Computational constraints on algorithmic governance

Elija Perrier

Centre for Quantum Software and Information, University of Technology (Sydney);

School of Economics, University of Sydney; School of Law, Macquarie University
elija.t.perrier@student.uts.edu.au; eper2139@uni.sydney.edu.au

26 August 2019

What this talk is about

Ethical governance of AI/algorithms requires awareness of limitations of what is possible:

What this talk is about

Ethical governance of AI/algorithms requires awareness of limitations of what is possible:

- computational constraints on regulation of algorithms

What this talk is about

Ethical governance of AI/algorithms requires awareness of limitations of what is possible:

- computational constraints on regulation of algorithms
- ethical impossibility theorems

What this talk is about

Ethical governance of AI/algorithms requires awareness of limitations of what is possible:

- computational constraints on regulation of algorithms
- ethical impossibility theorems
- probabilistic ethics and ethical inconsistency

What this talk is about

Ethical governance of AI/algorithms requires awareness of limitations of what is possible:

- computational constraints on regulation of algorithms
- ethical impossibility theorems
- probabilistic ethics and ethical inconsistency
- solutions drawn from technical literature and existing systems, such as legal

Summary

In this talk, I will argue:

1. **Proposition**: complexity of datasets and AI/algorithms necessitates *regulation of algorithms by algorithms*

Summary

In this talk, I will argue:

1. **Proposition**: complexity of datasets and AI/algorithms necessitates *regulation of algorithms by algorithms*
2. $\Longrightarrow$ ethical algorithmic governance is limited by computability and complexity constraints

Summary

In this talk, I will argue:

1. **Proposition**: complexity of datasets and AI/algorithms necessitates *regulation of algorithms by algorithms*
2. $\implies$ ethical algorithmic governance is limited by computability and complexity constraints
3. $\implies$ reliance upon heuristics and probabilistic reasoning

Summary

In this talk, I will argue:

1. **Proposition**: complexity of datasets and AI/algorithms necessitates *regulation of algorithms by algorithms*
2. $\implies$ ethical algorithmic governance is limited by computability and complexity constraints
3. $\implies$ reliance upon heuristics and probabilistic reasoning
4. $\implies$ potential for inconsistent ethical outcomes sought or uncertainty over or competition concerning classification as ethical

Summary

In this talk, I will argue:

1. **Proposition**: complexity of datasets and AI/algorithms necessitates *regulation of algorithms by algorithms*
2. $\implies$ ethical algorithmic governance is limited by computability and complexity constraints
3. $\implies$ reliance upon heuristics and probabilistic reasoning
4. $\implies$ potential for inconsistent ethical outcomes sought or uncertainty over or competition concerning classification as ethical
5. $\implies$ necessity of unavoidable ethical trade-offs and ethical inconsistency overall

Summary

In this talk, I will argue:

1. **Proposition**: complexity of datasets and AI/algorithms necessitates *regulation of algorithms by algorithms*
2. $\Longrightarrow$ ethical algorithmic governance is limited by computability and complexity constraints
3. $\Longrightarrow$ reliance upon heuristics and probabilistic reasoning
4. $\Longrightarrow$ potential for inconsistent ethical outcomes sought or uncertainty over or competition concerning classification as ethical
5. $\Longrightarrow$ necessity of unavoidable ethical trade-offs and ethical inconsistency overall
6. $\Longrightarrow$ regulatory architects and policymakers must consider these limitations when framing and implementing ethical AI regulation

Ethical computation

Question: What does it mean for an algorithm or AI to be 'ethical'?

Ethical computation

Question: What does it mean for an algorithm or AI to be 'ethical'?

- **Ethical computation**: the algorithmic computation consists of:

Ethical computation

Question: What does it mean for an algorithm or AI to be 'ethical'?

- **Ethical computation**: the algorithmic computation consists of:
  (i) *ethical decision-procedures* (is procedure provably ethical?) ("**ethical means**") or ("**algorithmic deontology**") prescribed/proscribed calculation methods

Ethical computation

Question: What does it mean for an algorithm or AI to be 'ethical'?

- **Ethical computation**: the algorithmic computation consists of:
  - (i) *ethical decision-procedures* (is procedure provably ethical?) ("**ethical means**") or ("**algorithmic deontology**") prescribed/proscribed calculation methods
  - (ii) *ethical outcomes* (is output provably ethical?) ("**ethical ends**") or ("**algorithmic consequentialism**")

Ethical computation

Question: What does it mean for an algorithm or AI to be 'ethical'?

- **Ethical computation**: the algorithmic computation consists of:
  (i) *ethical decision-procedures* (is procedure provably ethical?) ("**ethical means**") or ("**algorithmic deontology**") prescribed/proscribed calculation methods
  (ii) *ethical outcomes* (is output provably ethical?) ("**ethical ends**") or ("**algorithmic consequentialism**")
- **Auditing**: an algorithm is ethical if it only if it is provably (or perhaps probably) ethical i.e. if ethical status of its procedures/outputs can be audited

Ethical AI will be algorithmic

- **Ethical AI problems are complex**: big data (high volume/velocity of data; curse of dimensionality) and complex methods (supervised v unsupervised learning; deep learning):

Ethical AI will be algorithmic

- **Ethical AI problems are complex**: big data (high volume/velocity of data; curse of dimensionality) and complex methods (supervised v unsupervised learning; deep learning):
  - financial transactions
  - communications (e.g. social media)
  - cybersecurity
  - autonomous machines
  - complex and dynamic code structures (e.g. how to monitor millions of neural networks at any one time)

Ethical AI will be algorithmic

- **Ethical AI problems are complex**: big data (high volume/velocity of data; curse of dimensionality) and complex methods (supervised v unsupervised learning; deep learning):
  - financial transactions
  - communications (e.g. social media)
  - cybersecurity
  - autonomous machines
  - complex and dynamic code structures (e.g. how to monitor millions of neural networks at any one time)
- **Ethical AI will be algorithmic**: complex nature of algorithmic systems, large datasets and ubiquity of AI will necessitate that, by and large, *auditing/procedural* regulation of algorithms will need to be via computational means - e.g. algorithms regulating algorithms

The Four C's Framework

Four C's Framework: framework for considering impact of computational constraints upon algorithmic governance:

The Four C's Framework

Four C's Framework: framework for considering impact of computational constraints upon algorithmic governance:

- Computability

The Four C's Framework

Four C's Framework: framework for considering impact of computational constraints upon algorithmic governance:

- Computability
- Complexity

The Four C's Framework

Four C's Framework: framework for considering impact of computational constraints upon algorithmic governance:

- Computability
- Complexity
- Consistency

The Four C's Framework

Four C's Framework: framework for considering impact of computational constraints upon algorithmic governance:

- Computability
- Complexity
- Consistency
- Controllability

Computability and efficiency

- **Computability and efficiency**: two questions we should ask when debating ethical AI are:

Computability and efficiency

- **Computability and efficiency**: two questions we should ask when debating ethical AI are:
  (i) is the ethical computation (procedure/outcome) actually *computable*?

Computability and efficiency

- **Computability and efficiency**: two questions we should ask when debating ethical AI are:
  (i) is the ethical computation (procedure/outcome) actually *computable*?
  (ii) is the ethical computation *efficiently* (and feasible, given resource constraints) computable?

Computability

- **Computability**: regulation requires procedure for determining ethical status of algorithm. Issues:

Computability

- **Computability**: regulation requires procedure for determining ethical status of algorithm. Issues:
  (i) *decidability*: are ethical criteria well-defined, consistent and decidable? Can procedure/output actually be definitively classified as ethical or not? Is there a decision procedure to decide between competing ethical algorithms (e.g. competing ethical AI settings among driverless car manufacturers)?

Computability

- **Computability**: regulation requires procedure for determining ethical status of algorithm. Issues:
  (i) *decidability*: are ethical criteria well-defined, consistent and decidable? Can procedure/output actually be definitively classified as ethical or not? Is there a decision procedure to decide between competing ethical algorithms (e.g. competing ethical AI settings among driverless car manufacturers)?
  (ii) *deterministic v probabilistic*: is the assessment deterministic or probabilistic? Can algorithmic governance be ethical if ethical status is not computable or uncertain/probabilistic?

Computability

- **Computability**: regulation requires procedure for determining ethical status of algorithm. Issues:
  (i) *decidability*: are ethical criteria well-defined, consistent and decidable? Can procedure/output actually be definitively classified as ethical or not? Is there a decision procedure to decide between competing ethical algorithms (e.g. competing ethical AI settings among driverless car manufacturers)?
  (ii) *deterministic v probabilistic*: is the assessment deterministic or probabilistic? Can algorithmic governance be ethical if ethical status is not computable or uncertain/probabilistic?
  (iii) *uncertainty/risk thresholds*: if probabilistic, how are decisions around acceptable risk of unethical outcomes determined?

Complexity and feasibility

- **Complexity**: computational complexity considers how the resources needed to computationally solve a problem scale with the the input size n to a problem [Aaronson, 2011]

Complexity and feasibility

- **Complexity**: computational complexity considers how the resources needed to computationally solve a problem scale with the the input size n to a problem [Aaronson, 2011]
  (i) *Efficient*: run-time is upper-bounded by a *polynomial* function of n, solvable by classical computation

Complexity and feasibility

- **Complexity**: computational complexity considers how the resources needed to computationally solve a problem scale with the the input size n to a problem [Aaronson, 2011]
  (i) *Efficient*: run-time is upper-bounded by a *polynomial* function of n, solvable by classical computation
  (ii) *Inefficient*: run-time is lower-bounded by an *exponential* function of n, higher complexity class, cannot be solved by classical computation (e.g. EXPTIME problems)

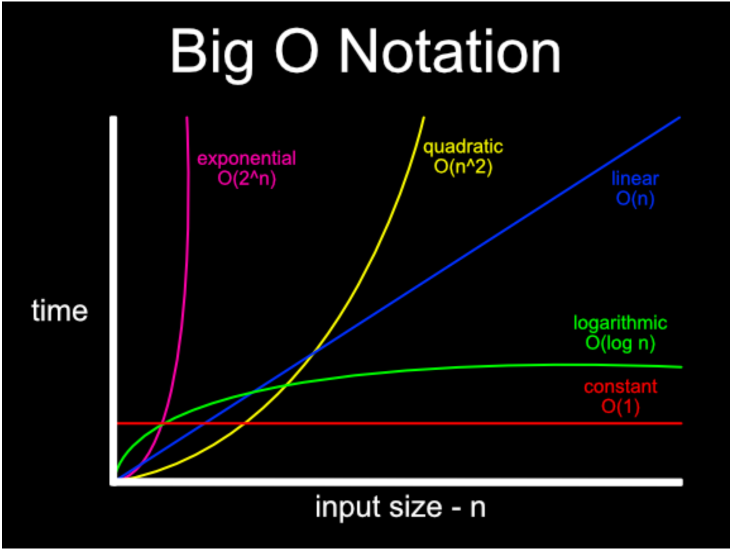
Complexity and feasibility

- **Complexity**: computational complexity considers how the resources needed to computationally solve a problem scale with the the input size n to a problem [Aaronson, 2011]
  (i) *Efficient*: run-time is upper-bounded by a *polynomial* function of n, solvable by classical computation
  (ii) *Inefficient*: run-time is lower-bounded by an *exponential* function of n, higher complexity class, cannot be solved by classical computation (e.g. EXPTIME problems)
- **Feasibility**: run-time is efficient but exceeds computational resources needed to carry out e.g. computation takes too long relative to use-case

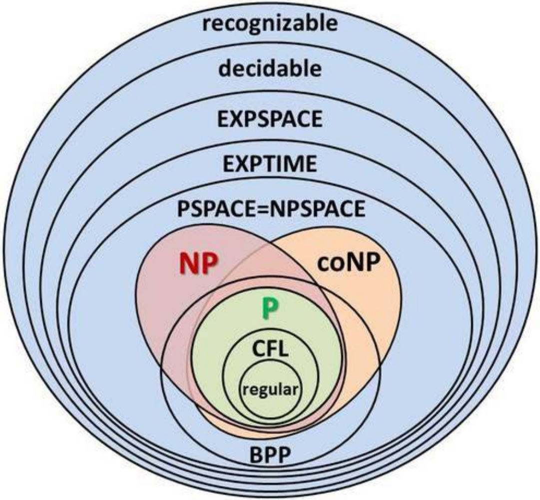
Complexity and feasibility

- **Complexity**: computational complexity considers how the resources needed to computationally solve a problem scale with the the input size n to a problem [Aaronson, 2011]
  (i) *Efficient*: run-time is upper-bounded by a *polynomial* function of n, solvable by classical computation
  (ii) *Inefficient*: run-time is lower-bounded by an *exponential* function of n, higher complexity class, cannot be solved by classical computation (e.g. EXPTIME problems)
- **Feasibility**: run-time is efficient but exceeds computational resources needed to carry out e.g. computation takes too long relative to use-case
- **Complexity zoo**: most problems in the universe are in fact not tractable - the size of higher complexity classes vastly outstrips that of P and NP for example

Complexity and feasibility



Complexity and feasibility



Consistency

- **Consistency**: computational consistency covers the extent to which ethical algorithmic decisions and procedures are *consistent*:

Consistency

- **Consistency**: computational consistency covers the extent to which ethical algorithmic decisions and procedures are *consistent*:
  (i) *process consistency*: are decision-procedures consistent, do / should similar decisions follow similar methodologies?

Consistency

- **Consistency**: computational consistency covers the extent to which ethical algorithmic decisions and procedures are *consistent*:
  (i) *process consistency*: are decision-procedures consistent, do / should similar decisions follow similar methodologies?
  (ii) *outcome consistency*: are outcomes consistent - must two algorithms making ethical decisions come to the same conclusion? Must the set of all ethical AI decisions be consistent with all ethical norms at all times? If not, how to decide what is in/out?

Consistency

- **Consistency**: computational consistency covers the extent to which ethical algorithmic decisions and procedures are *consistent*:
  (i) *process consistency*: are decision-procedures consistent, do / should similar decisions follow similar methodologies?
  (ii) *outcome consistency*: are outcomes consistent - must two algorithms making ethical decisions come to the same conclusion? Must the set of all ethical AI decisions be consistent with all ethical norms at all times? If not, how to decide what is in/out?
- **Maximal consistency**: must classifications by ethical algorithms form maximally consistent set?

Consistency

- **Consistency**: computational consistency covers the extent to which ethical algorithmic decisions and procedures are *consistent*:
  (i) *process consistency*: are decision-procedures consistent, do / should similar decisions follow similar methodologies?
  (ii) *outcome consistency*: are outcomes consistent - must two algorithms making ethical decisions come to the same conclusion? Must the set of all ethical AI decisions be consistent with all ethical norms at all times? If not, how to decide what is in/out?
- **Maximal consistency**: must classifications by ethical algorithms form maximally consistent set?
- **Inconvenient truths**: what happens if algorithmic approaches reveal inherent contradictions within ethical norms? Should such inconvenient truths be censored? Are there algorithmic results that dare not speak their name?

Controllability

- **Controllability**: is an algorithm ethically controllable?

Controllability

- **Controllability**: is an algorithm ethically controllable?
  (i) *control theory*: are controls are available to steer a system to a desired ethical state (e.g. ethical outcome or guarantees on use of ethical methods)?

Controllability

- **Controllability**: is an algorithm ethically controllable?
  (i) *control theory*: are controls are available to steer a system to a desired ethical state (e.g. ethical outcome or guarantees on use of ethical methods)?
  (ii) *ethically controllable*: what types of controls are sought/appropriate
    - '*black box*' methods - vary inputs for some desired output state, inside black box remains unknown
    - '*open box*' methods - control of procedures/calculation

Controllability

- **Controllability**: is an algorithm ethically controllable?
  (i) *control theory*: are controls are available to steer a system to a desired ethical state (e.g. ethical outcome or guarantees on use of ethical methods)?
  (ii) *ethically controllable*: what types of controls are sought/appropriate
    - '*black box*' methods - vary inputs for some desired output state, inside black box remains unknown
    - '*open box*' methods - control of procedures/calculation
  (iii) *Open (offline) v closed loop (online) control* - how/when should input external to algorithm be mandated:
    - open loop control - humans 'in the loop' (but what are trade-offs e.g. drop in efficiency/accuracy?)
    - closed-loop (online) control - (how do we assess risks associated with trusting system?)

Controllability

- **Controllability**: is an algorithm ethically controllable?
  (i) *control theory*: are controls are available to steer a system to a desired ethical state (e.g. ethical outcome or guarantees on use of ethical methods)?
  (ii) *ethically controllable*: what types of controls are sought/appropriate
    - '*black box*' methods - vary inputs for some desired output state, inside black box remains unknown
    - '*open box*' methods - control of procedures/calculation

  (iii) *Open (offline) v closed loop (online) control* - how/when should input external to algorithm be mandated:
    - open loop control - humans 'in the loop' (but what are trade-offs e.g. drop in efficiency/accuracy?)
    - closed-loop (online) control - (how do we assess risks associated with trusting system?)

  (iv) *Noise* - how to control system subject to 'noise' (errors/uncertainty in data) e.g. fairness under measurement error [Liu et al. 2019]

Computational complexity necessitates heuristics

- **Computational complexity necessitates heuristics**:computational complexity limits ethical algorithmic solutions:

Computational complexity necessitates heuristics

- **Computational complexity necessitates heuristics**:computational complexity limits ethical algorithmic solutions:
    (i) **ethical computation inefficient**: there may be no efficient (in a computational sense) algorithm for undertaking the ethical computation [Kearns et al. 2017]

Computational complexity necessitates heuristics

- **Computational complexity necessitates heuristics**:computational complexity limits ethical algorithmic solutions:
  (i) **ethical computation inefficient**: there may be no efficient (in a computational sense) algorithm for undertaking the ethical computation [Kearns et al. 2017]
  (ii) **computationally infeasible**: regardless of efficiency, complexity may grow rapidly to render ethical computation or its generalisation to novel contexts infeasible (high-degree polynomial runtime) given compute resources

Computational complexity necessitates heuristics

- **Computational complexity necessitates heuristics**:computational complexity limits ethical algorithmic solutions:
  (i) **ethical computation inefficient**: there may be no efficient (in a computational sense) algorithm for undertaking the ethical computation [Kearns et al. 2017]
  (ii) **computationally infeasible**: regardless of efficiency, complexity may grow rapidly to render ethical computation or its generalisation to novel contexts infeasible (high-degree polynomial runtime) given compute resources
  (iii) **probabilistic ethics**: use of heuristics/non-deterministic solutions can render ethical computational claims uncertain/probabilistic [Kearns et al. 2017]

Ethical trade-offs

- **Heuristic solutions necessitate ethical trade-offs**: constraints on ethical computation can expose inherent inconsistency of ethical imperatives and require risk-assessment of ethical trade offs

Ethical trade-offs

- **Heuristic solutions necessitate ethical trade-offs**: constraints on ethical computation can expose inherent inconsistency of ethical imperatives and require risk-assessment of ethical trade offs
    (i) *probably approximately ethical*: algorithmic learning of ethical labels approximate - acceptable error rates?

Ethical trade-offs

- **Heuristic solutions necessitate ethical trade-offs**: constraints on ethical computation can expose inherent inconsistency of ethical imperatives and require risk-assessment of ethical trade offs
  - (i) *probably approximately ethical*: algorithmic learning of ethical labels approximate - acceptable error rates?
  - (ii) not unbiased algorithms but choice of acceptable bias (Kleinberg 2016 - Inherent Trade-Offs in the Fair Determination of Risk Scores)

Ethical trade-offs

- **Heuristic solutions necessitate ethical trade-offs**: constraints on ethical computation can expose inherent inconsistency of ethical imperatives and require risk-assessment of ethical trade offs
  (i) *probably approximately ethical*: algorithmic learning of ethical labels approximate - acceptable error rates?
  (ii) not unbiased algorithms but choice of acceptable bias (Kleinberg 2016 - Inherent Trade-Offs in the Fair Determination of Risk Scores)
  (iii) deciding if appropriate heuristic (computationally) ethical problematic in own right
- **BUT** such challenges are ubiquitous - Four C's not in principle barriers to ethical AI

- Computability
- Complexity
- Consistency
- **Controllability**

Controllability

Example - want to control algorithm to prevent use of protected attributes in deep neural network model

Controllability

Example - want to control algorithm to prevent use of protected attributes in deep neural network model

- **Redaction**: there exist methods for redaction of features corresponding to protected class (e.g. race) [McNamara et a. 2019] (provably fair representation learning)

Controllability

Example - want to control algorithm to prevent use of protected attributes in deep neural network model

- **Redaction**: there exist methods for redaction of features corresponding to protected class (e.g. race) [McNamara et a. 2019] (provably fair representation learning)
- **Re-engineering**: however, neural network re-engineers features that resemble protected attributes from other attributes, uses these as new features in training set (Google/AIES'19)

Controllability

Example - want to control algorithm to prevent use of protected attributes in deep neural network model

- **Redaction**: there exist methods for redaction of features corresponding to protected class (e.g. race) [McNamara et a. 2019] (provably fair representation learning)
- **Re-engineering**: however, neural network re-engineers features that resemble protected attributes from other attributes, uses these as new features in training set (Google/AIES'19)
- **Controllability**: can this be controlled for in 'black box' scenario? That is, *how controllable* is the algorithm? Trade-offs between fairness of deep learning model versus how controllable

Conclusions

Conclusions

Takeaway: computational limitations and impossibility results are relevant to formulation of ethical AI and algorithmic governance.

Conclusions

Takeaway: computational limitations and impossibility results are relevant to formulation of ethical AI and algorithmic governance.

- **Ethical AI proposals should consider computational constraints**: ethical AI proposals and regulations consider four C's: *computability, complexity, consistency, controllability*

Conclusions

Takeaway: computational limitations and impossibility results are relevant to formulation of ethical AI and algorithmic governance.

- **Ethical AI proposals should consider computational constraints**: ethical AI proposals and regulations consider four C's: *computability, complexity, consistency, controllability*
- **Risk assessment of uncertainty important**: important to assess types of uncertainty and risk appetite given probabilistic or heuristic-basis for ethical classifications [Bogosian, K 2017]

Conclusions

Takeaway: computational limitations and impossibility results are relevant to formulation of ethical AI and algorithmic governance.

- **Ethical AI proposals should consider computational constraints**: ethical AI proposals and regulations consider four C's: *computability, complexity, consistency, controllability*
- **Risk assessment of uncertainty important**: important to assess types of uncertainty and risk appetite given probabilistic or heuristic-basis for ethical classifications [Bogosian, K 2017]
- **Solutions**: learn from existing systems, such as legal systems, how consistency/complexity challenges are handled