

✓ Congratulations! You passed!Grade received **60%** To pass 1% or higher[Go to next item](#)**Week 4 Quiz - Conclusion**

Latest Submission Grade 60%

1. Scenario A**0 / 1 point**(Answer the following 6 questions *based on the information given below and the material in this course.*)

You are a data scientist tasked with improving a government computer system. You are asked to analyze a dataset that contains results from a “report a problem with this system” form in the “Help” tab inside the computer software. The error reporting system also takes a snapshot of data about the user’s system at the time of the error report. Your goal is to suggest improvements to the system using these data, and to diagnose problems.

What is an example of a **sampling problem** you will encounter with these data?

- A) Your report of errors will be conflated with the user’s ability to report the error.
- B) You will have no data on errors so severe that the software doesn’t work at all.
- C) You may be more likely to find errors that affect people who like filling in forms and have the time to do so.
- D) Both A and B.
- E) All of the above.

**Incorrect**

Correct answer: All of the above.

All of these are sampling problems. Some errors will likely be harder to observe and report accurately, so this analysis is biased toward certain kinds of errors (answer A). Indeed, if the error causes the reporting page not to work it will not be reported in these data (answer B) so this analysis will miss very severe errors. Finally, the material about who fills in surveys might lead you to wonder if there are particular kinds of people more likely to make these reports, and this could interact with kinds of errors in an undesirable way (answer C). Note that sampling bias could also help you in this scenario – it could be that some errors are severe enough that they prevent people from doing their job or accessing some essential service. This would create a big incentive to report the error, skewing the data away from, say, minor annoyances.

2. What is an example of an **ethical problem related to **sampling** if the system were designed for the use of the **general public**?****0 / 1 point**

- A) Your suggestions may not be able to detect the needs of people who do not have a computer to access the software.
- B) You have violated your “duty of care.”
- C) A detailed snapshot of the user’s computer might include sensitive information.
- D) All of the above
- E) Both A and C

**Incorrect**

Correct answer: Your suggestions may not be able to detect the needs of people without computers

It is important to highlight that we consider this to be an **ethical** problem because it relates to questions of who is included and excluded in a system, and reminds us to take a wide view of the systems we create. Services provided by the government are special: we expect them to be provided equally to everyone (sometimes called “equal protection”). In this case people without computers may currently be non-users of the computer system, but if it’s a computer systems designed for the use of the general public they must be represented in the sample.

The other answers are incorrect. The “duty of care” was not covered in this course. Although “duty of care” sounds relevant, answer B is not particularly important for questions about this scenario and sampling -- at least unless we are given more information. Although answer C is also correct in the sense that it is a true statement, it is not relevant to sampling, which was the question asked in this case.

3. Imagine you were also offered an additional dataset from a company like Clearview AI. It consists of all of the times people have mentioned your government software on social media, along with data about who posted the messages/tweets/posts/photos/status updates/etc. However, it was gathered in violation of the policies of the social media companies involved. It was also gathered without user consent. What is the name for the general issue with**0 / 1 point**

using these data?

- A) Reflexivity and member-checking
- B) Provenance and disposition
- C) Aggregation
- D) Classification
- E) all of the above



Incorrect

Correct answer: Provenance and disposition

Don't overthink this one -- we were intending to check to be sure that you understand the course concept that the way data is obtained matters: that's provenance (a.k.a. origination) and it's important to note if the data are being moved to a context that is different from why they were generated in the first place: that's disposition. In other words: where the data came from and where the data ended up. Although "aggregation" might also be tempting as an answer, that word would apply only if you merged this dataset with the other one specified above. It isn't clear how you would merge the two datasets as they would be very different (one consists of form responses and measures about the state of a user's computer system, the other is a repository of social media posts and credentials). The other answers shouldn't be too tempting.

4. If you complete an analysis and you find a problem with the operation of the system, one way that you might determine that something **counts as a problem** at all is by comparing the system to what you believe ought to happen. If you object to the operation of the system by saying, "everyone has the right to equal access to government services," or "it is wrong to discriminate between groups," this is best described by what concept?

1 / 1 point

- A) Compliance
- B) Patient
- C) Power
- D) A norm



Correct

Correct answer: A norm

This question is intended to test your understanding of how you might make an ethical case for or against a particular situation. Usually we recommend linking such a claim to some more fundamental reasoning that we take for granted. In Johnny Cash's segment this week we proposed the UN Declaration of Human Rights as an example, but there may be an industry- or application-specific set of rights or principles that makes more sense for you. Both of these examples are from the UN Declaration (they relate to Article 7 and Article 21). We called this a "norm" but we could also refer to it as "grounds" or "rights."

Although you might have been tempted by "A" (compliance), norm is a better answer as these two statements are the grounds for ethical reasoning whether or not there is a law about them. Indeed, many kinds of discrimination are **legal**, even if they may be unethical. Patient and power are relevant vocabulary for the course but they do not apply to this situation.

5. Imagine that you raise an ethical objection to the operation of this system and are told by a manager: "Don't worry, this system uses only publicly available data." Based on the case studies reviewed this week, how should you feel about this statement? You should ____.

1 / 1 point

- A) Feel alarmed -- public data is more dangerous than private data.
- B) Feel reassured that there are no privacy problems.
- C) Feel reassured that there are no consent problems.
- D) Not be reassured, as this doesn't tell you very much.



Correct

Correct answer: Not be reassured, as this doesn't tell you very much

This week we discussed portability, disposition, and forgetting. All of these concepts and their related cases tried to make the case to you that whether or not the data was once public under some particular circumstances might not tell you very much about whether it is acceptable to use the data now for some other purpose. This week with the idea of provenance we tried to emphasize that you might need to know how the data were obtained.

Both public and private data present potential for danger or harms. Answer "A" is not credible, as we covered the dangers of private data as a source of harms in Week 1. Answer "B" or "C" are clearly not correct because the cases this week involved public data where there were privacy problems (e.g., mugshots.com) and consent problems (e.g., Clearview AI).

6. When analyzing the data about this system, you come across some timestamps revealing that some the data gathered by this "report a problem" system is extremely old. What would this week's materials suggest that you do in response as a general policy (even though these might not apply in all cases)?

0 / 1 point

- A) Suggest a policy requiring that old data are erased (retention/disposition).
- B) Ensure that there is a way for users to delete their own data (disposition/forgetting).
- C) Erase the timestamps to help anonymize the historicity of the data (temporality).
- D) Both A and B
- E) All of the above

 Incorrect

Correct answer: both A and B

Although these responses might not be applicable to every system, in this week's materials we discussed examples where there should be some attention paid to where data end up at the end of the data lifecycle. We emphasized the debate in data science about the importance of forgetting data. Answer "C" might be attractive because it contains the word "anonymize" -- which sounds like a good thing -- but in general we want to know how old data are. Therefore we don't want to promote a policy of erasing timestamps.

7. Scenario B

1 / 1 point

(Answer the next 4 questions based on the information given below and the material in this course.)

Imagine yourself to be a data scientist working in industry. You are assigned to an ethics board in your organization. Due to a series of terrible scandals, your board is tasked with improving data science ethics in general. Your recommendations might include any of the suggestions in the Loukides et al. book.

You want to build ethics into your company's culture. What would you propose to your company?

- A) Add ethical challenges in the hiring process.
- B) Cultivate a diverse and culturally sensitive team culture.
- C) Ensure that ethicists supervise data scientists.
- D) Both A and B
- E) All of the above

 correct

Correct answer: Both A and B

It is the position of this class and the readings that ethics is everyone's responsibility. During week 1 we argued that only data scientists can make effective ethical decisions in instances where expertise is required to diagnose problems and offer solutions. That's why data science ethics involves data scientists, and not, for example, outside philosophers or ethicists. This makes answer (C) incorrect. The first two answers are taken directly from the reading as suggestions.

8. Considering the general advice in the Loukides et al. book, Chapters 2 and 4, you may want to recommend an ethics **checklist** instead of an ethics **oath** or pledge because checklists ____.

1 / 1 point

- A) are more standardized.
- B) connect principle to practice.
- C) can be customized.
- D) cannot be issued by regulatory bodies.

 correct

Correct answer: connect principle to practice

Loukides et al. argue that checklists are more specific to situations that might arise in the workflow at your specific organization, so that is one reason to recommend them over oaths, which tend to be more general. The other answers are incorrect because both oaths and checklists are standardized (A) and can be customized (C). Checklists can be issued by regulatory bodies (D).

9. We should develop our checklist around:

1 / 1 point

- A) Simple and specific questions.
- B) Ensure all five "Cs" are represented in any checklist.
- C) Use only questions from the "Data Ethics Workbook" model checklist.
- D) All of the above



Correct

Correct answer: Simple and specific questions

This was not meant to be a tricky question -- hopefully it was easy. Loukides et al. emphasize simple, specific questions and encourage you to construct your own questions that are specific to your industry and situation. The five "Cs" are NOT supposed to be relevant to every situation. Indeed the authors emphasize writing focused checklists about a particular part of a workflow that can be repeated often, not a general adherence to principles like the five "Cs." The authors disagree with the "Data Ethics Workbook" (UK) and do not recommend it because of its "high overhead."

10. Part of an ethical culture is protecting ethical workers. Which of these proposals are consistent with Loukides et al.?

1 / 1 point

- A) Workers must have protections from retaliation.
- B) Workers must have a process to escalate issues.
- C) Any worker must be able to stop production.
- D) Both A and B
- E) All of the above



Correct

Correct answer: All of the above

These recommendations are taken directly from Chapter 4. Examples given in the text include the "Andon cord" pioneered within Japanese auto manufacturing. And the US State Department's "dissent channel." In some industries and situations, protections for workers raising an ethical problem -- such as whistleblowers -- may be written in the law.