**### Task 1 ###**

**Source Documents:**

Link to OpenReview: <https://openreview.net/forum?id=ByxPYjC5KQ>

**Meta-review:**

The paper received unanimous accept over reviewers (7,7,6), hence proposed as definite accept.

**### Task 2 ###**

**Source Documents:**

Link to OpenReview: <https://openreview.net/forum?id=HJgS7p4FPH>

**Meta-review:**

The paper presented a detailed discussion on the implementation of a library emulating Atari games on GPU for efficient reinforcement learning. The analysis is very thoroughly done. The major concern is whether this paper is a good fit to this conference. The developed library would be useful to researchers and the discussion is interesting with respect to system design and implementation, but the technical depth seems not sufficient.

**### Task 3 ###**

**Source Documents:**

Link to OpenReview: <https://openreview.net/forum?id=SJgaRA4FPH>

**Meta-review:**

The paper provides methods for training generative models by combining federated learning techniques with differentiable privacy. The paper also provides two concrete applications for the problem of debugging models. Even though the method in the paper seems to be a standard combination of DP deep learning and federated learning, the paper is well-written and presents interesting use cases.

**### Task 4 ###**

**Source Documents:**

Link to OpenReview: <https://openreview.net/forum?id=ry831QWAb>

**Meta-review:**

The paper proposes to study the impact of normalizing the gradient for each layer before applying existing techniques such as SG + momentum, Adam or AdaGrad. The study is done on a reasonable number of datasets and, after the reviewers' comments, confidence intervals have been added, although Table 1 puts results in bold but many of these results are not statistically significant.\n\nThe paper, however, lacks a proper analysis of the results. Two main things could be improved:\n- Normalization does not always have the same effect but the reasons for it are not discussed. This needs not be done theoretically but a more thorough analysis would have been appreciated.\n- There is no hyperparameter tuning, which means that the results are heavily dependent on which hyperparameters were chosen. Thus, it is hard to draw any conclusion.\n\nRegarding the seemingly conflicting remarks of the two reviewers, it all depends on what the paper is trying to achieve. If it tries to show that is it state-of-the-art, then comparing to state-of-the-art algorithms on every dataset is crucial. If it tries to study the impact of one specific change, in this case layer normalization, on the optimization, then comparing to the vanilla version is fine. The paper seems to try to address the latter so it is OK if it is not compared to all the state-of-the-art algorithms. However, proper tuning of existing methods is still required.\n\nUltimately, a better understanding of layer normalization could be useful but the paper is not convincing enough to provide that understanding. There is no need to increase the number of datasets but it should rather focus on designing setups to test and validate hypotheses.

**### Task 5 ###**

**Source Documents:**

Link to OpenReview: <https://openreview.net/forum?id=ryg7jhEtPB>

**Meta-review:**

The authors argue that directly optimizing the IS proposal distribution as in RWS is preferable to optimizing the IWAE multi-sample objective. They formalize this with an adaptive IS framework, AISLE, that generalizes RWS, IWAE-STL and IWAE-DREG. \n\nGenerally reviewers found the paper to be well-written and the connections drawn in this paper interesting. However, all reviewers raised concerns about the lack of experiments (Reviewer 3 suggested several experiments that could be done to clarify remaining questions) and practical takeaways. \n\nThe authors responded by explaining that \"the main \"practical\" takeaway from our work is the following: If one is interested in the bias-reduction potential offered by IWAEs over plain VAEs then the adaptive importance-sampling framework appears to be a better starting point for designing new algorithms than the specific multi-sample objective used by IWAE. This is because the former retains all of the benefits of the latter without inheriting its drawbacks.\" I did not find this argument convincing as a primary advantage of variational approaches over WS is that the variational approach optimizes a unified objective. At least in principle, this is a serious drawback of the WS approaches. Experiments and/or a discussion of this is warranted.\n\nThis paper is borderline, and unfortunately, due to the high number of quality submissions this year, I have to recommend rejection at this point.\n

**### Task 6 ###**

**Source Documents:**

Link to OpenReview: <https://openreview.net/forum?id=1HxTO6CTkz>

**Meta-review:**

The paper investigates various approaches, and a unifying framework, for sequence design. There were a variety of opinions about the paper. It was felt, after discussion, that the paper would benefit from a sharper focus, and somewhat suffers from being overwhelmed by various approaches, lacking a clear narrative. But overall all reviewers had a positive sentiment, and the paper makes a nice contribution to the growing body of work on protein design.