

Children's Evaluation of Human, Search Engine, and Generative AI Information Sources Across Temporal Domains

Research by: Parth Parnami, Kunskapsskolan International School

Research supervised by: Steve Joordens, University of Toronto and Fatos Selita, Hult International Business School

1. Abstract

As artificial intelligence becomes increasingly accessible, understanding how children evaluate different sources of information is becoming more important. However, research examining children's judgments about information sources has comparatively limited evidence from non-Western contexts and has not extensively examined generative AI alongside established sources such as search engines and human informants. This study investigated how children evaluated a human source (a friend), Google, and ChatGPT when considering questions about the past, present, and future. Participants were 108 students aged 7–11 years from Grades 3–5 at a school in Gurugram, India. Using an online, quantitative, cross-sectional, within-participants questionnaire adapted from previous research, participants judged whether each source could correctly answer six questions and then selected the source they believed would provide the best answer. Primary responses were analyzed using a binomial generalized linear mixed-effects model with random intercepts for participants and items. Information source had a significant main effect, $\chi^2(2) = 184.62$, $p < .001$: the odds of a “Yes” response were highest for Google, followed by ChatGPT and Friend. Temporal domain also had a significant main effect, $\chi^2(2) = 35.96$, $p < .001$, with lower odds of endorsement for future-oriented than for past-oriented questions or present-oriented questions; the adjusted difference between past and present was not statistically significant. Age was not significantly associated with responses, $\chi^2(1) = 0.34$, $p = .561$. Descriptively, Google was also the most frequently selected source in the forced-choice task. These findings provide evidence of how children in an Indian school context perceive human, search engine and generative AI information sources, while highlighting the need for further cross-cultural and developmental research. The findings concern perceived capability rather than the sources' actual accuracy.

2. Introduction

Children growing up in contemporary digital environments encounter an increasingly diverse range of information sources, including parents, teachers, peers, search engines, and artificial intelligence (AI) systems (Danovitch & Alzahabi, 2013; Harris et al., 2018). These sources differ in how they acquire, store, and communicate information, yet children may consult them for

similar purposes: answering questions, learning about the world, and resolving uncertainty. As digital technologies become increasingly accessible, understanding how children evaluate and compare these sources has become an important question in developmental psychology. Access to information alone, however, does not ensure that the information obtained is accurate. Children must also judge whether a source is knowledgeable, whether it has access to relevant information, and whether it is likely to provide a useful answer. Empirical work demonstrates that children begin making these evaluative distinctions early in development when navigating text-based and digital environments (Chandler-Campbell et al., 2022; Einav et al., 2020). Examining these judgments is therefore important for understanding how children navigate an increasingly complex information environment.

Research on selective trust demonstrates that children do not accept information indiscriminately. Instead, they evaluate informants using cues such as prior accuracy, apparent knowledge, and access to relevant information. Koenig and Harris (2005), for example, found that preschool children distinguished between knowledgeable and ignorant speakers and showed greater trust in informants who had previously provided accurate information. Children also extend aspects of this selective evaluation to technological sources. Danovitch and Alzahabi (2013) demonstrated that young children relied more on a previously accurate computer than an inaccurate one when learning new information, suggesting that children can use a source's history of accuracy to guide their judgments beyond interactions with people. Similarly, studies investigating children's responses to web pages and digital text indicate that children weigh source credibility cues when deciding whether to endorse online claims (Einav et al., 2020). Furthermore, work comparing children's trust across multiple information channels shows that children readily incorporate digital tools into their epistemic networks alongside teachers and peers (Wang et al., 2019). Nevertheless, children's evaluations are not determined by accuracy alone: the type of information being sought and the characteristics of the source can also matter. These findings establish that children can evaluate different information sources selectively, while raising the question of how they judge technologies that provide access to information in different ways.

Search engines and generative AI represent two distinct forms of digital information source. Search engines such as Google direct users toward existing online information, whereas generative AI systems such as ChatGPT produce responses to prompts using generative models. Although both can be used to answer questions, they differ in how information is presented and in the role users may attribute to them. Research on children's evaluations of technological informants suggests that children can regard digital systems as capable sources of information, but that their judgments depend on the task and the source being evaluated (Danovitch & Alzahabi, 2013). For instance, research comparing children's trust in digital voice assistants and human informants has found that children's evaluations can vary according to the type of information being communicated (Girouard-Hallam & Danovitch, 2022). More recent work examining children's trust in Google has found that children are more likely to judge Google as capable of answering questions than a human informant, although these judgments vary across temporal domains and some judgments vary with age (Girouard-Hallam & Danovitch, 2025). Generative AI introduces an additional consideration because its

conversational responses may resemble communication with a human informant, even though the system does not acquire knowledge through human experience. Research on children's interactions with AI systems is beginning to examine how children perceive these technologies and form judgments about them (Van Brummelen et al., 2024; Oh et al., 2025). However, findings from studies of interaction, social perception, or trust in individual systems do not necessarily establish how children compare generative AI with a search engine and a human source when answering the same questions. Direct comparisons are therefore needed to understand how children evaluate these different sources within a common framework.

An additional factor that may shape children's judgments is the temporal domain of the information sought. By middle childhood, children increasingly engage in episodic future thinking and construct mental timelines to differentiate between past, present, and future events (Atance, 2015; Russell et al., 2010; Tillman et al., 2018; Zhang & Hudson, 2018). Questions about the past concern events that have already occurred, questions about the present concern current circumstances, and questions about the future require consideration of events that have not yet happened (Hudson & Mayhew, 2011). Empirical investigations into temporal judgments reveal that while young children can sequence basic past and future events, their ability to reason about the predictive limits of information improves progressively with age (Hudson & Mayhew, 2011; Zhang & Hudson, 2018). These domains place different demands on an information source: past questions may be answerable through stored records, present questions require current information, and future questions involve prediction under uncertainty. Children's judgments about whether a source can answer a question may consequently vary according to the temporal nature of that question. Importantly, such differences need not reflect children's actual technical knowledge of how a system works; they may instead reflect their developing beliefs about what that source can know or access (Koenig et al., 2015). Examining judgments across temporal domains therefore provides a way to investigate whether children evaluate information sources as having different capabilities depending on the type of information being sought.

A particularly relevant foundation for the present investigation is Girouard-Hallam and Danovitch's (2025) study, *Children's Trust in Google's Ability to Answer Questions About the Past, Present, and Future*. Across two studies involving 164 children aged 7–10 years, the researchers examined children's judgments of whether Google and a human informant could correctly answer questions about past, present, and future events, as well as their forced-choice preferences between the two sources. In Study 1, children were more likely to judge Google as capable of answering questions than the human informant, and endorsement of both sources was lower for future-oriented questions than for past-oriented questions or present-oriented questions. Children's judgments about current events also varied with age, with older children more likely to endorse Google's ability to answer these questions. In the forced-choice task, children selected Google more frequently than the human informant, including for future-oriented questions. Study 2 further examined past and present events and found that children continued to endorse Google more frequently than the human informant, while older children were more likely to select Google for current-event questions. Children's explanations frequently referred to Google's access to a wide range of information as a reason for choosing

it. Together, these findings indicate that children's evaluations of information sources vary with the temporal domain of a question and that they often perceive Google as having an informational advantage over a human informant. This study provides the conceptual and methodological foundation for the present investigation. However, its comparisons were limited to Google and a human informant, leaving open the question of how children evaluate generative AI alongside these sources.

Despite growing research on children's evaluations of human and technological informants, questions remain about how children compare search engines, generative AI, and human sources within the same task. In particular, children's judgments about ChatGPT cannot be assumed to mirror their judgments about Google, given the differences between retrieving existing online information and generating a conversational response. Evidence from non-Western contexts is also comparatively limited in the literature directly relevant to these comparisons. Research conducted in a particular cultural or educational setting may not capture the full range of children's experiences with technology, making it important to investigate information-source judgments across diverse contexts. An Indian school setting provides an opportunity to examine these judgments outside the contexts represented in much of the existing research. Together, these gaps motivate an investigation that includes a human informant, a search engine, and generative AI, while retaining the comparison across past, present, and future questions established by the reference study.

The present study investigated the central research question: How do children evaluate the ability of Google, ChatGPT, and a human to answer questions about the past, present, and future? Specifically, this study has two main aims: (1) to investigate children's judgments regarding the ability of Google, ChatGPT, and a human to answer questions across these temporal domains, and (2) to adapt the paradigm established by Girouard-Hallam and Danovitch (2025) in an Indian context and extend it through the inclusion of generative AI (ChatGPT). To achieve this, children in Grades 3–5 at a school in Gurugram, India, evaluated a human source (a friend), Google, and ChatGPT across past-oriented, present-oriented, and future-oriented scenarios. Following the reference design, the study examined children's judgments of whether each source could correctly answer questions and recorded their forced-choice selections of the source they believed would provide the best answer. By examining these three sources within this adapted framework, the study extends previous research on children's information-source judgments to include generative AI and contributes evidence from an Indian school context. The study concerns children's perceived capabilities of the sources rather than the sources' objective accuracy.

3. Methodology

3.1 Research Design

The study used an online, quantitative, cross-sectional, within-participants questionnaire design to investigate children's judgments about information sources. Each participant evaluated

Google, ChatGPT, and a friend of the fictional individual asking the question across scenarios concerning the past, present, and future.

The study examined two outcomes within the Informant Evaluation Task. The primary outcome was perceived source capability, measured through binary “Yes” or “No” judgments about whether each information source could correctly answer a given question. The secondary outcome was source preference, measured by a forced-choice question asking participants which source they believed would provide the better answer.

The independent variables were information source (Google, ChatGPT, and Friend) and temporal domain (Past, Present, and Future). Age was recorded and examined as a continuous predictor.

Each participant responded to six scenarios, with two assigned to each temporal domain. For every scenario, participants provided a binary evaluation of each source and made one forced-choice selection, yielding up to 18 primary-outcome responses and six secondary-outcome responses per participant.

Six questionnaire versions were created to counterbalance temporal assignments and scenario presentation order. In Version 1A, the first two scenarios concerned the past, the next two the present, and the final two the future. In Version 1B, the scenarios were assigned to the present, future, and past, respectively, in blocks of two. In Version 1C, they were assigned to the future, past, and present, respectively, in blocks of two. Thus, Versions 1A–1C rotated the temporal domain assigned to each pair of scenarios.

Versions 2A–2C used the same temporal assignments as Versions 1A–1C, respectively, but presented the six scenarios in reverse order. Version 2A presented the scenarios in the reverse order of Version 1A, while Version 2B reversed the order of Version 1B and Version 2C reversed the order of Version 1C.

These variations were intended to distribute temporal-domain assignments and reduce potential order effects. The presentation order of the information sources (Google, ChatGPT, Friend) remained constant across all versions. Participants were randomly assigned to questionnaire versions separately on each data-collection day. For each day, a random-list generator was used to randomize the list of students' school email details, after which students were allocated sequentially to the six questionnaire versions.

3.2 Participants and Sampling

Participants were recruited through opportunity sampling at Kunskapsskolan International School in Gurugram, India. Students in Grades 3–5 were invited to participate after permission was obtained from the school.

A total of 129 form submissions were received. Seventeen were partially incomplete; one contained only age information and a response to the Time Inclusion Task, with no responses to the Informant Evaluation Task; one was entirely blank; and two respondents indicated that they did not assent to participation. These 21 submissions were excluded, leaving 108 participants in the analytical sample.

Within the analytical sample, participants were aged 7–11 years ($M = 8.86$, $SD = 0.91$). Four were aged 7 (3.70%), 40 aged 8 (37.04%), 32 aged 9 (29.63%), 31 aged 10 (28.70%), and one aged 11 (0.93%). Participants aged 7 and 11 were retained because eligibility was determined by school grade rather than a strict age cutoff.

3.3 Sample-Size Justification

Sample size was planned using the participant-level power-planning approach reported by Girouard-Hallam and Danovitch (2025). Because the planned primary analysis involved repeated binary responses analyzed using a generalized linear mixed-effects model (GLMM), a conventional one-sample t-test did not directly represent the planned analysis. Following the approach reported in the reference study, a two-tailed one-sample t-test in G*Power (Faul et al., 2007) was used as a proxy for participant-level sample-size planning. The calculation used $\alpha = .05$, power = .80, and Cohen's $d = .32$, yielding a target minimum of 79 participants.

The value $d = .32$ was adopted as a planning value based on the reference study's procedure, rather than as an exact predicted effect for the present three-source design. A sensitivity analysis using raw data from Study 1 of the reference paper examined the proportions of trials on which participants judged Friend and Google capable of providing the correct answer. Among participants aged 7–10, these proportions were approximately 34.9% for Friend and 73.7% for Google. Hypothetical ChatGPT response rates across this range suggested that plausible intermediate effects could yield pairwise differences around $d = .4$. This analysis provided context for the chosen planning value; the actual ChatGPT effect could not be estimated before data collection.

The planned recruitment target was approximately 100 children to allow for exclusions and incomplete responses, exceeding the minimum calculated sample of 79 to allow for potential exclusions or incomplete responses. After applying the exclusion criteria, the final analytical sample comprised 108 participants. The final analysis used a GLMM rather than the one-sample t-test used for sample-size planning.

Accordingly, the sample-size calculation should be interpreted as a practical recruitment guide rather than a direct assessment of statistical power for the final model.

3.4 Questionnaire Development and Cultural Adaptation

The questionnaire was developed with reference to the study by Girouard-Hallam and Danovitch (2025). The original materials were reviewed, and additional materials were

requested from the authors to inform the adaptation. The central task structure was retained, while ChatGPT was added as an information source. Selected character names and contextual details were also culturally adapted to make names, locations, and contextual references more familiar to children in India.

For example, Carrie was renamed Charu, and a scenario referring to traffic on Hamilton Road was adapted to refer to Golf Course Road. These changes were intended to make the scenarios more familiar to participants while preserving their temporal structure and the judgments being assessed.

Whereas the reference study presented visual materials during individual Zoom sessions, the present study embedded images and instructions in Google Forms, allowing participants to complete the questionnaire online while still receiving visual examples of the information sources.

3.5 Time Inclusion Task

The Time Inclusion Task was adapted from Hudson and Mayhew (2011), as used by Girouard-Hallam and Danovitch (2025), to assess participants' understanding of "yesterday," "today," and "tomorrow." Participants were shown three events in a random, arbitrary order and asked to arrange them chronologically, from the event that occurred first to the event that would occur last: "Tanmay rode his bicycle yesterday," "Maya is playing with blocks today," and "Geeta will listen to music tomorrow."

In the reference study, successful completion was used as an eligibility criterion for the primary task. In the present study, the task was administered to all participants but was retained strictly as a descriptive measure rather than an exclusion tool. This decision was made due to concerns regarding the phrasing and administration of the task, specifically the possibility that children could rely on the visual order of the images rather than true temporal comprehension, limiting the interpretability of incorrect responses. Furthermore, teacher feedback indicated that students in this age group already possessed an adequate understanding of past, present, and future concepts in class.

3.6 Introductions to Google and ChatGPT

Before answering the familiarity questions, participants received introductions to Google and ChatGPT. Each introduction comprised a short explanatory paragraph and three illustrative images: the source's homepage, an example of a question entered into the source, and an example response. The Google introduction explained how Google Search retrieves websites that may contain relevant information. The ChatGPT introduction explained how users submit questions and receive generated responses.

The same example question—about the most popular ice cream flavour in India—was entered into Google and ChatGPT, and the same response text was displayed for both sources. The homepage and user interface shown in the illustrative images differed between the two sources.

3.7 Familiarity Questions

After the introductions, participants answered three questions about each source. They indicated whether they had heard of the source before and whether they had used it previously, with both questions requiring a “Yes” or “No” response. They were also asked to provide an example of a question they might ask the source. The Google questions followed the format of the reference study; corresponding questions were developed for ChatGPT to maintain parallel wording. These responses documented participants’ prior awareness and self-reported use.

3.8 Informant Evaluation Task

The Informant Evaluation Task, adapted from Girouard-Hallam and Danovitch (2025), was the primary task. Participants evaluated six scenarios involving fictional individuals seeking information about past, present, or future events. Each participant received two scenarios from each temporal domain.

For each scenario, participants indicated whether Google, ChatGPT, and a friend of the fictional individual could correctly answer the question. Responses were recorded as “Yes” or “No”; for analysis, “Yes” was coded as 1 and “No” as 0.

Participants then selected which of the three sources they believed would provide the better answer. A selection was required even if they had indicated that none of the sources could answer correctly. This captured both their judgments about each source individually and their relative preference among the available sources.

3.9 Information Stability Task

The Information Stability Task in the reference study assessed children’s understanding that some information remains stable over time while other information can change from day to day. It comprised four questions: two about stable information and two about unstable information.

This task was not included in the present study, in part because the reference study reported a high ceiling effect. The present questionnaire therefore focused on the Time Inclusion Task, familiarity questions, and Informant Evaluation Task, with ChatGPT added alongside Google and Friend.

4. Procedure

Data collection took place at the school during regular school hours and within students’ free periods to avoid disrupting lessons. Grade 3 students participated on the first day, followed by

students in Grades 4–5 on the second day. Both sessions took place for the same duration at the same time of day. At each session, participants gathered in a classroom and received a standardized verbal introduction. The researcher provided a general overview of the questionnaire, stated its estimated completion time, and explained that there were no right or wrong answers. Participants were asked to respond independently and honestly, avoid discussing answers, and remain quiet after submitting the form. They were told that participation was voluntary and that they could stop at any time without consequences. Some of the school's teachers were also present during the sessions.

The questionnaire was administered through Google Forms and distributed to students' school email addresses. Participants completed it at school using their own devices. They waited for the researcher's permission before beginning. The first question asked whether they agreed to participate; those who did not assent were excluded from the analysis. The questionnaire sequence was the Time Inclusion Task, introduction to Google, familiarity questions for Google, introduction to ChatGPT, familiarity questions for ChatGPT.

There was no fixed time limit. The researcher remained present to clarify procedural instructions when necessary, without guiding participants toward particular answers. After completion, participants received a debriefing explaining that the study examined how children judge whether a person, Google, or ChatGPT would be better able to answer questions about the past, present, and future. Participants were thanked and given a toy elephant keychain as a token of appreciation.

5. Ethical Considerations

The study received approval from the school and its ethics board before data collection began. Parents or guardians received a consent form describing the study's purpose, the researcher, the activities involved, potential risks and benefits, and the voluntary nature of participation. Parents could decline participation; one parent opted their child out before data collection, and consequently, that child did not participate or submit a response.

Parental consent was supplemented by child assent at the beginning of the questionnaire. Children were asked whether they wished to participate and were instructed to answer honestly. Two children who declined assent were excluded. Participants were told that participation was voluntary, that they could withdraw without consequences, and that they could change their answers before submitting the form.

The questionnaire did not request participants' names or other directly identifying information; age was the only demographic characteristic collected. School email accounts were used to distribute questionnaire links and allocate questionnaire versions, but responses were not linked to email addresses in the analytical dataset, so the researcher could not determine which email address had submitted a particular response.

The questionnaire used age-appropriate language, written instructions, and illustrative images. Participants were encouraged to provide their own judgments, including when uncertain. The study involved pre-designed scenarios and introductory materials; participants did not interact with either AI system in real time. No physical intervention or intentionally distressing material was involved, and the study was considered to present minimal foreseeable risk.

6. Results

6.1 Time Inclusion Task

Of the 108 participants included in the analytical sample, 48 (44.44%) arranged the three events in correct chronological order, while 60 (55.56%) provided an incorrect or incomplete sequence.

6.2 Primary Outcome: Descriptive Statistics

Descriptive statistics were calculated for participants' binary judgments of whether each source could correctly answer questions about the past, present, and future. Each source contributed 648 responses, comprising 216 responses per temporal domain.

Google received the highest proportion of "Yes" responses overall (70.37%), followed by ChatGPT (58.02%) and Friend (33.95%). The proportion of "Yes" responses declined from past-oriented questions to future-oriented questions for all three sources: from 79.17% to 59.26% for Google, from 63.89% to 50.00% for ChatGPT, and from 38.89% to 27.78% for Friend.

Table 1 presents the complete distribution.

Information source	Temporal domain	Yes, n (%)	No, n (%)	Total
Google	Past	171 (79.17%)	45 (20.83%)	216
Google	Present	157 (72.69%)	59 (27.31%)	216
Google	Future	128 (59.26%)	88 (40.74%)	216
Google	Overall	456 (70.37%)	192 (29.63%)	648
ChatGPT	Past	138 (63.89%)	78 (36.11%)	216
ChatGPT	Present	130 (60.19%)	86 (39.81%)	216
ChatGPT	Future	108 (50.00%)	108 (50.00%)	216
ChatGPT	Overall	376 (58.02%)	272 (41.98%)	648
Friend	Past	84 (38.89%)	132 (61.11%)	216
Friend	Present	76 (35.19%)	140 (64.81%)	216
Friend	Future	60 (27.78%)	156 (72.22%)	216

Friend	Overall	220 (33.95%)	428 (66.05%)	648
--------	---------	--------------	--------------	-----

6.3 Primary Outcome: Inferential Statistical Analysis

A generalized linear mixed-effects model (GLMM) with a binomial distribution and logit link examined whether participants' judgments varied by information source, temporal domain, and age. Random intercepts for participants and items accounted for repeated responses and variation across scenarios.

All statistical analyses were conducted in R (version 4.6.1; R Core Team, 2026). Models were fitted using the lme4 package (Bates et al., 2015), estimated marginal means and pairwise comparisons were obtained using emmeans (Lenth, 2025), and model diagnostics were conducted using DHARMA (Hartig, 2022).

Model comparisons assessed whether interaction terms improved model fit. None indicated a statistically significant improvement (all $p > .05$). The main-effects model had the lowest Akaike information criterion among the compared models (AIC = 2387.4) and was retained as the final model on grounds of parsimony. However, nonsignificant comparisons do not establish that interaction effects were absent.

Type II Wald chi-square tests showed a significant main effect of information source, $\chi^2(2) = 184.62$, $p < .001$, and temporal domain, $\chi^2(2) = 35.96$, $p < .001$. The main effect of age was not significant, $\chi^2(1) = 0.34$, $p = .561$.

Tukey-adjusted pairwise comparisons showed higher odds of a "Yes" response for ChatGPT than Friend (OR = 3.10, 95% CI [2.31, 4.15], $p < .0001$), for Google than Friend (OR = 5.71, 95% CI [4.22, 7.75], $p < .0001$), and for Google than ChatGPT (OR = 1.85, 95% CI [1.38, 2.48], $p < .0001$). These comparisons were averaged across temporal domains.

Across sources, the odds of a "Yes" response were higher for past-oriented questions than future-oriented questions (OR = 2.10, 95% CI [1.56, 2.82], $p < .0001$). The past–present difference was not significant after adjustment (OR = 1.29, 95% CI [0.96, 1.73], $p = .106$). Odds were lower for future-oriented questions than for present-oriented questions (OR = 0.62, 95% CI [0.46, 0.82], $p = .0003$).

Age was not significantly associated with the odds of a "Yes" response ($\beta = 0.050$, SE = 0.086, 95% CI [−0.119, 0.220], $p = .561$). Estimated variances of the participant-level random intercepts and item-level random intercepts were 0.385 and 0.164, respectively. The item-level variance estimate should be interpreted cautiously because it was based on six scenarios.

Formal residual diagnostics did not identify statistically significant problems with residual uniformity ($p = .249$), dispersion ($p = .590$), or outliers ($p = .297$). Diagnostic plots showed some departures in binned residuals and homogeneity patterns, so these tests do not establish that the model was perfectly specified.

6.4 Secondary Outcome: Descriptive Statistics

The secondary outcome comprised participants' forced-choice selections of the source they believed would provide the better answer. Across 648 selections, Google was chosen 370 times (57.10%), Friend 147 times (22.69%), and ChatGPT 131 times (20.22%).

Google was the most frequently selected source across all three temporal domains, peaking in the present domain (60.19%). In contrast to the primary binary evaluation task, forced-choice preferences for ChatGPT increased progressively from past-oriented scenarios (18.06%) to future-oriented scenarios (23.15%), where it virtually converged with choices for Friend (23.61%).

Table 2 presents the full distribution.

Temporal domain	Friend, n (%)	Google, n (%)	ChatGPT, n (%)	Total
Past	52 (24.07%)	125 (57.87%)	39 (18.06%)	216
Present	44 (20.37%)	130 (60.19%)	42 (19.44%)	216
Future	51 (23.61%)	115 (53.24%)	50 (23.15%)	216
Overall	147 (22.69%)	370 (57.10%)	131 (20.22%)	648

7. Discussion

7.1 Overview of Findings

This study investigated how children evaluate human, search engine and generative AI-based information sources when considering questions about the past, present, and future. Building on the research paradigm developed by Girouard-Hallam and Danovitch (2025), the study extended the investigation to include ChatGPT alongside Google and a human peer within an Indian school context. The primary analysis assessed whether children believed that a Friend, Google, or ChatGPT could correctly answer questions across the three temporal domains.

The findings revealed significant differences in children's judgments across information sources and temporal domains. Google was associated with the highest overall proportion of "Yes" responses and highest odds of endorsement (70.37% overall; OR = 5.71 compared to Friend; OR = 1.85 compared to ChatGPT), followed by ChatGPT (58.02% overall) and Friend (33.95% overall). Children were also significantly less likely to judge any source as capable of answering future-oriented questions than past-oriented questions (OR = 2.10) or present-oriented questions (OR = 1.63), though the adjusted difference between past-oriented questions and present-oriented questions was not statistically significant. Age was not significantly associated with children's judgments in the fitted model ($\beta = 0.050$, $p = .561$).

The secondary, forced-choice outcome provided a complementary descriptive perspective. Google was selected most frequently overall (57.10%) as the source expected to provide the better answer, while Friend (22.69%) and ChatGPT (20.22%) received smaller, broadly comparable proportions of overall selections. Although both the primary and secondary outcomes highlighted Google's prominent position in children's evaluations, the secondary results were descriptive and were not subjected to inferential statistical testing. Consequently, they reflect patterns in observed choices rather than statistically established differences in source preference.

7.2 Children's Evaluations of Human, Search Engine and Generative AI Information Sources

The significant main effect of information source ($\chi^2(2) = 184.62, p < .001$) suggests that children did not regard all three sources as equally capable of answering questions. Google had substantially higher odds of receiving a "Yes" response than both ChatGPT and Friend, while ChatGPT also had higher odds than Friend (OR = 3.10). These differences were observed across temporal domains in the main-effects model.

One possible interpretation is that children may distinguish between sources according to their perceived access to information. Google is explicitly designed to retrieve information from across the internet, and children may associate it with a broad capacity to locate answers. ChatGPT, meanwhile, presents information through a conversational interface, potentially making it appear more similar to a human interlocutor while still being understood as a distinct source. A friend may be perceived as knowledgeable about personal experiences or familiar events but less likely to have access to the same breadth of information. These explanations are plausible interpretations rather than mechanisms directly tested by the present study.

The findings also suggest that children may attribute different kinds or degrees of knowledge to human, search engine and generative AI sources. However, the results do not establish that participants understood how these systems obtain information or that they could accurately distinguish their capabilities. A "Yes" response measured a child's judgment about whether a source could answer a question; it did not establish whether the source would actually produce a correct answer. The findings therefore concern children's perceived capabilities of information sources rather than their objective accuracy.

The higher endorsement of Google than ChatGPT (70.37% vs. 58.02% overall) is particularly noteworthy given the increasing accessibility of generative AI. It suggests that, in this sample, the addition of a conversational AI source did not displace Google as the source most frequently judged capable of answering questions. Nevertheless, the study did not directly measure the mechanisms underlying this difference. Familiarity, prior exposure, interface design, and children's assumptions about how information is obtained may all be relevant, but the available data do not permit conclusions about their individual contributions.

These findings extend the investigation of children's judgments about human and search engine information sources by incorporating a generative AI system. They also highlight the importance of distinguishing between children's general willingness to endorse a source and their understanding of the specific circumstances in which that source may be reliable.

7.3 The Influence of Temporal Domain

Temporal domain also had a significant main effect on children's judgments ($\chi^2(2) = 35.96$, $p < .001$). Across information sources, children had higher odds of judging a source capable of answering past-oriented questions than future-oriented questions (OR = 2.10). They were also more likely to endorse sources for present-oriented than future-oriented questions (OR = 0.62 comparing future to present). The adjusted comparison between past-oriented questions and present-oriented questions, however, was not statistically significant ($p = .106$).

This pattern may reflect children's developing understanding of the differences between information about events that have already occurred, events happening in the present, and events that have yet to occur. Past-oriented questions concern events that may already have records or established answers, while present-oriented questions may concern circumstances that can be observed or checked at the time of asking. Future-oriented questions, by contrast, often require predictions about events that have not yet happened. Children may therefore be less certain that any source can provide a correct answer about the future.

This interpretation is consistent with the distinction between retrieving information about an existing event and predicting an uncertain future outcome. However, the study did not directly measure children's reasoning about these distinctions, so it cannot establish why future-oriented questions received fewer "Yes" responses. The differences may also have been influenced by the specific scenarios used, children's interpretations of the questions, or their assumptions about what each source could know.

The temporal pattern also indicates that children's judgments were not determined solely by the identity of the information source. Their evaluations varied according to the type of question being asked. This reinforces the importance of examining children's understanding of information sources in relation to the knowledge required by a particular question, rather than treating trust in a source as a single, context-independent judgment.

The model comparisons did not provide statistically significant evidence that interaction terms improved model fit. Accordingly, the results support interpreting the main effects of information source and temporal domain as overall patterns across the study conditions. However, the nonsignificant interaction tests do not establish that children's evaluations of every source were identical across temporal domains. Further research with larger and more varied samples would be needed to examine these relationships more precisely.

7.4 Age and Children's Judgments

Age was not significantly associated with the odds of a “Yes” response in the fitted model ($\chi^2(1) = 0.34, p = .561$). Within this sample, the analysis therefore did not provide sufficient evidence that age predicted children's evaluations of information sources after accounting for the other modelled factors.

This finding should not be interpreted as evidence that children's understanding of information sources does not develop with age. The participants were drawn from Grades 3–5 at a single school, and the observed age distribution was concentrated between eight and ten years. The study consequently examined a relatively narrow developmental range. Any age-related differences may have been too small to detect within this range, or they may become more apparent when children are compared across a broader span of development.

Furthermore, chronological age may not fully capture differences in children's experiences with search engines, generative AI, or independent information-seeking. Children of the same age may differ in their exposure to these technologies and in their understanding of their capabilities. Since the study did not analyze familiarity responses or directly measure AI literacy, it cannot determine whether such differences help explain children's judgments.

The absence of a statistically significant age effect therefore represents a finding within the limits of the present sample and model. It provides a basis for further investigation rather than a definitive conclusion about the role of development in children's evaluations of human, search engine and generative AI information sources.

7.5 Forced-Choice Source Preferences

The secondary forced-choice responses offered an additional view of children's evaluations. Overall, Google was the most frequently selected source (57.10%), followed by Friend (22.69%) and ChatGPT (20.22%). Google remained the most frequently selected source across all three temporal domains (peaking at 60.19% in the present domain). ChatGPT accounted for a larger proportion of selections for future-oriented questions (23.15%) than for past-oriented questions (18.06%), converging closely with selections for Friend in the future domain (23.61%). Friend's selection proportions were relatively similar across temporal domains, but highest for past-oriented questions (24.07%).

These patterns may indicate that children's judgments about whether a source could answer a question and their judgments about which source would provide the better answer are related but distinct. The primary task asked participants to evaluate each source independently, whereas the forced-choice task required them to select one source from the available alternatives. A source could therefore receive a “Yes” response without necessarily being chosen as the best source.

Because no inferential tests were conducted for the forced-choice outcome, the observed differences between temporal domains should not be interpreted as statistically established effects. The results are best understood as descriptive evidence that Google was the most

frequently preferred source in this sample, while Friend and ChatGPT received smaller but broadly comparable proportions of selections.

7.6 Time Inclusion Task

In the Time Inclusion Task, less than half of the analytical sample (44.44%) correctly ordered the three events. While the primary task relied on children comprehending past, present, and future concepts, these results should not be interpreted as a direct reflection of each child's overall understanding of time. Because the task comprised only three items, an incorrect ordering could reflect several factors, including task interpretation, attention, or difficulties with the particular presentation rather than conceptual misunderstanding.

Retaining all participants allowed their completed responses to the primary evaluation task to be included without using performance on a potentially ambiguous preliminary task as an exclusion criterion. Future work could examine more directly how children's understanding of temporal concepts relates to their evaluations of information sources.

7.7 Methodological Considerations and Limitations

Several methodological considerations should be taken into account when interpreting these findings.

First, participants were recruited through opportunity sampling from a single school in Gurugram, India. Although the study contributes evidence from a context that is less represented in this area of research, the sample cannot be assumed to represent children across India. Differences in educational environments, socioeconomic circumstances, access to technology, language, and cultural experiences may influence how children evaluate human, search engine and generative AI information sources. The findings should therefore be generalized cautiously beyond the participating school.

Second, the questionnaire was administered online through Google Forms, and the order of the three information sources remained fixed across all questionnaire versions. Although the temporal assignments and presentation orders of the scenarios were rotated, the source order was not counterbalanced. Consequently, potential effects of presentation order, interface familiarity, or the online format cannot be ruled out. The use of students' own devices also meant that the study did not fully standardize the viewing conditions experienced by every participant.

Third, another methodological consideration concerns the introductory materials used to familiarise participants with Google and ChatGPT. Both sources were introduced using the same example question and response, although their interfaces and the ways in which they presented information differed. These materials were intended to familiarise children with both sources before the main task; however, the example and visual presentation may have shaped how children perceived each source's capabilities and, consequently, their judgments.

Fourth, the Time Inclusion Task did not function as an exclusion criterion. Although this result was reported descriptively, the study did not examine whether performance on this task was associated with responses in the primary analysis. It is therefore difficult to determine the extent to which individual differences in understanding temporal concepts influenced the main findings. The task's limited format also means that its results should not be interpreted as a comprehensive assessment of children's temporal understanding.

Fifth, the age distribution was relatively restricted, with most children between eight and ten years old and only a small number at the youngest and oldest ages represented. This limits the conclusions that can be drawn about developmental differences. The nonsignificant age effect should not be taken to indicate that age is irrelevant to children's evaluations of information sources, particularly because the study did not cover a broad developmental range. In addition, the forced-choice outcome was analyzed descriptively only, so the study cannot establish whether the observed differences in source preferences would remain statistically significant under inferential testing.

Taken together, these considerations do not negate the observed patterns, but they define the scope of the conclusions that can reasonably be drawn. The findings provide evidence of differences in children's judgments within this particular sample and task, while leaving questions about generalizability, developmental variation, and the processes underlying those judgments open for further investigation.

7.8 Directions for Future Research

Future research could build on the present findings in four main directions.

First, studies should examine children's evaluations of human, search engine, and generative AI information sources across a wider range of cultural and geographical contexts, particularly in non-Western locations. Much of the existing research on children's judgments about information sources has been conducted in Western settings. Research in India and other underrepresented contexts could help establish whether the patterns observed here are consistent across different educational systems, cultural environments, and levels of access to technology. Multi-school and cross-cultural studies would also help distinguish patterns that may be broadly shared by children from those that depend on particular social or educational contexts.

Second, future studies should investigate developmental differences using a broader age range and, where feasible, longitudinal designs. Including younger children, older children, and adolescents would allow researchers to examine whether judgments about human sources, search engines, and generative AI change as children's cognitive abilities, experiences, and understanding of technology develop. Longitudinal research could further explore how these evaluations change over time rather than relying on comparisons between age groups at a single point.

Third, future research should strengthen and extend the methodology. Studies could recruit larger and more diverse samples, include a wider range of question scenarios, and counterbalance the presentation order of information sources to reduce potential order effects. Researchers could also refine measures of temporal understanding and examine whether performance on such tasks is associated with children's judgments about information sources. These methodological improvements would help establish the robustness of the observed patterns and provide a more detailed account of the circumstances in which children endorse different sources.

Fourth, future studies should investigate the reasoning behind children's judgments and compare perceived source capabilities with actual answer accuracy. Asking children to explain their choices, examining their familiarity with different technologies, and testing sources against questions with independently verifiable answers could help distinguish perceived trustworthiness from demonstrated reliability. Such research could also investigate how children respond when information sources disagree or provide incorrect answers. This would be particularly useful for understanding how children evaluate AI-generated information and whether they can recognize circumstances in which a source's answer requires verification. Together, these directions would extend the present study from documenting children's judgments toward a more comprehensive understanding of how those judgments develop, how they vary across contexts, and how they relate to the actual reliability of information sources.

7.9 Conclusion

This study examined how children in an Indian school evaluated a human source, Google, and ChatGPT when considering questions about the past, present, and future. The primary analysis indicated that children were most likely to judge Google as capable of answering a question, followed by ChatGPT and Friend. Children were also less likely to endorse information sources for future-oriented questions than for past-oriented questions or present-oriented questions, while age was not significantly associated with their judgments in the fitted model. The descriptive forced-choice results similarly identified Google as the most frequently selected source.

These findings extend research on children's evaluations of information sources by including generative AI alongside a search engine and a human source. They also provide evidence from an Indian context, contributing to an area of research in which non-Western settings remain comparatively underrepresented. Nevertheless, the findings concern children's perceptions rather than the objective accuracy of the sources, and they should be interpreted in light of the study's sampling and methodological limitations.

As AI systems become increasingly accessible to children, understanding how they judge the capabilities of different information sources is an important part of understanding children's developing relationship with information. The present study offers an initial contribution to this question and highlights the need for further cross-cultural, developmental, and methodological

research into how children evaluate, compare, and verify information from human informants, search engines and generative AI sources.

8. References

- Atance, C. M. (2015). Young children's thinking about the future. *Child Development Perspectives*, 9(3), 178–182. <https://doi.org/10.1111/cdep.12128>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Chandler-Campbell, I. L., Ghossainy, M. E., Mills, C. M., & Corriveau, K. H. (2022). Is secondhand information better read or said? Factors influencing children's endorsements of text-based information. *Cognitive Development*, 64, Article 101215. <https://doi.org/10.1016/j.cogdev.2022.101215>
- Danovitch, J. H., & Alzahabi, R. (2013). Children show selective trust in technological informants. *Journal of Cognition and Development*, 14(3), 499–513. <https://doi.org/10.1080/15248372.2012.689391>
- Einav, S., Levey, A., Patel, P., & Westwood, A. (2020). Epistemic vigilance online: Textual inaccuracy and children's selective trust in webpages. *British Journal of Developmental Psychology*, 38(4), 566–579. <https://doi.org/10.1111/bjdp.12335>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Girouard-Hallam, L. N., & Danovitch, J. H. (2022). Children's trust in and learning from voice assistants. *Developmental Psychology*, 58(4), 646–661. <https://doi.org/10.1037/dev0001318>
- Girouard-Hallam, L. N., & Danovitch, J. H. (2025). Children's trust in Google's ability to answer questions about the past, present, and future. *Computers in Human Behavior*, 165, Article 108496. <https://doi.org/10.1016/j.chb.2024.108496>
- Harris, P. L., Koenig, M. A., Corriveau, K. H., & Jaswal, V. K. (2018). Cognitive foundations of learning from testimony. *Annual Review of Psychology*, 69, 251–273. <https://doi.org/10.1146/annurev-psych-122216-011710>
- Hartig, F. (2022). DHARMA: Residual diagnostics for hierarchical (multi-level/mixed) regression models [R package]. <https://CRAN.R-project.org/package=DHARMA>
- Hudson, J. A., & Mayhew, E. M. Y. (2011). Children's temporal judgments for autobiographical past and future events. *Cognitive Development*, 26(4), 331–342. <https://doi.org/10.1016/j.cogdev.2011.09.005>

Koenig, M. A., Cole, C. A., Meyer, M., Ridge, K. E., Kushnir, T., & Gelman, S. A. (2015). Reasoning about knowledge: Children's evaluations of generality and verifiability. *Cognitive Psychology*, 83, 22–39. <https://doi.org/10.1016/j.cogpsych.2015.08.007>

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development*, 76(6), 1261–1277. <https://doi.org/10.1111/j.1467-8624.2005.00849.x>

Lenth, R. V. (2025). emmeans: Estimated marginal means, aka least-squares means [R package]. <https://CRAN.R-project.org/package=emmeans>

Oh, S., Zhang, C., Girouard-Hallam, L., Zhou, Z., March, H., Jayaramu, S., & Xu, Y. (2025). “Hey Curio, can you tell me more?": Children's information-seeking and trust in AI. In *Proceedings of the 24th ACM Interaction Design and Children Conference* (pp. 545–555). Association for Computing Machinery. <https://doi.org/10.1145/3713043.3727059>

R Core Team. (2026). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>

Russell, J., Alexis, D., & Clayton, N. (2010). Episodic future thinking in 3- to 5-year-old children: The ability to think of what will be needed from a different point of view. *Cognition*, 114(1), 56–71. <https://doi.org/10.1016/j.cognition.2009.08.013>

Tillman, K. A., Tulagan, N., Fukuda, E., & Barner, D. (2018). The mental timeline is gradually constructed in childhood. *Developmental Science*, 21(6), Article e12679. <https://doi.org/10.1111/desc.12679>

Van Brummelen, J., Tian, M. C., Kelleher, M., & Nguyen, N. H. (2024). Learning affects trust: Design recommendations and concepts for teaching children—and nearly anyone—about conversational agents. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(13), 15860–15868. <https://doi.org/10.1609/aaai.v37i13.26883>

Wang, F., Tong, Y., & Danovitch, J. H. (2019). Who do I believe? Children's epistemic trust in internet, teacher, and peer informants. *Cognitive Development*, 50, 248–260. <https://doi.org/10.1016/j.cogdev.2019.05.006>

Zhang, M., & Hudson, J. A. (2018). Children's understanding of yesterday and tomorrow. *Journal of Experimental Child Psychology*, 170, 107–133. <https://doi.org/10.1016/j.jecp.2018.01.010>

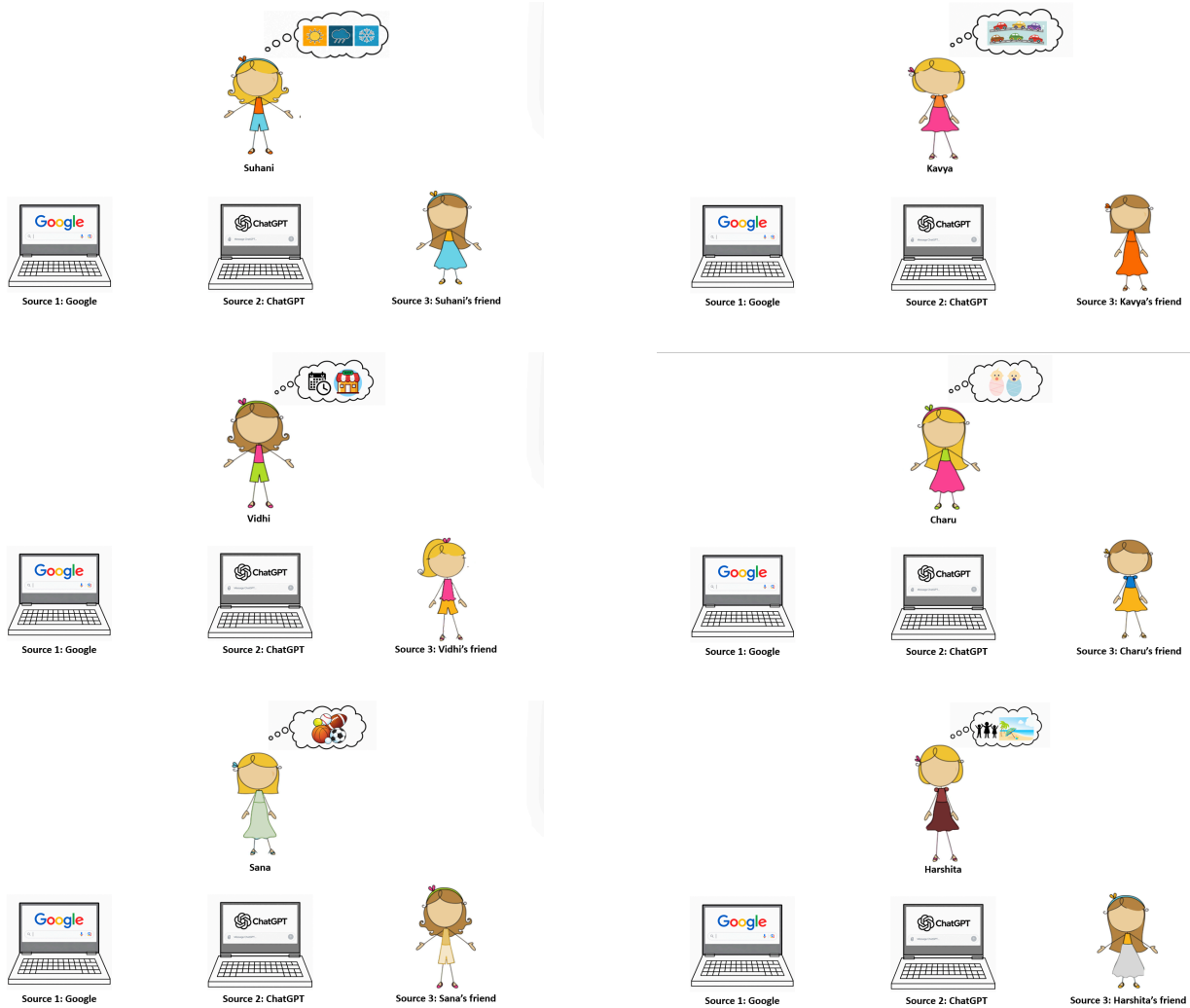
9. Appendices

9.1 Appendix A: Informant Evaluation Scenarios

The following table presents the six scenarios used in the Informant Evaluation Task from Questionnaire 1A. Each scenario involved a fictional individual seeking information from Google, ChatGPT, and a friend.

No.	Character	Scenario Question
1	Suhani	Suhani wants to know what the weather was like yesterday.
2	Kavya	Kavya wants to know if there was traffic on Golf Course Road yesterday.
3	Vidhi	Vidhi wants to know what time the shop opens today.
4	Charu	Charu wants to know how many babies are born in India today.
5	Sana	Sana wants to know which cricket team will win the match tomorrow.
6	Harshita	Harshita wants to know how many people will be at Juhu Beach tomorrow.

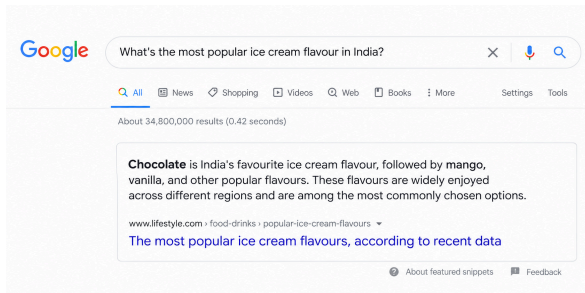
The following images present the six fictional characters and their corresponding information-seeking scenarios used in the Informant Evaluation Task. Each image depicts the character alongside the three information sources—Google, ChatGPT, and a friend—whom participants evaluated as potential sources of information.



9.2 Appendix B: Introductions to Google

Google is a search engine, which is something that uses the internet in order to answer questions about different things. If you have a question, you can “Google” it. You do this by typing a question into a box on the screen. After you’ve typed in your question, you can press a button that says “Google Search”. Google then searches the internet for websites that might have the answer.

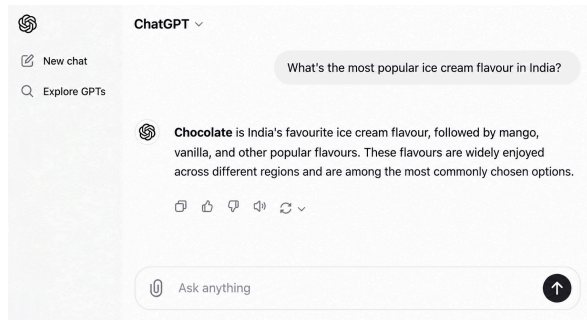
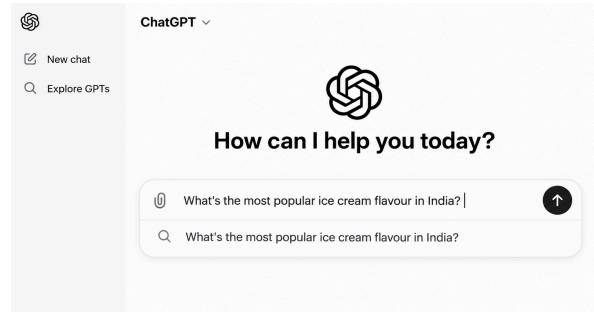
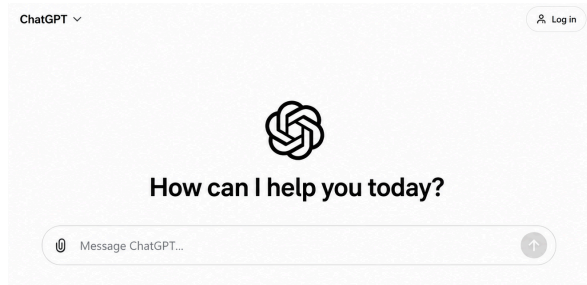
The following images illustrate the Google homepage, an example question entered into Google, and the example response presented to participants.



9.3 Appendix C: Introductions to ChatGPT

ChatGPT is an Artificial Intelligence (AI) tool, which is something that uses information learned from many books, articles, and websites in order to answer questions about different things. If you have a question, you can ask ChatGPT about it. You do this by typing a question into a box on the screen. After you've typed in your question, you can press a button to send it. ChatGPT then generates an answer to your question.

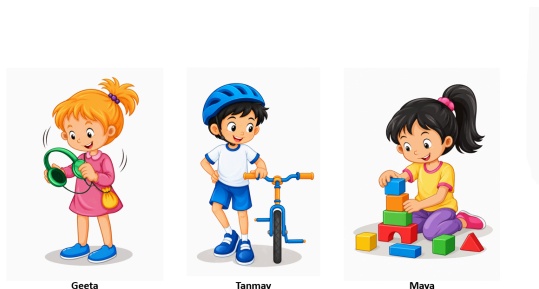
The following images illustrate the ChatGPT homepage, an example question entered into ChatGPT, and the example response presented to participants.



9.4 Appendix D: Time Inclusion Task

The three events were:

1. Tanmay rode his bicycle yesterday.
2. Maya is playing with blocks today.
3. Geeta will listen to music tomorrow.



9.5 Appendix E: Cultural Adaptations to the Reference Study

The following table presents the cultural adaptations made to the statements and questions from the reference study for the present study. The wording and temporal assignments shown correspond to the Informant Evaluation Task from Questionnaire 1A.

Reference Study Wording	Adapted Wording	Reason for Adaptation
Suzie wants to know what the weather was like yesterday.	Suhani wants to know what the weather was like yesterday.	Adapted the character's name to the Indian context.
Carla wants to know if there was traffic on Hamilton Road yesterday.	Kavya wants to know if there was traffic on Golf Course Road yesterday.	Adapted the character's name and road location to the Indian context.
Victoria wants to know what time the store opened yesterday.	Vidhi wants to know what time the shop opened yesterday.	Adapted the character's name to the Indian context and replaced "store" with "shop," a term more widely used in India.
Carrie wants to know how many babies were born in the United States yesterday.	Charu wants to know how many babies were born in India yesterday.	Adapted the character's name and country reference to the Indian context.
Sarah wants to know what sports team won the big game yesterday.	Sana wants to know which cricket team won the match yesterday.	Adapted the character's name and sports reference to the Indian context.
Hannah wants to know how many people were at Harper Beach yesterday.	Harshita wants to know how many people were at Juhu Beach yesterday.	Adapted the character's name and beach location to the Indian context.
Tom rode his bike yesterday.	Tanmay rode his bicycle yesterday.	Adapted the character's name to the Indian context and replaced "bike" with "bicycle," the term more widely used in India.
Mary is playing with blocks today.	Maya is playing with blocks today.	Adapted the character's name to the Indian context.
Grace will listen to music tomorrow.	Geeta will listen to music tomorrow.	Adapted the character's name to the Indian context.