

# Reading Response 2/23

Kritika Chugh  
SUID: 882046659

*kchugh@syr.edu*

## Talking Governance in Data Science

One thing to make sure there is governance is that as a data scientist community should always question their methods and practices. There are many such things an organization or a data science committee should consider such as : the use of data , security, privacy, bias, accountability, transparent decision making and many more such parameters.

Why is good governance and accountable algorithms required?

Because whatever they are doing with this data, there is broader impact on society not just an individual. So perspective should include not just the legal compliance but must also consider ethical and reputational considerations. Note that there is always privacy risk at play for the consumers when their data is used without them knowing about it. For example a purchasing pattern might reveal a lot about the medical conditions of a patient such as pregnancy.

That is why we need data governance:

- 1) **Information security and privacy of retained and non-retained data** – avoid retaining data if no longer required by the business and if retained then transform it (encrypt in rest or in transit) to minimize risk and sensitivity.
- 2) **Re-identification risk** – re-identification of non-sensitive data is possible under right circumstances.
- 3) **Review Board for data use** – they should be diverse and empowered enough to be able to accept or deny data collection requests, investigation of sensitive questions and deployment of insight from techniques used in investigation.
- 4) **Impact statements** – they structure the process for critical transparency and techniques, layout the analytical means to tradeoff certain approaches. Can be used for internal consumption or published to keep the engagement of the customers.
- 5) **Explain data driven approaches** – it helps to maintain transparency by accompanying data with information about provenance and processing.
- 6) **Auditing** – validate their predictions continually and performance even after the launch.
- 7) **Looking for systematic biases** – using data collected from a non-representative pattern.
- 8) **A feed back loop for process improvement**
- 9) **Framework for accountable algorithms** – an internal role may be designated that is responsible for owning to the outcomes of the automated process using the data. Notice that the dealing requires both individual-level and broader societal-level claims.

