Data science – an occupation should become a profession?

Beuth-Hochschule 10.12.2019

Ursula Garczarek & Detlef Steuer

Cytel Inc, Clinical Research Services ICC, Geneva, ursula.garczarek@cytel.com

Helmut-Schmidt-Universität, Hamburg, steuer@hsu-hh.de

What to expect in the next minutes

- Some definitions, but no mathematics!
- Original artwork copyright Ursula Garczarek!
- Some provocative statements.
- Interruptions (a real debate would be perfect).
- A line of reasoning, why it is essential that data science develops professional ethics and becomes a profession.

Definition of data science (Donoho 2017)

Data science is the science of learning from data; it studies the methods involved in the analysis and processing of data and proposes technology to improve methods in an evidence-based manner. The scope and impact of this science will expand enormously in coming decades as scientific data and data about science itself become ubiquitously available.

1. Data gathering, preparation, and exploration,
2. data representation and transformation,
3. computing with data,
4. data modelling,
5. data visualization and presentation,
6. science about data science.

David Donoho (2017), 50 Years of Data Science,
Journal of Computational and Graphical Statistics, 26:4, 745-766

Occupation vs Profession (Airaksinen, 2009)

A profession is different from an occupation through

- Scientific training

  Knowing what is to be done by understanding the rational, epistemological foundations of professional action

- Autonomy

  A profession can influence the social decisions that regulate its members' work and their related rights and obligations

- Professional ethics

  If the *public needs this expertise and therefore cannot unproblematically reject, challenge, or ignore the professional advice and the influence of their work, professional ethics becomes a key issue when the public evaluates the potential bias of professional work in relation to the quality of their life. (Trust!)*

Timo Airaksinen, The Philosophy of Professional Ethics, 2009,
INSTITUTIONAL ISSUES INVOLVING ETHICS AND JUSTICE, edited Robert Charles Elliot, Vol 1, in Encyclopedia of Life Support Systems

Does data science form a profession?

- The Code of Conduct of the RSS defines statistics as a profession.
- Data science is only in part statistics. It draws a lot from i.e. computer science.
- Data science was born as a tool in a commercial surrounding, where data was abundant, cheap to store, cheap to process.
- No common rational, epistemological foundation of data science education. Some start from the inferential framework of statistics, some start from computational learning theory in machine learning, or even economics!
- A very weak education in ethical aspects of data science, but improving. Most new study programs include some ethics courses.

No! But why not and why would it help?

Our statement for debate

Data science is in the focal point of current societal development.

Without becoming a profession with professional ethics,

data science will fail in building trust

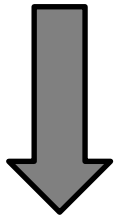
in its interaction with

and it's much needed contributions to

society!

Line of reasoning

Societal impact + Ethical issues + Influence



Responsibility

Lack of knowledge

Lack of skills

Lack of power

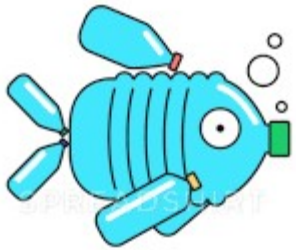
Lack of interest

Societal impact

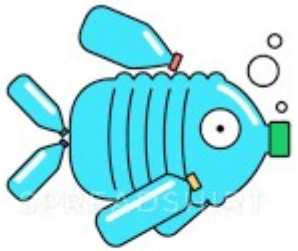
Ethical issues

Influence

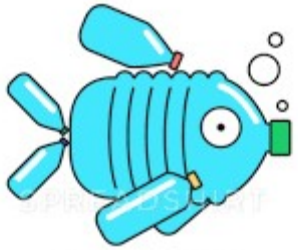
Responsibility



Lack of knowledge



Lack of Skills



Lack of power

Reasoning step by step

Agree or Disagree?

“Societal impact of data science is large!”

- Vast amount of data (on individuals and the general public)
- Increased computational power
- Fast, low cost processing of large databases
- Used for decision making on all aspects of (human) life

  (Weapons of Math Destruction, Cathy O’Neill)
- (In general the progress of computational power stresses human imagination and human scales of time. In my opinion this changes the **quality** of applying an algorithm, **not only** the **quantity**.)

Agree or disagree?

“Using algorithms can raise ethical issues!”

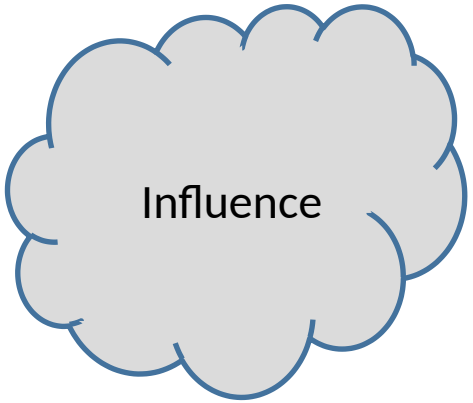
- Autonomous machines: a threat to free will and responsibility
- Bias, discrimination and exclusion (not only black box algorithms)
- Algorithmic profiling: personalisation versus collective benefits
- Preventing massive files while enhancing AI: seeking a new balance
- Quality, quantity, relevance: the challenges of data curated for AI
- Human identity before the challenge of artificial intelligence

Commission Nationale Informatique & Liberte, Algorithms and artificial intelligence: CNIL’s report on the ethical issues, 25 May 2018

Agree or disagree?

“Data scientists have influence on design and application of algorithms”

1. Data gathering, preparation, and exploration
2. data representation and transformation
3. computing with data
4. data modelling
5. data visualization and presentation
6. science about data science.

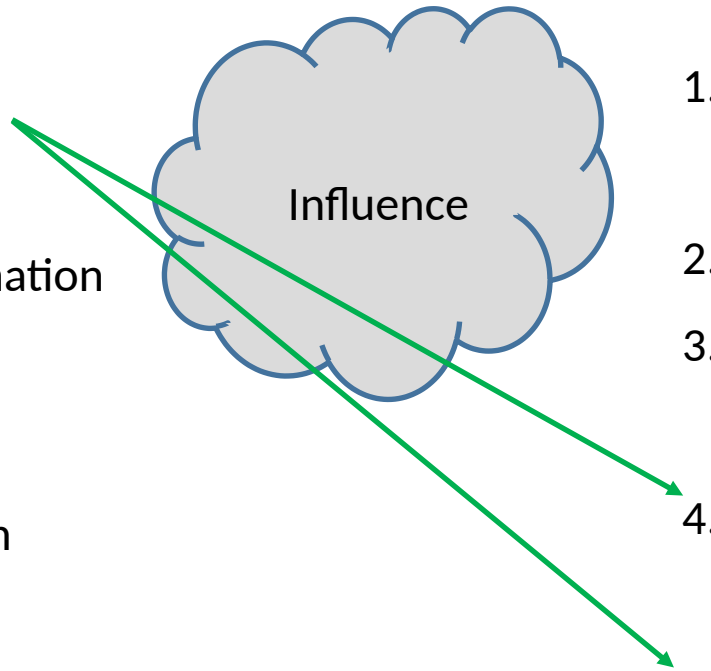


1. Autonomous machines: a threat to free will and responsibility
2. Bias, discrimination and exclusion
3. Algorithmic profiling: personalisation vs collective benefits
4. Preventing massive files while enhancing AI: seeking a new balance
5. Quality, quantity, relevance: the challenges of data curated for AI

Agree or disagree?

“Data scientists have influence on design and application of algorithms”

1. Data gathering, preparation, and exploration
2. data representation and transformation
3. computing with data
4. data modelling
5. data visualization and presentation
6. science about data science.

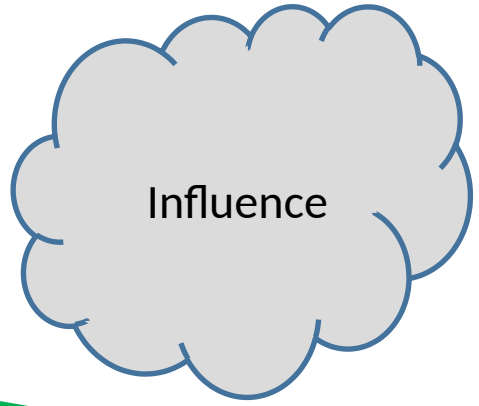


1. Autonomous machines: a threat to free will and responsibility
2. Bias, discrimination and exclusion
3. Algorithmic profiling: personalisation vs collective benefits
4. Preventing massive files while enhancing AI: seeking a new balance
5. Quality, quantity, relevance: the challenges of data curated for AI

Agree or disagree?

“Data scientists have influence on design and application of algorithms”

1. Data gathering, preparation, and exploration
2. data representation and transformation
3. computing with data
4. data modelling
5. data visualization and presentation
6. science about data science.

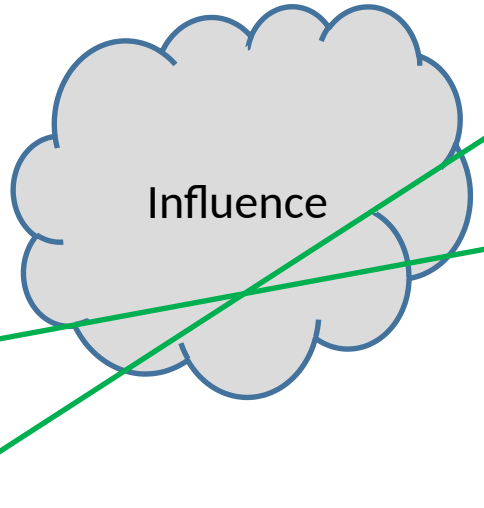


1. Autonomous machines: a threat to free will and responsibility
2. Bias, discrimination and exclusion
3. Algorithmic profiling: personalisation vs collective benefits
4. Preventing massive files while enhancing AI: seeking a new balance
5. Quality, quantity, relevance: the challenges of data curated for AI

Agree or disagree?

“Data scientists have influence on design and application of algorithms”

1. Data gathering, preparation, and exploration
2. data representation and transformation
3. computing with data
4. data modelling
5. data visualization and presentation
6. science about data science.

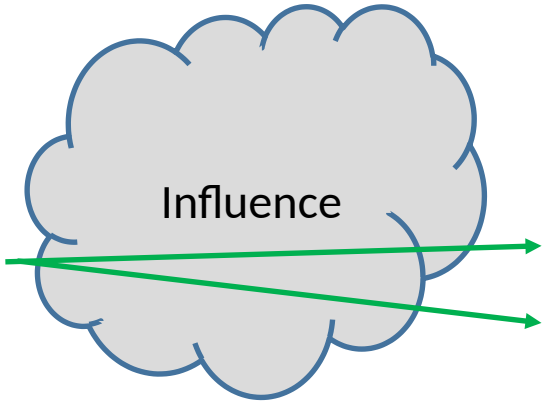


1. Autonomous machines: a threat to free will and responsibility
2. Bias, discrimination and exclusion
3. Algorithmic profiling: personalisation vs collective benefits
4. Preventing massive files while enhancing AI: seeking a new balance
5. Quality, quantity, relevance: the challenges of data curated for AI

Agree or disagree?

“Data scientists have influence on design and application of algorithms”

1. Data gathering, preparation, and exploration
2. data representation and transformation
3. computing with data
4. data modelling
5. data visualization and presentation
6. science about data science.

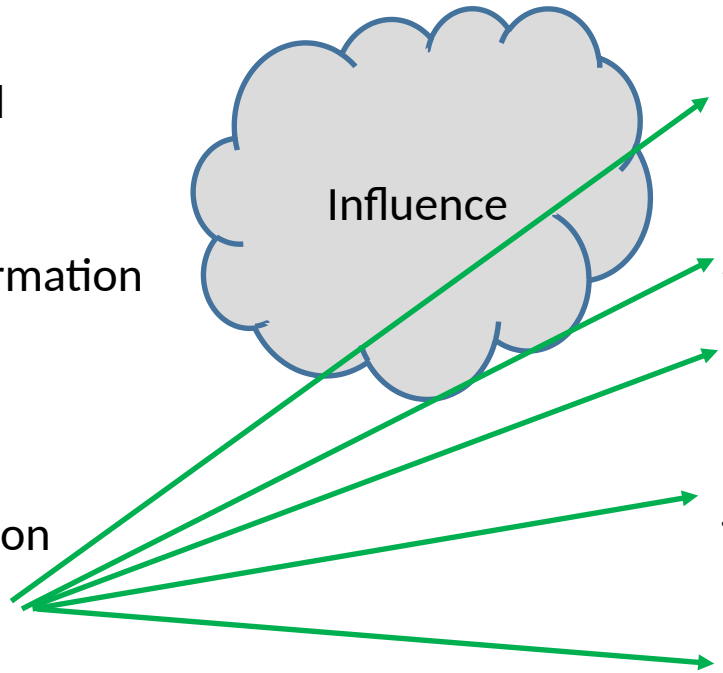


1. Autonomous machines: a threat to free will and responsibility
2. Bias, discrimination and exclusion
3. Algorithmic profiling: personalisation vs collective benefits
4. Preventing massive files while enhancing AI: seeking a new balance
5. Quality, quantity, relevance: the challenges of data curated for AI

Agree or disagree?

“Data scientists have influence on design and application of algorithms”

1. Data gathering, preparation, and exploration
2. data representation and transformation
3. computing with data
4. data modelling
5. data visualization and presentation
6. science about data science.



1. Autonomous machines: a threat to free will and responsibility
2. Bias, discrimination and exclusion
3. Algorithmic profiling: personalisation vs collective benefits
4. Preventing massive files while enhancing AI: seeking a new balance
5. Quality, quantity, relevance: the challenges of data curated for AI

Agree or disagree?

“Influence leads to responsibility”

- GI (Gesellschaft für Informatik): No data science specific aspects, „Do no harm!“
- GfKl (Data Science Society): Trying to establish a working group on ethical guidelines
- ASA: Because society depends on informed judgments supported by statistical methods, all practitioners of statistics—regardless of training and occupation or job title—have an obligation to work in a professional, competent, respectful, and ethical manner.
- ACM: Computing professionals' actions change the world. To act responsibly, they should reflect upon the wider impacts of their work, consistently supporting the public good.
- RSS/IfoA: Data science can be both beneficial and detrimental to individuals and/or society …members may seek to understand the impact of their work. (Nov 19)

Agree or disagree?

“Currently among those that apply data science there are too many that lack knowledge on data science and on ethics to act responsibly”

- Deficits in data science education on data science (esp. statistics)

  “*Knowing what is to be done by understanding the rational, epistemological foundations of professional action*”

  - Data scientists trained in 3 month courses
  - No „Science about data science“ at all for many
  - Statisticians SaDS on World → Data → Knowledge
  - Machine Learners SaDS covers Data → Representation → Knowledge → Representation
  - Engineers SaDS Algorithm? → World?

- Deficits in data science education on ethics
  - No training in ethics or ethical reasoning, (moral) debate, politics,…

Lack of knowledge

Agree or disagree?

“Data scientists need to become better in communicating with lay people on data science and related ethical issues to act responsibly”

- Those that have all the knowledge are still limited in their effectiveness to take responsibility if they can not talk with lay people about it.

➢ Talking about methods and methodology with lay people

  ➢ Education in general consulting skills is still neglected in academic training

➢ Discuss ethics and own moral views related to data science

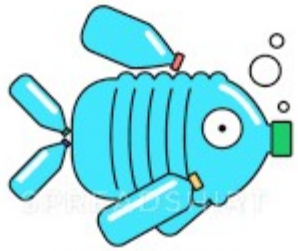
  ➢ Data scientists have to overcome shyness and ignorance

Lack of Skills

Agree or disagree?

“Currently, the community of those that want to practice responsible data science lack power to fight irresponsible data science”

- There are very many companies that sell expensive and less than useful data science services.
- Such poor data science **ruins** the reputation of data science and **ruins** companies and institutions.
- Data science sharlatanes also do harm to society! (No only outside academia!)
- In companies, departments for data science, analytics, statistics,etc. fight against each other for funding.
- In the academic world ML and statisticians often fight fruitless fights about who is fundamentally wrong or right.
- **Together we are strong!**



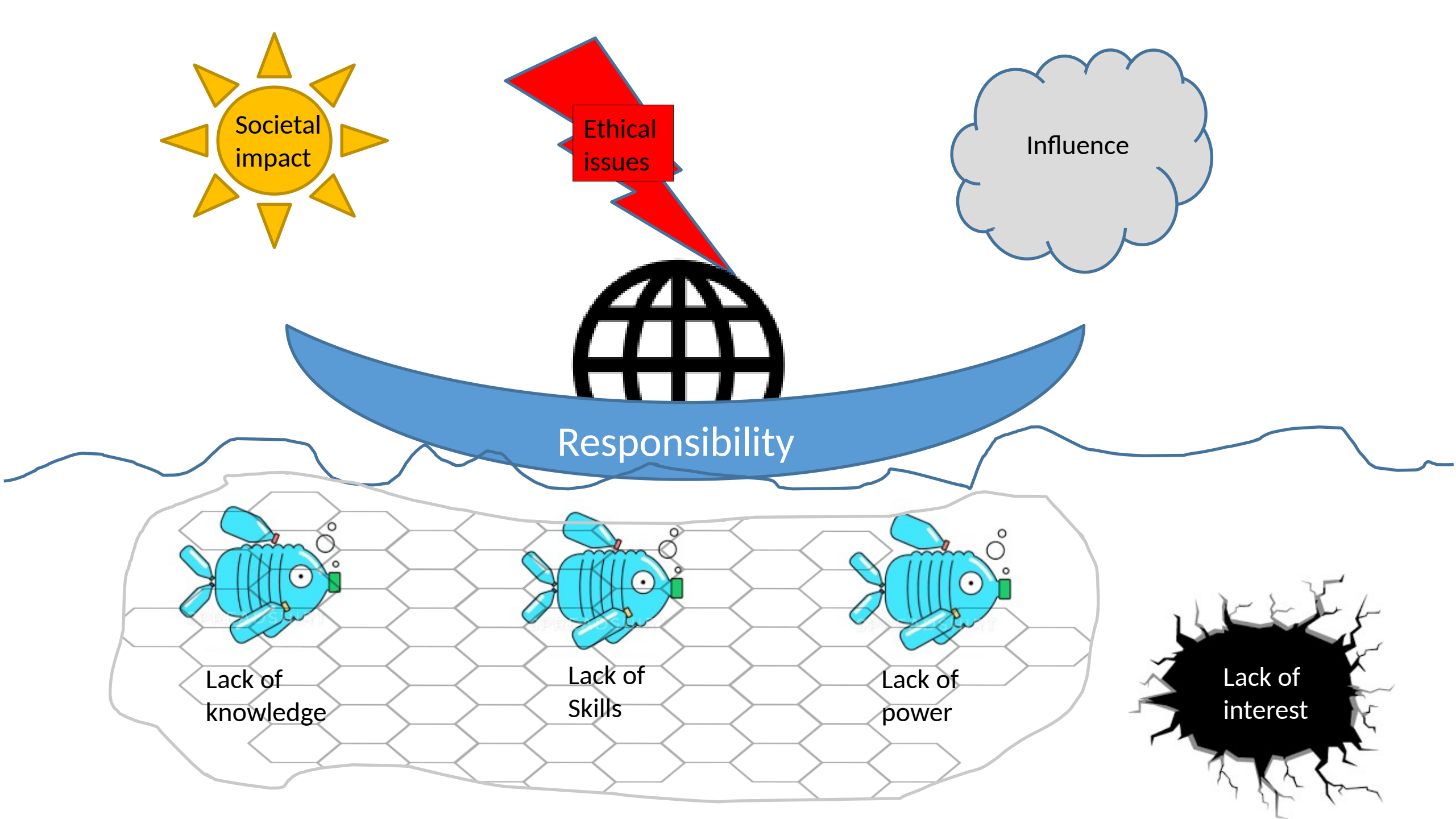
Lack of power

Agree or disagree?

“Lack of knowledge, skills, and power can be addressed by data science as a profession with professional ethics”

- Support the individual data scientist by education and guidelines in every day ethical decision making .
  - No need to re-invent the wheel for the individual.

- Written rules of conduct for data science services help to establish a relationship of trust between data scientists, their clients, their employers, and society.
- Being trusted as a profession increases social status, reputation and power for a professional.
- In case of conflicts of interests an ethical guideline under the maintainership of some professional society may offer an arbitration process.


Societal impact
Ethical issues
Influence
Responsibility
Lack of knowledge
Lack of Skills
Lack of power
Lack of interest

Agree with any of those reasons for lack of interest in professional ethics?

- Professional societies with memberships, codes of conducts asf are elitarian and limit scientific freedom and hinder innovation.
- What is right and wrong is defined by law.
- What is right and wrong is my personal belief.
- Corporate responsibility is more important than professional ethics, as those who implement algorithms are not responsible for their societal impact.
- Paper does not blush, and most code of conducts are hypocritical.
- My opinion: All those are ~~cheap~~ invalid excuses!

Help establishing the debate!

Data science is in the focal point of current societal development.

Without becoming a profession with professional ethics,

data science will fail in building trust

in its interaction with

and its much needed contributions to

society!

Thank you!