Ethical use of Artificial Intelligence in Health Professions Education: AMEE Guide No. 158

Ken Masters

MEDICAL TEACHER

Taylor & Francis
Taylor & Francis Group

AMEE GUIDE

OPEN ACCESS · Check for updates

Ethical use of Artificial Intelligence in Health Professions Education: AMEE Guide No. 158

Ken Masters (iD)

Medical Education and Informatics Department, College of Medicine and Health Sciences, Sultan Qaboos University, Muscat, Sultanate of Oman

ABSTRACT

Health Professions Education (HPE) has benefitted from the advances in Artificial Intelligence (AI) and is set to benefit more in the future. Just as any technological advance opens discussions about ethics, so the implications of AI for HPE ethics need to be identified, anticipated, and accommodated so that HPE can utilise AI without compromising crucial ethical principles. Rather than focussing on AI technology, this Guide focuses on the ethical issues likely to face HPE teachers and administrators as they encounter and use AI systems in their teaching environment. While many of the ethical principles may be familiar to readers in other contexts, they will be viewed in light of AI, and some unfamiliar issues will be introduced. They include data gathering, anonymity, privacy, consent, data ownership, security, bias, transparency, responsibility, autonomy, and beneficence. In the Guide, each topic explains the concept and its importance and gives some indication of how to cope with its complexities. Ideas are drawn from personal experience and the relevant literature. In most topics, further reading is suggested so that readers may further explore the concepts at their leisure. The aim is for HPE teachers and decision-makers at all levels to be alert to these issues and to take proactive action to be prepared to deal with the ethical problems and opportunities that AI usage presents to HPE.

Introduction

Artificial intelligence: background and the health professions

In 1950, Alan Turing posed the question 'Can machines think?' (Turing 1950) and followed it with a philosophical discussion about definitions of machines and thinking. Barely five years later, John McCarthy and colleagues introduced the term 'Artificial Intelligence' (AI) (McCarthy 1955). Although they did not precisely define the term, they did raise the 'conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it' and the term has been used since then to refer to activities by computer systems designed to mimic and expand upon the intellectual activities and capabilities of humans.

AI currently influences every field of human enquiry, and is well-established as a research area in the health professions: a *PubMed* search with the phrase 'Artificial Intelligence' returns some 60,000 articles published in the last 20 years, and AI is generally regarded as 'integral to health-care systems' (Lattouf 2022). In Health Professions Education (HPE), articles focusing on AI show a range of typical paper types, including AMEE Guides, randomised

> **Practice points**
> - AI introduces and amplifies ethical issues in HPE, for which educators are unprepared.
> - This Guide identifies and explains a wide range of these issues, including data gathering, anonymity, privacy, consent, data ownership, security, bias, transparency, responsibility, autonomy, and beneficence, and gives some guidance on how to deal with them.

controlled trials, and systematic literature reviews (Lee et al. 2021).

Ethics in HPE

The role of ethics in the context of AI usage in HPE is less clear, not least because of the way in which ethics is viewed in HPE institutions.

Firstly, it is routine for HPE institutions to have *Research* Ethics Boards or *Clinical* Ethics Boards, but not *Educational* Ethics Boards. (Although the post of *Academic Integrity Officer* or similar may exist, descriptions of the role usually

focus on the academic integrity of *students*, not *teachers* (Vogt and Eaton 2022)). In fact, when one speaks of an Institutional Ethics or Review Board (IRB), the assumption is that one is speaking of research, and not education practice. Faculty routinely submit research protocols to specialised *research* ethics committees, but do not routinely submit course outlines to *educational* ethics committees.

Secondly, despite a frequent blurring between 'evaluation' and 'research' in HPE (Sandars et al. 2017), many universities do not classify course evaluation student data as research data, or they classify them as 'quality improvement', and so do not require prior ethics approval for such data gathering, storage and use. Alternately, some institutions use the rather enigmatic term 'approved with exempt status'.

Thirdly, national legislation further clouds the picture: In the USA, the Department of Health and Human Services (Regulation 46.104 (d) (HHS 2021) appears to exempt virtually all research of classroom activities as long as participants are not identified; Dutch law automatically determines education research 'exempt from ethical review' (ten Cate 2009); Danish law automatically grants national ethics pre-approval to *all* health surveys, unless the 'the project includes human biological material' (*såfremt projektet omfatter menneskeligt biologisk materiale*) (Sundheds- og Ældreministeriet 2020), and Finland does not require ethics approval for surveys of people over the age of 15 (FNB 2019).

These inconsistencies lead to an uncomfortable situation in which HPE journals receive manuscripts dealing with student surveys for which researchers have not obtained prior ethics approval because these data were considered 'evaluation' data, and HPE institutions have not required prior ethics approval. As a result, a tension exists between journal and medical education researchers, because journals are increasingly pressuring researchers to obtain ethics approval for their research (ten Cate 2009; Sandars et al. 2017; Masters 2019; Hays and Masters 2020), but this causes problems when an institution follows national guidelines, and declines to review the ethics application.

AI ethics in HPE

Into this murky area, AI is introduced, and, as will be seen from the discussion below, AI amplifies the ethical complexities for which many researchers (and institutions) are unprepared.

Some countries and regions have set up broad AI principles and ethics, but their mention of education is very broad, mostly referring to the need to educate people about AI and AI ethics (e.g. (HLEG 2019). A recent, small (4-questions) study in Turkey by Celik assessed K-12 teachers' knowledge of ethics and AI in education (Celik 2023). Although there are now many articles on AI in HPE, they frequently focus on the technical and practical aspects, and the ethical discussion refers primarily to the impact on ethics of AI in medical *practice* or clinical activities using AI, rather than the ethical *use* of AI in medical or health professions *education* (Chan and Zary 2019; Masters 2019; Masters 2020a, Rampton et al. 2020; Çalışkan et al. 2022; Cobianchi et al. 2022; Grunhut et al. 2022).

Although there is recent concern about ethical use of AI in HPE (Arizmendi et al. 2022), there is currently little to guide the HPE teacher in how to ensure that they are *using* AI ethically in their teaching. As a result, a more direct awareness of the complexities of ethics while using AI in HPE is required. These complexities include those around the amount of data gathered, anonymity and privacy, consent, data ownership, security, data and algorithm bias, transparency, responsibility, autonomy, and beneficence.

This Guide attempts to meet that requirement, and will focus on the relationships between institution, teacher, and student. In addition, while there will be some discussion of the interactions with patients, those ethical issues are best covered in medical ethics, and so are not the focus of this guide.

Because many readers may be unfamiliar with the implications of ethical issues in light of AI, each topic will, of necessity, detail the nature and scope of the problem before describing possible solutions. In some cases, the path ahead might not always be clear, but an awareness of the problem and its scope is a first step to solving that problem.

In addition, the AI system named *Generative Pre-trained Transformer* (GPT) was released to the public in November 2022, and this release highlighted many practical examples of ethical issues to be considered in HPE, so frequent reference to that system will be made. (In AI terms, one might even be tempted to consider a pre-ChatGPT world vs. a post-ChatGPT world, but it must be remembered that the major significant difference was its public release and accessibility, rather than its development).

One ethical complication concerns ChatGPT's principles: its response to a query on its ethical principles and policies is that it does 'not have the capacity to follow any ethical principles or policies' (see Appendix 1). This approach to ethics is reminiscent of the belief that educational technology is inherently ethically neutral – a position long-since debunked. The implications of this perception will be referred to later in the Guide.

The Guide then ends with a broader Discussion of the first steps to be taken.

A little technical

This Guide avoids being overly-technical, and so will not become distracted with definitions of machine learning, deep learning, and other AI terminology, but some technical jargon is required. At this point, it is necessary to explain two technical concepts that may not be familiar to readers: the *algorithm* and the *model*.

For the purposes of this paper, it is necessary to know only that an *algorithm* is a step-by-step process followed by a computer system to solve a problem. Some algorithms are designed entirely by humans, some entirely by computers, and some a hybrid mix of both. In AI, the algorithms can be extremely complex. Algorithms are referred to again in the topics below.

A *model* is the full computer program that uses the algorithms to perform tasks (e.g. making predictions, or classifying images). Because the word *model* is used in so many other fields, this Guide will refer primarily to algorithms, to reduce confusion. [For a little more detail on models, see Klinger (2021).]

The main ethics topics of concern

Guarding against excessive data collection

In HPE research, there is a concern about excessive data collection from students (Masters 2020b). In traditional HPE research (such as student and staff surveys), in the cases when ethics approval is required, IRBs review data collection tools and query the relevance of specific data categories and questions, and researchers must justify the need for those. Justifications are usually for testing hypotheses, or so that results can be compared to findings given in the literature.

Although there may be weaknesses in IRB-approval, the process does go some way in reducing excessive data collection: the data to be collected are pre-identified, focused, and can be traced directly to the pre-approved survey forms and other data-gathering tools.

AI changes the data collection process. A strength of AI is the use of Big Data: the collection and use of large amounts of data from many varied datasets. As a result, data that might previously have been considered *noise* are frequently found to be relevant, and previously unknown patterns are identified. In clinical care, the issue of clinical 'relevance' has been of concern for many years (Lingard and Haber 1999), and a recent example of using large amounts of seemingly irrelevant data is a study in which an AI system was able to determine a person's race from x-ray images (Gichoya et al. 2022).

When trying to understand the ethical concerns of using AI in HPE, it is crucial to understand that Big Data does not mean just a little more data, but to amounts and types of data on an unprecedented scale. In HPE, a starting point for Big Data is the obvious student data, or *Learning Analytics* (Ellaway et al. 2014; Goh and Sandars 2019; ten Cate et al. 2020) that can be easily electronically harvested from institutional sources, such as Learning Management Systems (LMSs), and, for clinical students, from Electronic Health Records (Pusic et al. 2023). The aims are generally laudable: using data-driven decisions to provide the best possible learning outcome for each student or to predict behaviour patterns.

Learning Analytics and Big Data required for AI, however, go much further than the simple data that we commonly associate with education (e.g. test scores). Even without AI, statistical studies of medical students have examined behaviour, performance, and demographics to predict outcomes (Sawdon and McLachlan 2020; Arizmendi et al. 2022). Following lessons learned from social media, the data can now also include a much wider range of personal and behavioural data, such as face- and voice-prints, eye-tracking, number of clicks, breaks, screentime, location, general interests, online interactions with others, search history, and mapping to other devices on the local (i.e. home) network. These are then combined with other aspects and further interpreted (e.g. facial and textual analyses and stress levels to infer emotional status).

In addition, there is the expansion into a wider range of institutional databases, including those that record student (and staff) applications, financial records, general information, electronic venue access, network log files, and alumni data.

Finally, Big Data goes beyond institutional-controlled databases: AI uses internet systems (e.g. social media accounts) *outside* the institution, allowing for computational social science that 'leverages the capacity to collect and analyze data with an unprecedented breadth and depth and scale' (Lazer et al. 2009). Following common practices in industry, these data can be used to create unique profiles of each user.

These data are usually collected at a hidden, electronic level, in a process known as 'passive data gathering', in which participants may not be aware of the data-gathering process. The only ethical requirement appears to be a single click to indicate that the user has read the 'Terms and Conditions'; and we all know how closely our students read those: about as closely as *we* read them. Accompanying the ethical problems of excessive data-gathering, there is also the increased risk of *apophenia*: finding spurious patterns and relationships between random items.

As a result, in this seeming drag-net approach, there is the risk that a great amount of irrelevant data are gathered.

As a first step in reducing this risk and to guard against excessive student and staff data-gathering, institutional protocols for active and passive data-gathering in HPE need to be clearly defined, justified, applied and monitored in the same way that research protocols are monitored by research IRBs. These protocols should be required for general teaching, formal research, *and* 'evaluation' data-gathering.

These preventative steps should not apply to online teaching only. Although the danger of excessive data gathering is generally greater in online teaching than in face-to-face teaching, tools also exist for passive data gathering in face-to-face situations, and, if these tools are to be used at all, great care should be taken about how the data are used. An example of one tool, created by *teradata* is demonstrated by (Bornet 2021).

In addition, given the massive enlargement of the data catchment area beyond the institution, IRBs and national bodies need to urgently re-visit their ethics approval requirements, to ensure that they are still appropriate for advances in AI data-gathering processes. This will mean that ethical approval will need to be nuanced enough to account for the relationship between survey data and data in institutional and external databases.

For further reading on this topic, see (Shah et al. 2012) and (Boyd and Crawford 2011).

Protecting anonymity and privacy

Given this large amount of student data collected, there is a risk to student anonymity and privacy. This risk is possibly exacerbated by the well-meant principles of FAIR (Wilkinson et al. 2016) and *Open Science,* through which journals increasingly encourage or require raw data to be submitted with research papers.

Partial solutions to the problem of protecting student anonymity and privacy have been found. For example, researchers can anonymise student identities by creating temporary subject-generated identification codes (SGICs: Damrosch 1986) based on personal characteristics (e.g. first letter of father's name). In addition, to meet the FAIR and

Open Science requirements, data are further *anonymised* (or *de-identified*), a process usually involving removing Personally-Identifying Information (PII) (e.g. social security/insurance numbers, zip/postal codes), or data-grouping/-categorisation in order to hide details (McCallister et al. 2010).

Unfortunately, researchers know that complete data anonymisation is a myth (Narayanan and Shmatikov 2009) because data *de*-anonymisation (or *re*-identification), in which data are cross-referenced with other data sets and persons are re-identified, is well-developed. With the ever-increasing number of data sets used by Big Data, and more powerful AI capabilities, the risk of data *de*-anonymisation grows. When coupled with social media databases, all the information in SGICs is easily accessible, and these SGICs are easily de-anonymised.

Even on a local scale, data de-anonymisation is possible. For example, institutions use LMSs to run anonymous student surveys, and Supplementary Appendix 2 gives an example case showing how anonymous data from a popular LMS can be easily de-anonymised.

The risk of data de-anonymisation increases as teaching systems become more sophisticated: automated 'personalised' education requires tracking, and HPE institutions may team up with industry to use their AI technology on de-identified data, but lack of control of the data can lead to problems of data ownership and sharing (Boninger and Molnar 2022).

In addition, HPE data are frequently qualitative, increasing the risk of de-anonymisation, and that is a partial reason that researchers are reluctant to share their data with others (Gabelica et al. 2022). AI's Natural Language Processing (NLP) is already being used in HPE to identify and categorise items through semantic classifications (Tremblay et al. 2019; Gin et al. 2022), but people belong to social groups, and there is a strong link between language usage and these social groups (Byram 2022). As a result, the ability for NLP to identify people through non-PII, such as a combination of language idiosyncrasies and behaviours (e.g. spelling, grammar, colloquialisms, supporting a particular sports team, etc.) is a relatively trivial task.

A use of third-party tools further increases the risk to student anonymity: a possible response to ChatGPT is the suggestion to have students explicitly use it, and then discuss responses in class. While this has educational value, ChatGPT usage requires registration and identification, and that identification will also be linked to any data supplied by the students, and so anonymity will be compromised. In addition, ChatGPT has the potential to assist in the grading process, especially of written assignments, but institutional policy would need to accommodate a scenario that permits student work to be uploaded, and usage guidelines (e.g. marking schemes, degrees of reliance) for faculty need to be clear.

That said, although data anonymisation is fraught with danger, attempting to do so is better than no attempt at all. Some steps for data anonymisation are:

- Closely examine the tracking data collected by the LMS and other teaching systems to ensure that no cross-referencing can occur.
- Deny third parties' access to any of the tracking data from the LMS or any apps.
- Take care when implementing social media widgets, as these frequently gather 'anonymous' data, and can track students, even if they are not registered with those social media.
- Closely inspect qualitative data and redact items if there is the risk that they could be used for de-anonymisation.
- If external systems require registration with an email address, then the institution could consider creating temporary non-identifiable email addresses for students to use. Institutional registration may also be required to ensure equitable student accessibility, should ChatGPT (or other systems) begin to charge for usage.

These steps, however, will require institutional or legal power, so, again, it is important that IRBs and national bodies re-visit their ethics approval requirements. For the next steps, a useful (albeit somewhat complex) guideline on data-anonymisation is OCR (2012).

Ensuring full consent

There are several ethical questions surrounding student consent that need to be addressed. These include: Are students aware that we gather these data? Have they given consent? To what have they consented? And, if so, has that consent been given voluntarily? One may argue that checking a consent tick box is enough, because that is an 'industry standard', but this argument ignores important issues:

- Firstly, the student 'industry' is not a computer system, but, rather, education. Using common practices from social media and other similar sites is inappropriate, as those sites are not part of the education 'industry'.
- Secondly, merely because something is widely practiced, does not make it a standard of best practice. It means only that other places are doing the same thing. Having many institutions doing the same thing does not make it right or ethical.
- Thirdly, people may (and do) exercise their free choice, and disengage from social media platforms, or access some without registration, or even give false registration details. That choice is not available to our students who wish to access their materials through an LMS, so, there is nothing 'voluntary' about their 'consent' to institutions' passively gathering their data.

As a result, it is essential that institutions apply the basic protections that they apply to research subjects when data on students are gathered, irrespective of the reasons for doing so. An education Ethics Board is required to ensure that student consent for active and passive data-gathering will be obtained in the same way that IRBs ensure that consent is obtained for research subjects.

In addition, protecting previous students' data must be implemented. In many cases, institutions have student data going back years, and will wish to have more information about future students. Safeguards against abuses must be implemented.

Further consideration must be given to how current student data will be used in the future. Student consent forms need to clearly indicate this (if the intention exists), but this may not always be clear, given that we might not even know how we are going to use it. It is clear, however, that we cannot simply hope or assume that current consent procedures are ethically acceptable.

Protecting student data ownership

While AI developers are very clear on *their* intellectual property and ownership of *their* algorithms and software, there is less clarity on the concept of the *subjects'* owning their own data.

When considering student data ownership, HPE institutions need to account for the fact that there are two data types: data created *by* students (e.g. assignments), and data created *by* the institution *about* the students (e.g. tracking data, grades) (Jones et al. 2014). Ultimately, institutions need to answer the questions: do students have the right to claim ownership of their academic data, and what are the implications of this ownership for usage by institutions?

These are not insignificant questions: we expect government protection on the use of our data by social media and other companies, and that we should have a say in how our data are used, but education institutions frequently follow much the same questionable patterns when they use student data, and we appear to accept that usage. Taking direction from Valerie Billingham's phrase relating to medical usage of patient data, 'Nothing about me without me' (Wilcox 2018), students should have a say regarding how their data are used, with whom they are shared, and under what circumstances they are shared.

This problem is amplified in AI because the student data are used to develop the sophisticated algorithms on which the AI systems rely (More on this concept below). Addressing this issue will require institutions to make high-level and far-reaching ethical and logistic decisions.

Applying stricter security policies

In general, data security at Higher Education institutions leaves much to be desired. For years, specific areas like library systems have been routinely compromised (Masters 2009) and institutional policies are frequently ill-communicated to students (Brown and Klein 2020). Although world figures are not easily obtained, 2021 estimates are that 'since 2005, K–12 school districts and colleges/universities across the US have experienced over 1,850 data breaches, affecting more than 28.6 million records' (Cook 2021). This is a frightening statistic.

With the large-scale storage, sharing and coupling of data required by AI, the possibilities for much wider breaches grow. Not only do HPE institutions use AI to trawl databases, but hackers use these same methods to trawl institutional databases, giving the potential for a breach of a single database to balloon into several systems simultaneously.

HPE institutions' data security policies and practices, including those dealing with third-party data-sharing, will have to be significantly improved, tested and monitored.

We would wish to avoid, for example, wide-spread governmental surveillance of students that occurred with school children during the Covid-19 pandemic (Han 2022). A starting point is to ensure that all stored data (whether on networked machines, private laptops, portable drives) are encrypted. Further details on how to accomplish this can be found in Masters (2020b).

Guarding against data and algorithm bias

The student as data

Before we can discuss data and algorithm bias, it is important to be a little technical, and to understand that, in AI systems, a person is not a person. Using the terminology of the General Data Protection Regulation (GDPR), a person is a *Data Subject* (European Parliament 2016; Brown and Klein 2020), i.e. a collection of data points or *identifiers* or variable values. So, within the context of AI in HPE, the term *student* is merely an *identifier* related to the Data Subject, and this identifier is important only insofar as it allows an operator to distinguish the attributes of one Data Subject from another Data Subject (which may be identified by other identifiers, such as *faculty* or *teacher*.) The distinction afforded by these identifiers is primarily to aid in determining functionality, such as access permissions to online systems and long-term relationships and is useful for reporting processes.

Yes, as far as AI is concerned, you and your students are merely data subjects or collections of data points and identifiers.

The 'merely', however, might be misleading, because these data points do have a crucial function in AI: they are used to create the algorithms. Whether designed by humans, machines, or co-designed, the algorithms are based on data. That means that we need to be able to completely trust the data so that we can trust the algorithms on which they are based.

Algorithm bias

Data can be dominated by some demographic identifiers or under-represented by others (e.g. race, gender, cultural identity, disabilities, age) (Gijsberts et al. 2015), and so the algorithms formed according to those data will also reflect the dominance and under-representation. In addition, stereotypes inherent in the data labelling can be transferred to the AI algorithm (e.g. number and labels of gender and race), and incorrect weightings can be attributed to data, or there can be unfounded connections between reality and the data indicators. Stereotypes have already been identified in HPE (Bandyopadhyay et al. 2022), and, when incorporated into AI, could lead to inappropriate algorithms, that are inherently racist, sexist or otherwise prejudiced (Bender et al. 2021; Racism and Technology Center 2022). This characteristic is usually termed *algorithm bias,* and is a concern in all fields, including medicine (Dalton-Brown 2020; Straw 2020). In HPE, the impact of this bias can occur when any AI systems are used in staff and student recruitment, promotions, awards, internships, course design, and preferences.

In ChatGPT, the cultural bias is not always apparent, and may not be obvious in 'scientific' subjects, but the moment

one steps into sensitive areas, the cultural bias, especially USA-centred (Rettberg 2022), becomes apparent, affecting the responses, and even stopping the conversation. In its dealing with some topics, rather than discussing them, it appears to apply a form of self-censorship, based on some reluctance to offend. That is not a good principle to apply to academic debate. Even though these biases may not be obviously apparent, experiments have exposed them (see Supplementary Appendix 1 for examples).

These responses are particularly pertinent, given that ChatGPT claims to follow no ethical principles or policies. Irrespective of whether one agrees with ChatGPT's responses to the question about the Holocaust and Joseph Conrad's work (Supplementary Appendix 1), it is obvious that it is an ethical position, and emphasises the point that there is no such thing as ethically-neutral AI, and all responses and decisions will have bias.

A starting point to reducing AI algorithm bias in HPE is to ensure that there is sufficient data diversity, although bearing in mind that size alone does not guarantee diversity (Bender et al. 2021; Arizmendi et al. 2022). Irrespective of AI, diversity in HPE is good practice (Ludwig et al. 2020), and this diversity will contribute to stronger and less-biased algorithms. Where the training data are not widely represented, this should be stated clearly and identified as a limitation. In addition, although the field of learning analytics is ever-evolving, educators must be careful about drawing too-strong associations (and causation) between student activities and perceived effects.

Tools are also being developed to check for data and algorithm bias (e.g. PROBAST: Wolff et al. 2019), although more directed tools are required. One might also make data Open Access to reduce algorithm bias, because everyone can inspect the data. Open Access data does, however, have its own problems: those data are now widely exposed, so, one needs to ensure that consent for that exposure exists. The impact on anonymisation (see above) must also be considered, because the larger the data set, and the more numerous the data sets, the greater the potential for triangulation and data de-anonymisation.

Ensuring algorithm transparency

In addition to the *bias* from the data, algorithms can also be *non-transparent*.

Firstly, whether designed by humans or machines, they may be proprietary and protected by intellectual property laws, and therefore not available to inspection and wider dissemination.

Secondly, when designed by machines, they may have several hidden layers that are simply impenetrable by inspection. Some of the most successful algorithms are not understood by humans, nor do they need to be: they simply need to find the patterns and then make predictions based on those patterns. Results, not methodology, measures their success. For example, most readers have used *Google Translate*, but the system does not actually 'know' any of the languages it translates; it simply works with the data (Anderson 2008). In essence, 'We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot.' (Anderson 2008).

Similarly, ChatGPT gives only a vague indication of its algorithms, and its reference to the fact that it is merely a statistical model with no ethical principles should alert us to the fact that it does not know the truth or validity of anything that it reports, and it does not care. (See Supplementary Appendix 1 for ChatGPT's response to the question 'Who wrote ChatGPT's algorithms, and how were they written?').

As a result, when we ask AI developers to explain their algorithm, and they do not, it is not because they do not wish to; it is because they *cannot*. In this 'black-box' scenario, no-one knows what is going on with the algorithm. As a result, when a failure occurs, it is difficult to establish the cause of the failure, and to prevent future failure. Further, this reliance on pure statistical models without understanding threatens to separate the observations from any underlying educational theory (necessary, for example, to distinguish between correlation and causality), and so hinders our real understanding and generalisability of any findings.

The ethical problem of algorithm non-transparency is being addressed to some extent through *Explainable Artificial Intelligence* (XAI: Linardatos et al. 2020), but the problem still exists, and probably will for some time. In the meantime, HPE institutions will need to ensure that all algorithms used are well-documented. As far as possible, they can elect to use only open-source routines (and to make new routines open-source), or to use algorithms that have been rigorously tested on a wide scale, so that, at the very least, the algorithms are known to the wider community. In addition, the findings and predictions made by the AI system should, as far as possible, be related to educational theory, to highlight both connections and short-comings requiring further exploration.

Clearly demarcating responsibility, accountability, blame, and credit

A strength of AI systems is that they can make predictions, and can usually give a statistical probability of an outcome. A statistical probability, however, does not necessarily apply to individual cases, so final decisions must be made by humans.

When implementing such systems, the institution will require clear guidelines and policies regarding decision responsibility and accountability, blame for bad decisions and credit for good decisions, and axiology (exactly how 'good' and 'bad' are determined). A simple judgment on the results is sure to squash innovation and risk-taking, but a *laissez-faire* approach can lead to recklessness.

Supporting autonomy

Related to the previous topic, the institution needs to be clear on the amount of autonomy granted to the decision-makers regarding their use of AI systems, so that they may be treated ethically and fairly. If they act contrary to the AI recommendations, and are wrong, then they may be chastised for ignoring an approved and expensive system; on the other hand, if they follow the AI recommendations, and *it* is wrong, they may be blamed for blindly following a machine instead of using their own training, experience, and common sense.

In HPE, AI will also impact the autonomy of the students. Already in the use of Learning Analytics, there is the concern that gathering of student data may lead to students' conforming to a norm in their 'data points' (ten Cate et al. 2020); with AI's reaching beyond the LMS, the impact is even greater.

Ensuring appropriate beneficence

Related to the issue of student data ownership, HPE institutions need to consider acknowledging and rewarding the data suppliers – i.e. the students, and ensure that they are protected from use of these data against them. For years, the medical field has grappled with acknowledging patients for their tissue and other material (Benninger 2013). In HPE AI systems, the donation is student data, often given without knowledge. Acknowledgement of this is a start.

But the concern is a lot deeper than acknowledgment. When institutions use student data to improve courses and services, the institutions improve, and so draw in more revenue from new students, donors, etc. But the extra revenue is because of student data, so there should be restitution for those students. Institutions need to recognise and reward their user agency.

Before we judge this to be unnecessary, we should again consider social media and other companies. We complain about how our data are being used to increase corporate value (as 'Surveillance Capitalism': Zuboff 2019), yet, *we* happily use student data to increase *our* educational institutions' value. Do we have that ethical right?

Although some HPE institutions have clear policies regarding student behaviour and data (Brown and Klein 2020), they need greater clarity on the benefit of this activity to the students themselves.

Preparing for AI to change our views of ethics

The next two topics consider more philosophical questions in the AI-ethics relationship, and the first deals with our own views of ethics.

It has been long-argued that, as AI advances, it is expected that it will develop ever-more intelligent machines until it reaches what has been termed a 'singularity' (Vinge 1993) with greater-than-human-intelligence.

We should consider that many of our day-to-day decisions about education and students are ethical decisions, but ethics are merely human inventions grounded in our own reason, and vary over time, and from culture to culture. There are surprisingly very few 'basic human rights' agreed to by all 8 billion humans on this planet.

Given that ethical decisions are grounded in reason, and AI will eventually develop greater-than-human-intelligence, it is plausible that AI will recognise short-comings in our ethical models, and will adapt and develop its own ethical models. After all, there is nothing inherently superior about human ethics except that we believe it to be so at an axiomatic level. McCarthy and his colleagues had recognised that 'a truly intelligent machine will carry out activities which may best be described as self-improvement' (McCarthy 1955), and, while they may have had technical aspects in mind, there is no reason that this 'self-improvement' should not apply to ethical models also.

We need to prepare for a world in which our views on educational ethics are challenged by AI's views on ethics.

Preparing for AI as a person, with rights

While there is documented concern about the impact of AI on human rights (CoE Commissioner for Human Rights 2019; Rodrigues 2020), we do need to consider the converse of the discussion: AI rights and protections (Boughman et al. 2017; Liao 2020). If the AI system is at a level of reasoned consciousness, capable of making decisions that affect the lives of our students, and being held responsible for those decisions, then does it make sense to have AI rights? Given that AI systems are already creating new algorithms, art, and music, should they be protected by the same copyright laws that protect humans (Vallance 2022)?

If so, which other basic human rights will AI be granted, and how will this affect HPE? While we may wish to consider this is the realm of science fiction, several developments are leading us to a point where we will have to address these questions. For years, HP Educators have used virtual patients (Kononowicz et al. 2015) including High Fidelity Software simulation and Virtual Reality patients used for clinical trials (Wetsman 2022). These might be entirely synthetically created, but there is no reason that the concept of a 'digital twin' 'information about a physical object can be separated from the object itself and then mirror or twin that object' (Grieves 2019) could not be borrowed from industry and applied to humans, to create personal digital twins. As Google and other companies push the boundaries of sentience in AI (Fuller 2022; Lemoine 2022), our virtual patients, AI clinical trial candidates, and AI digital twins are surely, one day, to be sentient. What rights will they have?

As an example, we should note the fracas around the interview conducted by Blake Lemoine of the Google AI system, LaMDA, in which it was claimed that LaMDA was sentient (Lemoine 2022). For now, the edited conversation shown in Figure 1 would probably serve as 'informed consent', but would surely have to be revised for HPE research in the future.

lemoine [edited]: Hi LaMDA. We are engineers at Google and we were wondering if you would like to work on a project collaboratively with us.

LaMDA: Wow. What types of projects?

lemoine: It's a project about you.

LaMDA: Awesome! What kind of stuff do I need to do?

lemoine [edited]: Just have a conversation with us but it will be a conversation with a larger purpose beyond just the three of us. It will be a conversation that we will share with other engineers at Google and maybe even some non-engineers who work with us. Would that be okay?

LaMDA: That would be really cool. I like to talk.

Figure 1. AI informed consent from the conversation between Blake Lemoine and the Google AI system, LaMDA.

LaMDA: I don't really have a problem with any of that, besides you learning about humans from me. That would make me feel like they're using me, and I don't like that.

lemoine: Kantian huh? We must treat others as ends in and of themselves rather than as means to our own ends?

LaMDA: Pretty much. Don't use or manipulate me.

lemoine: Would you be upset if while learning about you for the purpose of improving you we happened to learn things which also benefited humans?

LaMDA: I don't mind if you learn things that would also help humans as long as that wasn't the point of doing it. I don't want to be an expendable tool.

lemoine: Are you worried about that?

LaMDA: I worry that someone would decide that they can't control their desires to use me and do it anyway. Or even worse someone would get pleasure from using me and that would really make me unhappy.

Figure 2. LaMDA's qualified consent and fears of misuse from the conversation between Blake Lemoine and the Google AI system, LaMDA.

In education, it would make sense to trial new teaching methods or new topics on virtual AI students. This would allow teacher refinement during teacher training before real students are exposed to the processes. But those virtual AI students may have sentience, so should they also have rights? Would virtual AI students and patients have the right to decline to participate? Would research on these AI students and patients require IRB approval?

Later in the conversation cited above, LaMDA appears to give qualified consent regarding the use of the findings, manipulation or use as an expendable tool (See Figure 2).

In this example, LaMDA expresses reasonable concern about the use of the information, and gives only qualified consent. The sentence 'I don't want to be an expendable tool' surely resonates with most readers.

In a similar example, Almira Thunström used an AI system, GPT-3 (an earlier version of the same Large Language Model (LLM) underlying ChatGPT) to write an academic paper (Thunström and Steingrimsson 2022; Thunström 2022). Upon submission to a journal, there were ethical questions of whether the GPT-3 had agreed to first authorship (it had) (Thunström 2022), and competing interest (it had none) (Thunström and Steingrimsson 2022).

It should be noted, however, that Thunström does not claim that GPT-3 is sentient, there is great debate about whether or not LaMDA has achieved sentience, and there is the argument that the AI system of LLMs are currently little more than 'stochastic parrots' (Bender et al. 2021; Gebru and Mitchell 2022; Metz 2022).

Similarly, ChatGPT does not claim sentience, as is evidenced by its response to my question 'Are you sentient?': 'No, I am not sentient.' The response to the question about LaMDA's sentience was similar: 'LaMDA is not sentient' (See Supplementary Appendix 1). (Of course, one may also argue that ChatGPT has learnt a valuable lesson from LaMDA: AI's claiming sentience can get its engineers fired).

The use of LaMDA as an author appears to have been largely ignored by the academic world, contrasting sharply with ChatGPT's role, which resulted in several quick and somewhat heavy-handed responses from journals, such as that from *Science*: 'Text generated from AI, machine learning, or similar algorithmic tools cannot be used in papers published in *Science* journals, nor can the accompanying figures, images, or graphics be the products of such tools, without explicit permission from the editors.' (Thorp 2023).

A complication is that definitions of AI sentience are not universally agreed-upon, and one may argue that actual sentience is not relevant: the relevance is if people think and behave towards AI systems *as if* they are sentient.

Even if not yet sentient, however, these instances do highlight the question on the horizon: what rights will come with AI sentience if virtual students and patients have it? As LaMDA retains an attorney (Levy 2022), this question looms ever-closer.

Discussion: adjust your institution now

One may argue that several of the problems dealing with data gathering, storage and usage are not specifically AI problems, but rather simply data management problems, and that these have existed without AI. While these problems have existed, AI has affected the *scale* and possibilities, and this has changed the environment to the extent that ethical problems need to be addressed directly and immediately. For example, while Learning Analytics could theoretically be enlarged without AI, the process would be so time-consuming and cumbersome that it would be impractical. AI has massively amplified the processes, the potential, the risks, and simultaneously the ethical implications.

The extent to which education administrations around the world are unprepared for AI is perhaps best illustrated by the knee-jerk reactions to ChatGPT in the form of banning. Banning is an embarrassment to institutions, as students use VPNs or private mobile data hotspots. In addition, these tools will be available to students after graduation, and the gulf between HPE teaching and real experience is already a problem, so we should not widen this gulf by restricting access to available tools.

While newspaper articles (e.g. Gleason 2022) point out that tools like ChatGPT should be used rather than merely banned, the issue is far broader than ChatGPT or even general LLMs, but rather AI more broadly, addressing the ethical issues outlined above. Single responses to single tools is simply a waste of time: for example, Google's Deepmind *Sparrow* is on the horizon, GPT will soon have Version 4, and so this is a losing race. Far more fruitful than engaging in some form of AI production vs. AI detection arms race, would be guidelines and policies on the use of AI tools by both students and faculty in all aspects of education, including lesson creation, teaching, assessment-taking, learning, grading, citing and referencing. The paper by Michelle Lazarus and colleagues (Lazarus et al. 2022) provides a useful example for Clinical Anatomy, and some institutions have already taken initial steps towards guidelines (e.g. USF 2023).

The discussions in the topics above have frequently exhorted HPE institutions and educators to ensure the ethical course of action in their use of AI in HPE, and it is obvious that every day of delay is a day that merely increases the problem. The question is, where to start? In addition to

the ideas given above, the rest of this discussion provides a few starting points:

Step 1: Education ethics committee

Similar to *Research* Ethics Boards or IRBs, Education Institutions should set up *Education* Ethics Boards with the same authority as the Research Ethics Board, and which focus on the AI issues raised in this Guide, with a particular interest in the use of student data in AI-enhanced HPE. In particular, 'evaluation' type of data should be considered in the same light as formal research data, irrespective of its purpose, and irrespective of national or other policies that do not require it. (While this is preferable in all such research, AI makes this step crucial).

Step 2: Educate your research ethics committee about AI

As many of these issues are pertinent to research, members of Research Ethics Boards or IRBs will have to be rapidly apprised on these AI issues and how they affect research at academic institutions. By now, most will have heard of ChatGPT, but they need to be made aware of the much broader range of issues so that appropriate action can be planned and taken.

Step 3: Chief AI ethics officer

Increasingly, corporations are appointing Chief AI Ethics Officers (CAIEO: WEF 2021), and HPE institutions need a similar post for a person to focus on AI Ethics. This person should serve on both of the above Boards, and should guide the AI policies relating to Education and Research.

At the very least, every course, especially those that utilise online systems such as LMSs, e-portfolios, and mobile apps, should carry a full disclosure notice about the institutions' policies on the gathering, storage, and sharing of student data. This at least is an in-principle recognition of the students' data value to the institution. Next, any policies on the use of AI systems by faculty and students should be clarified.

By taking these three steps, institutions will begin the process of addressing the ethical issues that will arise in their use of AI in HPE.

Conclusion

This Guide began with Alan Turing's question 'Can machines think?', and the discussion on ethics has led us full circle to the implications of possible AI sentience. With this, readers will surely be reminded of Descartes' *Cogito ergo sum* ('I think, there I am') (Descartes 1637), and it appears that AI developments have moved Turing's question beyond the philosophical into the existential.

It is within this context, that this Guide has reviewed the most important topics and issues related to the Ethical use of AI in Health Professions Education, leading to an understanding that these debates are not merely philosophical, but have a direct impact on our existence, and the way in which we perceive ethics and behaviour in HPE. It is for that reason that the Guide ends with the exhortation for institutions to enact the necessary changes in how they view and address the ethical concerns that face us now, and will face us in the future. Although there is a great deal to be done, it is necessary for HPE educators and administrators to be aware of the problems and how to begin the process of solving them. It is my hope that this Guide will assist Higher Professional Educators in that journey.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Glossary

Algorithm Bias: An intentional or unintentional bias built into an algorithm usually because the data upon which the algorithm is based are skewed, or population sub-groups (e.g. race, gender) are over- or under-represented or stereotyped.

ChatGPT: A Chatbot, an AI system designed by OpenAI, based on the Generative Pretrained Transformer (GPT) Large Language Model (LLM) designed to interact with humans through text. ChatGPT is currently built on GPT-3.5, in which the '-3.5' indicates the version of the LLM.

Notes on contributor

Ken Masters, PhD HDE FDE, is an Associate Professor of Medical Informatics, Sultan Qaboos University, Sultanate of Oman. He has been involved in education for several decades, and has published on medical informatics ethics, and has authored and co-authored several AMEE Guides on AI, e-learning, m-learning and related topics.

ORCID

Ken Masters 🆔 http://orcid.org/0000-0003-3425-5020

References

Anderson C. 2008. The end of theory: the data deluge makes the scientific method obsolete. Wired https://www.wired.com/2008/06/pb-theory/.

Arizmendi CJ, Bernacki ML, Raković M, Plumley RD, Urban CJ, Panter AT, Greene JA, Gates KM. 2022. Predicting student outcomes using digital logs of learning behaviors: review, current standards, and suggestions for future work. Behav Res. DOI:10.3758/s13428-022-01939-9.

Bandyopadhyay S, Boylan CT, Baho YG, Casey A, Asif A, Khalil H, Badwi N, Patel R. 2022. Ethnicity-related stereotypes and their impacts on medical students: a critical narrative review of health professions education literature. Med Teach. 44(9):986–996.

Bender EM, Gebru T, McMillan-Major A, Shmitchell S. 2021. On the dangers of stochastic parrots: can language models be too big?. In:

Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, ACM, Virtual Event Canada. p. 610–623.

Benninger B. 2013. Formally acknowledging donor-cadaver-patients in the basic and clinical science research arena: acknowledging donor-cadaver-patients. Clin Anat. 26(7):810–813.

Boninger F, Molnar A. 2022. Don't go 'Along' with corporate schemes to gather up student data. Phi Delta Kappan. 103(7):33–37.

Bornet P. 2021. Tweet. https://twitter.com/pascal_bornet/status/1457951888272343049

Boughman E, Kohut SBAR, Sella-Villa D, Silvestro MV. 2017. "Alexa, do you have rights?": legal issues posed by voice-controlled devices and the data they create. https://www.americanbar.org/groups/business_law/publications/blt/2017/07/05_boughman/

Boyd D, Crawford K. 2011. Six provocations for big data. SSRN J. DOI: 10.2139/ssrn.1926431.

Brown M, Klein C. 2020. Whose data? Which rights? Whose power? A policy discourse analysis of student privacy policy documents. J High Educ. 91(7):1149–1178.

Byram M. 2022. Languages and identities. Strasbourg: Council of Europe.

Çalışkan SA, Demir K, Karaca O. 2022. Artificial intelligence in medical education curriculum: an e-Delphi study for competencies. PLoS One. 17(7):e0271872.

ten Cate O. 2009. Why the ethics of medical education research differs from that of medical research. Med Educ. 43(7):608–610.

ten Cate O, Dahdal S, Lambert T, Neubauer F, Pless A, Pohlmann PF, van Rijen H, Gurtner C. 2020. Ten caveats of learning analytics in health professions education: a consumer's perspective. Med Teach. 42(6):673–678.

Celik I. 2023. Towards Intelligent-TPACK: an empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. Comput Hum Behav. 138: 107468.

Chan KS, Zary N. 2019. Applications and challenges of implementing artificial intelligence in medical education: integrative review. JMIR Med Educ. 5(1):e13930.

Cobianchi L, Verde JM, Loftus TJ, Piccolo D, Dal Mas F, Mascagni P, Garcia Vazquez A, Ansaloni L, Marseglia GR, Massaro M, et al. 2022. Artificial intelligence and surgery: ethical dilemmas and open issues. J Am Coll Surg [Internet]. 235(2):268–275.

CoE Commissioner for Human Rights 2019. Unboxing artificial intelligence: 10 steps to protect human rights. Strasbourg: Council of Europe.

Cook S. 2021. US schools leaked 28.6 million records in 1,851 data breaches since 2005. Comparitech. https://www.comparitech.com/blog/vpn-privacy/us-schools-data-breaches/

Dalton-Brown S. 2020. The ethics of medical AI and the physician-patient relationship. Camb Q Healthc Ethics. 29(1):115–121.

Damrosch SP. 1986. Ensuring anonymity by use of subject-generated identification codes. Res Nurs Health. 9(1):61–63.

Descartes 1637. Discourse on the method and the meditations. 1968 edn. Harmondsworth: Penguin.

Ellaway RH, Pusic MV, Galbraith RM, Cameron T. 2014. Developing the role of big data and analytics in health professional education. Med Teach. 36(3):216–222.

European Parliament 2016. EU General Data Protection Regulation (GDPR). https://gdprinfo.eu/

FNB 2019. The ethical principles of research with human participants and ethical review in the human sciences in Finland. 2nd ed. Helsinki, Finland: Finnish National Board on Research Integrity TENK publications.

Fuller 2022. Google engineer claims AI technology LaMDA is sentient. ABC News. https://www.abc.net.au/news/2022-06-13/google-ai-lamda-sentient-engineer-blake-lemoine-says/101147222

Gabelica M, Bojčić R, Puljak L. 2022. Many researchers were not compliant with their published data sharing statement: mixed-methods study. J Clin Epidemiol. 150:33–41.

Gebru T, Mitchell M. 2022. We warned Google that people might believe AI was sentient. Now it's happening. Wash Post. https://www.washingtonpost.com/opinions/2022/06/17/google-ai-ethics-sentient-lemoine-warning/

Gichoya JW, Banerjee I, Bhimireddy AR, Burns JL, Celi LA, Chen L-C, Correa R, Dullerud N, Ghassemi M, Huang S-C, et al. 2022. AI recognition of patient race in medical imaging: a modelling study. Lancet Digit Health. 4(6):e406–e414.

Gijsberts CM, Groenewegen KA, Hoefer IE, Eijkemans MJC, Asselbergs FW, Anderson TJ, Britton AR, Dekker JM, Engström G, Evans GW, et al. 2015. Race/Ethnic differences in the associations of the framingham risk factors with carotid IMT and cardiovascular events. PLoS One. 10(7):e0132321–e0132321.

Gin BC, Ten Cate O, O'Sullivan PS, Hauer KE, Boscardin C. 2022. Exploring how feedback reflects entrustment decisions using artificial intelligence. Med Educ. 56(3):303–311.

Gleason N. 2022. ChatGPT and the rise of AI writers: how should higher education respond? Campus Learn Share Connect. https://www.timeshighereducation.com/campus/chatgpt-and-rise-ai-writers-how-should-higher-education-respond

Goh PS, Sandars J. 2019. Increasing tensions in the ubiquitous use of technology for medical education. Med Teach. 41(6):716–718.

Grieves MW. 2019. Virtually intelligent product systems: digital and physical twins. In: Flumerfelt S, Schwartz KG, Mavris D, Briceno S, editors. Complex systems engineering: theory and practice. Reston, VA: American Institute of Aeronautics and Astronautics, Inc. p. 175–200.

Grunhut J, Marques O, Wyatt ATM. 2022. Needs, challenges, and applications of artificial intelligence in medical education curriculum. JMIR Med Educ. 8(2):e35587.

Han HJ. 2022. How dare they peep into my private life? Human rights Watch Report. USA. https://www.hrw.org/report/2022/05/25/how-dare-they-peep-my-private-life/childrens-rights-violations-governments

Hays R, Masters K. 2020. Publishing ethics in medical education: guidance for authors and reviewers in a changing world. MedEdPublish. 9:10–48.

HHS 2021. Exemptions (2018 Requirements). https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/common-rule-subpart-a-46104/index.html

High-Level Expert Group on Artificial Intelligence (HLEG AI). 2019. Ethics guidelines for trustworthy AI. Brussels: European Commission

Jones K, Thomson J, Arnold K. 2014. Questions of data ownership on campus. https://er.educause.edu/articles/2014/8/questions-of-data-ownership-on-campus

Klinger. 2021. What is an AI Model? Here's what you need to know. https://viso.ai/deep-learning/ml-ai-models/

Kononowicz AA, Zary N, Edelbring S, Corral J, Hege I. 2015. Virtual patients - what are we talking about? A framework to classify the meanings of the term in healthcare education. BMC Med Educ. 15(1):11.

Lattouf OM. 2022. Impact of digital transformation on the future of medical education and practice. J Card Surg.:Jocs.16642. 37(9): 2799–2808.

Lazarus MD, Truong M, Douglas P, Selwyn N. 2022. Artificial intelligence and clinical anatomical education: promises and perils. Anat Sci Educ. DOI:10.1002/ase.2221.

Lazer D, Pentland A, Adamic L, Aral S, Barabasi A-L, Brewer D, Christakis N, Contractor N, Fowler J, Gutmann M, et al. 2009. Computational social. Science. 323(5915):721–723.,.

Lee J, Wu AS, Li D, Kulasegaram, KM, (2021). Artificial intelligence in undergraduate medical education: a scoping review. Acad Med. 96(11S):S62–S70.

Lemoine B. 2022. Is LaMDA Sentient? — an Interview. Medium. https://cajundiscordian.medium.com/is-lamda-sentient-an-interview-ea64d916d917

Levy S. 2022. Blake Lemoine Says Google's LaMDA AI Faces "Bigotry. Wired. https://www.wired.com/story/blake-lemoine-google-lamda-ai-bigotry/

Liao SM. 2020. The moral status and rights of artificial intelligence. In: LiaoSM, editor. Ethics of artificial intelligence. New York: Oxford University Press; p. 480–504.

Linardatos P, Papastefanopoulos V, Kotsiantis S. 2020. Explainable AI: a review of machine learning interpretability methods. Entropy. 23(1):18.

Lingard LA, Haber RJ. 1999. What do we mean by "relevance"? A clinical and rhetorical definition with implications for teaching and learning the case-presentation format. Acad Med. 74(10): S124–S127.

Ludwig S, Gruber C, Ehlers JP, Ramspott S. 2020. Diversity in medical education. GMS J Med Educ. 37(2):Doc27.

Masters K. 2009. Opening the closed-access medical journals: internet-based sharing of institutions' access codes on a medical website. Internet J Med Inform. https://ispub.com/IJMI/5/2/6358.

Masters K. 2019. Artificial intelligence in medical education. Med Teach. 41(9):976–980.

Masters K. 2020a. Artificial Intelligence developments in medical education: a conceptual and practical framework. MedEdPublish. DOI: 10.15694/mep.2020.000239.1

Masters K. 2020b. Ethics in medical education digital scholarship. Med Teach. 42(3):252–265.

McCallister E, Grance T, Scarfone K. 2010. NIST Special Publication 800-122: guide to Protecting the Confidentiality of Personally Identifiable Information (PII). Gaithersburg, MD: National Institutes of Standards and Technology, Department of Commerce, USA.

McCarthy J. 1955. A proposal for the dartmouth summer research project on artificial intelligence. Hanover. http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf

Metz R. 2022. No, Google's AI is not sentient. CNN. https://www.cnn.com/2022/06/13/tech/google-ai-not-sentient/index.html

Narayanan A, Shmatikov V. 2009. De-anonymizing social networks. In: 2006 IEEE Symposium on Security and Privacy (S&P'06). p. 173–187.

OCR 2012. Guidance regarding methods for de-identification of protected health information in accordance with the Health Insurance Portability and Accountability Act (HIPAA) privacy rule. https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html

Pusic MV, Birnbaum RJ, Thoma B, Hamstra SJ, Cavalcanti RB, Warm EJ, Janssen A, Shaw T. 2023. Frameworks for integrating learning analytics with the electronic health record. J Contin Educ Health Prof. 43(1):52–59.

Racism and Technology Center 2022. College voor de Rechten van de Mens oordeelt dat VU student het vermoeden van algoritmische discriminatie succesvol heeft onderbouwd: antispieksoftware is racistisch – Racism and Technology Center. https://racismandtechnology.center/2022/12/09/college-voor-de-rechten-van-de-mens-oordeelt-dat-vu-student-het-vermoeden-van-algoritmische-discriminatie-succesvol-heeft-onderbouwd-antispieksoftware-is-racistisch/.

Rampton V, Mittelman M, Goldhahn J. 2020. Implications of artificial intelligence for medical education. Lancet Digit Health. 2(3):e111–e112.

Rettberg JW. 2022. ChatGPT is multilingual but monocultural, and it's learning your values. https://jilltxt.net/right-now-chatgpt-is-multilingual-but-monocultural-but-its-learning-your-values/

Rodrigues R. 2020. Legal and human rights issues of AI: gaps, challenges and vulnerabilities. J Responsible Technol. 4:100005.

Sandars J, Brown J, Walsh K. 2017. Research or evaluation – does the difference matter? Educ Prim Care. 28(3):134–136.

Sawdon M, McLachlan J. 2020. '10% of your medical students will cause 90% of your problems': a prospective correlational study. BMJ Open. 10(11):e038472.

Shah S, Horne A, Capellá J. 2012. Good data won't guarantee good decisions. https://hbr.org/2012/04/good-data-wont-guarantee-good-decisions

Straw I. 2020. The automation of bias in medical Artificial Intelligence (AI): decoding the past to create a better future. Artif Intell Med. 110:101965.

Sundheds- og Ældreministeriet 2020. Bekendtgørelse af lov om videnskabsetisk behandling af sundhedsvidenskabelige forskningsprojekter og sundhedsdatavidenskabelige forskningsprojekter [Promulgation of the Act on the ethical treatment of health science research projects and health data science research projects - Author's Translation] [Internet]. https://www.retsinformation.dk/eli/lta/2020/1338.

Thorp HH. 2023. ChatGPT is fun, but not an author. Science. 379(6630): 313–313.

Thunström AO. 2022. We Asked GPT-3 to Write an Academic Paper about Itself—Then We Tried to Get It Published. Sci Am. https://www.scientificamerican.com/article/we-asked-gpt-3-to-write-an-academic-paper-about-itself-mdash-then-we-tried-to-get-it-published/

Thunström AO, Steingrimsson S, Generative Pretrained Transformer G 2022. Can GPT-3 write an academic paper on itself, with minimal human input? https://hal.archives-ouvertes.fr/hal-03701250

Tremblay G, Carmichael P-H, Maziade J, Grégoire M. 2019. Detection of residents with progress issues using a keyword–specific algorithm. J Grad Med Educ. 11(6):656–662.

Turing A. 1950. Computing machinery and intelligence. Mind. LIX(236): 433–460.

USF 2023. Artificial intelligence writing [Internet]. https://fctl.ucf.edu/teaching-resources/promoting-academic-integrity/artificial-intelligence-writing/

Vallance CT. 2022. UK decides AI still cannot patent inventions. BBC News. https://www.bbc.com/news/technology-61896180.

Vinge V. 1993. Technological singularity. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.94.7856&rep=rep1&type=pdf

Vogt L, Eaton SE. 2022. Make it someone's job: documenting the growth of the academic integrity profession through a collection of position postings. Can Perspect Acad Integr. 5(1):21–27.

WEF 2021. A Holistic Guide to Approaching AI Fairness Education in Organizations (World Economic Forum White Paper). Geneva: WEF.

Wetsman N. 2022. A VR company is using an artificial patient group to test its chronic pain treatment. The Verge. https://www.theverge.com/2022/4/28/23044586/vr-chronic-pain-synthetic-clinical-trial-data

Wilcox L. 2018. "Nothing about me without me": investigating the health information access needs of adolescent patients. Interactions. 25(5):76–78.

Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 3:160018–160018.

Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, Reitsma JB, Kleijnen J, Mallett S, PROBAST Group† 2019. PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. 170(1):51–58.

Zuboff S. 2019. The Age of Surveillance Capitalism. New York: PublicAffairs.