**Execution Study Logistics and Consent Form**

Welcome to the study! In this onboarding document, we will go over all the study logistics including your tasks, the payment setup, and final deliverables. If you agree to participate, **please sign the consent form on the last page of this document, send the signed copy back to me, and tell me your preferred topic(s) to work on**, I’ll then follow up with the idea for you to execute. Please read all sections carefully before you start your execution task.

**Idea Assignment**

When sending the signed copy of this document back to me, please also indicate your preferred 1-2 topic(s) among the following:

1. Bias: novel prompting methods to reduce social biases and stereotypes of large language models

2. Coding: novel prompting methods for large language models to improve code generation

3. Safety: novel prompting methods to improve large language models’ robustness against adversarial attacks or improve their security or privacy

4. Multilingual: novel prompting methods to improve large language models’ performance on multilingual tasks or low-resource languages and vernacular languages

5. Factuality: novel prompting methods that can improve factuality and reduce hallucination of large language models

6. Math: novel prompting methods for large language models to improve mathematical problem solving

7. Uncertainty: novel prompting methods that can better quantify uncertainty or calibrate the confidence of large language models

We will randomly sample an idea from your selected topic for you to execute. We generally do not accept requests to switch the assigned ideas to avoid potential confounders, unless you have a strong reason in which case you should notify us and we will discuss case-by-case (e.g., if you think the idea is absolutely impossible for you to execute).

**Payment Structure**

Your payment will consist of three parts: the base hourly payment, the completion bonus, and the quality bonus.

Base Hourly Rate:

We will compensate for your time and effort by paying $20 per hour as the base payment. We will calculate and send the payment on a weekly basis. We will ask you to fill in a timesheet to record the number of hours you spent on the execution task every week. This is guaranteed payment regardless of the project outcome.

We expect all projects to be finished within 120 hours within a two-month window, so the max amount of base payment would be 120 hours x $20 / hour = $2400. Any extension beyond this should be discussed with us on a case-by-case basis.

Completion Bonus:

After you successfully complete the full project and submit all deliverables based on our requirements, we will pay you a completion bonus of $600. This serves as an additional incentive for you to finish the whole study.

Execution Quality Bonus:

We also set up an additional bonus to award the best executed projects. We will set up a review metric on how well the idea is executed into the project, and we will ask the blind reviewers to judge your submission based purely on the execution (rather than the underlying idea) for this metric. The top 5 projects under this metric will each get a $2000 bonus.

In summary, your total compensation will range from **$3000 to $5000** based on the quality of your execution.

The [Tremendous](https://www.tremendous.com/) platform will handle all the compensation. The platform will send you digital Visa cards or gift cards that you can use to purchase anything. This avoids the logistical troubles of involving bank transfers, so we can more easily include international participants in the study.

**Intermediate Check-ins**

Apart from the final deliverables, we will have weekly check-ins so that we can stay in sync on the progress. The check-in consists of two parts:

1. Code repo: We will request access to the private GitHub repo for the project so that we can track the code commits.
2. Written update: We will ask for a short update message at the end of every week to summarize the main tasks completed for the past week. This will serve as justification for your hourly payment.

We might request additional online meetings with you if we feel the need to. We understand that progress in executing a research project can sometimes be non-linear and we are happy to accommodate your pace, but we reserve the right to pause the payment if we have strong evidence that there has not been progress for several consecutive weeks.

**Execution Logistics**

1. Making modifications to the original idea:

Before you start writing the code, you should first take a look at the given idea and see if you want to propose any modifications. This involves fixing issues that could not be executed properly (e.g., infeasible baselines or experiments, inappropriate metrics or datasets), and adding additional details for parts that are originally underspecified or ambiguous.

We generally allow the following types of modifications:

* Delete [dataset / metric / baseline / model]
* Add [dataset / metric / baseline / model]
* Elaborate [implementation details]
* Change [implementation details]
* Extend [not necessary to make the original idea execute, but would be great extensions to make the project stronger]

Each action should be accompanied by a rationale. The high-level principle is that we want you to execute the given idea in its best form, as long as you are not turning it into a different idea. Please let us know once you have done a pass on the given idea and added your proposed modifications, we will take a look and approve them (we will approve most things as long as you are not making too drastic changes). You can still make further modifications after you start executing the idea, just make sure to keep us updated when you make new changes.

Please directly leave your proposed modifications as comments or suggestions on the idea Google doc. You can refer to this example here: <https://docs.google.com/document/d/1SzdD6CPYpT_4yQaCeT1xFLA56PrLr2W8EwFNmTlnsX4/edit?usp=sharing>.

1. Codebase structure:

You should maintain a README file with the full set of commands to run your codebase and reproduce your experiment results. You should also include the requirements.txt file in the repo. Ideally, you should also keep the experiment logs that contain the reported experiment results. See this for an example: <https://github.com/NoviScl/SRUQ>. In rare cases, using Jupyter Notebook is also fine if you think the idea can be implemented in a single notebook (possibly with some other util functions). Make sure to regularly push your commits to your private repo so that we can track the updates.

1. API Access:

We will reimburse all the API credits (OpenAI, Anthropic, etc.) that you spent on running the experiments for the project. Please give us an estimate before you start the actual experiments, and we will set an upper bound on the budget based on the estimate. After the experiments, please send us a record of the API usage. **We will directly add the reimbursement to your weekly compensation.**

1. Crowdsourcing Budget:

If your project requires human study or annotations (for example, by recruiting annotators from Amazon Mechanical Turk, Prolific, or Upwork), we can support that by reimbursing all the associated costs. You will be responsible for recruiting these annotators using your own account. In the end, you should submit the receipts to us and we will directly add the reimbursement to your final payment. Please let us know the estimated cost before launching the human study. Ideally, you should do some preliminary pilot study before the full study.

1. Dealing with ideas that can’t execute:

When we send the idea to you, please make sure to first take a close read and let us know within a week if you have strong reasons to believe that this idea is not feasible within our time and resource constraints. We will discuss together to determine if this is the case, and if so, we will swap to another idea for you.

In the unlikely event that you realize the given idea is not executable after you have already started it, you should let us know and we will discuss it together to determine whether the project is feasible with reasonable modifications. If not, we will terminate the execution of this project. In such cases, we will still pay you the hourly rate for the hours you have spent on the execution. Additionally, if you wish, you may request us to assign you another idea to execute.

**Final Deliverables**

The final deliverables include:

* The codebase: We will run your codebase to verify we can reproduce the reported experiment results. You can refer to our example repo here: <https://github.com/NoviScl/SRUQ>.
* The technical report: You should write a report to summarize your experiments. You should use the \*ACL two-column format ([overleaf template](https://www.overleaf.com/latex/templates/association-for-computational-linguistics-acl-conference/jvxskxpnznfj)). When you start writing, please notify us by sharing the overleaf link so that we can stay updated. We expect about 4-5 pages of content (3.5 pages minimum; 5 pages maximum) excluding references and the appendix. In general, try to write it as an \*ACL short paper submission, except that the Related Work section is optional for our report. We provide an example report here: <https://www.overleaf.com/read/jswkzmpmhwfn#fd1147>.

We will pay you the completion bonus once we have verified your deliverables, and enter your submission for the execution quality evaluation.

**Publication Policy**

If you think your executed project is of high quality (e.g., the performance of the proposed method is very good), we encourage you to convert the project into an actual conference submission.

We (Chenglei, Diyi, and Tatsu) would not be on the resultant paper since we are not directly contributing to it. If you are executing an idea submitted by our expert participants, we will notify you and put you in touch to discuss how to credit them in the paper submission. In any case, you can assume that you will be the first author of the resultant paper submission.

**Consent Form**

**Research Project Title**: Randomized Controlled Trial for Comparing Human and AI Research Ideation Capabilities (IRB# 74246)

**Investigators**: Chenglei Si, Tatsunori Hashimoto, Diyi Yang (CS Department, Stanford University)

**Contact**: clsi@stanford.edu

**What the Study is About**:

This research aims to compare the quality of AI-generated research ideas with those generated by human researchers. The goal is to understand the novelty, feasibility, and excitement of ideas generated by AI systems compared to human-generated ideas in the context of short-term research projects on the topic of Natural Language Processing.

**What We Will Ask You to Do**:

You will be assigned a random idea either from humans or from AI (you will be blind to the condition). You will then execute the given idea into a full project including the codebase, experiments, and a short report. You will then submit all the code and results to us for a round of blind review by expert reviewers.

**Risks and Discomforts**:

We do not anticipate any significant risks from participating in this research beyond those encountered in daily life.

**Benefits**:

Participating in the study also allows you to contribute to the advancement of knowledge in the field of AI and research methodology. That being said, we cannot and do not guarantee or promise that you will receive any benefits from this study.

**Payments:**

You will receive a base payment of $20 per hour when you are executing the project for a maximum of 120 hours. We will also give you an additional $600 bonus once you finish your project and submit all the deliverables. Lastly, we will give a $2000 bonus for the top 5 projects based on the execution quality as judged by the expert reviewers.

**Privacy and Confidentiality**:

Your privacy and confidentiality will be maintained to the fullest extent. All data will be de-identified (we will not reveal your names with the ideas), and any personal information will be kept separate from your research contributions. Access to this information will be limited to the research team. We do not do any recording, and you only need to submit the ideas and project plans to us.

In accordance with scientific norms, the data from this study may be used or shared with other researchers for future research (after removing personally identifying information) without additional consent from you.

**Taking Part is Voluntary**:

If you have read this form and have decided to participate in this project, please understand your participation is voluntary and you have the right to withdraw your consent or discontinue participation at any time without penalty or loss of benefits to which you are otherwise entitled. The alternative is not to participate. You have the right to refuse to answer particular questions. The results of this research study may be presented at scientific or professional meetings or published in scientific journals. Your individual privacy will be maintained in all published and written data resulting from the study.

**Consent**:

I have read the above information and have received answers to any questions I had. I consent to participate in the study.

**Participant Name:** \_\_\_\_Neeharika Gupta\_\_\_\_\_\_

Please save a copy of this form for your records. If you agree to participate, please fill in your full name above and send a copy of this form to Chenglei Si ([clsi@stanford.edu](mailto:clsi@stanford.edu)), we will then follow up with the execution idea.

**Contact Information:**

Questions: If you have any questions, concerns or complaints about this research, its procedures, risks and benefits, contact the Protocol Director: Chenglei Si, clsi@stanford.edu, 240-484-4880.

Independent Contact: If you are not satisfied with how this study is being conducted, or if you have any concerns, complaints, or general questions about the research or your rights as a participant, please contact the Stanford Institutional Review Board (IRB) to speak to someone independent of the research team at 650-723-2480 or toll free at 1-866-680-2906, or email at irbnonmed@stanford.edu. You can also write to the Stanford IRB, Stanford University, 1705 El Camino Real, Palo Alto, CA 94306.