

## **Session 28 – Repenser la responsabilité à l'heure de l'Intelligence Artificielle (IA) ?**

La responsabilité est, en droit, l'obligation de réparer le dommage que l'on a causé. Au Moyen âge, on le sait, des animaux pouvaient être traduits en justice et condamnés. Puis les conceptions juridiques ont évolué : la capacité de discernement est devenue indispensable à la responsabilité. On ne peut demander des comptes à l'animal, non doué de conscience de soi et agissant, comme le décrivait Descartes, en machine. Non plus qu'aux personnes mentalement déficientes et aux enfants incapables d'envisager les conséquences de leurs comportements. Le principe qui domine jusqu'à présent est ainsi que les dommages doivent être rapportés à la personne consciente, seule responsable : le parent, l'éducateur, l'adulte sain d'esprit. Alors que le monde est de plus en plus peuplé d'artefacts se déplaçant, interagissant avec les humains, prenant mille décisions à notre place, la notion de responsabilité entre dans une crise profonde. Les trois dimensions essentielles de nos sociétés sont menacées : la transaction, la relation et l'information.

Le premier défi est celui d'un brouillage des responsabilités lourdement handicapant pour l'économie. L'omniprésence des technologies multiplie les responsables potentiels. Il est très difficile d'en démêler l'écheveau. Un accident de voiture devra-t-il mettre en jeu la responsabilité du conducteur qui n'a pas su reprendre le contrôle du véhicule à temps, du fournisseur de l'algorithme, de son concepteur, du constructeur de la voiture, de l'institution qui commercialise la flotte automobile ? Les enquêtes complexes menées dans le cas des accidents aériens pour établir exactement la chaîne des erreurs dans les interactions entre l'humain, le matériel et le logiciel deviendront nécessaires dans le cas des accidents automobiles. Sans responsabilités claires, pas de confiance possible, et sans confiance pas d'échanges économiques. Les entreprises, dont la responsabilité pénale peut être fatale et qui ne pourront identifier leurs risques, auront du mal à développer leurs activités.

La crise de la responsabilité fait aussi peser une menace nouvelle sur la société. Cette dernière risque de voir lui échapper le contrôle des interactions qui font son existence. L'économie dépend de plus en plus profondément de processus automatisés réalisés sans l'intervention directe de l'homme. A mesure que nous entrons dans une nouvelle forme de capitalisme fondée sur la production et l'exploitation de masses considérables de données, seuls des processus informatiques, nommés par commodité intelligences artificielles, sont capables de traiter et valoriser adéquatement cette « matière première » digitale. Ce faisant, se pose de façon aiguë le problème du contrôle de ces traitements des données. A mesure qu'ils prennent une place prépondérante dans les modes de recrutement des collaborateurs, d'évaluation des contributions, de gestion des finances et des productions, d'attribution des financements, de sélection des contenus pertinents, il devient essentiel d'être capable d'en garantir le bon fonctionnement. Or l'intelligence artificielle n'est pas un simple algorithme. Elle ne contente pas d'appliquer un protocole décrit à l'avance. Elle fonctionne beaucoup plus comme un programme capable de faire évoluer lui-même ses propres paramètres, et cela de façon relativement incontrôlée et, pire, inconnue. C'est le cas par exemple des réseaux de neurones, qui font partie des techniques de « machine learning ». Le point d'arrivée de l'IA est à la fois imprévisible et difficilement explicable. Comment reconnaître une responsabilité à un programmeur qui n'avait aucun moyen de savoir le résultat du fonctionnement de son ouvrage ? Comment aussi rendre compte des décisions de la machine, de ses logiques propres, si le script en est tellement complexe que nul être humain ne peut en obtenir une claire vision ? Comment pourrons-nous établir des responsabilités dans un monde où à la complexité invraisemblable s'ajoute l'absence totale de transparence des processus ? Seules des machines pourront comprendre ce que font les autres machines. Des IA évalueront des IA.

Mais alors, selon la question bien connue de Juvénal, « qui gardera les gardiens ? ». Les principes moraux mêmes sur lesquels reposeront ces décisions pourront-ils au moins nous être connus — pour ne rien dire de leur contestabilité ? L’omniprésence des algorithmes de recommandation influence de façon peu visible mais déterminante les comportements et conditionne les relations sociales. Qui décide des critères de recommandation ? Quels effets réels ont-ils sur les choix de consommation, les informations sues ou ignorées, l’évolution des opinions et des groupes sociaux ?

La troisième dimension de la crise de la responsabilité est politique. Quels visages auront la liberté et les prérogatives régaliennes à l’ère d’une IA qui confisque une bonne partie de la maîtrise de l’information ? La manipulation des réseaux sociaux par des puissances étrangères est désormais une menace clairement identifiée lors des élections. Plus subtilement encore, le choix des plateformes de favoriser ou cacher certains contenus, pose la question des critères employés, de leur clarté et du processus par lequel ils sont établis et modifiés. L’État a déjà dans les faits délégué une partie de son pouvoir de censure aux réseaux sociaux qui sont expressément chargés de purger les contenus problématiques, précisément parce qu’ils sont reconnus responsables et ne sont pas, comme ils l’ont longtemps défendu, de simples hébergeurs. Dans quelles conditions s’opère cette censure ? Qui décide du vrai et du faux, de l’excessif et de l’inacceptable ? Le risque d’une censure *a priori* de tout contenu qui ne soit pas strictement dans le canon d’une nouvelle orthodoxie de la bienpensance n’est pas moins effrayant que la multiplication des contenus outranciers et des théories complotistes. Si « le code est la nouvelle loi », on peut s’inquiéter de l’absence de processus démocratique dans la façon dont elle est écrite. A quels tribunaux est-il possible d’avoir recours ? Au floutage des responsabilités correspond un affaiblissement préoccupant des mécanismes de séparation des pouvoirs et des droits des citoyens.

La crise de la responsabilité qui vient pourrait précipiter l’avènement d’ordres totalitaires. Dans une société dominée par le refus de l’imprévu, la haine de l’accident et plus généralement le fantasme de maîtrise, trouver le cerveau sain à qui attribuer la responsabilité est devenu un problème essentiel. Dans bien des cas, tout événement malheureux, même d’origine naturel, doit trouver un responsable. La foule le réclame. Si aucune règle n’a été transgessée, ce qui permettrait d’incriminer l’imprudent transgresseur, c’est que la règle manque. Le responsable par défaut sera ainsi la puissance publique ou l’institution qui aura manqué à son obligation d’encadrement des comportements et des interactions. Ce mécanisme a déjà mené à maintes interdictions. Pour éviter la responsabilité de pratiques incontrôlables, on interdit ou on limite les pratiques. La peur de la responsabilité pourrait nous précipiter dans l’excès inverse : un quadrillage excessivement serré des pratiques. La liberté d’expression connaît une réduction par un processus analogue : le politiquement correct est engendré par le souci de se protéger contre la responsabilité potentielle de propos jugés déviants. Réduire son discours au plus petit dénominateur commun est un moyen d’éviter les affres de la responsabilité. C’est le débat démocratique même qui se trouve ainsi stérilisé par l’effet indirect de la crise de la responsabilité.

Une dernière question vertigineuse se pose : celle de l’élargissement de la reconnaissance de la responsabilité à l’IA elle-même. A l’heure où les anti-spécistes soulignent l’absence de différences fondamentales entre les animaux et les humains les robots ne seront-ils pas les suivants à trouver des défenseurs voulant faire d’eux des sujets de droit ? Si être un « sujet sensible », c’est-à-dire doué de capacités de sentir le monde, est suffisant à faire accéder au statut de sujet de droit, les machines bardées de capteurs ne pourront-elles aisément y prétendre ? Le caractère anthropomorphe des humanoïdes rendra d’autant plus facile notre identification à eux. Le « moment Frankenstein », c’est-à-dire celui où nous parviendrons à créer de toutes pièces des êtres pensants et autonomes, ouvrira nécessairement la voie à la

reconnaissance de droits spécifiques et à la responsabilité des actes de ces artefacts. Si la question ne semble pas se poser dans l'immédiat, l'incertitude semble plus porter sur le délai (des décennies, des siècles ?) que sur sa possibilité.