



# Data privacy and ethical issues in data science

2021-10-04

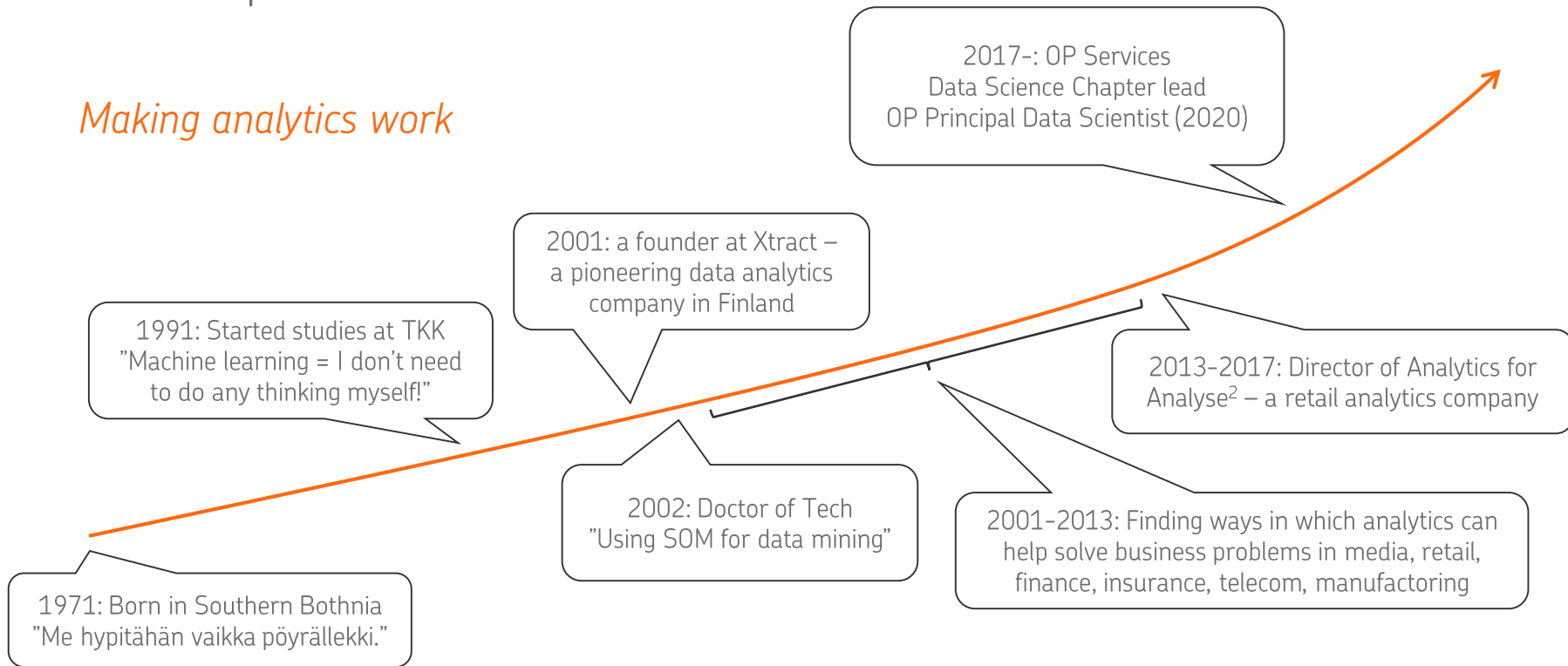
Dr Juha Vesanto / OP Principal Data Scientist



# Dr Juha Vesanto

## OP Principal Data Scientist

### *Making analytics work*



# Data privacy and ethical issues in data science

## Ethics ... what?

- Why is it an issue?
- Examples: how to do damage even if you don't mean to

## Ethical analytics

- What does ethics mean for data science?
- How to achieve it?
- Ethical design

## Enforcing ethics

- GDPR
- Data privacy
- Data protection



How to do damage even if you don't mean to

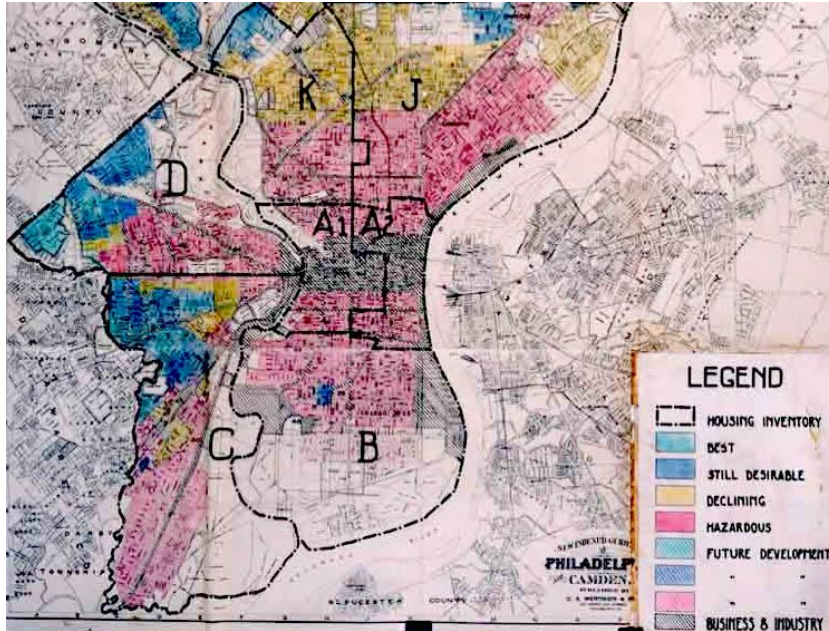
Data privacy and ethical issues in data science

This is math!  
What has ethics to do with it?

*Compare this statement to: "guns don't kill people, people do"*



# Example 1: Redlining



- Redlining
  - "refuse (a loan or insurance) to someone because they live in an area deemed to be a poor financial risk"
- Originates from 1960's with the practice of insurance companies to withhold house insurance (or banks refusing loans) from "bad" neighborhoods
- Result: the bad neighborhoods became increasingly worse

<https://en.wikipedia.org/wiki/Redlining>

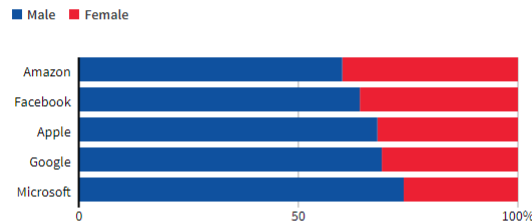
# Example 2: Amazon's recruitment algorithm

- Amazon used an analytical scoring method to pick candidates for interviews
- Because software engineers are dominantly male, the algorithm presumed that being female decreases the likelihood to be recruited
- Result: the women did not even get to the interview

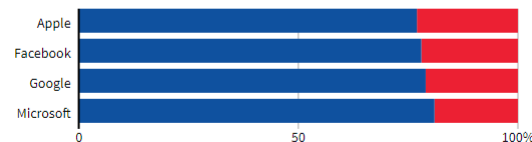
## Dominated by men

Top U.S. tech companies have yet to close the gender gap in hiring, a disparity most pronounced among technical staff such as software developers where men far outnumber women. Amazon's experimental recruiting engine followed the same pattern, learning to penalize resumes including the word "women's" until the company discovered the problem.

### GLOBAL HEADCOUNT



### EMPLOYEES IN TECHNICAL ROLES



Note: Amazon does not disclose the gender breakdown of its technical workforce.

Source: Latest data available from the companies, since 2017.

By Han Huang | REUTERS GRAPHICS

<https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>

# Example 3: British entrance exams 2020

## Boris Johnson blames 'mutant algorithm' for exams fiasco

Union accuses PM of 'shrugging away disaster' after Johnson says problems have been 'sorted out'



▲ Boris Johnson sits in on a class at Castle Rock school in Coalville, Leicestershire. Photograph: Jack Hill/AP

Boris Johnson got an angry response after telling school pupils that the exam results crisis was caused by a "mutant algorithm" and he was glad it had been "sorted out".

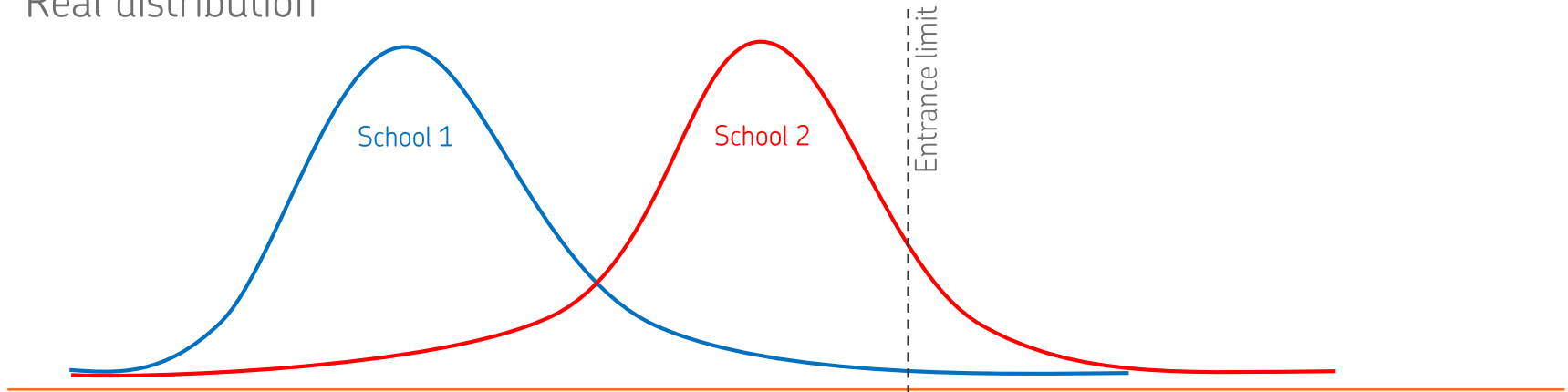
- Summer 2020 entrance exams were replaced with an algorithm:
  - Take into account the prior 3 year performance of each school
  - Ranks the student internally in each school. Rank schools wrt the history.
  - [https://en.wikipedia.org/wiki/Ofqual\\_exam\\_results\\_algorithm](https://en.wikipedia.org/wiki/Ofqual_exam_results_algorithm)
- Result: none of the pupils from worse (public) schools made it. Irrespective of individual performance.
  - The algorithm results were later scrapped, and teacher assesments were used instead...

<https://www.theguardian.com/politics/2020/aug/26/boris-johnson-blames-mutant-algorithm-for-exams-fiasco>



# Ofqual failure: dealing with the tails

Real distribution



Statistical estimate (box+whiskers)



# Example 4: Social media

- Social media algorithms
  - Optimize for click-through-rate → any kind of reaction is good
  - Conflict is good, because the reaction is strong
  - Create localized sub-spaces where all content enforces pre-existing opinions
- Result: world divides into sub-groups who do not communicate, or if they do, prefer discord and conflict
- Solutions:
  - Emphasizing external/random inputs (along with the 'preferred' inputs): YLE
  - ???



# Ethical analytics

Data privacy and ethical issues in data science

# Ethics

- Study on what is **good morality** ja **how to act right**
- Why is this relevant for **banks** (for example)?
  - Banks have an important role in society by **creating new opportunities** for individuals and companies to evolve and build. This is achieved by making it possible to make various **financial arrangements**. Typically loans.
  - Banks are dependent on their customers. This relationship is built **mutual trust** that both parties act as agreed.
  - Society has given banks some extra rights, which is also based on the mutual trust between the **bank and the society**.

Banking business is based on trust.

# Ethical action creates the basis for trust

Can I trust, that my  
information is correct?

Can I trust, that the bank  
operates fairly?

## Trust

Can I trust, that my  
Information stays secure?

Can I trust, that the analysis  
and decisions concerning me  
are done right?

# Artificial intelligence as a threat

Currently, a particular need for trust is concerned with **artificial intelligence**.

## Why is AI a threat?

- Why now?
  - Several component technologies have just matured to a level that makes them useful
  - Publicity has been raised by numerous well known leaders and scientists
- Why is so threatening about it?
  - New technology is always scary
  - People are afraid of becoming victims of artificial intelligence

## How to respond?

- Again, this is about trust: can one trust that the AI works correctly / fairly?
- The trust is created by
  - Communicating openly
  - Being honest
  - Emphasizing that a human is ultimately responsible



# OP's ethical guidelines

## People-first approach

- We will use data and AI responsibly and for the good of our customers. We will define the objectives guiding our use of AI clearly and refine them if necessary based on changed data, technical possibilities and the working environment.

## Transparency and openness

- We will act openly in our relations with customers, partners and stakeholders, ensuring sufficient transparency for the evaluation of the AI we have developed. We will discuss our use of AI use openly and subject our work to public scrutiny.

## Impact evaluation

- We will carefully study the impacts of the choices we make in our work on our customers and the society around us. Our choices regarding AI utilisation are always responsible.

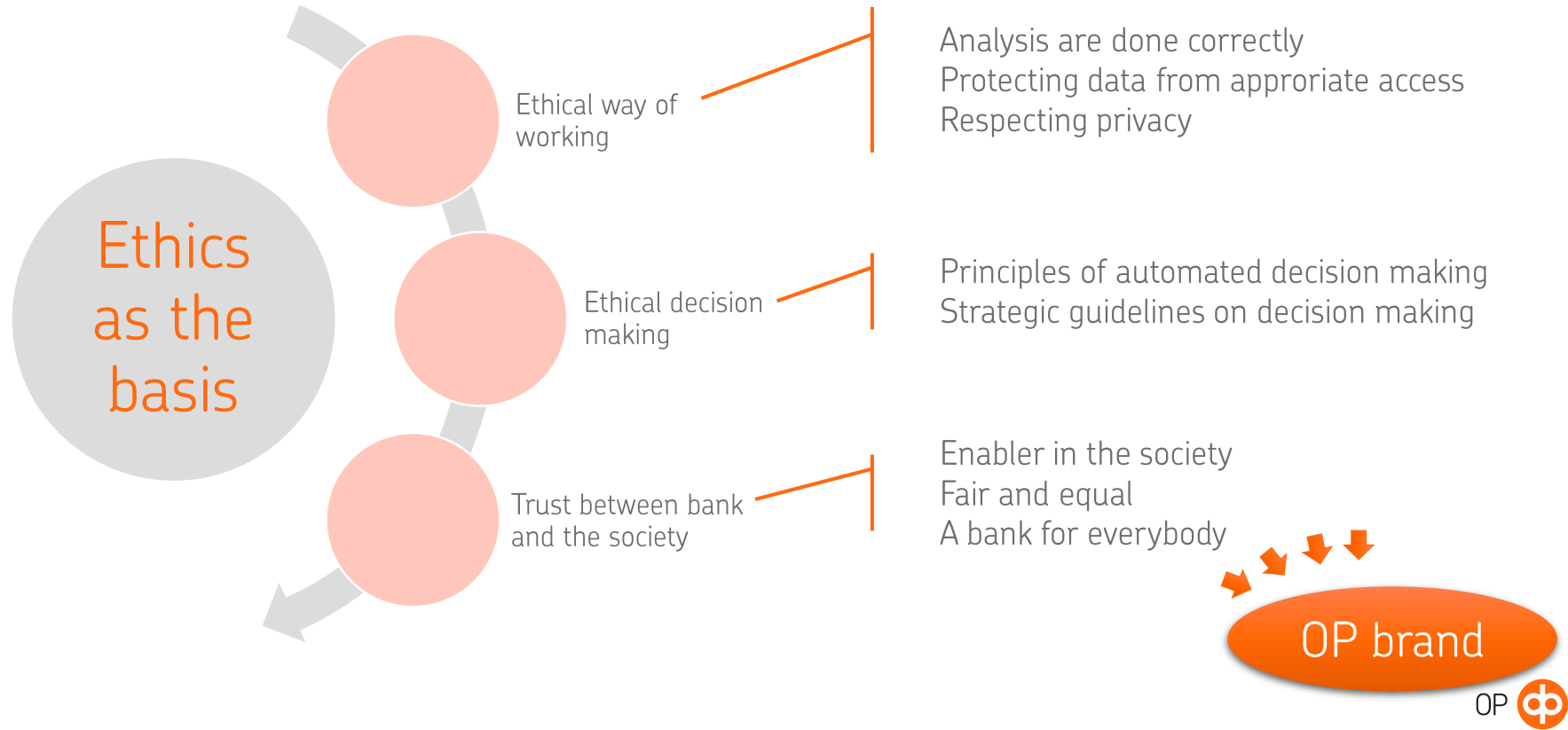
## Ownership

- We will define owners for the principles guiding our operations and for the algorithms we have developed, and will ensure the ethics of AI throughout the lifecycle.

## Privacy protection

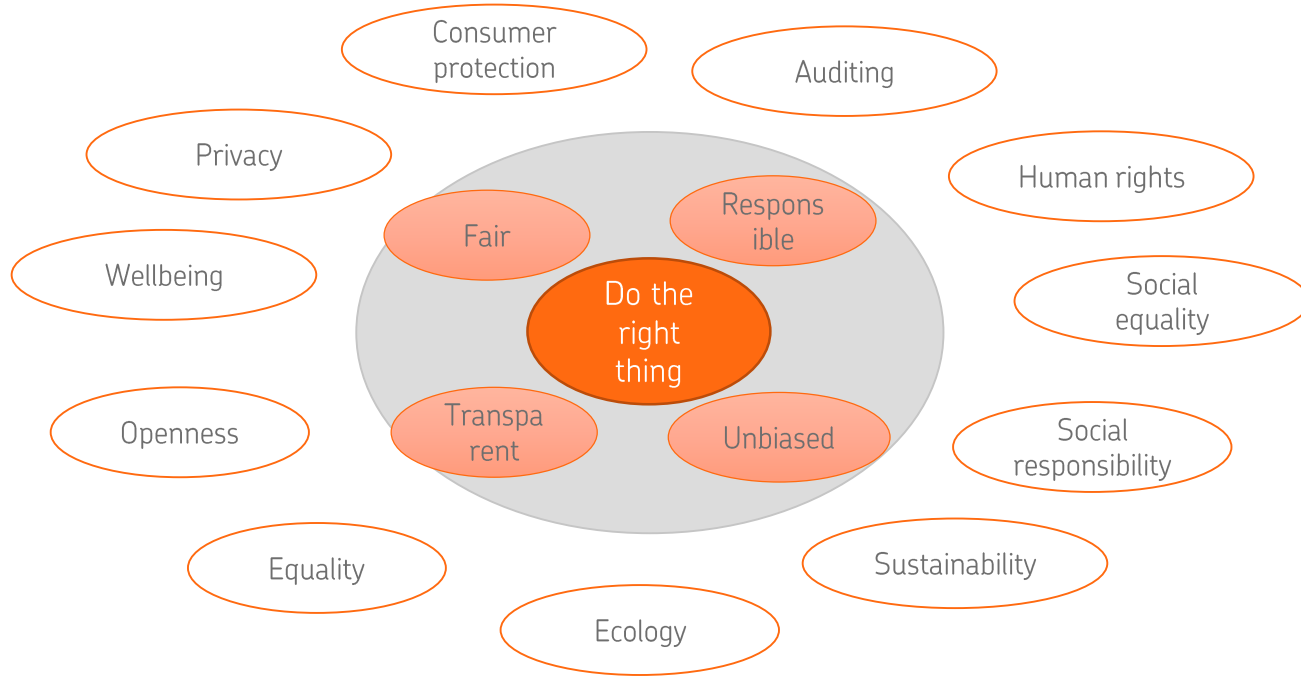
- We will guarantee privacy and personal data protection for the individuals represented in the data we use in accordance with our data protection principles.

# Ethical way of working



... so, what does this mean in practice?

# What is expected from an ethical company?



# What does "ethical analytics" mean?

## Own activities

- Making sure the data is good quality
- Taking care of data protection
- Respecting people's right to privacy

## Correctness of analysis

- Accuracy
- Timeliness
- Unbiased

## Strategic choices

- Treating people as individuals → personal level data & models
- Avoiding discrimination: eg gender-agnostic models
- Avoiding the usage of irrelevant data

# Ethical design

Digital Ethics

Ethical design

AI Ethics

Ethically Aligned  
Design

Working ethically is  
part of everyday work

- Are you and the company working on ethical solid ground?
- Are there any potentially unethical aspects to the product (being developed)?





## Enforcing ethics

Data privacy and ethical issues in data science

# Wishing for ethical behaviour is ... not enough

Society tries to enforce companies to act fairly

- Laws (and recommendations)
- Governmental supervision (eg European Central Bank, Finanssivalvonta)
- Promoting transparency (eg <https://ai.hel.fi/> )

... but also

- The fourth state power = journalism
- Ethical choices preferred by consumers (eg <https://saidot.ai/> )

# Some laws governing data science (in banking)

## GDPR

- One of the most advanced data protection and privacy laws
- Main points:
  - Openness and right to check your own data
  - Consumer rights wrt automated decision making
  - Concerned with personal information

## Tietosuoja-asetus (FI)

- Finnish regulation on using data
- Main principles:
  - Legality, moderation and and transparency
  - Dependency on the intended purpose
  - Data minimization
  - Accuracy & correctness
  - Confidentiality
  - Data protection
  - Limited storage time

## Yhdenvertaisuuslaki (FI)

- Law against discrimination
- Protection in particular concerns
  - Age
  - Nationality / origin
  - Language
  - Religion
  - Opinion
  - Political activity / trade union
  - Family relations
  - Health / disability
  - Sexual orientation

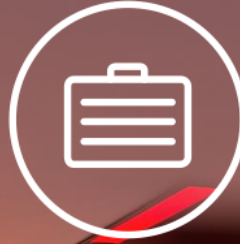
# Privacy protection

## Anonymous data

- Data either
  - Has no personal information at all (eg some IoT data with no connections to persons)
  - Has been aggregated so that any single item in the data is combination of several individuals
- Strong anonymity: requires that not even singular individuals are exposed, even with external data

## Using pseudonymns

- To be able to use data in data science, a weaker form of data protection is usually applied: pseudonymization
- Does not make data anonymous, but adds a layer of data protection (=at least individuals are not directly exposed)
- Techniques: Secret key; Hash function, possibly with salt, possibly with stored key; Deterministic encryption; Tokenization



# Summary of ethics in data science

Data privacy and ethical issues in data science

# Things to remember

- You can do damage even if you don't mean to. So, think it through:
  - How are the results going to be used?
  - What happens to the tails (rare cases)? Are they being treated fairly, or at all?
  - Are the outputs somehow skewed?
- Ethical analytics
  - Act professionally (know your data, be careful, check the results)
  - Respect privacy and protect the data
  - Keep your eyes open wrt discrimination
- Legal obligations
  - Data protection, GDPR
  - Transparency, openness



A young boy with brown hair, wearing a light blue and white striped long-sleeved shirt, dark green pants, and bright yellow rubber boots, is crouching on a grassy forest floor. He is holding a large, silver metal watering can with a long spout and pouring water onto the base of a small, young evergreen tree. The tree has green needles and small brown cones. The background shows a forest with bare trees and sunlight filtering through the leaves.

Thanks!

Dr Juha "juuso" Vesanto  
[juha.vesanto@op.fi](mailto:juha.vesanto@op.fi)