AI and Safety – Failure Modes of Automated Systems Stats and Facts

FACTS

1. Automated systems can fail when sensors provide incomplete, noisy, or incorrect data, causing the AI to misjudge distances, obstacles, or human presence.
2. Software bugs, outdated firmware, or flawed algorithms can cause automation to act unpredictably, perform unintended motions, or execute tasks at the wrong time.
3. Automation systems can fail during edge-case conditions—unusual lighting, weather, or object shapes—because the AI model has not been trained on those scenarios.
4. Network or communication interruptions can stop commands mid-operation, delay emergency signals, or cause robots to "freeze" in unsafe positions.
5. Autonomous systems can behave dangerously when safety logic is overridden, disabled, or improperly configured during maintenance or setup.
6. AI-driven systems can misclassify human limbs, PPE, or tools, creating hazardous situations where the robot "thinks" the area is clear when it is not.

STATS

- A U.S. analysis found that nearly 40% of automation-related injury events involved sensor or detection failures, such as a machine not recognizing a worker in the danger zone. (Center for Occupational Robotics Research, NIOSH)
- In the US, AI-related incidents in workplaces rose by 30% from 2023 to 2025, with failure modes like hallucinations and bias amplification contributing to 15% of reported

safety breaches in industrial automation.

- By 2025, 44% of Canadian organizations deploying AI automated systems experienced at least one failure mode event, such as system misinterpretation leading to operational downtime or near-misses.
- US manufacturing sectors saw a 25% reduction in accidents from AI robots between 2020-2025, but failure modes in predictive monitoring caused 10% of residual incidents, including unexpected autonomous decisions.
- From 2021-2024, credential phishing and adversarial attacks on AI systems surged 703% in the US, exploiting failure modes like insufficient transparency and resulting in workplace security compromises.
- In Canada, 32% of business email compromise incidents involving AI tools in 2024 stemmed from multi-factor authentication bypass failures, amplifying risks in automated decision-making processes.