Ethics and Information Technology (2022) 24:6
https://doi.org/10.1007/s10676-022-09621-6

ORIGINAL PAPER

# The emergence of "truth machines"?: Artificial intelligence approaches to lie detection

**Jo Ann Oravec**[1]

Accepted: 5 January 2022 / Published online: 24 January 2022


## Abstract

This article analyzes emerging artificial intelligence (AI)-enhanced lie detection systems from ethical and human resource (HR) management perspectives. I show how these AI enhancements transform lie detection, followed with analyses as to how the changes can lead to moral problems. Specifically, I examine how these applications of AI introduce human rights issues of fairness, mental privacy, and bias and outline the implications of these changes for HR management. The changes that AI is making to lie detection are altering the roles of human test administrators and human subjects, adding machine learning-based AI agents to the situation and establishing invasive data collection processes as well as introducing certain biases in results. I project that the potentials for pervasive and continuous lie detection initiatives ("truth machines") are substantial, displacing human-centered efforts to establish trust and foster integrity in organizations. I argue that if it is possible for HR managers to do so, they should cease using technologically-based lie detection systems entirely and work to foster trust and accountability on a human scale. However, if these AI-enhanced technologies are put into place by organizations by law, agency mandate, or other compulsory measures, care should be taken that the impacts of the technologies on human rights and wellbeing are considered. The article explores how AI can displace the human agent in some aspects of lie detection and credibility assessment scenarios, expanding the prospects for inscrutable, "black box" processes and novel physiological constructs (such as "biomarkers of deceit") that may increase the potential for such human rights concerns as fairness, mental privacy, and bias. Employee interactions with autonomous lie detection systems rather with than human beings who administer specific tests can reframe organizational processes and rules concerning the assessment of personal honesty and integrity. The dystopian projection of organizational life in which analyses and judgments of the honesty of one's utterances are made automatically and in conjunction with one's personal profile provides unsettling prospects for the autonomy of self-representation.

**Keywords** Lie detection · Self-representation · Autonomy · Honesty · Human resources · Human rights · Fairness · Mental privacy · Bias

Since the traditional polygraph emerged in the 1920s, lie detection systems have been construed as problematic by human resources (HR) administrators as well as by many courts and public policy analysts (Balmer, 2018; Bard, 2015). Some recent technological initiatives for the proposed improvement of lie detection focus on strategies that incorporate artificial intelligence (AI) approaches. In this article I show how these AI enhancements transform lie detection, followed with analyses as to how the changes can lead to moral problems. Specifically, I examine how these applications of AI introduce human rights issues of fairness, mental privacy, and bias and outline the implications of these changes for HR management. The changes that AI is making to lie detection are altering the roles of human test administrators and human subjects, adding machine learning-based AI agents to the situation and establishing invasive data collection processes as well as introducing certain biases in results. I project that the potentials for pervasive and continuous lie detection initiatives ("truth machines") are substantial, displacing human-centered efforts to establish trust and foster integrity in organizations. I argue that if it is possible for HR managers to do so, they should cease using lie detection systems entirely and work to foster trust

✉ Jo Ann Oravec
oravecj@uww.edu

[1] University of Wisconsin-Whitewater and –Madison, Whitewater, USA

and accountability on a human scale. However, if these AI-enhanced technologies are put into place by organizations by law, agency mandate, or other compulsory measures, care should be taken that the impacts of the technologies on human rights and wellbeing are monitored and considered.

In relation to HR, Singh and Doval (2019) declare in positive terms that AI "will automate time consuming, repetitive processes, enhance safety, eliminate hiring bias and further aid in training the hire" (p. 1). However, use of AI techniques in arenas with such sensitive implications for individuals as lie detection (dealing with subjects' personal integrity) can present various ethical as well as practical challenges in a growing assortment of organizational contexts. Consider the *EyeDetect* system, which "administers a 30-min test judging truthfulness based on a computer's observations of eye movement" (Melendez, 2018, para. 7). Its recent applications include the following, along with educational examination proctoring:

> Converus' technology, *EyeDetect*, has been used by FedEx in Panama and Uber in Mexico to screen out drivers with criminal histories, and by the credit-ratings agency Experian, which tests its staff in Colombia to make sure they aren't manipulating the company's database to secure loans for family members. In the U.K., police are carrying out a pilot scheme that uses *EyeDetect* to measure the rehabilitation of sex offenders. Other *EyeDetect* customers include the government of Afghanistan, McDonald's, and dozens of local police departments in the United States. (Katwala, 2019)

Output from *EyeDetect* was accepted as evidence by some courts (Melendez, 2018), though many judges have been reluctant participants in this arena.

I largely draw from US and UK examples in this article, but development of AI-enhanced lie detection technologies is growing in HR worldwide (Ayoub, 2018; Bergers, 2018). Alder (2009) writes of the "obsession" of the US with lie detection devices, but it is a passion increasingly shared with other nations, including China, which originally resisted the proliferation of lie detection technology (Zhang, 2011), and Germany (Fischer, 2020). Many uses of lie detection technologies in the US and other nations are restricted by law, but some applications have emerged in various police, military, and workplace contexts. This includes the post-conviction surveillance of sex offenders in the US and of potential sick leave falsifiers in other nations (Grubin et al., 2019; Kurland, 2019; Stathis & Marinakis, 2020); Mayoral et al. (2017) describe the use of lie detection technologies in theft investigation of employees in some US businesses. Voluntary uses of lie detection technologies are abundant in some workplace contexts, for example in attempts to support one's innocence if accused of a workplace malfeasance (Iacono & Patrick, 2018), which makes notions of transparency in processes and results especially relevant for HR managers.

## Varieties of AI-enhanced lie detection techniques

In this section I analyze the current and projected variations of AI-enhanced lie detection systems, after a brief examination of the lie detection technologies in place before the use of AI. Traditional polygraphy has played major roles in framing lie detection processes through the past decades, establishing a legacy as well as benchmarks for subsequent AI efforts. Polygraphy is "use of a physiological measurement apparatus with the explicit aim of identifying when someone is lying. This typically comes with specific protocols for questioning the subject, and the output is graphically represented" (Bergers, 2018, p. 1). The polygraph "measures galvanic skin response, blood pressure, heart and breathing rates, and perspiration as a proxy for nervous-system activity (primarily anxiety) as an (imperfect) proxy for deception" (Leonetti, 2017, p. 1). "Leakages" of various physiological cues (especially relating to the face and hands) can apparently signal increased levels of anxiety on the part of the subject relating to a particular topic but are not foolproof in providing the information needed for accurate lie detection (Burgoon, 2019; Denault & Dunbar, 2019). The requirement that individuals be physically strapped or otherwise attached to a lie detection apparatus has limited the variety of applications in which traditional polygraphy can play a part. However, the US Army's Preliminary Credibility Assessment Screening Systems (PCASS) are handheld polygraphs that are still in use for on-the-field lie detection efforts (Fuller, 2011; MacNeill & Bradley, 2016).

Below I describe assortments of emerging AI-enhanced approaches that are designed to overcome the obstacles in the kinds of lie detection directly performed by human agents and that require physical connection to the apparatus. AI technologies include a wide and growing assortment of methodologies, including pattern matching, profiling, and ontology construction (Domanski, 2019; Khatri, 2020), all of which are used in various lie detection applications. I contend that AI enhancements can potentially (1) shift the role of the human agent in relation to the subject of the investigation in favor of autonomous, robotic agents; (2) enable the remote and unannounced collection of subjects' data; (3) personalize lie detection analyses using big data-related profiling and surveillance techniques; (4) construct corpora of exemplars of "lying" so that machine learning devices can be trained; and (5) foster new varieties of multi-factored constructs and data mining routines related to human leakage and other physiological traces associated with lying. These various approaches can combine to facilitate the

development of perpetual and pervasive lie detection efforts. I provide some specifics below on how AI enhancement can change the character of lie detection initiatives:

*Role of the human agent*: With AI-enhanced systems, the human agent is often able to play a less obvious and visible role than with traditional polygraphs, changing the functions of the agent in lie detection efforts and presenting the potentials for more autonomous and less transparent lie detection (Gonzalez-Billandon et al., 2019). A number of skilled individuals may indeed be required to run the AI-enhanced system involved, but they generally do not play comparably direct roles with the subject than in previous kinds of systems.

*Remote, unobtrusive, and invasive collection of data*: New kinds of data collection devices and collection strategies are feasible with AI-enhanced system capabilities. One of the major concerns in many lie detection efforts is to reduce the potential for liars to evade detection through faking and coaching (Alliger & Dwight, 2000); with some of the AI-enhanced data collection systems, efforts at fakery are made more difficult because of the uncertainty about how, when, and what data are being collected. The modes for assimilating data for lie detection analysis have increasingly extended far beyond bulky sensors and also include instruments that collect data without the subject's close proximity or consent. For instance, such vehicles as wearable technologies, eye scanning, and webcams are being used to collect the data used for anti-deception initiatives (as with Converus Corporation's *EyeDetect*). Respiration rate detectors that do not require physical contact with subjects have also been developed (Prince et al., 2020). Other kinds of data sources are emerging: Maroulis (2014) outlines the potential for eye blinking patterns to be used in lie detection systems, and cognitive load considerations have been integrated into some systems in which the individuals' mental tasks are increased in ways that may reveal prevarication patterns (Bird et al., 2019; Stathis & Marinakis, 2020). Invasive approaches such as fMRI are also providing new, complex data sources that can require machine learning and big data analytical capabilities to interpret, potentially decreasing the transparency and openness of the systems involved (La Tona et al., 2020). Corporations have performed fMRI-based lie detection services for more than a decade (Moreno, 2009; Poldrack, 2018), although scientific support for their use is still emerging (Giattino et al., 2019).

*Profiling and the individuation of lie detection*: Profiling individuals (with the inclusion of demographic and behavioral information into AI analyses) has been utilized to improve lie detection (Singh, 2019). Predictive approaches can stem from such efforts to individuate (Kleinberg et al., 2019), presenting questions of whether the integrity-related behavior of individuals can (or should) be forecast. Accumulation of personalized "integrity scores" or other ways of profiling individuals over time in terms of their supposed propensity to lie has become a part of some recent research initiatives and technological development strategies in lie detection (Harding, 2019). Applications of the AI-enhanced methods and algorithms involved may indeed have particularly negative outcomes for individuals with certain demographic characteristics (as discussed in an upcoming section on bias); since these lie detection approaches are often used in security, wartime, and international border crossing contexts, such variations can be especially problematic in terms of human rights.

*Accumulating a "liar corpus"*: One of the recent approaches of AI researchers is to develop "corpora" of training examples for use in machine learning. For example, Takabatake et al. (2018) have constructed a "Liar Corpus" that collects for analysis various human expressions in situations that reportedly involve prevarication. Forms of bias can be introduced as items are selected for these training corpora that are skewed in various dimensions, such as in specific racial or gender directions (Hashemi & Hall, 2020; Tambe et al., 2019). HR managers can ask developers how the training corpora of their systems were compiled in order to mitigate potential problems, although training data are often generated through social media scraping, crowdsourcing, and other processes that can introduce bias in ways that may not be obvious even to developers.

*Developing lie detection-related constructs*: Another development in AI-enabled lie detection research is the crafting of complex constructs such as "micro-expressions" and "biomarkers of deceit" that would be difficult for those with limited technological support to utilize or challenge. In the case of micro-expressions, machine learning capabilities for analyzing large amounts of data about facial expressions have been designed to determine which subtle facial changes and combinations of physical cues are associated with lying. Barathi (2016) asserts that these supposedly unconscious micro-expressions are "involuntary reaction[s] that are impossible to fake" (p. 337) and are thus especially useful in lie detection efforts. Consider the following scenario involving silent talker, an early effort to incorporate AI into lie detection analysis:

> The Silent Talker consists of a digital video camera that is hooked up to a computer. It runs a series of programs called artificial neural networks… The camera records the subject in an interview and the artificial brain identifies non-verbal 'micro-gestures' on people's faces. These are unconscious responses that Silent Talker picks up on to determine if the interviewee is lying. Examples of micro-gestures include signs of stress, mental strain and what psychologists call 'duping delight'. This refers to the unconscious flash of a smile at the pleasure and thrill of getting

> away with telling a lie… One can imagine a near-future scenario… where every micro-gesture that "leaks" from your face is a response that flashes by [prospective employers'] eyes as "true" or "false" in real-time. (Kennedy, 2014, para. 5–8)

Some security and border control projects have recently segregated and labeled varieties of micro-expressions as "biomarkers of deceit," stirring some controversy and protest in part because of potential bias in their selection and implementation (Sánchez-Monedero, & Dencik, 2020).

I contend that many of these AI-related changes make the transparency and explicability of the lie detection initiatives more difficult for human audiences, creating forms of opaque "black boxes" (Pasquale, 2015). Various aspects of AI applications have been questioned as to their transparency, with algorithms and processes that are not readily interpretable for humans, especially in the realm of machine learning (Barn, 2019); for instance, the specific training data used by the systems (included in the corpus) are often unknown to the developer as well as the user. The physical and observable functionings of traditional polygraphs are being displaced by approaches that are seamlessly and often imperceptibly integrated into everyday workplace and community interactions, often to the detriment of transparency. Rules for building transparent and "trustworthy" AI (Floridi, 2019) are still emerging and basic security issues have yet to be resolved in many data capture and neuroscience arenas (Landau et al., 2020).

I argue that the AI enhancements described in this section have substantial impacts on decision making about lie detection. These initiatives have served to regenerate academic, corporate, police, and security interest in lie detection research and development as a whole, and also have apparently expanded the kinds of applications to which lie detection approaches can be integrated into everyday workplace settings (Heaven, 2018; Melendez, 2018) as well as airports and border crossings (Sánchez-Monedero & Dencik, 2020). For example, Bryant (2018) projects that such AI-enhanced technologies will "replace the polygraph" (para. 1). Issues of whether certain AI lie detection techniques are superior to the polygraph (which has served as a standard for lie detection for nearly a century) are common in evaluations of the systems in question (Meijer & Verschuere, 2017).

## Human rights concerns: fairness, mental privacy, and bias

In this section I address how AI-enhanced lie detection approaches and technologies present prospects that threaten psychological and social wellbeing with my efforts to link specific aspects of these approaches to fairness, mental privacy, and bias. The AI applications that result can have considerable implications for human rights (Aizenberg & van den Hoven, 2020) and can foster substantial concerns and tensions among employees, whether or not they can affect in significant ways employees' levels of honesty. I argue that lie detection introduces extraordinarily intimate and sensitive issues when primarily conducted with AI agents, who are often perceived as somehow "superintelligent" in their capabilities, and less easily challenged (Fischer, 2020; Natale, 2019; Poldrack, 2018). Being investigated for theft by a human interrogator who interprets human leakage for clues can be construed as substantially different from interactions with an AI-based agent, the output of which is less often questioned on critical concerns (Mayoral et al., 2017; Elkins et al., 2019). Fairness, mental privacy, and biases are among these emerging issues:

*Fairness*: I argue that a variety of forms of unfairness associated with AI-enhanced lie detection can diminish the autonomy of individuals and present human rights violations. For example, the prospects of being construed as guilty before having an opportunity to be proven innocent (with its associated unfairness) loom large in lie detection approaches that are rooted in autonomous and non-transparent processes in which the origins of the data involved cannot be inconclusively established. The use of individuated feedback and personalized profiles that calibrate some AI-enhanced lie detection devices has been linked with the notion of individuals "testifying against themselves" (Räikkä, 2017), triggering calls to expand the "right of silence" to AI-driven interrogation efforts (Thomasen, 2016). McAllister (2016) describes AI-driven questioning and interviewing as "stranger than science fiction" (p. 2527), requiring international discussions and agreements concerning human rights.

*Mental privacy*: Mental privacy deals with "people's right and ability to keep private what they think and feel" (Royakkers et al., 2018, p. 130). Many of the AI-enhanced lie detection systems described in this article have generated mental privacy concerns in relation to their data collection approaches (Wright, 2018). For example, the remote lie detection data collection initiatives I described in a previous section raise knotty issues about surreptitious data collection procedures and can complicate related organizational efforts to obtain informed consent. Brain scanning presents new concerns as well in this arena, imposing invasive data collection: the prospect that one's supposedly-private mental processes will be open to forms of scanning as an aspect of one's employment situation provides challenges to human rights (Burgoon, 2019; Farrell, 2009). These processes have the prospect to infringe on the autonomy of individuals' self-representations (Van den Hoven & Manders-Huits, 2008), with the subjects involved not having control or even knowledge of how their thoughts are being represented.

Mental privacy plays roles in human rights in affording individuals with adequate space to manifest personal autonomy and express themselves adequately in various situations. Mental privacy can also be construed as having organizational-level paybacks as well as benefits for employees, fostering the development of autonomous individuals capable of critical thinking. Some analysts have identified the "sanctity of the mind" (Reiner & Nagel, 2017, p. 108) as an important notion to defend for the purposes of reinforcing individual autonomy. Despite the dangers involved to human rights, many researchers are still apparently drawn to the "seductive allure" of neurotechnology and related AI-enhanced lie detection efforts in real-life organizational applications (Giattino et al., 2019, p. 397).

*Bias*: The problem of bias has been associated with an assortment of AI-enhanced systems, including facial recognition as well as lie detection (Bacchini & Lorusso, 2019); the quality of training data has been identified as one of the primary ways that AI-enhanced lie detection systems can produce biased results, although the machines can be faulty because of intentional misprogramming and other causes. Zou and Schiebinger (2018) state that "Most machine-learning tasks are trained on large, annotated data sets… such methods can unintentionally produce data that encode gender, ethnic and cultural biases" (p. 325). These data sets are often scraped from various social media and other internet sources, generally by outsourcers; HR managers may not be able to ascertain the quality of the data utilized. The kinds of biases that have been associated with some AI implementations (such as racial, gender, or disability-related skewing due to inappropriate choice of training data) could indeed have impacts upon how lie detection and credibility assessment systems are designed and implemented (Domanski, 2019; Trewin, 2019). Profiles of individuals that are built on these biased results can compound the damages associated with the biases. Efforts to eradicate system-imposed biases and isolate the damages involved can also be complicated by deficits in transparency in machine learning systems, so that debugging of the systems for potential problems is difficult if not impossible in some contexts.

In recommending that the use of AI-enhanced lie detection systems be ceased, I recognize that lie detection processes have often been problematic through the centuries, as well as directly associated with inhumane practices. Human interrogators have utilized such extreme and physically damaging measures as torture, sleep deprivation, and truth serums to elicit supposedly truthful statements and aid in the detection of lies (Alder, 2009; Winter, 2005). I argue that the damages involved in using AI-enhanced lie detection technologies described in this article may not involve physical pain but can result in the kinds of reputational and psychological harms that can have lasting impacts on an individual. I also recognize that some comparable harms can result from use of traditional polygraphs (such as unfair implementation), and that polygraphs also should be removed from organizations, as they already are in many contexts. Just construing and redeveloping lie detection technologies in terms of AI does not make the technologies more appropriate and humane.

## Future AI-related directions in lie detection research and applications

In this section, I project how the potential for perpetual, autonomously-controlled lie detection systems (or "truth machines") to become part of some organizational practices looms large for the foreseeable future. Many organizations have been damaged by extensive and uncontrolled prevarication among their participants in critical venues (Comer & Stephens, 2017; Noonan, 2018; Walczyk et al., 2005); some have looked to HR management for guidance in mitigating current or potential problems. Organizational insiders who misrepresent data for their personal gain can create problems for organizations far exceeding those fomented by malicious outsiders (Mecke, 2007). The social and cultural backings for lie detection technologies have varied in intensity, but their long roots in favorable film, television, and science fiction depictions of polygraphs and related technologies have had a sustained influence over time (Bunn, 2019); association of lie detection with AI has served to provide additional public support in some contexts (Pasquali et al., 2020).

I contend that enthusiasm for AI-enhanced approaches as potential solutions to these honesty-related problems can affect the judgment of researchers and practitioners toward the resulting systems. For example, research on potential neuroscientific lie detection applications has often been presented with an optimistic tone (La Tona et al., 2020; Meijer & Verschuere, 2017), with confident assessments including "One day cognitive neuroscientists might perform the magic of accurate mind reading" (Moreno, 2009, p. 737). There is a temptation to evaluate lie detection and cognitive engineering efforts in ways that are readily challenged but that are deemed acceptable because of the perceived security and economic implications that successful technological applications might entail (Strle & Markič, 2019); for example, Schauer asks in relation to lie detection approaches "can bad science be good evidence?" (2009, p. 1191). The capabilities for evaluating lies and assessing credibility that these emerging AI-enhanced technologies could provide may indeed engender radical changes in how organizations recruit, engage, and evaluate individuals. For instance, some neuroscientific approaches are working to expand the range of lie detection and even move toward cognitive engineering, in which the ways that individuals think in everyday contexts could be considerably influenced (Darby & Pascual-Leone,

2017). Maréchal et al. (2017) proposes ways to "increase honesty in humans with noninvasive brain stimulation," thus reducing the need for lie detection by decreasing the propensity to lie.

As I have described in this article, many AI-related research and development efforts are in use today despite the fact that they are in the early stages of testing and evaluation (Bittle, 2020). Some expressions of skepticism about the value of AI-enhanced lie detection are also emerging in security studies and legal research as scientific support for its use is often spotty (Jupe & Keatley, 2019). Some commentators identify lie detection as "little more than a racket" (Stroud, 2019). Along comparable lines, Laws (2020) characterizes lie detection efforts as "the bogus pipeline to the soul" and Fischer (2020) describes them as often akin to the mind reading tricks of magicians. Objective evaluation of credibility assessment systems may be complicated by the attitudes toward the systems (including fears of being displaced) of influential experts, as shown by Elkins et al. (2013). Increases in concern about the reliability of AI-enhanced systems may serve to expand pressures on organizations that use them with little scientific support and without safeguards (Dafoe, 2018). Evaluating the overall impact of new forms of lie detection systems requires time and perspective; unfortunately, many of the specific technologies involved have relatively short shelf-lives, with new approaches and developers emerging in quick succession.

## Approaches to containing and mitigating lie detection concerns

In this article I have argued that systems that are based on building human-centered integrity and trust are preferable to AI-enhanced lie detection systems, which are often lacking in transparency and fairness. However, many HR managers are faced in some compulsory manner with administering systems in which AI technologies are utilized for lie detection (such as in some security, classified materials, and police contexts). The following approaches may aid HR managers in identifying, containing, and mitigating the moral and human rights problems involved with these technologies:

*Providing training in various situational contexts*: HR managers and administrators, given adequate acclimation and training, could indeed help convey to employees some of the affordances and limitations of the AI-enhanced lie detection systems being utilized in organizational contexts. In this manner, systems implementations would incorporate more realistic notions of what is going on with AI-enabled lie detection as well as establish the infrastructure for obtaining informed consent. Implementing this strategy in HR contexts without substantial investment in staff training may have dangers, though. For instance, Masip et al. (2020) describe the difficulties of teaching students about basic lie detection issues and variations, and Khatri et al. (2020) characterize the "disruptive" impacts of AI on organizations with an emphasis on staff training. The technical skills involved in interpreting analyses of the validity of lie detection technology with precision are often lacking among the individuals who would be administering lie detection systems in organizational or agency contexts (Ben-Shakhar & Barr, 2018), possibly resulting in biases in implementation and interpretation. As a complicating factor, lie detection technologies are often utilized in tense, multi-dimensional security situations in which many issues and emotions are intertwined (Nahari et al., 2019); the uncertainties involved can enhance the challenges of system implementation.

*Obtaining and assessing critically specific forms of scientific support*: Accumulation of forms of scientific support (as well as critiques) for lie detection technologies is of substantial importance for legal and ethical purposes, and AI-enhanced lie detection systems present extraordinary challenges in this regard. Of special significance here is the transparency of operation of the devices involved (Watson & Nations, 2019). Although a "trick" or bogus lie detector can certainly be designed and sometimes utilized in practice (Laws, 2020), the physical connection of the subject with the machine provides at least some apparent visual support that the data involved were indeed at some point in time associated with the subject. AI-enhanced systems that collect data remotely and that are rooted in complex constructs (such as "biomarkers of deceit") can produce results that have less transparent and obvious connection to the subject with whom they are associated. New and complex issues of security also arise with neuroscientific initiatives in cognitive modification, with the potential for intricate cognitive interventions that could have unforeseen side effects (Landau et al., 2020; Poldrack, 2018). Other issues involve the choice of benchmarks for evaluation. Some evaluation contexts for lie detection compare results with conviction rates, possibly encountering societal bias concerns (Garrett, 2020) since conviction rates can vary significantly on racial and gender dimensions. Reducing the number of false positives should be a high priority for ethical organizations since being falsely accused of a lack of integrity can be highly traumatic for individuals as well as damaging to others involved in the processes. Also, many of the testing efforts associated with lie detection involve actors assuming the roles of liars, since the notion of obtaining "real liars" in particular experimental contexts is problematic (Burgoon, 2019); choice of actors involved can introduce bias.

*Reducing or containing AI "hype"*: The dangers of labeling technological applications as involving "AI" or "big data" without clarifying the impact of these approaches on organizational practices could increase employees' fears

and misconceptions about lie detection systems while not facilitating their uses in appropriate ways. Research on AI-enhanced lie detection is often characterized in ambitious terms: "your eyes never lie: a robot magician can tell if you are lying" (Pasquali et al., 2020, p. 392). Whether or not AI-enhanced lie detection devices indeed manifest the kinds of capabilities that are linked with them by developers, the uncertainties involved in their workplace applications could undermine the fairness- and transparency-related considerations of HR managers. The association of AI with supernatural and fantastic powers (which began in science fiction decades ago) has apparently grown stronger with the personalization initiatives that underpin the workings of many AI systems, often resulting in the assumption that "Amazon can read your mind" (Natale, 2019, p. 19). Exaggerated estimates of AI capabilities could play a role in deterrence: the deterrent effects of establishing lie detection systems have been shown to be consequential, as described in Peleg et al. (2019) and Witt and Neller (2018). Individuals' assumptions that AI-enhanced lie detection systems are somehow "superintelligent mind readers" could indeed be an aspect of deterrence, though they can also be problematic as managers attempt to inspire employee trust in the systems and in the organization as a whole.

*Limiting the use of large-scale and perpetual lie detection implementations*: Organizations that integrate lie detection systems into their operations on a large-scale incur social responsibilities that can be complex and difficult to address, since maintaining a culture of "automated honesty" can compete with institutional efforts supporting individual autonomy and mental privacy. Dystopian outcomes may indeed occur: with AI-enhanced capabilities, perpetual lie detection, in contrast with more context-sensitive and event-driven varieties, could be conducted in a covert manner and its results used in opportunistic ways to target certain individuals. Establishment of autonomous, AI-enhanced lie detection apparatuses that are widely implemented could replace the polygraph technician in a white coat with a pervasive, all-seeing presence. HR managers should work to eliminate both the AI-enhanced systems and the polygraph whenever possible.

In this section I characterized the difficult challenges many HR managers face in the compulsory use of lie detection technologies and endeavored to present some specific containment and mitigation strategies. HR managers are often called upon to clarify multifaceted situations in such arenas as hiring, security, and inventory control that can involve issues of honesty, and in some contexts (especially security and police work) the use of lie detection technologies is not likely to end soon (Vissak & Vadi, 2013). Despite the human rights challenges described in this article, Leonetti (2017) relates how widely-accepted AI-enhanced lie detection technology is still the "holy grail" of many corporations, military organizations, and security agencies, presenting the promise of a future in which organizations can mitigate the damage associated with lying. Katwala (2019) describes the "race to create a perfect lie detector" in organizational settings as incorporating AI approaches. Allan Dafoe (2018), in a report issued by the Oxford University's Future of Humanity Institute, identified the dangers of such lie detection efforts in the following stark terms: "robust totalitarianism could be enabled by advanced lie detection, social manipulation, autonomous weapons, and ubiquitous physical sensors and digital footprints" (p. 7). The organizational mandate for individuals to conform to the lie detection system's perceived requirements could unfortunately create severe anxieties as well as displace many creative and exploratory cognitive initiatives that may ultimately be of value to organizations.

## Conclusion

In this article, I have shown how AI enhancements are currently changing lie detection as well as how (with pervasive and perpetual lie detection systems) they may further transform it in the future. In addressing the human rights-related changes associated with recent lie detection technologies, I outlined how prospects for unfairness, bias, and violations of mental privacy are increased by many of the emerging AI-related developments, providing special challenges to HR managers in maintaining organizational transparency and trust. I argue that eliminating use of the systems entirely is the preferable approach to dealing with lie detection, given these human rights issues. However, in circumstances in which the use of the technologies is legally stipulated or compulsory through agency mandate, I have shown that these AI-related changes can lead to moral problems that in some small ways can be contained and mitigated with the vigilant efforts of HR managers. I have shown that remote data collection, moral neuroenhancement, and individuated integrity scores and profiles are continuing to expand the technological approaches of lie detection. These applications of AI-enhanced lie detection and credibility assessment technologies in HR management are just emerging, but have the potential to foment significant and potentially problematic transformations in workplaces. The unwritten agreements and understandings that bind individuals in organizations are increasingly including AI-enhanced agents, opaque entities that are not well understood by either their subjects or the HR managers who implement them. Hype and misunderstandings concerning AI capabilities could also play roles in distorting the human subject's perceptions of the lie detection processes involved, and possibly influence the deterrent capabilities of the systems as well. AI applications are certainly not omniscient, and machine learning systems

have biases based on how they are trained, so assumptions that the systems are without blemishes can be problematic.

Despite many technological and social advances through the decades, HR researchers and practitioners have not yet settled on a particular way to facilitate individuals in telling the truth, whether by using technologically-enhanced lie detection tests or using various kinds of threats and deterrence methods. I contend that an ideal scenario for HR managers would have organizations eliminating the use of lie detection technologies entirely, moving from forms of "automated honesty" to the building of trust and mutual respect among participants. In settings where technologically-supported lie detection is seen as a necessary factor by legal and administrative authorities, HR managers will need to make tough decisions concerning the validity and reliability of various AI-enhanced approaches. I have argued in this article that the pervasive or autonomous detection of lying may indeed free up HR staff to engage in other kinds of efforts, but will also introduce serious kinds of uncertainties and human rights challenges such as those relating to fairness, mental privacy, and bias.

## References


Aizenberg, E., & van den Hoven, J. (2020). Designing for human rights in AI. *Big Data & Society, 7*(2), 1–14. https://doi.org/10.1177/2053951720949566

Alder, K. (2009). *The lie detectors: The history of an American obsession*. University of Nebraska Press.

Alliger, G. M., & Dwight, S. A. (2000). A meta-analytic investigation of the susceptibility of integrity tests to faking and coaching. *Educational and Psychological Measurement, 60*(1), 59–72.

Ayoub, A., Rizvi, F., Akram, S., & Tahir, M. A. (2018). The polygraph and lie detection: A case study. *Arab Journal of Forensic Sciences & Forensic Medicine, 1*(7), 902–908.

Bacchini, F., & Lorusso, L. (2019). Race, again: How face recognition technology reinforces racial discrimination. *Journal of Information, Communication and Ethics in Society, 17*(3), 321–335. https://doi.org/10.1108/JICES-05-2018-0050

Balmer, A. (2018). *Lie detection and the law: Torture, technology and truth*. Routledge.

Barathi, C. S. (2016). Lie detection based on facial micro expression, body language, and speech analysis. *International Journal of Engineering Research & Technology, 5*(2), 337–343.

Bard, J. S. (2015). Ah yes, I remember it well: Why the inherent unreliability of human memory makes brain imaging technology a poor measure of truth-telling in the courtroom. *Oregon Law Review, 94*, 295–332.

Barn, B. S. (2019). Mapping the public debate on ethical concerns: Algorithms in mainstream media. *Journal of Information, Communication and Ethics in Society., 18*(1), 124–139. https://doi.org/10.1108/JICES-04-2019-0039

Ben-Shakhar, G., & Barr, M. (2018). *Science, pseudo-science, non-sense, and critical thinking: Why the differences matter*. Routledge.

Bergers, L. (2018). Only in America? A history of lie detection in the Netherlands in comparative perspective, ca. 1910–1980. Master's thesis, Utrecht University, The Netherlands

Bird, L., Gretton, M., Cockerell, R., & Heathcote, A. (2019). The cognitive load of narrative lies. *Applied Cognitive Psychology, 33*(5), 936–942. https://doi.org/10.1002/acp.3567

Bittle, J. (2020). Lie detectors have always been suspect. AI has made the problem worse. *Technology Review.* https://www.technologyreview.com/2020/03/13/905323/ai-lie-detectors-polygraph-silent-talker-iborderctrl-converus-neuroid/. Accessed 16 Jan 2022

Bryant, P. (2018). Will eye scanning technology replace the polygraph. *Government Technology*. Retrieved from http://www.govtech.com/public-safety/Will-Eye-Scanning-Technology-Replace-the-Polygraph.html. Accessed 16 Jan 2022

Bunn, G. C. (2019). "Supposing that truth is a woman, what then?": The lie detector, the love machine, and the logic of fantasy. *History of the Human Sciences, 32*(5), 135–163.

Burgoon, J. K. (2019). Separating the wheat from the chaff: Guidance from new technologies for detecting deception in the courtroom. *Frontiers in Psychiatry, 9*, 774–780. https://doi.org/10.3389/fpsyt.2018.00774

Comer, M. J., & Stephens, T. E. (2017). *Deception at work: Investigating and countering lies and fraud strategies*. Routledge.

Dafoe, A. (2018). *AI governance: A research agenda*. University of Oxford.

Darby, R. R., & Pascual-Leone, A. (2017). Moral enhancement using non-invasive brain stimulation. *Frontiers in Human Neuroscience, 11*, 77. https://doi.org/10.3389/fnhum.2017.00077

Denault, V., & Dunbar, N. E. (2019). Credibility assessment and deception detection in courtrooms: Hazards and challenges for scholars and legal practitioners. *The Palgrave handbook of deceptive communication* (pp. 915–935). Palgrave Macmillan.

Domanski, R. (2019). *The AI Pandorica: Linking ethically-challenged technical outputs to prospective policy approaches* (pp. 409–416). Association for Computing Machinery.

Elkins, A. C., Dunbar, N. E., Adame, B., & Nunamaker, J. F. (2013). Are users threatened by credibility assessment systems? *Journal of Management Information Systems, 29*(4), 249–262. https://doi.org/10.2753/MIS0742-1222290409

Elkins, A. C., Gupte, A., & Cameron, L. (2019). *Humanoid robots as interviewers for automated credibility assessment* (pp. 316–325). Springer.

Farrell, B. (2009). Can't get you out of my head: The human rights implications of using brain scans as criminal evidence. *Interdisciplinary Journal of Human Rights Law, 4*, 89–95.

Fischer, L. (2020). The idea of reading someone's thoughts in contemporary lie detection techniques. *Mind reading as a cultural practice* (pp. 109–137). Palgrave Macmillan.

Floridi, L. (2019). Establishing the rules for building trustworthy AI. *Nature Machine Intelligence, 1*(6), 261–262.

Fuller, C., Biros, D., & Delen, D. (2011). An investigation of data and text mining methods for real world deception detection. *Expert Systems with Applications, 38*, 8392–8398.

Garrett, B. L. (2020). Wrongful convictions. *Annual Review of Criminology*, *3*, 245–259.

Giattino, C. M., Kwong, L., Rafetto, C., & Farahany, N. A. (2019). *The seductive allure of artificial intelligence-powered neurotechnology* (pp. 397–402). Association for Computing Machinery (ACM).

Gonzalez-Billandon, J., Aroyo, A., Pasquali, D., Tonelli, A., Gori, M., Sciutti, A., Gori, M., Sandini, G., & Rea, F. (2019). Can a robot catch you lying? A machine learning system to detect lies during interactions. *Frontiers in Robotics and AI, 6*(64), 1–12. https://doi.org/10.3389/frobt.2019.00064

Grubin, D., Kamenskov, M., Dwyer, R. G., & Stephenson, T. (2019). Post-conviction polygraph testing of sex offenders. *International Review of Psychiatry., 31*(2), 141–148.

Harding, C. D. (2019). *Selecting the ethical employee: Measuring personality facets to predict integrity behavior*. Carleton University.

Hashemi, M., & Hall, M. (2020). Criminal tendency detection from facial images and the gender bias effect. *Journal of Big Data, 7*(1), 1–16.

Heaven, D. (2018). AI to interrogate travellers. *New Scientist, 240*(3202), 5.

Iacono, W. G., & Patrick, C. J. (2018). Assessing deception. In R. Rogers & S. D. Bender (Eds.), *Clinical assessment of malingering and deception.* Guilford Publications.

Jupe, L. M., & Keatley, D. A. (2019). Airport artificial intelligence can detect deception: Or am I lying? *Security Journal., 24*, 1–4.

Katwala, A. (2019). The race to create a perfect lie detector- and the dangers of succeeding. *The Guardian*. Retrived from https://www.theguardian.com/technology/2019/sep/05/the-race-to-create-a-perfect-lie-detector-and-the-dangers-of-succeeding. Accessed 16 Jan 2022

Kennedy, P. (2014). Artificial intelligence lie detector developed by imperial alumnus. *Imperial College London.* Retrived from https://www.imperial.ac.uk/news/144486/artificial-intelligence-detector-developed-imperial-alumnus/. Accessed 16 Jan 2022

Khatri, S., Pandey, D. K., Penkar, D., & Ramani, J. (2020). Impact of artificial intelligence on human resources. In D. Management (Ed.), *Analytics and innovation* (pp. 365–376). Springer.

Kleinberg, B., Arntz, A., & Verschuere, B. (2019). Detecting deceptive intentions: Possibilities for large-scale applications. *The Palgrave handbook of deceptive communication* (pp. 403–427). Palgrave Macmillan.

Kurland, J. (2019). Truth-detection devices and victims of sexual violence. *Family & Intimate Partner Violence Quarterly, 11*(4), 39–44.

La Tona, G., Terranova, M. C., Vernuccio, F., Re, G. L., Salerno, S., Zerbo, S., & Argo, A. (2020). Lie detection: fMRI. *Radiology in forensic medicine* (pp. 197–202). Springer.

Landau, O., Puzis, R., & Nissim, N. (2020). Mind your mind: EEG-based brain-computer interfaces and their security in cyber space. *ACM Computing Surveys (CSUR), 53*(1), 1–38.

Laws, D. R. (2020). *A history of the assessment of sex offenders: 1830–2020*. Emerald Publishing Limited.

Leonetti, C. (2017). Abracadabra, hocus pocus, same song, different chorus: The newest iteration of the science of lie detection. *Richmond Journal of Law & Technology., 24*(1), 1–35.

MacNeill, A. L., & Bradley, M. T. (2016). Temperature effects on polygraph detection of concealed information. *Psychophysiology, 53*(2), 143–150.

Maréchal, M. A., Cohn, A., Ugazio, G., & Ruff, C. C. (2017). Increasing honesty in humans with noninvasive brain stimulation. *Proceedings of the National Academy of Sciences, 114*(17), 4360–4364.

Maroulis, A. (2014). *Blinking in deceptive communication*. State University of New York at Buffalo.

Masip, J., Levine, T. R., Somastre, S., & Herrero, C. (2020). Teaching students about sender and receiver variability in lie detection. *Teaching of Psychology, 47*(1), 84–91.

Mayoral, L. P. C., Mayoral, E. P. C., Andrade, G. M., Mayoral, C. P., Helmes, R. M., & Pérez-Campos, E. (2017). The use of polygraph testing for theft investigation in private sector institutions. *Polygraph, 46*(1), 44–52.

McAllister, A. (2016). Stranger than science fiction: The rise of AI interrogation in the dawn of autonomous robots and the need for an additional protocol to the UN convention against torture. *Minnesota Law Review, 101*, 2527–2573.

Mecke, J. (2007). *Cultures of lying: Theories and practice of lying in society, literature, and film*. Galda & Wilch.

Meijer, E. H., & Verschuere, B. (2017). Deception detection based on neuroimaging: Better than the polygraph? *Journal of Forensic Radiology and Imaging, 8*, 17–21.

Melendez, S. (2018). Goodbye polygraphs: New tech uses AI to tell if you're lying. *Fast Company*. Retrieved from https://www.fastcompany.com/40575672/goodbye-polygraphs-new-tech-uses-ai-to-tell-if-youre-lying. Accessed 16 Jan 2022

Moreno, J. A. (2009). The future of neuroimaged lie detection and the law. *Akron Law Review, 42*, 717–737.

Nahari, G., Ashkenazi, T., Fisher, R. P., Granhag, P. A., Hershkowitz, I., Masip, J., Meijer, E. H., Nisin, Z., Sarid, N., Taylor, P. J., Vrii, A., & Verschuere, B. (2019). 'Language of lies': Urgent issues and prospects in verbal lie detection research. *Legal and Criminological Psychology, 24*(1), 1–23. https://doi.org/10.1111/lcrp.12148

Natale, S. (2019). Amazon can read your mind: A media archaeology of the algorithmic imaginary. In S. Natale & D. Pasulka (Eds.), *Believing in bits: Digital media and the supernatural* (pp. 19–36). Oxford University Press.

Noonan, C. F. (2018). *Spy the lie: Detecting malicious insiders (No. PNNL-SA-122655)*. Pacific Northwest National Lab (PNNL).

Pasquale, F. (2015). *The black box society*. Harvard University Press.

Pasquali, D., Aroyo, A. M., Gonzalez-Billandon, J., Rea, F., Sandini, G., & Sciutti, A. (2020). *Your eyes never lie: A robot magician can tell if you are lying* (pp. 392–394). ACM.

Peleg, D., Ayal, S., Ariely, D., & Hochman, G. (2019). The lie deflator-the effect of polygraph test feedback on subsequent (dis) honesty. *Judgment & Decision Making, 16*(6), 728–738.

Poldrack, R. A. (2018). *The new mind readers: What neuroimaging can and cannot reveal about our thoughts*. Princeton University Press.

Prince, P. G., Rajkumar, R. I., & Premalatha, J. (2020). Novel non-contact respiration rate detector for analysis of emotions. In D. J. Hemanth (Ed.), *Human behaviour analysis using intelligent systems* (pp. 157–178). Springer.

Räikkä, J. (2017). Privacy and self-presentation. *Res Publica, 23*(2), 213–226.

Reiner, P. B., & Nagel, S. K. (2017). Technologies of the extended mind: defining the issues. *Neuroethics: Anticipating the future* (pp. 108–122). Oxford University Press.

Royakkers, L., Timmer, J., Kool, L., & van Est, R. (2018). Societal and ethical issues of digitization. *Ethics and Information Technology, 20*(2), 127–142.

Sánchez-Monedero, J., & Dencik, L. (2020). The politics of deceptive borders: "Biomarkers of deceit" and the case of iBorderCtrl. *Information, Communication & Society.* https://doi.org/10.1080/1369118X.2020.1792530

Schauer, F. (2009). Can bad science be good evidence? Neuroscience, lie detection, and beyond. *Cornell Law Review, 95*(6), 1191–1219.

Singh, E., & Doval, J. (2019). Artificial intelligence and HR: Remarkable opportunities, hesitant partners. In *Proceedings of the 4th National HR Conference on Human Resource Management Practices and Trends*. Retrived from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3553448. Accessed 16 Jan 2022

Singh, R. (2019). Profiling and its facets. In R. Singh (Ed.), *Profiling humans from their voice* (pp. 3–26). Springer.

Stathis, M. J., & Marinakis, M. M. (2020). Shadows into light: The investigative utility of voice analysis with two types of online child-sex predators. *Journal of Child Sexual Abuse*. https://doi.org/10.1080/10538712.2019.1697780

Strle, T., & Markič, O. (2019). Looping effects of neurolaw, and the precarious marriage between neuroscience and the law. *Balkan Journal of Philosophy, 10*(1), 17–26.

Stroud, M. (2019). *Thin blue lie: The failure of high-tech policing*. New York: Metropolitan Books.

Takabatake, S., Shimada, K., & Saitoh, T. (2018). *Construction of a liar corpus and detection of lying situations* (pp. 971–976). IEEE Press.

Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review, 61*(4), 15–42.

Thomasen, K. (2016). *Examining the constitutionality of robot-enhanced interrogation*. Edward Elgar Publishing.

Trewin, S., Basson, S., Muller, M., Branham, S., Treviranus, J., Gruen, D., Hebert, D., Lyckowski, N., & Manser, E. (2019). Considerations for AI fairness for people with disabilities. *AI Matters, 5*(3), 40–63. https://doi.org/10.1145/3362077.3362086

Van den Hoven, J., & Manders-Huits, N. (2008). The person as risk, the person at risk'. *ETHICOMP 2008: Living working and learning beyond technology* (pp. 408–14). SAGE.

Vissak, T., & Vadi, M. (2013). *(Dis) honesty in management: Manifestations and consequences*. Emerald Group Publishing.

Walczyk, J. J., Schwartz, J. P., Clifton, R., Adams, B., Wei, M. I. N., & Zha, P. (2005). Lying person-to-person about life events: A cognitive framework for lie detection. *Personnel Psychology, 58*(1), 141–170. https://doi.org/10.1111/j.1744-6570.2005.00484.x

Watson, H. J., & Nations, C. (2019). Addressing the growing need for algorithmic transparency. *Communications of the Association for Information Systems, 45*(1), 26. https://doi.org/10.17705/1CAIS.04526

Winter, A. (2005). The making of "truth serum". *Bulletin of the History of Medicine*, *79*(3), 500–533.

Witt, P. H., & Neller, D. J. (2018). Detection of deception in sex offenders. In R. Rogers & S. D. Bender (Eds.), *Clinical assessment of malingering and deception* (pp. 401–421). The Guilford Press.

Wright, E. (2018). The future of facial recognition is not fully known: Developing privacy and security regulatory mechanisms for facial recognition in the retail sector. *Fordham Intellectual Property Media & Entertainment Law Journal, 29*(2), 611–685.

Zhang, X. (2011). The evolution of polygraph testing in the People's Republic of China. *Polygraph, 40*(3), 181–193.

Zou, J., & Schiebinger, L. (2018). AI can be sexist and racist—it's time to make it fair. *Nature, 559*, 324–326.