Emotional AI and EdTech: Serving the Public Good?

McStay, Andrew

Learning, Media and Technology

Peer reviewed version

Cyswllt i'r cyhoeddiad / Link to publication

27. Aug. 2022

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Emotional AI and EdTech: Serving the Public Good?

Abstract

Education Technology (EdTech) companies are deploying emotional AI to quantify social

and emotional learning. Focusing on facial coding emotional AI that uses computer vision

and algorithms to see, recognise, categorise and learn about facial expressions of emotion,

this paper evaluates nascent usage of these technologies in education. To do this, it assesses

the nature of child rights, the history and modern usage of face-based emotional AI,

methodology and efficacy, and what this paper sees as a clash of private and public interests.

Concern is shown to be two-fold: first is on method, especially given scope for material

effects on students; second are ethical and legal concerns. While proposing a list of

considerations for any implementation for these technologies in the classroom, the paper

concludes that significant risks exist in deployment of these technologies in the classroom.

Key words: Affective computing, creepiness, EdTech, emotional AI, human rights.

Introduction [A]

Emotional AI is a collect-all term to refer to affective computing techniques, machine learning and Artificial Intelligence (AI) (AUTHOR). This is a weak form of AI in that these technologies read and react to emotions through text, voice, computer vision and biometric sensing, but they do not have sentience or undergo emotional states themselves. Ultimately based on mathematics, analytics are applied to describe and predict behavioural patterns among people by generating data points from people's words, pictures, intonation, gestures, physiology and facial expressions.

In education, the application of emotional AI promises to assist with personalised learning, development of social and emotion learning, to understand if students are struggling with class material, and which students need to be challenged further by class content. This idea is not new: back in 2001, Rosalind Picard (the originator of the term and practice of affective computing), recognised the scope for emotion-sensing in education. Indeed, her team traces the development of computerised tutors that are sensitive to emotion and affect back to 1988 (Lepper and Chabay 1988, Picard 2001a). Noting the diversity of emotional and affective behaviours in the classroom (frustration, dejection, hopelessness, curiosity, interest, exploration and enjoyment), Picard et al. (2001) sought to build a 'computerized learning companion that facilitates the child's own efforts at learning'. The goal of the companion was to improve pedagogical techniques by using computer vision techniques to watch and respond to the affective states of children. The context of this observation was two-fold: that the interplay between emotions and learning is highly important; and that education systems do not focus sufficiently on *how* rather than *what* students learn (Picard et al. 2001a).

This paper anticipates the consequences of deploying emotional AI-based Education

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Technologies (EdTech) in the classroom. Influenced by postphenomenology that is keen to avoid dystopian narratives of technology or progressivist utopianism (Idhe 2002), the paper is interested in the empirics of face-based emotional AI employed in EdTech. It is also sensitive to pragmatism where moral truths (potentially fiercely fought for) are contingent upon period and context (Rorty 2010 [1993]). As such, it does not see the datafication of human emotion as innately wrong, especially given wider writing on emotional AI and its potential benefits in human-computer interaction (AUTHOR). To assess emotional AI that makes use of facial coding techniques, it examines these in relation to internationally agreed principles laid out in the *United Nations Convention on the Rights of the Child* (OHCHR 1990), an influential visionary report by the World Economic Forum (WEF 2016), and critical literature on emotional AI and EdTech pedagogy. It answers whether schools that employ companies using products based on emotional AI are serving the public good. The paper concludes with a list of considerations that should be passed before emotional AI-based EdTech is employed in education.

Rights of the child [A]

To consider the merits of deploying emotion-sensing technologies in education settings, it is useful to ask whether there exists a variance between social values and the technological capability of emotional AI. In other words, is there an inherent *creepy factor* (Tene and Polonetsky 2014)? In the case of facial coding emotional AI in the classroom, it appears that there may well be a divergence. The technologies will be unpacked below, but to understand the social values part of the moral equation it is useful to understand in some detail the *United Nations Convention on the Rights of the Child*, an agreed set of principles ratified by the majority of the international community for persons under the age of 18 (OHCHR 1990). While support of the United Nations amongst its members is not equal, it is still notable that

this community spans normative, social, cultural, legal, economic and political differences.

The Convention primarily details the most fundamental rights of a child (such as parental access, sanitation, nutrition, care and physical and mental wellbeing), but also states that signatories should act in the best of interest of children, that children have the right to have their privacy protected, and that their lives not be subject to excessive interference. There are at least six relevant articles including: the need to act in the child's best interests (Art. 3); the child's right to freedom of thought (Art. 14) and to privacy (Art. 16): the child's right to physical, mental, spiritual, moral and social development (Art. 27); the child's right to dignity when being disciplined (Art. 27); the development of the child's personality, talents and mental and physical abilities to their fullest potential (§1 Art. 29); the child's right to liberty (§2 Art. 29); and the child's right to be protected from economic exploitation and from performing any work 'likely to be hazardous or to interfere with the child's education, or to be harmful to the child's health or physical, mental, spiritual, moral or social development' (Art. 32).

In addition to technological capability and social values, a normative assessment should also consider the motive to deploy technology. In the case of EdTech companies, this is primarily financial. This contrasts with Art. 3 that requires that the interest of children be the primary consideration. Some pragmatism and appreciation of educational specifics are required, not least because there are many commercially developed products used in education and children's services. Rather, the presence of private interests in public service needs to be seen both in context of the technology itself (that uses biometric data about children) and how EdTech that uses emotional AI fares with other aspects of the Convention. While there is a good argument to be made that business and educational interests may align in service

provision, it is also correct to say that child wellbeing must be a secondary consideration for businesses because their primary goal must be revenue (for profit and to pay staff and shareholders). Next, on Art. 16 and Art. 32, it appears that this profit motive connects with arbitrary interference with privacy because power and technology is used in such a way that clearly advances the interest of a company, yet the benefit for children remains unclear. Art. 27 and Art. 29 also have application, especially given interest in 'mental, spiritual, moral and social development'. The scope for 'chilling effects', and how young people experience their own emotional and affective states, is relevant here, especially definitions that ground chilling effects as when a person may feel deterred even when the activity they are engaging in is not the one an organisation seeks to stop (Schauer 1978)

Face-based emotional AI [B]

For those unfamiliar with facial coding, the historical context is useful to understand that these logics are not new. The clear forerunner to modern facial coding is Guillaume-Benjamin-Amand Duchenne de Boulogne. Duchenne (1990 [1862]) is a 19th century neurologist who used cutting-edge technology of his time to track and understand facial muscles, create a semiotic taxonomy of facial behaviour, and then capture emotion expressions and findings through photography. Not only did he map and record emotional expressions, he set in place the intellectual discourse that today permeates facial coding applications and arguably wider emotional AI. For Duchenne there was no possibility that emotional expressions could be culturally contingent, arguing, 'Education and civilization only develop or modify' abilities to express and monitor emotions (ibid: 29). The logics and technics are important because, even around 150 years ago, forerunners of emotional AI sought to *gauge* facial expressions as universalized and autonomic *leaks* of emotion (Duchenne 1990 [1862]). Other applications tried to *meter* pleasure (Edgeworth 1881),

attention and learning (Wundt 1902).

In the 1960s and 70s Paul Ekman and Wallace Friesen drew upon Tomkins (1962, 1963), and less directly Darwin (2009 [1872]), to continue Duchenne's arguments. They posited a universal 'Basic Emotions' view that suggests existence of biologically given programs that control emotional reactions, and that facial muscles function as a feedback system for emotional experience (Ekman and Friesen 1971, 1978). Famously and controversially, in their New Guinea study, Ekman and Friesen showed photographic plates of white westerners and their facial expressions to local people assumed to be untainted by media images. The goal was to assess whether they would recognise these expressions, which they believed would prove a universal language of emotional expression. Technologically, the research technique employed static photography and then-recent hand-held colour video cameras to provide a dynamic sense of how test subjects reacted and formed their own facial expressions in response to the stimuli. This work led to the development of the Facial Action Coding System (FACS), the measuring of facial movement in humans and the refining of a taxonomy of human emotions and facial expressions (Ekman and Friesen 1978). FACS is based on Ekman's identification of seven facial expressions of primary emotion (joy, surprise, sadness, anger, fear, disgust and contempt); three overall sentiments (positive, negative, and neutral), advanced emotions (such as frustration and confusion) and 19 Action Units (AUs). Today, the FACS system that details the facial action units, head positions and movements, eye positions and related movements, descriptors, behaviours and codes to map faces involves more than 7000 action unit combinations (McDuff 2014).

As will be developed, the nature of emotional life is by no means settled and there is extensive controversy about Basic Emotions and categorical approaches. By contrast, the

dimensional approach rejects the idea of Basic categories. It argues for broad dimensions of experience based on *valence* and whether the experience is pleasant or unpleasant, and how *aroused* a person is. Taken together valence and arousal provide a two-dimensional axis on which emotional experiences can be mapped. Anxiety and distress, for example, map onto the same quarter of the valence/arousal axis (Zelenski and Larsen 2000). The dimensional approach rejects the idea of upfront programs of emotion, seeing emotions as labels (which are linguistically and socially constructed) that we attribute to affective states. However, the FACS, Basic and categorical approach is appealing to technologists because it reduces the complexity of understanding what emotions are and how they can be diagnosed.

Emotional AI itself derives from affective computing techniques and advances in machine learning and 'artificial intelligence' (AI). Taken together AUTHOR labels these as emotional AI. This is a weak, task-based, non-general and narrow form of AI (Warwick 2012) that is reliant on machine learning, a sub-domain of wider AI research (Cawsey 1998). The reading of expressions involves machine training and reading of words and images, seeing and sensing facial expressions, gaze direction, gestures and voice. It also encompasses machines sensing and learning about heart rate, body temperature, respiration and the electrical properties of our skin, among other bodily behaviours. While there exists a variety of emotional AI techniques, it is FACS-based facial analytics that is receiving the most interest from EdTech companies and legacy technology companies, such as Microsoft and Intel, that see opportunity in the education "market". Face-based emotional AI is typically reliant on computer vision (a perceptual aspect of AI) and supervised machine learning. Unlike systems (such as video) that simply record and store visual information, computer vision also extracts information from images, video and what cameras see. This is not straightforward and relies on mathematical and probabilistic means to identify three-dimensional shapes and objects in

images (Szeliski 2010). Although computer vision has roots in the late 1960s and 1970s as the attempt to extract three-dimensional understanding of the world for robots (Boden 2006), in the 2000s vision recognition research began applying machine learning techniques to computer vision problems (Szeliski 2010). Here, in context of computer vision that seeks to identify emotion expressions from faces, machines are first shown training examples of named facial emotions. Coded by humans, this entails people trained in the Facial Action Coding System (FACS) taxonomy to deconstructing facial behaviour into anatomic elements (action units) and thereafter identify what given emotional expression has taken place. This raises a key concern: who is doing the annotating and who are the people whose expressions are being read? As identified in AUTHOR ****, there is evidence that it is typically white Westerners that hand-annotate the facial expressions used as the basis for international facial coding applications. Although harms regarding bias and ethnicity have not been established in relation to facial coding, the potential is clear: training data in some cases is racially distorted.

Having been provided with classifications and labels, and trained, computers see with cameras, receive input features, register pixelated facial elements, process these against labelled training data and generate outputs of named emotional states. For example, computer vision may track the movement of lip corners, the speed with which this occurs, and the length of time corners of the mouth are moved from their usual position. Nose wrinkling represents emotions such as disgust, and depressions of the corner of lips are connected with sadness. Of course, people have unique faces, they execute expressions in different ways, they are not always directly facing a camera and they appear in different contexts. The learning element takes place when computers are given many faces (potentially millions) and software is written in such a way to allow itself to be trained by incoming data. This is done

by software auto-adjusting its algorithms to improve detection of expressions in unfamiliar

circumstances and situations. These include naturalistic settings such as classrooms where

lighting might be poor, children may be wearing glasses, moving their heads, or they may be

positioned against complex backgrounds.

Commoditising emotion [B]

As noted above, a normative assessment should consider motives to deploy technology as

well as technological capability and social values. Undoubtedly most (if not all) working

directly in education will have a foremost interest in child wellbeing. However, connecting

with wider interests in labour-based critique of AI and companies that employ machine

learning (Crawford and Joler 2018), it should not be missed that inferences about student's

emotions are used to train the neural networks owned by the EdTech providers.

This is significant because both well-known and lesser-known companies use facial coding

for ends beyond education. All are thirsty for training data to improve their algorithmic

services in other business contexts. The result is that data about child emotions are

commoditised to improve the company product, create competitive difference (in terms of

how many faces are analysed) and serve business and strategic contexts for which the

classroom data was not intended (such as in-car analysis, retail, out-of-home advertising,

market research, user experience, legal depositions, media and entertainment, and more).

The business model is significant in that technology firms license their facial coding software

and the resultant data teaches the vendor's system to more accurately judge emotional

expressions. Applied to this paper's interest in education, this means that students have scope

to become a population resource to not only sense, analyse, mine and learn, but commodify.

This is because prowess developed from education and children may be applied in other

contexts. Again, echoing Crawford and Joler (2018), modern AI is rooted in, functions, and

grows by means of, exploitation of the human body. Indeed, although this paper has restricted

its interest to facial coding applications, more physically invasive applications are on the

socio-technical horizon-line. In addition to seeing the face (which is literally a surface of the

head), computer vision may also be used to gauge emotions through *remote* ocular

measurement of heart rate and heart rate variability in real-time (Gupta and McDuff 2016).

This is now being actively pursued at Microsoft through research on how camera image

sensors may measure variations in light absorption related to the delivery of the blood

volume pulse (Blackford et al. 2018).

A question of method [A]

There is then the question of whether these technologies may actually interpret emotions.

This is a multifaceted question, first because what emotions are is not settled (Lewis et al.

2010). A multitude of theories of emotion are summarized by Scherer (2010: 10-14) and

further paraphrased below:

- *Adaption*: an *evolutionary conception* where emotions prime an organism to deal with

    challenges to its well-being.

- *Dimensional*: conceived of emotions on a pleasant-unpleasant continuum (*valence*) and

    whether an organism is activated or remains passive (*arousal*)[1].

- *Appraisal*: recognizes that there are opportunities for many emotional types as conditions

    change and new situations emerge. Emotions here are *determined by conditions* and what

    the organism makes of a given situation. Some emotion theorists argue there are as many

---

[1] The valence/arousal approach to psychology originally belongs to Wundt (1897).

emotions as there are possible situations. Others argue that the multitude of possible emotional responses are fashioned out of more Basic Emotions.

- *Motivation*: a psychobiological approach by which emotions are derived from *primitive emotional states* (e.g. fear/danger = need for flight; anger = preparation for flight).

- *Circuit*: an approach to emotions strongly influenced by the neuroscientific and neuroanatomical outlook. This equates emotions with genetically *coded neural circuits*.

- *Discrete*: this points to the existence of *Basic Emotions* as the blocks with which more complex emotions are built. Each emotion is said to have its own physiological, expressive and behavioural patterns.

- *Lexical*: this involves working backwards from language and expression to *examine metaphors* we use to describe experiences. The analyst then organises metaphors to build structures of emotional domains and subdivisions of emotional types.

- *Social constructionist*: this is the socio-cultural outlook whereby *societal structures* are held as the main influence on emotional make-up of people. This sees emotions defined in terms of collective norms and meaning to reactions generated by the socio-cultural context in which they appear.

While the nature of emotions remains disputed, this paper opts towards an amalgamation of psychobiological and sociocultural components in line with dimensional and appraisal-based theorisation. On discrete and Basic emotion, these have long been challenged, especially by social constructionists. Seeing emotions in discursive terms, they are said to exist in relation to systems of expression and understanding that 'make sense' because we interpret them according to learned social rules (Fridlund 1991, Russell 1994, Barrett 2006, Gendron et al. 2014). This invites questions about the social construction of labels by which people categorise their core affective state and the systems of expression and signification connected

with these. Closely related, the leak-based view misses a great deal regarding the

communicative and interactionist role of emotional behaviour (Fridlund 1995, Leys 2012).

As a result, for constructionists, social labelling of affects and communicative aspects of

emotion places emotional expression within the domain of language and intentionality

(Fridlund 1991, Leys and Goldman 2010, Leys 2011). They argue that the ability to recognise

emotions is not prelinguistic, that emotions are a combination of bodily affects and the socio-

cultural labels we attach to these, and that they are not simply expressions because emotion

also has social and communicative functions. If compromise between Basic Emotions and

constructionist views is required, the constructionist argument can be sub-divided into strong

and weak varieties (Harré 1986, Jasper 1998). Strong constructionism argues that there are no

basic emotions, whereas weak constructionism (which this paper prefers) recognises the

existence of basic emotional behaviour but argues that this explain little. For example,

whereas anger, surprise and Basic Emotions of the Ekman sort are directly connected with

bodily states, more complex emotions such as jealousy, compassion or shame are more

closely associated with the social and cultural contexts in which they are experienced (Thoits

1989, Jasper 1998). Keenly aware of social constructionist approaches to emotion, in the

afterword to Darwin's (2009 [1972]) *The Expression of the Emotions in Man and Animals*,

Ekman directly attacks cultural anthropologist Margaret Mead's argument that social

behaviour is determined by culture. Instead Ekman argues that while expressions may be

filtered through culturally defined *display rules*, Basic Emotions exist and that these are

leaked through facial behaviour (Ekman 1977).

What appears to be clear however, especially given Barrett et al.'s (2019) recent work, is that

the Basic Emotions approach does not capture how people convey, or interpret, emotion on

faces. For example, a smile can express more than one emotion depending on the situation,

the individual or the culture. Moreover, given that 'similar configurations of facial

movements variably express instances of more than one emotion category' (ibid), what is

clear is that more detail on the context of the situation is required to understand the emotion.

Echoing the weak constructionist position of this paper, as Barrett et al. (2019) argues, this

should not be taken to mean that expressions are meaningless. While facial movement should

not be read as being correspondent with a named emotion, expressions are useful if

information about context is present. In other words, expressions are more contextually

contingent than the Basic view suggests. Lastly, while many commercial applications are

reliant on Basic approaches, we should not be misled into the over-simplistic belief that

everyone interested in affective computing and synthetic emotions would agree with the

Basic Emotions view. Many would argue that single labels do not always capture complex

emotional and affective behaviour, such as embarrassment (Baron-Cohen and Tead 2003,

Gunes and Pantic 2010). To end this section, the answer to the question posed of whether

technologies may actually interpret emotions is that Basic approaches are deeply flawed.

Contextual and appraisal-based approaches (in theory) have a much better chance, but this

requires a large-scale revision of approach to begin with local particulars (and potentially

self-annotation) over universal understandings.

Deterritorialising emotion [B]

In the late nineteenth century, the potential for seduction by data representations of emotion

was captured and unwittingly illustrated by Edgeworth, who poetically remarked:

From moment to moment the hedonimeter varies; the delicate index now flickering with

the flutter of the passions, now steadied by intellectual activity, low sunk whole hours in

the neighbourhood of zero, or momentarily springing up towards infinity. The continually

indicated height is registered by photographic or other frictionless apparatus upon a

uniformly moving vertical plane. Then the quantity of happiness between two epochs is

represented by the area contained between the zero-line, perpendiculars thereto at the

points corresponding to the epochs, and the curve traced by the index [...] (1881: 101).

What this poetic fetish shows is that the role of media technologies (gauges, meters,

photography, video and now computer vision and algorithms) to mediate emotion should not

be missed because it is a vital element in each wave of facial coding applications. In line with

Verbeek (2005) and Idhe (2002), this is to consider how technologies and scientific

instruments frame how we perceive the world and ourselves. It is also to ask about

'hermeneutic relations' (with hermeneutics being the theory and methodology of

interpretation) suggested by Idhe (1990). As Verbeek (2011: 8) points out, a thermometer

does not convey what is to be hot or cold, but it provides a representation of these states.

How then does face-based emotional AI filter and constrain what we ask, know and believe?

The beginning of the answer, as argued above, is in the limits of the technology.

Emotional AI based on computer vision and supervised machine learning that requires a

broadly universal and homogenous view of emotions (although this may allow for regional

variations in relation to the intensity of the expression of a given basic emotion). What it does

not *currently* allow for is heterogenous and context-specific accounts of emotion, as required

by constructivist critique. Next is the issue of children's expression, which raises a range of

problems for application of basic emotions and facial coding to children. Central is that

children are by definition at an earlier stage of emotional and communicative maturity.

Indeed, Barrett et al. (2019) state that, 'In young children, instances of the same emotion

category appear to be expressed with a variety of different muscle movements, and the same

muscle movements occur during instances of various emotion categories, and even during

nonemotional instances' (2019: 27). They point out that young children's facial movements

lack strong reliability and specificity, giving examples of fear. In FACS this involves wide-

eyed and gasping facial configuration, but this has rarely been reported in young infants.

Other differences between FACS for adults and young children's behaviour includes

expression of anger (infants do not reliably produce a scowling FACS facial configuration) or

even smiles. Again, they have not been reliably observed to smile when they are in visually

engaging or mastery situations, or even when they are in pleasant social interactions. When

the child was angry (such as when being refused an object they wanted or being made to do

something against their will), a variety of facial movements were observed, including frowns,

furrowed brows, and eyerolls. Further, expressions of anger were more often vocal than facial

(ibid: 23-27).

Also, simplicity when backed up by [big] data can be seductive and misleading. Affectiva is a

leading facial emotion recognition company that has shown a keen interest in education.

Their product, Affdex, uses cameras in personal devices or public spaces to capture facial

expressions of people as they view or interact with content or objects. Depending on the

capacity of the reading device, multiple faces can be scanned at over 20 processed frames-

per-second. Affdex employs FACS to assess the movement of 45 different facial muscles and

key feature points, such as the eyes and mouth. The system analyzes each pixel in the region

to describe the colour, texture, edges and gradients of the face. It maps these and classifies

arrangements of pixels into named emotions. The big data lure is the most prominent feature

of Affectiva's home page stating that (as of August 28th 2019) they have analysed 8,333,646

faces. When there is attractive visualisation backed by corporate clout and big data, does it

matter that the fundamentals of the method might be wrong?

As facial coding deterritorialises emotional life, this also raises questions about the ways in which these articulations and representations: a) are said to constitute students' emotional experience for learning and interaction; and b) how they may lead children and older students to internalise these articulations of their affective states which, to an extent, would deterritorialise existing understanding about emotions. This is echoed by Kennedy et al. (2016) whose study on visualisation and representation also identifies questions of power, knowledge and simulation in the design conventions of data visualisation. Similarly, in an educational context, Williamson (2018a) raises alarm about dashboards, visual analytics technologies, the ushering in of a hidden infrastructure of big data techniques, and the eventual market reform of education. Here, through overly simple emotional taxonomies and the social semiotics of data visualization, education professionals are working with an account of emotional life that appears quantifiable and true, but is actually established on convenient, commercial and highly debatable foundations. Ultimately this expedience is grammatical (Agre 1994) as emotional life is being reorganized around the limited units and measures that computer vision and emotional AI can process.

Operationalising SEL through AI [A]

Given a topic as innately sensitive as biometrics and emotions, it is all too easy to point to neo-behaviourist industrial technics of extraction, mining, behaviourism and social optimisation, rather than to assess emotional AI itself and how it alters the relationships between technology and society (Achterhuis 2001). However, as we have seen above, knowledge of empirics and specifics may also lead to ethical understanding. The same can be applied to education in that while there are identified harms, the goal of emotional AI enabled EdTech is not malevolent: it is to improve student learning by teaching software to identify

the emotional and affective states of children. The rationale is that academic skills need to be

complemented with social and emotional skills and related competencies. This sees a 'whole

child' approach that connects success with ability to pay attention, understand and manage

emotions, and work effectively in groups (Aspen Institute 2019). The appeal of emotional AI

is clear in that it facilitates tracking of expressions and, in theory, emotions, attention and

interpersonal behaviour. Applications of these emergent technologies should be seen in the

context of ongoing datafication of schoolchildren, behaviour and performance (Lupton and

Williamson 2017, Williamson 2018). Such "disruptive innovation" to education from start-

ups and legacy technology companies alike raises questions, not least how solutionist

discourses around learning casts the student as a 'user', which in turn entails language of

commodification and depoliticization of education (Ramiel 2017). Neither EdTech or public-

private tensions are new, progress through technology and business narratives has a history.

These include Edison's vision of film as an educational technology (1920s); Pressey's

"machine for testing intelligence" (1920s); and Skinner's mechanical and electronic

"teaching machines" that, following logics of operant conditioning, sought to speed-up

student feedback (Watters 2014, Selwyn 2017). Modern EdTech, including that which

employs emotional AI, remains a social force in which private interests play a prominent role

in public education. It is also a manifestation of technological solutionism whereby

questionable methods are used to solve structural problems of how schools are funded

(Morozov 2013, Williamson 2015, Wright and Peters 2017).

Such initiatives are some distance away from an enlightenment-based Humboldtian model of

education. Rather, they represent a preparatory stage for the labour market (Peters 2016).

This applies to soft as well as hard skills, especially because social and emotional skills

(especially those connected with creativity and resilience) are seen as critical to the

workforce of the future (Organization for Economic Co-operation and Development 2018).

An extended report by the World Economic Forum (WEF 2016), 'New Vision for Education:

Fostering Social and Emotional Learning through Technology' identifies a current lack of

capacity to quantify and measure social and emotional learning. The cites Affectiva's Affdex

as a 'specific high-potential example' of how face-based emotional AI is being deployed for

educational purposes (2016: 15). In an educational context it is used to gauge attention,

emotion and comprehension, to use data to augment teaching methods, and to allow

educators to modulate delivery and class schedules. The significance of the World Economic

Forum is that it connects business, government and other relevant stakeholders, and has

significant influence in creating and driving high-level agendas in education. Certainly, one

expects the World Economic Forum to emphasize and potentially champion the role of

business in education, but their pronouncements on issues involving personal data and

privacy is usually done with greater understanding of involved technologies.

Given ambiguity (at best) around facial coding and children, to proceed from weak scientific

foundations is problematic for evidence-based education. However, this uncertainty is not

present in influential pronouncements, as the World Economic Forum champions affective

computing to 'recognize, interpret and simulate human emotions' (2016: 15). Indeed, this

'New Vision for Education' is seemingly entirely unaware of debates regarding the nature of

emotions. Interest in social and emotional learning has created a business opportunity for AI

companies that use computer-sensing to infer emotional states. The vision is to progress

EdTech from being a means of making class-time more productive, to using technology to

enhance Social and Emotional Learning. This is the process through which children and

adults understand and manage emotions so as to enjoy personal wellbeing, empathise, and

establish and maintain positive relationships (Philibert 2017). The report cites affective

computing as an 'emerging set of innovations that allow systems to recognize, interpret and simulate human emotions', also stating that these show 'great promise for developing social and emotional skills such as greater empathy, improved self-awareness and stronger relationships' (WEF 2016: 15). Vendors of these applications promise to identify students in need of attention, to help diagnose the nature of the problem, predict future failures, identify risks and tailor learning materials.

In discussion of the marketisation of education and technologies that make use of psycho-physiological data about the brain, Williamson (2018) draws on Whitehead et al. (2018) to account for this in terms of 'neuroliberalism'. This helps ground numerous issues: marketisation of education, quantification, and understanding brain functioning to engineer neuroplasticity and shape behaviour. This is not idle critique, especially given that the World Economic Forum report states:

> Developers can embed SEL into traditional academic ed-tech products and reach large untapped markets. And developers can push the boundaries by integrating SEL into products at the forefront of technology. As they do, they are taking action to spread, reinforce and advance SEL. (2016: 15)

Beyond profit motives, other financial factors play a role. Schools around the world have relatively limited resources for teaching and professional development related to social and emotional learning. As such, one can see the appeal of technologies that promise intimacy-at-scale to solve a longstanding problem in education: how to provide personalised attention to packed classrooms. This is less about automation and replacing human teachers, but detection of emotions and affective states to tailor learning approaches. The problems, as detailed,

involve issues of method, coercion, choice, accuracy, material consequences, commodification of child data, chilling effects (feeling inhibited and excessively self-conscious), emotional self-governance, relationships with peers, un-intruded upon mental space, control over what to reveal, and what here is considered as prima facie right not to have emotions profiled. While children do not have adult autonomy rights, privacy still plays an important role. This is because excessive, inescapable and recorded observation of emotional behaviours does not align with the need for a significant degree of mental and emotional reserve to ensure human flourishing.

Although this paper focuses on emotional AI that senses and makes use of data about facial expressions, it should be seen in the context of wider psycho-physiological and biopolitical interest in augmentation of child behaviour in education (Williamson 2017). On facial coding, there are variety of start-ups such as Affectiva and SensorStar Labs that use cameras to capture students' emotional expressions to gauge attention and classify student engagement. A key recent development however is the increased interest from legacy technology companies. The education branch of the technology company Intel states that they are researching how recognition of affect may personalise learning experiences and provide 'adaptive learning' (Intel Education, 2014). Key factors for Intel are scope to recognize individual learners' expressions of excitement, frustration and boredom, so educators may modulate and optimise the difficultly levels of content. Intel also seek to track 'patterns of student motivation' and 'gauge emotional investment.' Microsoft Azure (2018) offers similar facial coding products which are marketed alongside their face identification and detection products. Beyond sensing and detection, users of their Application Programming Interface (API) also grant capacity to search, identify and match faces within databases of up to 1 million people. These may be segmented by features, similarities and behaviour. This

portends not only class and school-level analytics, but searchability across much larger

groups, potentially for wider social scoring purposes.

Are the technologies in tension with the UN's stance? [A]

This paper appreciates the 'multistability' of technology (Idhe, 2002) in that technologies are

not innately bad or good, but neither are they neutral. Their moral charge comes out of its

relationship with the context in which it is applied (and how this takes place). The nature of a

technology is not fixed but comes into being in relation to people and usage. While this paper

does not take a dystopian or utopian view of the technologies themselves (it sees scope for

positive usage of wider and emergent forms of emotional AI in human-computer interaction,

entertainment and even art), it is concerned about deployment in education.

Practically and legally, emotional AI raises questions about the role of third parties, sensitive

personal data in the form of biometric data, knowing where and how data is processed, and

allowance of data erasure, but it is Art. 5(1) of the European Union's General Data Protection

Directive on 'data minimisation' that has special relevance due to its intention to curtail over-

use of personal data through big data techniques. This provides guidance on a question that

underlies many parts of this paper: *Is emotional AI necessary for successful education?*

Specifically, Art. 5(1) states that lawful bases of using personal data requires 'data

minimisation' and that processing is 'necessary' for a specific purpose. Given that it is

reasonable to suggest that teachers can empathise and monitor class reactivity without

processing data about emotions, the lawful basis to process personal data about emotions is

doubtful. This doubt is amplified by debatable methodology, current effectiveness and scope

to inhibit human flourishing through surveillance.

Beyond Europe and other regions with an interest in maintaining "adequacy" with European data protection principles, the issue of human flourishing is key given its centrality to the *United Nations Convention on the Rights of the Child* (OHCHR 1990). As detailed earlier, Art. 27 and Art. 29 have special application given their interest in mental, spiritual, moral and social development. Other articles apply though, including the need to act in the child's best interests (Art. 3), the child's right to freedom of thought (Art. 14) and privacy (Art. 16), the right to develop to full potential (§1 Art. 29); the child's right to liberty (§2 Art. 29); and the child's right to be protected from economic exploitation (Art. 32).

However, Art.29 §1 of the Convention raises a counterargument to that being advanced in this paper, because it could be argued that applications of AI in the classroom will help and not hinder a child's ability to reach their 'fullest potential'. Not without merit, this argument would be based on accusations of ideological antagonism to market involvement in education and that luddism blocks scope for children's development and flourishing. Connected, Art.29 §2's interest in non-interference to 'establish and direct educational institutions' should also be seen as having scope to provide a normative defence for EdTech companies. Anticipated, the response to the charge is that: a) the science of emotion needs to be sound; and b) evidence needs to be provided *before* deployment that concerns about harms to children's flourishing argued in this paper are incorrect.

Although the applications detailed by EdTech providers and legacy companies are clearly at odds with the UN's stance, this paper's postphenomenological orientation (that involves a need for empirics and understanding of context) keeps it open to the role of emotional AI in education. This would not be FACS-based due to methodological concerns, but the idea of using emotion and affect-sensitive technologies is not entirely dismissed. Close attention to

child benefits, efficacy, methodology, privacy, data protection, scope to flourish and

exploitation would all be factors in a future assessment. These are criteria that would apply to

many emergent technologies and practices. A more novel issue stems from the fact that

schools, to an extent, play an 'in loco parentis' function.[2] To help understand whether the

public good is being served by emotional AI technologies, a key question is this: *Do schools*

that employ emotional AI companies behave as any reasonable parent would do in promoting

the welfare and safety of children in their care? This is to ask whether parents would

outsource development of their child's social and emotional skills to companies that use

resultant data to improve their algorithms to serve other clients, including those from

commercial and security sectors. Beyond the question of the commensurability of the private

interests of emotional AI companies with the public good, child and educational benefits,

methodology, privacy, data protection, scope to flourish and absence of exploitation, this

paper suggests the following considerations for any deployment of emotional AI by

educators, whether this be based on facial analytics or a different sort (see Figure 1):

Conclusion [A]

Focusing on face-based approaches that employ Basic Emotions methodology, the goal of

this paper has been to consider benefit and risks of emotional AI-based EdTech. As a prima

facie principle, the suggestion that technologies may assist teachers in their activity is not

problematic. Emotional AI is also not seen as innately unethical, although the paper is

mindful of that existing applications (such as in advertising and retail are taking place with

questionable consideration of ethical factors such as consent). Centrally, this paper did not

---

[2] In the UK, under the Children Act 1989, teachers have a duty of care towards their pupils. This is referred to as 'in loco parentis'. Legally, while not bound by parental responsibility, teachers must behave as any reasonable parent would do in promoting the welfare and safety of children in their care.

begin with the assumption that private interests and serving the public good are

incommensurable. However, it finds numerous methodological, legal and normative

problems with emotional AI-based EdTech, especially applications based on facial coding.

These are: 1) serious questions about effectiveness, validity and representativeness of training

data; 2) financial incentives and the wellbeing of school-children do not align in this

situation; 3) mining the emotional lives of children is normatively wrong, especially when the

value extraction does not serve the wellbeing of those children; 4) it is problematic to use

inferences about children's emotions to train neural networks deployed for other commercial

purposes (such as advertising); 5) scope for mission creep where in-class data may be used

for other socially determining purposes (such as social scoring); 6) the social desirability of

chilling effects in the classroom; 7) existence of data minimisation principles that should ask

if emotional AI is necessary for successful education; and, lastly, 8) contravention of the

creepiness principle. This paper concludes that, in the case of emotional AI and facial coding

in the classroom, these technologies are incommensurable with current and near future social

values. The test posed for any deployment of emotional AI by educators is conceived as a

good start, rather than the definitive test. Ideally it will be amended and improved through the

work of many hands to help guard the public good.

References

Achterhuis, H. 2001. *American Philosophy of Technology: The Empirical Turn*.

Bloomington: Indiana University Press.

Agre, P. 1994. "Surveillance and Capture: Two Models of Privacy." *The Information Society*

10: 2, 101-127. Doi: 10.1080/01972243.1994.9960162

Aspen Institute. 2019. *National Commission on Social, Emotional, and Academic*

Development, https://www.aspeninstitute.org/programs/national-commission-on-social-

emotional-and-academic-development.

Barrett, L.F. 2006. "Are Emotions Natural Kinds?" *Perspectives on Psychological Science* 1:

28–58. doi: 10.1111/j.1745-6916.2006.00003.x

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. 2019. "Emotional

Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial

Movements." *Psychological Science in the Public Interest*, 20(1), 1–

68. https://doi.org/10.1177/1529100619832930

Blackford, E.B., Estepp, J.R. and McDiff, D. 2018. "Remote Spectral Measurements of the

Blood Volume Pulse with Applications for Imaging Photoplethysmography." *Microsoft*

Research. doi: 10.1117/12.2291073

Boden, M.A. 2006. *Mind as Machine: A History of Cognitive Science*. New York: OUP.

Cawsey, A. 1998. *The Essence of Artificial Intelligence*. Harlow: Prentice Hall.

Crawford, K. and Joler, V. 2018. *Anatomy of an AI System: The Amazon Echo as an anatomical map of human labor, data and planetary resources*, https://anatomyof.ai.

Darwin, C. 2009 [1872]. *The Expression of the Emotions in Man and Animals*. London: Harper.

Duchenne de Boulogne, G.B. 1990 [1862]. *The Mechanism of Human Facial Expression*. Cambridge: Cambridge University Press.

Edgeworth, F.Y. 1881. *Mathematical Psychics*. London: Kegan Paul.

Ekman, P. and Friesen, W.V. 1971. "Constants Across Cultures in the Face and Emotion." *Journal of Personality and Social Psychology*. 17 (2): 124–29. doi: 10.1037/h0030377.

Ekman, P. 1977. "Biological and Cultural Contributions to Body and Facial Movement." In *The Anthropology of the Body*, edited by J. Blacking, 34-84. London: Academic Press. www.paulekman.com/wp-content/uploads/2013/07/Biological-And-Cultural-Contributions-To-Body-And-Facial-Mov.pdf.

Ekman, P. and Friesen, W.V. 1978. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto, CA: Consulting Psychologists Press.

Fridlund, A.J. 1991. "The sociality of solitary smiles: Effects of an implicit audience." *Journal of Personality and Social Psychology* 60: 229–40. doi: none.

Fridlund, A.J. (1995) *Human Facial Expression: An Evolutionary View*. San Diego, CA: Academic.

Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L.F. 2014. "Perceptions of Emotion from Facial Expressions Are Not Culturally Universal: Evidence from a Remote Culture". *Emotion*, 14, 251-262. doi: http://10.1037/a0036052

Goldman, A.I. 2008. *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. New York: Oxford University Press.

Gupta, O. and McDuff, D. 2016. *Real-Time Physiological Measurement and Visualization Using a Synchronized Multi-Camera System*, http://alumni.media.mit.edu/~djmcduff/assets/publications/Gupta_2016_Realtime.pdf.

Harré, R. 1986. *The Social Construction of Emotions*. London: Basil Blackwell.

Idhe, D. 2002. *Bodies in Technology*. Minneapolis: Minnesota Press.

Intel Education. 2014. *Emerging Technologies: Accelerants for Deep Learning*,

https://www.intel.com/content/dam/www/public/us/en/documents/training/emerging-technologies-lesson-plan.pdf.

Jasper, J. 1998. "The Emotions of Protest: Affective and Reactive Emotions in and Around Social Movements." *Sociological Forum* 13 (3): 397–424. doi: none.

Kennedy, H., Hill, R.L., Aiello, G. and Allen, W. 2016. "The Work That Visualisation Conventions Do". *Information, Communication & Society* 19 (6): 715-735. doi: http://10.1080/1369118X.2016.1153126

Lepper. M.R. and R.W. Chabay. 1988. "Socializing the Intelligent Tutor: Bringing Empathy to Computer Tutors." In *Learning Issues for Intelligent Tutoring Systems*, edited by H. Mandl and A. Lesgold, 242-257. New York: Springer-Verlag.

Lewis, M., Haviland-Jones, J.M. and Barrett, L.F. 2010. *Handbook of Emotions*. London: Guildford Press.

Leys, R. 2012. Both of Us Disgusted in My Insula': Mirror Neuron Theory and Emotional Empathy, *Nonsite.org*, http://nonsite.org/article/%E2%80%9Cboth-of-us-disgusted-in-my-insula%E2%80%9D-mirror-neuron-theory-and-emotional-empathy.

Leys, R. and Goldman, M. 2010. "Navigating the Genealogies of Trauma, Guilt, and Affect: An Interview with Ruth Leys." *University of Toronto Quarterly* 79 (2): 656–79. doi: http://10.3138/utq.79.2.656

Leys, R. 2011. "The Turn to Affect: A Critique." *Critical Inquiry* 37 (3): 434–72. doi: 10.1086/659353

Lupton, D. and Williamson, B. 2017. "The Datafied Child: The Dataveillance of Children and Implications for Their Rights." *New Media and Society* 19 (5), 780-794. doi: 10.1177/1461444816686328

McDuff, D. 2014. *Crowdsourcing Affective Responses for Predicting Media Effectiveness*, PhD Thesis, http://affect.media.mit.edu/pdfs/14.McDuff-Thesis.pdf.

Microsoft Azure. 2018. *Face*, https://azure.microsoft.com/en-gb/services/cognitive-services/face.

Morozov, E. 2013. *To Save Everything, Click Here: Technology, Solutionism and The Urge to Fix Problems That Don't Exist*. London: Penguin.

Office of the United Nations High Commissioner for Human Rights. 1990. *Convention on the Rights of the Child*, https://www.ohchr.org/Documents/ProfessionalInterest/crc.pdf.

Organization for Economic Co-operation and Development. 2018. *World Class How to Build a 21st-Century School System*, http://www.oecd.org/education/world-class-9789264300002-en.htm.

Peters, M. 2016. "Inside the Global Teaching Machine: Moocs, Academic Labour and the Future of The University." *Learning and Teaching* 9 (2): 66-88. doi: http://10.3167/latiss.2016.090204

Philibert, C.T. 2017. *Everyday SEL in Early Childhood: Integrating Social-Emotional Learning and Mindfulness into Your Classroom*. Abingdon: Routledge.

Picard, R., Cassell, J., Kort, B., Reilly, R., Bickmore, T., Kapoor, A., Mota, S. and Vaucelle, C. 2001. *Affective Learning Companion*, https://affect.media.mit.edu/projectpages/lc/.

Picard, R., Kort, B. and Reilly, R. 2001a. *Exploring the Role of Emotion in Propelling the SMET Learning Process*, https://affect.media.mit.edu/projectpages/lc/nsf1.PDF.

Ramiel, H. 2017. "User or Student: Constructing the Subject in EdTech Incubator." *Discourse: Studies in the Cultural Politics of Education* 1-13. doi: 10.1080/01596306.2017.1365694

Rorty, R. (2010 [1993]) 'Human Rights, Rationality, and Sentimentality' in C.J. Voparil and R.J. Bernstein, eds. *The Rorty Reader*. Chichester: Wiley-Blackwell. Pp. 351-365.

Russell, J.A. 1994. "Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies." *Psychological Bulletin* 115 (1): 102–41. doi: 10.1037/0033-2909.115.1.102

Schauer, F. 1978. "Fear, risk and the first amendment: unraveling the chilling effect." *Boston University Law Review* 58/5: 685–732. doi: none.

Selwyn, N. 2017. *Education and Technology: Key Issues and Debates, 2nd ed*. London: Bloomsbury.

Szeliski, R. 2010. *Computer Vision: Algorithms and Applications*. Dordrecht: Springer.

Tene, O. and Polonetsky, J. 2014. "A Theory of Creepy: Technology, Privacy, and Shifting Social Norms." *Yale Journal of Law and Technology* 16 (1): 59-102. doi: none.

Thoits, P.A. 1989. "The Sociology of Emotions." *Annual Review of Sociology* 15: 317–42. doi: http://10.1146/annurev.so.15.080189.001533

Tomkins, S.S. 1962. *Affect, Imagery, Consciousness: Vol. 1. The Positive Affects*. New York: Springer.

Tomkins, S.S. 1963. *Affect, Imagery, Consciousness. Vol. 2: The Negative Affects*. New York: Springer.

Warwick, K. 2012. *Artificial Intelligence: The Basics*. Oxon: Routledge.

Watters, A. (2014) *Teaching Machines: A Brief History of "Teaching at Scale"* *#t509Massive*, http://hackeducation.com/2014/09/10/teaching-machines-teaching-at-scale.

WEF. 2016. *New Vision for Education: Fostering Social and Emotional Learning through Technology*, http://www3.weforum.org/docs/WEF_New_Vision_for_Education.pdf.

Whitehead, M., Jones, R., Lilley, R., Pykett, J., & Howell, R. 2018. *Neuroliberalism: behavioural government in the twenty-first century*. London: Routledge.

Williamson, B. 2015. "Silicon startup schools: technocracy, algorithmic imaginaries and venture philanthropy in corporate education reform." *Critical Studies in Education* 59 (2): 218-236. doi: 10.1080/17508487.2016.1186710

Williamson, B. 2018. "Brain Data: Scanning, Scraping and Sculpting the Plastic Learning Brain Through Neurotechnology." *Postdigital Science and Education* 1-22. doi: http://10.1007/s42438-018-0008-5

Williamson, B. .2018. "The Hidden Architecture of Higher Education: Building a Big Data Infrastructure for the 'Smarter University." *International Journal of Educational Technology in Higher Education* 15: 1-26. doi: http://10.1186/s41239-018-0094-1

Wright, N., & Peters, M. 2017. "Sell, Sell, Sell or Learn, Learn, Learn? The Edtech Market in New Zealand's Education System – Privatisation by Stealth?" *Open Review of Educational Research* 4 (1): 164-176. doi: none.

Wundt, W. 1902. *Principles of Physiological Psychology*. London: Wilhelm Englemmann.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Zelenski, J.M. and Larsen, R.J. (2000) The Distribution of Basic Emotions in Everyday Life: A State and Trait Perspective From Experience Sampling Data, *Journal of Research in Personality*, 34 (2): 178–97.

Figure 1 Considerations for deployment of emotional AI by educators

Concern	Issue
Scientific scrutiny	Does the technology do what its proponents claim? Has the data been trained on a suitably diverse dataset? Has independent scientific scrutiny been solicited?
Universality	Proponents of facial coding are adamant that Basic Emotions are universal, but what of: a) ethnocentric considerations; b) representative training data; c) location variation in emoting; d) individual-level variation in affective reactions and emoting?
Methodological expediency	Emotions are articulated in a way that makes sense to machines. Has consideration been given to the risk of creating an understanding of emotional life that suits data analytics, but not people?
Relationships with self	Has due consideration been given to how these technologies may impact on students' own emotional understanding?
Mental health	Are policies in place if the technology detects ill mental health and how will these be balanced with privacy and data protection concerns?
Disavowing self-reporting	Will students be believed if they complain that they *were* paying attention?
Students and parents: may they opt-out without penalty?	Power asymmetries and lack of choice of schools conspire to make meaningful consent impossible. What is the policy and rationale if parents or children do not want to be subject to sensing by emotional AI?
Trust	Has the trusting nature of people and their willingness to engage with non-human actors been considered?

Review	What processes and checks are in place to review the implementation of the emotional AI application? Have suitable independent experts been solicited and are they available to assist where answers to this test are unclear?