JOURNAL OF APPLIED SCIENCE AND SOCIAL SCIENCE
eISSN 2229-3113 pISSN 2229-3205
Volume 15 Issue 10, October 2025
Impact factor: 2019: 4.679 2020: 5.015 2021: 5.436, 2022: 5.242, 2023: 6.995, 2024 7.75

DETECTING THREATS TO HUMAN HEALTH AND SAFETY IN REAL-WORLD SITUATIONS USING ARTIFICIAL INTELLIGENCE: OPPORTUNITIES, CHALLENGES AND FUTURE DIRECTIONS

Azimov Sarvarbek Ravshanbek ugli
Assistant professor, Department of General Technical Sciences, Asia International University

Annotation: In an era of rapid technological advancement, artificial intelligence (AI) is increasingly being harnessed in healthcare, occupational safety and public health to identify and mitigate threats to human health and safety in real-world situations. This article reviews current applications of AI for threat detection across clinical, workplace and public domains, analyses key benefits and risks, and outlines a conceptual framework for deploying AI in threat detection systems. We highlight major advances such as predictive analytics for adverse events, real-time monitoring of occupational hazards, and anomaly detection in health data. Simultaneously, we examine challenges including data bias, transparency, regulatory gaps, privacy concerns and system integration in complex settings. The paper proposes guidelines and future research directions to ensure AI-driven threat-detection systems are effective, ethically sound and resilient.

Keywords: artificial intelligence, threat detection, human health, patient safety, occupational safety, real-world monitoring, anomaly detection, ethics, regulation

Introduction. The 21st century has seen an unprecedented increase in the deployment of artificial intelligence (AI) technologies across multiple domains. In the field of health and safety, real-world situations present continuous, dynamic threats — from adverse drug events and falls in hospital settings, to occupational hazards and large-scale public health emergencies. Traditional systems for identifying and mitigating such threats often rely on reactive reporting and human judgement; by contrast, AI offers the promise of proactive, continuous and scalable detection of risks.

For example, within healthcare settings, the monitoring of clinical alarms, drug safety reports and error-events remains challenging due to large volume and heterogeneity of data. A systematic review found that AI-enabled decision support systems can aid error detection, patient stratification and drug management.[1]

In occupational health and safety (OHS) as well, AI is being utilized to monitor workplace hazards in real time, identify ergonomic risks or detect anomalies in industrial operations.[2]

Yet, these opportunities are accompanied by substantial challenges: model bias, lack of standard benchmarks, transparency and accountability, privacy and regulatory frameworks.

In this article, we examine how AI is being applied to detect threats to human health and safety in real-world settings, review the evidence, analyse the gaps, and propose a roadmap for future deployment.

AI applications for threat detection in healthcare. A major harm domain in healthcare is patient safety — the avoidance of preventable harm to patients during care.[1] In a scoping review by Bates et al., eight major harm domains (including adverse drug events, falls, infections, diagnostic errors) were mapped to AI applications.[7] Their review identified hundreds of studies

JOURNAL OF APPLIED SCIENCE AND SOCIAL SCIENCE
eISSN 2229-3113 pISSN 2229-3205
Volume 15 Issue 10, October 2025
Impact factor: 2019: 4.679 2020: 5.015 2021: 5.436, 2022: 5.242, 2023: 6.995, 2024 7.75

in which AI methods — ranging from machine learning (ML) to natural language processing (NLP) — were used to predict, detect or prevent harmful events. These included monitoring vital signs, wearables, pressure sensors, computer vision, electronic health records (EHRs) and other novel data streams.

Meanwhile, Choudhury et al.'s review found that while AI-enabled decision support systems offer promise in enhancing patient safety (error detection, stratification, drug management), major gaps remain: in standardisation, heterogeneity of reporting, and real-world validation.[1]

More recently, De Micco et al. reported that AI systems can detect adverse events, predict medication errors, assess fall risk and support incident reporting automation—but socio-technical, implementation and standardisation barriers persist.[3]

On the risk side, AI tools can themselves produce unpredictable errors, raise privacy concerns, create bias and discrimination, and challenge patients' rights. Botha et al. (2024) summarised these threats in a scoping review of 80 articles.[5]

Thus, in the healthcare domain, the literature indicates substantial potential but also significant limitations in deploying AI for threat detection.

AI in occupational health and safety and other real-world settings. Beyond hospitals, workplaces and industrial settings also present threats to human health and safety: ergonomic injuries, exposure to hazardous substances, fatigue, falls, equipment malfunction, etc. The review by El-Helaly (2024) demonstrates that AI can support real-time hazard monitoring, proactive prevention and predictive analytics in occupational health and safety.[2] The use of wearable sensors, computer vision in factories, analytics for near-miss incidents and anomaly detection in industrial systems have begun to emerge.

Also, research suggests that AI in cybersecurity and device safety (e.g., medical devices) is becoming important: for example the cybersecurity of AI medical devices highlights risks that could impact patient safety and system integrity. Thus, the broader "real-world" threat detection domain for AI spans clinical, workplace and infrastructural settings, offering a spectrum of applications.

Key benefits and enablers. From the literature, several key benefits of AI for threat detection emerge:

- **Early detection and prediction**: AI models can identify patterns preceding adverse events (e.g., patient de-compensation, sudden fall risk, equipment failure) enabling proactive intervention. [7]
- **Continuous monitoring of large data volumes**: Unlike manual processes, AI can process streaming data from wearables, sensors, EHRs or connected devices, enabling real-time insight.
- **Automation of error-prone or repetitive tasks**: For example, classification of safety reports, anomaly detection in sensor data, alert generation.[3]
- **Support for decision-making**: AI may assist clinicians, safety officers or system managers by providing risk scores, stratification, and decision support.[1]

Together, these enablers support a vision of smart, data-driven health and safety ecosystems.

Key challenges, threats and gaps. However, the literature also emphasises major challenges:

JOURNAL OF APPLIED SCIENCE AND SOCIAL SCIENCE
eISSN 2229-3113 pISSN 2229-3205
Volume 15 Issue 10, October 2025
Impact factor: 2019: 4.679 2020: 5.015 2021: 5.436, 2022: 5.242, 2023: 6.995, 2024 7.75

- **Model bias, transparency and explainability**: AI systems may embed bias, lose transparency (so-called "black box" systems) and be difficult to interpret, especially in high-stakes settings. [6]
- **Data quality, generalisability and validation**: Many studies use retrospective, controlled datasets; fewer prospective, real-world validations exist. Bates et al. found most algorithms were not externally validated. [7]
- **Regulatory, ethical, privacy and legal issues**: Use of patient/workplace data, informed consent, regulation of AI in safety systems, liability for AI-driven decisions are emerging concerns.[5]
- **System integration and socio-technical barriers**: Incorporating AI into existing workflows, ensuring human–machine collaboration, managing alert fatigue and ensuring usability are non-trivial.[3]
- **Threats from misuse or unintended harm**: AI systems may themselves cause harm (e.g., mis-predictions), might be attacked (cybersecurity risks), or exacerbate disparities and inequalities.[5]

These gaps indicate that despite the promise, deployment in real-world threat detection requires careful design, oversight and governance.

Conceptual Framework for AI-based Threat Detection in Real-World Health & Safety. Building on the literature, we propose a conceptual framework for deploying AI to detect threats to human health and safety in real-world situations. The framework comprises four core components:

1. **Data Acquisition & Integration**
   - Real-time and historical data from diverse sources: EHRs, wearable sensors, environmental sensors, computer vision feeds, workplace logs, incident reports.
   - Integration of structured and unstructured data (e.g., sensor time-series, free-text reports, videos).
   - Ensuring data quality, annotation, interoperability and appropriate preprocessing.

2. **Analytics & Modeling**
   - Application of machine learning (ML), deep learning (DL), natural language processing (NLP), anomaly detection and predictive modelling.
   - Use of algorithms to detect deviations (anomalies), predict high-risk states, classify events and prioritise threats.
   - Incorporation of explainable AI (XAI) to enable transparent decision-making and human oversight.

3. **Decision Support & Intervention**
   - Output risk-scores, alerts or dashboards integrated into workflows for clinicians, safety officers or system managers.
   - Human–AI collaboration: AI supports rather than replaces human actors; decisions remain interpretable and actionable.
   - Real-time intervention capability (e.g., triggering alarms, initiating protocols) and

JOURNAL OF APPLIED SCIENCE AND SOCIAL SCIENCE
eISSN 2229-3113 pISSN 2229-3205
Volume 15 Issue 10, October 2025
Impact factor: 2019: 4.679 2020: 5.015 2021: 5.436, 2022: 5.242, 2023: 6.995, 2024 7.75

feedback loops.

4. **Governance, Ethics & Continuous Improvement**
- Framework for oversight, regulation, data privacy, fairness, accountability and legal compliance.
- Continuous monitoring of model performance, bias auditing, system retraining and validation in real-world settings.
- Stakeholder engagement (patients, workers, safety personnel) and transparent reporting of outcomes and harms.

This framework is designed to ensure that AI systems for threat detection are technically robust, ethically acceptable and operationally effective. It emphasises that threat detection is not only about prediction/modeling, but also about integration into the socio-technical environment of health and safety systems.

Real-World Use Cases and Evidence. Here we highlight representative use cases from literature and practice across domains, illustrating how the framework above applies.

Healthcare: Adverse drug event and fall-risk prediction. In healthcare, algorithms have been developed to flag patients at risk of adverse drug events (ADEs) or falls. For example, the scoping review by Bates et al. identified numerous studies using AI to monitor vital signs, sensor data and EHR entries to predict falls or other harm.[7]
These systems offer the potential to reduce incidence of events such as pressure ulcers, hospital-acquired infections or surgical complications. However, as noted, most remain at retrospective proof-of-concept stage with limited large-scale deployment.

Occupational Health & Safety: Real-time monitoring of workplace hazards. In industrial and workplace settings, AI has been used to analyse sensor feeds (accelerometers, video), detect unsafe behaviours, monitor ergonomic load or detect equipment malfunction. For instance, El-Helaly's review highlights real-time monitoring, proactive risk identification and preventive analytics in OHS. [2]
Real-world deployment includes systems that identify fatigue, alert to near-miss incidents, or monitor environmental exposures. Such systems can potentially reduce injuries, downtime and safety incidents.

Public Health / Infrastructure: Anomaly detection & system safety. Beyond individual health settings, AI can detect threats to human safety in public health or infrastructure contexts (e.g., detecting anomalies in sensor networks, cybersecurity threats to medical devices). A review on AI medical device cybersecurity raises the importance of protecting AI systems themselves from attacks that could compromise safety. [8]
Although direct large-scale applications in public health infrastructure remain emerging, the potential for AI to monitor population-level threats (e.g., outbreak detection, environmental hazards) is increasingly recognized.

JOURNAL OF APPLIED SCIENCE AND SOCIAL SCIENCE
eISSN 2229-3113 pISSN 2229-3205
Volume 15 Issue 10, October 2025
Impact factor: 2019: 4.679 2020: 5.015 2021: 5.436, 2022: 5.242, 2023: 6.995, 2024 7.75

Discussion. The foregoing review and framework highlight that AI-driven threat detection in real-world health and safety contexts offers significant promise—but also requires mindful navigation of multiple issues.

Translating from research to deployment. A key challenge lies in moving from retrospective modelling to prospective, real-time deployment in complex, dynamic environments. Many studies remain proof-of-concept, lacking external validation and real-world performance data.[7] Implementation involves data integration, system interoperability, workflow redesign and staff training. Socio-technical factors such as alert fatigue, user trust, integration into decision pathways are critical. [3]
Therefore, organisations seeking to implement AI threat-detection systems must emphasise change management, user engagement and real-world validation.

Ethical, legal and governance considerations. As AI systems play an increasingly influential role in safety-critical domains, ethical and regulatory oversight becomes imperative. Issues include algorithmic bias (which may exacerbate inequities), transparency/explainability (critical for trust and regulatory compliance), privacy/data protection (especially for health and occupational data) and liability when AI decisions cause harm. Botha et al. emphasise the threats to patient rights from AI tool deployment (e.g., unpredictable errors, inadequate regulation, data security). [5]
Hence, a governance framework must accompany technical deployment: clear accountability, audit trails, human-in-the-loop systems, informed consent where relevant, and policy/regulation aligned with local/national legal regimes.

Risk of AI systems themselves becoming a threat. An often overlooked dimension is that AI systems may themselves introduce new risks. For instance, models may malfunction, be attacked (data poisoning, adversarial examples), or operate outside safe boundaries. Banja (2020) highlights that AI integration into healthcare will likely heighten the magnitude of risk—even if new types of risks. [6]
Therefore, threat-detection systems must include robustness, cybersecurity, adversarial resistance, continuous monitoring and fail-safe mechanisms.

Equity, access and global implementation. While many AI applications are piloted in high-resourced healthcare or industrial settings, the global spread of health and safety threats means that low- and middle-income countries (LMICs) must also benefit. Challenges such as data scarcity, infrastructure limitations, workforce skills and regulatory gaps may limit AI deployment in resource-constrained contexts. Botha et al. note that factors like poverty, power supply, internet infrastructure may hamper AI adoption in global south settings. [5]
Hence, global implementation strategies must consider context-appropriate solutions, capacity building and equitable access.

Conclusion. The application of artificial intelligence to detect threats to human health and safety in real-world situations holds great promise. From clinical inpatient settings to industrial workplaces and public health infrastructure, AI offers the potential for early detection, continuous monitoring and actionable decision support. However, realising this potential

JOURNAL OF APPLIED SCIENCE AND SOCIAL SCIENCE
eISSN 2229-3113 pISSN 2229-3205
Volume 15 Issue 10, October 2025
Impact factor: 2019: 4.679 2020: 5.015 2021: 5.436, 2022: 5.242, 2023: 6.995, 2024 7.75

requires more than sophisticated algorithms: success depends equally on robust data ecosystems, user-centred design, ethical and regulatory oversight, human–machine collaboration and context-aware implementation. By adhering to the conceptual framework and recommendations outlined above, organisations can deploy AI-driven threat-detection systems that are effective, safe, equitable and sustainable. The future of health and safety lies in intelligent, integrated systems that enhance human capability—rather than replace it—and safeguard wellbeing in an increasingly complex world.

References

1. Choudhury A, et al. Role of Artificial Intelligence in Patient Safety Outcomes. PMC. 2020. (PMC)
2. El-Helaly M. Artificial Intelligence and Occupational Health and Safety: A Systematic Review. PMC. 2024. (PMC)
3. De Micco F, et al. Artificial intelligence in healthcare: transforming patient safety. Frontiers in Medicine. 2025. (Frontiers)
4. Khan MM, et al. Towards secure and trusted AI in healthcare: A systematic review. ScienceDirect. 2024. (ScienceDirect)
5. Botha NN, et al. A scoping review of perceived threats to patient rights and safety from AI tool use in healthcare. Arch Public Health. 2024. (BioMed Central)
6. Banja J. How might artificial intelligence applications impact risk management? AMA Journal of Ethics. 2020. (Journal of Ethics)
7. Bates DW, et al. The potential of artificial intelligence to improve patient safety. NPJ Digital Medicine. 2021. (Nature)
8. Biasin E, et al. Cybersecurity of AI medical devices: risks, legislation, and challenges. arXiv. 2023. (arXiv)