

SÛRETÉ DE FONCTIONNEMENT

**« SYSTEMES AUTONOMES, VALIDATION
& QUALIFICATION »**

Différences entre Systèmes Autonomes & Systèmes Automatisés (1)

- Un système automatisé
 - ▣ Connaissance a priori de son environnement
 - ▣ Connaissance des règles de prise de décision
 - Si vecteur X en entrée, vecteur Y pour action en sortie
 - ▣ Reproductibilité de la décision

Différences entre Systèmes Autonomes & Systèmes Automatisés (2)

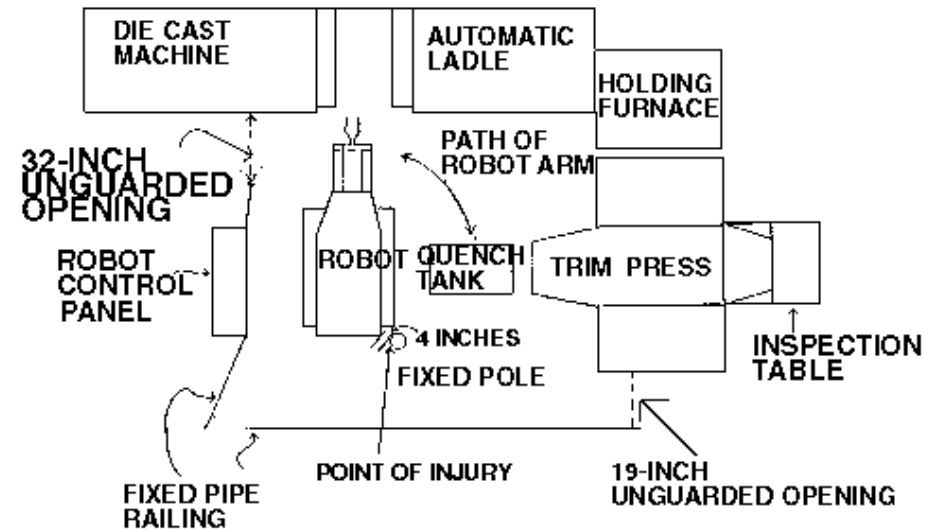
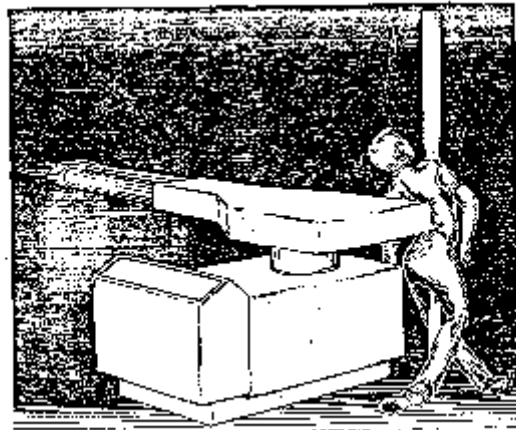
- Un Système Autonome
 - ▣ Absence de connaissance a priori de l'environnement
 - ▣ Règles de prise de décision
 - Soit déterministe, pour une situation donnée, génération de la même décision
 - Soit non-déterministe, pour une situation donnée, génération de plusieurs décisions possibles
 - ▣ Problème de l'explicabilité de la décision
 - Explicable : il est possible de comprendre le calcul ayant conduit à la décision
 - Non-Explicable : la n'est pas possible d'inférer le pourquoi de la décision a posteriori.

Limites des systèmes automatisés

- Environnement prédéfini
 - ▣ Ontologie permettant de caractériser les informations à posséder et à réaliser.
 - ▣ Ensemble de capteurs/estimateurs nécessaires à la détermination des informations pertinentes
 - ▣ Ensemble des actionneurs nécessaires à la réalisation des opérations nécessaires.
- Décision
 - ▣ Identification du domaine de validité du moteur de décision
 - ▣ Capacité de détecter les sorties de domaine => mise en sécurité du système.

Origine des fautes dans les Systèmes Automatisés (1)

- Violation d'une contrainte de fonctionnement
 - ▣ Ex: obstacle sur la voie quand la voie est supposée vide.
 - ▣ Ex: non-respect d'une règle d'exploitation.



Origine des fautes dans les Systèmes Automatisés (2)

- Flux d'informations erronées
 - ▣ Dysfonctionnement d'un capteur, d'un estimateur de mesure.
 - ▣ Corruption de données.
 - ▣ Injection de fausses données
- Erreur de prise de décision
 - ▣ Erreur d'implantation logicielle (Ariane 501)
 - ▣ Erreur de la logique de prise de décision



Origine des fautes dans les Systèmes Automatisés (2)

- Défaillance de l'interaction « Homme-Machine »
 - Confusion (Mauvaise Analyse ou mauvaise compréhension des informations)
 - Mauvaise interaction
 - Mauvaise décision



Validation & Vérification des Systèmes Automatisés

(1)

- Validation & Vérification du bon fonctionnement
 - ▣ Tests sur un ensemble de scénarii représentatifs.
 - ▣ Preuves formelles si nécessaires.
- Règles d'exploitation
 - ▣ Garantir l'absence de violation du domaine d'exploitation
- Mise en place de systèmes de protections
 - ▣ Prévenir les violations des règles d'exploitation.
 - ▣ Identifier les violations et mettre en sécurité.
 - ▣ Monitorer les prises de décisions.

Validation & Vérification des Systèmes Automatisés

(2)

- Améliorer le design en fonction de l'incidentologie
 - ▣ Analyse des défaillances, natures
 - ▣ Prise en compte dans le système ou ses évolutions
- Bilan:
 - ▣ Concevoir un Système Automatisé est maîtrisé mais peut-être très couteux.
 - ▣ La concurrence des fonctions automatisés peut conduire à l'émergence de comportements non-désirables qui doivent être identifiés lors de la conception.

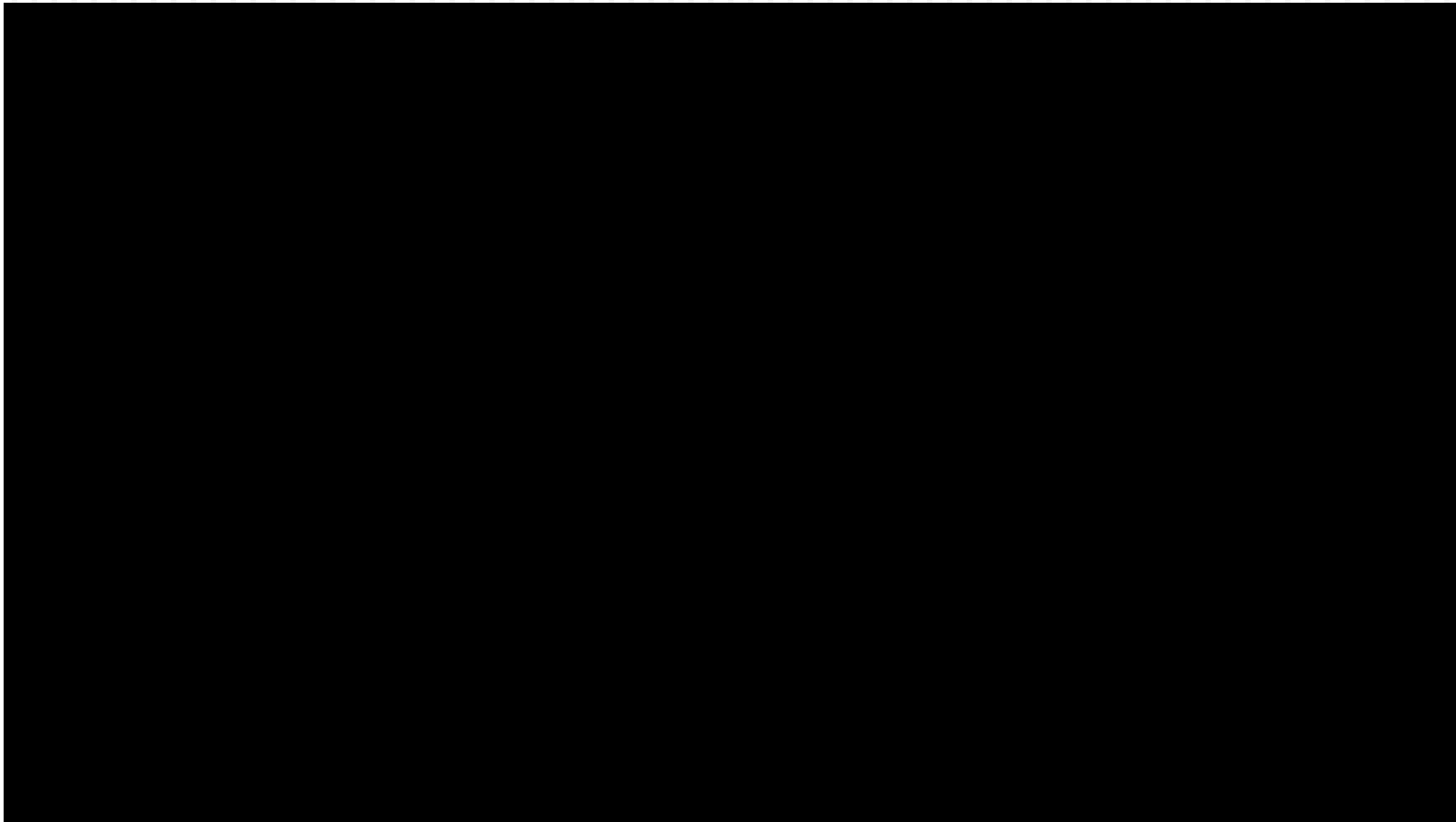
Quand parlons nous de Systèmes Autonomes

- Augmentation de la capacité à prendre une décision
 - ▣ Dans un environnement « nouveau », des conditions « nouvelles ».
 - ▣ Dans un environnement complexe (nombres d'interactions importants, acteurs de nature différentes).
- La décision peut-être :
 - ▣ Prédéfinie (ie. l'algorithme n'évolue pas dans le temps).
 - ▣ Conditionnée (par un apprentissage par exemple).

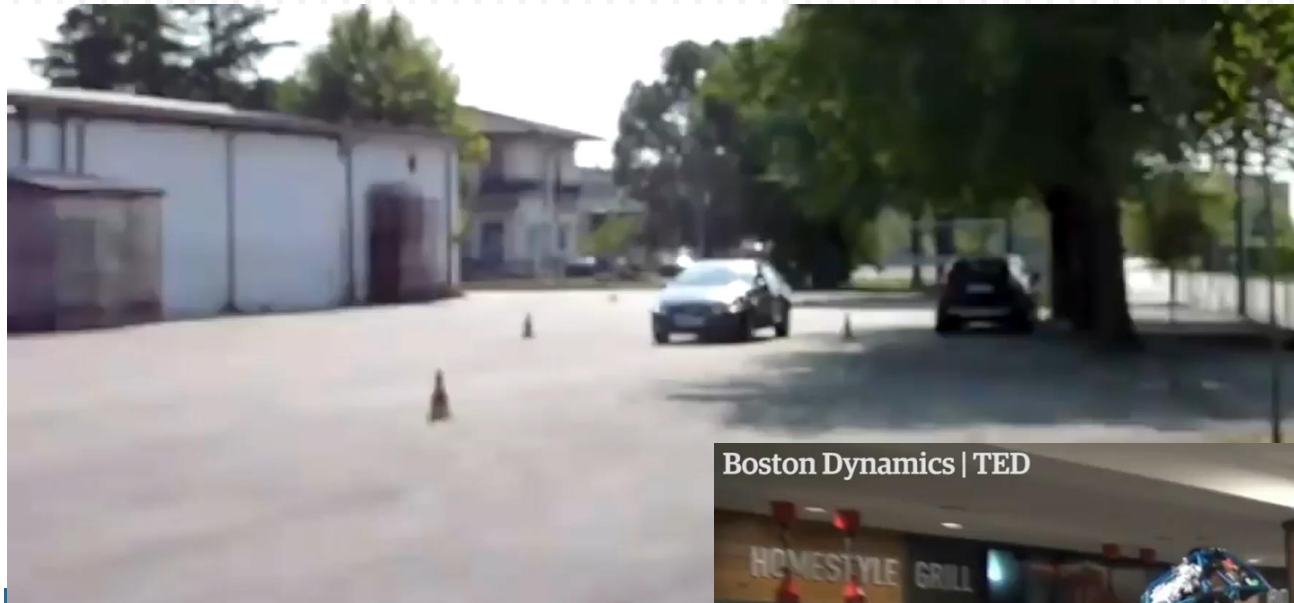
Les Systèmes Autonomes & l'IA

- Systèmes Autonomes n'impliquent pas IA
 - ▣ Peut effectuer des traitements complexes sans IO
 - Traitement & analyse d'images
 - Traitement du signal
 - ▣ Peut prendre une décision sans IA
 - Système expert (Rule Based)
 - MDP
 - ...
 - ▣ Intérêt de l'IA pour les systèmes autonomes
 - Richesse et performance de l'analyse de la décision.
 - Simplification de l'implantation !
 - Apprentissage peut-être plus « simple » que développer un algorithme de contrôle-commande complexe de coordination par exemple.

Apprentissage par Renforcement & Génération de contrôleurs



Cependant l'erreur est aussi « Robotique »



Boston Dynamics | TED



Origine des Fautes dans les Systèmes Autonomes

**TOUTES LES FAUTES DES SYSTEMES
AUTOMATISES**

+

**TOUTES LES FAUTES LIEES A LA PRISE DE
DECISION EN ENVIRONNEMENT INCERTAIN**

Problème de la décision dans les Systèmes Autonomes

- Comment caractérise-t-on une bonne décision ?
 - En terme de performances ?
 - En terme d'occurrence de risques ?
 - En terme d'éthique ?
- Comment mesure-t-on la performance d'une décision ?
 - Vitesse d'exécution ?
 - Perception par rapport à ce que ferait l'humain ?
 - Coût/Bénéfice ?

Problème : Optimisation multicritère

- Optimiser la performance
 - Possibilité de rouler sur une route de 2m de large à 200 km/h
 - Problème: Survenance d'une défaillance d'un pneumatique ?
- Optimiser la safety
 - Prendre en compte l'ensemble des évènements quelqu'en soit l'occurrence
 - Problème: comportement « sous-efficace », éventuellement comportement aberrant.
- Optimiser la consommation
- Optimiser la perception sensorielle

Pistes pour définir la qualité d'une décision

- Prendre une « bonne décision humaine » comme référence :
 - ▣ Satisfaisante relativement au niveau de prise de risques.
 - ▣ Performantes par rapport au critère des « opérateurs qualifiés ».
 - ▣ Répondant aux attentes :
 - En terme de confiance,
 - En terme d'appréciation subjective.
 - ▣ S'appréciant dans un domaine normatif donné.

Piste pour définir la qualité d'une décision

- Mesurer la différence entre la décision proposée par les système et celle proposée par l'humain
 - ▣ Simple dans le cas de jeux
- Plus délicate dans les cas complexes comme :
 - ▣ Copilote virtuel
 - ▣ Véhicule Autonome



Problématique de la métrique de la performance

- Exemple : Véhicule Autonome
 - ▣ Métrique actuelle : nombres d'accidents par millions de km parcourus.
- Limite de la métrique
 - ▣ Ne prend pas en compte les défaillances systémiques (a priori sans objet pour l'humain).
 - ▣ Ne prend pas en compte la notion de performance de la conduite, son intégration.
 - ▣ Ne prend pas en compte la QoS de la conduite autonome.
 - ▣ Ne prend pas en compte les variétés de missions réalisées par le véhicule.

Problématique de la mesure de la métrique

- Qualification « a priori »
 - ▣ L'estimation de la mesure doit être réalisée a priori.
 - ▣ **Problème :**
 - Si la métrique consiste à un nombre de kms parcourus sans accident, comment effectuer ces tests ?
 - Si la métrique consiste en un certain nombre de situations à gérer, comment effectuer ces tests ?

Les problèmes de la décision

- Performances d'un système type reconnaissance de chiffres manuscrits (obtenu par mesure sur un dataset)
 - ▣ **Meilleur algorithme** : 0,23% (Deep Learning)
- Problème :
 - ▣ Comparaison avec l'humain (à peu près au même niveau de non reconnaissance)
 - ▣ L'humain peut prendre une décision supplémentaire
 - Je ne sais pas déchiffrer.
 - ▣ Prise en compte de l'évolution de l'écriture dans le temps.
- Performance bien plus mauvaise pour l'analyse de situation plus complexe
 - ▣ 80 à 90 % de reconnaissance des objets dans les meilleures situations.

Les premières réponses (1)

- La validation par les tests réels
 - ▣ Mise en condition réelle.
 - ▣ Comparer le niveau de performance avec la performance humaine.
 - ▣ Estimer le delta objectif et subjectif.
- Limites
 - ▣ Uniquement quelques scénarii.
 - ▣ Uniquement quelques contextes.

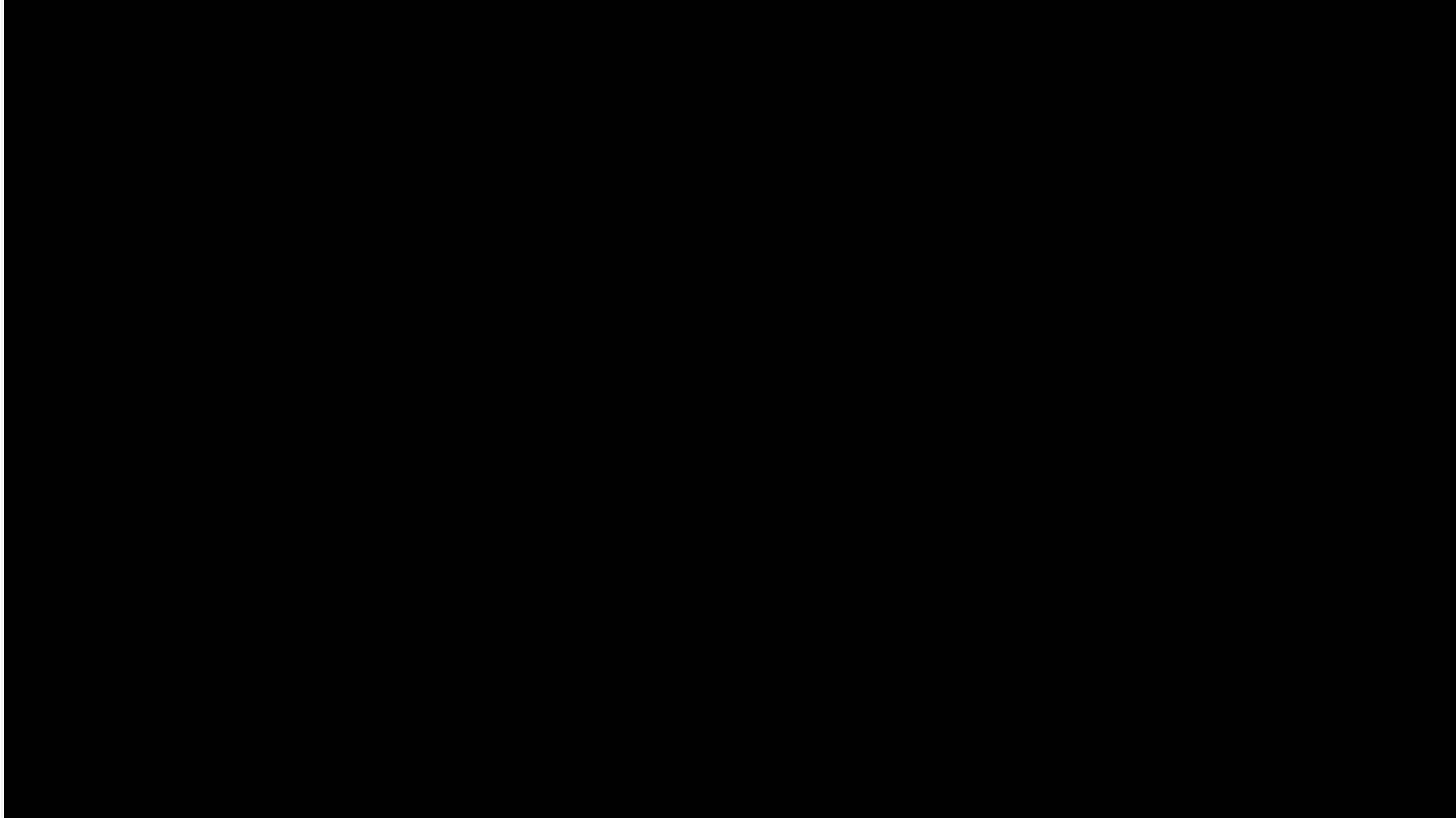
Les premières réponses (2)

- La validation par simulation
 - ▣ Réaliser un ensemble de tests dans un simulateur.
 - ▣ Comparer le niveau de performance avec la performance humaine.
 - ▣ Estimer le delta objectif et subjectif.
- Avantages
 - ▣ Réalisation de nombreuses situations.
 - ▣ Coût d'évaluation plus modéré.
- Limites
 - ▣ Biais induit par l'environnement simulé.
 - ▣ Uniquement quelques contextes.

Le biais induit par la simulation

- Différence Environnement simulé/Environnement réel
 - ▣ Moins de variabilité.
 - ▣ Moins d'information.
 - ▣ Structure répétitive.
 - ▣ Biais induit sur les capteurs.
- Génération de situations pertinentes/représentatives
 - ▣ Simulation de situations dangereuses uniquement -> Difficilement comparable en terme d'accident par km parcourus.
 - ▣ Simulation aléatoire de parcours -> problème de représentativité du parcours.

Monde virtuel/Monde réel



Les approches formelles

- **Stabilité de de la réponse** : deux situations relativement proches doivent engendrer la même décision
- Comment garantir une telle stabilité de réponse ?
 - Deep Neural Network : a priori pas de solution lors de la conception.
- Proposition:
 - Introduire un modèle formel approché du modèle de réseaux de neurones employés vérifiant que toute propriété de stabilité dans le modèle formel est correct dans le modèle approché.
 - Démontrer la stabilité du modèle approché
 - En déduire la stabilité du modèle implanté.
- Limite / Défaut
 - Force de l'IA : tenir compte des signaux faibles.
Risque de masquer certains signaux faibles pour conserver la stabilité.

Les approches systèmes

- Panacher plusieurs systèmes de prise de décision [Diversité d'implantation]
- Adjoindre pour chacune des chaînes de prise de décision un estimateur de confiance [Moniteur de QoS]
- Effectuer un voir sur la solution en prenant les différences entre les solutions pondérées par l'estimateur de confiance dans la solution. [Générateur de décision/Détection d'une absence de capacité à répondre]

Les problèmes résiduels

Absence de solutions « satisfaisantes »

« La voiture autonome sera-t-elle programmée pour tuer son conducteur ? »

Pour éviter une collision imminente avec une foule de personnes, quelle décision ?

- ▣ **braquer brusquement** au risque de **tuer son conducteur ...**
- ▣ continuer et **percuter** la foule ?

Conclusions (1)

- Nécessité de développer une quantification de la performances de la décision
 - ▣ Définition de nouvelles métriques.
 - ▣ Définition de nouvelles méthodes de tests pour estimer cette performance.
 - ▣ Comparaison de la métrique humaine vs. la métrique système
- Séparer les deux problématiques de l'implantation et de la décision
 - ▣ L'implantation -> Problématique des Systèmes Automatisés
 - ▣ La qualification de la décision -> Problématique de la qualité de la décision, validation de celle-ci (à développer)

Conclusions (2)

- Définir la notion de GAME entre Décision d'un Système Autonome et Décision Humaine.
- Développer une nouvelle incidencologie
 - Distinguer Erreur de Décision/Défaillance de l'implantation.