**Frameworks for understanding and managing AI/data science’s opportunities and challenges**

**DJ’s “5 Cs”** — “a framework for implementing the golden rule of data” (treat others’ data as you’d like your own data to be treated; what dimensions do people care about when sharing their personal data?)

1. Consent — Informed agreement on what data is to be collected and to what end; notions of consent and its limits can vary by geography (e.g., EU and GDPR).
2. Clarity — Even when there appears to be consent, how confident are we that end-users understand the terms? This can be affected by technical complexity of usages and by legalese documents, such as form Terms and Conditions.
3. Consistency (and trust) — Ensuring systems operate as expected and are reliable, so that users have accurate mental models of how their data will be used and the data-holder’s responsibilities.
4. Control (and transparency) — Understanding what data is held, on what dimensions, and to what use, on an ongoing basis.
5. Consequences — Anticipating how data might be misused or abused, either by malevolent actors or by well-intentioned actors with unforeseen circumstances.

**DJ’s “data science project checklist”**— what are the questions a data science team should ask before undertaking or launching a project, akin to other product review questions a team may undergo (e.g., will this new launch cannibalize our existing products)

*1. Have we listed how this technology can be attacked or abused?*

*2. Have we tested our training data to ensure it is fair and representative?*

*3. Have we studied and understood possible sources of bias in our data?*

*4. Does our team reflect diversity of opinions, backgrounds, and kinds of thought?*

*5. What kind of user consent do we need to collect to use the data?*

*6. Do we have a mechanism for gathering consent from users?*

*7. Have we explained clearly what users are consenting to?*

*8. Do we have a mechanism for redress if people are harmed by the results?*

*9. Can we shut down this software in production if it is behaving badly?*

*10. Have we tested for fairness with respect to different user groups?*

*11. Have we tested for disparate error rates among different user groups?*

*12. Do we test and monitor for model drift to ensure our software remains fair over time?*

*13. Do we have a plan to protect and secure user data?*

I should note as well here that while cybersecurity is directly present in a few of the above questions (for instance, in questions 8, 9, and 13), this was also a central focus of DJ’s talk, e.g., in describing [‘red-team’](https://en.wikipedia.org/wiki/Red_team) approaches to both infiltrating data systems and to how systems may be misused or abused.

https://medium.com/@sjgadler/care-about-ai-ethics-what-you-can-do-starting-today-882a0e63d828