Forensic Anthropology Vol. 7, Nos. 2–3: 164–176
DOI: 10.5744/fa.2023.0030

RESEARCH ARTICLE

We Think There's Been a Glitch: Artificial Intelligence and Machine Learning in Forensic Anthropology

Micayla C. Spiros[a,*] • Sherry Nakhaeizadeh[b,c]

ABSTRACT: Forensic anthropology plays a pivotal role in the medicolegal system analyzing skeletal remains in forensic casework. With the increased incorporation of artificial intelligence (AI) and machine learning (ML), their continued application in the field promises enhanced efficiency and accuracy in identification processes, trauma analyses, and decision-making. However, the integration of AI/ML also raises critical ethical concerns. The use of any technology involving human skeletal remains poses challenges related to privacy, consent, and the potential for dehumanization. Furthermore, the risk of bias within algorithms cannot be ignored—specifically, the inadvertent perpetuation of existing biases in training data, which can lead to incorrect identifications or skewed results. Addressing these concerns requires a balanced approach. Transparency in AI/ML design, clear guidelines for data collection and integration, and ongoing evaluation of algorithmic performance are essential. Additionally, interdisciplinary collaboration between anthropologists, ethicists, computer scientists, and legal experts would be a crucial step forward to ensure that the benefits of AI/ML are maximized while capturing the accountability and responsibilities within a legal and ethical context. Only through such an integrated approach can the potential of AI/ML in forensic anthropology be utilized as a tool responsibly, preserving the dignity of the deceased.

KEYWORDS: artificial intelligence; machine learning; bioethics; decision-making

Introduction

Technological developments have significantly advanced with substantial improvement in the way people perform different tasks (Frey & Osborne 2017). Machine learning (ML) and artificial intelligence (AI) in today's world are progressing rapidly with new advanced innovation impacting our everyday decisions—text prediction (Devlin et al. 2018), social media algorithms (Medvedev 2019), recommendations from Netflix (Lamkhede & Kofler 2021), personalized music playlists on Spotify (Jebara 2019), or even our e-mail's spam folder (Aski & Sourati 2016). It is evident that ML and AI are predicted to grow pervasively with technological advancements. Therefore, it is imperative to discuss the ramifications of how we build and interact with these technologies that will be the next step forward.

Within the criminal justice system, the use of ML and AI in assessment, detection, and prevention of crime is becoming increasingly technologically sophisticated (Galante et al. 2023; Završnik 2020). In forensic science, research has shown how ML, particularly AI and deep learning (DL), could assist with classifying bloodstain spatter (Liu et al. 2020), footwear impressions (Budka et al. 2021), maximizing accuracy in digital forensics (Jarrett & Choo 2021), predicting precision in forensic odontology (Khanagar et al. 2021), aiding in age estimation in forensic radiology (Li et al. 2019), and assisting with identification in the biometric fields (Kaur et al. 2020). While current literature has shown that it is possible to utilize the approaches, there is also a notion that the accuracy and performance of the algorithms currently used are relatively complex with varying degrees of explanation accuracy (Lo et al. 2023; Rudin 2019). The need for transparent and explainable models is akin to the call for transparency in forensic science in terms of understanding the role of the expert and their decision-making and the inferences and conclusions drawn from the evidence (Earwaker et al. 2020; Smit et al. 2016). Even though we have seen the positive attributes of ML and AI across various disciplines, universally we need to constantly proceed with calculated caution when building and applying these systems within a forensic context.

In forensic anthropology, the utilization of ML algorithms has been applied across biological profile analyses (Bertsatos et al. 2020; Darmawan et al. 2015; Hefner &

[a]Department of Anthropology, Michigan State University, East Lansing, MI, USA
[b]Department of Security and Crime Science, University College London, London, UK
[c]Centre for the Forensic Sciences, University College London, London, UK
*Correspondence to: Micayla Spiros, Michigan State University, Department of Anthropology, 655 Auditorium Drive, East Lansing, MI 48824
E-mail: spirosmi@msu.edu

Received 13 September 2023; Revised 15 December 2023; Accepted 19 January 2024

Ousley 2014; Langley et al. 2018; Nikita & Nikitas 2020; Spiros & Hefner 2020; Toneva et al. 2021). More recently, studies have begun to focus on the application of AI and DL in classification decisions through imaging (Bewes et al. 2019; Cao et al. 2022; Ortega et al. 2021), although the exploration of ML and AI began over 20 years ago in forensic anthropology. In 2001, McBride and colleagues used a decision tree algorithm to estimate sex of an individual from the os coxae. Konigsberg and colleagues (2002) responded with commentary on the importance of the expert and the need for probabilistic algorithms rather than deterministic methods. They argue that the methods used in the original article did not exceed the experts, but they did take into consideration that humans lack the computational power of processing the often massive amounts of data needed. As such, they acknowledged that computers can be programmed quite effectively in probabilistic statements and even highlighted the possible future of computing and ML in forensic anthropology (Konigsberg et al. 2002). These papers arguably set the stage for a critical examination of the role of experts and ML in forensic anthropology. However, despite the massive developments of the application of ML and AI models/algorithms in the field and society at large, a significant glitch in the perception of AI and ML has persisted.

We propose that the glitch is intimately tied to the state of education (or lack thereof) on the subject and lies at the root of many challenges we face in this field. We are not proposing to test if computers can be indistinguishable from humans in the field of forensic anthropology, as is the aim of the Turing (Keskinbora 2019) and Lovelace tests (Kalpokiene & Kalpokas 2023). Instead, we believe that machines, by means of AI and ML, can and should be used in tandem with experts in forensic investigations but only when ethically modeled, transparent in the build, and consistently tested for biases. In this article, we therefore aim to give a brief overview of ML and AI, their potentials within a forensic investigation and forensic anthropology specifically as well as their current biases and pitfalls, how they can play a role in decision-making processes, and a path forward for ethical utility in forensic anthropology.

Artificial Intelligence and Machine Learning

Turing (1950) laid the foundation for AI by positing the question, "Can machines think?" His theories led to what we recognize today as AI, even recognizing potential creator biases. He stated that if these "thinking machines" he was proposing were created by just one sex, they "would not really be satisfactory" (Turing 1950:435–436) due to human biases, biases that have been acknowledged and deemed a priority today. To diminish these issues that Turing predicted, invoking diverse teams to build, train, and deploy AI/ML models from the start begins to consider the societal risks and systemic biases (Schwartz et al. 2022). As Buolamwini (2023:54) states, "Neural does not equate to neutral."

In 1956, John McCarthy coined the term "artificial intelligence," and four years later, Arthur Samuel defined the term "machine learning" (Joshi 2020). Whereas AI is the field of computer science that explores the ability for machines to exhibit behavior resembling human intelligence, ML focuses on the development of the algorithms and statistical models that allow computers to learn through training and decision-making, which can then be used by AI (Joshi 2020).

ML can broadly be divided into supervised and unsupervised learning (Alloghani et al. 2020; Berry et al. 2019; Hastie et al. 2009). "Supervised learning" depends on labeled input data (independent variables), creating an algorithm that learns to make predictions or classifications based on the inputted features and known outputs (dependent variables) (Alloghani et al. 2020; Berry et al. 2019; Hastie et al. 2009). It trains by visualizing the relationship between the given input and output data (training set), allowing the model to classify new data into one of the known outputs (Alloghani et al. 2020; Berry et al. 2019; Hastie et al. 2009). "Unsupervised learning" excludes preconceived labels, allowing the model to visualize the input data and cluster the data based on patterns or relationships it discovers, without a specific given output or classification (Alloghani et al. 2020; Berry et al. 2019; Hastie et al. 2009). Although there are pros and cons with both approaches, the learning algorithms in both supervised and unsupervised models benefit from high quantity of data to avoid underfitting and overfitting problems (Lever et al. 2016). Underfitting results from the learning algorithm being unable to capture the variability of the data, causing reduced predictive performances across training, testing, and new data sets (Jabbar & Khan 2015). Overfitting occurs when the model learns and includes the details and noise of irrelevant components to the extent that it negatively impacts the performance of the model on new data and thereby impacts the model's ability to generalize (Ying 2019).

Black box algorithms are types of ML models where the decision-making processes and specific calculations are hidden from the user such that the inputs and outputs are clear but details of how the machine got to that answer are obscured. Black box algorithms include random forest models, artificial neural networks, support vector machines, and clustering algorithms (Hastie et al. 2009). DL is a subset of ML that uses neural networks with multiple hidden layers to learn how the test data are represented through features and patterns, such as image or voice recognition processes (Deng 2016).

Turing (1950) proposed that there are three factors that are needed to imitate an adult human mind: (a) the "blank" slate of the mind at birth, (b) the education of the individual, and (c) the experiences of the individual. Education (training) and experience (iterative processing and testing) are the key features of ML, but both are also a consistent discussion for when it comes to experts and expertise in forensic anthropology (Bartelink et al. 2020; Boyd et al. 2020; Kranioti & Paine 2011; Passalacqua & Pilloud 2020, 2021; Spiros et al. 2023). Visual approaches in forensic anthropology (e.g., trauma analysis and morphological sex/age/population affinity techniques) are arguably inherently subjective (Lesciotto 2023). In the past, methods have been said to be seen as being just as much an art as a science (Hefner 2009; Maples 1989; Rhine 1990), relating to subjective, qualitative analyses such as trait list approaches. The field has progressed to adding the additional level of qualitative analyses to the subjective methods, but they still rely on the experience of the observer to accurately score various features of the skeleton (Kamnikar et al. 2018; Klales & Kenyhercz 2015; Nakhaeizadeh et al. 2020). Subjective classifications, such as any nonmetric approach to sex, age, population affinity, taphonomy, or trauma analysis, rely on observation rather than more objective measurements, highlighting the importance of experience, education, and interobserver and intraobserver error testing. To enhance the transparency within the current methods and minimize the subjective elements involved, machines should complement and enhance human capabilities as a tool, rather than replacing humans, as has been positively explored in forensic science at large.

Forensic Sciences

Within recent years, new innovative approaches to AI have been proposed across forensic science (Galante et al. 2023). The research conducted to date has given insight into tackling some of the challenges forensic science has faced, with the application of AI showing the potential of, for example, training and enhancing decision outcomes (Khanagar et al. 2021), automating manually consuming tasks (Trigueros et al. 2021), increasing efficiency in decision-making (Mesejo et al. 2020), and overcoming subjective limitations in methodology (Bewes et al. 2019). Arguably, the principle aims of integrating AI and automation in forensic investigations includes (but is not limited to) accuracy, efficiency, and cost reduction (Jarrett & Choo 2021). This has been prominent in digital forensics specifically, where the utilization of AI-powered technology has shown efficiency in powering through massive amounts of data in a short amount of time, enabling more accurate and impactful digital investigations (Hall et al. 2022; Jarrett & Choo 2021). With electronic evidence and digital data being a vital part of almost all criminal activities, digital forensics is a branch of forensic science that focuses on uncovering and interpreting electronic data and digital evidence (Interpol 2023). The definition of digital forensics has expanded from computer forensics to include the forensics of all digital technology (e.g., smartphones, remote storage) (Carew et al. 2021; Kohn et al. 2013).

Time-consuming tasks conducted in forensic laboratories have also benefited from AI approaches, such as automated DNA extractions and purifications (Silva 2003). Similar observations have been made within the field of forensic genetics, where probabilistic genotyping software programs are being used as a tool to assist the forensic geneticist (Coble & Bright 2019). In the field of biometrics, artificial neural networks, specifically deep neural networks, have shown superiority over traditional systems in their ability to, for example, match facial images taken from multiple angles (Berghoff et al. 2021). Furthermore, DL methods have been applied to more complex and inconsistent patterns such as shoeprints, which have aided in enhancing the retrieval of shoeprints at crime scenes (Cui et al. 2019), as well as age prediction and gait analysis (Hassan et al. 2021). Even within the more traditional fields applying automatic identification systems (such as fingerprint identification), the function of image technology based on DL has opened new methodologies of identification algorithms (Deshpande et al. 2020). For many of the classification decisions used in forensic science, ML and DL have been proposed to tackle some of the challenges regarding subjectivity, misclassification, and human interpretations (Acampora et al. 2014). For example, studies on the classification of blood stain patterns have shown potential to further develop new automated methods with increased accuracy for bloodstains produced by different forces, as well as being able to study the difference in the accuracy of bloodstain patterns depending on distance and the type of target surface (Liu et al. 2020).

The emerging application of AI across a variety of human identification fields has also been emphasized, with a growing interdisciplinary approach to methods used within the biological profile (Galante et al. 2023; Hassan et al. 2021). In a recent systematic review by Galante et al. (2023), the authors highlighted research utilizing AI in age, sex, population affinity, and postmortem identifications across fields such as forensic odontology (De Tobel et al. 2017; Patil et al. 2020; Vila-Blanco et al. 2020), forensic pathology (Garland et al. 2021; Yilmaz et al. 2017; Zhou et al. 2019), forensic genetics (Vidaki et al. 2017), and forensic anthropology (Bewes et al. 2019). Many of the methods used across these fields share commonalities such as methods being founded on visual observations and classifying decision outcomes, which therefore may benefit from an automated approach.

Forensic Anthropology

As previously stated, the conversation surrounding ML/AI in forensic anthropology began in the early 2000s with a discussion on the application of decision trees (Konigsberg et al. 2002; McBride et al. 2001), but it was not until 2009 when researchers first incorporated and mentioned the utility of statistical decision analyses into forensic anthropology research (Hefner 2009; Stevenson et al. 2009), and within a couple of years, there was an increase in the use of decision trees (Fun et al. 2011; Love et al. 2012; Moore & Schaefer 2011). This was arguably the first step into more consistently using ML in the field.

In 2014, forensic anthropology started exploring more rigorous ML methods (Hefner & Ousley 2014). The first big push that was seen to minimize the subjectivity of the expert by utilizing ML in forensic anthropology was for (macro)morphoscopic traits of the cranium (Hefner & Ousley 2014). Hefner and Ousley (2014) exemplified this by testing multiple types of ML models on the same data set. They applied artificial neural networks, random forest models, and support vector machines to cranial macromorphoscopic traits to assess classification rigor and feasibility of cranial variations in population affinity estimation. To move beyond the antiquated typological ideology that a practitioner could estimate a person's population affinity based on a trait list approach alone, Hefner (2009) standardized the scoring technique of cranial morphological variations. Using these traits, in tandem with ML algorithms, Hefner and Ousley (2014) were able to establish a way to adhere to the standards of empirical testing, peer review, validation, and error rates (Christensen & Crowder 2009; Daubert v. Merrell Dow Pharmaceuticals 1993; Ousley & Hollinger 2012). Applying ML methods increased the accuracy of the previously subjective trait list approach by allowing the decision-making to be transparent and increase the objectivity of the process.

More recently, we have seen a variety of ML algorithms being applied throughout the skeleton and employed across the field within trauma analysis (Kyllonen et al. 2022), radiographic analysis (Lo et al. 2023; Ortiz et al. 2020), age (Gámez-Granados et al. 2022; Joshi & Tallman 2023; Navega & Cunha 2020; Shan et al. 2022), sex (Bidmos et al. 2023; Coelho & Curate 2019; Curate et al. 2017; Imaizumi et al. 2020; Navega et al. 2015; Ortega et al. 2021; Ortiz et al. 2020; Toy et al. 2022), and population affinity (Algee-Hewitt et al. 2020; Hefner & Ousley 2014; Hefner et al. 2014; Hefner et al. 2015; Navega et al. 2015; Spiros & Hefner 2020) estimations.

Beyond the use of ML to help decrease the subjectivity of visual techniques and increase the accuracy of the current methods used in forensic anthropology, within recent years, the application of DL has started to push the techniques used even further by utilizing forensic imaging (Bewes et al. 2019; Yang et al. 2020). Bewes et al. (2019) utilized computed tomography (CT) scans of 900 skulls to train a convolution neural network, a type of DL model, to estimate sex of the individuals. Instead of applying the traditional traits analysis used in forensic anthropology, the neural network did not rely on the preconceived sex-related traits. Instead, the decision-making process of the neural network works to self-identify the regions of the skulls that are most useful to classify the individuals into one of the two groups explored in this study. Testing the model on 100 individuals, the algorithm had an overall correct classification of 95%. Though the black box nature of this model does not allow one to understand which features the algorithm is relying on for this estimation, it does begin to explore how we can utilize these processes to continue to push techniques and methods forward.

The importance of the interaction between humans and the algorithm has been highlighted across numerous fields (Swofford & Champod 2021). Human experts are creative and critical thinking beings who are able to draw inferences that are context specific. Compared to a human expert, ML/AI, to date, may not be able to capture all the minuities and nuances of details included in the interpretation of skeletal remains within the context of a case. As proposed in the medical field, these systems should contribute to the decision-making process but not overrule the expert's judgment (Morgan & Mates 2023). Moreover, ML systems need to operate on vast databases for the machine to learn from as they follow predefined algorithms. When such data are lacking, the ML models are (compared to an expert) limited in articulating why certain exclusions have been made (Hefetz 2023).

Take Fordisc (Jantz & Ousley 2005) as an example. Fordisc is a software that uses statistical methods and data from reference samples of known individuals to create discriminant functions that can help in making these estimations based on the measurements and characteristics of skeletal remains (Ousley & Jantz 2012). Although Fordisc does not utilize ML, as with all technologies, this is an example of how human interpretation should not be excluded. Within the graphical user interface (GUI), users can simply enter craniometrics into the program, press "Process," and expect a result. This, however, is arguably not how one would properly use the program. There is an ethical responsibility to learn how the program works and what its limitations are, to understand the statistical processing, and then to interpret and integrate the results within the context of the case, rather than just accepting the group into which the individual is classified on the first run of the program. Human involvement and oversight are therefore critical in terms of understanding the limitations and risks of any ML model. Despite implementing algorithms in forensic anthropology as a tool for interpretation, there is diversity in how often and how willing practitioners are in applying these models operationally as well as how applicable these models are universally. For example, Fordisc is a system that is very much used in forensic anthropology casework, but it is primarily applicable within a US context. Furthermore, Swofford and Champod (2021) emphasized in their article the lack of implementations of ML and AI models in forensic practical settings. Some of the arguments highlighted were practitioners' reservations and concerns as well as lack of trust regarding its application in casework. This shows some of the misconceptions that practitioners in forensic science might have regarding AI/ML applications. We need to work with the machines, not rely on them, to ensure accountability, transparency, and liability. This is especially important to tackle some of the ethical challenges and potential bias that may be introduced at various levels of the analysis.

AI and ML Bias

Along with multiple advantages, AI systems and computational models also create new challenges. Research has shown that there is, arguably, a level of subjectiveness naturally inherent in methods used in science in general as discussed within the context of the myth of objectivity (Leavy et al. 2020). Forensic anthropologists rely on human interpretations and the experience of the observer for almost all techniques in one way or another. Studies within cognitive bias in forensic anthropology specifically (Hartley et al. 2022; Nakhaeizadeh et al. 2018) and the criminal justice system broadly have repeatedly shown different extents to which cognitive biases may influence the decision-making process and the interpretations of evidence showing how human decision-making can be shaped and impacted by context, society, and other stimuli (Buolamwini 2023; Cooper & Meterko 2019). Since ML algorithms learn to consider only the variables that improve their predictive accuracy, in many cases, AI can therefore arguably reduce human subjective interpretation of data (Buolamwini 2023; Hefetz 2023), but just as humans can be influenced by a variety of biases (Cooper & Meterko 2019; Davidson et al. 2023; Klales & Lesciotto 2016; Nakhaeizadeh & Morgan 2009; Nakhaeizadeh et al. 2014a, 2014b; Nakhaeizadeh et al. 2015; Nakhaeizadeh et al. 2018; Nakhaeizadeh et al. 2020; Warren et al. 2018), so can an algorithm that is built by humans.

Various issues regarding including type I/II errors (false positives and false negatives, respectively) for diagnoses or discrimination have been seen as a result of bias in AI/ML, including, but not limited to, reinforcing societal stereotypes and sexist/racist risk assessments (Bertrand et al. 2022; Buolamwini 2017, 2023; Celi et al. 2022; Cooper & Meterko 2019; DeBrusk 2018; Fletcher et al. 2021; Fraser et al. 2022; Koenecke et al. 2020; Schwartz et al. 2022; Yapo & Weiss 2018). These issues can appear either within the data (systemic biases), within the algorithm (statistical/computational biases), or by the user themselves (human biases) and have inspired groups such as the Algorithmic Justice League (Buolamwini 2023). Although statistical models and algorithms can help in providing an empirical foundation to experts' subjective conclusions (Swofford & Champod 2021), there is a misconception that AI will eliminate human biases. Technochauvinism is this idea that technology is always the solution (Broussard 2019; Schwartz et al. 2022). It is important to acknowledge that ML and AI rely heavily on data that are either collected via systems created by humans or generated by humans. Arguably, whatever biases may or may not exist in humans, their input may be reproduced or even increase when modeling, using, and interpreting AI systems (Bertrand et al. 2022; Buolamwini 2023).

Looking at the data is the first step to interpreting the potential for bias when it comes to these systems. It has been argued that some of the current data sets used in research that are limited in size are not necessarily representative of real-world forensic investigations, with more empirical studies needed to increase accuracy as well as its ecological validity (Galante et al. 2023). A specific type of ML that is utilized due to its high accuracy are black box models, such as neural networks. These models are not easily explainable due to the hidden layers of predictive steps and decisions, which, in turn, lack an element of transparency (Benjamins et al. 2019; Bertrand et al. 2022; Buolamwini 2017; DeBrusk 2018; Franzoni 2023; Hefetz 2023; Keskinbora 2019; Nguyen et al. 2023; Samek et al. 2017; Solanke 2022; Stahl et al. 2023; Završnik 2020; Zhang & Zhang 2023).

Many studies in forensic science conducted to date have primarily focused on using black box algorithms such as random forest, neural networks, and support vector machines, which arguably have unexplainable elements to their procedures when used in isolation (Hefner & Ousley 2014; Samek et al. 2017; Spiros & Hefner 2020). Therefore, caution has been raised for the use of AI models in critical contexts such as the justice system (Nowotko 2021). The skepticism expressed by legal practitioners, courts, and the public over AI evidence used in criminal proceedings has mainly focused on the lack of understanding and transparency in how the AI system reached its decision (Atkinson et al. 2020; Solanke 2022).

For example, random forest models result in slightly different results every time if a random seed is not set (Genuer et al. 2020). When a random seed is set to ensure reproducibility, this fixes the variation in the classificatory statistics being variable but removes part of the inherent randomness. Thus, it might be best for classification for seeds not to be set. While this might minimize reproducibility of the statistics, it may make the model more realistic for future unknown data when the seed is not set to optimize the training data set or even if a random seed is utilized. As with all black box algorithms, it is important to be able to explain why the method and why the results may vary. If the exact decision-making processes of the machine are unable to be explained, such as black box algorithms, the processes of how the models are built should still be explained. Working in a field where methodology needs to be explainable (e.g., judges, jurors), this is not a call for excluding black box methods in forensic science but, rather, opening a broader field discussion of how to appropriately represent the methods that are being utilized to the general public.

Explainable AI (XAI) is a growing field that aims to make AI systems and the data they utilize transparent by "glass-boxing" the system's functioning components (Guidotti et al. 2018; Samek et al. 2017). However, maintaining a good balance between performance, transparency, and accuracy might be challenging when applying XAI, with some scholars arguing for systems that are instead interpretable (Rudin 2019) and responsible (Benjamins et al. 2019). Furthermore, explainability is related to the concept that AI models and their outputs can be rationally explained in a way acceptable for human understanding (Solanke 2022). This might be problematic in a forensic context where some ML models, although more straightforward, might perform lower compared to DL models that might perform significantly better but be considerably more complex to explain to a judge or a juror. Interpretability in AI, on the other hand, might also be challenging in a forensic context as understanding the inner workings of a model and communicating an explanation that is comprehensible will arguably be domain specific (Solanke 2022). There is therefore a need to develop AI models that are precise, understandable, and objective in their decision-making, which might be a challenge in forensic anthropology for a variety of reasons.

When we look at potential biases related to the models themselves, there are various examples that can already be highlighted as potential issues in forensic anthropology. Labeling bias can be identified when assigned categories or labels for data points are influenced by subjective decisions— although this is not an issue if utilizing an unsupervised learning algorithm for classification (Alloghani et al. 2020; Berry et al. 2019). Labeling bias in forensic anthropology can be culturally constructed ideologies. For example, in biological profile analyses, the individual reporting the age, sex, stature, or population affinity of a missing person has the ability to make a decision on this label—leading to a potential for error. Likewise, if an individual receiving a forensic anthropological report inputs the estimated information of the unknown individual into a missing person database improperly, this would be an example of labeling bias.

Selection/sample bias can be seen when the sample used is not representative of the broader population (Dankwa-Mullan & Weeraratne 2022; Fletcher et al. 2021; Mehrabi et al. 2021; Olubeko 2021; Pan et al. 2021; Park & Hu 2023; Schwartz et al. 2022; Zhang et al. 2022). While unsupervised learning is beneficial in countering labeling bias, if the sample itself is skewed, even unsupervised learning outputs can become problematic. Moreover, a machine cannot have an outcome of what it is not fed in training. This is common within forensic anthropology as data are frequently limited due to bias in donated collections (Winburn et al. 2020), lack of diverse data available (Winburn et al. 2020), and the impact of the osteological paradox (Alves Cardoso & Henderson 2018). Due to intrinsic limitations in sex, population affinity, and age estimations (see Galante et al. 2022 for full review) as well as limitations in acquiring large enough samples of skeletal remains from known modern populations, transparency is key in beginning to address these issues.

Data collection bias, on the other hand, is when the error stems from the collection of the data themselves (Fletcher et al. 2021; Mehrabi et al. 2021; Pannucci & Wilkins 2010), for example, a human mismeasuring or misapplying a technique (Adams & Byrd 2002; Christensen et al. 2014; Jantz et al. 1995; Lesciotto & Doershuk 2018). Statistical safeguards may be able to pick up these outliers if the error occurs once or if the data collection procedure is compared through interobserver tests, but without these checks to mitigate these issues, we cannot rely on the algorithms themselves to understand these errors in isolation.

When it comes to algorithmic bias, one element that should be explored when utilizing AI/ML models is feedback loop bias (Mehrabi et al. 2021; Pan et al. 2021; Schwartz et al. 2022). Feedback loop bias is seen when the user introduces bias in the labeling, selection, or sampling stages (as previously discussed), but this bias is then perpetuated by the algorithm itself (Mehrabi et al. 2021; Pan et al. 2021; Schwartz et al. 2022). Algorithms can then, in turn, amplify societal and structural biases as patterns of objective truth when they are not (Schwartz et al. 2022; Yapo & Weiss 2018). Nevertheless, similar to the medical field, the initiatives in using digital data in developing new methods in human skeletal identification to produce accurate estimation/prediction models using medical images are growing, showing promising preliminary results and innovation in the field of forensic anthropology (Bewes et al. 2019; Lo et al. 2023).

Current attempts for tackling the detrimental effects of ML/AI and bias continue to focus on computational factors, such as representativeness of data sets and fairness of ML algorithms (Schwartz et al. 2022). In forensic anthropology, we see model evaluation in reporting correct classification rates and confusion matrices (Joshi & Tallman 2023;

Kamnikar 2022; Patil et al. 2020; Ortega et al. 2021). Further, tests of algorithmic performance include utilizing holdout samples and cross-validating samples in order to test result consistency and overfitting the model (Algee-Hewitt et al. 2020; Hastie et al. 2009; Hefner & Ousley 2014).

Although these are important factors to consider, human, institutional, and societal factors tend to be overlooked. Moreover, much of the discussion on approaches for AI and bias mitigation is focused on the preprocessing methods (focusing on the data), in-processing methods (focusing on the algorithm), and postprocessing methods (focusing on the model). To address bias within AI successfully, we need to expand our perspectives beyond the ML pipeline with empirical research looking at the impact of AI on human decision-making. Kliegr and colleagues (2021) provided a study on the effect of cognitive bias on the interpretation of AI models and human understanding, outlining over 20 different biases. Although informative and extremely useful, the article focused only on rule-based explanations, in which if–then rules are used to illustrate conditions that must be met for a decision to be made to enhance the interpretability of the model. Further studies are therefore needed that incorporate a broader aspect of human integration—for example, cognitive biases related to XAI techniques (Bertrand et al. 2022). Taking into consideration a holistic approach to human aspects will allow us to address the challenges of how these technologies are not only created but also interpreted by experts and how the outcomes of these models impact on our society from legal and ethical perspectives.

In order to mitigate against ML/AI algorithms being biased, there are various controls that can be taken. These include, but are not limited to, evaluating the data collection procedures, sample assessment, preprocessing quality assurance, algorithm design, model testing, regular evaluation, interdisciplinary collaboration, and incorporating ethical frameworks (Flores-Vivar & García-Peñalvo 2023; Morgan & Mates 2023; Noseworthy et al. 2020; Solanke 2022; Zhang et al. 2022; Zhang & Zhang 2023; Zimmerman et al. 2023). As AI and ML stand, especially with the black box approaches, experts should not solely rely on a machine's results. There should be a certain level of human logic at play when utilizing these methods. Beyond the decision process of the user, these controls should be incorporated due to the possibility of skewed results since these controls are not necessary for the algorithms to "work." See "NIST Towards a Standard for Identifying and Managing Bias in Artificial Intelligence" (Schwartz et al. 2022) for a comprehensive list and glossary of the human, systemic, and statistical/computational biases that may influence ML/AI. This leads to the need for ethical use and intentionality creating, applying, and analyzing AI and ML throughout the entire process.

Ethics and Intentionality

Ethics issues are seen in various fields related to AI and ML (Bali et al. 2019; Bankins & Formosa 2023; Hawkins & Mittelstadt 2023; Hefetz 2023; Hosseini et al. 2023). Just this year, the U.S. Copyright Office declared that images created using the AI-powered image generators should not be granted copyright protection (Copyright Office 2023). Beyond legality, the current iterations of AI "art" generators are built on art that is posted online, regardless of copyright (Vinchon et al. 2023), an exploitation that is using preexisting artwork regardless of copyright without the ability for artists who post their work online to opt out from their art being web-scrapped and used in these training sets (Salkowitz 2022). This poses one of the important ethical questions in terms of where the training data originated from (Anshari et al. 2023).

As we see with other integrations of technology and the digital world with human osteology, transparency, permissions, and respect should be recognized as integral, especially when creating data sets for training algorithms (Errickson & Thompson 2019; Spiros et al. 2022). Legal standards alone are not enough. Legal systems struggle to keep up with the fast-paced digital world, while simultaneously not holding decedents to the same standards of respect as living human beings. Looking at ethical guidelines for AI from a global landscape, many of these generally include regulations around ethical principles such as fairness, justice, transparency, privacy, and responsibility (Hefetz 2023). However, there is a lack of consensus as to how these principles are applied, interpreted, and implemented (Jobin et al. 2019). What has been highlighted is that within criminal investigations, many of the protocols and frameworks used may not consider the challenges that are pertinent within law enforcement and forensic contexts. As we see in the use of skeletal collections and other uses of human remains online, laws fail to keep up with the research, protocols, and pedagogy surrounding the deceased. This is why ethics, as well as a case-by-case discussion, is important for digital data sets. Although the research in AI and forensic anthropology has been impressive, there is arguably a lack of ethical discussions in the field regarding the compatibility of these models in a legal investigation as well as how to integrate these models in institutional settings without compromising on transparency, accuracy, and public trust (Hefetz 2023).

In response to the lack of legal timeliness, the medical field has already begun to discuss the impact of AI focusing on bioethics. Global bioethics center on the principles of autonomy, justice, beneficence, and nonmaleficence (Blumenthal-Barby 2023; Keskinbora 2019; Lewis 2020; Nabi 2018; Tai 2020; Zimmerman et al. 2023). While there have also been discussions on how these medical bioethical tenets can be problematic if only understanding the terms

TABLE 1—*Principles of Trust in AI Systems (Keskinbora 2019)*

Principle	Definition
Transparency	Operations are visible to the user
Credibility	Outcomes are acceptable
Auditability	Efficiency can be easily measured
Reliability	AI systems perform as intended
Recoverability	Manual control can be assumed if required

from a Western lens (Bali 2018; Spiros et al. 2022; Takala 2001; Tosam 2020; Turner 2005), these principles can be utilized as a beginning guide to having broader discussions on the ethical use of ML/AI in forensic anthropology. There are a multitude of ethics questions that are raised in the medical system when it comes to the utility of ML/AI in decision-making processes—specifically, how the models are built and subsequently are incorporated. Five elements, seen in Table 1, have been proposed as being necessary in order for these processes to be seen as "trustworthy": transparency, credibility, auditability, reliability, and recoverability (Francesca 2016; Keskinbora 2019).

Similar to the standards that forensic anthropological methodology is held to (Christensen 2004; Christensen & Crowder 2009; Grivas & Komar 2008; Lesciotto 2015; Ousley & Hollinger 2012), the additional "recoverability" principle is added when it comes to machines. Recoverability is the idea that humans can override the control of the system, if needed. As previously discussed, transparency is key within the forensic context and plays a pivotal role in understanding the decision-making processes (Bali et al. 2020; Chin et al. 2022; Chin & Ibaviosa 2022; Fleischman et al. 2019; Hartley et al. 2022; Houck 2019; Klales & Lesciotto 2016; Nakhaeizadeh et al. 2018; Nakhaeizadeh et al. 2020; Passalacqua et al. 2019; Spiros et al. 2023; Warren et al. 2018). Transparency goes beyond algorithms and decision-making, but understanding the source of the data is also of utmost importance. This is where there is a discordance between transparency of the methods and privacy of the individual. While open data sets would be ideal, working with human remains makes open data more difficult. When the data themselves cannot be open-sourced (i.e., forensic casework), there becomes a black box effect in the research itself that proves to be difficult to overcome. Once again, permission (Errickson & Thompson 2019; Spiros et al. 2022) and privacy (Keskinbora 2019; Nabi 2018; Nguyen et al. 2023; Skorburg et al. 2020) are integral but make complete transparency a hurdle. While various fields discuss copyright and infringement, there are minimal legal standards for utilizing deceased human subjects in research, only ethical principles (Errickson & Thompson 2019; Spiros et al. 2022).

Following practices utilizing digital osteology, we propose guidelines to consider when incorporating decedents into ML/AI models: (a) do not share the data openly without permissions from the decedents; (b) do not incorporate the data into larger, open source data sets that could be pulled from web-scraping without permissions; (c) focus on respect and privacy of the individuals (deidentification); and (d) obtain permission, when possible, from the decedent or next of kin for use of their remains in research and educational practices.

Beyond the data alone, the U.S. government has now created an "AI Bill of Rights" that proposed five principles that should guide the creation and utility of ML and AI (Blumenthal-Barby 2023; OSTP 2022). These include (a) safe and effective systems, (b) algorithmic discrimination protections, (c) data privacy, (d) notice and explanation, and (e) human alternatives, consideration, and fallback (Blumenthal-Barby 2023; OSTP 2022). As computing power is increasing, the need to understand the symbiotic relationship of humans and computers with education and ethics guiding our practices is pivotal.

Conclusion

With the increased incorporation of AI and ML systems in forensic anthropology, there needs to be an increased consideration in education, ML/AI training, and ethical discussions when bridging the two fields. Transparency in ML/AI design, clear guidelines for data collection and integration, and ongoing evaluation of algorithmic performance are essential. Rigorous bias and ethics are needed throughout and beyond the processing pipeline but also into the human decision-making realms. Only through such an integrated approach can the potential of AI/ML in forensic anthropology be utilized responsibly, preserving the dignity of the deceased. Like the progress seen across the medical domain and clinical decision-making (Dietvorst et al. 2018), statistical modeling and ML/AI algorithms should be part of one entity with human decision-making and not arguably seen as two separate solutions to the challenges facing forensic anthropology specifically and forensic science more broadly. Therefore, to increase the benefit and mitigate risks, AI/ML should be utilized as a tool— not replacing but extending the utility of experts. Rigorous safeguards and discussions need to be had now to make sure the incorporation of machines continues to develop to improve and not to harm individuals. Thus, Turing's assertion still rules in the field of machines: "We can only see a short distance ahead, but we can see plenty there that needs to be done" (Turing 1950:460).

Acknowledgments

We thank Joseph Hefner and Richard Jantz for organizing and including us in this Festschrift honoring Steve Ousley. We also thank *Forensic Anthropology* editors for publishing this special issue.

References

Acampora G, Vitiello A, Di Nunzio C, Saliva M, Garofano L. Bloodstain pattern analysis. In: Proceedings of the International Joint Conference on Computational Intelligence, October 22–24, 2014; Rome, Italy.

Adams BJ, Byrd JE. Interobserver variation of selected postcranial skeletal measurements. *Journal of Forensic Sciences* 2002;47(6):1193–1202.

Algee-Hewitt B, Coelho C, Navega D, Cunha E. Statistical approaches to ancestry estimation: New and established methods for the quantification of cranial variation for forensic casework. In: Obertová Z, Stewart A, Cattaneo C, eds. *Statistics and Probability in Forensic Anthropology*. London: Elsevier; 2020:227–247.

Alloghani M, Al-Jumeily D, Mustafina J, Hussain A, Aljaaf A. A systematic review on supervised and unsupervised machine learning algorithms for data science. In: Berry M, Mohamed A, Yap B, eds. *Supervised and Unsupervised Learning for Data Science*. Cham: Springer; 2020:3–21.

Alves Cardoso F, Henderson CY. Identified skeletal collections: The testing ground of anthropology? In: Henderson CY, Cardoso FA, eds. *Identified Skeletal Collections: The Testing Ground of Anthropology?* Oxford: Archaeopress Archaeology; 2018:1–198.

Anshari M, Hamdan M, Ahmad N, Ali E, Haidi H. COVID-19, artificial intelligence, ethical challenges and policy implications. *AI & Society* 2023;38(2):707–720.

Aski AS, Sourati NK. Proposed efficient algorithm to filter spam using machine learning techniques. *Pacific Science Review A: Natural Science and Engineering* 2016;18(2):145–149.

Atkinson K, Bench-Capon T, Bollegala D. Explanation in AI and law: Past, present and future. *Artificial Intelligence* 2020;289: 103387.

Bali AS, Edmond G, Ballantyne KN, Kemp RI, Martire KA. Communicating forensic science opinion: An examination of expert reporting practices. *Science & Justice* 2020;60(3):216–224.

Bali J, Garg R, Bali RT. Artificial intelligence (AI) in healthcare and biomedical research: Why a strong computational/AI bioethics framework is required? *Indian Journal of Ophthalmology* 2019;67(1):3.

Bali M. The "unbearable" exclusion of the digital. In: Kim D, Stommel J, eds. *Disrupting the Digital Humanities*. Brooklyn, NY: Punctumbooks; 2018:295–319.

Bankins S, Formosa P. The ethical implications of artificial intelligence (AI) for meaningful work. *Journal of Business Ethics* 2023;185:1–16.

Bartelink EJ, Boyd DC, France DL, Pokines JT, Prince-Zinni D. The American Board of Forensic Anthropology turns 40: Historical perspectives and current trends in certification for forensic anthropology. *Forensic Anthropology* 2020;3(2): 103–111.

Benjamins R, Barbado A, Sierra D. Responsible AI by design in practice. arXiv preprint arXiv:1909.12838. 2019.

Berghoff C, Neu M, von Twickel A. The interplay of AI and biometrics: Challenges and opportunities. *Computer* 2021;54(9): 80–85.

Berry MW, Mohamed A, Yap BW. *Supervised and Unsupervised Learning for Data Science*. Cham: Springer; 2019.

Bertrand A, Belloum R, Eagan JR, Maxwell W. How cognitive biases affect XAI-assisted decision-making: A systematic review. In: Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society; August 2022; Oxford.

Bertsatos A, Chovalopoulou M, Brůžek J, Bejdová Š. Advanced procedures for skull sex estimation using sexually dimorphic morphometric features. *International Journal of Legal Medicine* 2022;134:1927–1937.

Bewes J, Low A, Morphett A, Pate FD, Henneberg M. Artificial intelligence for sex determination of skeletal remains: Application of a deep learning artificial neural network to human skulls. *Journal of Forensic and Legal Medicine* 2019;62: 40–43.

Bidmos MA, Olateju OI, Latiff S, Rahman T, Chowdhury MEH. Machine learning and discriminant function analysis in the formulation of generic models for sex prediction using patella measurements. *International Journal of Legal Medicine* 2023;137(2):471–485.

Blumenthal-Barby J. An AI Bill of Rights: Implications for health care AI and machine learning—A bioethics lens. *The American Journal of Bioethics* 2023;23(1):4–6.

Boyd DC, Bartelink EJ, Passalacqua NV, Pokines JT, Tersigni-Tarrant M. The American Board of Forensic Anthropology's certification program. *Forensic Anthropology* 2020;3(2):112.

Broussard M. *Artificial Unintelligence*. Cambridge, MA: MIT Press; 2019.

Budka M, Ul Ashraf AW, Bennett M, Neville S, Mackrill A. Deep multilabel CNN for forensic footwear impression descriptor identification. *Applied Soft Computing* 2021;109:107496.

Buolamwini JA. *Gender Shades: Intersectional Phenotypic and Demographic Evaluation of Face Datasets and Gender Classifiers* [MS thesis]. Cambridge: Massachusetts Institute of Technology; 2017.

Buolamwini JA. *Unmasking AI: My Mission to Protect What Is Human in a World of Machines*. New York, NY: Random House; 2023.

Cao Y, Ma Y, Yang X, Xiong J, Wang Y, Zhang J, et al. Use of deep learning in forensic sex estimation of virtual pelvic models from the Han population. *Forensic Sciences Research* 2022;7(3):540–549.

Carew RM, French J, Morgan RM. 3D forensic science: A new field integrating 3D imaging and 3D printing in crime reconstruction. *Forensic Science International: Synergy* 2021;3: 100205.

Celi LA, Cellini J, Charpignon M, Dee EC, Dernoncourt F, Eber R, et al. Sources of bias in artificial intelligence that perpetuate healthcare disparities—A global review. *PLoS Digital Health* 2022;1(3):e0000022.

Chin JM, Growns B, Sebastian J, Page MJ, Nakagawa S. The transparency and reproducibility of systematic reviews in forensic science. *Forensic Science International* 2022;340: 111472.

Chin JM, Ibaviosa CM. Beyond CSI: Calibrating public beliefs about the reliability of forensic science through openness and transparency. *Science & Justice* 2022;62(3):272–283.

Christensen AM. The impact of Daubert: Implications for testimony and research in forensic anthropology (and the use of frontal sinuses in personal identification). *Journal of Forensic Sciences* 2004;49(3):427–430.

Christensen AM, Crowder CM. Evidentiary standards for forensic anthropology. *Journal of Forensic Sciences* 2009;54(6): 1211–1216.

Christensen AM, Crowder CM, Ousley SD, Houck MM. Error and its meaning in forensic science. *Journal of Forensic Sciences* 2014;59(1):123–126.

Coble MD, Bright J. Probabilistic genotyping software: An overview. *Forensic Science International: Genetics* 2019;38: 219–224.

Coelho J, Curate F. CADOES: An interactive machine-learning approach for sex estimation with the pelvis. *Forensic Science International* 2019;302:109873.

Cooper GS, Meterko V. Cognitive bias research in forensic science: A systematic review. *Forensic Science International* 2019;297:35–46.

Copyright Office, Library of Congress. *88 FR 16190: Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence*. Washington, DC: Library of Congress: Federal Register; 2023.

Cui J, Zhao X, Liu N, Morgachev S, Li D. Robust shoeprint retrieval method based on local-to-global feature matching for real crime scenes. *Journal of Forensic Sciences* 2019; 64(2):422–430.

Curate F, Umbelino C, Perinha A, Nogueira C, Silva AM, Cunha E. Sex determination from the femur in Portuguese populations with classical and machine-learning classifiers. *Journal of Forensic and Legal Medicine* 2017;52:75–81.

Dankwa-Mullan I, Weeraratne D. Artificial intelligence and machine learning technologies in cancer care: Addressing disparities, bias, and data diversity. *Cancer Discovery* 2022;12(6):1423–1427.

Darmawan, MF, Yusuf SM, Kadir MRA, Haron H. Comparison on three classification techniques for sex estimation from the bone length of Asian children below 19 years old: An analysis using different group of ages. *Forensic Science International* 2015;247:130.e1–130.e11.

Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579. 1993.

Davidson M, Nakhaeizadeh S, Rando C. Cognitive bias and the order of examination in forensic anthropological non-metric methods: A pilot study. *Australian Journal of Forensic Sciences* 2023;55(2):255–271.

De Tobel J, Radesh P, Vandermeulen D, Thevissen PW. An automated technique to stage lower third molar development on panoramic radiographs for age estimation: A pilot study. *The Journal of Forensic Odonto-Stomatology* 2017;35(2):42.

DeBrusk C. The risk of machine-learning bias (and how to prevent it). *MIT Sloan Management Review* 2016;15:1.

Deng L. Deep learning: From speech recognition to language and multimodal processing. *APSIPA Transactions on Signal and Information Processing* 2016;5:e1.

Deshpande UU, Malemath VS, Patil SM, Chaugule SV. CNNAI: A convolution neural network-based latent fingerprint matching using the combination of nearest neighbor arrangement indexing. *Frontiers in Robotics and AI* 2020;7:113.

Devlin J, Chang M, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805. 2018.

Dietvorst BJ, Simmons JP, Massey C. Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science* 2018; 64(3):1155–1170.

Earwaker H, Nakhaeizadeh S, Smit NM, Morgan RM. A cultural change to enable improved decision-making in forensic science: A six phased approach. *Science & Justice* 2020;60(1): 9–19.

Errickson D, Thompson TJU. Sharing is not always caring: Social media and the dead. In: Squires K, Errickson D, Márquez-Grant N, eds. *Ethical Approaches to Human Remains*. Cham, Switzerland: Springer; 2019:299–313.

Fleischman JM, Pierce ML, Crowder CM. Transparency in forensic anthropology through the implementation of quality assurance practices. In: Fulginiti LC, Hartnett-McCann K, Galloway A, eds. *Forensic Anthropology and the United States Judicial System*. Oxford: Wiley; 2019:70–88.

Fletcher RR, Nakeshimana A, Olubeko O. Addressing fairness, bias, and appropriate use of artificial intelligence and machine learning in global health. *Frontiers in Artificial Intelligence* 2021;3:1–17.

Flores-Vivar J, García-Peñalvo F. Reflections on the ethics, potential, and challenges of artificial intelligence in the framework of quality education (SDG4). *Comunicar* 2023; 31(74):37–47.

Franzoni V. From black box to glass box: Advancing transparency in artificial intelligence systems for ethical and trustworthy AI. In: Gervasi O, Murgante B, Rocha AMAC, Garau C, Scorza F, Karaca Y, Torre CM, eds. *Computational Science and Its Applications–ICCSA 2023 Workshops*. Cham: Springer Nature Switzerland; 2023.

Fraser KC, Kiritchenko S, Nejadgholi I. Diversity is not a one-way street: Pilot study on ethical interventions for racial bias in text-to-image systems. In: Proceedings of the 14th International Conference on Computational Creativity, June 2023; Waterloo, ON, Canada.

Frey CB, Osborne MA. The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change* 2017;114:254–280.

Fun PS, Osman K, Hamzah SPAA, Hamzah NH. Stature and sex estimation using foot measurements for Malays and Chinese in Malaysia. *Jurnal Sains Kesihatan Malaysia* 2011;9(2): 23–28.

Galante N, Cotroneo R, Furci D, Lodetti G, Casali MB. Applications of artificial intelligence in forensic sciences: Current potential benefits, limitations and perspectives. *International Journal of Legal Medicine* 2023;137(2):445–458.

Garland J, Hu M, Kesha K, Glenn C, Morrow P, Stables S, et al. Identifying gross post-mortem organ images using a pre-trained convolutional neural network. *Journal of Forensic Sciences* 2021;66(2):630–635.

Genuer R, Poggi JM, Genuer R, Poggi JM. Introduction to random forests with R. In: Genuer R, Poggi J-M, eds. *Random Forests with R*. Cham: Springer; 2020:1–8.

Grivas CR, Komar DA. Kumho, Daubert, and the nature of scientific inquiry: Implications for forensic anthropology. *Journal of Forensic Sciences* 2008;53(4):771–776.

Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D. A survey of methods for explaining black box models. *ACM Computing Surveys (CSUR)* 2018;51(5):1–42.

Hall SW, Sakzad A, Choo KR. Explainable artificial intelligence for digital forensics. *Wiley Interdisciplinary Reviews: Forensic Science* 2022;4(2):e1434.

Hartley S, Winburn AP, Dror IE. Metric forensic anthropology decisions: Reliability and biasability of sectioning-point-based sex estimates. *Journal of Forensic Sciences* 2022; 67(1):68–79.

Hassan M, Wang Y, Wang D, Li D, Liang Y, Zhou Y, et al. Deep learning analysis and age prediction from shoeprints. *Forensic Science International* 2021;327:110987.

Hastie T, Tibshirani R, Friedman JH, Friedman JH. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. New York: Springer; 2009.

Hawkins W, Mittelstadt B. The ethical ambiguity of AI data enrichment: Measuring gaps in research ethics norms and practices. In: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, June 12–15, 2023; Chicago, IL.

Hefetz I. Mapping AI-ethics' dilemmas in forensic case work: To trust AI or not? *Forensic Science International* 2023;350: 111807.

Hefner JT, Ousley SD. Statistical classification methods for estimating ancestry using morphoscopic traits. *Journal of Forensic Sciences* 2014;59(4):883–890.

Hefner JT, Pilloud MA, Black CJ, Anderson BE. Morphoscopic trait expression in "Hispanic" populations. *Journal of Forensic Sciences* 2015;60(5):1135–1139.

Hefner JT, Spradley MK, Anderson B. Ancestry assessment using random forest modeling. *Journal of Forensic Sciences* 2014;59(3):583–589.

Hefner JT. Cranial nonmetric variation and estimating ancestry. *Journal of Forensic Sciences* 2009;54(5):985–995.

Hosseini M, Resnik DB, Holmes K. The ethics of disclosing the use of artificial intelligence tools in writing scholarly manuscripts. *Research Ethics* 2023;19(4):449–465.

Houck MM. Open, transparent science helps promote justice. *Forensic Science International: Synergy* 2019;1:A275–A276.

Interpol—Digital Forensics. https://www.interpol.int/en/How-we-work/Innovation/Digital-forensics. Accessed December 11, 2023.

Imaizumi K, Bermejo E, Taniguchi K, Ogawa Y, Nagata T, Kaga K, et al. Development of a sex estimation method for skulls using machine learning on three-dimensional shapes of skulls and skull parts. *Forensic Imaging* 2020;22:1–9.

Jabbar H, Khan RZ. Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study). *Computer Science, Communication and Instrumentation Devices* 2015;70(10.3850):978–981.

Jantz RL, Hunt DR, Meadows L. The measure and mismeasure of the tibia: Implications for stature estimation. *Journal of Forensic Sciences* 1995:40(5):758–761.

Jantz RL, Ousley SD. *FORDISC 3.0: Personal Computer Forensic Discriminant Functions*. Knoxville: University of Tennessee, Knoxville; 2005.

Jarrett A, Choo KR. The impact of automation and artificial intelligence on digital forensics. *Wiley Interdisciplinary Reviews: Forensic Science* 2021;3(6):e1418.

Jebara T. Personalization of Spotify Home and TensorFlow. *O'Reilly*. October 31, 2019. https://www.oreilly.com/radar/personalization-of-spotify-home-and-tensorflow/.

Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. *Nature Machine Intelligence* 2019;1(9):389–399.

Joshi AV. *Machine Learning and Artificial Intelligence*. Cham, Switzerland: Springer Nature Switzerland; 1962.

Kalpokiene J, Kalpokas I. Creative encounters of a posthuman kind–anthropocentric law, artificial intelligence, and art. *Technology in Society* 2023;72:102197.

Kamnikar KR. *Cranial Metric and Nonmetric Variation in Southeast Mexico and Guatemala: Implications for Population Affinity Assessment in the United States* [PhD dissertation]. East Lansing: Michigan State University; 2022.

Kamnikar KR, Plemons AM, Hefner JT. Intraobserver error in macromorphoscopic trait data. *Journal of Forensic Sciences* 2018;63(2):361–370.

Kaur P, Krishan K, Sharma SK, Kanchan T. Facial-recognition algorithms: A literature review. *Medicine, Science and the Law* 2020;60(2):131–139.

Keskinbora KH. Medical ethics considerations on artificial intelligence. *Journal of Clinical Neuroscience* 2019;64:277–282.

Khanagar SB, Vishwanathaiah S, Naik S, Al-Kheraif AA, Divakar DD, Sarode SC, et al. Application and performance of artificial intelligence technology in forensic odontology–A systematic review. *Legal Medicine* 2021;48:101826.

King M. Harmful biases in artificial intelligence. *The Lancet Psychiatry* 2022;9(11):e48.

Klales AR, Kenyhercz MW. Morphological assessment of ancestry using cranial macromorphoscopics. *Journal of Forensic Sciences* 2015;60(1):13–20.

Klales AR, Lesciotto KM. The "science of science": Examining bias in forensic anthropology. In: Proceedings American Academy of Forensic Sciences, February 22–27, 2016; Las Vegas, NV.

Kliegr T, Bahník Š, Fürnkranz J. A review of possible effects of cognitive biases on interpretation of rule-based machine learning models. *Artificial Intelligence* 2021;295:103458.

Koenecke A, Nam A, Lake E, Nudell J, Quartey M, Mengesha Z, et al. Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences* 2020;117(14):7684–7689.

Kohn MD, Eloff MM, Eloff JH. Integrated digital forensic process model. *Computers & Security* 2013;38:103–115.

Konigsberg LW, Herrmann NP, Wescott DJ. Commentary on: McBride DG, Dietz MJ, Vennemeyer MT, Meadors SA, Benfer RA, Furbee NL. Bootstrap methods for sex determination from the os coxae using the ID3 algorithm. *Journal of Forensic Sciences* 2002;47(2):424–426.

Kranioti E, Paine R. Forensic anthropology in Europe: An assessment of current status and application. *Journal of Anthropological Sciences* 2011;89:71–92.

Kyllonen KM, Monson KL, Smith MA. Postmortem and antemortem forensic assessment of pediatric fracture healing from radiographs and machine learning classification. *Biology* 2022;11(5):749.

Lamkhede SD, Kofler C. Recommendations and results organization in Netflix search. In: Proceedings of the 15th ACM Conference on Recommender Systems, September 13, 2021; Amsterdam.

Langley NR, Dudzik B, Cloutier A. A decision tree for nonmetric sex assessment from the skull. *Journal of Forensic Sciences* 2018;63(1):31–37.

Leavy S, O'Sullivan B, Siapera E. Data, power and bias in artificial intelligence. arXiv preprint arXiv:2008.07341. 2020.

Lesciotto KM. The impact of Daubert on the admissibility of forensic anthropology expert testimony. *Journal of Forensic Sciences* 2015;60(3):549–555.

Lesciotto KM, Doershuk LJ. Accuracy and reliability of the Klales et al. (2012) morphoscopic pelvic sexing method. *Journal of Forensic Sciences* 2018;63(1):214–220.

Lever J, Krzywinski M, Altman N. Points of significance: Model selection and overfitting. *Nature Methods* 2016;13(9):703–705.

Lewis A. Where bioethics meets machine ethics. *The American Journal of Bioethics* 2020;(11):22–24.

Li Y, Huang Z, Dong X, Liang W, Xue H, Zhang L, et al. Forensic age estimation for pelvic X-ray images using deep learning. *European Radiology* 2019;29:2322–2329.

Liu Y, Attinger D, De Brabanter K. Automatic classification of bloodstain patterns caused by gunshot and blunt impact at various distances. *Journal of Forensic Sciences* 2020;65(3):729–743.

Lo M, Mariconti E, Nakhaeizadeh S, Morgan RM. Preparing computed tomography images for machine learning in forensic and virtual anthropology. *Forensic Science International: Synergy* 2023;6:100319.

Love JC, Derrick SM, Wiersema JM, Peters C. Validation of tool mark analysis of cut costal cartilage. *Journal of Forensic Sciences* 2012;57(2):306–311.

Maples WR. The practical application of age-estimation techniques. In: Iscan MY, ed. *Age Markers in the Human Skeleton*. Springfield, IL: Charles C Thomas; 1989:319–324.

McBride DG, Dietz MJ, Vennemeyer MT, Benfer RA, Furbee NL, Meadors SA. Bootstrap methods for sex determination from the os coxae using the ID3 algorithm. *Journal of Forensic Sciences* 2001;46(3):427–431.

Medvedev I. Powered by AI: Instagram's Explore recommender system. Instagram Engineering. Last modified November 26, 2019. https://instagram-engineering.com/powered-by-ai-instagrams-explore-recommender-system-7ca901d2a882.

Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)* 2021;54(6):1–35.

Mesejo P, Martos R, Ibáñez Ó, Novo J, Ortega M. A survey on artificial intelligence techniques for biomedical image analysis in skeleton-based forensic human identification. *Applied Sciences* 2020;10(14):4703.

Moore MK, Schaefer E. A comprehensive regression tree to estimate body weight from the skeleton. *Journal of Forensic Sciences* 2011;56(5):1115–1122.

Morgan MB, Mates JL. Ethics of artificial intelligence in breast imaging. *Journal of Breast Imaging* 2023;5(2):195–200.

Nabi J. How bioethics can shape artificial intelligence and machine learning. *Hastings Center Report* 2018;48(5):10–13.

Nakhaeizadeh S, Dror IE, Morgan RM. Cognitive bias in forensic anthropology: Visual assessment of skeletal remains is susceptible to confirmation bias. *Science & Justice* 2014;54(3):208–214.

Nakhaeizadeh S, Hanson I, Dozzi N. The power of contextual effects in forensic anthropology: A study of biasability in the visual interpretations of trauma analysis on skeletal remains. *Journal of Forensic Sciences* 2014;59(5):1177–1183.

Nakhaeizadeh S, Dror IE, Morgan RM. The emergence of cognitive bias in forensic science and criminal investigations. *British Journal of American Legal Studies* 2015;4:527.

Nakhaeizadeh S, Dror IE, Morgan RM. Cognitive bias in sex estimation: The influence of context on forensic decision-making. In: Klales AR, ed. *Sex Estimation of the Human Skeleton*. London: Elsevier; 2020:327–342.

Nakhaeizadeh S, Morgan RM. Forensic anthropology and cognitive bias. In: Jamieson A, Moenssens A, eds. *Wiley Encyclopedia of Forensic Science*. Hoboken, NJ: Wiley-Blackwell; 2009:1–8.

Nakhaeizadeh S, Morgan RM, Olsson V, Arvidsson M, Thompson T. The value of eye-tracking technology in the analysis and interpretations of skeletal remains: A pilot study. *Science & Justice* 2020;60(1):36–42.

Nakhaeizadeh S, Morgan RM, Rando C, Dror IE. Cascading bias of initial exposure to information at the crime scene to the subsequent evaluation of skeletal remains. *Journal of Forensic Sciences* 2018;63(2):403–411.

Navega D, Coelho C, Vicente R, Ferreira MT, Wasterlain S, Cunha E. AncesTrees: Ancestry estimation with randomized decision trees. *International Journal of Legal Medicine* 2015;129:1145–1153.

Navega D, Vicente R, Vieira DN, Ross AH, Cunha E. Sex estimation from the tarsal bones in a Portuguese sample: A machine learning approach. *International Journal of Legal Medicine* 2015;129:651–659.

Nguyen A, Ngo HN, Hong Y, Dang B, Nguyen BT. Ethical principles for artificial intelligence in education. *Education and Information Technologies* 2023;28(4):4221–4241.

Nikita E, Nikitas P. On the use of machine learning algorithms in forensic anthropology. *Legal Medicine* 2020;47:101771.

Noseworthy PA, Attia ZI, Brewer LC, Hayes SN, Yao X, Kapa S, et al. Assessing and mitigating bias in medical artificial intelligence: The effects of race and ethnicity on a deep learning model for ECG analysis. *Circulation: Arrhythmia and Electrophysiology* 2020;13(3):e007988.

Nowotko PM. AI in judicial application of law and the right to a court. *Procedia Computer Science* 2021;192:2220–2228.

Ortega RF, Irurita J, Campo EJE, Mesejo P. Analysis of the performance of machine learning and deep learning methods for sex estimation of infant individuals from the analysis of 2D images of the ilium. *International Journal of Legal Medicine* 2021;135:2659–2666.

Ortiz AG, Costa C, da Silva RHA, Biazevic MGH, Michel-Crosato E. Sex estimation: Anatomical references on panoramic radiographs using machine learning. *Forensic Imaging* 2020;20:200356.

Ousley SD, Hollinger RE. The pervasiveness of Daubert. In: Dirkmaat D, ed. *A Companion to Forensic Anthropology*. Chichester: John Wiley & Sons; 2012:654–665.

Ousley SD, Jantz RL. Fordisc 3 and statistical methods for estimating sex and ancestry. In: In: Dirkmaat D, ed. *A Companion to Forensic Anthropology*. Chichester: John Wiley & Sons; 2012:311–329.

Pan W, Cui S, Wen H, Chen K, Zhang C, Wang F. Correcting the user feedback-loop bias for recommendation systems. arXiv preprint arXiv:2109.06037. 2021.

Pannucci CJ, Wilkins EG. Identifying and avoiding bias in research. *Plastic and Reconstructive Surgery* 2010;126(2):619.

Park Y, Hu J. Bias in artificial intelligence: Basic primer. *Clinical Journal of the American Society of Nephrology* 2023;18(3):394–396.

Passalacqua NV, Pilloud M. Education and training in forensic anthropology. *Forensic Anthropology* 2020;3(2):66.

Passalacqua NV, Pilloud MA. The need to professionalize forensic anthropology. *European Journal of Anatomy* 2021;25:35–47.

Passalacqua NV, Pilloud MA, Belcher WR. Scientific integrity in the forensic sciences: Consumerism, conflicts of interest, and transparency. *Science & Justice* 2019;59(5):573–579.

Patil V, Vineetha R, Vatsa S, Shetty DK, Raju A, Naik N, et al. Artificial neural network for gender determination using mandibular morphometric parameters: A comparative retrospective study. *Cogent Engineering* 2020;7(1):1723783.

Office of Science and Technology (OSTP). Blueprint for an AI bill of rights: Making automated systems work for the American people. 2022. www.whitehouse.gov/ostp/ai-bill-of-rights.

Rhine S. Non-metric skull racing. In: Gill GW, Rhine S, eds. *Skeletal Attribution of Race: Methods for Forensic Anthropology*. Albuquerque, NM: Maxwell Museum Anthropological Papers No. 4.; 1990:9–20.

Rossi F. Artificial intelligence: Potential benefits and ethical considerations. Policy Department: Citizens' Rights and Constitutional Affairs, European Parliament, PE 571.380. 2016. www.europarl.europa.eu/RegData/etudes/BRIE/2016/571380/IPOL_BRI(2016)571380_EN.pdf.

Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence* 2019;1(5):206–215.

Salkowitz R. Midjourney founder David Holz on the impact of AI on art, imagination and the creative economy. *Forbes*. 2022. https://www.forbes.com/sites/robsalkowitz/2022/09/16/midjourney-founder-david-holz-on-the-impact-of-ai-on-art-imagination-and-the-creative-economy/?sh=708b6ed22d2b.

Samek W, Wiegand T, Müller K. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. arXiv preprint arXiv:1708.08296. 2017.

Schwartz R, Vassilev A, Greene K, Perine L, Burt A, Hall P. Towards a standard for identifying and managing bias in artificial intelligence. NIST special publication 2022:1270 (10.6028). https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf.

Silva JT. Automated DNA extraction from frozen blood using the Abbott M1000 for HLA-DRB sequencing-based typing. *Human Immunology* 2003;64:95.

Skorburg JA, Sinnott-Armstrong W, Conitzer V. AI methods in bioethics. *AJOB Empirical Bioethics* 2020;11(1):37–39.

Smit NM, Lagnado DA, Morgan RM, Fenton NE. Using Bayesian networks to guide the assessment of new evidence in an appeal case. *Crime Science* 2016;5(1):1–12.

Solanke AA. Explainable digital forensics AI: Towards mitigating distrust in AI-based digital forensics analysis using interpretable models. *Forensic Science International: Digital Investigation* 2022;42:301403.

Spiros MC, Hefner JT. Ancestry estimation using cranial and postcranial macromorphoscopic traits. *Journal of Forensic Sciences* 2020;65(3):921–929.

Spiros MC, Plemons AM, Biggs JA. Pedagogical access and ethical considerations in forensic anthropology and bioarchaeology. *Science & Justice* 2022;62(6):708–720.

Spiros MC, Nakhaeizadeh S, Thompson TJU, Morgan RM, Arvidsson M, Olsson V, et al. Using eye-tracking technology to quantify the effect of experience and education on forensic anthropological analyses. *Forensic Anthropology* 2023;6(1): 25–35.

Stahl BC, Antoniou J, Bhalla N, Brooks L, Jansen P, Lindqvist B, et al. A systematic review of artificial intelligence impact assessments. *Artificial Intelligence Review* 2023;56:1–33.

Stevenson JC, Mahoney ER, Walker PL, Everson PM. Technical note: Prediction of sex based on five skull traits using decision analysis (CHAID). *American Journal of Physical Anthropology* 2009;139(3):434–441.

Swofford H, Champod C. Implementation of algorithms in pattern & impression evidence: A responsible and practical roadmap. *Forensic Science International: Synergy* 2021;3:100142.

Tai MC. The impact of artificial intelligence on human society and bioethics. *Tzu-Chi Medical Journal* 2020;32(4):339.

Takala T. What is wrong with global bioethics? On the limitations of the four principles approach. *Cambridge Quarterly of Healthcare Ethics* 2001;10(1):72–77.

Toneva D, Nikolova S, Agre G, Zlatareva D, Hadjidekov V, Lazarov N. Machine learning approaches for sex estimation using cranial measurements. *International Journal of Legal Medicine* 2021;135(3):951–966.

Tosam MJ. Global bioethics and respect for cultural diversity: How do we avoid moral relativism and moral imperialism? *Medicine, Health Care and Philosophy* 2020;23:611–620.

Toy S, Secgin Y, Oner Z, Turan MK, Oner S, Senol D. A study on sex estimation by using machine learning algorithms with parameters obtained from computerized tomography images of the cranium. *Scientific Reports* 2022;12(1):4278.

Trigueros DS, Meng L, Hartnett M. Generating photo-realistic training data to improve face recognition accuracy. *Neural Networks* 2021;134:86–94.

Turing AM. I. Computing machinery and intelligence. *Mind* 1950:LIX(236):433–460. doi: 10.1093/mind/LIX.236.433.

Turner L. From the local to the global: Bioethics and the concept of culture. *The Journal of Medicine and Philosophy* 2005; 30(3):305–320.

Vidaki A, Ballard D, Aliferi A, Miller TH, Barron LP, Syndercombe Court D. DNA methylation-based forensic age prediction using artificial neural networks and next generation sequencing. *Forensic Science International: Genetics* 2017; 28:225–236.

Vila-Blanco N, Carreira MJ, Varas-Quintana P, Balsa-Castro C, Tomas I. Deep neural networks for chronological age estimation from OPG images. *IEEE Transactions on Medical Imaging* 2020;39(7):2374–2384.

Vinchon F, Lubart T, Bartolotta S, Gironnay V, Botella M, Bourgeois-Bougrine S, et al. Artificial intelligence & creativity: A manifesto for collaboration. *The Journal of Creative Behavior*. 2023;57(4):472–484.

Warren MW, Friend AN, Stock MK. Navigating cognitive bias in forensic anthropology. In: Boyd CC Jr, Boyd DC, eds. *Forensic Anthropology: Theoretical Framework and Scientific Basis*. Hoboken, NJ: John Wiley; 2018;39–51.

Winburn AP, Jennings AL, Steadman DW, DiGangi EA. Ancestral diversity in skeletal collections: Perspectives on African American body donation. *Forensic Anthropology* 2020;5(2): 1–12.

Yang P, Baracchi D, Ni R, Zhao Y, Argenti F, Piva A. A survey of deep learning-based source image forensics. *Journal of Imaging* 2020;6(3):9.

Yapo A, Weiss J. Ethical implications of bias in machine learning. In: Proceedings of the 51st Hawaii International Conference on System Sciences, January 3-6 2018; Waikoloa Village, HI.

Yilmaz R, Erkaymaz O, Kara E, Ergen K. Use of autopsy to determine live or stillbirth: New approaches in decision-support systems. *Journal of Forensic Sciences* 2017;62(2):468–472.

Ying X. An overview of overfitting and its solutions. *Journal of Physics* 2019;1168(2):1–6.

Završnik A. Criminal justice, artificial intelligence systems, and human rights. *ERA Forum* 2020;20:567–583.

Zhang J, Zhang Z. Ethics and governance of trustworthy medical artificial intelligence. *BMC Medical Informatics and Decision Making* 2023;23(1):7.

Zhang K, Khosravi B, Vahdati S, Faghani S, Nugen F, Rassoulinejad-Mousavi SM, et al. Mitigating bias in radiology machine learning: 2. Model development. *Radiology: Artificial Intelligence* 2022;4(5):e220010.

Zhou Y, Zhang J, Huang J, Deng K, Zhang J, Qin Z, et al. Digital whole-slide image analysis for automated diatom test in forensic cases of drowning using a convolutional neural network algorithm. *Forensic Science International* 2019;302: 109922.

Zimmerman A, Janhonen J, Saadeh M, Castelyn C, Saxén H. Values in AI: Bioethics and the intentions of machines and people. *AI and Ethics* 2023;3(3):1003–1012.