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

FAITS

1. Les systèmes automatisés peuvent échouer lorsque les capteurs fournissent des données incomplètes, bruitées ou incorrectes, ce qui conduit l'IA à mal évaluer les distances, les obstacles ou la présence humaine.
2. Les bogues logiciels, les micrologiciels obsolètes ou les algorithmes défectueux peuvent entraîner un comportement imprévisible de l'automatisation, des mouvements involontaires ou l'exécution de tâches au mauvais moment.
3. Les systèmes d'automatisation peuvent tomber en panne dans des conditions extrêmes (éclairage, conditions météorologiques ou formes d'objets inhabituels) car le modèle d'IA n'a pas été entraîné à ces scénarios.
4. Les interruptions de réseau ou de communication peuvent interrompre les commandes en cours d'exécution, retarder les signaux d'urgence ou provoquer le « gel » des robots dans des positions dangereuses.
5. Les systèmes autonomes peuvent se comporter de manière dangereuse lorsque la logique de sécurité est outrepassée, désactivée ou mal configurée pendant la maintenance ou la configuration.
6. Les systèmes basés sur l'IA peuvent mal classer les membres humains, les EPI ou les outils, créant ainsi des situations dangereuses où le robot « pense » que la zone est dégagée alors qu'elle ne l'est pas.

STATISTIQUES

- Une analyse américaine a révélé que près de 40 % des accidents liés à l'automatisation impliquaient des défaillances de capteurs ou de détection, par exemple une machine ne reconnaissant pas un travailleur dans la zone dangereuse. (Centre de recherche sur la robotique professionnelle, NIOSH)
- Aux États-Unis, les incidents liés à l'IA sur les lieux de travail ont augmenté de 30 % entre 2023 et 2025, les modes de défaillance tels que les hallucinations et l'amplification des biais contribuant à 15 % des violations de sécurité signalées dans le domaine de l'automatisation industrielle.
- En 2025, 44 % des organisations canadiennes déployant des systèmes automatisés basés sur l'IA ont connu au moins un incident lié à un mode de défaillance, tel qu'une mauvaise interprétation du système entraînant un arrêt opérationnel ou un quasi-accident.
- Les secteurs manufacturiers américains ont enregistré une réduction de 25 % des accidents liés aux robots IA entre 2020 et 2025, mais les modes de défaillance

dans la surveillance prédictive ont causé 10 % des incidents résiduels, y compris des décisions autonomes inattendues.

- Entre 2021 et 2024, le phishing d’identifiants et les attaques adversaires contre les systèmes d’IA ont augmenté de 703 % aux États-Unis, exploitant des modes de défaillance tels que le manque de transparence et compromettant la sécurité sur le lieu de travail.
- Au Canada, 32 % des incidents de compromission des e-mails professionnels impliquant des outils d’IA en 2024 résultaient d’échecs de contournement de l’authentification multifactorielle, amplifiant les risques dans les processus décisionnels automatisés.