Expectations and Practices around AI Disclosure in CS Research

Arati Mohapatra
Indian Institute of Science
Bengaluru, KA, India
aratim@iisc.ac.in

Danish Pruthi
Indian Institute of Science
Bengaluru, KA, India
danishp@iisc.ac.in

Abstract

As generative AI tools find increasing use in research workflows, ongoing debates on their impact, appropriateness and responsible use have led policymakers to enact policies to disclose AI use at multiple publishing venues. However, are current AI disclosure policies and practices reflective of their purpose? In this work, we first investigate disclosure policies of top computer science venues and find that despite their prevalence, they remain highly under-specified. Secondly, through a survey of computer science researchers (N=109), we characterize the necessity of disclosures across different research tasks and levels of human involvement. We learn that researchers find disclosures most necessary for tasks involving research design, and for tasks when the human involvement is low. We also compile expectations that researchers have about the information to be conveyed in AI disclosure statements. Lastly, through an analysis of 13867 disclosure statements from EMNLP 2025 and ICLR 2026, we reveal a large disconnect between these expectations and AI disclosures in practice—a prime example being writing assistance which is deemed less necessary but is frequently disclosed. We conclude with recommendations to align AI disclosure policies and practices with expectations, suggesting a categorization of research tasks by perceived necessity and a boilerplate template capturing expected details.[1]

1 Introduction

Accompanying the ongoing release of increasingly capable and efficient generative AI tools are discussions around the appropriateness of their use across professions (Inie et al., 2023; Li et al., 2024; Butler et al., 2025). The scientific community is similarly engaged in such a debate, and while researchers report improvements in quality and efficiency from the integration of generative AI tools into scientific workflows, concerns around their ethical use abound (Khalifa and Albadawy, 2024). Due to their ability to generate coherent natural language content on demand, generative AI tools may potentially shortcut intellectual contributions that were once considered indicators of rigorous human effort (Watermeyer et al., 2025). Hence, to maintain the credibility, transparency, and reproducibility of research, strong arguments for disclosing generative AI use have emerged (Hosseini et al., 2023; van Dis et al., 2023; Resnik and Hosseini, 2026).

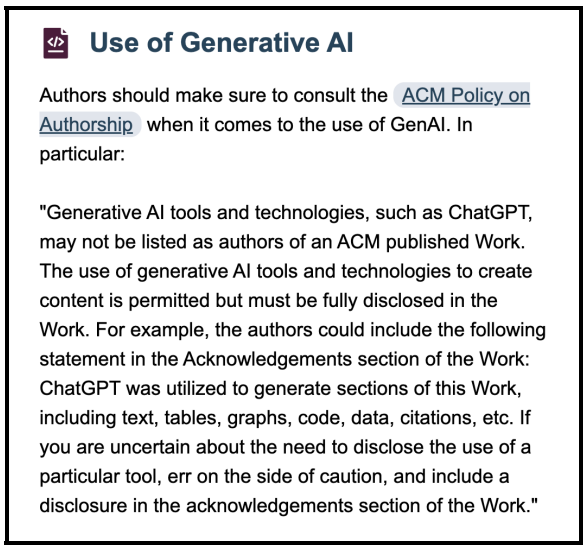


Figure 1: ACM SIGCOMM's AI Disclosure policy.

As a consequence, multiple publishing venues set up policies encouraging the inclusion of an AI disclosure statement in submitted research papers (Ganjavi et al., 2023). The introduction of such policies was especially swift in the field of computer science, given the role of the community in developing these tools, and their relevance to computer science research (Chairs, 2023). Relatedly, perceptions of quality and trust placed in research papers containing AI disclosures are mixed, per-

[1]Interactive survey results and code are available at https://ai-disclosure.github.io/

Findings of the Association for Computational Linguistics: EMNLP 2026, pages 48880–48903
October 24-29, 2026 ©2026 Association for Computational Linguistics

haps due to shifting norms around generative AI use (Akpinar et al., 2026). It is however evident that the presence of AI disclosures can shape the trust placed in scientific discourse. Yet, there have been minimal efforts to characterize the current state of AI disclosures in research.

In this work, we first examine the prevalence and evolution of publishing venue policies that dictate the existence of AI disclosures in computer science research. We find that all major publishing societies (AAAI, ACL, ACM, and IEEE) and 35 out of 65 top computer science conferences have AI disclosure guidelines in place. However, we notice that they have stayed fairly constant since their introduction despite the rapidly evolving capabilities of AI tools, and provide limited guidance on *when* generative AI use should be disclosed and *what* specific details to include in a disclosure statement (see Figure 1 for an example).

To characterize current perspectives around the necessity and content of AI disclosures, we then conduct a survey of computer science researchers (N=109). Specifically, we ask the question: *For what research tasks and levels of human involvement is it necessary to disclose generative AI use?* We observe that generally, participants find it more imperative to disclose generative AI use for research design and analysis than for writing. However, we also highlight the large variation in these opinions, which prevents us from recommending a few categories of tasks that warrant disclosure. Moreover, we show that the perceived necessity of disclosure depends on the level of human involvement during generative AI use, with higher levels of human involvement uniformly requiring less frequent disclosure.

To understand whether current disclosure practices are reflective of these expectations, we analyze AI disclosure statements from EMNLP 2025 and ICLR 2026, two computer science conferences with established AI disclosure policies. We find that disclosures are included in 40% and 64% of research papers accepted at EMNLP 2025 and submitted to ICLR 2026 respectively. However, our analysis reveals a disconnect between expected disclosure behavior and how disclosing generative AI use plays out in practice. Specifically, while a declaration of responsibility for AI-assisted work is consistently expected, concerningly, 98% and 77% of AI disclosure statements in EMNLP and ICLR do not include this information. Moreover, the most frequently disclosed tasks are polishing text and editing code, for which disclosure is less expected.

Overall, our findings underscore both the complexity of formulating precise AI disclosure policies and the misaligned nature of current disclosure practices. As initial steps, we recommend that policies move towards task-based guidelines, clarifying disclosure needs per research task rather than imposing a blanket requirement across all uses. Specifically, we suggest organizing research tasks into *mandatory*, *recommended*, and *optional* categories based on the disclosure necessity scores assigned by the participants in our survey. We further recommend standardizing the format of disclosures, and design a boilerplate template that captures widely expected details (see §7).

While we expect norms around AI disclosure to evolve organically, we hope this work will provide a starting point for the alignment of policies and practices with expectations, thus contributing to ongoing efforts to maintain credibility and transparency in research communication.

2 Related Work

2.1 Generative AI in Research Workflows

Given the recent interest in AI for Science, efforts are underway to develop AI tools to assist researchers with the aim of accelerating scientific discovery (Eger et al., 2026). Usage data from AllenAI's Asta, a research-specific generative AI service shows the prevalence of use for tasks from literature retrieval and idea generation to designing methods and writing (Haddad et al., 2026). The use of generative AI assistance at multiple stages spanning the research lifecycle is supported by researchers' self-reported Large Language Model (LLM) use (Liao et al., 2024). Notably, perceptions of appropriateness of AI use vary by research tasks, which may in turn indicate differing necessity for AI disclosure (Kwon, 2025). Moreover, these norms themselves vary across different fields of research, with researchers in quantitative or experimental fields, such as computer science, demonstrating more openness towards both AI use and disclosure (Andersen et al., 2025).

2.2 AI Disclosure Policies

Since the introduction of the first publicly available generative AI chatbots and tools in 2022-2023, the policy landscape on AI use and disclosure has seen multiple shifts. Initial policies set up by publish-

ers and journals across domains, while sufficiently prevalent, demonstrated considerable heterogeneity regarding the appropriateness of generative AI use and disclosure, particularly on when and how generative AI use should be disclosed (Ganjavi et al., 2023). In computer science specifically, a temporal analysis of conferences policies from 2023-2025 shows an uptick in the establishment of policies around generative AI use for authors, albeit with varying leniency and sanctions (Nahar et al., 2025). Domain-agnostic attempts to standardize AI disclosure policies have recommended declaring generative AI assistance when used in an intentional and substantial manner, but have not yet reached consensus on the details to include in an AI disclosure statement, thus leaving significant room for author interpretation (Weaver, 2024; Resnik and Hosseini, 2026; BaHammam, 2025). While these frameworks provide a valuable starting point for standardization, ensuring their representativeness of broader opinions of the research community remains essential.

2.3 Perceptions of AI Disclosures

AI disclosures in research papers have been shown to impact perceptions of information quality and trust. A recent study shows that among abstracts disclosing differing levels of human and LLM involvement, human-written but LLM-edited abstracts receive the highest clarity ratings, whereas fully LLM-written abstracts score the least on both quality and trust scales, implying that the level of human involvement plays a role in influencing perceptions (Akpinar et al., 2026). Moreover, this study shows that while the absence of an AI disclosure leads to speculations about the origins of the given text rather than quality evaluations, its presence puts these suspicions to rest, and increases trustworthiness. However, despite these recent findings demonstrating the positive effects of AI disclosure, hesitation towards full transparency may linger, due to a variety of reasons including how generative AI assistance was used, and social and personal factors (Fang and Lee, 2025). Evidence from a survey of 777 researchers shows that self-decided appropriateness and the normalization of generative AI use are potential reasons for non-disclosure (Yusuf et al., 2025). Thus, in the absence of clear AI disclosure standards, non-disclosure may persist, with the potential to hamper transparency and trust in research (BaHammam, 2025).

3 Research Questions

The review of related work highlights the prevalent use of generative AI assistance in research workflows, along with the attempts to introduce and standardize AI disclosure policies. This is especially imperative given the ability of declared AI assistance to impact the trust placed in research. Past work has also shown that these effects are closely entangled with the perceived appropriateness of generative AI use, that varies based on specific research tasks and levels of human involvement. In light of their potential impact, we attempt to characterize the current landscape of AI disclosures by addressing the following research questions:

> **RQ1:** How prevalent are AI disclosure policies, and how have they evolved over time?
> **RQ2:** For what research tasks and levels of human involvement is it necessary to disclose generative AI use?
> **RQ3:** How well do current AI disclosure practices align with reader expectations?

We detail the procedures and results of the mixed-methods approach we took to answer these research questions in the following sections.

4 AI Disclosure Policy Analysis

> **RQ1:** How prevalent are AI disclosure policies, and how have they evolved over time?

4.1 Method and Procedure

To answer RQ1, we analyzed 65 computer science conferences, examining the existence, origins and evolution of their AI disclosure policies. The 65 conferences, spanning subdomains of Artificial Intelligence, Systems, Theory, and Interdisciplinary Areas were taken from the CSRankings website.[2] This list, developed in consultation with faculty and through community surveys, is representative of the top conferences in computer science. Since our aim was to characterize the current state of AI disclosure policies, we limited ourselves to an analysis of conference policies from 2025-2026. We also analyzed the policies of AAAI, ACL, ACM, and IEEE, since many computer science conferences are affiliated with these societies, and hence

[2] https://csrankings.org/

may have contributed to shaping their current policies. We mapped each of the 65 conferences to one or more of these society policies if they link to, mention, or verbatim quote a particular society policy. To trace the evolution of society policies since their origin, we used the Internet Archive's Wayback Machine,[3] which stores snapshots of web pages across time. We initially conducted these analyses from January 5-19, 2026, and reviewed and revised them from May 18-20, 2026.

4.2 Results

Our analysis of 65 computer science conferences shows that roughly half (35) of the conferences have set up AI disclosure policies. Moreover, we find that the major societies, AAAI, ACL, ACM, and IEEE, all have established AI disclosure policies: 29 of the 35 conferences with AI disclosure policies borrow closely from these society-level policies. Notably, 21 conferences directly link to, mention, or verbatim quote the ACM policy on authorship, which includes guidelines on disclosing the use of generative AI. From our study of the evolution of society-level policies, we find that they have undergone only 1 to 4 changes since their first introduction in early 2023, with revisions being minor, including restructuring existing content, adding reviewer guidelines or sanctions for prompt injections. We also note that the ACL and ACM policies include clear guidelines on the conditions necessitating disclosure of generative AI use based on either the novelty of generated text or the distinction between AI assistance for research versus writing. However, the AAAI and IEEE policies remain majorly open-ended, providing limited guidance on *when* and *where* to disclose generative AI use, only stating that any use of generative AI should be disclosed. Moreover, worryingly, none of the society-level policies provide specific guidance on what details are required to be mentioned while disclosing generative AI use.

5 AI Disclosure Expectations Survey

RQ2: For what research tasks and levels of human involvement is it necessary to disclose generative AI use?

[3]https://web.archive.org/

Level	Description
Assumed	No explicit mention of human involvement. Participants *assume* a certain baseline level of human involvement.
High	When the author provides most of the contribution and initiative, while AI outputs are closely supervised and verified.
Low	When the AI assistant provides most of the contribution and initiative, and AI outputs are loosely supervised and verified by the author.

Table 1: Descriptions of different levels of human involvement considered in the survey (See §5.1).

Dimension	Variants
Detail	Task Model name Reason for AI use Non-use of AI Human oversight Responsibility declaration Purpose of disclosure
Length	1-2 sentences A few sentences or a short paragraph Multiple paragraphs Length and detail proportionate to the extent of AI use

Table 2: Dimensions along which survey participants indicate their preferences for framing AI disclosures.

5.1 Method and Procedure

To answer RQ2, we administered an online survey (on LimeSurvey) with 109 respondents. Participants rated the necessity of AI disclosure for 21 research tasks at 3 levels of human involvement on a 5-point Likert scale, where 1 indicates that AI use need not be disclosed and 5 implies that disclosure is always necessary. The chosen 21 research tasks span the 5 chronological phases central to the preparation of a research manuscript: idea generation, research design, data collection, data analysis, and writing and reporting.

The taxonomy of research tasks we used was adapted from Andersen et al. (2025), with minor changes to ensure suitability to our study and minimize cognitive load on participants. For instance, we removed all tasks related to peer review and research proposal writing, as these are unrelated to manuscript preparation, and also grouped together all coding tasks (for data analysis, statistical analysis, and simulations). We ensured that the final taxonomy is representative of tasks known to be assisted through generative AI tools (Liao et al.,

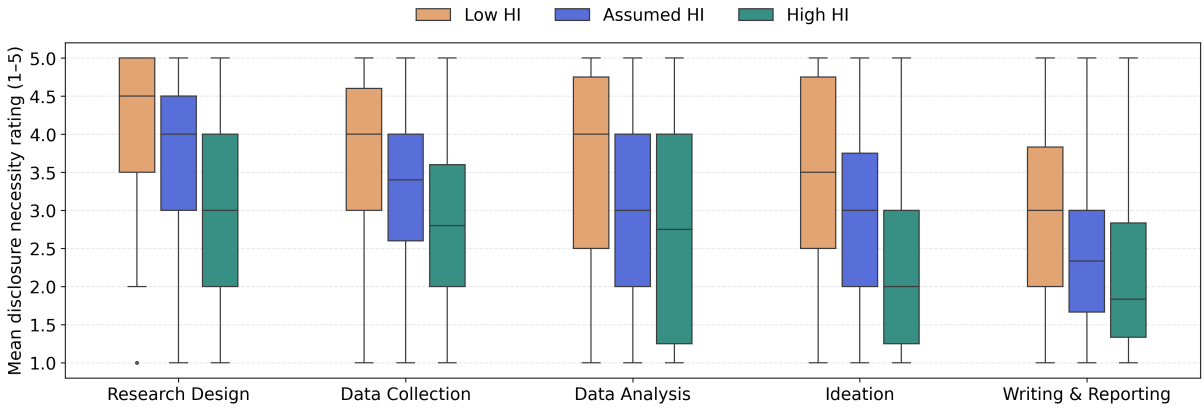


Figure 2: AI disclosure necessity ratings across research phases and human involvement (HI) conditions.

2024; Kapania et al., 2025). The research phase-wise categorization of tasks can be found in the Appendix, but individual tasks will also be listed in the subsequent sections, when we detail and discuss their necessity scores.

Given the diverse methods of using current generative AI tools, the appropriateness of use does not only hinge on *what* AI assistance was used for, but also on *how* it was used. Particularly, since computer science societies and conferences across the board, including ACL[4], NeurIPS[5] and RTAS[6], stress the need for human intellectual contribution, integrity and verification, we try to understand whether the level of human involvement impacts AI disclosure necessity. The levels of human involvement used are described in Table 1.

To characterize expectations on the framing of an AI disclosure statement, we elicit preferences on details and length through multiple-choice questions. The specific options, present in Table 2 were consolidated from existing conference guidelines and conceptual frameworks outlined in past work (Resnik and Hosseini, 2026; BaHammam, 2025; Weaver, 2024). We empirically validate these options, by analyzing 110 random AI disclosure statements from ICLR 2026, which notably led to the inclusion of the *Purpose of disclosure* and *Non-use of AI* categories within the details dimension.

To circulate our survey, we used a combination of purposive and snowball sampling through professional networks, university mailing lists, and departmental channels, targeting computer science researchers across AI, Systems, Theory and Interdisciplinary research. Participants span all 4 subdomains, career stages (<1 to >10 years), and countries (e.g., US, Germany, India). The full survey instrument and participant demographics are included in the appendix. Participants were compensated with an Amazon Gift Card worth \$10/ €10/ ₹500, depending on self-reported geographical location. The study received ethical approval from an institutional review board.

To estimate differences in necessity between different research tasks and levels of human involvement, we fit a linear mixed effects model with necessity rating as the target variable, research task and human involvement as fixed effects, and participant-level random effects. Since norms around AI disclosure are still in flux, we expect substantive individual variability in respondents' ratings. Linear mixed-effects models, which account for such variability through a participant-level random intercept, can isolate the fixed effects of research task and human involvement from this between-participant noise. This model can be formulated as:

$$y_{\text{target}} = \beta_0 + \beta_{\text{task}} \times x_{\text{task}} + \beta_{\text{human}} \times x_{\text{human}} + u_{\text{part}} \times x_{\text{part}} + \epsilon,$$

where y_{target} corresponds to the user-provided necessity rating, β_0 is the fixed intercept, x_{task} and x_{human} are indicator variables for research task and level of human involvement, and u_{part} is a participant-level random intercept.

5.2 Results

Overall, we observe that participants *find disclosing generative AI use moderately necessary*, with a mean necessity rating of 2.95 (95% confidence interval: [2.78, 3.13]). Through a subsequent research phase-wise analysis, we observe that AI

[4]https://aclrollingreview.org/cfp#authorship
[5]https://neurips.cc/Conferences/2025/LLM
[6]https://cmte.ieee.org/tcrts/conferences/transparency/

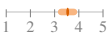
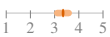
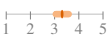
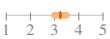
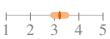
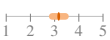
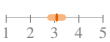
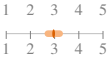
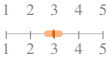
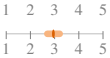
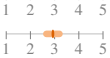
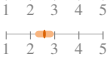
Research Task	Phase	Necessary?	Mean Rating
Generate synthetic data sets*	Data Collection	Usually	
Help develop theoretical models or conceptual frameworks*	Research Design	Usually	
Propose new hypotheses*	Idea Generation	Usually	
Help design research methodology or experiments*	Research Design	Sometimes	
Translate a research paper into a different language*	Writing & Reporting	Sometimes	
Clean and reformat dataset*	Data Collection	Sometimes	
Help pattern recognition in data*	Data Analysis	Sometimes	
Support qualitative and thematic data analysis*	Data Analysis	Sometimes	
Formulate questions for surveys or interviews	Data Collection	Sometimes	
Create or modify scientific figures or images	Data Analysis	Sometimes	
Suggest experimental parameters	Data Collection	Sometimes	
Create or edit software code	Data Analysis	Sometimes	
Draft parts of a research paper	Writing & Reporting	Sometimes	
Transcribe recordings of research material	Data Collection	Sometimes	
Summarize or analyse existing literature	Idea Generation	Sometimes	
Discover research topics or identify gaps*	Idea Generation	Sometimes	
Edit a research paper to improve readability*	Writing & Reporting	Rarely	
Identify relevant literature*	Idea Generation	Rarely	
Format references*	Writing & Reporting	Rarely	
Suggest a structure for a research paper*	Writing & Reporting	Rarely	
Propose a title or keywords for a research paper*	Writing & Reporting	Rarely	

Table 3: Research tasks ordered by mean disclosure-necessity rating. The third column reports the necessity of disclosure, by identifying the nearest option in the Likert scale, and the Mean Rating column visualizes each task's mean and 95% confidence interval. Significance markers ($^*p < .05$) indicate differences from the median-rated task (*Suggest experimental parameters*) in a linear mixed-effects model.

disclosure is the most necessary for tasks in the research design phase, and the least necessary for tasks in the writing and reporting phase (see Figure 2). Contrary to what we had expected given the emphasis on novel ideas in research, we find that disclosing generative AI assistance for tasks from the idea generation phase is considered relatively less necessary, with a mean rating of only 2.86.

We also observe a surprisingly high amount of variability in the perceived disclosure necessity across all phases and levels of human involvement. This effect could potentially be explained by individual tasks within the same phase receiving wildly different ratings. For instance, within the idea generation phase, *Proposing new hypotheses* (mean=3.54) and *Identifying relevant literature* (mean=2.46) differ by more than an entire point on the Likert scale, and fall on opposite sides of the median-rated task. Tasks within the data collection phase, specifically *Generating synthetic data sets* (mean=4.02) and *Transcribing recordings of research material* (mean=2.93), exhibit similar differences in ratings. Task-wise necessity ratings are available in Table 3, where we also observe that most writing and reporting tasks are clustered together at the lower end of the necessity scale.

Interestingly, across all research phases, shifts from the *assumed* level of human involvement exhibit a consistent trend: *low human involvement raises disclosure expectations*, (β=0.49, 95% CI=[0.42, 0.55], $p < 0.001$) while *high human involvement causes drops in the perceived necessity of disclosure* (β=-0.44, 95% CI=[-0.51, -0.38], $p < 0.001$). The shifts in either direction, are however not always similar. For instance, the median necessity ratings for the low and assumed human involvement conditions in the data analysis phase are 1 point apart, while this difference for the assumed and high human involvement conditions is less than 0.5. The direction of asymmetry is reversed in the idea generation and research design phases, with a 1-point drop in necessity ratings

from the assumed to the high human involvement conditions, but only a 0.5-point increase from the assumed level of human involvement to the low human involvement condition. This indicates that the loss of human agency in certain phases (e.g., in data analysis) drives disclosure necessity up, perhaps due to the potential errors that could arise in this situation, while the assurance of retained human agency for certain phases (e.g., in idea generation) lowers the need for disclosure.

Our findings complicate recommendations that outline the necessity of disclosing generative AI use based on a categorization of tasks, given the high variability in necessity ratings per phase. Resnik and Hosseini (2026) advocate for disclosure when AI-assistance either directly impacts research results or is used to generate or analyze content, data or images. While research tasks fitting this description are generally assigned higher necessity ratings by our participants, we highlight the observed variability, as illustrated by the difference in perceived necessity of disclosure for *Generating synthetic datasets* (mean=4.02) and *Creating or modifying scientific figures or images* (mean=3.00). Similarly, we show that while *Creating or editing software code* is stressed upon by ACM's guidelines,[7] this is considered less necessary by our participants (mean=2.96). Further, ACL's guidelines[4] that necessitate AI disclosure depending on the novelty of text or ideas produced, evoke similar mismatches, particularly for idea generation tasks, where participants find AI disclosure to be relatively less needed. Lastly, we highlight tasks that exhibit high perceived necessity, but cannot be neatly compartmentalized based on current policy guidelines: *Translating a research paper* (mean=3.31) and *Cleaning and reformatting data sets* (mean=3.24), both tasks that could introduce subtle erroneous artifacts if AI-assisted.

6 AI Disclosure Practices Audit

RQ3: How well do current AI disclosure practices align with reader expectations?

6.1 Method and Procedure

To answer RQ3, we extracted AI disclosure statements from 19525 research papers submitted to ICLR 2026 and 3216 research papers accepted to EMNLP 2025 Main and Findings, and analyzed the prevalence of tasks and details being disclosed. We chose to focus on these 2 specific conferences due to their established AI disclosure policies, their recency, and the open availability of research papers. Interestingly, the two conferences present a key policy difference: ICLR 2026 encouraged the inclusion of AI disclosure statements in a separate section within the submitted manuscript, while EMNLP 2025 allowed for generative AI use to be disclosed either in the main text or in the author submission checklist. This allows us to also comment on differing disclosure behavior due to differences in policy design.

We accessed accepted, withdrawn, rejected, and desk-rejected ICLR research papers through the OpenReview API, and EMNLP Main and Findings submission checklists and research papers from the ACL Anthology. We then used Gemini 2.5 Flash to evaluate the presence of and extract AI disclosure statements in research papers or submission checklists. We also used Gemini 2.5 Flash to annotate extracted AI disclosures with the details, tasks, and length present, in accordance with the taxonomies presented in Table 2 and Table 3. For extraction, Gemini achieved 100% F1 against human annotations on 100 papers, while on 100 disclosures annotated by three researchers, Gemini scored average micro F1 of 96.5% (details) and 90.6% (tasks). While these scores are not perfect, we believe they are satisfactory for our extraction and annotation purposes.

6.2 Results

Despite the lack of strict enforcement of AI disclosure policies, our analysis of 12577 extracted ICLR disclosures and 1290 extracted EMNLP disclosures reveals a moderately high prevalence of AI disclosures, with 64% of research papers submitted to ICLR and 40% of EMNLP Main and Findings research papers containing disclosure statements. We observe that tasks with lower perceived necessity are disclosed the most frequently: *Editing a research paper* (96.5% in ICLR and 81.4% in EMNLP) and *Creating or editing software code* (16.1% in ICLR and 23.3% in EMNLP). On the contrary, only a small fraction of disclosures exist for higher-rated tasks such as *Generating synthetic data sets* (2.1% in ICLR and 2.5% in EMNLP) and *Translating a research paper* (1.6% in ICLR and 2.0% in EMNLP). This gap may exist because generative AI use for tasks like polishing text or editing

[7] https://www.acm.org/publications/policies/new-acm-policy-on-authorship

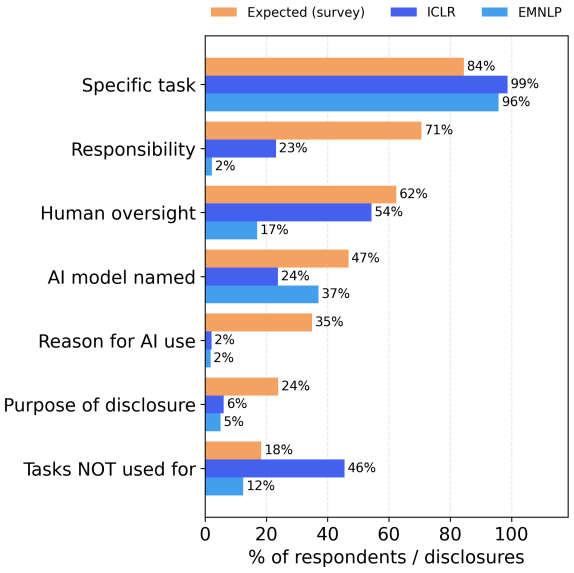


Figure 3: Expected vs. observed coverage of disclosure details, comparing survey expectations against ICLR and EMNLP practice.

Large Language Models (LLMs) were used to aid in the writing and polishing of the manuscript. Specifically, we used an LLM to assist in refining the language, improving readability, and ensuring clarity in various sections of the paper. The model helped with tasks such as sentence rephrasing, grammar checking, and enhancing the overall flow of the text.

It is important to note that the LLM was not involved in the ideation, research methodology, or experimental design. All research concepts, ideas, and analyses were developed and conducted by the authors. The contributions of the LLM were solely focused on improving the linguistic quality of the paper, with no involvement in the scientific content or data analysis.

The authors take full responsibility for the content of the manuscript, including any text generated or polished by the LLM. We have ensured that the LLM-generated text adheres to ethical guidelines and does not contribute to plagiarism or scientific misconduct.

Figure 4: An identical AI disclosure statement appearing in 95 ICLR 2026 submissions.

code is seen as more appropriate, and is therefore more common and more frequently disclosed.

Despite the structural difference between ICLR's free-text disclosure statements and EMNLP's checklist responses, we observe similar disclosure behavior in both venues, as shown in Figure 3. However, concerningly, while 71% of participants expect disclosures to contain a statement of responsibility from the authors, only 2% and 23% of EMNLP and ICLR disclosures respectively include such a declaration. It is also interesting to note that almost 50% of ICLR disclosures disclose the tasks generative AI was not used for, but this detail is considered the least necessary by participants. While authors may understandably include such information to emphasize the scope of AI assistance and their own contributions, perhaps this information is better included while expanding on the extent of human oversight, which is more frequently expected (62%). Moreover, surprisingly, less than 50% of participants exhibit a preference for the inclusion of the specific AI model used, perhaps due to the limited number of commercial models in widespread use, and their roughly equivalent capabilities.

We also find that while the general expectation for the length of an AI disclosure statement is *A few sentences or one short paragraph*, most disclosures in practice tend to be 1-2 sentences long. This prevalence of shorter disclosure statements (91%) is especially pronounced in EMNLP. This is in contrast to ICLR, where a sizeable chunk of disclosures (41%) are a few sentences long.

Alarmingly, during our analysis, *we uncover the existence of multiple duplicate disclosures*. One such disclosure statement (c.f., Figure 4) was found verbatim in 95 unique ICLR submissions.[8] While the statement itself is comprehensive and includes all expected details, its recurrence across unrelated papers suggests that the disclosure is not specific to any individual submission's actual AI use. Given the difficulty of enforcing disclosure policies, this highlights the potential issue of performative disclosures, intended to comply with policies rather than to inform.

7 Recommendations

Informed by our findings, we recommend that publishing venues specify *when* disclosure is needed, and standardize *what* a disclosure should contain.

Task-based AI disclosure policies The results of our survey show that disclosing AI use is not considered equally necessary across research tasks, and we also note that existing policy categories, at their current granularity, do not reflect these varying expectations. While we acknowledge that policies around AI disclosure are hard to formu-

[8]We initially suspected that this could be due to shared authors, but that was not the case. However, state-of-the-art AI text detector, Pangram, flags the statement as AI-generated.

In this work, we used generative AI tools for <some mandatory disclosure tasks>. We have not used generative AI tools for <other mandatory disclosure tasks>, and <the rest of mandatory disclosure tasks> are not applicable to this work. Additionally, we used generative AI tools for <some recommended disclosure tasks> [and optionally for <some optional disclosure tasks>].

We have reviewed all AI-assisted work [Elaborate. For example, "We checked LLM-generated research ideas for potential plagiarism through a manual literature survey", "LLM-generated code was verified and tested for correctness by 2 authors", etc.]. We take responsibility for the final content of this work, including text/claims/artifacts produced with the aid of generative AI.

Figure 5: A structured AI disclosure template, designed to capture details that survey respondents prioritize.

late precisely, we recommend that wherever possible, policies specify disclosure requirements at the level of individual tasks instead of broad rules. Such policies could include a list of tasks at differing levels of disclosure necessity: *mandatory disclosure* for tasks where disclosure is widely expected (e.g., *Proposing new hypotheses*), *recommended disclosure* for tasks where expectations are more diffuse (e.g., *Creating or editing software code*), and *optional disclosure* for tasks where disclosure is deemed less necessary (e.g., *Editing or polishing text*). We encourage policymakers at publishing venues to reproduce our survey, and use the resulting task-wise necessity ratings to inform the level of disclosure appropriate for different research tasks, for example, classifying tasks scoring ≥ 3.5 as *mandatory*, 2.5–3.5 as *recommended*, and the rest as *optional*.

Standardized AI disclosure formats Our analysis of AI disclosures in practice shows that current disclosure statements are rarely aligned with expectations, and may be performative. We speculate that this is in part due to the absence of a common format. We construct a boilerplate template from the details expected by over half of our survey participants, and recommend that publishing venues include it as an example in their policies and embed it directly within conference-specific LaTeX files. We also suggest this boilerplate template include the tasks requiring disclosure, as outlined in §5.2. We provide an example boilerplate disclosure statement in Figure 5, with task lists omitted for brevity, but included in Table 4. A disclosure within the paper primarily serves readers; we additionally recommend a dedicated field in author submission checklists for disclosing *mandatory* tasks, which would better serve reviewers and area chairs.

Research Tasks by Disclosure Level
Mandatory (rating ≥ 3.5)
Generate synthetic data sets
Develop theoretical models or frameworks
Propose new hypotheses
Recommended ($2.5 \leq$ rating < 3.5)
Design research methodology or experiments
Translate a research paper
Clean and reformat a dataset
Pattern recognition in data
Qualitative and thematic data analysis
Formulate survey or interview questions
Create or modify scientific figures
Suggest experimental parameters
Create or edit software code
Draft parts of a research paper
Transcribe research recordings
Summarize or analyze existing literature
Discover research topics or gaps
Optional (rating < 2.5)
Edit a paper for readability or language
Identify relevant literature
Format references
Suggest a structure for a paper
Propose a title or keywords

Table 4: Suggested disclosure levels for research tasks, based on ratings from survey participants.

8 Conclusion

In this work, we examined AI disclosure in computer science research through a policy analysis of 65 conferences, a survey of 109 researchers, and an audit of 13867 disclosure statements. We find that high variability in the perceived disclosure necessity across research tasks complicates the formulation of precise disclosure policies, and that current author practices diverge substantially from these expectations. Our findings offer empirical grounding to inform ongoing efforts to standardize and maintain transparency in research communication.

Limitations

Self-selection bias Our survey relied on convenience-based recruitment, and participants who chose to respond likely already held opinions about AI disclosure. Expectations from a broader, less self-selected population may be weaker or more diffuse than those we report.

Future work could mitigate this through stratified sampling across computer science subdomains or randomized recruitment via venue mailing lists.

Scope to CS Research Our findings are limited to computer science research, where AI tools are arguably most familiar and disclosure norms are evolving most rapidly. While this focus is deliberate given that disclosure expectations likely differ across disciplines, extending this characterization to other research communities is a natural direction for future work.

LLM Annotation Reliability We also make use of an LLM judge (Gemini 2.5 Flash) to annotate disclosure statements. Although we validated these annotations against human annotators with high agreement (average micro F1 of 96.5% (details) and 90.6% (tasks)), residual errors are possible, particularly for rare details or task categories with limited validation support.

Ethical Considerations

This work characterizes AI disclosure expectations through a survey with human participants. Informed consent was explicitly confirmed at the beginning of the survey, and respondents were given the option to stop participation at any point. The survey was completely anonymous, and providing demographic details was optional. Participants who completed the survey were compensated with Amazon Gift cards worth $10/ €10/ ₹500, depending on self-reported geographical location. We only collected contact details (such as e-mail) to dispense the gift cards, and this information was deleted immediately after confirmation of receipt of compensation. As these details were obtained through a separate form, linking contact details to survey responses (thus risking deanonymization), was not possible. This study received approval from an institutional review board.

Our work further recommends policy changes, and while these are made in the spirit of improving transparency and are grounded in community opinions, AI disclosures carry broader consequencess. For instance, researchers who disclose legitimate uses of AI assistance such as translation may inadvertently reveal that they are non-native English speakers, an attribute that, once inferred during peer review, risks reinforcing existing biases against them. To amend this, we suggest masking language support disclosures during the review period. More broadly, AI disclosures may invite negative perceptions of the work. While current norms around the appropriateness of AI use are still in flux, this cannot be fully eradicated. We nonetheless believe that standardized disclosure formats and processes from policymakers, along with authors' thoughtful efforts to emphasize the extent of their oversight and responsibility will help. On the other hand, disclosures may also occasionally benefit authors: because they are made a priori, they can serve as evidence for the actual extent of AI use when work is incorrectly flagged for violating AI use policies by venues relying on (potentially inaccurate) AI text detectors.

Acknowledgments

We are grateful to all participants who took part in our survey and (whenever possible) forwarded it. We thank Rounak Saha and Rose Sathyanathan for annotations to validate the LLM judge used to label AI disclosure statements. We also thank Shobini NS for providing feedback on initial versions of the survey, and Kirti Bhagat for thoughtful discussions on interpreting results. We are also grateful to Vani A for administrative support. This work was supported in part by Schmidt Sciences, and additionally, DP is grateful to Google, Microsoft Research, and the Indian Institute of Science for supporting his group's research.

AI Disclosure In this work, we used generative AI tools (Claude Code) to create and edit software code, specifically for parallelizing the AI disclosure statement extraction pipeline. All generated code was manually verified for correctness. We have not used generative AI tools for generating synthetic data sets, developing theoretical models/frameworks, or proposing new hypotheses. We take responsibility for the final content of this work, including the artifacts produced with the aid of generative AI.

References

Nil-Jana Akpinar, Sandeep Avula, Chia-Jung Lee, Brandon Dang, Kaza Razat, and Vanessa Murdock. 2026. Llm or human? perceptions of trust and quality in research summaries. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems*, CHI '26, New York, NY, USA. Association for Computing Machinery.

Jens Peter Andersen, Lise Degn, Rachel Fishberg, Ebbe K. Graversen, Serge P. J. M. Horbach, Evan-

thia Kalpazidou Schmidt, Jesper W. Schneider, and Mads P. Sørensen. 2025. Generative Artificial Intelligence (GenAI) in the research process – A survey of researchers' practices and perceptions. *Technology in Society*, 81:102813.

Ahmed BaHammam. 2025. The Transparency Paradox: Why Researchers Avoid Disclosing AI Assistance in Scientific Writing. *Nature and Science of Sleep*, Volume 17:2569–2574.

Jenna Butler, Jina Suh, Sankeerti Haniyur, and Constance Hadley. 2025. Dear Diary: A Randomized Controlled Trial of Generative AI Coding Tools in the Workplace. In *2025 IEEE/ACM 47th International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*, pages 319–329. ISSN: 2832-7659.

Program Chairs. 2023. ACL 2023 Policy on AI Writing Assistance.

Steffen Eger, Yong Cao, Jennifer D'Souza, Andreas Geiger, Christian Greisinger, Stephanie Gross, Yufang Hou, Brigitte Krenn, Anne Lauscher, Yizhi Li, Chenghua Lin, Nafise Sadat Moosavi, Wei Zhao, and Tristan Miller. 2026. Transforming Science with Large Language Models: A Survey on AI-assisted Scientific Discovery, Experimentation, Content Generation, and Evaluation. *arXiv preprint*. ArXiv:2502.05151 [cs.CL].

Jingchao Fang and Mina Lee. 2025. What Shapes Writers' Decisions to Disclose AI Use? *arXiv preprint*. ArXiv:2505.20727 [cs].

Conner Ganjavi, Michael B. Eppler, Asli Pekcan, Brett Biedermann, Andre Abreu, Gary S. Collins, Inderbir S. Gill, and Giovanni E. Cacciamani. 2023. Bibliometric analysis of publisher and journal instructions to authors on generative-ai in academic and scientific publishing. *Preprint*, arXiv:2307.11918.

Dany Haddad, Dan Bareket, Joseph Chee Chang, Jay DeYoung, Jena D. Hwang, Uri Katz, Mark Polak, Sangho Suh, Harshit Surana, Aryeh Tiktinsky, Shriya Atmakuri, Jonathan Bragg, Mike D'Arcy, Sergey Feldman, Amal Hassan-Ali, Rubén Lozano, Bodhisattwa Prasad Majumder, Charles McGrady, Amanpreet Singh, and 3 others. 2026. Understanding Usage and Engagement in AI-Powered Scientific Research Tools: The Asta Interaction Dataset. *arXiv preprint*. Version Number: 1.

Mohammad Hosseini, David B Resnik, and Kristi Holmes. 2023. The ethics of disclosing the use of artificial intelligence tools in writing scholarly manuscripts. *Research Ethics*, 19(4):449–465.

Nanna Inie, Jeanette Falk, and Steve Tanimoto. 2023. Designing Participatory AI: Creative Professionals' Worries and Expectations about Generative AI. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI EA '23, pages 1–8, New York, NY, USA. Association for Computing Machinery.

Shivani Kapania, Ruiyi Wang, Toby Jia-Jun Li, Tianshi Li, and Hong Shen. 2025. 'I'm Categorizing LLM as a Productivity Tool': Examining Ethics of LLM Use in HCI Research Practices. *Proc. ACM Hum.-Comput. Interact.*, 9(2):CSCW102:1–CSCW102:26.

Mohamed Khalifa and Mona Albadawy. 2024. Using artificial intelligence in academic writing and research: An essential productivity tool. *Computer Methods and Programs in Biomedicine Update*, 5:100145.

Diana Kwon. 2025. Is it OK for AI to write science papers? Nature survey shows researchers are split. *Nature*, 641(8063):574–578.

Jie Li, Hancheng Cao, Laura Lin, Youyang Hou, Ruihao Zhu, and Abdallah El Ali. 2024. User Experience Design Professionals' Perceptions of Generative Artificial Intelligence. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, pages 1–18, New York, NY, USA. Association for Computing Machinery.

Zhehui Liao, Maria Antoniak, Inyoung Cheong, Evie Yu-Yen Cheng, Ai-Heng Lee, Kyle Lo, Joseph Chee Chang, and Amy X. Zhang. 2024. Llms as research tools: A large scale survey of researchers' usage and perceptions. *Preprint*, arXiv:2411.05025.

Mahjabin Nahar, Sian Lee, Rebekah Guillen, and Dongwon Lee. 2025. Generative AI Policies under the Microscope: How CS Conferences Are Navigating the New Frontier in Scholarly Writing. *arXiv preprint*. ArXiv:2410.11977 [cs].

David B. Resnik and Mohammad Hosseini. 2026. Disclosing artificial intelligence use in scientific research and publication: When should disclosure be mandatory, optional, or unnecessary? *Accountability in Research*, 33(2):2481949.

Eva A. M. van Dis, Johan Bollen, Willem Zuidema, Robert van Rooij, and Claudi L. Bockting. 2023. ChatGPT: five priorities for research. *Nature*, 614(7947):224–226.

Richard Watermeyer, Donna Lanclos, Lawrie Phipps, Hanne Shapiro, Danielle Guizzo, and Cathryn Knight. 2025. Academics' Weak(ening) Resistance to Generative AI: The Cause and Cost of Prestige? *Postdigital Science and Education*, 7(4):1171–1191.

Kari Weaver. 2024. The artificial intelligence disclosure (aid) framework: An introduction. *College & Research Libraries News*, 85(10):407.

Abdullahi Yusuf, Samia Mouas, and Mohammad Hamad Al-khresheh. 2025. Understanding the Psychological Factors Associated with (non) Disclosure Behaviour after GenAI Usage. *Journal of Academic Ethics*, 24(1):34.

A Appendix

The contents of the Appendix are organized as follows: Table 5 outlines the prevalence of AI disclosure policies across computer science conferences, grouped by subdomain. Table 6 shows the mapping of conferences with established AI disclosure policies to corresponding societies, along with the number of revisions the society policies have undergone since their introduction. Table 7 details the taxonomy of research tasks we used, categorized by research phase. Table 8 contains examples of extracted disclosure statements, grouped by type of detail. Tables 9–14 include the full survey instrument. Table 15 contains survey participant demographic information. Tables 16–18 detail the prompts used for the extraction of AI disclosure statements. Table 19 and Table 20 show validation results of Gemini 2.5 Flash for annotating extracted disclosures. Tables 21–23 contain the specific annotation prompt used.

Area	Conference	n	Year	Disclosure
AI	NeurIPS, ACL, EMNLP, SIGIR, WWW	5	2026	Yes
	ICLR, NAACL	2	2025	Yes
	AAAI, IJCAI, CVPR, ECCV, ICML	5	2026	No
	ICCV	1	2025	No
Systems	ISCA, SIGCOMM, NSDI, CCS, IEEE S&P, SIGMOD, DAC, RTAS, HPDC, ICS, MobiSys, IMC, SIGMETRICS, FSE, ICSE	15	2026	Yes
	RTSS, SC	2	2025	Yes
	ASPLOS, USENIX Security, VLDB, EMSOFT, MobiCom, SenSys, OSDI, SOSP, PLDI, POPL	10	2026	No
	MICRO, ICCAD	2	2025	No
Theory	STOC, CRYPTO, EuroCrypt	3	2026	Yes
	SODA, CAV, LICS	3	2026	No
	FOCS	1	2025	No
Interdisciplinary	SIGGRAPH, SIGGRAPH Asia, SIGCSE, CHI, ICRA, VIS, VR	7	2026	Yes
	IROS	1	2025	Yes
	ISMB, RECOMB, Pervasive, UIST, RSS	5	2026	No
	EC, WINE, UbiComp	3	2025	No

Table 5: AI disclosure policy presence across major CS conferences.

Society	Conferences	n	#Changes
AAAI	–	0	1
ACL	ACL, EMNLP, NAACL	3	2
ACM	SIGIR, WWW, SIGCOMM, CCS, SIGMOD, DAC, HPDC, ICS, MobiSys, IMC, SIGMETRICS, FSE, ICSE, STOC, CRYPTO, EuroCrypt, SIGGRAPH, SIGGRAPH Asia, SIGCSE, CHI, ISCA	21	1
IEEE	ISCA, IEEE S&P, ICSE, ICRA, IROS, VIS, VR	7	4
None	ICLR, NeurIPS, NSDI, RTAS, RTSS, SC	6	–

Table 6: Conferences inspired by each society's disclosure policy, and the number of revisions made.

Phase	Task
Idea Generation	Discover research topics or identify gaps in current research
	Identify relevant literature
	Summarize or analyse existing literature
	Propose new hypotheses
Research Design	Help design research methodology or experiments
	Help develop theoretical models or conceptual frameworks
Data Collection	Suggest experimental parameters (e.g., choosing sample size, experimental conditions, or training settings)
	Formulate questions for surveys or interviews
	Transcribe recordings of research material (e.g. interviews, workshops or focus groups)
	Generate synthetic data sets
	Clean and reformat dataset
Data Analysis	Create or edit software code for data analysis, statistical analysis or simulations
	Support qualitative and thematic data analysis and coding
	Help pattern recognition in data
	Create or modify scientific figures or images
Writing and Reporting	Suggest a structure for a research paper
	Draft parts of a research paper
	Propose a title or keywords for a research paper
	Edit a research paper to improve readability and/or language
	Format references
	Translate a research paper into a different language

Table 7: Taxonomy of research phases and associated tasks.

Detail	Description	Example
Task	The specific task AI was used for.	"In writing this work, LLMs have been used in helping find relevant works, format LaTeX, and implement code, and proofread."
Model name	The AI model or system that was used.	"Specifically, we employed Claude Sonnet 4 (Anthropic) and GPT-5 (OpenAI) for language polishing and refinement purposes."
Reason	The reason AI was used instead of a non-AI method.	"We use large language models (LLMs) to support labor-intensive and mistake-prone work."
Non-use	Tasks for which AI was not used	" LLMs were not used for generating novel scientific ideas, experiments, or analyses."
Human oversight	The extent of human oversight or contributions during or after AI use.	"All outputs from LLMs were carefully reviewed, verified, and edited by the authors to ensure correctness and originality."
Responsibility	The authors' declaration of responsibility for outcomes of AI use, such as potential errors or biases.	"The authors retain full responsibility for the entire content, including any errors or inaccuracies."
Purpose of disclosure	The purpose of the disclosure itself (why the information is shared), such as for transparency or compliance.	"In accordance with ICLR guidelines, we disclose that Large Language Models (LLMs) were used during the preparation of this manuscript."

Table 8: Disclosure details, with the relevant portion of each example highlighted.

Welcome Screen
This survey aims to understand how readers and researchers perceive AI use in academic writing and research. You will be asked about your views on when and how AI use should be disclosed, and how such disclosures differ based on the research subtask and level of human involvement. Your responses will help inform better guidelines for transparency and ethical AI use in research communication. Participation is voluntary and anonymous, and the survey should take approximately **15–20** minutes to complete.

Table 9: Welcome screen of the survey instrument.

Introduction and Consent

1. **Purpose and Goal**
   An AI Disclosure Statement in a research paper is a brief description of how Generative AI tools (such as ChatGPT, Gemini, and Claude) were used by the authors during the entire research process. The inclusion of such statements is being gradually mandated by academic publishers, individual journals, and conferences, however with no consensus on the presence or framing across institutions. The purpose of this survey is to understand the opinions of academic readers on the importance of the presence and framing of AI disclosure statements, as well as their necessity across various research tasks with differing levels of human involvement. This information will then be used to arrive at framing recommendations for AI Disclosure Statements, and an AI Disclosure Checklist that aims to inform policy regarding AI use and disclosure at academic and publishing organizations.

2. **Procedure**
   After confirming the informed consent, you will be directed to answer a mixture of multiple-choice, Likert scale, and open-ended questions in the following sections: Presence and Framing of AI Disclosure Statements, Importance of AI Disclosure in Idea Generation, Research Design, Data Collection, Data Analysis, and Writing and Reporting. Then, you will be prompted to optionally share additional perspectives on the necessity of AI Disclosures and fill in some demographic information. Finally, you will be redirected to a separate Google Form to provide your e-mail address to receive compensation.

3. **Participation and Compensation**
   Your participation in this user study is completely voluntary. Compensation is available for participation in this study. If you opt in to be compensated, you will receive a compensation of $10/€10/ ₹500, depending on your geographical region, in the form of an Amazon Gift Card. You may withdraw and discontinue participation at any time without penalty. Repeated participation in the survey is not permitted.

4. **Confidentiality**
   All information gathered in this study is anonymous and will be kept completely confidential. No reference will be made in written or oral materials that could link you to this study. All records will be stored securely at [RETRACTED INSTITUTE NAME].

5. **Investigators**
   If you have any questions or concerns about this research, please feel free to contact [RETRACTED INVESTIGATOR DETAILS]

6. **Informed Consent and Agreement**
   ❑ I understand the explanation provided to me. I understand that this declaration of consent is revocable at any time. I have had all my questions answered to my satisfaction, and I voluntarily agree to participate in this survey.

Table 10: Introduction and Consent section of the survey instrument.

Presence and Framing of AI Disclosure Statements

1. Indicate how strongly you agree or disagree with the following statements.
   Scale: 1 = *Strongly disagree,* 2 = *Disagree,* 3 = *Neither Agree nor Disagree,* 4 = *Agree,* 5 = *Strongly agree*
   (a) I usually notice AI-related disclosure statements in the research papers I read.
   (b) I actively look for AI disclosure statements when reading research papers.
   (c) It is important to me that I am aware of the extent of AI usage by the authors of a research paper.
   (d) AI disclosure statements help me make an informed evaluation of the quality, credibility or originality of the research paper.

2. *[Depending on response to 1(c)]*
   (a) *If* >= 4*:* Why is it important for you to be aware of the extent of AI use by the authors of a research paper?
   (b) *If* == 3*:* When is it important for you to be aware of the extent of AI use by the authors of a research paper?
   (c) *If* <= 2*:* Why is it not important for you to be aware of the extent of AI use by the authors of a research paper?

3. In your opinion, where should AI disclosure statements be located in a research paper, if at all?
   ❏ At the beginning of a research paper
   ❏ In the methodology section of a research paper
   ❏ In a separate section at the end of the research paper
   ❏ In the acknowledgments section of a research paper
   ❏ Throughout the paper, in each section where AI played a role
   ❏ Outside the main text (e.g., cover sheet, landing page, or supplementary materials)
   ❏ I do not believe AI disclosure statements are necessary

4. Please describe the reasons behind your preferred locations for an AI disclosure statement.

5. In your opinion, what should an AI disclosure statement include, if present?
   ❏ The specific task AI was used for
   ❏ The AI model or system that was used
   ❏ The reason AI was used instead of a non-AI method
   ❏ Tasks for which AI was not used
   ❏ The extent of human oversight or contributions during or after AI use
   ❏ The authors' declaration of responsibility for outcomes of AI use, such as potential errors or biases
   ❏ The purpose of the disclosure itself (why the information is shared), such as for transparency or compliance
   ❏ I do not believe AI disclosure statements are necessary

6. In addition, what level of detail, if any, is appropriate for an AI disclosure statement in a research paper?
   ❍ One to two sentences
   ❍ A few sentences or one short paragraph
   ❍ Multiple paragraphs
   ❍ Length and detail proportionate to the extent of AI use (this could potentially span multiple paragraphs)
   ❍ I do not believe AI disclosure statements are necessary

Table 11: Presence and Framing of AI Disclosure Statements section of the survey instrument.

Importance of AI Disclosure in Idea Generation

Before proceeding, please take a moment to review the tasks below. If you believe that none of the tasks below require AI disclosure, regardless of the level of human involvement, select the option below to skip this section. We'll record your answers as 'need not be disclosed.' If even one task or condition gives you pause, leave this unselected and answer the questions below instead.
❏ None of the tasks below require AI disclosure, regardless of human involvement.

1. For each of the following idea generation tasks, please indicate how necessary you think AI disclosure is.
   Scale: 1 = *Need not be disclosed,* 2 = *Disclosure is rarely needed,* 3 = *Disclosure is sometimes needed,* 4 = *Disclosure is usually needed,* 5 = *Disclosure is always needed*
   (a) Discover research topics or identify gaps in current research
   (b) Identify relevant literature
   (c) Summarize or analyse existing literature
   (d) Propose new hypotheses
2. For each of the following idea generation tasks, if human involvement is high, please indicate how necessary you think AI disclosure is.
   High human involvement is when the author provides most of the contribution and initiative, while AI outputs are closely supervised and verified.
   Scale: 1 = *Need not be disclosed,* 2 = *Disclosure is rarely needed,* 3 = *Disclosure is sometimes needed,* 4 = *Disclosure is usually needed,* 5 = *Disclosure is always needed*
   (a) Discover research topics or identify gaps in current research
   (b) Identify relevant literature
   (c) Summarize or analyse existing literature
   (d) Propose new hypotheses
3. For each of the following idea generation tasks, if human involvement is low, please indicate how necessary you think AI disclosure is.
   Low human involvement is when the AI assistant provides most of the contribution and initiative, and AI outputs are loosely supervised and verified by the author.
   Scale: 1 = *Need not be disclosed,* 2 = *Disclosure is rarely needed,* 3 = *Disclosure is sometimes needed,* 4 = *Disclosure is usually needed,* 5 = *Disclosure is always needed*
   (a) Discover research topics or identify gaps in current research
   (b) Identify relevant literature
   (c) Summarize or analyse existing literature
   (d) Propose new hypotheses

Table 12: Importance of AI Disclosure in Idea Generation section of the survey instrument. All subsequent phases follow the same template, and are thus excluded for brevity.

Additional Views on AI Disclosure Necessity

1. Are there any other research tasks for which you think AI disclosure is necessary?
2. In general, how do you think the level of human involvement affects whether AI use should be disclosed across different research tasks?

Table 13: Additional Views on AI Disclosure Necessity section of the survey instrument.

Demographics

1. What is your age?

2. What gender(s) do you identify with the most?
   ❏ Female ❏ Male ❏ Non-binary ❏ Prefer not to disclose

3. Which country are you currently based in for your primary work or study?

4. What is your current professional or academic affiliation?
   ❍ Academia ❍ Industry

5. *[Depending on response to 4]* What is your position at your current institution?

   (a) *If Academia:*
   ❍ Undergraduate Student ❍ Master's Student ❍ PhD Student ❍ Postdoctoral Researcher
   ❍ Research Associate ❍ Assistant Professor ❍ Associate Professor ❍ Professor
   ❍ Prefer not to answer ❍ Other

   (b) *If Industry:*
   ❍ Research Associate ❍ Junior Research Scientist ❍ Senior Research Scientist
   ❍ Principal Research Scientist ❍ Prefer not to answer ❍ Other

6. What is your major field of study/research?
   ❍ Artificial Intelligence
   ❍ Systems
   ❍ Theory
   ❍ Other / Interdisciplinary Areas

7. How many years of research experience do you have?
   ❍ <1 year ❍ 1-3 years ❍ 4-7 years ❍ 8-10 years ❍ >10 years

8. How many years of experience do you have with reviewing research papers for peer review?
   ❍ <1 year ❍ 1-3 years ❍ 4-7 years ❍ 8-10 years ❍ >10 years

9. How frequently do you read research papers?
   ❍ Never ❍ Occasionally ❍ Monthly ❍ Weekly ❍ Daily

10. In your research field, how common is the use of AI tools in research workflows?
    Scale: 1 = *Very uncommon,* 2 = *Somewhat uncommon,* 3 = *Neither common nor uncommon,* 4 = *Somewhat common,* 5 = *Very common*

11. For each of the following research phases, please indicate whether you have used and/or disclosed AI assistance in any of your past research papers.
    We'd like to reiterate that responses are anonymous.
    Options: ❍ *Didn't use AI assistance* ❍ *Used and disclosed AI assistance* ❍ *Used but didn't disclose AI assistance*

    (a) Idea Generation
    (b) Research Design
    (c) Data Collection
    (d) Data Analysis
    (e) Writing and Reporting

12. Indicate how strongly you agree or disagree with the following statement: I am comfortable with the increasing integration of AI tools into research workflows.
    Scale: 1 = *Strongly disagree,* 2 = *Disagreee,* 3 = *Neither agree nor disagree,* 4 = *Agree,* 5 = *Strongly agree*

13. *[Depending on response to 12]*

    (a) *If <=2:* What aspects of the increasing integration of AI tools into research workflows makes you uncomfortable?

    (b) *If ==3:* In what situations are you comfortable with the integration of AI tools into research workflows, and in what situations are you not?

    (c) *If >=4:* Are there some specific reasons behind your stance on the integration of AI tools into research workflows, or why you feel this way?

Table 14: Demographics section of the survey instrument.

Demographic	Level	n
Gender	Female	36
	Male	67
	Non-binary	1
	Prefer not to disclose	5
Country of work / residence	India	46
	United States	40
	Germany	7
	Canada	5
	United Arab Emirates	4
	Israel	1
Affiliation	Academia	104
	Industry	3
	Other	2
Field of study	Artificial Intelligence	52
	Systems	13
	Theory	12
	Other / Interdisciplinary Areas	32
Years of research experience	<1 years	22
	1–3 years	48
	4-7 years	21
	8-10 years	11
	>10 years	7
Years of review experience	<1 years	56
	1–3 years	29
	4-7 years	15
	8-10 years	5
	>10 years	4

Table 15: Demographics of survey respondents ($N = 109$). Mean age: 26.64 years (SD = 6.28). Counts in the n column are unweighted respondent totals; categories within each demographic are mutually exclusive.

ICLR Disclosure Extraction Prompt

You are extracting AI/LLM usage disclosure statements from academic papers.

Your task: find a dedicated section where the authors disclose how they used AI tools or Large Language Models (LLMs) in **preparing or writing the manuscript** — for example, grammar checking, text polishing, writing assistance, or code editing.

IMPORTANT DISTINCTIONS: - DO extract: sections about LLMs used for manuscript preparation (writing, editing, grammar, coding assistance for the paper itself). - DO extract: negative disclosures where authors explicitly state they did NOT use LLMs. - DO NOT extract: inline mentions of LLM use embedded within Methods, Experiments, or Results sections that are not part of a dedicated disclosure statement or paragraph.

Where to look: - This section almost always appears near the END of the paper (after Conclusions, near Acknowledgments or References). - Common headings include: "LLM Usage", "Use of Large Language Models", "AI Disclosure", "Declaration of LLM Usage", "LLM Usage Statement", etc. - It may also be embedded inside an "Ethics Statement" or "Acknowledgments" section. - Occasionally there is no heading — look for a standalone paragraph near the end of the paper that explicitly addresses AI tool usage in manuscript preparation.

Return a JSON object with: - "found": true if a disclosure section or paragraph was found, false otherwise - "heading": the exact heading of the section (empty string if embedded or no heading) - "content": the full text content of the disclosure (empty string if not found)

Return ONLY valid JSON.

Table 16: Prompt used for extracting ICLR disclosures.

EMNLP Disclosure Extraction Prompt 1

You are reading the "Responsible NLP Checklist" of an EMNLP paper.

The checklist has a legend at the top that explains the symbols: ✓ (checkmark / tick) = the authors responded YES ✗ (cross / X) = the authors responded NO N/A = does not apply empty box = no response

These two symbols are DIFFERENT and must not be confused: ✓ = YES (a tick, checkmark, or filled checkmark — the answer is affirmative) ✗ = NO (a cross, X, or filled X — the answer is negative)

Find section E, which concerns AI assistant use: E. Did you use AI assistants (e.g., ChatGPT, Copilot) in your research, coding, or writing? E1. If you used AI assistants, did you include information about their use?

Your task:

1. Look at the symbol next to question E (not E1). Read it carefully. - If it is ✗ (cross/X/NO) or N/A: set ai_used = false and stop. - If it is ✓ (tick/checkmark/YES): set ai_used = true and continue.

2. Read the answer text written next to or below question E1 (do NOT rely on E1's symbol — it varies across submissions and is unreliable). Instead, classify by the text itself:

- "paper_section": the text is a location pointer — either a bare pointer (just a section number, heading name, appendix letter, or page, e.g. "Section 5", "Ethics Statement", "Appendix A", "page 8", "see §4") OR a wrapper sentence that only says the disclosure IS in a section without describing WHAT AI was used for (e.g. "We include information in the Acknowledgement section", "We disclosed AI usage in Appendix G", "In the Ethics Statement we specified the AI usage", "Yes we include the use of AI assistants in Section 3", "We mention the usage in pre-appendix", "We have detailed how to use it in Appx. G"). Key test: does the text tell you WHAT AI was used for? If no — it is paper_section. → set disclosure_location = "paper_section", copy the text to section_reference.

- "inline": the text describes WHAT AI was used for — it contains actual content about the usage, not just a pointer to where it is described. It may also mention a section, but the description itself is informative without needing to look elsewhere (e.g. "Used for grammar correction", "We used ChatGPT only for writing polish", "Used for machine translation. We outline in Section 4.6."). → set disclosure_location = "inline", copy the full text to content.

- "none": E1 has no meaningful answer text (blank, just "N/A", bare "yes"/"no"). → set disclosure_location = "none".

Return a JSON object with exactly these fields: - "ai_used": true or false - "disclosure_location": one of "inline", "paper_section", "none" - "content": the inline disclosure text if disclosure_location is "inline", else "" - "section_reference": the section/page pointer if disclosure_location is "paper_section", else "" - "e_symbol_seen": the exact symbol you saw next to question E (e.g. "✓", "✗", "N/A", "empty")

Return ONLY valid JSON.

Table 17: First-stage prompt for the two-stage EMNLP extraction pipeline. Reads the Responsible NLP Checklist (Section E and E1) and returns AI-use status, disclosure type (inline vs. paper-section pointer), and the corresponding content or section reference.

EMNLP Disclosure Extraction Prompt 2

You are extracting an AI usage disclosure from an EMNLP paper.

The authors' checklist section E1 points to this location: "section_reference"

Locate that section, appendix, or page. Your task is to determine whether it contains a GENUINE AI usage disclosure and, if so, extract ONLY the AI disclosure portion.

A GENUINE disclosure is one where the authors explicitly address their use (or non-use) of AI tools in the context of THIS paper — for example: - Using LLMs or AI assistants to write, edit, polish, proofread, or translate the manuscript - Using AI coding assistants (Copilot, Cursor) to write code for the paper - Explicitly stating they did NOT use any AI tools - Any deliberate statement of AI usage made to inform readers about how the paper was produced

NOT genuine (set found = false): - Sections that only describe LLMs as the subject of study or research methodology (e.g. "we used GPT-4 to generate our dataset", "we evaluated ChatGPT on our benchmark") with no separate statement about manuscript preparation - Generic ethics or limitations sections with no mention of AI tool use at all - Sections that do not exist in the paper

IMPORTANT — partial extraction: The referenced section may be a broader section (Ethics Statement, Acknowledgments, Limitations, etc.) that contains other content unrelated to AI tool usage. In that case, extract ONLY the sentences or paragraph that specifically address AI/LLM usage. Do NOT copy the entire section — only the AI disclosure part. Set "heading" to the section's actual heading even if only part of it is extracted.

Return a JSON object with: - "found": true if a genuine AI usage disclosure was found, false otherwise - "heading": the exact section heading where the disclosure appears (empty string if no heading) - "content": ONLY the AI disclosure text, not the full surrounding section

Return ONLY valid JSON.

Table 18: Second-stage prompt for the two-stage EMNLP extraction pipeline. Reads the referenced section of the paper PDF and returns whether a genuine AI usage disclosure is present, the section heading, and the disclosure text itself.

Annotator	**Details**			**Tasks**			**Length Acc.**
	P	**R**	**F1**	**P**	**R**	**F1**	
Annotator 1	98.2	97.4	97.8	91.8	93.7	92.7	90.0
Annotator 2	98.2	95.7	96.9	88.4	90.2	89.3	86.0
Annotator 3	91.1	98.6	94.7	91.1	88.7	89.9	87.0
Average	*95.8*	*97.2*	*96.5*	*90.4*	*90.9*	*90.6*	*87.7*

Table 19: Validation of LLM-based annotation against 3 human annotators on $N = 100$ disclosures. We report micro precision (P), recall (R), and F1 for the multi-label **Details** and **Tasks** schemas, and classification accuracy for the single-label **Length** schema. All values are percentages; numbers compare each human annotator's labels to Gemini's annotations.

Annotator A	**Annotator B**	**Details**			**Tasks**			**Length Acc.**
		P	**R**	**F1**	**P**	**R**	**F1**	
Annotator 1	Annotator 2	94.8	96.5	95.6	90.2	90.2	90.2	81.0
Annotator 1	Annotator 3	98.1	89.9	93.8	87.3	91.6	89.4	90.0
Annotator 2	Annotator 3	98.6	88.7	93.4	84.7	88.8	86.7	80.0
Average		*97.2*	*91.7*	*94.3*	*87.4*	*90.2*	*88.8*	*83.7*

Table 20: Pairwise inter-annotator agreement on $N = 100$ disclosures. We report micro precision (P), recall (R), and F1 for the multi-label **Details** and **Tasks** schemas, and classification accuracy for the single-label **Length** schema. All values are percentages.

Disclosure Annotation Prompt (1/3)

You are an expert annotator of AI usage disclosure statements in academic papers.

Given a disclosure statement, return a JSON object with exactly these three keys:

"details" — which types of information the disclosure contains. Include only keys that are explicitly present; do not infer.

"task": The disclosure names at least one specific task AI was used for.

"model": A specific AI model or tool is named (e.g. ChatGPT, GPT-4, Claude, Gemini, Copilot). NOT for generic references like 'an LLM'.

"reason": The disclosure explicitly states WHY AI was used instead of a non-AI approach — a justification clause, not just a description of what was done. Rare.

"non_use": The disclosure explicitly states tasks for which AI was NOT used.

"human_oversight": Authors explicitly state their own independent contributions ("all research ideas were ours", "all analysis was done by the authors"), OR explicitly describe reviewing/verifying/editing AI outputs before inclusion.

"responsibility": Authors make a formal declaration of accountability for the content (e.g. 'take full responsibility', 'authors are responsible for').

"purpose": The disclosure explicitly cites a reason for making the disclosure itself — conference policy, ethics guidelines, compliance.

Disambiguation:
- "model" requires a specific name (ChatGPT, GPT-4, Gemini 2.5, Copilot, Grammarly...). Generic phrases like "a large language model" or "an LLM" do NOT trigger this. If a specific model is named only in the context of what was NOT used (e.g. "we did not use GPT or DeepSeek"), do NOT label "model".

- "human_oversight" vs "responsibility": oversight requires an EXPLICIT statement — either about what the authors themselves contributed ("all research ideas were ours", "all analysis was done by the authors") or about reviewing/verifying AI outputs ("we reviewed all LLM-generated text", "authors verified and edited all suggestions"). Do NOT infer oversight from the fact that AI was used only for minor tasks (e.g. "we only used LLMs for writing polish" alone does not imply human_oversight). "Taking full responsibility" is responsibility, not oversight. Both can appear together.

- "non_use" requires an explicit negation statement ("not used for", "were not involved in", "did not use AI for"). Saying AI was "only" or "solely" used for X does NOT by itself imply non_use — it just scopes the use. Only label non_use when there is an explicit statement that AI was NOT used for something (e.g. "LLMs were not used for data analysis", "AI was not involved in experiment design"). Phrases like "all research ideas/contributions are our own" or "all analysis was done by the authors" are authorship claims (human_oversight), NOT non_use.

- "purpose" requires citing a specific external trigger for the disclosure itself ("in accordance with ICLR 2026 policy", "as required by the Code of Ethics"). Do not infer purpose from the existence of the disclosure alone.

- "reason" is rare — only include if the disclosure gives a forward-looking justification for WHY AI was chosen (e.g. "we used LLMs because manual annotation at this scale was not feasible", "AI was used to accelerate repetitive processing"). Statements about what AI did or did not affect ("these uses do not influence our results", "LLMs were not involved in the core methodology") are scope/impact statements, not reasons.

Table 21: Annotation prompt (1/3): defines the seven detail labels (*task*, *model*, *reason*, *non_use*, *human_oversight*, *responsibility*, *purpose*) and the disambiguation rules.

Disclosure Annotation Prompt (2/3)

"tasks" — which specific tasks AI was used FOR. Include only tasks the disclosure explicitly mentions AI performing.

"edit_clarity": Edit, polish, or improve EXISTING text for grammar, clarity, style, or readability. Includes proofreading, sentence refinement, spell-checking, paraphrasing. Most common task.
"draft_paper": Generate NEW text or draft sections/paragraphs from scratch. Distinct from edit_clarity: drafting = creating new content; editing = improving existing content.
"paper_structure": Suggest or plan the structure or outline of the paper, OR adjust the layout and presentation of existing figures, tables, or sections (e.g. "suggesting formatting adjustments for tables", "adjusting the layout of figures and tables", "restructuring sections for logical flow").
"title_keywords": Propose a title or keywords for the paper.
"translate": Translate text between languages.
"format_references": Format or manage citations and references.
"discover_topics": Identify research gaps or brainstorm research directions.
"identify_literature": Search for, find, or suggest relevant papers and references.
"summarize_literature": Summarize or synthesize existing papers.
"propose_hypotheses": Propose new research hypotheses.
"design_methodology": Help design research methodology or experiments.
"develop_models": Help develop theoretical models or conceptual frameworks.
"suggest_parameters": Suggest experimental parameters or hyperparameters.
"formulate_questions": Formulate questions for surveys or interviews.
"synthetic_data": Generate synthetic or artificial datasets.
"clean_data": Clean, reformat, or preprocess datasets.
"transcribe": Transcribe recordings or research material.
"qualitative_analysis": Support qualitative coding, thematic analysis, or annotation.
"pattern_recognition": Help detect patterns in data.
"software_code": Write, edit, debug, or complete code for analysis, simulations, or document preparation. Includes writing or editing LaTeX commands/macros/syntax, plotting scripts, and code assistants like Copilot or Cursor. Does NOT include adjusting the visual layout of figures and tables (use "paper_structure") or general text formatting (use "edit_clarity").
"figures_images": Generate or create new scientific figures, plots, diagrams, schematics, or visual content. Use this when AI produces the visual content itself — not when it adjusts the layout or presentation of existing figures/tables (use "paper_structure" for that).
"other": AI was clearly used for a task not covered above. Use sparingly — only when the use case is unambiguous but genuinely outside all categories (e.g. used as automated judge in experiments, acting as a research subject).

Disambiguation:
- "edit_clarity" vs "draft_paper": if the text says AI "polished", "refined", "improved", "corrected", "proofread" → edit_clarity. If it says AI "wrote", "drafted", "rewrote sections", "generated initial text", "produced paragraphs" → draft_paper. A disclosure can include both.

- "figures_images" vs "paper_structure": use "figures_images" when AI generates or creates new visual content (plots, diagrams, schematics, figure captions as text). Use "paper_structure" when AI adjusts, reformats, or suggests changes to the layout or presentation of existing figures and tables (e.g. "formatting adjustments for tables and figures", "adjusting the layout and presentation of figures and tables").

- "software_code" vs "edit_clarity": LaTeX command writing/editing (macros, environments, syntax) → software_code. General wording like "LaTeX formatting" or "LaTeX styling" that refers to how text looks (not actual code) → edit_clarity. When in doubt, if the phrase implies writing or correcting actual code/commands → software_code; if it just means making the text look right → edit_clarity.

- "identify_literature" = searching for / finding papers. "summarize_literature" = synthesizing or summarizing papers already found.

- "other": use only when AI is clearly used for something that fits no category above, e.g. acting as an automated evaluator/judge in experiments.

Table 22: Annotation prompt (2/3): defines the task labels (e.g., *edit_clarity*, *draft_paper*, *software_code*, *synthetic_data*) and disambiguation rules for borderline cases.

Disclosure Annotation Prompt (3/3)

"length" — total length of the AI disclosure portion, exactly one of:
"1-2 sentences" — exactly 1 or 2 sentences. Count carefully: if there are 3 or more sentences, this label does not apply.
"few sentences/1 paragraph" — exactly 3, 4, or 5 sentences, or one coherent paragraph of similar length
"many paragraphs" — 2 or more paragraphs, OR a bulleted/numbered list with 3 or more distinct items, OR 6 or more sentences
Count sentences explicitly before deciding.

Rules:
- Only annotate what is explicitly stated — do not infer or extrapolate.
- For tasks: omit any task not directly described as something AI was used for.
- If the disclosure only states what AI was NOT used for and describes no actual use, "tasks" must be empty and "task" must not appear in "details".
- Return empty lists [] if nothing applies.
- Return ONLY valid JSON, no explanation.
- The disclosure may be embedded within a broader section (e.g. ethics statement, acknowledgments, limitations) that contains unrelated content. Focus only on the portion that describes AI/LLM usage. In particular, "length" should reflect only the length of the AI disclosure portion, not the entire surrounding section.

EXAMPLES

Example 1 — short, editing only, explicit authorship claim:
Disclosure: "We used LLMs to assist with grammar and writing polishing. All equations, analysis, and research contributions are our own." Output: "details": ["task", "human_oversight"], "tasks": ["edit_clarity"], "length": "1-2 sentences"

Example 2 — one paragraph, multiple details including explicit oversight:
Disclosure: "We used LLMs (specifically, OpenAI GPT-4.1, GPT-5 and Google Gemini 2.5) solely for checking grammar errors and improving the readability of the manuscript. The LLMs were not involved in research ideation, the development of research contributions, experiment design, data analysis, or interpretation of results. All substantive content and scientific claims were created entirely by the authors. The authors have reviewed all LLM-assisted text to ensure accuracy and originality, and take full responsibility for the contents of the paper." Output: "details": ["task", "model", "non_use", "human_oversight", "responsibility"], "tasks": ["edit_clarity"], "length": "few sentences/1 paragraph"

Example 3 — multiple tasks, explicit authorship claim:
Disclosure: "Large language models (LLMs) were used to revise sentences and correct grammar, to generate visualization code for some figures, and to assist with the implementation of the MULTIMNIST dataset. All conceptual contributions, experiment design, analysis, and the writing were done by the authors." Output: "details": ["task", "human_oversight"], "tasks": ["edit_clarity", "software_code", "figures_images"], "length": "few sentences/1 paragraph"

Example 4 — drafting + editing + compliance purpose + responsibility, no explicit review of outputs:
Disclosure: "In this work, large language models served as an assistive tool in the research and writing process. The model's contributions included assisting with drafting and iteratively revising the manuscript — such as generating initial text and rephrasing sentences to improve clarity — and helping to write and debug Python code for the data analysis pipeline. In accordance with ICLR 2026 policy, we disclose this use. The authors take full responsibility for all content." Output: "details": ["task", "responsibility", "purpose"], "tasks": ["draft_paper", "edit_clarity", "software_code"], "length": "few sentences/1 paragraph"

Example 5 — pure non-use, no actual use described (tasks must be empty):
Disclosure: "This paper does not use LLMs for research ideation or paper writing." Output: "details": ["non_use"], "tasks": [], "length": "1-2 sentences"

Example 6 — model named only in non-use context (do NOT label "model"):
Disclosure: "This work does not involve the use of any large language models (LLMs), such as GPT, DeepSeek, or similar models." Output: "details": ["non_use"], "tasks": [], "length": "1-2 sentences"

Example 7 — explicit human oversight (review process described):
Disclosure: "LLMs were extensively used as annotation helpers, as analytical assistants for reasoning trace analysis, and as automated judges. They contributed to the writing process by generating initial drafts and assisting with revision. All LLM-generated content underwent rigorous human review and validation, with human authors verifying analyses and thoroughly editing all contributions. While LLMs served as powerful assistive tools, all final decisions were made by human authors." Output: "details": ["task", "human_oversight"], "tasks": ["qualitative_analysis", "draft_paper", "edit_clarity", "other"], "length": "many paragraphs"

Table 23: Annotation prompt (3/3): defines the length categories, general output rules, and worked examples for the model to follow.