DATAX201
Week 12 – Lecture 2

Beyond data: Principles of Responsible Algorithms

Previous Lecture(s)

→ Data Practices can perpetuate & amplify historic injustices & systemic biases.

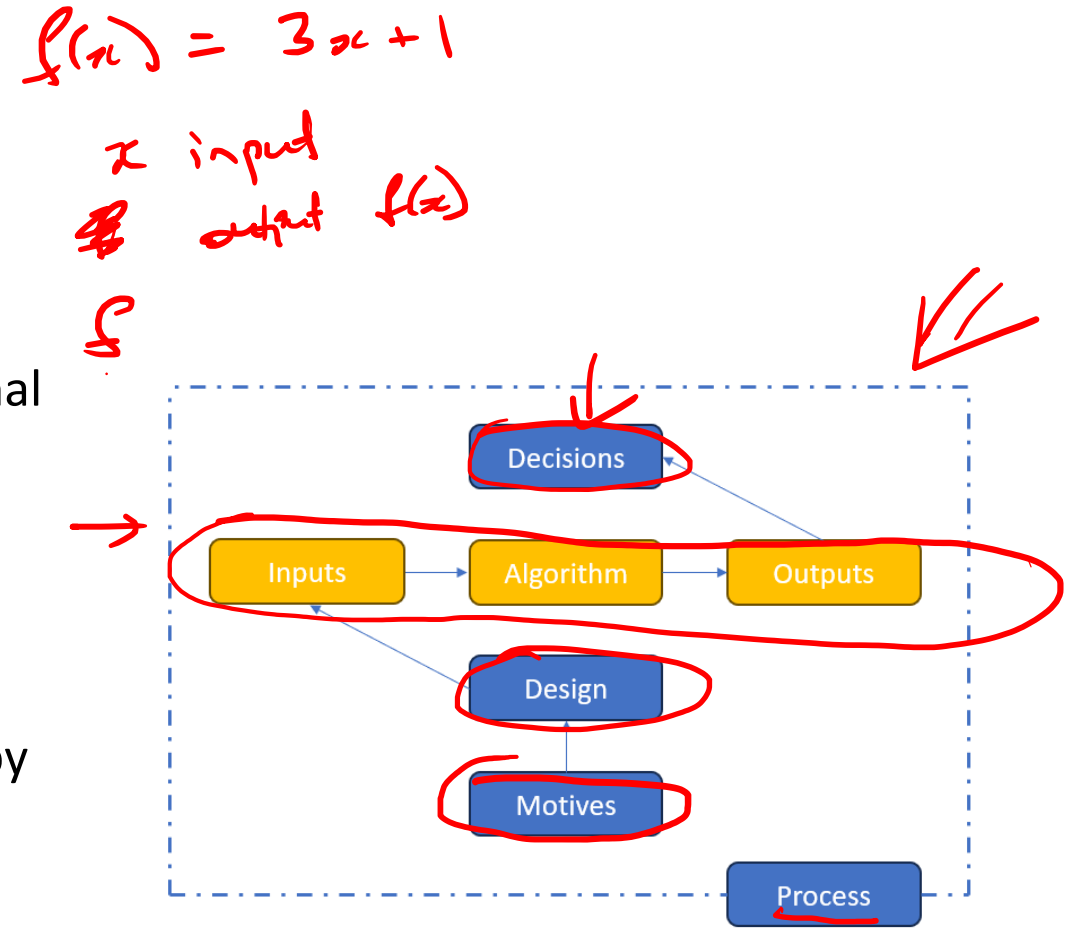
→ Good Data Practice

- Integrity
- Responsibility

What is an Algorithm?

- Computational Algorithms
  - An abstract, formalised description of a computational procedure
  - A machine that turns inputs into outputs

- Algorithmic systems
  - An iterative, decision-making process that is driven by humans, data, and computational algorithms
  - Synonyms: AI, AI tech, automated decision-making processes, models ++



Generic structure of an Algorithmic system

Algorithmic Bias

Definition: When the outputs of an algorithm benefit or disadvantage individuals or groups over others without justification or reason

- Cause of algorithmic bias - Technical perspective
  - Input bias
  - Training bias
  - Programming bias

- Cause of algorithmic bias – deeper dive
  - Purpose of algorithm does not allign with ideas of fairness and equity, or justice for marginalised people
  - Data used may have inherent biases embedded within – perpetuates already established biases
  - Values embedded in the system are from those who are involved over the process
    - No indigenous input – no indigenous values!!

- Assessment of algorithms and bias can be hard!
  - Lack of transparency – very opaque
  - Explainability issues – understanding what goes on “under the hood” can be very difficult

NZ Algorithm Assessment Report 2018

- A cross-government report of how government agencies use (operationalised) algorithms to improve the lives of NZers
  - https://data.govt.nz/docs/algorithm-assessment-report/

- Assessment of algorithms is required – this lead to the NZ Algorithmic Charter 2020
  - https://data.govt.nz/toolkit/data-ethics/government-algorithm-transparency-and-accountability/algorithm-charter/

Principles of Responsible Algorithms

- Governments and agencies across the (mostly developed) world have been putting out ethics guidelines for algorithms

- A metastudy conducted by Jobin *et al.,* (2019) looked at 84 ethics frameworks for AI and found
  - 5 principles that were common in over 50% ~~fe~~ of frameworks
  - 6 other principles that came through in a few studies

Principles of Responsible Algorithms (Top 5)

- **Transparency**: Algorithms must be open, explainable, and explicit regarding their purpose, development, use and maintenance. Efforts should be made to increase the explainability of how it functions, interpretability of outputs, and understandability of the system.

- **Fairness/Justice**: Outputs of algorithms should be free of algorithmic bias, or at the very least tested for bias and disclose findings. Purpose and use of the outputs must allign with the ideas of fairness, equity, and inclusion, and should be explainable enough so that outputs may be challenged.

Principles of Responsible Algorithms (Top 5)

(3) • **Non-Maleficence (Do No Harm)**: Algorithms should be safe and secure and should not purposely cause foreseeable or unintentional harm (discrimination, physical harm, violation of privacy etc.).

(4) • **Responsibility**: Use of algorithms should be done with integrity, and the allocation of responsibility, obligation, and legal liability should be clear in all parts of the process.

(5) • **Privacy**: Algorithms require data, and privacy (data security and protections) is a value to uphold as a protected right.

both individuals & collectives

Principles of Responsible Algorithms (Others)

- **Beneficience**: Promote human well-being and flourishing communities
- **Freedoms**: Promote freedoms, empowerment, autonomy and self-determination
- **Trust**: Noble purpose, used by trusted experts from reputable organisations
- **Sustainability**: Considered protection of the environment
- **Dignity**: Uphold human rights
- **Solidarity**: benefits must go towards strong safety nets, re-distribution of wealth

An Indigenous Perspective

- Those involved in development and use of algorithms embed their value systems in those processes
- Development of operational/government algorithms in NZ lack meaningful partnership and participation of Māori
- Idea: algorithmic outputs are a function of data.
  - If Māori data is used in algorithms – the ideas of Māori data sovereignty apply!