When Postal Codes Matter More Than Studying: Algorithms as Arbiters of Fair Assessment

Kristjan Kikerpill, Andra Siibak* and Maris Männiste

Abstract

During the Covid-19 pandemic, educational systems around the world became heavily reliant on data-intensive technologies and algorithmic decision-making systems. In the UK, concerns over grade inflation for A-levels qualifications led the overseeing authority to using a grading algorithm in the hopes of providing objective, neutral and fair assessment. In reality, the algorithm displayed bias by downgrading nearly 36% of the grades of students coming from more marginalised groups, resulting in an 'A-levels fiasco'.

We applied a critical discursive psychology approach to explore how students, teachers, parents and policymakers impacted by algorithm-based grading expressed and relied on different senses of fairness (formal, implied contractual, relational, retributive) when commenting on the incident in international news media (N=329).

Our findings indicate that while the affected agents evoked and clung to different senses of fairness, the algorithm itself also emerged as a meta-agent of fairness in the discursive constructs of social reality. The analysis further illustrates how public authorities held overly optimistic and techno-solutionist imaginaries about the neutrality and objectivity of AI-based systems, revealing a crucial need for oversight bodies that guide public institutions in responsibly deploying safe and trustworthy data technologies.

Keywords: algorithms, education, assessment, algorithmic grade, critical discursive psychology.

Introduction

It is known that parental resources influence the higher education attainment of the child (Lauri & Saar, 2022), and cumulative advantages and disadvantages derived from those resources tend to transmit intergenerationally. Research has shown that initial inequalities in social characteristics produce increasingly greater inequalities in the same or other characteristics (e.g. Lauri & Saar, 2022). In the context of education, mitigating inequalities and facilitating higher education attainment comes in the form of egalitarian educational policies; that is, initial social inequalities that might hinder education attainment must and can be positively influenced by proper educational policies (Lauri & Saar, 2022). However, as education systems increasingly adopt data-based technologies, it becomes essential to consider how such technologies interact with, reinforce, or potentially challenge these longstanding social inequalities.

Datafied education refers to the expanding use of automated data processing to assess student progress (see Jarke & Breiter, 2019), often at the expense of aspects not easily captured by data (Knox et al., 2020). It follows the logic of *governance by numbers* (Neumann, 2019) and is increasingly reliant on algorithmic outputs to provide policy-relevant inputs. By their very essence and function, algorithms reinforce and perpetuate any inequalities that may be gleaned from their input data (see Zajko, 2022). Rushed and poorly tested algorithm-based solutions deployed at large often yield undesirable social results (see Burgess et al., 2023), raising grave concerns over fairness for those subjected to algorithmic decision-making (Starke et al., 2022). Yet, data-hungry solutions reliant on algorithms and artificial intelligence are heralded as the way towards a more personalised learning experience in education (see also Li & Wong, 2023).

* E-mail of corresponding author: andras@ut.ee

In many ways, the Covid-19 pandemic was a turning point for the rapid uptake of data-based technologies and algorithmic decision-making in educational settings (e.g., see Selwyn et al., 2023; Männiste & Siibak, 2025). Our article focuses on the events that took place in the UK, where concerns over grade inflation for A-levels qualifications led the overseeing authority to use a grading algorithm in the hopes of providing objective, neutral and fair assessment. In reality, the algorithm displayed bias by downgrading nearly 36% of the grades of students from more marginalised groups (Kelly, 2021), resulting in an 'A-level grading fiasco' (Kolkman, 2020). The controversy brought renewed attention to how algorithmic bias can undermine visions of equitable, just and accountable data-driven education.

In our article, we applied critical discursive psychology to explore how students, teachers, parents and policymakers impacted by algorithm-based grading expressed and relied on different senses of fairness – formal, implied contractual, relational, retributive (Nisbet & Shaw, 2020) – when commenting on the incident in the international news media (N=329). Our findings indicate that while the affected agents evoked and clung to different senses of fairness, the algorithm itself also emerged as a meta-agent of fairness in the discursive constructs of social reality. The analysis further illustrates how public authorities held techno-solutionist (Morozov, 2013; Olojede, 2024; Sadowski, 2025) imaginaries; in other words, they believed that technology would provide a solution to every problem. Furthermore, they held woefully naïve views about the neutrality and objectivity of AI-based systems, revealing a crucial need for oversight bodies that guide public institutions in responsibly deploying safe and trustworthy data technologies.

Literature review

Technological solutionism in education

Jathan Sadowski (2025) argues that 'our cultural lexicon for talking about social change is dominated by technological disruption' (p.197). We can see this being true in different life spheres where the implementation of algorithmic tools are believed to disrupt how things are done and change the old ways. In education, techno-solutionism manifests through the increasing digitalisation of classrooms, the proliferation of virtual and e-learning tools, and persistent efforts to implement new technologies, all grounded in the belief that appropriate technological use can effectively remedy the various challenges that afflict educational systems (Thibaud, 2025). Technological determinism (Sadwoski, 2025, p.197) is oftentimes driven by the political thought that if we do not build and use this technology, then someone else will: this mindset assigns agency to technology rather than society itself. One of the ethical concerns associated with the increasing deployment of AI in education is the tendency to oversimplify complex educational and social problems by reducing them to a single algorithmic output (Zytek et al., 2022). When decision-making is mediated by AI systems, nuanced pedagogical, cultural, and social contexts are frequently translated into abstract data categories, which can erase lived experience in favour of computational efficiency. This reductionist logic reinforces the assumption that more data and better models will necessarily produce better outcomes, thereby strengthening the techno-solutionist mentality described by Sadowski (2025). Rather than engaging with structural problems of inequality, access, or educational justice, data-driven systems risk narrowing the scope of inquiry to what can be measured, predicted and optimised.

According to Simon Lindgren (2024), this reductionist framing is particularly evident in how power relations are conceptualised within AI systems. Lindgren (2024) argues that when socially and historically situated structures of domination are encoded into algorithmic systems, they are frequently reinterpreted as mere *bias* within the model. As a result, political and structural inequality is reduced to a technical flaw that can supposedly be corrected through better design, improved data, or refined training processes. This approach, Lindgren (2024) contends, reflects an AI-first perspective in which the necessity of AI is taken for granted, and criticism is redirected toward optimisation rather than questioning whether AI should be used at all. Lindgren (2024) further explains that forms of structural oppression are not minor distortions within otherwise

neutral systems but are instead deeply embedded in historical conditions, language, ideology, and institutional structures, and that reducing them to bias depoliticises injustice by making systemic violence appear as a technical problem rather than a social and moral crisis. While data-driven approaches in education are commonly justified through appeals to access, fairness and innovation, such narratives risk obscuring the political consequences of technological adoption, including how inequality, institutional interests, and historical power relations become embedded in seemingly neutral systems.

It is precisely because injustice is reframed as a technical problem that the concept of algorithmic fairness becomes central to understanding how inequality is managed, limited, and sometimes obscured within AI governance.

Algorithmic fairness

In a thorough review of the topic, Mitchell et al. (2021, p. 142) define algorithmic fairness as a field of study pertaining to '[u]ncovering and rectifying … biases in statistical and machine learning models'. Fairness in technology, in general, encapsulates the concerns about whose values are embedded, who is excluded, and whether this exclusion is legally, morally, and socially justified (Porayska-Pomsta et al., 2023). Reasons for engaging in the aforementioned activities are legion as prediction-based automated decision-making has rapidly infiltrated numerous domains of public life; for example, immigration (Koulish & Evans, 2021), public health (Mhasawade et al., 2021), and welfare eligibility (Eubanks, 2018; Sleep & Redden, 2024). However, this also means that 'in the automated order of algorithmic modernity, human agency is increasingly outsourced to intelligent machines' (Elliot, 2024, p. 21), and we have started to witness the 'dramatic transmutation of agency' (Elliot, 2024, p. 76).

In popular understanding, the aim of algorithmic fairness is to rectify situations where a certain 'model's performance (however defined) unjustifiably differs along social axes such as race, gender, and class' (Mitchell et al., 2021, p. 142). Within the relatively loose confines of predictive decision-making as such, the field of education, and testing more specifically, has been a focal point of interest for decades (see Zwick & Dorans, 2016). Although not for a lack of critical inquiries into educational policy (see Gulson et al., 2022), the area of education governance is becoming increasingly dominated by technological means (Williamson, 2016), which is driven, in part, by the vast amounts of (multimodal) data produced in the processes of providing education, managing its provision and assessing the outcomes (Gulson et al., 2022).

With the introduction of data technologies at all levels of education (see Kikerpill & Siibak, 2023a), the agential borders between humans and machines become muddled within complex interplays that ultimately generate policy outcomes. According to Gulson et al. (2022, p. 37), emerging systems of thought in educational governance embed a multitude of subject-becoming-object and object-becoming-subject processes into the wider plan of educational policy considerations. Put differently, while a seemingly endless stream of data about agents within the education system turn human actions into data objects, the algorithms and technologies that rely on said data obtain a form of meta-agency. The latter, of course, does not refer to any subjectivity or autonomy in a conventional sense, but rather foregrounds the 'black-box' characteristic (see e.g., Williamson, 2016) of such technologies. From there, meta-agency might be gleaned and discursively constructed via a simple lack of comprehension of the technology's inner workings (Watson, 2019), which make it seem as if it is undertaking independent *thinkings* and *doings*. In this context, Masso and Kasapoglu (2020, p. 1215) propose a framework of 'triple agency' that comprises the agency of experts, data subjects, and algorithms. The framework aims to help unpack the complex interactions and blurred lines of accountability in data technologies. What may seem like algorithmic agency is, in fact, often an extension of state power, where technology embodies and enacts policy (Masso & Kasapoglu, 2020).

According to Broussard (2019, p. 129), 'a computer "knows" what it has been told'. Thus, when problems emerge with automation or algorithms, these cannot be considered as mere glitches in the system, which are unexpected but inconsequential; instead, such glitches are a result of human decisions that are shaped by social, cultural and political contexts (Broussard, 2023). This notion

aligns with data justice scholarship (Dencik et al., 2019; Dencik, 2025), which seeks to move the conversation beyond the narrow concerns of efficiency and privacy to ask what is fundamentally at stake with datafication (Hintz et al., 2018). These matters are shaped and enabled by particular forms of political and economic organisation that promote a normative vision of how social issues should be understood and resolved (Dencik & Sanchez-Monedero, 2022, p.3). Within this broader framing, an important question is how fairness is defined – either from the bottom up, based on people's perceptions, or from the top down, in line with legal frameworks (Heeks & Renken, 2018). These differing approaches highlight ongoing tensions between legal standards and people's lived experiences of fairness in data practices (Heeks & Renken, 2018).

Fairness in assessment

Asking whether a certain use of data-driven technologies could be considered fair in education can only be answered by first defining what fairness means and how it is understood. As Moden et al. (2025) note, it is important to consider that when asking if something is fair, we should also ask *who* defines fairness, for *whom* and from *what* position of power, especially when talking about the use of data-driven technologies in education. Drawing on Bøyum's (2014) analysis of OECD documents, they (Moden et al., 2025) argue that fairness in education is often framed primarily as equal treatment, particularly equal opportunities to participate. However, questions of fairness arise from resource constraints and conflicting norms for institutional allocation and entitlement.

These contested meanings of fairness have direct implications for educational assessment, where they must be operationalised as concrete design and implementation choices. In assessment specifically, fairness now sits with validity and reliability as a core principle (Worrell, 2016). Nisbet and Shaw (2020, pp. 3–4) propose four primary senses of fairness:

- the formal sense: denoting 'accuracy or the appropriate application of a rule or design';
- the implied contractual sense: something is considered fair 'if it meets the legitimate expectations of those affected';
- the relational sense: 'treating (relevantly) like cases alike';
- the retributive sense: something is considered fair if 'it is an appropriate reward (or penalty) for what has gone before'.

The formal sense of fairness is, in and of itself, mostly presented as value neutral. Put differently, the fact of applying a rule says little about the suitability or goodness of the rule (see Nisbet & Shaw, 2020) – when specific conditions are met, a rule is applied. The retributive sense of fairness addresses what subjects of assessment 'deserve' based on their performance, while the retributive sense of fairness may become particularly vulnerable in times of instability such as during the Covid-19 pandemic (Shaw & Nisbet, 2021), because rapid changes brought on by extreme or unprecedented circumstances can lead to questions of deservedness in (temporarily) altered assessment practices.

In addition to the four primary senses of fairness, Nisbet and Shaw (2020) further mention two approaches that may be better understood from the perspective of unfairness – the consequential sense and the retrospective sense. The consequential sense of (un)fairness emerges where assessment may be 'judged as unfair if its outcomes might be used as a basis for unfair actions in the future' (Nisbet & Shaw, 2020, p. 147), while the retrospective sense concerns outcomes that are considered unfair due to past social or structural injustice (Nisbet & Shaw, 2020, p. 5). Together these two additional senses of (un)fairness add a temporal dimension to critically addressing fairness in assessment by linking present outcomes to both future consequences and historical conditions.

Ultimately, fairness in assessment includes technical aspects but is also socially situated, its different senses are evoked by various agents with their own agendas, and the outcomes of assessment matter to a wider group of actors than just those being assessed.

Data and methods

To study media discussions centring on the 'A-level fiasco', we collected topic-relevant international news articles using a custom search tool for listing results. The tool was created by implementing the *pygooglenews* Python library, which allowed us to make time- and keyword-restricted searches. Taking into account the particular social context and high profile of the incident, we ultimately opted to use a relatively broad search phrase 'A-levels' for the time period from July 2020 to October 2021. Since our aim was to capture the discussion in as wide a manner as possible, we concluded that the custom search tool served our research endeavours best when used with a more general keyword. The retrieved search results (N = 461) were imported into a table, which comprised the title of the news article, its date of publication, and the associated hyperlink.

We excluded articles that were only accessible for paying customers or failed to address the central topic because relevant discourse was required for our subsequent analysis and the entire article was considered an initial unit of analysis. After exclusions, our final sample comprised 329 English-language news media articles (2020 = 169; 2021 = 160) from which most were news reports, but op-eds and reported reader letters were also included. The sample includes articles from 122 different media outlets, which cover the topic in the United Kingdom as well as its current (Wales, Northern Ireland, Scotland) and former territories (e.g., United Arab Emirates) that were affected by the decisions made in the context of A-levels grading. The greatest number of articles were published in *The Guardian* (N = 76), *BBC News* (N = 29), *Wales Online* (N = 16) and *Daily Mail* (N = 10).

We used the critical discursive psychology (CDP) approach for analysing the media texts. Since agents who use talk and utterances to (re)construct their version of social events (Locke & Budds, 2020) do not solely draw on a singular large discourse, the CDP approach is particularly interested in the interpretive repertoires; in other words, 'recognisable routines of arguments, descriptions, and evaluations found in people's talk' (Seymour-Smith et al., 2002). Therefore, interpretive repertoires can be understood as forming the connection between micro-level talk and larger socio-cultural discourses (see Kikerpill & Siibak, 2023b). The core of interpretive repertoires comprises the ways in which people make use of specific terminology, phrasing, examples and jargon when assessing and addressing social reality (Locke & Budds, 2020; Kikerpill & Siibak, 2023b). Due to the action-orientation aspect of talk and utterances, people use discourse to position themselves in relation to certain social phenomena or topics; for example, choosing from the pool of available positions to either affirm or resist particular ideas or stances (Kikerpill & Siibak, 2023). Furthermore, talk and utterances are used to engage in a variety of actions, such as justifying, excusing, and blaming (Locke & Budds, 2020), or supporting and affirming aspects of experienced social reality (Kikerpill & Siibak, 2023). These actions can also be achieved via remaining silent or not addressing a topic or parts of it; that is, what agents neglect to mention is as important as what is mentioned (Billig, 1991).

Although interpretive repertoires are usually analysed as they emerge from people's talk and utterances (see Locke & Budds, 2020), we decided to take a somewhat different approach in our application of CDP. Because the descriptions of the senses of fairness suggested by Nisbet and Shaw (2020) provide a highly relevant framework for analysing the discourse about the issues pertaining to the 'A-level fiasco', our analysis focused on whether and (if yes) how agents engage with these categories of *fair* in their everyday explanations of what occurred (Seymour-Smith, 2002; Locke & Budds, 2020). Therefore, while the four senses of fairness as proposed by Nisbet and Shaw (2020) were used as a general guide in our analysis, we remained open to the possibility that the theoretical conceptualisations of the different senses of fairness may not emerge from people's utterances in a 'clean-cut' manner.

Results

Following the analytical approach described above, we established five different categories of results pertaining to the discourse on fairness. Four categories relate to the primary senses of fairness

suggested by Nisbet and Shaw (2020): formal, relational, implied contractual, and retributive. The fifth category emerged from the data as the discursive construction of the algorithm as a relevant and impactful meta-agent. We consider the grading algorithm as a meta-agent because although it lacks the ability for self-expression, the algorithm was discursively constructed as *doing things*. Hence, the algorithm could be construed as partly mimicking complete agency in a discursive manner.

The formal sense of fairness

In public discussions, the formal sense of fairness was linked not only to the rules themselves but also how they were changed and applied in practice. A major concern was whether new rules still rewarded student effort in the same way as before.

> *[X] is a GCSE pupil at Bexleyheath Academy in south London. He says the process will be unfair. 'With previous years, it was always the people who put in more work before the actual exam who were expected to get a better grade because they worked harder for it, but now it's like you can do less work, and then just before the exam you can see what you're supposed to work on and it feels like it's going to be almost too easy this year,' he said.* (Chowdhury, 2020: paras. 8–9)

In the excerpt above, the student describes how a rule change could possibly devalue prior work. Following the 2020 grading fiasco, a rule change was implemented, which was to grant students more generous grades and allow them to preview exam topics. Since the application of the rule was outside the norm, it would effectively dismantle prior approaches to exam preparation. Most importantly, however, it raises questions about whether, or to what extent, the formal sense of fairness in assessment carries independent significance in discursive constructs of events.

Essentially, there are two ways to approach the formal sense of fairness: strict and procedurally augmented. The strict sense would involve a value-neutral application of a rule, while the procedurally augmented formal sense of fairness connects the rule and its application to the social context into which the rule is created. Once created, the application of a rule may be considered fair in the formal sense. Yet, it is difficult, if not impossible, to detach the application of a rule from the relevant context in social practice, which is clearly shown in the provided excerpt. Even if the rule is applied strictly as prescribed, the process of how the rule came to be, and the impact its application will likely continue to exert, can only be separated from the strict formal sense of fairness analytically.

The procedurally augmented formal sense of fairness – not the strict formal sense – can also be witnessed when rules are discussed from an action-perspective; that is, 'ruling something out' as was the case in 2021 when the grading of A-level exams was again a task for the teachers:

> *[Minister of education] Mr Weir said that 'more weight' would be given to the 'professional judgement of teachers' in awarding qualifications in 2021. […] 'This year there will be no statistical standardisation using an algorithm,' he said. […] He said that the grades awarded to pupils would be based on work they had completed in school.* (Meredith, 2021: paras. 12–14)

The controversy surrounding an algorithmic grading solution was serious enough to lead to the system being 'ruled out'. While the phrase 'ruled out' appears in the title of the excerpted news story, it also described the actions in the story. The professional judgement of teachers was also reviewed to 'ensure fairness and consistency' (Meredith, 2021: para. 2). Hence, the social reality, as constructed in the news article, portrayed algorithmic grading as sufficiently unfair to warrant exclusion. Therefore, the experienced social injustice itself shaped the ruling and expectations about any future impact. Discursively, such rulings are not value-neutral and distinguishing between the path to rule, its moment of application and the expected outcomes is, again, analytical. The procedurally augmented formal sense of fairness, therefore, is more useful because it connects rule-making, implementation and outcomes. In effect, the (strict) formal sense of fairness in assessment remains attached to the retrospective or consequential senses of fairness.

The implied contractual sense of fairness

The implied contractual sense of fairness pertains primarily to legitimate expectations. Even though what exactly constitutes feasible in terms of expectations varies from one situation to another, the discourse produced in the context of the A-levels fiasco suggests a number of typical cases. First, the implied contractual sense of fairness appears with respect to what students had expected from their final grades:

> *'To be quite frank with you, I burst into tears in front of everyone. In front my headteacher, everyone because it was a very traumatic experience really', she told The Independent. […]'I was not expecting that. I had put my faith in my teachers and the system that I would be treated fairly. That the government would ensure that I was able to access my university course that I need for my future, and I was severely let down by the system.'* (Cowburn, 2020: paras. 17–18)

In the extract above, the expectation of fairness was addressed to the system and its decision-making agents. After receiving disappointing exam results, it was claimed that students would be able to appeal the results, but the appeals process had not been elaborated by the state at that point (Cowburn, 2020). This illustrates how structural bias in algorithmic decision-making became an individual burden, as the responsibility for challenging the outcome fell on the student alone. The absence of a functioning appeals process also reflects the procedurally augmented formal sense of fairness, since the system failed to provide the necessary procedural step for the affected individual. This situation also highlights a wider issue raised by researchers: the need to make algorithmic systems more transparent and open to scrutiny (Pasquale, 2015). As Porayska-Pomsta et al. (2023, p. 579) point out, it is not enough for systems to be technically transparent – they also need to be understandable to users. Without clear explanations and accountability, individuals are left to deal with complex outcomes alone, without a fair opportunity to challenge the decisions made.

Second, by standing between students and educational bureaucracy, teachers have a mid-level view on the effects that official positions create for students and the school:

> *'We've got amazing students, some of whom do incredibly well, some of whom have had childhood experiences and traumas which mean they don't do as well, but they all work hard, and the school is on an improvement journey,' she said, adding that the results were, quite simply, 'not fair'.* (Ferguson & Savage, 2020: para. 2)

This excerpt reveals that expectations are oftentimes already adjusted to the particular situation of an individual, while the importance of acknowledging effort is emphasised throughout. When individualised expectations are widely met with the reaction of 'unfair', it points to a systemic miscommunication between those performing assessments and the individuals being assessed. In the A-levels downgrading case, the use of an algorithm facilitated this communicative process, and revealed how manufactured aggregates of situated populations devoured and marginalised individual efforts:

> *The school or college's past results have no bearing on an individual student's academic performance this year. To apply this in the form of an algorithm to remotely adjust the grades of individual students is both grossly unfair and perverse. Why can't the government trust teachers, universities and employers to work out how best to support the ambitions of today's young people?* (Chris Pratt as cited in Letters, 2020: para. 5)

Here, the issue is twofold: where the expectations come from and who is affected by them. The former pertains to the crucial distinction between individual performance even where this must be adjusted to account for current circumstances and relying on aggregated data which reduces students to an institutional average. The latter concerns replacing human judgement with algorithmic objectivity when interpreting individual effort. This is of particular significance in an assessment which is likely to impact people's lives long after the exam itself (see Wright, 2021). Because the algorithm could not account for exceptional circumstances or apply leniency, it failed to meet these expectations.

The relational sense of fairness

In a pertinent editorial, Bart van der Sloot (2023) conveyed the idea that differentiating, or discriminating, on legitimate grounds is a common necessity to arrive at the correct decision in a particular situation. It is, however, the arbitrary use of power and mismatches between grounds for discrimination and the outcomes to be achieved that should be alarming (van der Sloot, 2023). As mentioned previously, Nisbet and Shaw (2020) accounted for the same reality in educational assessment: discrimination and differentiation occur, the issue to be mindful of pertains to the legitimacy of the grounds for such discrimination. In our analysis, one notion of arbitrary algorithmic discrimination emerged with reference to a 'postcode lottery':

> *[X] said his peers deserve better than to be graded by a postcode lottery. He said: 'I'm very conscious that something which works as a good system for me as someone who goes to a west London comprehensive school, with some of the best teachers I could possibly ask for and a great academic history, is probably not representative of what's good for the rest of the country.'* (Mohdin, 2020: para. 5)

The idea captured in the excerpt was similarly expressed with respect to the 'implied contractual sense of fairness'. Yet, the relational sense of fairness unravels the bases from which legitimate expectations emerge, and how these expectations are violated once outcomes show that legitimate grounds were not abided by. In this case, the so-called postcode lottery functioned as the epitome of the A-levels fiasco. First, it captured the cold objectivity of algorithmic decision-making that, in this instance, could not account for individual effort, and instead generated downgrades based on historical results received in specific schools. As expressed in the excerpt, a good school meant a better historical record of results, which subsequently came to be reflected in the algorithmic outcomes. Second, the excerpt indirectly revealed that algorithms are no more than an extension of what humans view as the necessary elements in calculating certain outcomes. If school history is to be included in the reflection of personal effort, then individuals are, by default, morphed into representatives of a population, and everyone included must bear the weight of a datafied history in (quiet) solidarity. Also, the excerpt pointed to systemic issues of human-infrastructural inequality with respect to the degree of instruction and guidance available to the involuntary participants in the 'postcode lottery'. In effect, the use of a poorly constructed algorithm only re-emphasised the broader problems present in modern education, which merely became crystallised and put on clear display following the algorithm's use.

The use of the algorithm, as such, was particularly peculiar, given that education officials were already aware of the risks of bias and groundless discrimination; for example, as expressed by the Education Committee chairman:

> *'We have serious worries about the fairness of the model developed by Ofqual.[…] 'There is a risk it will lead to unfair bias and discrimination against already disadvantaged groups.' […] He also questioned the fairness of the appeals system, which he said seemed to favour 'the well-heeled and sharp-elbowed', who know how to navigate the system. […] He added that there was a potential for the system to resemble the 'Wild West of appeals' with different systems being used by different exam boards.* (Richardson, 2020: paras. 16–19)

To an extent, this may be explained by the fear that the inherent unconscious human bias (see Richardson, 2020), which is present in the school system 'anyways', would have impacted students' final grades unfairly. Yet, the end result was to replace one form of bias for another, which raises further questions about the feasibility of using large-scale algorithmic decision-making in already contested areas of social life. Put differently, the perceived novelty of using algorithms in decision-making contexts, along with the cold objectivity that algorithms are thought to introduce, merely mirrors existing social structures rather than transforming them (see Broussard 2023). In the algorithmic considerations pertaining to A-level grading decisions, the notion that something could be objectively unfair gets lost. There is no equals sign between objectivity and fairness because fairness is a conceptual container with highly mobile contents.

The retributive sense of fairness

Expectations created on the basis of promises, guides and, in recurring situations, the experiences of preceding cohorts set in motion what succeeding cohorts might consider as fair in the retributive sense. In connection with the sense of fairness analysed above, the retributive sense emerged clearly from potentially not being able to sit exams and have the results substituted with calculated grades:

> *Another A-level student […] backed the decision to press ahead with exams as fairest option. 'I would be very disappointed not to be given the opportunity to prove my ability to exam boards to achieve the grade I truly deserve, not one generated by an algorithm or even allocated by a teacher.'* (Coughlan, 2020: paras. 23–24).

The excerpt above addressed two issues with respect to the retributive sense of fairness (Nisbet & Shaw, 2020). First, the possibility of having to deal with calculated grades was viewed from the perspective of what opportunities others had in the past. Since others were provided with the opportunity to exhibit individual effort and knowledge, making a change in the assessment was considered an unfair consequence to something that the students did not bring about themselves. Second, the historical background of different schools would have played an overemphasised and unfair role in the calculated grades. As with the previous senses of fairness analysed above, the cold objectivity of algorithmic assessment would incorporate historical unfairness into completely new results that ought to be based on what a specific individual achieved. Moreover, the choice of algorithmic grading also impacted other members of the educational ecosystem; for example, by shutting out and overwriting teachers' expertise, we also see how human agency is outsourced to automatic machines and turned into a mechanistic approach to fairness:

> *'We're questioning how a change that significant can be fair,' said Hannafin. 'Teachers have spent a lot of time calculating the centre-assessed grade with evidence, and they have got the professional expertise to do that. We can't have students getting grades just because they fit the past profile of schools' results. That is unfairness. We have to rectify that.'* (Ferguson & Savage, 2020: para. 14).

Hence, the use of an algorithm to assign grades was viewed widely as an unfair penalty for something which students and teachers had no control over, like the pandemic circumstances. More broadly, this signifies human reactions to being effectively shut out from a familiar ecosystem by a machine replacement on which the government had placed overbroad hopes of objective fairness. It was this particular sense in which the algorithm became personalised as something that overtook humans in terms of power and control related to the circumstances. In short, while teachers experienced unwelcome thoughts of displacement, students were left to experience unwanted feelings of dependence. While the use of the algorithm was broadly perceived as a poor decision that was inappropriate even in the circumstances of the pandemic, the algorithm itself also became a connective embodiment of unfairness in the government's treatment of the individual.

The algorithm as a meta-agent

Focus on the grading algorithm as a meta-agent came in two main forms. The first of these saw the grading algorithm as an absorbing agent the perceived social malfunction of which served to devalue and lessen the agency of the government. The second perspective on algorithmic meta-agency came in the form of what an algorithm ought to be like without referencing the human involvement in effecting such changes or standards.

The 'black box' nature of algorithmic systems, and the associated issues of transparency and accountability were expressed in the discourse as follows:

> *This points to a shift in the government's perspective of, and expectations for, accountability. Algorithmic systems are opaque and complex "black boxes" that enable powerful political decisions to be made based on mathematical calculations, in ways not always clearly tied to legal requirements. […] Failure of public authorities to ensure that algorithmic systems are accountable is at worst a deliberate attempt to hinder democratic processes by shielding algorithmic systems*

> *from public scrutiny. And at best, it represents a highly negligent attitude towards the responsibility of government to adhere to the rule of law, to provide transparency, and to ensure fairness and the protection of human rights.* (Harkens, 2020: paras 12, 14)

The choice of words in the excerpt above clearly suggests that an agential struggle was observed to have taken place between the government and its decision to use the grading algorithm. In this way, discursively expressed power imbalances between those who develop and set policies, and the measures ultimately taken to bring said policies to bear, include the potential to confuse the addressees as to who or what holds power over the consequences of the aforementioned policies. Moreover, engaging such discourses creates the possibility for misattributing responsibility; on one hand, mistakenly relieving some pressure from policymakers and, on the other hand, attributing the responsibility to an inanimate object. Even so, the algorithm had become recognisable enough to merit utterances addressed more to the algorithm as such rather than its creators:

> *The exam grading algorithm may not be sophisticated AI, but to be ethical it should still adhere to certain principles, including that it should impact positively on teaching and learning; be fair and explainable to the people whom it impacts; and that it should use data that is not biased towards or against any particular group of people.* (Letters: para 1)

The expressed attention to what was wrong with the algorithm situates it as a cultural object, but the wording also hints at meta-agency; for example, the algorithm or "it" ought to impact positively on teaching and learning. Moreover, the utterance claims that the algorithm must be fair, although this characteristic is presented as some inherent quality of an algorithm, which could be erroneously construed as being detached from the functions that the algorithm is expected to perform by its creators.

Conclusion

As governments increasingly move towards automating core functions of implementing policies, including educational policies, the issue of algorithmic fairness has emerged as an element that requires further attention. In this article we have explored how different target groups impacted by the algorithmic grade fiasco that occurred in the UK in 2020, expressed and relied on different senses of fairness when commenting on the incident in international news media (N=329).

Our analysis indicated that the four senses of fairness – formal, relational, implied contractual and retributive (Nisbet & Shaw, 2020) – were present and active in the utterances of those impacted by the application of the grading algorithm. Moreover, the algorithm itself emerged from the discourse as a meta-agent that, through discourse, was made to carry some of the responsibility. These findings resonate with Masso and Kasapoglus' (2020) notion of triple agency, where algorithms, data subjects, and experts co-produce decision-making processes and accountabilities. In our case, the grading algorithm was not merely treated as a passive tool but was constructed discursively as a meta-agent. This aligns with Watson's (2019) argument that technological opacity allows systems to appear as independent actors. Such a perception enabled state actors to assign partial responsibility to the algorithm itself, which complicates traditional understandings of agency and fairness. The way fairness was framed, particularly through the relational and implied contractual senses, suggests that stakeholders did not view the algorithm in isolation, but as part of a broader socio-technical system in which human and technological actors jointly shape outcomes. This supports the idea that algorithmic systems can take on policy-enacting roles (Masso & Kasapoglu, 2020), while also being shaped by the very discourses they are embedded in. It also reflects ideas from data justice research, which questions what is really at stake when decisions are made using data technologies (Dencik et al., 2019). There was a clear tension between how fairness was understood by people affected and how it was defined by rules and systems (Heeks & Renken, 2018). Ultimately, our critical analysis of media discourses revealed that technological shortcuts often have an effect contrary to what was initially imagined by those responsible for implementing policy. In educational contexts, flawed and biased input data is bound to result in flawed and

biased output, potentially creating situations where postal codes matter more than studying.

Furthermore, in the context of education, our findings extend existing research on algorithmic fairness by highlighting how fairness is framed not only through technical accuracy or outcomes, but also through expectations of merit, trust, and responsibility. The analysis shows that students and educators draw heavily on relational and implied contractual notions of fairness, which are largely absent from formal policy and technical design frameworks. This gap helps to explain why algorithmic systems may be experienced as unjust even when they are formally compliant with policy objectives.

References

Airoldi, M. (2021). *Machine Habitus: Toward a Sociology of Algorithms*. John Wiley & Sons.

Billig, M. (1991). *Ideology and opinions: Studies in rhetorical psychology*. Sage Publications, Inc.

Bøyum, S. (2014). Fairness in education – a normative analysis of OECD policy documents. *Journal of Education Policy*, *29*(6), 856–870. https://doi.org/10.1080/02680939.2014.899396

Broussard, M. (2023). *More than a glitch. Confronting race, gender, and ability bias in tech*. MIT Press.

Broussard, M. (2019). *Artificial unintelligence. How Computers misunderstand the world*. MIT Press.

Burgess, M., Schot, E., & Geiger, G. (2023, March 6). This algorithm could ruin your life. *Wired*. https://www.wired.com/story/welfare-algorithms-discrimination/

Chowdhury, S. (2020, December 3). COVID-19: More generous grades, crib sheets and fewer topics: New rules for exams in England. *Sky News*. https://news.sky.com/story/covid-19-more-generous-grades-crib-sheets-and-fewer-topics-new-rules-for-exams-in-england-12149671

Coughlan, S. (2020, October 12). Next year's exams in England delayed but still going ahead. *BBC News*. https://www.bbc.com/news/education-54508851

Cowburn, A. (2020, August 13). A-level results: Gavin Williamson facing backlash over grading system as private schools see biggest increase in top grades. *The Independent*. https://www.independent.co.uk/news/uk/politics/a-level-results-private-school-state-gavin-williamson-grades-a9669571.html

Dencik, L. (2025). 'Rescuing' data justice? Mobilising the collective in responses to datafication. *Information, Communication & Society, 28*(6), 1023–1038.

Dencik, L., & Sanchez-Monedero, J. (2022). Data justice. *Internet Policy Review*, 11(1), 1–16. https://doi. org/10.14763/2022.1.1615

Dencik, L., Hintz, A., Redden, J., & Treré, E. (2019). Exploring data justice: Conceptions, applications and directions. *Information, Communication & Society*, 22(7), 873–881. https://doi. org/10.1080/1369118X.2019.1606268

Elliot, A. (2024). *Algorithms of anxiety. Fear in the digital age*. Polity.

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St.Martin's.

Fendler, L. (2016). Ethical implications of validity-vs.-reliability trade-offs in educational research. *Ethics and Education*, *11*(2), 214–229. https://doi.org/10.1080/17449642.2016.1179837

Ferguson, D., & Savage, M. (2020, August 16). Autumn term chaos feared over exam resits and appeals. *The Guardian*. https://www.theguardian.com/education/2020/aug/16/autumn-term-chaos-feared-over-exam-resits-and-appeals

Gulson, K. N., Sellar, S., & Webb, P. T. (2022). *Algorithms of education: How datafication and artificial intelligence shape policy*. University of Minnesota Press.

Harkens, A. (2020, September 3). Not just A-levels: Unfair algorithms are being used to make all sorts of government decisions. *The Conversation*. https://theconversation.com/not-just-a-levels-unfair-algorithms-are-being-used-to-make-all-sorts-of-government-decisions-145138

Heeks, R., & Renken, J. (2018). Data justice for development: What would it mean?. *Information Development, 34*(1), 90–102.

Jarke, J., & Breiter, A. (2019). The datafication of education. *Learning, Media and Technology, 44*(1), 1–6. https://doi.org/10.1080/17439884.2019.1573833

Kelly, A. (2021). A tale of two algorithms: The appeal and repeal of calculated grades systems in England and Ireland in 2020. *British Educational Research Journal*, *47*(3), 725–741. https://doi.org/10.1002/berj.3705

Kikerpill, K., & Siibak, A. (2023a). AI in schools and universities: Mapping central debates through enthusiasms and concerns. In S. Nah (Ed.), *Research Handbook on AI and Communication* (pp. 94–107). Edward Elgar Publishing.

Kikerpill, K., & Siibak, A. (2023b). Schools engaged in doom-monitoring students' online interactions and content creation: An analysis of dominant media discourses. *Child and Adolescent Mental Health*, *28*(1), 76–82. https://doi.org/10.1111/camh.12621

Knox, J., Williamson, B., & Bayne, S. (2020). Machine behaviourism: Future visions of 'learnification' and 'datafication' across humans and digital technologies. *Learning, Media and Technology*, *45*(1), 31–45. https://doi.org/10.1080/17439884.2019.1623251

Kolkman, D. (2020, August 26). "F**k the algorithm"?: What the world can learn from the UK's A-level grading fiasco. *LSE Impact Blog*. https://blogs.lse.ac.uk/impactofsocialsciences/2020/08/26/fk-the-algorithm-what-the-world-can-learn-from-the-uks-a-level-grading-fiasco/

Koulish, R., & Evans, K. (2021). Punishing with impunity: The legacy of risk classification assessment in immigration detention. *Georgetown Immigration Law Journal*, *36*(1). https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=6813&context=faculty_scholarship

Lauri, T., & Saar, E. (2022). Cumulative advantages and disadvantages in attainment of higher education: Set-analytic comparison of asymmetric inequalities in six European countries. *International Journal of Comparative Sociology*, *63*(1–2), 51–88. https://doi.org/10.1177/00207152221092152

Letters. (2020, August 14). Damage and disillusion caused by A-level downgrades. *The Guardian*. https://www.theguardian.com/education/2020/aug/14/damage-and-disillusion-caused-by-a-level-downgrades

Li, K. C., & Wong, B. T. (2023). Artificial intelligence in personalised learning: a bibliometric analysis. *Interactive Technology and Smart Education*, *20*(3), 422–445. https://doi.org/10.1108/itse-01-2023-0007

Lindgren, S. (2024). *Critical theory of AI*. John Wiley & Sons.

Locke, A. (2015). Agency, 'good motherhood' and 'a load of mush': Constructions of baby-led weaning in press. *Women's Studies International Forum*, *53*, 139–146. https://doi.org/10.1016/j.wsif.2014.10.018

Locke, A., & Budds, K. (2020). Applying critical discursive psychology to health psychology research: A practical guide. *Health Psychology and Behavioral Medicine, 8,* 234–247. https://doi.org/10.1080/21642850.2020.1792307

Masso, A., & Kasapoglu, T. (2020). Understanding power positions in a new digital landscape: perceptions of Syrian refugees and data experts on relocation algorithm. *Information, Communication & Society, 23*(8), 1203–1219. DOI: 10.1080/1369118X.2020.1739731

Meredith, R. (2021, February 2). Education minister rules out stats tools for exam grades. *BBC News*. https://www.bbc.com/news/uk-northern-ireland-55895279

Mhasawade, V., Zhao, Y., & Chunara, R. (2021). Machine learning and algorithmic fairness in public and population health. *Nature Machine Intelligence*, *3*(8), 659–666. https://doi.org/10.1038/s42256-021-00373-4

Mitchell, S., Potash, E., Barocas, S., D'Amour, A., & Lum, K. (2020). Algorithmic Fairness: Choices, assumptions, and definitions. *Annual Review of Statistics and Its Application*, *8*(1), 141–163. https://doi.org/10.1146/annurev-statistics-042720-125902

Mohdin, A. (2020, August 13). Downgraded A-level students urged to join possible legal action. *The Guardian*. https://www.theguardian.com/education/2020/aug/13/downgraded-a-level-students-urged-to-join-possible-legal-action

Morozov, E., & Myers, J.J. (2013). *To save everything, click here: The folly of technological solutionism*. Public Affairs. https://media-1.carnegiecouncil.org/import/studio/To_Save_Everything_Click_Here.pdf

Männiste, M., & Siibak, A. (2025). AI magnified inequalities: bias, (un)fairness, and discrimination resulting from the use of AI-based technologies in the education sector. In T.-A. Wilska & J. Nyrhinen (Eds.). *Young People in Digital Environments. Agency, Risks and Opportunities* (pp. 219–233). Edvard Elgar.

Neumann, E. (2019). Setting by numbers: Datafication processes and ability grouping in an English secondary school. *Journal of Education Policy, 36*(1), 1–23. https://doi.org/10.1080/02680939.2019.1646322

Nisbet, I., & Shaw, S. (2020). *Is assessment fair?* Sage.

Pasquale, F. (2015). *The black box society.* Harvard university press.

Porayska-Pomsta, K., Holmes, W., & Nemorin, S. (2023). The ethics of AI in education. In *Handbook of Artificial Intelligence in Education* (pp. 571–604). Edward Elgar Publishing.

Richardson, H. (2020, July 11). GCSE and A-level results 'could be affected by bias'. *BBC News*. https://www.bbc.com/news/education-53364485

Sadowski, J. (2025). *The Mechanic and the Luddite: A Ruthless Criticism of Technology and Capitalism*. Univ of California Press.

Selwyn, N., O'Neill, C., Smith, G., Andrejevic, M., & Gu, X. (2023). A necessary evil? The rise of online exam proctoring in Australian universities. *Media International Australia*, *186*(1), 149–164. https://doi.org/10.1177/1329878X211005

Seymour-Smith, S., Wetherell, M., & Phoenix, A. (2002). 'My wife ordered me to come!': A discursive analysis of doctors' and nurses' accounts of men's use of general practitioners. *Journal of Health Psychology*, *7*(3), 253–267. https://doi.org/10.1177/1359105302007003220

Shaw, S., & Nisbet, I. (2021). Attitudes to fair assessment in the light of COVID-19. *Research Matters*, 31, 6–21. https://www.cambridgeassessment.org.uk/Images/research-matters-31-attitudes-to-fair-assessment-in-the-light-of-covid-19.pdf

Sleep, L., & Redden, J. (2024). Reimagining failed automation: from neoliberal punitive automated welfare towards a politics of care. In *Handbook on Public Policy and Artificial Intelligence* (pp. 366–382). Edward Elgar Publishing.

Starke, C., Baleis, J., Keller, B., & Marcinkowski, F. (2022). Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature. *Big Data & Society*, *9*(2). https://doi.org/10.1177/20539517221115189

Thibaud, E. (2025). Reflections on techno-solutionism in education: Manifestations and causes. *Educational Philosophy and Theory*, 1–12. https://doi.org/10.1080/00131857.2025.2528852

Olojede, H. T. (2024). Techno-solutionism a Fact or Farce? A Critical Assessment of GenAI in Open and Distance Education. *Journal of Ethics in Higher Education, 1*(2024), 193–216. https://doi.org/10.26034/fr.jehe.2024.5963

Van der Sloot, B. (2023). Editorial. *European Data Protection Law Review*, *9*(2), 93–97. https://doi.org/10.21552/edpl/2023/2/3

Watson, D. (2019). The rhetoric and reality of anthropomorphism in artificial intelligence. *Minds and Machines*, *29*(3), 417–440. https://doi.org/10.1007/s11023-019-09506-6

Wetherell, M. (2015) Discursive psychology: Key tenets, some splits, and two examples. In I. Parker (ed.) *Handbook of Critical Psychology* (pp. 315–324). Routledge.

Williamson, B. (2016). Digital education governance: An introduction. *European Educational Research Journal*, *15*(1), 3–13. https://doi.org/10.1177/1474904115616630

Worrell, F. (2016). Commentary on 'perspectives on fair assessment'. In N. J. Dorans & L. L. Cook (Eds.), *Fairness in Educational Assessment and Measurement* (pp. 283–293). Routledge.

Wright, J. (2021, January 29). Pupils should have the RIGHT to repeat academic year because lockdown will leave them 'scarred for life' by exam grades - after PM said schools will stay shut until at least March 8. *MailOnline*. https://www.dailymail.co.uk/news/article-9199651/Ministers-urged-act-prevent-major-public-policy-disaster-exams.html

Zajko, M. (2022). Artificial Intelligence, Algorithms, and Social Inequality: Sociological Contributions to Contemporary Debates. *Sociology Compass*, *16*(3). https://doi.org/10.1111/soc4.12962

Zytek, A., Liu, D., Vaithianathan, R., & Veeramachaneni, K. (2021). Sibyl: Understanding and addressing the usability challenges of machine learning in high-stakes decision making. *IEEE Transactions on Visualization and Computer Graphics, 28*(1), 1161–1171.

Zwick, R., & Dorans, N. J. (2016). Philosophical perspectives on fairness in educational assessment. In N. J. Dorans & L. L. Cook (Eds.), *Fairness in Educational Assessment and Measurement* (pp. 267–282). Routledge.

Biographical statements

Kristjan Kikerpill is a lecturer in Information Law and Digital Sociology at the Institute of Social Studies, University of Tartu, Estonia. His main research areas are the social and communicative aspects of cybercrime and critical data studies with a focus on privacy and surveillance in everyday life.

Andra Siibak is a Professor of Media Studies and Deputy Head of Research at the Institute of Social Studies, University of Tartu, Estonia. Her research focuses on the opportunities and risks associated with the use of AI-based technologies and the internet. Together with Giovanna Mascheroni she has co-authored "Datafied Childhoods: Data Practices and Imaginaries in Children's Lives" (Peter Lang, 2021), and "Children and AI: Changing Digital Childhoods" (Palgrave, 2026). She serves as Governing Body member for the European Communication Research and Education Association (ECREA) and is currently the Vice President of the Association of Internet Researchers (AoIR).

Maris Männiste is a lecturer in Critical Data Studies at the Institute of Social Studies, University of Tartu, Estonia. Her research focuses on the intersection of critical data and algorithm studies, media and communication, and public administration, with a particular emphasis on welfare automation. She investigates how data-driven and automated decision-making in welfare services reshapes citizen-state relations and what challenges increasing automation poses at individual, institutional and societal levels. Her work also engages with the datafication of education and the use of artificial intelligence in educational contexts, examining how similar dynamics unfold across domains.