

Ethical Aspects of Algorithms and Scoring in Pedagogical and Social Work Contexts

Nadia Kutscher

Examples of algorithm-based decision-making and scoring in social work and pedagogical contexts

Algorithms and Scoring play an increasing role in social work as well as in pedagogical contexts. In China, facial recognition systems, partially connected with WeChat accounts which are linked to the social scoring system, check who is attending courses in university (in 2018, in a Swedish high school used as well facial recognition technology to check the attendance of students). In 500 schools in China, researchers from Jiao Tong University Shanghai use cameras and software to measure whether children are bored or overstrained, supported by the German Research Center for Artificial Intelligence (DFKI). The aim of recognizing boredom or overstrain is to find out what learning skills a child has, how it solves its problems and what causes it difficulties. As soon as the facial recognition registers that a child looks bored, it is assigned new tasks on its computer and if the child is overwhelmed, the system offers him or her additional help.

Matthias Burchardt speaks of „Digital Panopticism“ (Burchardt 2018, 109) when analyzing developments in the context of learner’s scoring described in a paper of the Bertelsmann Stiftung on the use of the software Knewton: „Knewton scans everyone who uses the tutorial. The software meticulously observes and stores what, how and at what speed a student learns. Every reaction of the user, every mouse click and every keystroke, every right and wrong answer, every page call and every abort is recorded. Every day we collect thousands of data points from each student‘ says Ferreira proudly. This data is analysed and used to optimise personal learning paths. Complex algorithms put together individual learning packages for each individual student, whose content and tempo continually adapt, if necessary at minute intervals. Knewton already reliably calculates the probability of correct and incorrect answers as well as the grade that a student will achieve at the end of a course. One day there will probably be no need for exams; the computer already knows what the result will be.“ (Dräger/Müller-Eiselt 2015, S. 24f).

Asylum management as „techno-humanitarianism“ (Garelli/Tazzioli 2018) in Europe uses biometric data, name, age and data points related to a person’s vulnerability, relationship status and geographic location to score the entitlement for housing of asylum-seekers or the availability of cash assistance for them according to geographical restrictions (Metcalfe/Dencik 2019). In Germany and Austria, mobile phones of refugees can be searched for metadata to find out about migrants’ identities for deciding about entitlements to asylum and residents permits. EURODAC fingerprint checks serve differentiation procedures of migrants into categories such as 1 - „person as an applicant for international protection“, 2 - „person as having crossed, or attempted to cross, a border illegally“, and 3 - „being a potential illegal immigrant“ (Metcalfe/Dencik 2019).

Access to social services is being facilitated by scoring systems such as – depicted in Virginia Eubank’s book „Automating Inequality“, a first documentation of the uses and consequences of algorithms in social services. Examples focus on the „Service Prioritization Assistance Tool“ used for the „Homeless Management Information system“ (HMIS) in Los Angeles to rate entitlements of homeless people to housing programs or the automatization or the privatization of the eligibility processes for the welfare system in the state of Indiana checking the entitlement to Medicaid benefits in the US and cutting it when mistakes are made. In this context, real mistakes or errors were assessed by the algorithmic

system as fraud and it became the people's responsibility to prove that the system was wrong. Only resourceful people could manage to go against the system. Joanna Redden reports a similar phenomenon from Little Rock, Arkansas where „an algorithm introduced by the state's Department of Human Services was blamed for unjustly cutting the home care hours of people with severe disabilities“ when „some „weekly home care hours [were] cut by more than 30 percent“ whereas the „automated system ensures that assignments of home care hours are fair and objective“ (Redden 2018).

Eubank's third example focuses the child protection context but in Germany, there are similar developments: In some cities in Germany, the public child and youth welfare administration (Jugendamt) has installed software to support the decision-making by professionals in the context of child protection cases. After some children died in the year 2006 and following years, nationwide efforts were undertaken to set up monitoring systems that should prevent this from happening again. Those systems aim at risk assessment and controlling on the one hand side and flexible support and close monitoring on the other hand side. Risk assessment and control focus on preventive medical check-ups with paediatricians, screenings in maternity clinics, software for the calculation of risk threshold values in youth welfare offices and we could watch an increase of evidence-based and standardized programs. Flexible support and close monitoring comprise welcome visits and welcome packages for families getting a baby, social pedagogical family support, parents cafés, setting up family centers and enforcing parent and family work in early childhood education centers.

In some welfare administrations on the communal level¹, if someone calls to notice the authorities that there could be child in danger of mistreatment e.g. by parents, the social workers have to fill in the software with information such as who made the announcement, what is going on, checking on certain categories defined as relevant for child wellbeing such as criteria to measure a possible withdrawal of necessities of life: body hygiene, clothing, housing, protection against danger, economic livelihood. The software then calculates a threshold value of endangerment of the child and thus 'supports' the decision-making of the professionals e.g. if a child should be taken away from the family, if other support systems or interventions should be taken or if the family could be left alone without intervening as a public authority.

In social work discipline and practice there is a debate about whether this should be regarded as helpful or not. Daniela Schneider and Udo Seelmeyer point to the fact that „support by software can lead both to the empowerment of professionals and to de-professionalisation, for example by limiting the scope for discretion (Ley/Seelmeyer 2014). Categorizations and typifications that are recorded in the software, e.g. via selection fields or standardized diagnostic manuals, can direct the view to aspects that would otherwise be overlooked, but can also narrow the view and thus complicate a holistic view of the case or promote labeling attributions. In particular, if categorisations and standardisations at the level of the addressees are combined with standardisation at the level of the assistance services, i.e. if decisions on the selection and design of assistance are automatically derived from certain 'diagnoses', this leads to particularly serious restrictions on the scope for discretion“ (Schneider/Seelmeyer 2018). The debate goes on about where a manualization in social work leads to which does not anymore need professionals who based on professional knowledge and reflection are able to assess a case and instead evidence-based criteria based on statistical knowledge are being taken as the basis for standardized and thus seemingly ‚objective‘ decisions without further

¹ e.g. JusIT in Hamburg - Based on IBM Watson Health Solution for Child Welfare

reflection of the single case. As Marl Schrödter et al. point out, "the items are selected exclusively for the limited purpose of risk assessment and empirically tested for their prognostic power" (Schrödter et al. 2018). But the accurate risk prediction is often erroneously equated with a statement about an intervention decision (Schrödter et al. 2018). A statistical entity is then equated to a single case. By that, correlations are used wrongly as causal relations, as Cukier and Meyer-Schönberger (2013) put it.

Further, the statistically based criteria reproduce data biases from databases available for the scoring procedure. In Germany as well as in the US American examples, parents with psychosocial problems or living from public benefits are – based on statistical data - regarded as risky and by that, this information matters in the risk assessment procedures trying to find out the so-called „high risk families“ (Hensen 2010). Virginia Eubanks speaks in the context of the Allegheny Family Screening Tool (AFST) of „poverty profiling. Like racial profiling, poverty profiling targets individuals for extra scrutiny based not on their behavior but rather on a personal characteristic: living in poverty. Because the model confuses parenting while poor with poor parenting, the AFST views parents who reach out to public programs as risks to their children.“ (Eubanks 2018). In Eubanks' example, the AFST „is run on every member of a household, not only on the parent or child reported to the hotline. Under the new regime of prediction, you are impacted not only by your own actions, but by the actions of your lovers, housemates, relatives, and neighbors. Prediction, unlike classification, is intergenerational.“ (Eubanks 2018).

There is a „discursive shift from probabilistic actuarial methods for the assessment of knowable risks in society, towards more loosely conjured assumptions about possibilistic threats posed by chronically uncertain, unknowable, risks.“ (Pithouse et al. 2011, p. 162). Here, predictive algorithmic systems play an important role. In the context of an investment state, the question comes up how to define outcomes and impacts of interventions as well as a rational relation between costs and outcomes of social work (Zetino/Mendoza 2019, 411). The debate about the advantages of these systems is still going on, showing that classification errors within the instruments as well as different perspectives on core problems from the professionals' or the service users' side „make an objectification of the assessments by standardized diagnostic instruments at least questionable“ (Ley 2019).

Ethical Aspects of Algorithms and Scoring in Social Work Contexts

'Objectivity' of Technology - Disguise of and Implicit Normativity and Normalism

By following a logic of normalism (Link 2006), decision-making turns to technical solutions instead of normative decisions – seemingly. We can speak of at least an implicit normativity as certain ways of life are being valued within the system as risky - and mainly it is the life of underprivileged persons depicted as statistically risky. In case diagnosis, an ethical aspect seems which normative assumptions are inscribed in the respective diagnostic instruments. This can be reconstructed by looking at what is considered "normal", "appropriate", "problematic" etc. in the respective instrument (e.g. with regard to educational or care behaviour or developmental steps). The same question also arises with analogous forms of diagnostics, but with a software-based form of diagnostics, subjective responsibility for technology is potentially pushed into the background. While conducting case assessment within software, the initially subjective assessment transforms into a technically-informatized calculation that 'faces' the professional then as an 'objective' software-based decision. The underlying subjective evaluation, which led to clicking on certain risk values (and which can potentially differ from professional to professional), becomes an 'objectively verified recommendation'

and thus in a double sense a technique of normalisation. The judgement of the software - for example also in the case of a risk threshold calculation in child protection - implies objectivity and unambiguity. However, the fact that the categories are connected with certain data values as well as the imposing of certain categories aiming at individuals' behaviour as relevant for assessing "safe" or "endangering" practices are both bound to (more or less implicit) normative decisions (Gillingham 2019). Even though recommendations of the software may seem more reliable than volatile assessments of professionals (Bastian/Schrödter 2015), the question remains as to how much statistical-actuarial approaches to case assessment can do justice to the complexity and uniqueness of each individual case.

In addition, there is an as yet unresolved question as to what the mere existence of a software-based decision or recommendation means for the implicit narrowing of professional decision-making and discretion. Practical experiments show that, in view of the responsibility for risk, experts hardly dare to act against the software decision, since in case of doubt it is granted an objectifying status that "occurs" as an external authority when a conflict occurs or when damage has occurred and the question of responsibility is asked. This algorithm-based classification has ambivalent implications. In criminology, the principle of these evidence-based control mechanisms is actuarial justice (Balzer 2015). This "objectifies dangers to risks and thus does not operate by means of morality. Thus, the morally deviating 'evil criminal' of the old penology becomes a bearer of risk characteristics through objective facts" (Balzer 2015, 80). Non-moral evaluations that do not depend on subjective assessments promise a higher degree of objectivity. At the same time, they are based on the simultaneous measurement and establishment of "normality" and deviation.

Subjectivation and Moral Delegitimation of 'Underclass'

The actuarial logic of risk assessment systems raises the question what cases but especially what subjects are being constructed by using scoring criteria differentiating people into reliable or non-reliable resp. deserving and undeserving ones. Moreover, it can be questioned which classifications and underlying hypotheses form these judgements and on which assumptions, measurabilities they are based on and which intended or unintended consequences are resulting from this. The seeming rise of safety or reliability of decision-making and minimizing risks suggests objectivity, blinding out on underlying bias and consequent stigmatization as well as the subjective perception which the single database entries e.g. of professionals are based on. The orientation in diagnostic instruments (diagnostic sheets, classifications, checklists, inventories, manuals, etc.) focusing on measurable and statistically based criteria which follow certain theoretical and also normative assumptions or empirical biases can thus lead to a general suspicion towards underprivileged persons (as also Nicholas Kayser-Bril points out in his essay, targeting vulnerable populations). By that, inequality is virtually being de-thematized. Instead, a classification of the poor, exclusion from services, discrimination and reproduction of inequality are being promoted in the guise of objective measurements and classifications.

What Works Logic, Recognition of the Individual Situation and Stigmatization

Moreover, also justice issues are being de-thematized and replaced by efficiency and efficacy logics, following a widespread logic of „what works“ which takes statistical evidence as single case solving evidence. Bastian et al (2020) warn that a recent longitudinal study in England convincingly showed that, „despite increasingly precise forecasting procedures, a large proportion of abused and neglected

children receive no social assistance at all, while a large proportion of children who are not affected by abuse and neglect have had to undergo unnecessary testing processes (Devine 2017, p. 7).“ The reason for that was that the categories imposed focused the statistically risky people so that „prognoses can more effectively organize the exclusion of those who are considered dangerous, criminal, needy or otherwise deviant, although these populations have been constructed as such by surveillance technologies in the first place" (Bastian et al. 2020). This means that there is not only the risk of stigmatization but also of neglecting risks based on statistical knowledge.

The Dilemma of Low-Threshold and Reaching Target Groups of Services vs. Client Data Protection

Empirical findings on "algorithmic bias" show that the discrimination of certain social groups in the population is reproduced again within the framework of algorithmic calculations, since either the algorithms are designed in this way or the data they process carry this bias within them (Angwin et al. 2016). Against this background, the risk of structural and potentially non-transparent reproduction of social inequalities - also in access to social services - is a problem to which Virginia Eubanks points (Eubanks 2018). So as soon as users of social services visit, for example, the Facebook page of a social work institution or contact a specialist via WhatsApp, metadata is created that is combined and evaluated by the providers with personal data and identifies them as potential clients of psychosocial services. This means that the possibility of being accessible as a help facility via digital, widespread channels also means that precarious metadata about the service users will inevitably be produced, which can reduce the current and future freedoms and accesses of the service users to information and resources. Suggestions to deal with this point to the explicit dilemma between reaching out and being accessible and divulging clients' data to third parties (Dolinsky/Helbig 2015).

Autonomy or Subjectivation

In the context of "Liquid Surveillance" inscribed in digital media (Bauman/Lyon 2013), these include the possibility that a subjectively perceived gain in autonomy does not exclude or even conceal subjection to a future - or even current - powerful lack of freedom on the basis of digital data. In the sense of a "governmediality" (Traue 2009), the users of digital media experience themselves as autonomous actors, but at the same time submit to the structures and forms of representation of the media structures. Thus the use of digital media brings with it challenges within the framework of antinomic constellations of (not only governmental) power, discipline, standardization and technologies of the self and raises the question of the relationship between autonomy and the possibility of participation as the object of a digital-reflective ethical-moral debate also in the context of social work (Kutscher 2020).

Dencik et al. (2019, 17) point to the fact that the logic of attaching risk factors to individual characteristics and behaviour might divert focus away from structural causes, such as issues of inequality, poverty or racism. This goes along well with the individualisation of responsibility as a core element of an activating welfare state. Which is installed in Germany since 2006 and started to develop since the end of the 1990ies. The activating and investing welfare state sets up structures of normalism, implies a rhetoric of ‚freedom of agency‘ (but in fact only under the condition of individual responsabilization and the postulate of caring for the community) and a strong moral and punitive perspective towards the ‚underclass‘. By that, scoring promotes the establishment of governmentality in a Foucauldian sense with perfection, in fact: technologies of the self. This governance of responsabilization policies aiming at individualization and economization (Brown 2015, 131) imposes

also a reduction of autonomy in two ways: 1) Reducing the discretion of professionals, by objectivating decision-making and setting criteria based on statistical probabilities, partly connected with an 'automation bias' (Cummins, 2004) but also with contradictory and divergent practices (Ley 2019) and 2) by subjectivating individuals as datafied subjects (Allert et al, 2018, 153f.) based on intransparent decisions and leaving to the subjects to deal with this framework by strategies such as affirmation, adaptation, playing around and subversion (Allert et al 2018, 153) in the context of massive asymmetries in knowledge and the resulting power (Zuboff 2019) where the demand for 'informed consent' (Reamer 2013) seems preposterous.

References

- Allert, H./Asmussen, M./Richter, C. (2018): Formen von Subjektivierung und Unbestimmtheit im Umgang mit datengetriebenen Lerntechnologien – eine praxistheoretische Position. In: Zeitschrift für Erziehungswissenschaft (2018) 21, pp.142–158.
- Angwin, J./Larson, J./Mattu, S./Kirchner, L. (2016): Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks. URL: www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
- Balzer, A. (2015): Im Netz der Kontrolle. Gilles Deleuze' Kontrollgesellschaft im Blick der Governmentality Studies. Bamberger Beiträge zur Soziologie. Band 15. URL: https://opus4.kobv.de/opus4-bamberg/files/44371/BBzS15Balzeropusfinse_A3a.pdf
- Bastian, P./Schrödter, M. (2015): Risikotechnologien in der professionellen Urteilsbildung der Sozialen Arbeit. In: Kutscher, N./Ley, T./ Seelmeyer, U. (Hrsg.): Mediatisierung (in) der Sozialen Arbeit, Baltmannsweiler: Schneider Hohengehren, pp. 192–207.
- Bauman, Z./Lyon, D. (2013): Liquid surveillance: A conversation. Malden: Polity Press.
- Brown, W. (2015): Undoing the Demos: Neoliberalism's Stealth Revolution. New York: Zone Books.
- Burchardt, M. (2018): Big brother is teaching you – Schule total digital? In: Vierteljahrsschrift für wissenschaftliche Pädagogik 94 (2018), S. 102-112.
- Cukier, K./Mayer-Schönberger, V. (2013). Big data: A revolution that will transform how we live, work and think. New York, NY: John Murray.
- Dencik, L./Hintz, A./Redden, J./Warne, H. (2018): Data Scores as Governance: Investigating Uses of Citizen Scoring in Public Services. Project Report. Data Justice Lab, URL: <https://datajustice.files.wordpress.com/2018/12/data-scores-as-governance-project-report2.pdf>
- Dencik, L./Hintz, D./Redden, J./Treré, E. (2019): Exploring Data Justice: Conceptions, Applications and Directions, Information, Communication & Society, 22:7, pp. 873-881
Devine 2017
- Dolinsky, H. R./Helbig, N. (2015): Risky Business: Applying Ethical Standards to Social Media Use with Vulnerable Populations. In: Advances in Social Work, Vol 16 No 1 (2015): Special Issue: Technology, the Internet & Social Work Practice, pp. 55-66. URL: <http://journals.iupui.edu/index.php/advancesinsocialwork/article/view/18133/19920>
- Dräger, J./Müller-Eiselt, R. (2015): Die digitale Bildungsrevolution. München: dva.

- Eubanks, V. (2018): *Automating Inequality*. New York: Macmillan.
- Gillingham, P. (2019): *Decision Support Systems, Social Justice and Algorithmic Accountability in Social Work. A New Challenge*. Practice 3, pp. 1–14.
- Hensen, G. (2010): Risikofamilien. Wie Probleme fachlichen Handelns einzelnen Familien als Eigenschaft zugeschrieben werden. *Sozial Extra*, (3,4), pp. 16–19.
- Kutscher, N. (2020): Ethische Fragen Sozialer Arbeit im Kontext von Digitalisierung. In: Kutscher, N./Ley, T./Seelmeyer, U./Siller, F./Tillmann, A./Zorn, I. (Hrsg.): *Handbuch Digitalisierung und Soziale Arbeit*. Weinheim: Beltz Juventa.
- Ley, T. (2019): *Zur Informatisierung Sozialer Arbeit - Eine qualitative Analyse sozialpädagogischen Handelns im Jugendamt unter Einfluss von Dokumentationssystemen*. Bielefeld: Universität Bielefeld (Dissertation). Weinheim: Beltz Juventa.
- Ley, T./Seelmeyer, U. (2014): Dokumentation zwischen Legitimation, Steuerung und professioneller Selbstvergewisserung. Zu den Auswirkungen digitaler Fach-Anwendungen. In: *Sozial extra. Zeitschrift für soziale Arbeit & Sozialpolitik* 38(4), pp. 51-55.
- Link, J. (2006): *Versuch über den Normalismus. Wie Normalität produziert wird*. Göttingen: Vandenhoeck & Ruprecht.
- Metcalf, P./Dencik, L. (2019): The politics of big borders: Data (in)justice and the governance of refugees. **First Monday**, apr. 2019. Available at: <https://firstmonday.org/ojs/index.php/fm/article/view/9934/7749>
- Pithouse, A./Broadhurst, K./Hall, C./Peckover, S./Wastell, D./White, S. (2011): Trust, risk and the (mis)management of contingency and discretion through new information technologies in children's services. In: *Journal of Social Work* 12 (2), pp. 158-178.
- Reamer, F. G. 2013. Social Work in a Digital Age: Ethical and Risk Management Challenges. In: *Social Work*, 58 (2), pp. 163–172.
- Redden, J. (2018): The Harm That Data Do. Paying attention to how algorithmic systems impact marginalized people worldwide is key to a just and equitable future. In: *The Scientific American*. URL: <https://www.scientificamerican.com/article/the-harm-that-data-do/>
- Schneider, D./Seelmeyer, U. (2018): Der Einfluss der Algorithmen. Neue Qualitäten durch Big Data Analytics und Künstliche Intelligenz. In: *Sozial Extra* 42, H. 3, pp. 21–24
- Schrödter, M./Bastian, P./Taylor, B. (2020): Risikodiagnostik und Big Data Analytics in der Sozialen Arbeit. In: Kutscher, N./Ley, T./Seelmeyer, U./Siller, F./Tillmann, A./Zorn, I. (Hrsg.): *Handbuch Digitalisierung und Soziale Arbeit*. Weinheim: Beltz Juventa.
- Schrödter, M./Bastian, P./Taylor, B. (2018): Risikodiagnostik in der Sozialen Arbeit an der Schwelle zum »digitalen Zeitalter« von Big Data Analytics. URL: https://www.researchgate.net/publication/323267949_Risikodiagnostik_in_der_Sozialen_Arbeit_an_der_Schwelle_zum_digitalen_Zeitalter_von_Big_Data_Analytics
- Zetino, J./Mendoza, N. (2019): Big Data and Its Utility in Social Work: Learning from the Big Data Revolution in Business and Healthcare, *Social Work in Public Health*, 34:5, pp. 409-417
- Zuboff, S. (2018): *Das Zeitalter des Überwachungskapitalismus*. Frankfurt, New York: campus.