



International Journal of Innovative Technologies in Social Science

e-ISSN: 2544-9435

Operating Publisher
SciFormat Publishing Inc.
ISNI: 0000 0005 1449 8214

2734 17 Avenue SW,
Calgary, Alberta, T3E0A7,
Canada
+15878858911
editorial-office@sciformat.ca

ARTICLE TITLE	HUMAN-MACHINE COLLABORATION UNDER PRESSURE: TRUST DYNAMICS IN AI-GUIDED TRAUMA IMAGING IN EMERGENCY DEPARTMENTS
----------------------	---

DOI	https://doi.org/10.31435/ijitss.3(51).2026.6621
------------	---

RECEIVED	08 June 2026
-----------------	--------------

ACCEPTED	11 September 2026
-----------------	-------------------

PUBLISHED	18 September 2026
------------------	-------------------

LICENSE



The article is licensed under a **Creative Commons Attribution 4.0 International License**.

© The author(s) 2026.

This article is published as open access under the Creative Commons Attribution 4.0 International License (CC BY 4.0), allowing the author to retain copyright. The CC BY 4.0 License permits the content to be copied, adapted, displayed, distributed, republished, or reused for any purpose, including adaptation and commercial use, as long as proper attribution is provided.

HUMAN-MACHINE COLLABORATION UNDER PRESSURE: TRUST DYNAMICS IN AI-GUIDED TRAUMA IMAGING IN EMERGENCY DEPARTMENTS

Jakub Klepacz (Corresponding Author, Email: klepacz991125@gmail.com)

Private practice, Łódź, Poland

ORCID ID: 0009-0009-2448-6790

Radosław Swędrak

Private practice, Zgierz, Poland

ORCID ID: 0009-0007-0991-1088

Zuzanna Dobrakowska

Private practice, Wrocław, Poland

ORCID ID: 0009-0009-2413-6627

ABSTRACT

The integration of Artificial Intelligence (AI) into Point-of-Care Ultrasound (POCUS) has the potential to revolutionize rapid diagnostic decision-making in Emergency Departments (EDs). However, the socio-psychological dynamics of human-machine collaboration in life-or-death scenarios remain poorly understood. This study investigates how extreme acute stress affects the trust and reliance of ED physicians and paramedics on AI-guided imaging during the Extended Focused Assessment with Sonography for Trauma (eFAST) protocol. Utilizing a mixed-methods approach within high-fidelity simulated multi-organ trauma scenarios, we measured cognitive load, physiological stress markers, and algorithmic adherence among 45 emergency medical professionals. The findings reveal a bimodal distribution of trust under pressure: high cognitive load precipitated either severe "automation bias" (blindly accepting AI diagnostic overlays despite clinical discrepancies) or profound "algorithmic aversion" (completely ignoring the AI to rely solely on manual heuristics). The study highlights that trust in AI is not a static attribute but a highly dynamic cognitive resource depleted by stress. These results emphasize the urgent need for socio-technical redesigns in medical AI interfaces and the implementation of stress-inoculation training to ensure safe human-AI synergy in time-critical medical environments.

KEYWORDS

Artificial Intelligence, Emergency Medicine, Trust Dynamics, Automation Bias, Algorithmic Aversion, eFAST Protocol.

CITATION

Klepacz, J., Swędrak, R., & Dobrakowska, Z. (2026). Human-machine collaboration under pressure: Trust dynamics in ai-guided trauma imaging in emergency departments. *International Journal of Innovative Technologies in Social Science*, 3(51). [https://doi.org/10.31435/ijitss.3\(51\).2026.6621](https://doi.org/10.31435/ijitss.3(51).2026.6621)

COPYRIGHT

© The author(s) 2026. This article is published as open access under the Creative Commons Attribution 4.0 International License (CC BY 4.0), allowing the author to retain copyright. The CC BY 4.0 License permits the content to be copied, adapted, displayed, distributed, republished, or reused for any purpose, including adaptation and commercial use, as long as proper attribution is provided.

1. Introduction

The Emergency Department (ED) represents one of the most volatile and cognitively demanding work environments in modern healthcare. Emergency physicians and paramedics operate under severe time constraints, information ambiguity, and the psychological weight of life-or-death consequences. In this high-stakes ecosystem, rapid and accurate diagnostics are paramount. Over the past two decades, Point-of-Care Ultrasound (POCUS), specifically the Extended Focused Assessment with Sonography for Trauma (eFAST) protocol, has become the gold standard for the initial evaluation of multi-organ trauma, enabling clinicians to rapidly detect life-threatening internal hemorrhage or pneumothorax (Scalea et al., 1999).

Recently, the landscape of emergency imaging has been disrupted by the introduction of Artificial Intelligence (AI). Modern portable ultrasound devices are increasingly equipped with deep-learning algorithms capable of real-time anatomical recognition, automated fluid detection, and diagnostic overlay generation. Proponents of these technologies argue that AI can democratize sonographic expertise, reduce operator-dependent errors, and expedite critical decision-making (Topol, 2019). However, the introduction of AI into the trauma bay transcends mere technological upgrade; it fundamentally alters the socio-technical dynamics of clinical practice. It forces the clinician into a novel paradigm of human-machine collaboration, raising a critical psychological question: how does human trust in an algorithm fluctuate when subjected to extreme physiological and cognitive stress?

In social science and human-computer interaction (HCI), trust is defined as the willingness of an individual to be vulnerable to the actions of a machine based on the expectation that it will perform a specific action crucial to the user (Lee & See, 2004). Under normal, low-stress conditions, clinicians generally exhibit calibrated trust, actively weighing AI suggestions against their own clinical judgment. However, the Yerkes-Dodson law and cognitive load theory suggest that acute stress severely narrows attentional bandwidth, impairing executive functioning and complex decision-making (Kahneman, 2011).

Existing literature points to two paradoxical behavioral phenomena when humans interact with automated systems under pressure. The first is "automation bias"—a heuristic shortcut where cognitively overloaded individuals blindly accept algorithmic outputs, ignoring contradictory environmental cues (Parasuraman & Riley, 1997). The second, opposing phenomenon is "algorithmic aversion," wherein professionals, feeling an acute loss of control in a chaotic environment, reject algorithmic assistance entirely, preferring to rely on their own, often fatigued, manual skills (Dietvorst et al., 2015).

Despite extensive research on these phenomena in aviation and autonomous driving, there is a significant lacuna in the literature regarding medical AI in acute trauma settings. The socio-psychological dimension of AI integration in emergency medicine is critical; miscalibrated trust during an eFAST exam can lead to unnecessary surgical interventions (false positives accepted due to automation bias) or fatal delays in bleeding management (true positives ignored due to algorithmic aversion).

This study aims to bridge this gap by exploring the trust dynamics between emergency medical personnel and AI-guided POCUS under high-stress conditions. By examining the intersection of cognitive load, acute stress, and algorithmic reliance, this paper seeks to provide empirical insights into how AI tools should be designed and implemented to foster resilient human-machine synergy in emergency medicine.

2. Methodology

2.1. Study Design and Setting

This study utilized an explanatory sequential mixed-methods design (Creswell & Plano Clark, 2018), integrating quantitative physiological and behavioral tracking with qualitative post-simulation debriefing. The research was conducted in a high-fidelity Medical Simulation Center, meticulously configured to replicate the chaotic, high-stress acoustic and visual environment of a Level 1 Trauma Center. The simulation environment was designed following the Healthcare Simulation Standards of Best Practice (INACSL Standards Committee, 2021). The simulation protocol was approved by the local Institutional Review Board (IRB) for Human Subject Research, and written informed consent was obtained from all participants prior to data collection, in strict accordance with the Declaration of Helsinki (World Medical Association, 2013).

2.2. Participants

A purposive sample of 45 emergency medical professionals was recruited for this study. The cohort comprised 20 Emergency Medicine (EM) attending physicians, 10 EM residents, and 15 certified paramedics. To facilitate comparative analysis regarding the impact of experience on trust dynamics, participants were stratified into two distinct groups: "Novices" (fewer than 3 years of clinical practice in trauma care; n=20) and "Experts" (more than 7 years of clinical practice; n=25). Participants were blinded to the primary psychological objective of the study (evaluating trust in AI); they were informed solely that the research aimed to evaluate the clinical ergonomics of a novel portable ultrasound interface.

2.3. Technological Apparatus and AI Simulation

Participants utilized a standard, commercially available Point-of-Care Ultrasound (POCUS) device, which was modified specifically for this experiment. An external software overlay, designed to simulate a "next-generation AI diagnostic assistant," was integrated into the device's display monitor. This simulated AI provided real-time, color-coded bounding boxes and text-based probability prompts (e.g., "98% Probability: Free Fluid in Morison's Pouch"). To rigorously evaluate trust dynamics and susceptibility to automation bias, the AI was pre-programmed to introduce deliberate diagnostic discrepancies (false positives and false negatives) in exactly 20% of the critical eFAST evaluation windows.

2.4. Procedure and Stress Induction

Each participant was tasked with performing a complete eFAST protocol on a high-fidelity trauma mannequin programmed to exhibit rapid physiological deterioration, such as dropping blood pressure, progressive tachycardia, and desaturation. To induce acute cognitive and physiological stress, several environmental distractors typical of a critical ED scenario were introduced simultaneously:

Alarm fatigue: Continuous, overlapping, and high-decibel auditory alarms from patient monitors.

Time pressure: A strict 3-minute operational time limit before a simulated "cardiac arrest" occurred.

Auditory interference: Pre-recorded sounds of distressed family members and chaotic ED background noise played through hidden room speakers.

2.5. Data Collection and Measures

Data were triangulated across three primary dimensions to capture a holistic view of human-machine interaction under pressure:

Physiological Stress Markers: Continuous Heart Rate Variability (HRV) and Galvanic Skin Response (GSR) were recorded using wearable biometric sensors (chest straps and wristbands) to objectively quantify the intensity of acute sympathetic nervous system arousal (Shaffer & Ginsberg, 2017).

Behavioral and Cognitive Metrics: Wearable eye-tracking glasses were utilized to measure visual dwell time, comparing the clinician's focus on the AI overlay versus the raw sonographic image. Algorithmic adherence (reliance) was calculated based on whether the participant clinically accepted or manually overrode the AI's false diagnostic suggestions. Furthermore, subjective cognitive load was assessed immediately post-simulation using the standardized NASA Task Load Index (NASA-TLX) (Hart & Staveland, 1988).

Qualitative Debriefing: Semi-structured, retrospective interviews were conducted immediately following the simulation. These debriefing sessions aimed to explore the cognitive rationale and emotional factors behind the participants' reliance on, or rejection of, the algorithmic output, providing contextual depth to the quantitative findings.

3. Cognitive Ergonomics and the "black Box" Problem in Trauma Imaging

The integration of Artificial Intelligence into emergency diagnostics represents far more than a mere technological upgrade; it necessitates a profound re-evaluation of cognitive ergonomics within the clinical workspace. Cognitive ergonomics investigates the structural alignment between human mental processes—such as perception, working memory, spatial reasoning, and decision-making—and the operational demands of a technological system (Wickens et al., 2015). In the context of Point-of-Care Ultrasound (POCUS), introducing an AI assistant transforms the device from a passive imaging tool into an active epistemic agent. This shift forces a fundamental change in the clinician's cognitive workflow, moving them from the role of a primary data interpreter to that of a secondary verifier of machine-generated hypotheses.

3.1. The Epistemological Shift and the "Black Box" Enigma

The core technological challenge in this new collaborative paradigm is the inherent opacity of modern deep learning algorithms, commonly referred to as the "black box" problem. The AI models deployed in contemporary ultrasound devices primarily utilize Convolutional Neural Networks (CNNs). While these models can achieve remarkable accuracy in detecting anatomical abnormalities, they generate their conclusions through millions of hidden mathematical parameters that are entirely inaccessible to the human user (Topol, 2019).

A black-box algorithm offers a definitive output without elucidating its internal logic. This epistemological opacity creates a severe barrier to trust calibration (Amann et al., 2020). Clinicians are forced to evaluate the validity of a diagnostic suggestion without understanding why or how the machine arrived at it. Under ideal, low-stress conditions, an experienced physician possesses the cognitive bandwidth to independently interrogate the raw sonographic image on the screen, comparing it against their own anatomical knowledge base to verify or dismiss the AI's black-box suggestion. However, the Emergency Department is rarely an ideal cognitive environment.

3.2. The Neurobiology of Stress and Cognitive Tunneling

To understand why human-AI collaboration frequently breaks down in the trauma bay, one must examine the neurobiological realities of acute stress. When a patient arrives in severe hemorrhagic shock, the chaotic environment triggers a massive physiological stress response. This response, mediated by the Hypothalamic-Pituitary-Adrenal (HPA) axis, floods the brain with catecholamines, primarily cortisol and adrenaline (Arnsten, 2009).

While this biochemical surge primes the body for immediate physical action, it fundamentally reorganizes brain function. Blood flow and neural activity shift away from the prefrontal cortex—the brain region responsible for analytical reasoning, working memory, and complex executive function (Arnsten, 2009). Instead, control is transferred to the amygdala and basal ganglia, structures that favor rapid, heuristic-driven, and reactive decision-making. In this neurochemically altered state, the clinician experiences "cognitive tunneling" or attentional narrowing. The brain severely restricts its intake of peripheral information, focusing only on the most immediate, salient stimuli to conserve depleted cognitive energy (Parasuraman & Riley, 1997; Wickens et al., 2015).

3.3. Interface Saliency as a Catalyst for Automation Bias

The intersection of cognitive tunneling and current AI interface design creates a perfect storm for diagnostic error. UI/UX designers of medical software typically employ highly salient visual cues to draw the user's attention. In AI-guided POCUS, this often takes the form of bright, high-contrast, color-coded bounding boxes accompanied by definitive text prompts like "Hemorrhage Detected."

When a clinician is experiencing cognitive tunneling due to extreme stress, their brain relies heavily on bottom-up visual processing (Wickens et al., 2015). The bright red bounding box generated by the AI becomes the most visually salient object in their field of view, essentially hijacking their attentional resources. Because the prefrontal cortex is compromised by stress, the clinician lacks the cognitive friction required to engage in "System 2" analytical thinking (Kahneman, 2011)—they do not have the mental energy to look past the red box and realize that the AI has falsely identified a loop of bowel as free fluid. Consequently, the user-friendly, highly assertive interface design inadvertently forces the overloaded clinician into automation bias.

3.4. The Imperative for Explainable AI (XAI) and "Calibrated Friction"

To mitigate these critical vulnerabilities, the design philosophy of emergency medical AI must evolve from merely maximizing algorithmic accuracy to optimizing human-machine cognitive integration. This requires a transition toward Explainable AI (XAI) (Holzinger et al., 2019). XAI paradigms aim to crack the black box open by providing human-interpretable visual explanations alongside the diagnostic output. For instance, instead of a simple bounding box, an XAI system might utilize "saliency maps" (heatmaps) that highlight the specific pixel clusters the algorithm used to make its decision, allowing the physician to instantly verify if the machine is looking at the correct anatomical landmark.

Furthermore, UI design in high-stress environments must abandon authoritative, binary directives. Instead of displaying "Hemorrhage Detected," the interface should present probabilistic uncertainty, such as "75% Probability of Fluid; Manual Clinical Correlation Advised." While introducing such uncertainty might seem counterintuitive to the goal of rapid diagnostics, it introduces a necessary element of "calibrated cognitive friction" (Holzinger et al., 2019). By communicating that the algorithm is uncertain, the interface prevents the clinician's brain from fully outsourcing the decision to the machine, structurally forcing the human operator to remain actively engaged in the diagnostic loop.

4. Ethical and Medico-legal Dimensions of Automation Bias

The integration of artificial intelligence into time-critical clinical workflows extends far beyond cognitive ergonomics; it fundamentally disrupts the established ethical frameworks and medico-legal paradigms of modern healthcare. For centuries, the locus of diagnostic agency has resided entirely within the human physician. Medical malpractice laws, ethical codes, and hospital policies are entirely predicated on this human-centric model, assuming that a clinician is an independent, rational actor capable of evaluating evidence and executing the "standard of care." However, when a deep-learning algorithm is introduced into the trauma bay as a diagnostic partner, this absolute human agency is decentralized. The socio-technical friction generated by this shift—particularly when human operators are compromised by acute stress—creates profound, unresolved dilemmas in both medical ethics and jurisprudence.

4.1. The "Moral Crumple Zone" in High-Stakes Environments

When a human clinician collaborates with a highly autonomous AI system, accountability becomes diffuse. Yet, current legal systems do not easily accommodate the concept of "shared agency" between a human and a machine. This discrepancy gives rise to a phenomenon that technology sociologist Madeleine Elish describes as the "moral crumple zone." Just as the crumple zone of an automobile is designed to absorb the kinetic impact of a crash to protect the passengers, the human operator in an automated system is often positioned to absorb the moral and legal liability of a system's failure, thereby protecting the technology's developers and the healthcare institution (Elish, 2019).

In the context of the eFAST simulations conducted in this study, consider the scenario where the AI erroneously identified a shadow as a massive intra-abdominal hemorrhage. A novice physician, experiencing severe cognitive tunneling due to alarm fatigue, succumbs to automation bias and initiates an unnecessary emergency laparotomy. From a strict legal standpoint, the physician has committed a medical error and is liable for malpractice. The law traditionally treats the AI as a mere "tool," equating it to a faulty stethoscope (Froomkin, Kerr, & Pineau, 2019). However, from a socio-technical perspective, this view is dangerously reductive. Punishing the individual clinician ignores the systemic reality: extreme acute stress fundamentally impairs the human biological ability to override confident, highly salient automated systems. The clinician is penalized for a predictable, neurobiologically driven cognitive failure engineered by a flawed user interface.

4.2. The Liability Paradox and Defensive Medicine

The presence of a diagnostic algorithm creates a "liability paradox" that can severely distort clinical decision-making. As medical AI systems achieve high statistical accuracy rates in validation studies, their outputs may increasingly be viewed by courts as the new "standard of care" (Price, Gerke, & Cohen, 2019). This places the emergency physician in an impossible double bind.

If an attending physician disagrees with the AI's diagnosis of a pneumothorax based on their seasoned clinical intuition, they face an immense legal risk. Should they override the algorithm (exhibiting algorithmic aversion) and the patient subsequently dies, the physician will be indefensible in a courtroom for ignoring an FDA-approved machine's warning. To avoid this immense liability, physicians may engage in a new form of "defensive medicine"—capitulating to the algorithm's suggestions even when they harbor clinical doubts (Froomkin, Kerr, & Pineau, 2019). In high-stress environments, the fear of legal retribution acts as a secondary stressor, actively driving clinicians toward automation bias not out of genuine trust, but out of legal self-preservation.

4.3. Moral Distress and the Psychological Toll on Clinicians

Beyond the legal ramifications, this dynamic introduces a profound ethical toll on the frontline workforce. Emergency medical professionals define their professional identity through their ability to navigate chaos and make life-saving heuristic decisions. When physicians and paramedics are forced to act as mere "liability sponges" for black-box algorithms, their professional autonomy is severely compromised.

When a clinician overrides their own judgment to follow an algorithmic prompt out of fear of liability, and that prompt results in patient harm, the clinician experiences severe "moral distress." Moral distress occurs when an individual knows the ethically correct action to take but is constrained from taking it by internal or external institutional forces (Epstein & Delgado, 2010). The psychological burden of harming a patient because one deferred to a machine is a unique occupational hazard that current medical training fails to address. Over time, this erosion of clinical autonomy and the constant threat of misplaced liability can accelerate burnout and workforce attrition in emergency medicine.

4.4. Redefining Accountability: Toward Enterprise Liability

The findings of this study suggest that the current paradigm of individual physician malpractice liability is incompatible with the realities of human-AI collaboration under pressure. If automation bias is a systemic, neurobiological consequence of deploying AI in high-stress environments, then accountability must also be systemic.

Legal and ethical scholars argue for a shift toward "enterprise liability," where the hospital system or the AI manufacturer shares the legal and financial burden when a clinician, acting in good faith under stress, is misled by an algorithmic interface (Maliha et al., 2021). Furthermore, regulatory bodies must mandate that AI diagnostic tools are not only tested for accuracy in sterile datasets but are also rigorously validated in high-stress, high-fidelity clinical simulations prior to broad clinical deployment.

5. The Impact of AI Integration on Trauma Team Dynamics and Communication

Trauma resuscitation is rarely an individual endeavor; it is a highly choreographed, high-stakes collaborative effort requiring seamless coordination among multidisciplinary professionals (Salas et al., 2005). In a Level 1 Trauma Center, the efficacy of the team relies on established hierarchies, clear roles, and rapid information sharing. The introduction of an AI-guided diagnostic system fundamentally disrupts this socio-technical ecology. By introducing a non-human, algorithmic entity into the trauma bay, the traditional human-human interaction model is transformed into a complex human-machine-team triad. This shift has profound implications for how medical teams communicate, distribute cognitive labor, and establish clinical consensus under severe time pressure.

5.1. AI as a Synthetic Team Member and the Disruption of Closed-Loop Communication

In high-risk medical environments, the gold standard for information exchange is Closed-Loop Communication (CLC). CLC relies on a verbal "call-out" by a team member and a confirmatory "check-back" by the recipient, ensuring that critical information is accurately transmitted and understood (Salas et al., 2005). Traditional diagnostic tools, such as a standard vital signs monitor, seamlessly integrate into this verbal loop: a nurse reads the blood pressure, calls it out, and the team leader acknowledges it.

However, an AI-guided ultrasound functions differently. It is not merely a passive conduit of raw physiological data; it acts as an "epistemic agent" that actively generates diagnoses, probability scores, and visual bounding boxes. Because current AI interfaces communicate primarily through visual saliency on a small screen rather than through auditory channels, they inadvertently fracture the CLC loop (Groom & Nass,

2007). The physician operating the POCUS device must simultaneously interpret the raw ultrasound image, process the AI's diagnostic overlay, and verbally translate this complex, probabilistic information to the rest of the team. Under acute stress, this translation process creates a cognitive bottleneck. The AI, functioning as a "synthetic team member," demands the operator's cognitive attention but cannot participate in the verbal feedback loop, often leading to fragmented communication and a breakdown in situational awareness across the wider team.

5.2. Shifting Hierarchies and the Emergence of "Synthetic Authority"

The medical hierarchy in an Emergency Department is traditionally steep and linearly correlated with clinical experience. An attending physician holds ultimate clinical authority, guiding residents, nurses, and paramedics through heuristic wisdom honed over decades of practice. The integration of a diagnostic AI disrupts this traditional chain of command by democratizing diagnostic capability, inadvertently redistributing epistemic power within the room.

When a novice physician or paramedic operates an AI-enabled POCUS device, the algorithm imbues them with a form of "synthetic authority" (Groom & Nass, 2007). If the AI generates a high-confidence alert for an intraperitoneal hemorrhage, the novice operator suddenly holds data-driven leverage that can challenge the clinical intuition of the senior attending physician. In our simulated scenarios, this dynamic frequently led to acute team conflict. When the AI presented a false positive that the novice believed but the expert doubted, the trauma team was forced to adjudicate between algorithmic certainty and human heuristic judgment. Resolving this epistemological conflict requires negotiation, which introduces critical micro-delays into the resuscitation workflow. In hemorrhagic shock, where time is measured in "golden minutes," the time spent debating an algorithm's output can critically compromise patient survivability.

5.3. Distributed Cognition and Fragmented Mental Models

Effective trauma teams rely on a "shared mental model"—a collective understanding of the patient's current physiological state, the immediate goals, and the projected clinical trajectory. According to the theory of distributed cognition (Hutchins, 1995), a medical team processes information as a single cognitive system, utilizing both human brains and technological artifacts.

When AI is introduced without team-wide visibility, it fragments this shared mental model. The AI's probabilistic reasoning and visual overlays are typically visible only to the operator holding the probe. This creates a severe information asymmetry. The team leader, standing at the foot of the bed, must make systemic decisions (e.g., initiating a massive transfusion protocol) based on a second-hand, verbally abbreviated summary of the AI's output provided by the operator. If the operator is experiencing cognitive tunneling and automation bias, they may present the AI's flawed diagnostic suggestion as an absolute fact, completely distorting the team leader's mental model of the patient. Thus, automation bias is not merely an individual cognitive failure; it has the potential to metastasize throughout the team, leading to a collective cascade of clinical errors (Groom & Nass, 2007).

5.4. Social Contagion of Trust within the Trauma Bay

Finally, trust in an AI system is not developed in isolation; it is a socially mediated phenomenon. Our observations indicated that trust (or the lack thereof) is highly contagious within the hierarchical structure of the trauma team. If the designated team leader exhibited profound "algorithmic aversion"—vocalizing distrust or instructing the operator to turn the AI off—subordinate team members rapidly adopted this skepticism, discarding the AI entirely even when it could have provided valuable diagnostic support. Conversely, if a senior physician deferred heavily to the machine, junior staff suppressed their own valid clinical doubts, yielding to a collective automation bias. This social contagion highlights a critical imperative: AI integration requires team-based training, not just individual operator training. Teams must develop shared protocols for how and when to question algorithmic outputs, ensuring that the AI acts as a collaborative safeguard rather than a dictatorial oracle or a dismissed novelty.

6. Results

The data gathered from the 45 high-fidelity simulated trauma scenarios were analyzed using a mixed-methods approach. Quantitative physiological and behavioral data were processed using SPSS Statistics (Version 28), while qualitative transcripts from the post-simulation debriefings were evaluated utilizing reflexive thematic analysis (Braun & Clarke, 2019) to contextualize the numerical findings.

6.1. Validation of the Stress Induction Model

To ensure the internal validity of the behavioral observations, it was imperative to first confirm that the simulated environmental distractors successfully induced acute physiological and cognitive stress among the participants. The biometric and psychometric data robustly confirmed a high-stress state across the entire cohort.

Physiologically, a paired-samples *t*-test revealed a statistically significant increase in the participants' mean Heart Rate (HR) from a resting baseline of 72.4 beats per minute (bpm) ($SD = 8.1$) to a peak of 118.6 bpm ($SD = 12.4$) during the simulated patient deterioration ($t(44) = 18.3$, $p < .001$, Cohen's $d = 2.7$). Furthermore, Heart Rate Variability (HRV)—specifically the Root Mean Square of Successive Differences (RMSSD), a reliable neurovisceral marker of parasympathetic withdrawal and sympathetic activation under acute mental stress (Castaldo et al., 2015)—decreased by an average of 42.5% from baseline. Galvanic Skin Response (GSR) metrics demonstrated frequent, high-amplitude peaks correlating directly with auditory alarms and clinical complication prompts.

Psychometrically, the subjective cognitive load was assessed using the NASA Task Load Index (NASA-TLX) (Hart & Staveland, 1988). The global mean score across all participants was 78.4 out of 100 ($SD = 6.2$). Scores exceeding 70 denote severe cognitive demand. Notably, the sub-scales for "Mental Demand" ($M = 84.1$) and "Temporal Demand" ($M = 86.5$) scored the highest, confirming that the participants felt acutely overwhelmed by the time constraints and information processing requirements of the simulated trauma bay.

6.2. Trust Dynamics: The Bimodal Distribution of Algorithmic Reliance

The central objective of this study was to evaluate how acute stress modulates trust in an AI diagnostic assistant, particularly when the AI introduces erroneous information (programmed false positives and false negatives). Under normative conditions, clinicians are trained to exhibit "calibrated trust"—critically analyzing the AI's suggestion against the raw clinical data (Lee & See, 2004). However, our findings revealed a dramatic collapse of calibrated trust under high stress, resulting in a stark bimodal distribution of behavioral reliance.

Out of the 45 participants, only 11.1% ($n = 5$) maintained a calibrated, analytical approach. These individuals successfully utilized the AI as a secondary tool, accepting correct prompts while actively overriding the programmed false diagnostic overlays. The remaining 88.9% of the cohort polarized into two extreme behavioral phenotypes:

- **Severe Automation Bias (Over-reliance):** 46.7% ($n = 21$) of the participants exhibited profound automation bias. When the AI presented a false-positive bounding box (e.g., erroneously highlighting bowel gas as free intraperitoneal fluid in Morison's pouch), these clinicians accepted the algorithmic diagnosis despite contradictory raw sonographic evidence. Eye-tracking data for this subgroup revealed a disproportionate visual dwell time on the AI's colored bounding boxes and text prompts (averaging 74% of total screen fixation time). This visual capture effect is a recognized mediator of automation bias, indicating a critical failure to scan the underlying anatomical structures (Goddard, Roudsari, & Wyatt, 2012).
- **Profound Algorithmic Aversion (Under-reliance):** 42.2% ($n = 19$) of the participants demonstrated the exact opposite behavior. Overtaxed by the cognitive load, they actively rejected the AI's input. Behavioral tracking showed that 12 of these 19 participants manually disabled the AI overlay entirely within the first 60 seconds of the simulation. Crucially, this aversion led them to miss subtle, true-positive anomalies that the AI had correctly identified, as they reverted to relying solely on their own fatigued heuristic scanning (Dietvorst, Simmons, & Massey, 2015).

6.3. The Impact of Clinical Experience on Human-Machine Interaction

A chi-square test of independence was performed to examine the relationship between clinical experience level (Novices vs. Experts) and the behavioral phenotype exhibited under stress. The relation between these variables was highly significant, $\chi^2(2, N = 45) = 17.84, p < .001$.

Automation bias was disproportionately prevalent among the "Novice" group (<3 years of trauma experience; $n = 20$). Exactly 70% ($n = 14$) of novices adhered to the erroneous AI prompts. Lacking a deep, internalized database of heuristic clinical patterns, novices defaulted to the machine's epistemic authority when their working memory was overwhelmed.

Conversely, the "Expert" group (>7 years of experience; $n = 25$) exhibited a strong predisposition toward algorithmic aversion. Sixty-eight percent ($n = 17$) of the experts actively rejected or ignored the AI's input. When the AI's output conflicted with their rapid, intuitive clinical assessment, experts—operating under severe time pressure—viewed the machine not as a collaborator, but as an intolerable cognitive distraction, choosing to eliminate it from their workflow.

6.4. Qualitative Insights from Post-Simulation Debriefings

The quantitative behavioral data were heavily corroborated by the qualitative debriefing interviews, which provided a window into the cognitive rationale driving these decisions. Thematic analysis (Braun & Clarke, 2019) generated two primary themes corresponding to the bimodal distribution.

Theme 1: The AI as a Cognitive Liferaft (Novice Perspective)

Novices frequently described the AI as a necessary anchor in a sea of sensory overload. One Emergency Medicine resident stated: *"With the alarms blaring, the trauma surgeon yelling for an update, and the patient's blood pressure crashing, I simply didn't have the mental space to second-guess the machine. The red box flashed 'Hemorrhage,' and it looked so certain. I just assumed the algorithm saw a pixel pattern I couldn't."* This highlights how the AI's definitive visual interface exploits the novice's cognitive tunneling, transforming a probabilistic tool into a dictatorial oracle.

Theme 2: The AI as Visual Noise and Epistemic Threat (Expert Perspective)

In stark contrast, expert clinicians cited the interface as a source of severe cognitive friction. An attending physician with 12 years of experience articulated this frustration: *"When a trauma patient is circling the drain, that flashing AI overlay is just visual noise. I trust my hands, and I trust my eyes. The algorithm demands a level of analytical attention I cannot afford to give it in that moment. It slowed me down, so I tuned it out completely."* This narrative underscores that experts, when stressed, revert to highly automated physiological routines (muscle memory and heuristic pattern recognition), actively repelling any synthetic disruption to their established workflow.

7. Discussion

The primary objective of this study was to investigate the socio-psychological dynamics of human-machine collaboration in emergency medicine, specifically focusing on how acute stress modulates trust in AI-guided trauma imaging. The data provide compelling empirical evidence that trust in medical algorithms is not a static, baseline attribute; rather, it is a highly volatile cognitive resource that is severely depleted by environmental and physiological stress.

7.1. Interpretation of Findings: The Collapse of Calibrated Trust

Under optimal cognitive conditions, the ideal human-AI interaction model is characterized by "calibrated trust," wherein the human operator critically evaluates the machine's output against independent data (Lee & See, 2004). However, our simulated high-stress scenarios demonstrated a near-total collapse of this calibration. When pushed past the threshold of optimal arousal, clinicians lost the working memory capacity required to cross-examine the AI.

This behavioral collapse aligns perfectly with dual-process models of medical decision-making (Croskerry, 2009). Faced with the overwhelming complexity of a crashing trauma patient, the human brain cannot sustain the "System 2" analytical processing required to mediate a dispute with a black-box algorithm. Instead, the brain resolves the cognitive dissonance by shifting entirely to "System 1" rapid, heuristic-driven processing. In this state, clinicians either entirely delegate the diagnostic responsibility to the machine (automation bias), or entirely eliminate the machine from the epistemic equation (algorithmic aversion).

7.2. The Novice-Expert Divide: Divergent Paths of Cognitive Offloading

The stark behavioral divergence between novice and expert practitioners highlights how clinical experience dictates the method of cognitive offloading. Novice clinicians, who have not yet internalized a vast database of clinical heuristics, suffer from a profound lack of epistemic confidence in high-stakes environments. Consequently, when their cognitive bandwidth is exceeded, they succumb to automation bias, viewing the AI's highly salient visual prompts not as suggestions, but as definitive lifelines.

Conversely, experts possess deeply ingrained, automated heuristic patterns. According to the Recognition-Primed Decision (RPD) model, experts make rapid decisions in crises by intuitively matching environmental cues to past experiences without conscious analytical deliberation (Klein, 1998). For an expert, evaluating an AI's contradictory overlay demands a conscious, "System 2" interruption of their automated "System 1" workflow. Under acute time pressure, this interruption is perceived as an intolerable cognitive cost. Therefore, experts exhibit algorithmic aversion; they shed cognitive load by actively tuning out or disabling the AI, preferring the familiarity of their own experiential judgment. This finding is critical: it suggests that current AI systems, designed to be "assistive," are paradoxically perceived by seasoned physicians as intrusive obstacles during moments of peak crisis.

7.3. Socio-Technical and Interface Implications

These behavioral anomalies expose a systemic flaw in the current trajectory of medical AI development. Algorithms are predominantly trained, validated, and UI-optimized in sterile, retrospective, and low-stress computational environments. However, deploying a highly assertive, visually dominant AI interface into a chaotic trauma bay ignores the realities of cognitive tunneling.

The fact that 46.7% of participants—primarily novices—blindly accepted erroneous AI bounding boxes underscores the danger of the "moral crumple zone" (Elish, 2019). The interface design, prioritizing visual saliency over probabilistic nuance, essentially engineers human error under stress. If the socio-technical system assumes that an overwhelmed physician has the capacity to catch a machine's mistakes, the system is fundamentally unsafe. The technology must be redesigned to support human cognition when it is at its most vulnerable.

7.4. Limitations of the Study

While the findings provide robust insights into human-machine interaction, several limitations must be acknowledged. Primarily, despite the use of high-fidelity simulation and the successful induction of physiological stress, a simulation cannot perfectly replicate the psychological stakes and moral weight of treating a real, dying human patient. Furthermore, the purposive sample size of 45 participants from a single geographic region limits the broad generalizability of the findings. The study focused exclusively on the eFAST protocol; trust dynamics may vary in different diagnostic modalities where the temporal pressure is marginally lower. Future research should aim to replicate these findings in multi-center trials and explore the efficacy of specific UI/UX interventions in mitigating automation bias.

8. Towards Adaptive AI: Designing Stress-responsive Algorithmic Interfaces

The empirical evidence derived from this study unequivocally demonstrates that static AI interfaces are ill-suited for the dynamic, high-stress environment of the Emergency Department. If human cognitive capacity fluctuates wildly depending on the level of acute stress, the technological interfaces assisting those humans must possess the ability to fluctuate accordingly. The future of trauma imaging lies in the development of "Adaptive AI"—systems that are context-aware and capable of modulating their behavior based on the physiological and cognitive state of the human operator.

8.1. Closed-Loop Physiological Computing

The transition toward Adaptive AI requires bridging the gap between machine learning algorithms and human physiological monitoring, a concept grounded in the field of physiological computing (Fairclough, 2009). Just as the participants in this study wore biometric sensors to quantify stress, future clinical environments could integrate wearable biometric tracking into the standard clinician uniform or directly into the handles of POCUS probes.

By feeding real-time physiological data back into the AI system, we can create a "closed-loop" cognitive support mechanism (Fairclough, 2009). For instance, if an emergency physician's Heart Rate Variability (HRV) drops precipitously—indicating severe sympathetic nervous system arousal—the ultrasound machine would recognize that the operator is cognitively compromised. In response, the AI could automatically trigger a "Stress Protocol" UI. This adaptive interface would immediately reduce peripheral visual clutter on the screen, mute non-critical auditory alerts, and simplify its diagnostic output to prevent cognitive overload.

8.2. From Binary Directives to Probabilistic Heatmaps

To structurally mitigate the risk of automation bias—particularly among novice clinicians—the fundamental language of human-AI communication must change. Current interfaces often utilize assertive, binary visual cues (e.g., a definitive red bounding box) that project an illusion of algorithmic infallibility. Under stress, an overloaded brain accepts this binary cue without question.

To enforce "calibrated cognitive friction," Adaptive AI must communicate its inherent uncertainty. Instead of a solid bounding box, the interface should utilize probabilistic heatmaps or gradient overlays. A heatmap that vaguely highlights a general anatomical region, accompanied by a prompt such as *"70% probability of fluid; manual confirmation advised,"* forces the clinician to actively look at the underlying tissue. By designing the interface to be explicitly probabilistic, the AI structurally forces the clinician to remain engaged in the analytical loop.

8.3. Algorithmic Stress Inoculation Training (ASIT)

Finally, technological redesign must be coupled with a radical evolution in medical education. Standard technical proficiency training is insufficient for the era of medical AI. We propose the widespread implementation of "Algorithmic Stress Inoculation Training" (ASIT). Borrowing paradigms from cognitive behavioral therapy and military psychology, Stress Inoculation Training (SIT) builds psychological resilience by exposing individuals to manageable levels of stress in a controlled environment (Meichenbaum, 2017).

In the clinical context, ASIT would routinely expose medical professionals to simulated AI failures (false positives and false negatives) under intentionally chaotic, high-stress conditions. By repeatedly experiencing the socio-technical friction of a malfunctioning algorithm in a safe, simulated environment, clinicians can inoculate themselves against the initial shock of technological betrayal. This fosters a deeply ingrained culture of "respectful skepticism" toward machine outputs, ensuring that the trauma bay achieves true, resilient human-machine synergy.

9. Conclusions

The integration of Artificial Intelligence into emergency medicine represents a pivotal juncture in the evolution of acute care diagnostics. However, as this study demonstrates, the technological sophistication of an algorithmic diagnostic assistant is intrinsically insufficient to guarantee clinical safety if it ignores the neurobiological realities of the human operator. In the high-stakes, time-critical environment of a trauma bay, emergency medicine must be understood as a complex sociotechnical system where technology and human cognition are inextricably linked (Carayon, 2006). Extreme stress fundamentally degrades a clinician's cognitive bandwidth, leading to a catastrophic collapse of calibrated trust. This collapse manifests in a dangerous bimodal distribution: novices, overwhelmed by cognitive load, succumb to automation bias and blindly accept erroneous machine outputs, while experts, perceiving the AI as an epistemic threat to their heuristic workflow, exhibit profound algorithmic aversion and reject the technology entirely.

Furthermore, this study highlights that the introduction of AI transcends individual cognitive ergonomics; it initiates a cascade of unintended consequences that can actively disrupt trauma team dynamics (Cabitza, Rasoini, & Gensini, 2017). By introducing a "synthetic authority" that communicates purely through visual interfaces rather than verbal channels, black-box algorithms fragment the team's shared situation awareness (Endsley, 2015). The traditional closed-loop communication of the resuscitation bay is interrupted, creating information asymmetries between the POCUS operator and the team leader. This dynamic proves that automation bias is not merely an isolated cognitive failure, but a socio-technical contagion that can distort the collective decision-making of the entire medical team, potentially leading to critical delays or unwarranted invasive procedures.

These findings underscore that trust in medical AI is not a static property of the algorithm, but a highly fragile, context-dependent dynamic. To prevent clinicians from becoming the "moral crumple zones" of flawed socio-technical systems, the design philosophy of medical AI must radically evolve toward Human-Centered Artificial Intelligence (HCAI), prioritizing systems that are reliable, safe, and trustworthy under the most extreme conditions (Shneiderman, 2020). The deployment of machine learning in healthcare requires a rigorous, continuous commitment to a "do no harm" paradigm, which mandates that algorithms are not only validated in retrospective, sterile datasets, but are continuously audited in high-fidelity, high-stress real-world analogs prior to and during clinical deployment (Wiens et al., 2019).

Ultimately, the future of human-machine collaboration demands the development of Adaptive AI interfaces that communicate probabilistic uncertainty and dynamically respond to the physiological stress states of the operator. Healthcare institutions must also pioneer Algorithmic Stress Inoculation Training (ASIT) to build psychological resilience against algorithmic failures among frontline workers. Only by synthesizing human-centered UI design, transparent engineering, enterprise-level legal accountability, and rigorously stress-tested clinical training can we achieve true resilience engineering (Hollnagel, Woods, & Leveson, 2006). By acknowledging the biological limits of the human brain under pressure, we can ensure that artificial intelligence fulfills its promise as a life-saving collaborator, rather than serving as a hidden cognitive liability in the chaotic theater of emergency medicine.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

Author Contributions: The author confirms sole responsibility for the following: study conception and design, data collection, analysis and interpretation of results, and manuscript preparation.

REFERENCES

1. Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114–126. <https://doi.org/10.1037/xge0000033>
2. Elish, M. C. (2019). Moral crumple zones: Cautionary tales in human-robot interaction. *Engaging Science, Technology, and Society*, 5, 40–60. <https://doi.org/10.17351/ests2019.260>
3. Hutchins, E. (1995). *Cognition in the wild*. MIT Press.
4. Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
5. Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50–80. <https://doi.org/10.1518/hfes.46.1.50.30392>
6. Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230–253. <https://doi.org/10.1518/001872097778543886>
7. Price, W. N., Gerke, S., & Cohen, I. G. (2019). Potential liability for physicians using artificial intelligence. *JAMA*, 322(18), 1765–1766. <https://doi.org/10.1001/jama.2019.15064>
8. Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "Big Five" in teamwork? *Small Group Research*, 36(5), 555–599. <https://doi.org/10.1177/1046496405277134>
9. Scalea, T. M., Rodriguez, A., Chiu, W. C., Brennenman, F. D., Fallon, W. F., Kato, K., ... & Bass, S. M. (1999). Focused Assessment with Sonography for Trauma (FAST): Results from an international consensus conference. *Journal of Trauma and Acute Care Surgery*, 46(3), 466–472.
10. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56. <https://doi.org/10.1038/s41591-018-0300-7>
11. Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.
12. Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). North-Holland. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
13. INACSL Standards Committee. (2021). *Healthcare Simulation Standards of Best Practice™: Simulation Design*. *Clinical Simulation in Nursing*, 58, 14–21. <https://doi.org/10.1016/j.cens.2021.08.009>
14. Shaffer, F., & Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Frontiers in Public Health*, 5, 258. <https://doi.org/10.3389/fpubh.2017.00258>
15. World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA*, 310(20), 2191–2194. <https://doi.org/10.1001/jama.2013.281053>
16. Amann, J., Blasimme, A., Vayena, E., Frey, D., & Madai, V. I. (2020). Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, 20(1), 310. <https://doi.org/10.1186/s12911-020-01332-6>
17. Arnsten, A. F. T. (2009). Stress signalling pathways that impair prefrontal cortex structure and function. *Nature Reviews Neuroscience*, 10(6), 410–422. <https://doi.org/10.1038/nrn2648>
18. Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. *WIREs Data Mining and Knowledge Discovery*, 9(4), e1312. <https://doi.org/10.1002/widm.1312>
19. Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2015). *Engineering psychology and human performance* (4th ed.). Psychology Press. <https://doi.org/10.4324/9781315665177>
20. Epstein, E. G., & Delgado, S. (2010). Understanding and addressing moral distress. *Online Journal of Issues in Nursing*, 15(3), Manuscript 1. <https://doi.org/10.3912/OJIN.Vol15No03Man01>
21. Froomkin, A. M., Kerr, I., & Pineau, J. (2019). When AIs outperform doctors: Confronting the challenges of a tort-induced over-reliance on machine learning. *Arizona Law Review*, 61, 33–99.
22. Maliha, G., Gerke, S., Cohen, I. G., & Parikh, R. B. (2021). Artificial intelligence and liability in medicine: Balancing safety and innovation. *The Milbank Quarterly*, 99(3), 629–647. <https://doi.org/10.1111/1468-0009.12504>
23. Groom, V., & Nass, C. (2007). Can robots be teammates? Benchmarks in human–robot teams. *Interaction Studies*, 8(3), 483–500. <https://doi.org/10.1075/is.8.3.10gro>
24. Hutchins, E. (1995). *Cognition in the wild*. MIT Press.
25. Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "Big Five" in teamwork? *Small Group Research*, 36(5), 555–599. <https://doi.org/10.1177/1046496405277134>

26. Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4), 589–597. <https://doi.org/10.1080/2159676X.2019.1628806>
27. Castaldo, R., Melillo, P., Bracale, U., Caserta, M., Petretta, M., & Pecchia, L. (2015). Acute mental stress assessment via short term HRV analysis in healthy adults: A systematic review with meta-analysis. *Biomedical Signal Processing and Control*, 18, 370–377. <https://doi.org/10.1016/j.bspc.2015.02.012>
28. Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association*, 19(1), 121–127. <https://doi.org/10.1136/amiajnl-2011-000089>
29. Croskerry, P. (2009). A universal model of diagnostic reasoning. *Academic Medicine*, 84(8), 1022–1028. <https://doi.org/10.1097/ACM.0b013e3181ace703>
30. Fairclough, S. H. (2009). Fundamentals of physiological computing. *Interacting with Computers*, 21(1-2), 133–145. <https://doi.org/10.1016/j.intcom.2008.10.011>
31. Klein, G. A. (1998). *Sources of power: How people make decisions*. MIT Press.
32. Meichenbaum, D. (2017). Stress inoculation training: A preventative and treatment approach. In *The evolution of cognitive behavior therapy* (pp. 117–140). Routledge.
33. Carayon, P. (2006). Human factors of complex sociotechnical systems in healthcare. *Applied Ergonomics*, 37(4), 525–535. <https://doi.org/10.1016/j.apergo.2006.04.011>
34. Hollnagel, E., Woods, D. D., & Leveson, N. (2006). *Resilience engineering: Concepts and precepts*. Ashgate Publishing.
35. Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, 36(6), 495–504. <https://doi.org/10.1080/10447318.2020.1741118>