Critical Algorithm Studies

ACCOUNTABILITY, BIAS AND TRANSPARENCY

A Critical Look at the AMS Algorithm
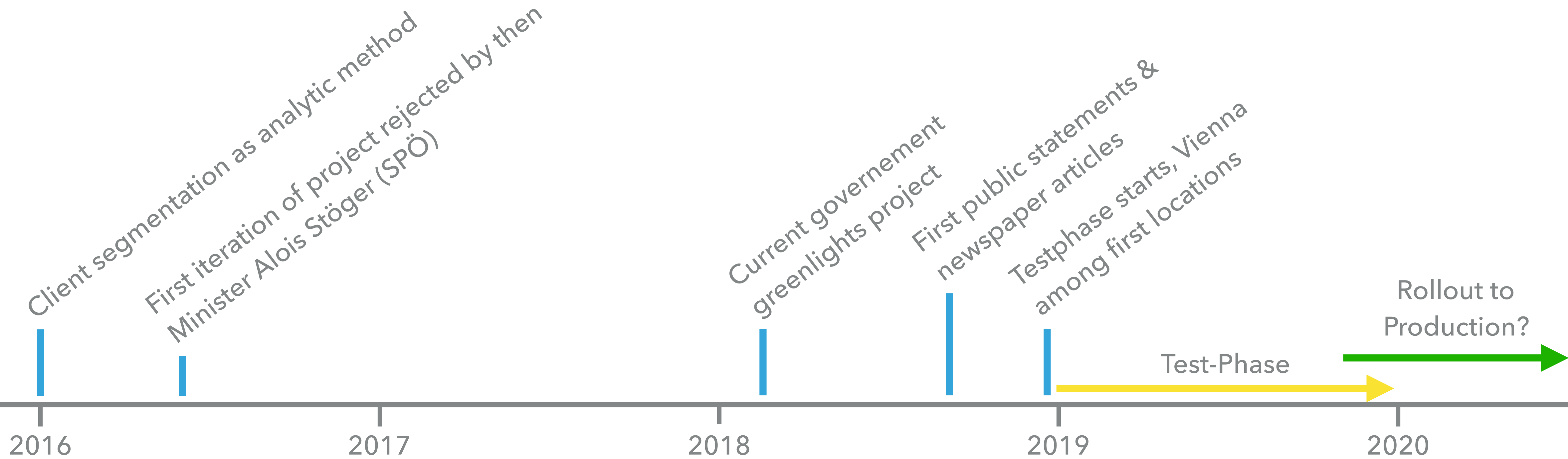
Florian Cech
Centre for Informatics and Society

Centre for Informatics and Society @ TU Wien
Critical Algorithm Studies

▶ Research / Knowledge Centre

▶ General focus: *Digital Transformation*

  ¬ … interfacing technology and society

▶ Special Focus: *Critical Algorithm Studies*

  ¬ Critical analysis of algorithmic systems

  ¬ Issues include accountability, transparency, bias, discrimination & ethics

  ¬ Practical focus on transparency and accountability improvements for algorithmic systems

Centre for Informatics and Society
der Fakultät für Informatik an der TU Wien

AMS Algorithm - Timeline

Client segmentation as analytic method

First iteration of project rejected by then Minister Alois Stöger (SPÖ)

Current governement greenlights project

First public statements & newspaper articles

Testphase starts, Vienna among first locations

Test-Phase

Rollout to Production?

2016   2017   2018   2019   2020

Das AMS-Arbeitsmarkt-chancen-Modell

Jürgen Holl
Günter Kernbeiß
Michael Wagner-Pinter

Dokumentation zur Methode

AMS Algorithm - Information Sources

‣ „Specification" document published by the Synthesis GmbH, a third party data analysis and research company

‣ Interviews and Statements by Johannes Kopf

‣ Internal AMS presentations

  ⌐ ... partly accidentally published?

‣ Reports by Austrian Court of Audit

‣ OECD Technical Workshop Presentation on Profiling Tools in the labor market

Centre for Informatics and Society
der Fakultät für Informatik an der TU Wien

# „Machine Learning" - AMS Algorithm

Quelle:
XKCD by Randall Munroe
https://xkcd.com/1838/

▸ Statistical profiling tool for client segmentation

▸ Logistic regression predicts job-seeker's chances in the labor market based on prior observations

▸ Training dataset consists of AMS client's PII

  ⌐ … at least partially self-reported data!

▸ Prediction classifies into three categories (low, medium and high chances) with two target functions

  ⌐ Short-Term: 90 days of employment w/in 7 months

  ⌐ Long-Term: 180 days of employment w/in 24 months

Centre for Informatics and Society
der Fakultät für Informatik an der TU Wien

Input Data - Client Information via AMS Data Warehouse

▸ Gender [m | f]

▸ Age [0-29 | 30-49 | 50+]

▸ Nationality [AT | EU | Non EWR]

▸ Education
[mandatory school | vocational school | AHS, FH, University]

▸ Health impairments [y | n]

▸ Care obligations - only for women! [y | n]

▸ Type of occupation [production | service industry]

▸ Type of regional labor market [5 types - unclear!]

▸ Prior occupational career

⌐ Days of gainful employment within 4 years prior [>= 75% | < 75%]

⌐ Number of opened cases with AMS within 4 x 1 year prior intervals
[0 cases | 1 case | 1+ cases in 2 intervals | 1+ cases in 3+ intervals]

⌐ Duration of opened cases > 180 days
[0 cases | 1+ cases]

⌐ Support measures claimed
[0 | 1+ supportive | 1+ educational | 1+ subsidized]
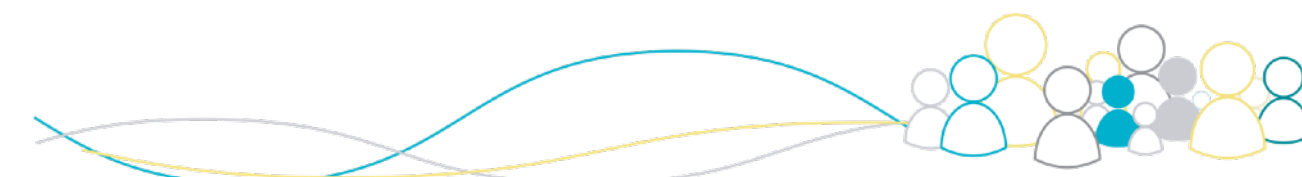
Impact coefficients of PII attributes - example short term model

▸ Impact of PII attributes based on observations in previous years

▸ **Positive** impacts: frequency of business cases, occupation in manufacturing/industry, education in vocational schools

$$BE\_INT$$
$$= f \, (\, 0{,}10$$
$$- \, 0{,}14 \times GESCHLECHT\_WEIBLICH$$
$$- \, 0{,}13 \times ALTERSGRUPPE\_30\_49$$
$$- \, 0{,}70 \times ALTERSGRUPPE\_50\_PLUS$$
$$+ \, 0{,}16 \times STAATENGRUPPE\_EU$$
$$- \, 0{,}05 \times STAATENGRUPPE\_DRITT$$
$$+ \, 0{,}28 \times AUSBILDUNG\_LEHRE$$
$$+ \, 0{,}01 \times AUSBILDUNG\_MATURA\_PLUS$$
$$- \, 0{,}15 \times BETREUUNGSPFLICHTIG$$
$$- \, 0{,}34 \times RGS\_TYP\_2$$
$$- \, 0{,}18 \times RGS\_TYP\_3$$
$$- \, 0{,}83 \times RGS\_TYP\_4$$
$$- \, 0{,}82 \times RGS\_TYP\_5$$
$$- \, 0{,}67 \times BEEINTRÄCHTIGT$$
$$+ \, 0{,}17 \times BERUFSGRUPPE\_PRODUKTION$$
$$- \, 0{,}74 \times BESCHÄFTIGUNGSTAGE\_WENIG$$
$$+ \, 0{,}65 \times FREQUENZ\_GESCHÄFTSFALL\_1$$
$$+ \, 1{,}19 \times FREQUENZ\_GESCHÄFTSFALL\_2$$
$$+ \, 1{,}98 \times FREQUENZ\_GESCHÄFTSFALL\_3\_PLUS$$
$$- \, 0{,}80 \times GESCHÄFTSFALL\_LANG$$
$$- \, 0{,}57 \times MN\_TEILNAHME\_1$$
$$- \, 0{,}21 \times MN\_TEILNAHME\_2$$
$$- \, 0{,}43 \times MN\_TEILNAHME\_3)$$

▸ Multiple models for different populations

  ⌐ Separation by PII and data quality / completeness

▸ **Negative** impacts: gender 'female', clients with health issues, background in immigration, obligations for care

Quelle: http://www.forschungsnetzwerk.at/downloadpub/arbeitsmarktchancen_methode_%20dokumentation.pdf

Centre for Informatics and Society
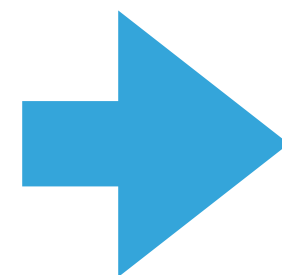der Fakultät für Informatik an der TU Wien

Example Calculation

▸ Sample Persona 1

Female	-0,14
32 years old	-0,13
EU citizen	0,16
Bachelor's degree	0,01
1 prior case < 180 days	0,65
Responsibility f. Care	-0,15
Total + 0.10	**0,5**

Logistic Transformation

$$f(x) = \frac{L}{1 + e^{-k \cdot (x - x_0)}}$$
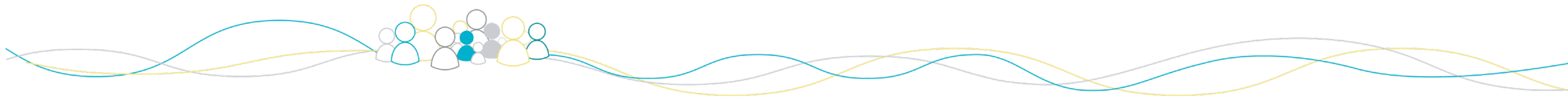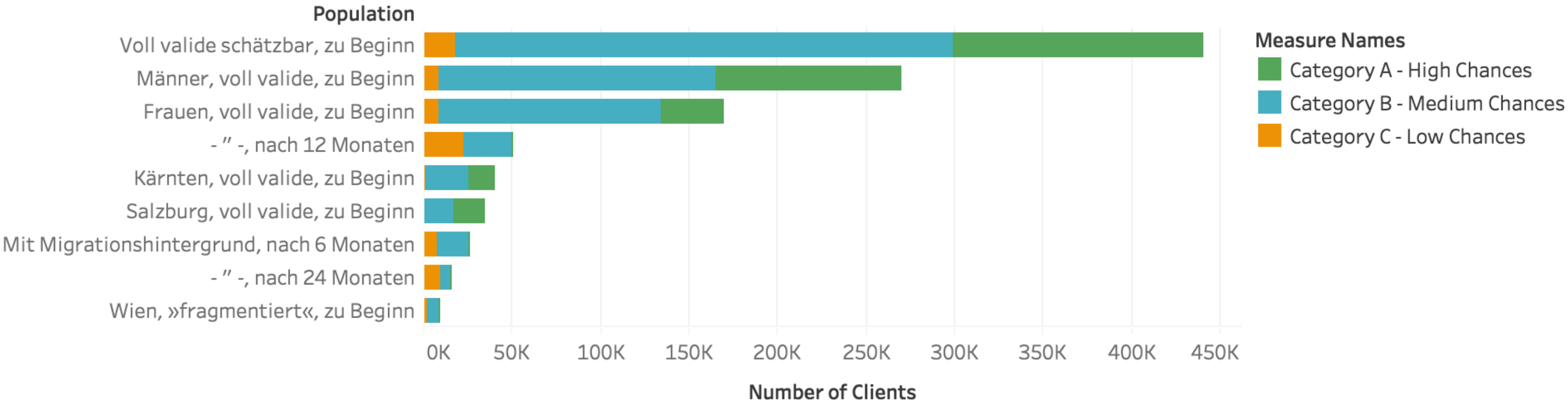
$$f(0.5) = \frac{1}{1 + e^{-1 \cdot (0.5 - 0)}} = 60\,\%$$

Group A:	66% - 100%
Group B:	25% - 66%
Group C:	0% - 25%

Centre for Informatics and Society
der Fakultät für Informatik an der TU Wien
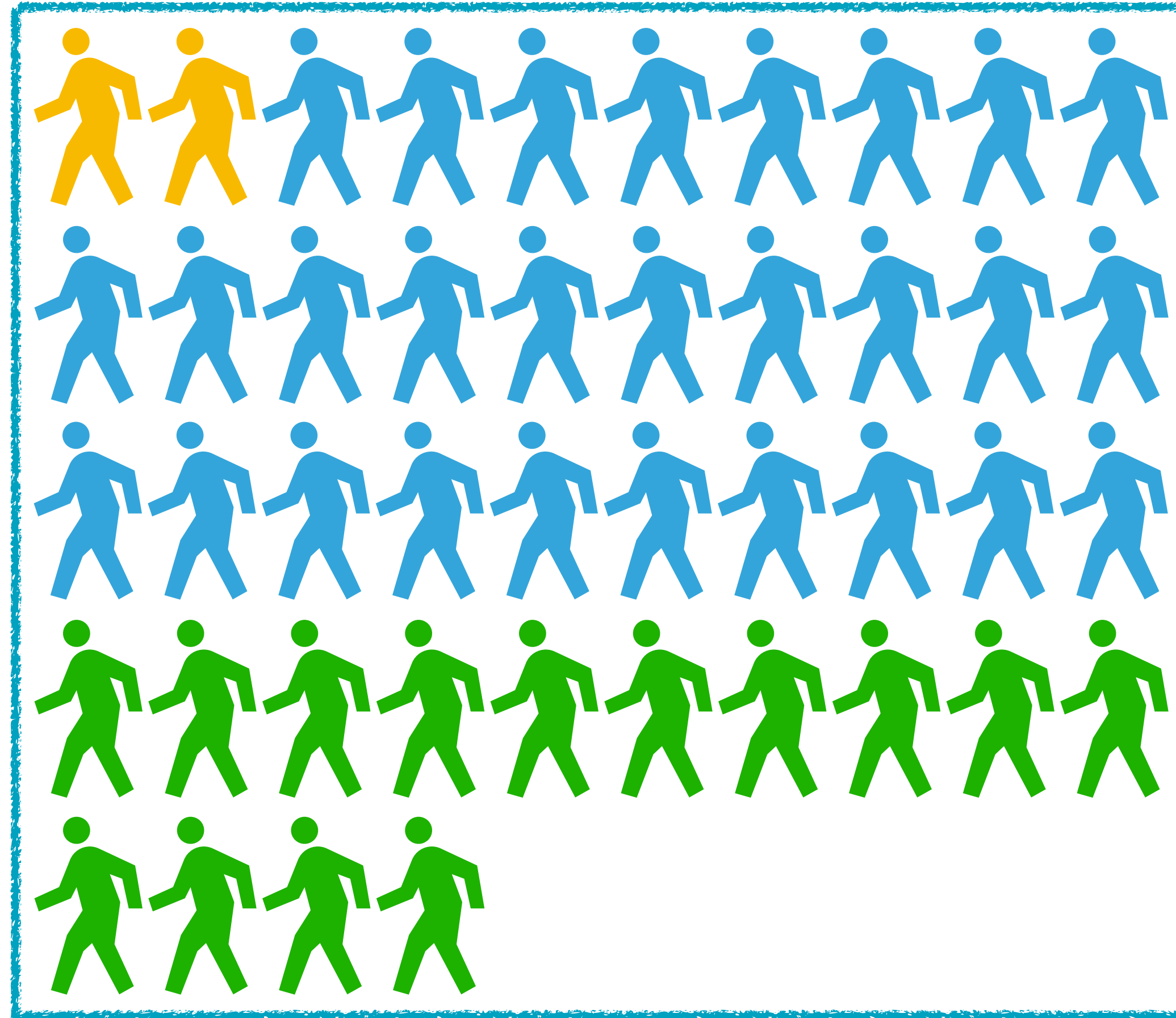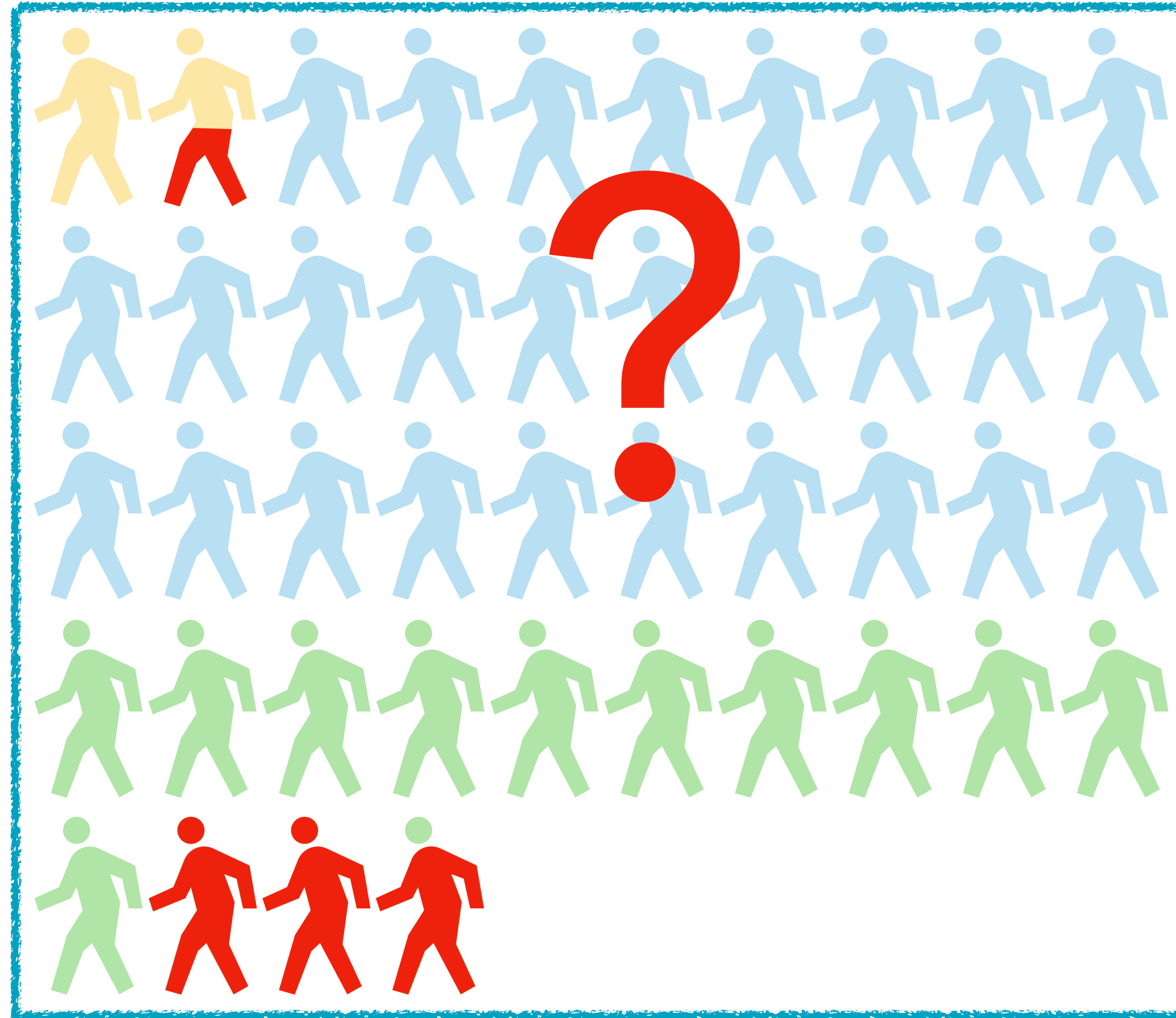
In absolute numbers

Error rate

= 10.000 clients

440.000 in total

Category „A"
High Chance

Category „B"
Medium Chance

Category „C"
Low Chance

Error rate

= 10.000 clients

440.000 in total

Category „A"
High Chance

Category „B"
Medium Chance

Category „C"
Low Chance

Falsely classified:

2.640 „C"

28.160 „A"

?????? „B"
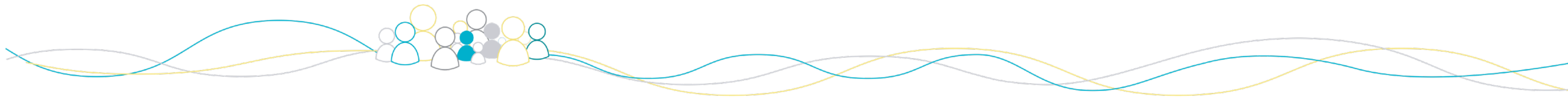
AMS Algorithm as prime example for 'Mathwashing'

„Models are opinions embedded in mathematics."

Cathy O'Neil,
„Weapons of Math Destruction"

▸ Seemingly „neutral" mathematics / statistics hide …

… human decisions & prejudice

… discrimination & bias

▸ Pretense of objectivity allows *delegation of responsibility* towards automated decision making systems

AMS & Accountability

▸ Responsibility/Liability in case of errors unclear

　▸ … once occured it's mostly too late!

▸ No transparency of operative rules and guidelines

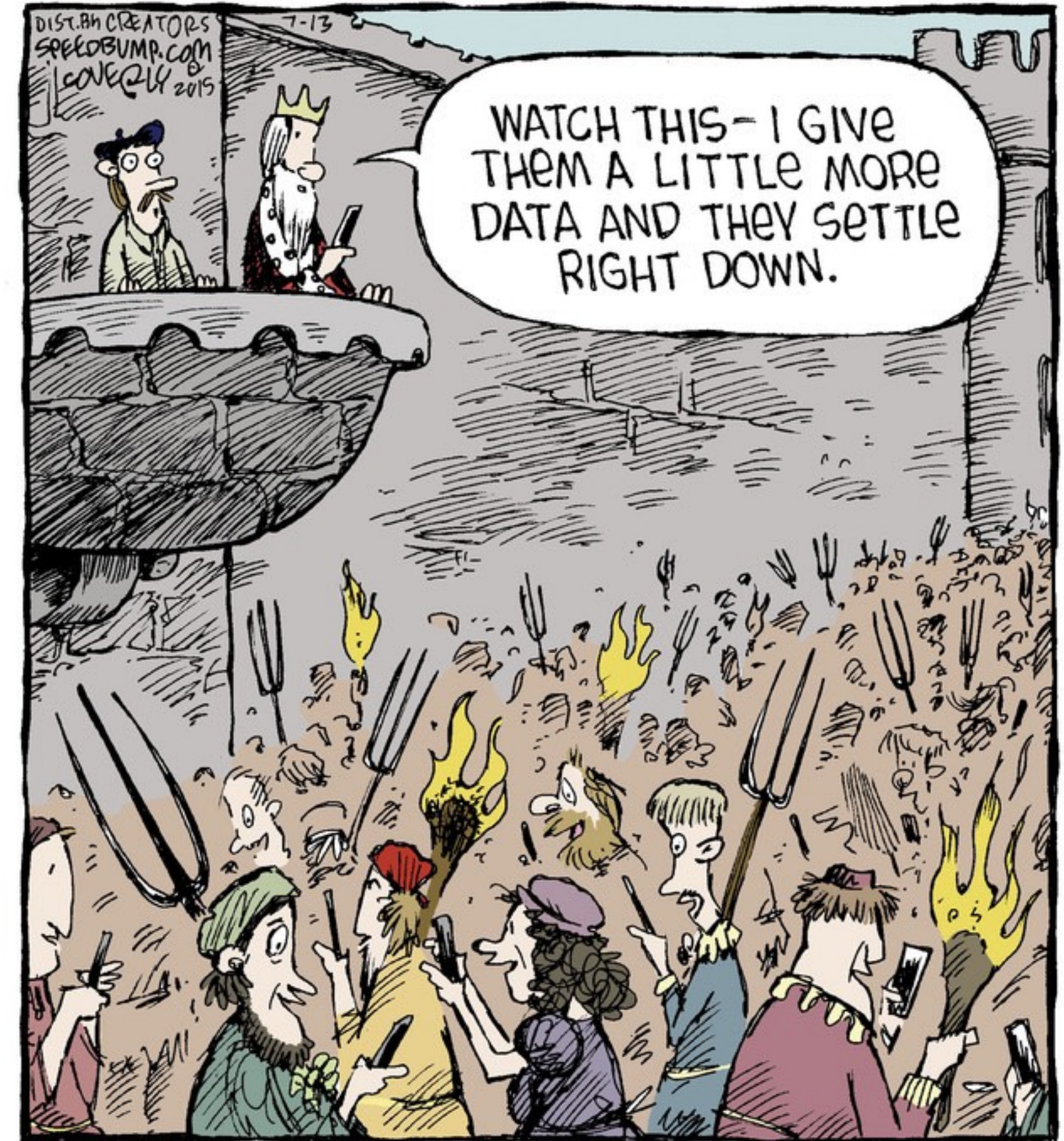▸ Promise of 'explanations' for classification

　￢ Most likely by showing largest impact factors

▸ Regulatory oversight of algorithmic systems and their use generally lacking

Source: http://www.roystoncartoons.com/2011/08/boardroom-cartoon-game-on.html

Centre for Informatics and Society
der Fakultät für Informatik an der TU Wien

**Discriminating
Labor Market**

**Error rate,
algorithmic illiteracy,
stigmatisation &
implicit biases**
affect discriminated
groups disproportionally

Feedback Loops

Algorithmic System
trains on and replicates
bias

Decisions
are based on biased results

For more, see (for instance): Barocas S and Selbst AD (2016) Big data's disparate impact. California Law Review 10: 671–732.

Centre for Informatics and Society
der Fakultät für Informatik an der TU Wien

Other Critical Issues

TRANSPARENCY

▸ Calculations and training data are opaque

▸ Results not traceable / explainable

    ⌐ … almost certainly not for AMS clients and workers!

▸ As of December 2018:
staff reported no training or guidance on the system

▸ Unclear procedural consequences for clients

ALGORITHMIC LITERACY

Claim:

▸ „The algorithm doesn't decide - AMS workers can ignore the suggestions!"

Reality:

▸ Lack of algorithmic literacy limits understanding for data biases & error rates

▸ Suggestions of algorithmic systems tend to be taken with little question - „Automation Bias"

▸ **Pressure to explain when disagreeing with the algorithm**
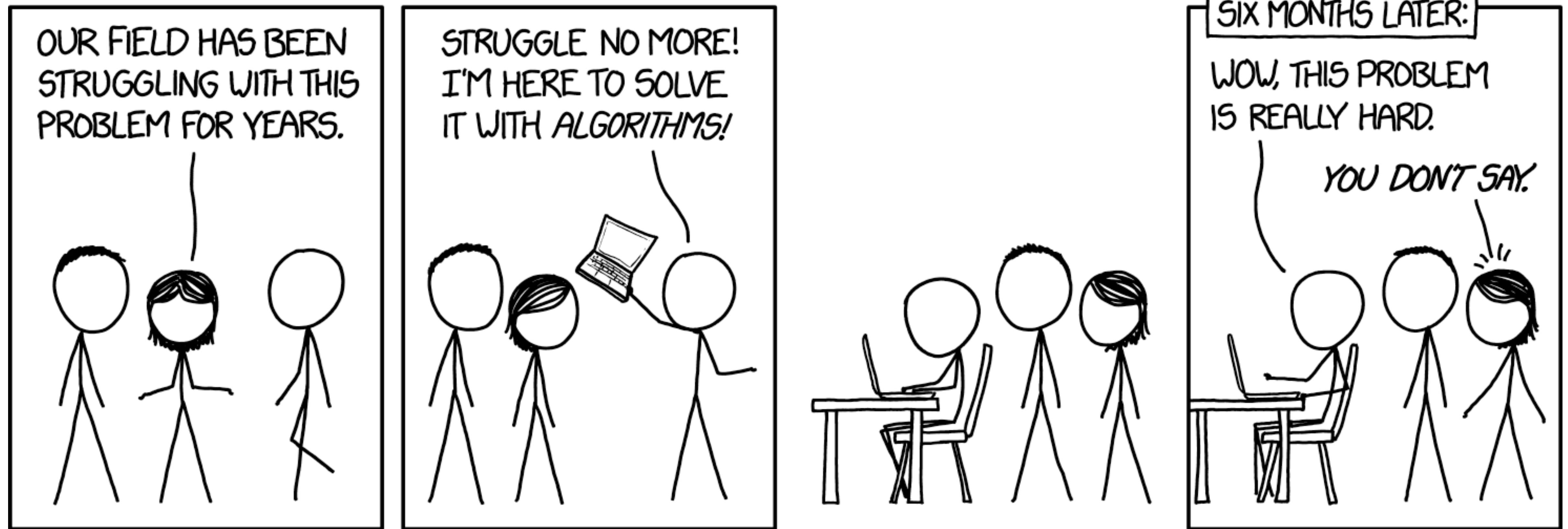
Statistical Profiling across OECD Countries[1]

| | Outcome (probability of) | Data derived from | Type of data sources (in addition to socioeconomic info) | | | | | Statistical model | Accuracy | Compulsory/voluntary use by | |
| | | | Job readiness | | | Motivation | Opportunities | | | | |
			Labour market history[a]	Hard skills	Soft skills	Jobseekers' behaviour	Regional labour market info			Jobseekers	Caseworkers
Australia	Long-term unemployed (12 months)	Personal interview; online trial ongoing	Yes	Yes	No	No	No	Logistic regression		Compulsory	Compulsory
Austria	LM integration probability[b]	Administrative data	Yes	Yes	No	No	Yes	Logistic regression	71%-87%	Compulsory	
Belgium (Flanders)	Long-term (>6 months) unemployed	Administrative data; "click" data	Yes	Yes	No	Yes	No	Random forest model	67% (AUC ~76%)	Compulsory	Compulsory
Denmark	Long-term (>26 weeks) unemployed	Online questionnaire; Administrative data	Yes	Yes	Yes	Yes	No	Big data model	>60%	Voluntary	Voluntary
Ireland	Probability of exit to employment within 12 months	Questionnaire as part of benefit claim process, administrative data	Yes	Yes	No	Yes	Yes	Probit regression	70% - 86%	Compulsory	Compulsory
Italy	Long-term unemployed (12 months)	Administrative data	Yes	Yes	No	No	Yes	Logistic regression		Compulsory	Compulsory
Latvia	Long term unemployed (12 month)	Personal (individual) interview, questionnaire, administrative data	Yes	Yes	Yes	Yes	Yes	Factor analysis	No data yet	Compulsory at PES & voluntary online	Compulsory (advisory for now)
Netherlands	Long-term unemployed (12 months)	Online questionnaire		Yes	Yes	Yes		Logistic regression	70%	Voluntary	Compulsory
New Zealand	Lifetime income support costs (LET), change in lifetime income support and staff costs from receiving a case management service (SEM)	SEM/LET are based on administrative data	Yes	Yes	No	No	No	Random forest (LET), Gradient boosting (SEM)	AUC: 0.63 - 0.83%	Compulsory for jobseekers; opt-in for other PES clients	Compulsory
Sweden	LTU (6 months)	Administrative data	Yes	Yes	No	No	Yes	Logistic regression			Voluntary
US	Exhausting the 26-week entitlement to UI benefits	Online questionnaire; Administrative data		Yes	No	No	Yes	Logistic regression		Compulsory	Compulsory

Source:
Desiere, S., Langenbucher, K., Struyven, L.: Statistical profiling in public employment services. An international comparison. (2018).

Centre for Informatics and Society
der Fakultät für Informatik an der TU Wien

Moral of the story

Hard Problems + Algorithms != Easy Problems

Centre for Informatics and Society
der Fakultät für Informatik an der TU Wien