



Quelques Questions Ethiques en Intelligence Artificielle et en Robotique

Raja Chatila

Institut des Systèmes Intelligents et de Robotique (ISIR)

CNRS et Université Pierre et Marie Curie, Paris

Membre de la CERNA

Commission de réflexion sur l'éthique de la recherche en sciences et technologies du numérique d'Allistene



Le robot: un paradigme de l'IA « incorporée »

- Un « robot » est une *machine matérielle* munie de capteurs et d'actionneurs et contrôlée par des logiciels implémentés sur des ordinateurs.
- Elle possède des capacités de perception de son environnement, d'action et de mouvement,
- Elle possède des capacités de prise de décision pour effectuer des tâches (auto) assignées.
- Elle est (éventuellement) capable de communiquer et d'apprendre.

Remarques

- Ces capacités peuvent être développées à divers degrés de **complexité** et confèrent aux robots des niveaux d'**autonomie** différents.
- Ces capacités peuvent être intégrées dans un seul « corps » ou réparties (*i.e.*, perception ou décision à distance par rapport à l'action).
- Il faut aussi considérer les systèmes multi-robots et les essaims (comportement collectifs et émergents).

Etat de la Robotique

- La recherche a atteint un certain degré de maturité pour certaines fonctions opérationnelles en perception, planification et contrôle du mouvement, interaction humain-robot, ...
- L'opération autonome est possible dans des cas relativement bien définis.
- Des applications dans plusieurs domaines deviennent possibles.

Applications de la Robotique: production industrielle, milieux dangereux, transport, défense, agriculture, construction, santé, loisirs, ...

- Remplacer les humains



- Assister et servir les humains



- Rehabilitier/augmenter les humains



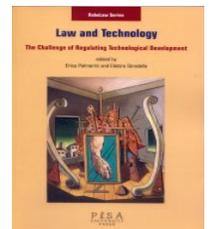
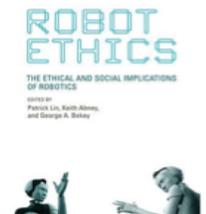
Prise de conscience éthique

- Questionnement sur les problèmes éthiques, légaux et sociétaux (ELS) dans l'usage des robots depuis une douzaine d'années.
- Questionnements récents (ou renouvelés) sur l'emploi, sur les dangers potentiels de l'IA (singularité, ...) et la de la robotique.
- Questions soulevées par des domaines particuliers comme l'usage d'armes létales *autonomes* (SALA).

Le grand public est souvent mal informé du véritable état de l'art.

Quelques sujets ELS

- Prise de décision autonome
- Responsabilité morale et juridique
- Protection de la vie privée et surveillance
- Dignité humaine
- Dépendance, isolation
- Liens affectifs
- Augmentation de l'homme
- Imitation du vivant
- Identité humaine
- Ethique de l'usage des robots
- Statut du robot dans la société



Ethique de la recherche, de l'usage, de la machine

- Des questions liées:
 - Ethique de la recherche: respect de préconisations ou de règles par le concepteur, le chercheur (p. ex le clonage); *recherche responsable*.
 - Ethique de l'usage: utilisation respectant des règles éthiques
 - Ethique des systèmes : respect de règles ou de choix éthiques dans dans le comportement du système (p. ex: dilemmes)

- CDEFI
- CEA
- CNRS
- CPU
- Inria
- Institut Mines-télécom

CERNA

Commission de réflexion sur l'Éthique de la Recherche
en sciences et technologies du Numérique
d'Allistene

Ethique de la Recherche en Robotique: Préconisations de la CERNA



(Novembre 2014)

http://cerna-ethics-allistene.org/digitalAssets/38/38704_Avis_robotique_livret.pdf

La CERNA

<http://cerna-ethics-allistene.org>

- Commission de réflexion sur l’Ethique de la Recherche en sciences et technologies du Numérique d’Allistene
- Installée en Octobre 2012, renouvelée en janvier 2016
- Un vingtaine de membres
- Emet des avis à partir de saisines
- Organise des séminaires ouverts

(p. ex « Gouvernance des algorithmes »
1/02/2016, Paris)

Saisine à l’initiative du Cocor d’Allistene : Ethique de la recherche en robotique

La recherche sur les systèmes autonomes et les robots constitue un domaine important des sciences du numérique. Son impact sur la société à travers une vaste gamme d’applications techniques promet de modifier plusieurs secteurs industriels et économiques et s’étend jusqu’aux aspects privés de la vie humaine.

L’utilisation des robots à des fins militaires, de soins palliatifs, de soins pour les enfants ou afin d’accroître le plaisir dans une pratique de divertissement pose d’ores et déjà des problèmes éthiques. Sur le plan de la recherche, cela demande de réfléchir aux limites de l’autonomie des robots, de leur pouvoir décisionnel, de l’affectivité intégrée et de la simulation des émotions par une machine. On recensera de tels cas concrets et on se demandera dans quelle mesure et de quelle façon le chercheur doit intégrer les contraintes éthiques, y compris les problèmes envisageables à l’avenir, aux stades de définition et d’exécution de son projet de recherche.

On se posera la question de responsabilité du chercheur pour les comportements et les actions d’un robot ou d’un système autonome dont il a assumé la conception. On analysera la pertinence de la législation en vigueur et on formulera des recommandations en vue de l’évolution des normes juridiques.

Méthode de la CERNA

- Etude de cas posant des questions ELS, p. ex:
 - Défense et sécurité
 - Réhabilitation et augmentation
 - Robots assistants pour personnes vulnérables
 - Robot compagnons pour le grand public
- 4 capacités soulevant des questions éthiques au **chercheur**:
 - Autonomie
 - Augmentation de l'homme
 - Emotions et liens émotionnels
 - Imitation du vivant

Défense et sécurité

- Drones, essaims de robots, armes autonomes, véhicules robotisés



Predator



Crusher, CMU

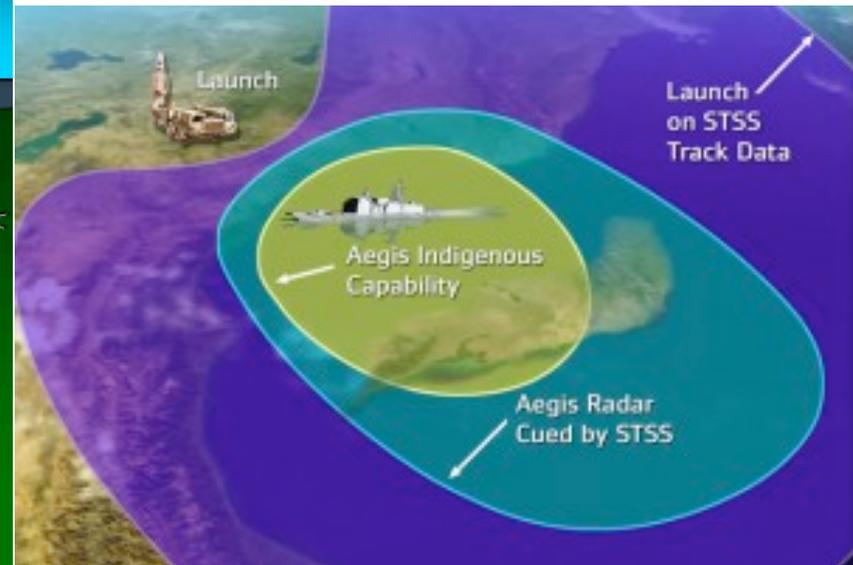
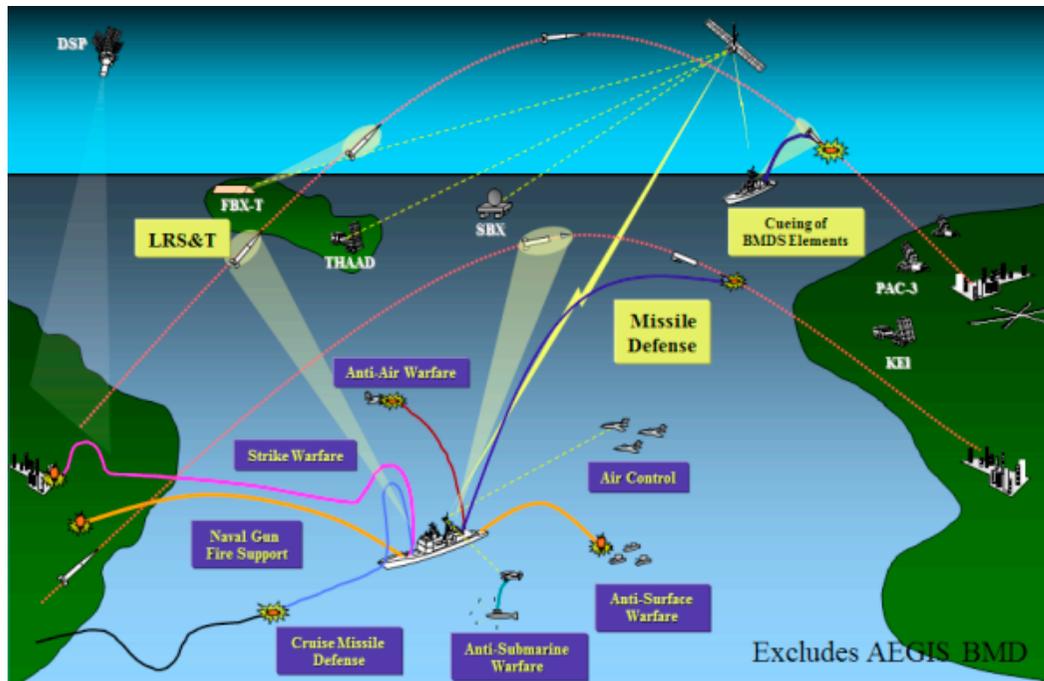


Drones for security monitoring

Packbot, IRobot

Systemes d'Armes Létales Autonomes (?)

Systeme AEGIS de defense anti-missile (US D.o.D)



Systemes d'Armes Létales Autonomes (?)



Phalanx
(US D.o.D)



Samsung SGR-A1

Systemes d'Armes Létales Autonomes (?)



Drone Reaper

How drones work



SAL « A »

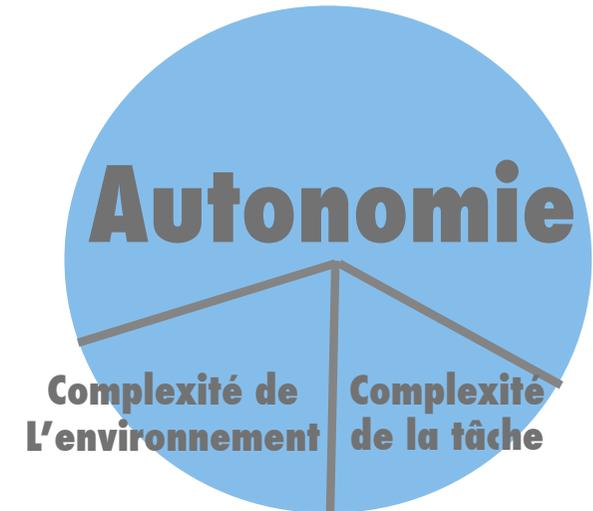
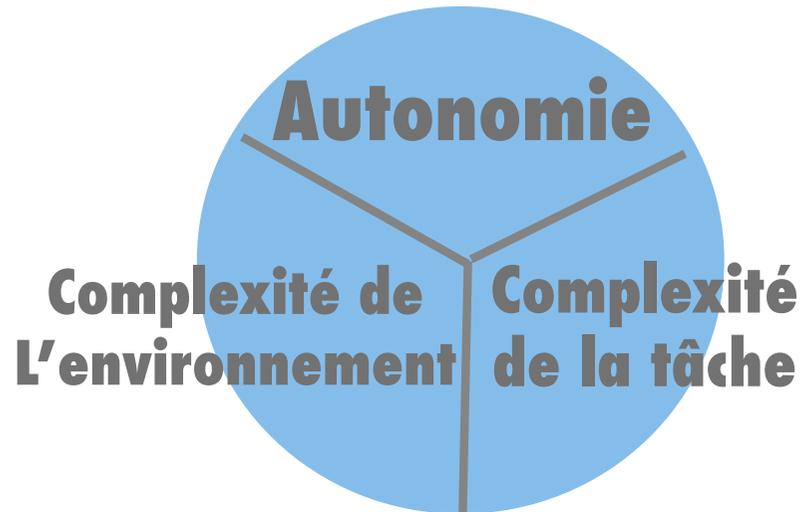
- Distinguer les systèmes *automatiques* des systèmes *autonomes*
- L'autonomie est réduite à la **désignation** et au traitement de l'objectif
- La définition de caractéristiques (ou signature) de l'objectif incombe à l'homme
- La capacité de **discerner correctement** un objectif dépendra des facultés de perception et d'interprétation de situations de la machine qui sont limitées par le complexité du contexte.

L'autonomie

Capacité d'un agent à prendre des décisions sans l'assistance d'un autre agent.

- Autonomie opérationnelle:
 - Concerne le traitement des données et leur interprétation, des décisions simples, la commande et l'exécution des actions.
 - Nécessaire dans tous les systèmes
- Autonomie décisionnelle: Concerne une perception et l'interprétation plus sémantique, l'interprétation de situations et la prise de décisions non triviales.
- *Il y a un continuum entre le système automatique et le système autonome ouvert.*

L'autonomie est un concept relatif



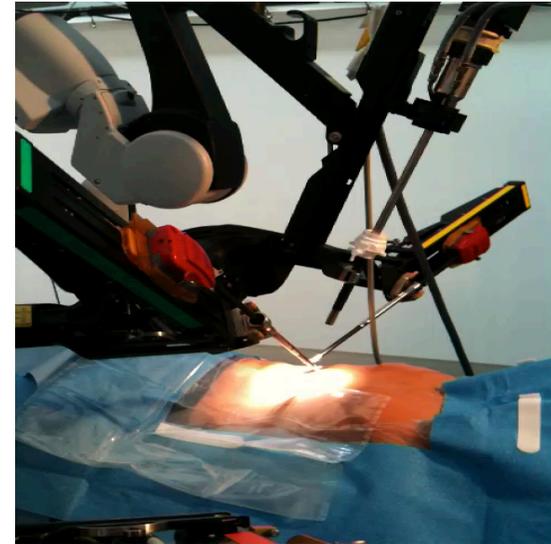
- L'autonomie atteignable dépend de la complexité de l'environnement et de celle de la tâche.
- La complexité de l'environnement peut être mesurée par la quantité (et variabilité) de l'information et son flux.
- La complexité de la tâche dépend de la dimension et de la structure de l'espace d'état du processus de décision.

Exemples

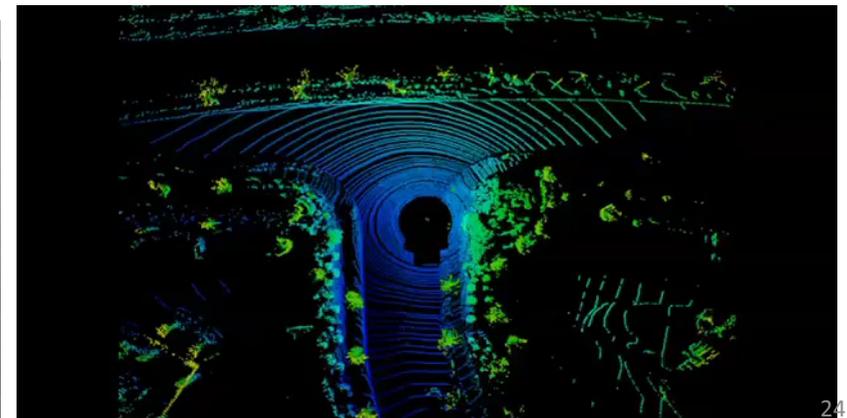


Autonomie opérationnelle

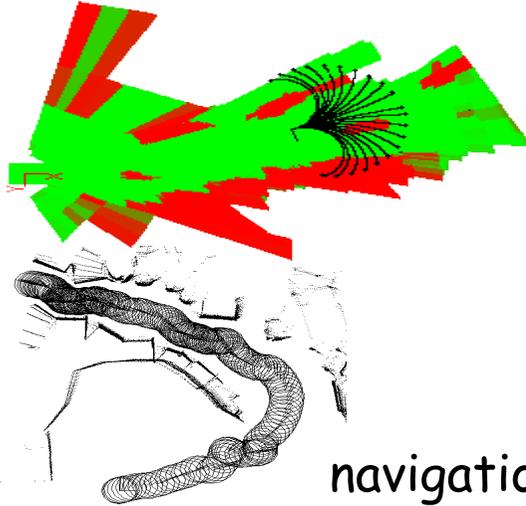
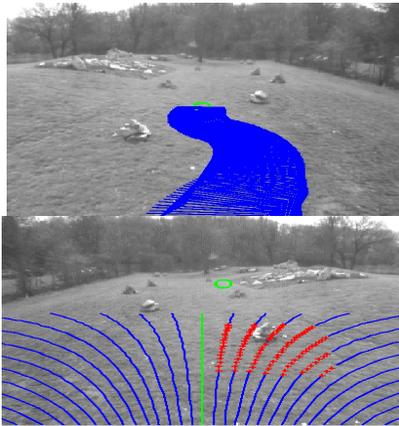
Téléopération:
Contrôle humain



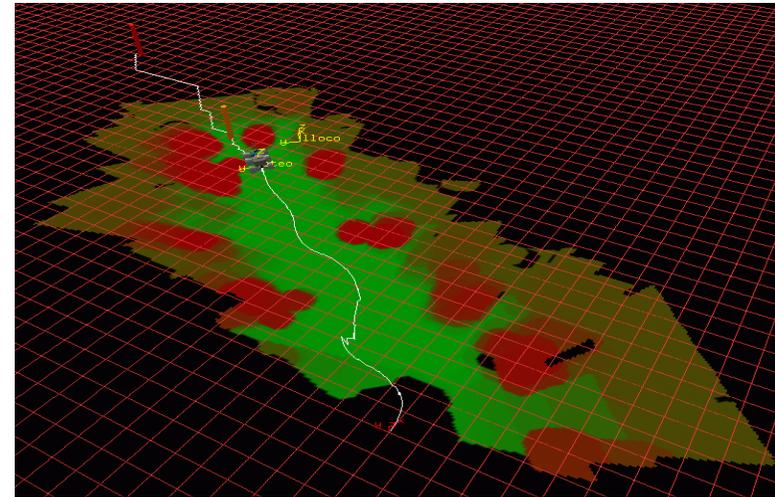
24/03/2016



Exemple d'autonomie opérationnelle: la navigation



navigation locale



Stratégie de navigation
Planification de trajectoire



24/03/2016

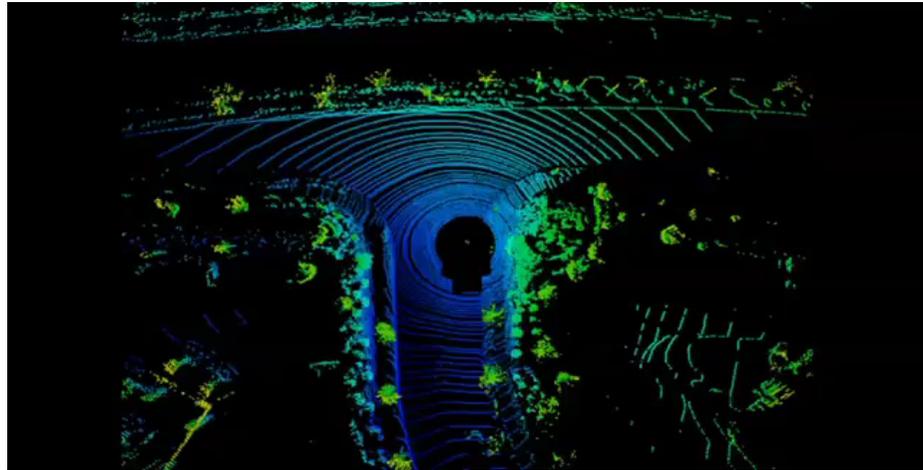


Véhicules autonomes

- Aspects légaux et dilemmes éthiques

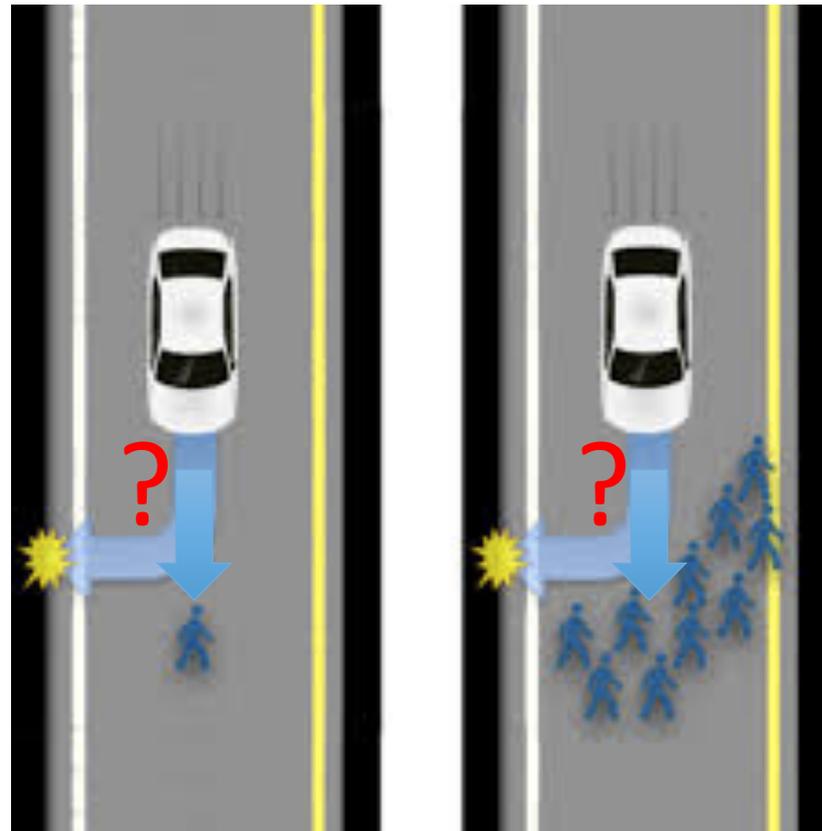


Google

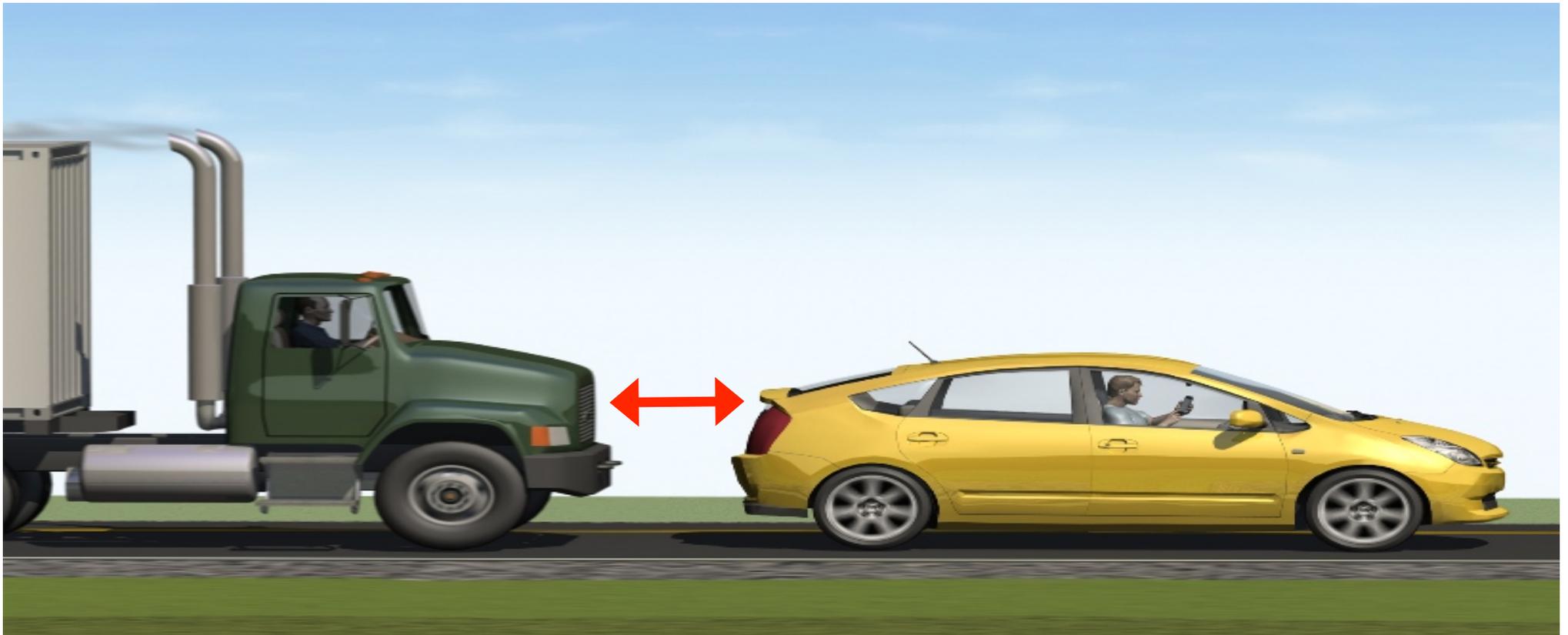


Nissan

Dilemmes éthiques



Influence du contexte



Situations prévisibles



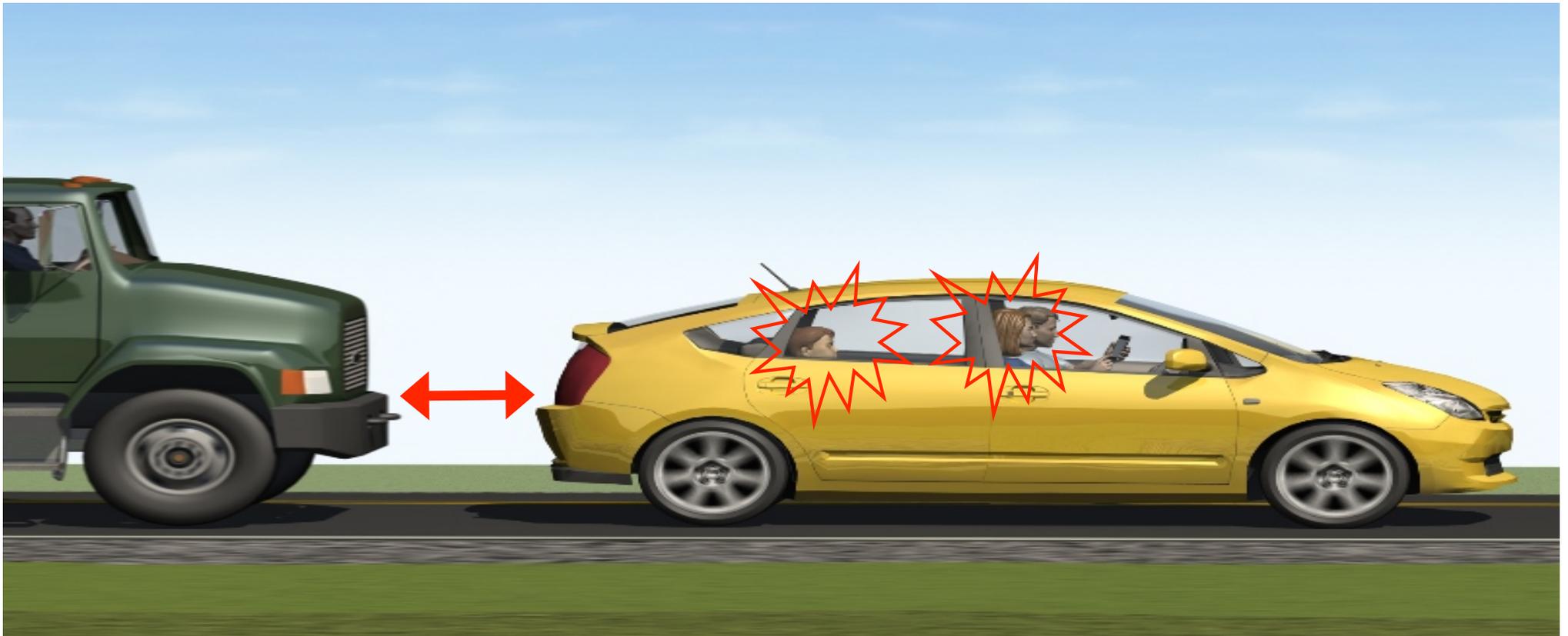
Obstacles imprévus



Evénements imprévus avec suite prévisible



Influence du contexte



Autonomie et partage d'autorité

- Machine
 - Capacités limitées par sa programmation
 - Incertitudes de perception et d'interprétation (ex., combattant/non-combattant) ;
 - Effectue l'action (avec incertitude parfois)
 - Rapidité de traitement
- Operateur humain
 - Occupé, faillible, stressé, émotif (e.g., : attentional tunnelling)
 - Distanciation par rapport au théâtre d'opération et aux actions de la machine (buffer moral)
 - Procédures non forcément suivies selon le contexte
 - Accomplissement de scénario
- Interaction
 - Perturbation ou absence de communications
 - Biais d'automatisation: surestimation de la confiance en la machine
 - Surprises de l'automatisation: ignorance de l'état exact de la machine

Autonomie et capacités décisionnelles 1/4

- **[AUT-1] Reprises en main**

Le chercheur doit se poser la question des reprises en main que l'opérateur ou l'utilisateur peut effectuer (au détriment du robot) et que la machine peut effectuer (au détriment de l'humain), des circonstances qui les permettent ou les rendent obligatoires. Il doit également étudier la possibilité ou non laissée à l'humain de « débrayer » les fonctions autonomes du robot.

- **[AUT-2] Décisions à l'insu de l'opérateur**

Le chercheur doit faire en sorte que les décisions du robot ne soient pas prises à l'insu de l'opérateur afin de ne pas créer de ruptures dans sa compréhension de la situation (c'est-à-dire afin que l'opérateur ne croie pas que le robot est dans un certain état alors qu'il est dans un autre état).

Autonomie et capacités décisionnelles 2/4

- **[AUT-3] Influences sur le comportement de l'opérateur**

Le chercheur doit être conscient des phénomènes de biais de confiance, c'est-à-dire la tendance de l'opérateur à s'en remettre aux décisions du robot, et de distanciation morale (« Moral Buffer ») de l'opérateur par rapport aux actions du robot.

- **[AUT-4] Limites des programmes**

Le chercheur doit être attentif à évaluer les programmes de perception, d'interprétation et de prise de décision et à en expliciter les limites. En particulier, les programmes qui visent à conférer une conduite morale au robot sont soumis à de telles limites.

Autonomie et capacités décisionnelles 3/4

- **[AUT-5] Caractérisation d'une situation**

En ce qui concerne les logiciels d'interprétation du robot, le chercheur doit évaluer jusqu'à quel point ceux-ci peuvent caractériser correctement une situation et discriminer entre plusieurs situations qui semblent proches, surtout si la décision d'action prise par l'opérateur ou par le robot lui-même est fondée uniquement sur cette caractérisation. Il faut en particulier évaluer comment les incertitudes sont prises en compte.

- **[AUT-6] Prévisibilité du système humain-robot**

De manière plus globale, le chercheur doit analyser la prévisibilité du système humain-robot considéré dans son ensemble, en prenant en compte les incertitudes d'interprétation et d'action, ainsi que les défaillances possibles du robot et celles de l'opérateur, et analyser l'ensemble des états atteignables par ce système.

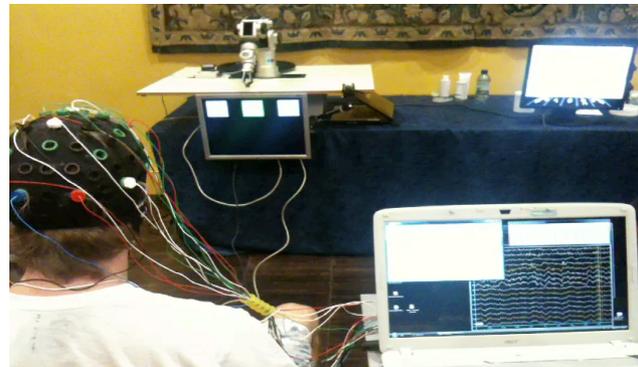
Autonomie et capacités décisionnelles 4/4

- **[AUT-7] Traçage et explications**

Le chercheur doit intégrer des outils de traçage dès la conception du robot. Ces outils doivent permettre d'élaborer des explications, même limitées, à plusieurs niveaux selon qu'elles s'adressent à des experts de la robotique, à des opérateurs ou à des utilisateurs.

Réhabilitation et augmentation de l'homme

EPFL



U. Saragozza



RIC Institute



Ekso
Bionics

Centre d'Alembert, Orsay



RB3D

24/03/2016

Assistance aux personnes vulnérables

- Personnes âgées, handicapées, enfants



S. Korea



Telenoid - Osaka U.



Paro



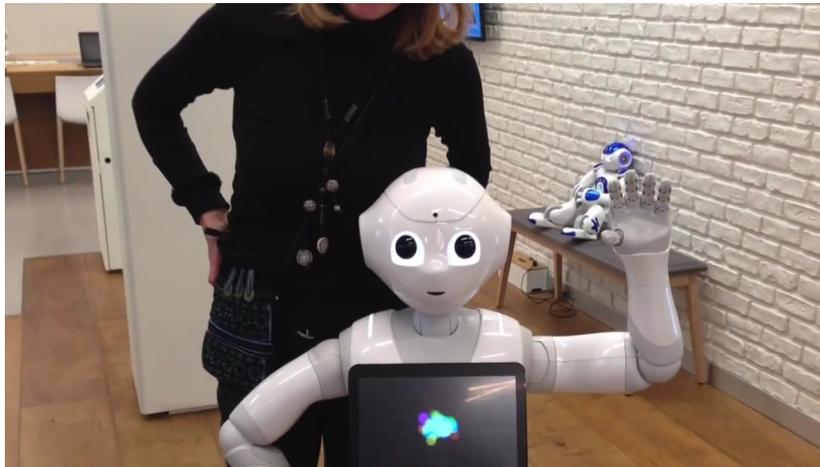
Aldebaran Romeo

Robots compagnons et émotionnels

Aldebaran Nao



Aldebaran (Softbank)
Pepper



The Big Bang Theory

Bio-mimétisme



24/03/2016

Osaka U
Centre d'Alembert, Orsay

42

L'imitation du vivant et l'interaction affective et sociale avec les humains 1/3

- **[IVI-1] Utilité au regard des finalités**

Le chercheur étudiera, au regard des fonctions utiles du robot, la pertinence et la nécessité de susciter des émotions et la pertinence et la nécessité de l'aspect ou du comportement biomimétiques, notamment dans les cas de forte ressemblance visuelle ou comportementale entre un robot et un être vivant. Dans les cas où l'apparence ou la voix humaines sont imitées, le chercheur s'interrogera sur les effets que pourrait avoir cette imitation, y compris hors des usages pour lesquels le robot est conçu.

- **[IVI-2] Frontière vivant - artefact**

Si une ressemblance quasi parfaite est visée, le chercheur doit avoir conscience que la démarche biomimétique peut brouiller la frontière entre un être vivant et un artefact. Le chercheur consultera sur ce brouillage le comité opérationnel d'éthique de son établissement.

L'imitation du vivant et l'interaction affective et sociale avec les humains 2/3

- **[IVI-3] Étude des effets**

Pour les projets de recherche qui ont trait au développement de la robotique affective, le chercheur s'interrogera sur les répercussions éventuelles de son travail sur les capacités de socialisation de l'utilisateur.

- **[IVI-4] Interaction enfant-robot**

Pour les projets qui mettent en présence des enfants et des robots, le chercheur doit se poser la question de l'impact de l'interaction enfant-robot sur le développement des capacités émotionnelles de l'enfant, tout particulièrement dans la petite enfance.

L'imitation du vivant et l'interaction affective et sociale avec les humains 3/3

- **[IVI-5] Évaluation**

Pour les projets de recherche relatifs à des robots susceptibles d'avoir des effets sur l'affectivité des utilisateurs et de susciter leur attachement, le chercheur devra élaborer un protocole de conception et d'évaluation en veillant à impliquer les compétences multidisciplinaires nécessaires et des utilisateurs potentiels.

- **[IVI-6] Communication**

Le chercheur doit être prudent dans sa communication sur les capacités émotionnelles des robots et sur l'imitation de la nature et du vivant, notamment parce que l'expression des émotions, au sens humain, par un robot, est un leurre, et parce que l'imitation du vivant peut amener, volontairement ou pas, à prêter à l'artefact des caractéristiques du vivant.

Réparation et augmentation de l'humain par la machine 1/3

[RAH-1] Éthique médicale

Les chercheurs en robotique réparatrice ou d'assistance doivent appliquer, en coordination avec les professionnels de santé, les aidants et les patients, les principes d'éthique en usage dans le secteur médical afin d'arbitrer entre les exigences d'efficacité et de sécurité des soins, celles d'autonomie et d'intégrité de la personne et, enfin, de protection de la vie privée. Ce sont là des questions qui relèvent de l'éthique et non uniquement du droit en cela qu'elles demandent à être arbitrées dans chaque cas particulier et qu'elles ne reçoivent pas de réponse générale. Pour en traiter, il faudra prendre avis auprès de comités opérationnels d'éthique des sciences médicales et veiller à ce que les compétences technologiques y soient étroitement associées.

Réparation et augmentation de l'humain par la machine 2/3

- **[RAH-2] Autonomie et intégrité de l'individu**

Dans le cas des organes robotisés à vocation réparatrice, le chercheur aura le souci de la préservation de l'autonomie de l'individu équipé, à savoir de la maîtrise qu'il conservera autant que faire se peut sur ses actions, et de la conservation de l'intégrité des fonctions autres que celles concernées par la réparation.

- **[RAH-3] Réversibilité de l'augmentation**

Dans le cas des dispositifs robotisés visant l'augmentation, le chercheur veillera à la réversibilité de celle-ci : les dispositifs doivent être amovibles sans dommage pour la personne, autrement dit, sans que la personne perde l'usage de ses fonctions initiales.

Réparation et augmentation de l'humain par la machine 3/3

- **[RAH-4] Discrimination induite par l'augmentation**

Le chercheur se posera la question de l'incidence de l'augmentation des facultés et des capacités humaines induites par les dispositifs qu'il développe sur le comportement social de ceux qui en bénéficient ainsi que, symétriquement, de ceux qui n'en bénéficient pas.

Remarques de conclusion

- L'éthique pose des questions prégantes dans certains domaines: véhicules autonomes, SALA, personnes vulnérables.
- Les autorités sont intéressées par le cadrage des questions ELS en robotique et IA (Parlement européen, français - OPECST).
- Accélération des réflexions sur le plan international (France, Japon, USA, UK, Europe, IEEE, ...)
- Vers des chartes ou des standards respectant des préconisations éthiques