Large Language Label

Ekanem Okeke
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA 02139
eokeke@mit.edu

Alex Stewart
Department of Architecture
Massachusetts Institute of Technology
Cambridge, MA 02139
alexstew@mit.edu

Kyrylo Beskorovainyi
Harvard Nieman Fellowship
Harvard University
Cambridge, MA 02139
kyrylob@mit.edu

Abstract

The label is often the first place one looks to make a judgement about something. However, labels can also be places of both information scarcity and overload, limiting their efficacy, and forcing people to do their own research. What new understandings about our environment could be found by simply including the appropriate and relevant key information right there on the label? Large Language Label (LLL) enables people to evaluate their surroundings through a new lens by highlighting the benefits and negative consequences of objects, locations, and even people. LLL also gives agency to the user, allowing them to select the lens to evaluate the object through, and generates a physical label for it, a permanent marker of information about the object. Ultimately this project poses the following questions: what don't people consider about the things around them, and how can labels shape better understandings about the world around us.

1 Project motivation

Labels have long been utilized to help users make informed decisions, most commonly in the form of nutrition facts and pharmaceutical labels. However, their effectiveness can vary widely. For example, pharmaceuticals often feature "laundry lists" of side effects that make it difficult for drug patients to effectively determine risk and/or use medical terms that the broader public does not understand [Schwartz and Woloshin, 2013]. Ultimately, there is a gap between the intent and impact of these labels, limiting the patient's ability to analyze risk and thus reducing their agency.

Large Language Label is an innovative project exploring the use of AI in facilitating risk-benefit analysis, leveraging the ability of LLMs to synthesize information designed to facilitate decision-making by providing users with AI-powered label creation for risk-benefit analysis. Each label features the three most relevant benefits and side effects of the chosen item. By enabling individuals to create concise, context-specific labels, LLL addresses the challenges of information overload and miscommunication that often accompany complex decision-making scenarios. With its straightforward process, users can simply snap a photo, select what they wish to label, choose an appropriate analysis scope (environmental, social, emotional, or physical), define their audience (children, scholars, etc.), and print their label, bringing information from the digital space into the real world.

As highlighted in the paper "Large Language Objects: The Design of Physical AI and Generative Experiences", AI and LLMs are limited in their impact due to their inherent separation from the

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Creative AI Track.

physical world [Coelho and Labrune, 2024]. As Coelho and Labrune suggest, this calls for a new class of objects that can bring AI into the real world, known as "Large Language Objects" [Coelho and Labrune, 2024]. With this in mind, using LLL results in a printed physical label that users can stick to the object itself. This invites users to share their label with others and gives the object a permanent context. As labels are a familiar low-tech solution to informing the public, the design leans on retro aesthetics of vintage game controller design, using simple interfaces and tactile inputs.

Beyond simply creating more comprehensive and effective labels, we saw an opportunity with LLL to encourage users to see objects through a different lens. Nutrition facts labels are designed to help people consider how consuming products can affect their physical bodies, but what if one wants to consider the environmental impact of purchasing cow's milk, or how your lunch choices could affect your social life at work? Leveraging the use of Large Language Models, LLL can draw on cultural contexts and emotional reasoning, and communicate the variety of ways one can perceive an object.

As such, Large Language Label draws out subconscious cultural norms and biases. Any misinformation, bias, or stereotype present within the LLM can be exposed in the label, forcing the user to contend with stereotypes they feel and/or disagree with. As an example, a sample label for the social impact of ballet shoes listed concerns of being viewed as elitist, an idea that not everyone may agree with. Moreover, as the labels are generated in real-time by an LLM, each label is effectively a snapshot in time. Relabeling over time can reveal how perspectives and available information have changed, providing further insight into human values and the ways that we evaluate our surroundings.

2 Technical description

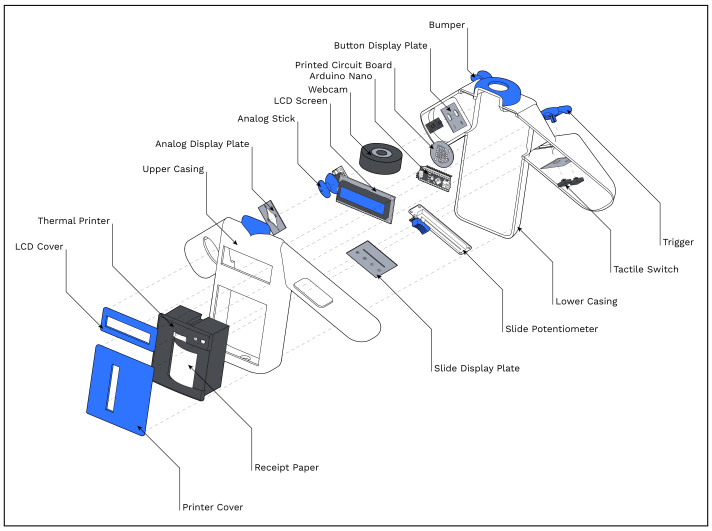

Figure 1: Exploded view of the components in Large Language Label

At the center of the Large Language Label tool is an Arduino Nano microcontroller, which handles the signals from all of the physical inputs (buttons, joystick, slide potentiometer, rotary encoders, and a thermal printer module). A computer is used to interface with the OpenAI API, and temporarily stores photographs. As such, the code features multiple USB serial handoffs between the Arduino Nano and a computer. ChatGPT is used for object detection in the photographs and generates the list of benefits and side effects on each label. At the time of development, the GPT-4 series was the most advanced ChatGPT models available, and 4o-mini was chosen for its speed. However, the computing power necessary still limits the portability, and there is a latency in generating labels, which we seek to improve.

The process begins when the user presses the capture button on the device. This action initiates communication between the Arduino and the computer, which triggers the camera to take a JPEG photo and overwrite any previous photographs. The photo is base64-encoded on the computer and sent to ChatGPT for object detection with the fixed prompt, "Provide a list of up to 10 of the most prominently featured things/people/places in the photo, including brand names/proper names if

visible/identifiable. Each list item should be no more than 28 characters, including spaces." This character limit is based on the space available on the LCD screen. Although other models such as Yolov5 were considered for object detection, ChatGPT was found to provide more granular details (i.e., colors, brands, etc.) in our experience. Using ChatGPT, Large Language Label is able to identify popular brands, decipher text, and identify celebrities, providing more specific answers.

After selecting the desired item, the user uses the joystick and slider to set the label's parameters. When the user presses print, their parameters are compiled into a fixed prompt template: "Provide a label for '[word]' from a [mode] perspective, so everything should relate to its [mode] impact. Describe it to me like I am a [age]. Be specific to [word].", which is sent to ChatGPT to generate the label. Beyond this prompt, this model leverages in-context learning by including example labels within the prompt history. Next, the text is screened for unprintable characters and sent back to the Arduino. Lastly, the label is reformatted for printing and printed using Adafruit's thermal printer Arduino library, effectively bridging the gap between digital insights and physical reality.

3 Author Biographies

3.1 Ekanem Okeke

Ekanem Okeke is a Canadian mechanical engineer and designer of Nigerian origin. She recently graduated with her Bachelor of Engineering from MIT, and is currently pursuing her Masters in Mechanical Engineering. As part of her Masters program, she is a researcher in MIT's Learning Engineering and Practice (LEAP) Group, designing solutions for improving manufacturing and mechanical engineering education. As a designer and engineer, she seeks to understand and leverage new technologies to re-think design possibilities. She has worked in a wide range of positions; designing gear for high-performance athletes at the Australian Institute of Sport, designing shoe soles with MIT's Self-Assembly Lab, and researching the advancements of AI in design.

3.2 Alex Stewart

Alex Stewart is a multidisciplinary designer and Morningside Academy for Design Fellow currently pursuing a Master of Architecture with MIT's Class of 2026. Originally from Weeki Wachee, Florida, he earned his Bachelor of Arts in Art, Film, and Visual Studies from Harvard University in 2022, where he graduated with highest honors and was awarded the Peter Cai Prize. Alex has worked with studios such as Dami Lee's NolliStudio in Vancouver, Meaningful in Paris, and KieranTimberlake in Philadelphia. In these positions, he has used a range of research, sketching, modeling, rendering, animation, and fabrication techniques to contribute to a wide range of projects across mediums and scales, including housing, university buildings, a design book, and products.

3.3 Kyrylo Beskorovainyi

Kyrylo Beskorovainyi is the co-founder and publisher of the Ukrainian popular science media company Kunsht. As a media entrepreneur and science journalist, he produces podcasts, teaches science journalism at the Ukrainian Catholic University in Lviv and trains independent journalists. He has organized projects to counter disinformation and exhibitions on digital freedom and in 2022, he co-founded Science at Risk!, an initiative to help Ukrainian scientists and institutions cope with challenges during the war in Ukraine. In 2016, he served as Ukraine's youth delegate to the United Nations. In 2020 he received a UNICEF Youth Changemaker in Education and Science award and in 2020, he was named to Kyiv Post's Top 30 under 30 list. His 2020 children's book about a boy from a family of astrophysicists who tries to get to know more about the universe, was named as a finalist for the BBC Book of the Year award in 2020. He became a Nieman Fellow at Harvard in 2024-2025.

4 Acknowledgements

We would like to thank Marcelo Coelho, Chenyue "xdd" Dai, Sergio Mutis, Bill McKenna and Quincy Lau for their advice and support throughout this project.

5 References

References

Marcelo Coelho and Jean-Baptiste Labrune. Large language objects: The design of physical ai and generative experiences. *Interactions*, 31(4):43–48, June 2024. ISSN 1072-5520. doi: 10.1145/3672534. URL https://doi.org/10.1145/3672534.

Lisa M. Schwartz and Steven Woloshin. The drug facts box: Improving the communication of prescription drug information. *Proceedings of the National Academy of Sciences*, 110(supplement_3):14069–14074, 2013. doi: 10.1073/pnas.1214646110. URL https://www.pnas.org/doi/abs/10.1073/pnas.1214646110.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract reflects the motivation and content of the creative work.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: It is included within the technical description.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

Justification: The paper does not include theoretical results.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [NA]

   Justification: The paper is a description of a creative work, and does not include experiments. However, relevant prompts are disclosed for possible reproduction of results.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [NA]

Justification: The paper is a description of a creative work, and does not include experiments.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [NA]

   Justification: The paper is a description of a creative work, and does not include experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [NA]

   Justification: The paper is a description of a creative work, and does not include experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.

- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [NA]

   Justification: The paper is a description of a creative work, and does not include experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: Photographs taken by LLL are only stored temporarily, and are overwritten with each use.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: It is discussed within the Project motivation section.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: The project does not pose such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: LLM models that were used in the project are credited.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: The project does not include new assets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: There was no crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: There was no crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [Yes]

Justification: LLM use for the project is detailed in the paper.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.