Prof. K.J. Ray Liu
IEEE President 2022
Department of Electrical and Computer Engineering
2232 Kim Bldg
University of Maryland
College Park, MD 20742
Email: kjrliu@umd.edu
Phone: (301) 405-6619

Aug. 21, 2022

Subject: Misconducts in "Deep Learning" and Tolerance to new science of AMD (Autonomous Mental Development)

Dear Prof Liu,

Prof. Jim Kelly, the president of IEEE Computational Intelligence Society (CIS), informally suggested to me during IJCNN 2022 that for such scales of problems, I should contact an IEEE level leader like you.

There are two problems, (1) misconducts in AI known as so-called "deep learning" and (2) lack of tolerance to new science AMD that does not suffer from the misconducts in "deep learning". I did not claim that (2) is due to (1), but the Ad Hoc Committee timely appointed by Jim Kelly seems to have distorted my two claims by only denying (2) is due to (1). Of course, I should not have the burden to prove (2) is due to (1). The Ad Hoc Committee did not address (1) or (2) individually. I am not informed whether any systematic measures about either (1) or (2) would take place in CIS.

In terms of (1), let me provide some text in my popular science book titled "Conscious Learning: Humans and Machines", that will be published by WSPC.

== Start of the quoted text ==

The author believes that the main published performance figures of almost all "deep learning" projects, including many media hyped ones, such as AlphaGo, AlphaGoZero, AlphaZero, AlphaFold, and the IBM Debater, appear to have been grossly inflated. Namely, so called "deep learning" is in a controversial status, as we expose below.

Such misconducts composed of a combination of two types, "data deletion" and "test on training data".

The First Misconduct: Data Deletions

Data deletion is serious research misconduct, since the person who conducts an experiment deletes data that he dislikes. A proper conduct is to report all data that were generated. If the number of performance data is too large to report all data, he must at least provide a characteristic of the data distribution. For performance data, the reported information for distribution should include the worst case, the average case and the best case.

For example, for a teacher to report his teaching performance of his class for a standard test (e.g., TOEFL, Test Of English as a Foreign Language), he should not only report the highest score of his class, but also the minimum score and the average score of the class. Data deletion happens when he reports only the highest score, or few high scores, but deletes all other scores from the report.

In almost all so-called "deep learning" projects, each project needs to trained multiple networks, often many. Each network starts from a different set of random network weights and other parameters. The performances of these networks are very different, although they use the same training data set. The person who

Brain-Mind Institute

BMI Press

Juyang (John) Weng
President

4460 Alderwood Drive
Okemos, Michigan
48864
USA

juyang.weng@gmail.com
Tel: (517) 980-6270
http://www.brain-mind-institute.org/

conducted the experiments deleted many performance data from all trained networks. He reports only the performance of the luckiest network. This is a gross violation of well-established statistical protocol that requires reporting the average performance of all trained networks, instead of the luckiest. For the same reason, in a lottery, one should report the average chance for one lottery ticket to win money from the lottery.

We call the stage of selecting a trained network from multiple trained network "Post-Selection" stage, selection after training.

The Second Misconduct: Tests on Training Data Sets

Furthermore, those who picked up the luckiest network probably based on a test data set, instead of only training data set. This is because a trained network that fits the training set well does not mean that it will fit the test set well.

For the same reason, should we test students using examples that are exactly the same as taught examples? No. This is because students must demonstrate their capability to extend taught examples to similar but different from taught examples.

Post-Selection using a test set is another violation of well-established statistical protocols. This is like testing a student using a training set, so he does not need to understand. All he needs is to memorize all the training examples.

In summary, "deep learning" suffers from two compounding misconducts in the Post-Selection stage: the first, data deletion; the second, tests on training sets.

The first misconduct and the second misconduct are both research misconducts. To reach the level of extreme exaggeration of performance data that has misled the public in various countries today, it is necessary to superimpose these two misconducts together. The absence of any one misconduct of the two is far from enough to reach such a serious level.

The principles of both misconducts are well-understood by the scientific community. However, engineers unintentionally or intentionally ignored it. However, when my paper about the misconducts was published and my report was submitted to the journals that published the misconduct papers, the correspondence authors of these papers should have known their behaviors were wrong. But they still evaded the truth to the public.

As far as I am aware, the only neural network series that does not have the Post-Selection problem is the series from Cresceptron to Developmental Network. Because freedom from Post-Selection requires a solution to a million-dollar problem—the local-minima problem in a high dimensional space, that is why almost all other neural network projects, including those in large publicly listed companies like Google and IBM and all those that the author is aware of, suffer from the local-minima problem.

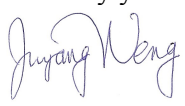
Furthermore, the author has published a paper titled "20 Million-Dollar Problems for Any Brain Models and a Holistic Solution: Conscious Learning" in Proceedings of International Joint Conference on Neural Networks (IJCNN) 2022. It reasons that any brain model, like DN-3 below, must have solved at least 20 million-dollar problems. The local minima problem is one of them.

== End of the quoted text==

For (2), please read Attachment 1.

Please acknowledge your receipt of this letter via email and kindly let me know your planned actions before the end of your 2022 term.

Sincerely yours,



Juyang (John) Weng

Att: 1: 2022-06-15-CIS-Conferences.pdf; 2: 2022-07-28-AdHoc-Committee.pdf