Peer Review Checklist for the Project Paper
Writing Workshop #1

Name of writer: Boyan Li

Comments by: Wentao Mo

Attention! Writing Workshop #1 focuses on the <u>big picture</u> of the essay. In other words, we pay attention to <u>large-scale issues,</u> such as tightening up or shifting focus, clarifying purpose, reorganizing paragraphs, etc. (Grammar, word choices, sentences are for later revision because it is a waste of time to polish a sentence that you later decide you don't need.)

[Clear ideas and clear connections between ideas are critical for a successful paper. As an author, the structure of your paper might be clear to you, but unclear to readers. This activity of writing a reverse outline provides the opportunity to see what the reader is getting out of the paper and whether the key ideas are being communicated effectively.]

Create a reverse outline of your partner's essay. Begin by numbering the paragraphs. For each paragraph, write a one-sentence description or use key words in a bullet-point of the main idea contained in that paragraph. (For an example of a reverse outline, please refer to Week 8 lecture slides, #47.)

As you are outlining, pay special attention to:
- Paragraphs where the main idea is unclear
- Places where the connections between paragraphs are unclear

<u>Reverse outline</u>:

①. Data fueled the modern society's progress, and different types of data require different treatment in the data collection supervision.
②. The data collection situation is complex, where Internet penetrates unavoidably into daily life and people are more and more concerning data privacy while willing to share their information online (like on SNS).
③. It is hard to draw a definite line between public/private data.
④. As consequentialism indicates, some people fear data leakage that triggers deanonymizing and personal discrimination. Also, the advanced algorithm may implicitly leak information from existing data.
⑤. Data sharing benefits in some aspects (like medication) overwhelms its opposite.
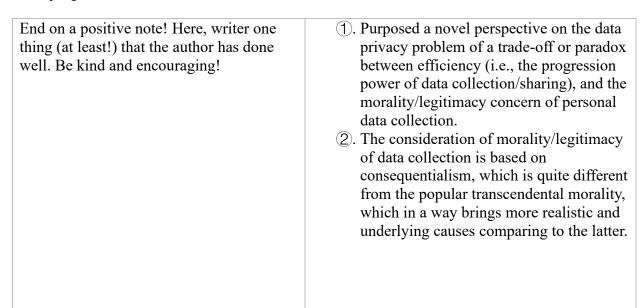⑥. Summary…

Carefully read your group member's draft essay. Based on your reading, address the following questions through constructive comments and suggestions. When you comment, keep in mind what the author did well, as well as provide constructive comments on possible improvements. Your comments are as specific as possible.

Consider the following questions:

I. Overall structure	Comments / Suggestions
Looking at the essay as whole, do you think the author express an insightful perspective towards the topic?	Yes, the author suggests that there are multiple benefits and harm of (big) data sharing, and people might have wrong evaluations upon data privacy problems. I suggest the author include more materials supporting the benefits of data sharing, especially those involving big data.
Are the needs of the audience kept in mind? For instance, do some concepts or words need to be defined?	For most concepts, the definitions are clear for me. Maybe some further explanation and details of the so-called advanced algorithm in analyzing the big data(one example is enough).
Does the essay display a logical progression of ideas, with a clear structure that enhances the thesis?	The essay does exhibit a logical progression, from the concern from the multitude and the ambiguity of the data privacy threshold to the benefits over the harm of big data sharing. Finally, it leads to the conclusion that big data collection is susceptible but will be acceptable with care and supervision.
II. Thesis statement	
What's the thesis (main point)? Please underline the thesis on your partner's draft and copy to here. If the main idea is implied only, jot down what you perceive to be the thesis here.	The data privacy paradox can be seen in 2 aspects. First, the real big data benefits are few and we need a further supervisory system. Second, people often wrongly evaluate data privacy.
Does the essay have a debatable and focused thesis?	Yes, the essay tried to focus on the paradox or trade-off between efficiency and morality/legitimacy in big data collection.
Does the thesis state a compelling position?	Yes, the thesis statement attempts to clarify the benefits and harm of data sharing.
Are key supporting points provided in the thesis statement?	I'd rather say the author might have tried to convey the idea in the first half of the essay, that transparent personal data privacy (i.e., big data privacy) raised concern among people. But later

	in the essay when it tried to tell the benefits (of doing so), the author used examples of COVID genome sequence data and personal data collection for scientific purposes (namely the genomics data) to explain the benefits of big data sharing, but the sharing of COVID genome data is rather some kind of knowledge sharing, but not big data. And the genomics data collection is not so correlated with data sharing, since it is not shared but only used on the patients themselves. Furthermore, the share of medical data shall not be commercialized or publicized (in the US) without an agreement, as the HIPAA act suggests. So, I don't see these statements as a strong support to the benefit of big data sharing.
III. Body	
How does each paragraph relate to the essay's main idea or the previous paragraph?	The 1st paragraph purposed the paradox of the efficiency and morality/legitimacy in data collection concern. The 2-4 paragraphs analyzed the source of personal data privacy problem nowadays, and then the 5th paragraph talked about the benefits of data sharing, and then it concludes and summarizes in the 6th paragraph (the thesis part). The connection might be a little vague between the 2-3 paragraph discussing the complex borderline between public/private data, and the latter analysis in the 4th paragraph involving the cause of the data privacy problem.
Does the author consider and critically evaluate alternative/opposing points of view? *Specify the paragraph(s) where this strategy is used.*	The essay is rather informative with many positive support and materials and is very analytical rather than very argumentive. The strategy using these opposing points is not used.
Is there any unnecessary information throughout the body of the paper, such as excessive quotation or unsupported claims?	Quotations are all useful under the author's intention, while the only less supportive part is on the benefits of data sharing in the 5th paragraph.
IV. Overall evaluation	
How can the essay be further improved? Provide at least TWO constructive suggestions.	①. Is personal data privacy borderline so hard to define? At least we can state the exact part of data that is private/public, and the first may include your personal

photo/diary/memo recorded and stored only locally, and the latter may include the photo you uploaded to Facebook or Twitter, which you deliberately convey to the public Internet.

②. In my opinion, personal data *is* big data. Data privacy concern *is* personal data privacy problem. So the knowledge sharing case (i.e., COVID) might not be suitable for big data sharing, nor the genomics information collection (maybe it involves big data but I didn't really get it from your text).

③. Structurally, the discussion about the unequal/asymmetric algorithms may be separated as a new paragraph with the discussion of consequentialism. And the thesis points finally purposed in the last paragraph could be moved forward into the first paragraph, to make the essay thesis clearer for the readers. Also, the connection might be a little vague between the 2-3 paragraph discussing the complex borderline between public/private data, and the latter analysis in the 4th paragraph involving the cause of the data privacy problem.

④. (Personal comment) In the 4th paragraph, Google and Facebook do not have dominancy in user's text information, or at least not in the NLP field. Actually, they often use unbiased, clean, and public text sources, like Wikipedia, for better performances. They may utilize their data supremacy in the recommendation system, where the personalized text is the key to generate recommendations. And these biases in the dataset also intrigued the research of ethical AI/ML, that how can an AI/ML algorithm learn from biased/dirty/discriminatory data and gives an unbiased/clean/in-discriminatory result?

End on a positive note! Here, writer one thing (at least!) that the author has done well. Be kind and encouraging!	①. Purposed a novel perspective on the data privacy problem of a trade-off or paradox between efficiency (i.e., the progression power of data collection/sharing), and the morality/legitimacy concern of personal data collection. ②. The consideration of morality/legitimacy of data collection is based on consequentialism, which is quite different from the popular transcendental morality, which in a way brings more realistic and underlying causes comparing to the latter.

Please add any other suggestions or comments that you feel would be useful for the author: