', one of their hobbies, making challenging scientific concepts more relatable and easier to understand. Similarly, P14 found some personalizations funny as they incorporated the safety and security aspects of their hometown and cultural background (see translation of P14 in Figure 7).

> P10: I like that it started with 'running', it brings it up here [in the translation]. But then it continued on the science [content]. Yeah, the science to me is usually the harder part. So I like that. It explained it [science] all at once, and then it started trying to simplify and use analogies.

Furthermore, these participants felt that highly personalized translations offered detailed scientific content, including additional findings (P8) and detailed explanations of implications (P14).

> P14: I feel like this last point [implication] was pretty good. Because it says, I'm not sure if the other ones really talked about it that much, but that the cleaner energy could be causing pollution. So the implication of the study is kind of good [useful].

These participants expressed dissatisfaction with translations of lower degrees. They felt that less personalized translations reduced the depth of scientific information (*e.g.,* fewer science concepts and facts addressed in the original article). They also noted that lower degrees of personalization often employ analogies drawn from contexts that were intended to be more universally familiar to general audiences. However, these analogies turned out to be unfamiliar to many participants, ironically making it more challenging to comprehend the scientific articles (*e.g.,* P9).

6.2.2 Participants Who Preferred Lower Degree Liked Concise Translations and Subtle Personalization. Some participants (*e.g.,* P1, P2, P4, P5) preferred translations with lower degrees of personalization primarily because they demanded less mental load to comprehend. They emphasized the cognitive load required to process long personalized translations, stating that visibly longer translations immediately diminished their interest and willingness to read the material (*e.g.,* P2, P5, P13). For instance, translations with lower degrees of personalization tend to be more concise (Figure 5), which likely contributed to their reduced mental effort.

> P5: I feel like I have to put in more effort to understand the analogy in order to understand the actual content. ... it takes a long time, I think for me the length matters. I want it to be personalized, but I don't want it lengthy.

---

[4]We emphasize that our quantitative analysis is limited by the small sample size and participant homogeneity regarding education levels. Hence, it should be treated as preliminary. We encourage a robust analysis before such a system is deployed.

	PID	Degree	Translation
Liked	P5	21 (Low)	While we aim to reduce carbon emissions by using LiBs, we might unintentionally introduce these resilient chemicals into our water and soil. Think of it as trading one problem for another, like **solving traffic jams by building more roads, only to end up with more pollution**. It underscores the need for better recycling and treatment solutions for these chemicals, helping ensure that our efforts to go green don't backfire.
			→ **Reason:** subtle analogy, concise, and clear
	P8	100 (High)	Think of it like noticing that broken **bat** pieces are scattered all over the field and beyond. Researchers found bis-FASIs in the environment, checked how toxic they are to tiny water creatures similar to minnows, and tested ways to clean them up. But the study also raises concerns: **while we're reducing CO2 emissions to help the planet, we might be unintentionally introducing new persistent pollutants**. This is like trying to win the game but cheating, which could ruin the sport itself.
			→ **Reason:** Liked the baseball analogy and informative
	P9	73 (High)	Imagine **baking a cake with a special ingredient** that makes it taste wonderful, but over time, it starts leaking something harmful into the environment. Lithium-ion batteries, like that ingredient, are fantastic for powering our gadgets and cars, but they use a group of chemicals called PFAS, which can be unsafe for the environment if they're not handled properly. Just like how you need to make sure your cake ingredients are safe and stored correctly, we need to be careful about how these PFAS chemicals are used, recycled, and disposed of in lithium-ion batteries.
			→ **Reason:** Liked the baking analogy
	P14	100 (High)	Imagine basketball-friendly streets in **Berkeley**. **Imagine there are always two security guards on watch** who ensure that no one interrupts your game. Now, suppose one of those guards gets distracted by something shiny (let's call this shiny thing MGO), leaving the field unprotected momentarily. Smash! Someone throws a rock, ruining your perfect game. This is like what happens in cells with the protein BRCA2. Usually, both guards (or both copies of BRCA2) have to be taken out for cancer to intrude.
			→ **Reason:** Hilarious due to accurate descriptions about Berkeley area
Disliked	P5	100 (High)	If you love **browsing social media** and keeping up with your favorite Italian food hashtags, the hidden mess in our environment can sneak into the water or surroundings where our food is grown or produced. Think of it like finding out that your favorite pizza place might toss the leftover dough down the street—eventually, it could end up covering a beautiful garden. Scientists are making sure these battery chemicals don't end up like that dough, messing up things we care about. They're working to find out how to manage and clean up these tiny chemicals so we can keep powering our devices without harming our planet.",
			→ **Reason:** Not relatable profile and the convoluted translation
	P9	20 (Low)	Imagine there's a security guard named BRCA2 who is very good at protecting the important blueprints of a building. Normally, for something bad to happen in the building, the security guard would need to be taken out of action twice (**like a double knockout in a boxing match**). However, there's a sneaky chemical called MGO that can temporarily distract the guard, making it easier for the building to suffer some damage without the guard being knocked out completely.
			→ **Reason:** Not relatable analogy (boxing)
	P10	75 (High)	Imagine you are **prepping sushi** in your kitchen. You've got all the ingredients ready, and you start rolling. But what if some of the ingredients are harmful? Lithium-ion batteries (like the ones in your phone or electric car) need powerful materials to keep them charged. These materials, including a special kind called PFAS, are amazing at their job but can be really bad for the environment. Just like **adding a harmful substance into your sushi would make it unsafe to eat**, using PFAS in batteries means they can end up polluting our water when batteries are made, used, or thrown away.
			→ **Reason:** Analogy deviates from core idea

Fig. 7. Several examples of preferred translations. While each personalized translation is longer (two paragraphs), we present specific parts that participants pointed out to highlight what they liked or disliked.

> P4: What I liked about less personal ones was that they felt more direct, versus this [high degree personalization] was like reading a lot of text to get to the main point. Yeah, the personalization makes the actual message longer.

They also appreciated that the analogies were subtly integrated into lower degrees, making them easy to understand and requiring minimal cognitive effort.

> P5: If I'm ever reading a scientific article, the goal is not to read a lot, but [to figure out] mostly in the shortest amount of time, what is the most information I can take away. So I think having more subtle analogies allowed me to get to the understanding better, but still have the key takeaways.

6.2.3 Experimenting with Profile Edits. Despite acknowledging the benefits of personalized translations, regardless of their preferred degrees, participants reported that the translations were sometimes unnatural and difficult to understand. Specifically, they found that certain aspects of their background (personalization profile) were not appropriate to explain the given scientific

content. Profile editing enabled them to address this issue by experimenting with different profiles. Participants enjoyed experimenting with their profiles to explore various personalization options. They found it entertaining and fun as they could tell that profile modifications directly influenced the translations. For example, P2 pretended to be a five-year-old Chinese boy, and P3 added "going to a bar" as a hobby to receive bar-related analogies.

> P3: Bar one. I picked the right analogy... And I didn't mean to. I was just like having fun with it. But that's a really good analogy, for PFAS [Per- and Polyfluoroalkyl Substances] I would say.

Participants noted that hobby-related information influenced the analogies used in translations, while age settings impacted the overall tone.

> P6: Oh, I thought it was pretty good. I noticed, specifically, when I changed the profile I liked seeing that drastic differentiation. As you mentioned, age was potentially a factor. I'm glad to see that [age] was taken into consideration in the message tone.

6.3 Impact of Interactive Reading Features on User Experience (RQ2)

We analyze participants' responses to two tool features: generating multiple translations per article and slider interaction to control the personalization degree.

6.3.1 Compounding Understanding with Multiple Translations. Most participants reported several benefits of reading multiple translations, describing it as assembling a complete picture from partial pieces. They also mentioned that reading multiple translations enabled a cumulative understanding of complex scientific concepts iteratively (*e.g.,* P1, P5, P7, P8, P10) and even helped correct initial misunderstandings (*e.g.,* P2). For example,

> P10: As I read more [translations], I understand the topic more. It was helpful as I kept going. Because I feel like I was picking up some small details here and there, and it pieced it together at the end. It definitely wasn't distracting, like completing a big picture like by collecting more positive pieces.

Furthermore, through an iterative process of generating and reading multiple translations, P2 discovered inconsistencies in their initial understanding. This process helped them identify and correct conceptual misunderstandings that they had not initially recognized, ultimately leading to a more accurate comprehension of the material.

> P2: I thought the sugar was BRCA2 [BReast CAncer gene 2]. It's actually not. [I noticed from the later translations that] the protein is the BRCA2 and the sugar bypasses the BRCA2.

However, despite such benefits, a few participants (*e.g.,* P2, P3) expressed reluctance to explore multiple translations in realistic scenarios, especially for reading multiple versions of scientific articles that did not align with their interests.

> P3: In real life, I wouldn't want to read the same thing in a different analogy. If the 1st one didn't work, I'd be like, 'Okay, whatever' you know, I wouldn't rerun it. I don't have enough time. However, I've read different political articles this year, especially with an election year. If there's something I don't understand, I'll go find information elsewhere. In the same way, I would potentially rerun this to see a different analogy or different versions [if I'm interested in the topic].

This shows that articles that deeply interest participants are a strong motivator for them to read multiple translations and gain a better understanding. While *TranSlider* could offer valuable

benefits for comprehension, user engagement may vary depending on individual interests and the perceived relevance of the content.

6.3.2 Balancing Control vs. Automation Afforded by the Slider. Participants generally recognized and appreciated the gradient differences in personalized translations across the 0-100 spectrum on the slider control. All participants found the slider interaction to be simple, straightforward, and enjoyable, requiring minimal effort to modify the translations of scientific articles.

> P15: I think that it gave me a lot of control, and I like the control [the agency] that it gave me. I'm selecting what I want to see, when the model is doing [translation] for you. I think that the more you can interact with it [through slider] and kind of control, it makes you feel welcome.

While participants agreed that the slider was simple and engaging, their desired level of control differed. Some participants expressed interest in having more granular control over the translations (*e.g.,* P3, P7, P8, P9, P10), suggesting that the slider alone might not be sufficient. For example, they wanted to manually type specific numbers to modify the degrees of personalization in addition to slider interaction (*e.g.,* P9) or prompt interactions to specify more or less personalization.

> P10: It would be nice if there were an area of more personalization where I tell them what I like about it, and then they do that [personalization] more or do it less. Like, 'I would like you [the AI model] to specify less, or I would like you to specify more on either analogy or on the science of it.'

In particular, P10 proposed a feature to allow users to highlight specific parts of the personalized translations and modify them instead of changing the entire translation (*e.g.,* selective editing features [57]). They envisioned that this feature could enable users to request more detailed elaboration or varying degrees of personalization for each selected part, providing more granular control over how the tool interprets and modifies different text parts.

In contrast, others preferred minimal control or even automated translations without any user intervention. Several participants (*e.g.,* P2, P13) found the slider feature overwhelming as it provides too many options (*i.e.,* 100 different versions in the range 0 to 100). For example, P13 suggested 10 or even fewer options for personalization rather than a scale of 0-100. Some proposed having just two versions of the article: original and translated, with an on/off toggle.

Furthermore, some participants (*e.g.,* P2) suggested a completely automated personalization system. Rather than steering various personalization degrees, they wanted the AI to automatically deliver optimally personalized translations based on user preferences and article topics. Further, the perceived benefits from the system outweighted privacy concerns.

> P2: Maybe that [the optimal degree of personalization] should be already part of your profile – how much personalization do you want. Then it's like, I just want the optimal translation. It's easier for someone to go on a blog and just have it [the optimally personalized translation] straight up there.

6.4 Potential Use Cases and Concerns

Participants identified several potential use cases and some concerns. Below, we highlight two use cases: learning unfamiliar content and tailoring their ideas to different audiences. We then present two concerns: the reliability of malleable translations and reduced access to the source article.

6.4.1 Use Case I: Learning Unfamiliar Content for Various User Motivations. Participants envisioned using this type of tool to learn unfamiliar topics that they were interested in [intrinsic motivation] such as technology (P1), political pledges (P3), economic content (P10), and how-to documents

(P15). Particularly, P10 envisioned using this tool for expanding their knowledge boundaries by learning about unfamiliar or new domains.

> P10: I tend to dodge articles about Econ and business. They are not my strong suit. I feel like I would be much more willing to read those [through TranSlider]. Because usually they're filled with so much jargon, so like it now opens up a field for me of something that I'm interested in learning.

Furthermore, P3 and P9 envisioned using this tool to learn something about an unfamiliar topic (*e.g.,* airline compensation policy documents (P3)), which they are not intrinsically motivated to learn, but need to know about [extrinsic motivation].

> P9: Maybe when I'm trying to find an article that I'm not actually interested in, but I need to get specific information about the topic that I'm not too familiar with where I need the translation. That's where I would see myself using something like this.

Overall, participants viewed *TranSlider* as a learning tool to explore various unfamiliar topics and new domains. Moreover, for topics they found less interesting, the tool could help them quickly grasp essential concepts and key takeaways.

6.4.2   Use Case II: Tailoring Ideas to Communicate to Different Audiences. Participants (*e.g.,* P1) also envisioned using this tool to learn how to communicate complex ideas to diverse audiences. For example, the different degrees of personalization could provide tailored translations of how to explain technical products or business ideas to general audiences. P1 envisioned saving multiple translations to their personal library and using them during business meetings.

> P1: If I need to explain this topic to somebody who's in the industry is very different from how I will explain this to somebody who's not technical. So it's nice to know how I should explain this to somebody who has some knowledge and how to explain it to somebody who has no knowledge like the board of directors for their company.

6.4.3   Concern I: Reliability of Malleable Translations from AI. While most participants focused more on the benefits, one participant (P8) raised concerns about the reliability of AI-generated personalized translations. They argued that personalized translations can vary with each interaction, raising questions about their consistency and trustworthiness.

> P8: If I read something and then have to apply that information in my work, it's a little more concerning. I might not trust it because it keeps changing. If I had to cite it, I probably wouldn't trust this. It [AI] might misunderstand something.

They further stated that this concern would motivate them to check out the original paper for a complete and accurate understanding. While they would not fully rely on the translations in their professional context, they acknowledged that the tool would be useful in deciding whether to read the original document.

6.4.4   Concern II: Reduced Reference to Original Source Material. One participant (P5) raised a contrasting concern: she worried that AI-personalized translations might discourage people from referring to original documents. People might rely on these personalized translations rather than engaging with the original material because personalized translations are easier to digest. This could lead to a dependence on AI-processed content at the expense of direct engagement with source materials.

> P5: The downside is that nobody would probably read the full paper outside of people within that industry or field. But I don't think that's necessarily bad, because the alternative is you have more people understanding key concepts.

P5 summarized the trade-off: the tool could broaden superficial knowledge access across wider ranging topics but hinder deep understanding and further exploration. Conversely, P5 weighed the risk of shallow learning against the benefit of increased accessibility, noting that without the tool, fewer individuals would gain any exposure to such knowledge.

7 Discussion

7.1 Implications for Science Communication and the CSCW Community

Our study demonstrated the feasibility of addressing the *translation* challenge in science communication. Rather than pursuing a single 'optimal' translation, *TranSlider* generates and enables exploring multiple personalized translations. This approach produced two key benefits: i) a compounding understanding of scientific content, where each translation contributed like a puzzle piece to build a more complete picture of the underlying content, and ii) enabling multi-perspective comprehension, where varied viewpoints conveyed different aspects of the content.

7.1.1 Towards Collective Understanding of Science with General Audiences. Our approach could address the *dissemination* issue in science communication: scientists' reluctance to disseminate their work on digital media. Their concerns stem from the potential misinformation arising during translation, which could make audiences misinterpret simplified explanations [67, 93]. By enabling readers to explore multiple personalized translations, the *TranSlider* might create an opportunity for them to build a comprehensive understanding while potentially reducing the risk of misinformation. This approach aligns with established science communication practices, where journalists often craft multiple re-expressions when content might be unfamiliar to general audiences [44].

We envision collective knowledge sharing spaces inspired by existing AI-generated content platforms, such as Midjourney [3] and DeviantArt [1]. The Midjourney system shares users' generated images throughout its community by default, allowing individuals to inspire each other to create new and unique creative works along with constructive discussions within the community [39]. Applying similar dynamics to science communication, platforms like Google Scholar [2] or Semantic Scholar [4] could integrate our approach to let general audiences explore scientific articles and share personalized translations across their community spaces (*e.g.,* OtherTube [15]). Creating such environments could foster diverse perspectives and collaborative comprehension of complex scientific materials, and further the larger vision of making science accessible at scale.

Such collective community spaces could also create valuable opportunities for scientists to observe how general audiences digest and interact with their research. Traditionally, scientists have limited visibility into the public perception and potential implications of their work, largely due to a lack of active engagement with broader audiences [92]. However, within this community setting, researchers could preview how individual and collective understandings develop around their work. These insights could influence or engender new research directions. Additionally, such feedback might help scientists identify and even prevent potentially problematic research activities that they would not otherwise anticipate without such collective reflection through public interpretation [65].

However, designers should be aware that malicious actors could exploit such community ecosystems to manipulate the collective understanding of certain topics. Furthermore, some organizations (*e.g.,* 'Big Brother' [77]) might leverage these platforms to influence science activities, as public interest often drives funding allocation, and without proper financial support, some research endeavors may not remain sustainable. We see this line of research as a promising new direction.

7.1.2 Enhancing Communication by Surfacing Diverse Perspectives. The benefit of having varied perspectives could be useful in contexts such as education and cross-disciplinary collaboration.

Education has grown increasingly complex, particularly with the widespread adoption of online learning [61]. While online platforms provide greater scalability, instructors face persistent challenges in delivering effective instruction to large, diverse student cohorts [56]. AI-powered personalization tools could help address these challenges by offering instructors tailored translation suggestions adapted to different students. For example, systems could let children input their favorite animals, foods, or objects relevant to the given learning context, and generate tailored explanations and learning materials to better engage students. Future research should explore the impact of these personalized translations on learning outcomes, such as learning performance.

Likewise, in cross-disciplinary contexts, effective communication is often challenging due to several obstacles, such as divergent terminology [43], varying data interpretations [52], and different motivations [68]. Research shows that new ideas are better understood when presented through familiar examples [52]. Cross-disciplinary collaborative activities (*e.g.,* drafting documents, building systems, and shaping policy letters) require clear and constant communication among collaborators [85, 95]. While real-time conversations offer opportunities to address misunderstandings immediately, scheduling synchronous meetings among collaborators inevitably creates communication delays that can slow project progress. Our interface could benefit such asynchronous communication scenarios by allowing collaborators to generate personalized translations of complex content, helping team members from different disciplinary backgrounds understand shared documents without waiting for clarification meetings.

However, these approaches may lead to over-reliance on AI-personalization, as indicated by a participant (P5). In a future where AI tools work exceptionally well at addressing communication barriers, individuals might prefer interacting with these interfaces rather than directly communicating with students, instructors, or collaborators. The convenience of AI-generated 'perfect' translations might even diminish individuals' efforts and motivation to develop clear communication skills and articulate their ideas in accessible ways. This raises crucial concerns about balancing between technological assistance and the development of fundamental human agency in communication capabilities, suggesting that such tools should be designed to augment rather than replace human roles. Further research is warranted on how to design and develop such AI-driven cross-disciplinary collaborative interfaces while balancing the perceived benefits, potential long-term harms, and ethical concerns.

7.2 Interface Design Enabling User Control for Human-AI Alignment

We designed an interactive interface that empowers users to steer and align AI outputs to their preferences, as opposed to optimizing the AI models directly [78]. While many participants appreciated the controllability of AI output, we found some users may not want such control in-the-wild. Participants also showed varied responses about the reliability of AI outputs.

7.2.1 Flexible AI Interfaces for Balancing User Control and Automation. Our interface provides three key features to support user agency: a) a slider that allows users to adjust the degree of personalization, b) a profile editing function, and c) a history box. These features enable users to control AI outputs without technical expertise, save personalized preferences, and iteratively explore various outputs until they find desirable results. While highly accurate AI models might better serve diverse user needs, achieving such accuracy often requires extensive user data collection, raising concerns of cost, privacy, and selection bias amongst annotators [86]. Our design approach provides an alternative: rather than relying on AI-powered systems to perfectly model the plurality of user preferences, we empower users to actively align AI outputs to their individual preferences. This approach not only preserves user privacy but also maintains human agency in human-AI interaction, addressing critical concerns about AI overreliance in our society [21, 75].

However, we should acknowledge a potential limitation of our approach. Our interaction design assumes users want to preserve their agency over AI systems, which may not always be true depending on individuals and contexts. Indeed, a participant (P2) stated they do not want to interact with systems in-the-wild (Section 6.3.2). As our study suggests, in designing systems for mass audiences, we must recognize that a portion of users, like P2, may prefer fully automatic personalization without manual adjustments. Also, in different contexts such as entertainment and creative art domains, users may prefer consuming the contents passively, rather than proactively interacting with systems [50]. Therefore, future systems should balance automation and user control, perhaps by offering flexible interfaces (*e.g.,* [72]) providing both *automatic* and *control* modes that users can switch between depending on their needs.

7.2.2 The Reliability of AI-Personalized Translations. Participants expressed contrasting concerns about the reliability of AI-generated translations for science communication. While some participants were reluctant to rely on the translations, others were worried that people might depend too heavily on them. As AI technology becomes more sophisticated, the higher accuracy of translations with fewer instances of incorrect or hallucinated content might mitigate such reliability concerns [94]. However, regardless of translation quality, concerns about AI-processed output and the trustworthiness of AI may persist. The fact that AI-generated translations are not human-authored continues to affect the perceived trust [83]. The perceived mistrust could stem from an ambiguity in responsibility, specifically regarding the onus for mistranslations and their resultant ramifications. For instance, a participant mentioned that they would not trust AI outputs for work-related tasks without verifying the source, as they are personally accountable for any errors in the work.

To mitigate risks, verifying translations against original source materials could be an approach. However, who should bear the responsibility: readers or authors? Putting the onus on the reader defeats the purpose of personalization for making science accessible as original scientific articles use overly complex language and jargon. Our findings and prior work indicate that readers do not want to invest significant mental effort parsing complex texts, resulting in the risk of over reliance on AI-generated translations [21]. A potential solution could be for authors (*i.e.,* scientists) to verify several versions of AI translations before publishing to the general public [50]. For example, scientists could go through 5–10 different translations with varying degrees of personalization and approve a subset as verified translations. The balance between utilizing AI translations and encouraging verification of original sources presents a unique challenge, particularly given the current state of scientific communication. Further explorations are needed to find scalable approaches such that the author verification does not become cumbersome.

7.3 Limitations and Future Work

Our study had several limitations. First, it was conducted in a small controlled setting. Specifically, a pool of 15 participants were asked to explore two pre-selected science articles from health and environment domains using our tool. Future studies should consider a large-scale in-the-wild study where participants could choose from a wider range of scientific articles. Secondly, our participants might not fully represent the general audience with varying degrees of formal education. More research is needed by studying participants with diverse educational levels (*e.g.,* high school graduates or those with lower educational attainment), to better understand how people with lesser exposure to formal education could draw benefit from our tool. Similarly, our participant demographics were limited, with most participants (12/15) in their twenties and recent college graduates. Only three participants were from other age groups. This age-homogeneous participant pool may have influenced the findings of our study. Future studies should include participants from diverse backgrounds in terms of age, education level, and culture to capture a broader range of perspectives

on AI-personalized translation tools. For instance, the impact of AI-personalized translation tools in high school science learning classes [58] or children-parent learning contexts [24] could be different from our findings. Lastly, our study setup required participants to provide feedback based on brief interactions with the tool. The long-term effects of interactive reading experiences with AI-personalized translation tools may differ from our findings. Future longitudinal studies are warranted to provide a more comprehensive understanding of how users interact with and benefit from these types of tools over extended periods.

8 Conclusion

Our study investigated the potential of AI-personalized translations to enhance science communication in digital media platforms. We developed *TranSlider*, an AI-powered reading interface that generates personalized translations of scientific text for general audiences. The tool's key feature involves an interactive slider that allows users to adjust the degree of personalization to generate tailored translations. Through an exploratory study with 15 participants, we found that participants preferred varying degrees of personalization for the tailored translations, from deeply relatable and contextualized translations to short and slightly contextualized ones. Interestingly, reading multiple translations enabled them to have a compounding understanding of scientific content. We present several implications for science communication and for supporting communication across individuals with diverse backgrounds in collaborative contexts. We also discuss designing flexible interfaces that allow users to preserve control over AI tools to steer desired outcomes. We hope our study opens a new avenue to promote science communication for general audiences.

Acknowledgments

References

[1] 2025. DeviantArt. https://www.deviantart.com/
[2] 2025. Google Scholar. https://scholar.google.com/
[3] 2025. Midjourney. https://www.midjourney.com/home
[4] 2025. Semantic Scholar. https://www.semanticscholar.org/
[5] Eytan Adar, Carolyn Gearig, Ayshwarya Balasubramanian, and Jessica Hullman. 2017. PersaLog: Personalization of News Article Content. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3188–3200. doi:10.1145/3025453.3025631
[6] Sarabjot Singh Anand and Bamshad Mobasher. 2007. Introduction to intelligent techniques for Web personalization. *ACM Trans. Internet Technol.* 7, 4 (Oct. 2007), 18–es. doi:10.1145/1278366.1278367
[7] Anthropic. 2024. Claude 3.5 Sonnet. https://claude.ai/
[8] Akari Asai, Jacqueline He*, Rulin Shao*, Weijia Shi, Amanpreet Singh, Joseph Chee Chang, Kyle Lo, Luca Soldaini, Sergey Feldman, Tian, D'arcy Mike, David Wadden, Matt Latzke, Minyang, Pan Ji, Shengyan Liu, Hao Tong, Bohao Wu, Yanyu Xiong, Luke Zettlemoyer, Dan Weld, Graham Neubig, Doug Downey, Wen-tau Yih, Pang Wei Koh, and Hannaneh Hajishirzi. 2024. OpenScholar: Synthesizing Scientific Literature with Retrieval-Augmented Language Models. *Arxiv* (2024).
[9] Tal August, Lauren Kim, Katharina Reinecke, and Noah A. Smith. 2020. Writing Strategies for Science Communication: Data and Computational Analysis. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu (Eds.). Association for Computational Linguistics, Online, 5327–5344. doi:10.18653/v1/2020.emnlp-main.429

[10] Tal August, Kyle Lo, Noah A. Smith, and Katharina Reinecke. 2024. Know Your Audience: The benefits and pitfalls of generating plain language summaries beyond the "general" audience. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 14, 26 pages. doi:10.1145/3613904.3642289

[11] Tal August, Lucy Lu Wang, Jonathan Bragg, Marti A. Hearst, Andrew Head, and Kyle Lo. 2023. Paper Plain: Making Medical Research Papers Approachable to Healthcare Consumers with Natural Language Processing. *ACM Trans. Comput.-Hum. Interact.* 30, 5, Article 74 (sep 2023), 38 pages. doi:10.1145/3589955

[12] Calvin Bao, Yow-Ting Shiue, Marine Carpuat, and Joel Chan. 2025. Words as Bridges: Exploring Computational Support for Cross-Disciplinary Translation Work. In *Proceedings of the 30th International Conference on Intelligent User Interfaces (IUI '25)*. Association for Computing Machinery, New York, NY, USA, 1598–1623. doi:10.1145/3708359.3712110

[13] Larry Bell. 2008. Engaging the Public in Technology Policy: A New Role for Science Museums. *Science Communication* 29, 3 (2008), 386–398. doi:10.1177/1075547007311971

[14] John C. Besley, Anthony Dudo, and Martin Storksdieck. 2015. Scientists' views about communication training. *Journal of Research in Science Teaching* 52, 2 (2015), 199–220. doi:10.1002/tea.21186

[15] Md Momen Bhuiyan, Carlos Augusto Bautista Isaza, Tanushree Mitra, and Sang Won Lee. 2022. OtherTube: Facilitating Content Discovery and Reflection by Exchanging YouTube Recommendations with Strangers. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 204, 17 pages. doi:10.1145/3491102.3502028

[16] Sid Black, Leo Gao, Phil Wang, Connor Leahy, and Stella Biderman. 2021. *GPT-Neo: Large Scale Autoregressive Language Modeling with Mesh-Tensorflow*. doi:10.5281/zenodo.5297715 If you use this software, please cite it using these metadata..

[17] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. doi:10.1191/1478088706qp063oa

[18] Wayne Breslyn and Amy E. Green. 2022. Learning science with YouTube videos and the impacts of Covid-19. *Disciplinary and Interdisciplinary Science Education Research* 4, 1 (2022), 13. doi:10.1186/s43031-022-00051-4

[19] Michael Brüggemann, Ines Lörcher, and Stefanie Walter. 2020. Post-normal science communication: exploring the blurring boundaries of science and journalism. *JCOM* 19, 03 (2020), A02. doi:10.22323/2.19030202

[20] M. Bucchi and B. Trench. 2014. *Routledge Handbook of Public Communication of Science and Technology: Second edition*. Taylor & Francis. https://books.google.com/books?id=KrLcAwAAQBAJ

[21] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z. Gajos. 2021. To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 188 (April 2021), 21 pages. doi:10.1145/3449287

[22] T. W. Burns, D. J. O'Connor, and S. M. Stocklmayer. 2003. Science Communication: A Contemporary Definition. *Public Understanding of Science* 12, 2 (2003), 183–202. doi:10.1177/09636625030122004

[23] John Joon Young Chung and Max Kreminski. 2024. Patchview: LLM-powered Worldbuilding with Generative Dust and Magnet Visualization. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 77, 19 pages. doi:10.1145/3654777.3676352

[24] Drew Cingel and Anne Marie Piper. 2017. How Parents Engage Children in Tablet-Based Reading Experiences: An Exploration of Haptic Feedback. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 505–510. doi:10.1145/2998181.2998240

[25] Chad E. Cook, Neil E. O'Connell, Toby Hall, Steven Z. George, Gwendolen Jull, Alexis A. Wright, Enrique Lluch Girbés, Jeremy Lewis, and Mark Hancock. 2018. Benefits and Threats to Using Social Media for Presenting and Implementing Evidence. *Journal of Orthopaedic & Sports Physical Therapy* 48, 1 (2018), 3–7. doi:10.2519/jospt.2018.0601 PMID: 29291281.

[26] Hope B. Corrigan, Georgiana Craciun, and Allison M. Powell and. 2014. How Does Target Know So Much About Its Customers? Utilizing Customer Analytics to Make Marketing Decisions. *Marketing Education Review* 24, 2 (2014), 159–166. doi:10.2753/MER1052-8008240206

[27] Debarati Das, David Ma, and Dongyeop Kang. 2023. Balancing the Effect of Training Dataset Distribution of Multiple Styles for Multi-Style Text Transfer. In *Findings of the Association for Computational Linguistics: ACL 2023*, Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (Eds.). Association for Computational Linguistics, Toronto, Canada, 3932–3943. doi:10.18653/v1/2023.findings-acl.243

[28] Munmun De Choudhury, Meredith Ringel Morris, and Ryen W. White. 2014. Seeking and sharing health information online: comparing search engines and social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 1365–1376. doi:10.1145/2556288.2557214

[29] C Dimarco, P Bray, HD Covvey, DD Cowan, V Diciccio, E Hovy, Joan Lipa, and C Yang. 2006. Authoring and generation of individualized patient education materials. *AMIA Annual Symposium Proceedings* 2006 (2006), 195–199.

[30] Zijian Ding, Arvind Srinivasan, Stephen Macneil, and Joel Chan. 2023. Fluid Transformers and Creative Analogies: Exploring Large Language Models' Capacity for Augmenting Cross-Domain Analogical Creativity. In *Proceedings of the 15th Conference on Creativity and Cognition* (Virtual Event, USA) *(C&C '23)*. Association for Computing Machinery, New York, NY, USA, 489–505. doi:10.1145/3591196.3593516

[31] Shiran Dudy, Steven Bedrick, and Bonnie Webber. 2021. Refocusing on Relevance: Personalization in NLG. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih (Eds.). Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 5190–5202. doi:10.18653/v1/2021.emnlp-main.421

[32] Baruch Fischhoff. 2013. The sciences of science communication. *Proceedings of the National Academy of Sciences* 110, supplement_3 (2013), 14033–14039. doi:10.1073/pnas.1213273110

[33] Raymond Fok, Hita Kambhamettu, Luca Soldaini, Jonathan Bragg, Kyle Lo, Marti Hearst, Andrew Head, and Daniel S Weld. 2023. Scim: Intelligent Skimming Support for Scientific Papers. In *Proceedings of the 28th International Conference on Intelligent User Interfaces* (Sydney, NSW, Australia) *(IUI '23)*. Association for Computing Machinery, New York, NY, USA, 476–490. doi:10.1145/3581641.3584034

[34] Andrew Gelman and John Carlin. 2014. Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. *Perspectives on Psychological Science* 9, 6 (Nov. 2014), 641–651. doi:10.1177/1745691614551642 PMID: 26186114.

[35] Katy Ilonka Gero, Vivian Liu, Sarah Huang, Jennifer Lee, and Lydia B. Chilton. 2021. What Makes Tweetorials Tick: How Experts Communicate Complex Topics on Twitter. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 422 (Oct. 2021), 26 pages. doi:10.1145/3479566

[36] James E. Grunig. 1979. Research on Science Communication: What is Known and What Needs To Be Known. *Journal of Applied Communications* 62, 4 (1979). doi:10.4148/1051-0834.1882

[37] Ziwei Gu, Ian Arawjo, Kenneth Li, Jonathan K. Kummerfeld, and Elena L. Glassman. 2024. An AI-Resilient Text Rendering Technique for Reading and Skimming Documents. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 898, 22 pages. doi:10.1145/3613904.3642699

[38] Jennifer L. Guelfo, P. Lee Ferguson, Jonathan Beck, Melissa Chernick, Alonso Doria-Manzur, Patrick W. Faught, Thomas Flug, Evan P. Gray, Nishad Jayasundara, Detlef R. U. Knappe, Abigail S. Joyce, Pingping Meng, and Marzieh Shojaei. 2024. Lithium-ion battery components are at the nexus of sustainable energy and environmental release of per- and polyfluoroalkyl substances. *Nature Communications* 15, 1 (2024), 5548. doi:10.1038/s41467-024-49753-5

[39] Qingyu Guo, Kangyu Yuan, Changyang He, Zhenhui Peng, and Xiaojuan Ma. 2024. Exploring the Evolvement of Artwork Descriptions in Online Creative Community under the Surge of Generative AI: A Case Study of DeviantArt. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 151, 7 pages. doi:10.1145/3613905.3650851

[40] Andrew Head, Kyle Lo, Dongyeop Kang, Raymond Fok, Sam Skjonsberg, Daniel S. Weld, and Marti A. Hearst. 2021. Augmenting Scientific Papers with Just-in-Time, Position-Sensitive Definitions of Terms and Symbols. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 413, 18 pages. doi:10.1145/3411764.3445648

[41] Alison J. Head, Michele Van Hoeck, and Kirsten Hostetler. 2017. Why blogs endure: A study of recent college graduates and motivations for blog readership. *First Monday* 22, 10 (Oct. 2017). doi:10.5210/fm.v22i10.8065

[42] Chris Hogan. 2024. Science Daily. https://www.sciencedaily.com/

[43] Youyang Hou and Dakuo Wang. 2017. Hacking with NPOs: Collaborative Analytics and Broker Roles in Civic Data Hackathons. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 53 (Dec. 2017), 16 pages. doi:10.1145/3134688

[44] Jessica Hullman, Yea-Seul Kim, Francis Nguyen, Lauren Speers, and Maneesh Agrawala. 2018. Improving Comprehension of Measurements Using Concrete Re-expression Strategies. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3173574.3173608

[45] Maia Jacobs, James Clawson, and Elizabeth D. Mynatt. 2014. Cancer navigation: opportunities and challenges for facilitating the breast cancer journey. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Baltimore, Maryland, USA) *(CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 1467–1478. doi:10.1145/2531602.2531645

[46] Paige Brown Jarreau and Lance Porter. 2018. Science in the Social Media Age: Profiles of Science Blog Readers. *Journalism & Mass Communication Quarterly* 95, 1 (2018), 142–168. doi:10.1177/1077699016685558

[47] Shagun Jhaver, Alice Qian Zhang, Quan Ze Chen, Nikhila Natarajan, Ruotong Wang, and Amy X. Zhang. 2023. Personalizing Content Moderation on Social Media: User Perspectives on Moderation Choices, Interface Design, and Labor. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2, Article 289 (Oct. 2023), 33 pages. doi:10.1145/3610080

[48] Ridley Jones, Lucas Colusso, Katharina Reinecke, and Gary Hsieh. 2019. r/science: Challenges and Opportunities in Online Science Communication. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3290605.3300383

[49] Hita Kambhamettu, Danaë Metaxa, Kevin Johnson, and Andrew Head. 2024. Explainable Notes: Examining How to Unlock Meaning in Medical Notes with Interactivity and Artificial Intelligence. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 449, 19 pages. doi:10.1145/3613904.3642573

[50] Taewook Kim, Hyomin Han, Eytan Adar, Matthew Kay, and John Joon Young Chung. 2024. Authors' Values and Attitudes Towards AI-bridged Scalable Personalization of Creative Language Arts. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 31, 16 pages. doi:10.1145/3613904.3642529

[51] Taewook Kim, Jung Soo Lee, Zhenhui Peng, and Xiaojuan Ma. 2019. Love in Lyrics: An Exploration of Supporting Textual Manifestation of Affection in Social Messaging. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 79 (nov 2019), 27 pages. doi:10.1145/3359181

[52] Yea-Seul Kim, Jake M Hofman, and Daniel G Goldstein. 2022. Putting scientific results in perspective: Improving the communication of standardized effect sizes. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 625, 14 pages. doi:10.1145/3491102.3502053

[53] Yea-Seul Kim, Jessica Hullman, and Eytan Adar. 2015. DeScipher: A Text Simplification Tool for Science Journalism. In *Computation+Journalism Symposium*.

[54] Yea-Seul Kim, Jessica Hullman, and Maneesh Agrawala. 2016. Generating Personalized Spatial Analogies for Distances and Areas. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 38–48. doi:10.1145/2858036.2858440

[55] Li Ren Kong, Komal Gupta, Andy Jialun Wu, David Perera, Roland Ivanyi-Nagy, Syed Moiz Ahmed, Tuan Zea Tan, Shawn Lu-Wen Tan, Alessandra Fuddin, Elayanambi Sundaramoorthy, Grace Shiqing Goh, Regina Tong Xin Wong, Ana S.H. Costa, Callum Oddy, Hannan Wong, C. Pawan K. Patro, Yun Suen Kho, Xiao Zi Huang, Joan Choo, Mona Shehata, Soo Chin Lee, Boon Cher Goh, Christian Frezza, Jason J. Pitt, and Ashok R. Venkitaraman. 2024. A glycolytic metabolite bypasses "two-hit" tumor suppression by BRCA2. *Cell* 187, 9 (2024), 2269–2287.e16. doi:10.1016/j.cell.2024.03.006

[56] Chinmay Kulkarni, Julia Cambre, Yasmine Kotturi, Michael S. Bernstein, and Scott R. Klemmer. 2015. Talkabout: Making Distance Matter with Small Groups in Massive Classes. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 1116–1128. doi:10.1145/2675133.2675166

[57] Philippe Laban, Jesse Vig, Marti Hearst, Caiming Xiong, and Chien-Sheng Wu. 2024. Beyond the Chat: Executable and Verifiable Text-Editing with LLMs. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 20, 23 pages. doi:10.1145/3654777.3676419

[58] Irene Lee and Beatriz Perret. 2022. Preparing High School Teachers to Integrate AI Methods into STEM Classrooms. *Proceedings of the AAAI Conference on Artificial Intelligence* 36, 11 (Jun. 2022), 12783–12791. doi:10.1609/aaai.v36i11.21557

[59] A. Leerink, M. Bos, D. Reijnders, and E. van Sebille. 2024. The (non)effect of personalization in climate texts on the credibility of climate scientists: a case study on sustainable travel. *Geoscience Communication* 7, 3 (2024), 201–214. doi:10.5194/gc-7-201-2024

[60] Anna Lenhart, Sunyup Park, Michael Zimmer, and Jessica Vitak. 2023. "You Shouldn't Need to Share Your Data": Perceived Privacy Risks and Mitigation Strategies Among Privacy-Conscious Smart Home Power Users. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2, Article 247 (Oct. 2023), 34 pages. doi:10.1145/3610038

[61] Tiffany Wenting Li, Karrie Karahalios, and Hari Sundaram. 2021. "It's all about conversation": Challenges and Concerns of Faculty and Students in the Arts, Humanities, and the Social Sciences about Education at Scale. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW3, Article 216 (Jan. 2021), 37 pages. doi:10.1145/3432915

[62] Yusheng Li, Lijiang Shen, James Price Dillard, and Siyoung Sue Li. 2024. A Content Analysis of Online Messages about Sugar-Sweetened Beverages. *Nutrients* 16, 7 (29 Mar 2024), 1005. doi:10.3390/nu16071005 PMID: 38613038; PMCID: PMC11013644.

[63] Soo Lim, So-Youn Kim, Jung Im Kim, Min Kyung Kwon, Sei Jin Min, Soo Young Yoo, Seon Mee Kang, Hong Il Kim, Hye Seung Jung, Kyong Soo Park, Jun Oh Ryu, Hayley Shin, and Hak Chul Jang. 2011. A Survey on Ubiquitous Healthcare Service Demand among Diabetic Patients. *Diabetes & Metabolism Journal* 35, 1 (Feb. 2011), 50–57. doi:10.4093/dmj.2011.35.1.50

[64] Yuhan Liu, Aadit Shah, Jordan Ackerman, and Manaswi Saha. 2025. Exploring the Design Space of Real-time LLM Knowledge Support Systems: A Case Study of Jargon Explanations. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '25)*. ACM, Yokohama, Japan, 20. doi:10.1145/3706598.3714262

[65] Carolina Llorente, Gema Revuelta, Mar Carrió, and Miquel Porta. 2019. Scientists' opinions and attitudes towards citizens' understanding of science and their role in public engagement activities. *PLOS ONE* 14 (11 2019), 1–20. doi:10.1371/journal.pone.0224262

[66] Duri Long, Takeria Blunt, and Brian Magerko. 2021. Co-Designing AI Literacy Exhibits for Informal Learning Spaces. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 293 (Oct. 2021), 35 pages. doi:10.1145/3476034

[67] Maria Loroño-Leturiondo and Sarah R. Davies. 2018. Responsibility and science communication: scientists' experiences of and perspectives on public communication activities. *Journal of Responsible Innovation* 5, 2 (2018), 170–185. doi:10.1080/23299460.2018.1434739

[68] Yaoli Mao, Dakuo Wang, Michael Muller, Kush R. Varshney, Ioana Baldini, Casey Dugan, and Aleksandra Mojsilović. 2019. How Data Scientists Work Together With Domain Experts in Scientific Collaborations: To Find The Right Answer Or To Ask The Right Question? *Proc. ACM Hum.-Comput. Interact.* 3, GROUP, Article 237 (Dec. 2019), 23 pages. doi:10.1145/3361118

[69] Curtis Martin and Bertrum H. MacDonald. 2020. Using interpersonal communication strategies to encourage science conversations on social media. *PLOS ONE* 15, 11 (11 2020), 1–32. doi:10.1371/journal.pone.0241972

[70] Susana Martinez-Conde. 2016. Has Contemporary Academia Outgrown the Carl Sagan Effect? *Journal of Neuroscience* 36, 7 (2016), 2077–2082. doi:10.1523/JNEUROSCI.0086-16.2016

[71] Meta. 2024. Meta Llama 3. https://llama.meta.com/

[72] Bryan Min, Allen Chen, Yining Cao, and Haijun Xia. 2025. Malleable Overview-Detail Interfaces. In *CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–25. doi:10.1145/3706598.3714164

[73] Jacob M. Montgomery, Brendan Nyhan, and Michelle Torres. 2018. How Conditioning on Posttreatment Variables Can Ruin Your Experiment and What to Do about It. *American Journal of Political Science* 62, 3 (2018), 760–775. doi:10.1111/ajps.12357

[74] Ha Nguyen, Victoria Nguyen, Saríah López-Fierro, Sara Ludovise, and Rossella Santagata. 2024. Simulating Climate Change Discussion with Large Language Models: Considerations for Science Communication at Scale. In *Proceedings of the Eleventh ACM Conference on Learning @ Scale* (Atlanta, GA, USA) *(L@S '24)*. Association for Computing Machinery, New York, NY, USA, 28–38. doi:10.1145/3657604.3662033

[75] Chinasa T. Okolo, Dhruv Agarwal, Nicola Dell, and Aditya Vashistha. 2024. "If it is easy to understand then it will have value": Examining Perceptions of Explainable AI with Community Health Workers in Rural India. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 71 (April 2024), 28 pages. doi:10.1145/3637348

[76] OpenAI. 2022. ChatGPT. https://openai.com/chatgpt/

[77] G. Orwell. 2024. *1984*. Modernista. https://books.google.com/books?id=ET0EEQAAQBAJ

[78] Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback. arXiv:2203.02155 [cs.CL] https://arxiv.org/abs/2203.02155

[79] Hans Peter Peters. 2013. Gap between science and media revisited: Scientists as public communicators. *Proceedings of the National Academy of Sciences* 110, supplement_3 (2013), 14102–14109. doi:10.1073/pnas.1212745110

[80] Habeeb Ibrahim Abdul Razack, Sam T. Mathew, Fathinul Fikri Ahmad Saad, and Saleh A. Alqahtani. 2021. Artificial intelligence-assisted tools for redefining the communication landscape of the scholarly world. *Science Editing* 8, 2 (2021). doi:10.6087/kcse.244

[81] Ronald E. Rice and Howard Giles. 2017. The Contexts and Dynamics of Science Communication and Language. *Journal of Language and Social Psychology* 36, 1 (2017), 127–139. doi:10.1177/0261927X16663257

[82] Mark O. Riedl. 2010. Scalable personalization of interactive experiences through creative automation. *Comput. Entertain.* 8, 4, Article 26 (dec 2010), 3 pages. doi:10.1145/1921141.1921146

[83] Mark Ryan. 2020. In AI We Trust: Ethics, Artificial Intelligence, and Reliability. *Science and Engineering Ethics* 26, 5 (2020), 2749–2767. doi:10.1007/s11948-020-00228-y

[84] Carl Sagan, Ann Druyan, and Steven Soter. 1980. Cosmos: A Personal Voyage. Television series, 13 episodes.

[85] Manaswi Saha, Devanshi Chauhan, Siddhant Patil, Rachel Kangas, Jeffrey Heer, and Jon E. Froehlich. 2021. Urban Accessibility as a Socio-Political Problem: A Multi-Stakeholder Analysis. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW3, Article 209 (Jan. 2021), 26 pages. doi:10.1145/3432908

[86] Nithya Sambasivan, Shivani Kapania, Hannah Highfill, Diana Akrong, Praveen Paritosh, and Lora M Aroyo. 2021. "Everyone wants to do the model work, not the data work": Data Cascades in High-Stakes AI. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing

Machinery, New York, NY, USA, Article 39, 15 pages. doi:10.1145/3411764.3445518

[87] Mike S. Schäfer. 2017. How Changing Media Structures Are Affecting Science News Coverage. In *The Oxford Handbook of the Science of Science Communication*. Oxford University Press. doi:10.1093/oxfordhb/9780190497620.013.5

[88] Marén Schorch, Lin Wan, David William Randall, and Volker Wulf. 2016. Designing for Those who are Overlooked: Insider Perspectives on Care Practices and Cooperative Work of Elderly Informal Caregivers. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (San Francisco, California, USA) *(CSCW '16)*. Association for Computing Machinery, New York, NY, USA, 787–799. doi:10.1145/2818048.2819999

[89] Leona Yi-Fan Su, Michael A. Cacciatore, Dietram A. Scheufele, Dominique Brossard, and Michael A. Xenos. 2014. Inequalities in Scientific Understanding: Differentiating Between Factual and Perceived Knowledge Gaps. *Science Communication* 36, 3 (2014), 352–378. doi:10.1177/1075547014529093

[90] Times. 2024. Science Times. https://www.sciencetimes.com/

[91] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou. 2022. Chain-of-thought prompting elicits reasoning in large language models. In *Proceedings of the 36th International Conference on Neural Information Processing Systems* (New Orleans, LA, USA) *(NIPS '22)*. Curran Associates Inc., Red Hook, NY, USA, Article 1800, 14 pages.

[92] Spencer Williams, Ridley Jones, Katharina Reinecke, and Gary Hsieh. 2022. An HCI Research Agenda for Online Science Communication. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 490 (Nov. 2022), 22 pages. doi:10.1145/3555591

[93] Shaomei Wu, Lindsay Reynolds, Xian Li, and Francisco Guzmán. 2019. Design and Evaluation of a Social Media Writing Support Tool for People with Dyslexia. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3290605.3300746

[94] Shaomei Wu, Jeffrey Wieland, Omid Farivar, and Julie Schiller. 2017. Automatic Alt-text: Computer-generated Image Descriptions for Blind Users on a Social Network Service. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 1180–1192. doi:10.1145/2998181.2998364

[95] Dong Whi Yoo, Hayoung Woo, Sachin R. Pendse, Nathaniel Young Lu, Michael L. Birnbaum, Gregory D. Abowd, and Munmun De Choudhury. 2024. Missed Opportunities for Human-Centered AI Research: Understanding Stakeholder Collaboration in Mental Health AI Research. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 95 (April 2024), 24 pages. doi:10.1145/3637372

[96] Richard Zhang and Emőke-Ágnes Horvát. 2025. The Unwanted Dissemination of Science: The Usage of Academic Articles as Ammunition in Contested Discursive Arenas on Twitter. *Proceedings of the ACM on Human-Computer Interaction* CSCW (2025), 16 pages.

[97] Tiancheng Zhao and Kyusong Lee. 2020. Talk to Papers: Bringing Neural Question Answering to Academic Search. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, Asli Celikyilmaz and Tsung-Hsien Wen (Eds.). Association for Computational Linguistics, Online, 30–36. doi:10.18653/v1/2020.acl-demos.5

[98] Minjun Zhu, Yixuan Weng, Linyi Yang, and Yue Zhang. 2025. Personality Alignment of Large Language Models. In *The Thirteenth International Conference on Learning Representations*. https://openreview.net/forum?id=0DZEs8NpUH

[99] Yimei Zhu and Kingsley Purdam. 2017. Social media, science communication and the academic super user in the United Kingdom. *First Monday* 22, 11 (Oct. 2017). doi:10.5210/fm.v22i11.7866

[100] Thomas P Zollo, Andrew Wei Tung Siah, Naimeng Ye, Ang Li, and Hongseok Namkoong. 2025. PersonalLLM: Tailoring LLMs to Individual Preferences. In *The Thirteenth International Conference on Learning Representations*. https://openreview.net/forum?id=2R7498e2Tx

A   Original Science Articles

A.1   Health [55]

Title: A glycolytic metabolite bypasses "two-hit" tumor suppression by BRCA2

Abstract: Knudson's "two-hit" paradigm posits that carcinogenesis requires inactivation of both copies of an autosomal tumor suppressor gene. Here, we report that the glycolytic metabolite methylglyoxal (MGO) transiently bypasses Knudson's paradigm by inactivating the breast cancer suppressor protein BRCA2 to elicit a cancer-associated, mutational single-base substitution (SBS) signature in nonmalignant mammary cells or patient-derived organoids. Germline monoallelic BRCA2 mutations predispose to these changes. An analogous SBS signature, again without biallelic BRCA2 inactivation, accompanies MGO accumulation and DNA damage in Kras-driven, Brca2-mutant murine pancreatic cancers and human breast cancers. MGO triggers BRCA2 proteolysis, temporarily disabling BRCA2's tumor suppressive functions in DNA repair and replication, causing functional haploinsufficiency. Intermittent MGO exposure incites episodic SBS mutations without permanent BRCA2 inactivation. Thus, a metabolic mechanism wherein MGO-induced BRCA2 haploinsufficiency transiently bypasses Knudson's two-hit requirement could link glycolysis activation by oncogenes, metabolic disorders, or dietary challenges to mutational signatures implicated in cancer evolution.

A.2   Environment [38]

Title: Lithium-ion battery components are at the nexus of sustainable energy and environmental release of per- and polyfluoroalkyl substances

Abstract: Lithium-ion batteries (LiBs) are used globally as a key component of clean and sustainable energy infrastructure, and emerging LiB technologies have incorporated a class of per- and polyfluoroalkyl substances (PFAS) known as bis-perfluoroalkyl sulfonimides (bis-FASIs). PFAS are recognized internationally as recalcitrant contaminants, a subset of which are known to be mobile and toxic, but little is known about environmental impacts of bis-FASIs released during LiB manufacture, use, and disposal. Here we demonstrate that environmental concentrations proximal to manufacturers, ecotoxicity, and treatability of bis-FASIs are comparable to PFAS such as perfluorooctanoic acid that are now prohibited and highly regulated worldwide, and we confirm the clean energy sector as an unrecognized and potentially growing source of international PFAS release. Results underscore that environmental impacts of clean energy infrastructure merit scrutiny to ensure that reduced CO_2 emissions are not achieved at the expense of increasing global releases of persistent organic pollutants.