



## Les risques catastrophiques mondiaux et les institutions de la liberté

par  
*Stephen Davies*



**Stephen Davies** est Docteur en Histoire Médiévale et Moderne. Maître de conférences à la Manchester Metropolitan University jusqu'en 2009, il est à ce jour responsable du département éducation à l'Institute of Economic Affairs et Chercheur Associé au John Locke Institute. Il est l'auteur de plusieurs livres, dont *Empiricism and History* (Palgrave Macmillan, 2003), avec Nigel Ashford, *Dictionary of Conservatism and Libertarian Thought* (Routledge, 1991) et en 2024 *Apocalypse Next : The Economics of Global Catastrophic Risks*, Institute of Economic Affairs.



## Introduction

Il y a eu récemment de nombreuses discussions autour des menaces associées au développement de l'intelligence artificielle ; menaces pouvant aller jusqu'à l'extinction de l'espèce humaine. Ce n'est là qu'un cas parmi d'autres d'un phénomène plus large, celui des risques catastrophiques mondiaux (RCM). Ces derniers n'étaient autrefois la préoccupation que d'une poignée d'universitaires regroupés dans des instituts spécialisés, et des lecteurs de revues telles que le *Bulletin of Atomic Scientists*. Aujourd'hui, cependant, ces risques suscitent un bien plus grand intérêt et constituent une vraie source de préoccupation, pour le monde universitaire et bien au-delà.

Mais qu'entendons-nous au juste par RCM ? Nous devons pour commencer donner une définition plus précise que « quelque chose de très grand et de très mauvais ». Plusieurs chercheurs ont avancé des propositions et notamment le philosophe d'Oxford, Nick Bostrom. L'étape suivante consistera à réfléchir plus sérieusement à la problématique du risque (et de sa mesure probabiliste) et à la façon dont nous devons prendre en compte les coûts potentiels calculés à partir de ces probabilités. Ce n'est qu'une fois que cela aura été fait que nous pourrons aborder la question centrale : que pouvons-nous ou devrions-nous faire à propos de ces RCM ? Pour répondre correctement à cette question sans doute faudra-t-il la contribution des économistes familiarisés avec la théorie des choix et les calculs coûts-avantages.

Plus important encore, les RCM soulèvent des problèmes très graves liés à des questions de gouvernance et à la préservation de nos libertés. Là encore les économistes peuvent apporter une contribution vitale. Et il y a urgence, car les conclusions auxquelles aboutissent le plus souvent les débats sur les RCM annoncent un avenir sombre pour la liberté humaine bien comprise. Cela n'est pas seulement mauvais en soi ; ça anéantit tout réel espoir de trouver une solution au défi que représentent ces RCM pour notre humanité.



## Une définition du Risque Catastrophique Mondial

Bostrom, avec d'autres collègues scientifiques, définit les Risques Catastrophiques Mondiaux comme des événements possibles qui ont des effets sur le monde entier (ils affectent toute la planète et sa population) et sont catastrophiques dans leurs conséquences. Catastrophique implique ici qu'il soit mis fin de façon définitive à l'épanouissement et au développement de l'humanité ; un renoncement définitif à des formes d'existence humaine et de développement que, selon des attentes raisonnables, nous aurions pu connaître. Deux catastrophes possibles qui auraient de tels effets sont l'extinction de l'espèce humaine, ou un effondrement irréversible de la civilisation mondiale. En d'autres termes, un RCM est un événement qui conduit l'histoire humaine, ou du moins l'histoire de la civilisation avancée et développée technologiquement, à une fin ou une stagnation complète et irréversible.

Les RCM doivent donc être distingués de risques appartenant à l'ensemble plus large des « très mauvais événements ». Parfois, vous avez un événement qui est catastrophique par ses conséquences mais qui n'affecte qu'une partie du monde ou de sa population. Il en serait ainsi, par exemple, d'une hypothétique guerre nucléaire qui rendrait une partie du monde à jamais inhabitable et anéantirait sa population. Ces événements sont catastrophiques mais pas mondiaux. De même vous avez des événements qui affectent le monde entier sans pour autant conduire à un résultat catastrophique du style extinction de l'humanité ou fin de la civilisation. Se trouve dans cette catégorie une guerre mondiale conventionnelle ou une dépression économique mondiale ou encore la plupart des pandémies, y compris des cas historiques tels que la peste noire. Cela signifie que les événements hypothétiques auxquels il est possible d'associer un RCM (y compris un risque mettant en péril notre existence) sont en nombre plus limité qu'on ne le croit habituellement.



## L'évaluation des risques

Un élément clé dans la réflexion sur les RCM est celui de la probabilité de réalisation du risque. Fort heureusement, il s'agit toujours d'événements de faible, voire, dans la plupart des cas, de très faible probabilité. Cela ne signifie pas pour autant que nous devons les ignorer et ne pas y penser ; et cela pour deux raisons. Premièrement, bien que la probabilité de la plupart des RCM soit faible, elle n'est pas « extrêmement » faible. Les risques d'un niveau de probabilité trivial *peuvent être ignorés*. Cependant, la plupart des RCM ont des probabilités beaucoup plus élevées que cela, ils ne sont pas triviaux et peuvent aller jusqu'à 1% de chance de réalisation sur une année voire, dans certains cas, plus encore. Pour la plupart des individus cela peut néanmoins paraître insignifiant : si un événement a une chance sur mille ou une chance sur trois cents de se produire au cours d'une année donnée, il n'y a pas de quoi en perdre le sommeil ! Mais ce raisonnement est fallacieux. Lorsque les conséquences de cet événement de faible probabilité sont fatales pour l'espèce humaine ou privent cette dernière de la possibilité d'un futur développement, il convient de prendre au moins un minimum de mesures pour se couvrir contre ce risque ; un peu comme les ménages qui souscrivent une assurance pour la maison, même si les chances que leur maison soit totalement détruite sont très faibles.

La seconde raison de ne pas négliger ces RCM est que, dans certains cas, il est difficile d'estimer leur probabilité, *avec un danger très réel de sous-estimation*. Cette difficulté est due au fait que les événements catastrophiques sont des événements rares, peu fréquents, de telle sorte que leur distribution de probabilité est « à queue large » : il n'est pas facile d'établir une estimation de cette probabilité à partir de l'observation des données du passé, à moins de disposer de données sur le très long terme, et même dans ce cas la fiabilité de l'estimation est faible. Il n'est pas exclu que le passé puisse ne pas contenir l'un de ces grands événements et c'est pourquoi utiliser l'écart type pour estimer la probabilité produira une estimation beaucoup trop faible.

Considérant les conséquences dévastatrices de ces événements, nous devrions plutôt pécher par excès de prudence. Notre première



réponse collective aux prémices de la réalisation d'un RCM devrait être la panique, ou au moins une mise en alerte radicale. Nous aurons toujours la possibilité de nous calmer plus tard si des informations nouvelles indiquent que nous sommes confrontés à un problème gérable et, bien que grave, limité dans ses effets. Si nous attendons d'obtenir plus d'informations avant de faire quoi que ce soit, il sera presque certainement trop tard ; si l'on a affaire à un RCM ce sera la fin de tout.

De plus, outre les caractéristiques statistiques des RCM, il y a autre chose qui rend très difficile l'établissement d'une probabilité précise à partir de données passées : la grande majorité des RCM qui nous concernent aujourd'hui sont d'un type nouveau. Il ne s'agit pas de risques qui ont toujours existé tout au long de l'histoire de l'humanité. Il s'agit plutôt de choses qui ne sont apparues que maintenant ou dans un passé récent, en raison du progrès technologique ou de la manière dont les sociétés modernes se sont développées. L'IA en est un exemple : ce n'est qu'au cours de la dernière décennie que des avancées technologiques ont suscité des craintes réelles et crédibles que l'IA engendre un RCM .

### Les risques nouveaux

Depuis l'apparition du genre humain en Afrique, nous avons toujours vécu sous la menace d'événements susceptibles d'anéantir notre espèce. Ces menaces étaient toutes d'origine naturelle, indépendantes de l'intervention humaine et avec une très faible probabilité d'occurrence. Parmi elles se trouvaient la possibilité d'une collision majeure avec un astéroïde, semblable à la collision qui a anéanti les dinosaures à la fin du Crétacé. Une autre possibilité était une éruption super-volcanique majeure ; une éruption volcanique si grande qu'elle pulvérise plus de mille kilomètre cubes de roches dans l'atmosphère planétaire. Ces menaces sont toujours présentes.

Cependant, nous nous trouvons désormais dans une situation différente. Le nombre et le type de RCM auxquels nous sommes confrontés ont clairement augmenté en raison de l'impact potentiel des



nouvelles technologies et du développement continu des sociétés modernes. Il est donc difficile, dans certains cas, d'estimer leur probabilité, car il n'existe pas de données antérieures pouvant servir de guide. En outre, dans plusieurs de ces nouveaux cas, la probabilité, pour autant que nous puissions l'établir, va dans la mauvaise direction – celle d'une plus grande probabilité. En outre, certains des risques naturels qui ont toujours existé deviennent désormais beaucoup plus probables qu'auparavant en raison de l'activité humaine – les deux principaux cas ici sont le changement climatique brutal et les pandémies. Tout cela se produit pour quatre raisons.

Premièrement, l'activité humaine a augmenté la probabilité de certains risques naturels. L'exemple classique est celui des pandémies. Il y a eu au plus trente véritables pandémies au cours des deux derniers millénaires. Parmi ces événements, seuls quatre ou cinq se sont produits avant 1818 (début de la première pandémie de choléra). Aujourd'hui, les différentes régions du monde sont plus connectées, par le commerce et les migrations, qu'au cours de la majeure partie de l'histoire, et il s'agit d'une réalité permanente contrairement à des « invasions barbares » massives et occasionnelles. Les gens voyagent en plus grand nombre, plus fréquemment, sur de plus longues distances et, surtout, à des vitesses beaucoup plus élevées. Tout cela accroît considérablement le risque de propagation d'une épidémie locale à l'échelle mondiale.

Simultanément, le développement de l'élevage moderne et « industriel » intensif a produit un environnement qui aurait pu être conçu pour produire de nouveaux agents pathogènes et faciliter leur transmission des animaux aux humains. L'utilisation généralisée d'antibiotiques dans ce type d'agriculture ajoute au danger car elle crée le risque d'un pathogène bactérien résistant à la plupart des médicaments. À cela s'ajoute le risque aigu posé par les expériences biologiques sur les agents pathogènes, en particulier les virus, notamment ce que l'on appelle les expériences de « gain de fonction » conçues pour produire des variantes virales plus virulentes et plus transmissibles. Ceci est censé permettre d'étudier les moyens de lutter contre les agents pathogènes naturels émergents, mais les risques sont énormes, étant donné la fréquence des fuites dans les laboratoires et



autres failles de sécurité. La menace ultime réside dans la production d'une arme biologique extrêmement mortelle qui serait ensuite accidentellement libérée.

Deuxièmement, la plus grande connectivité de la société contemporaine fait que les risques sont bien plus susceptibles d'avoir un impact mondial plutôt que local. Dans le passé, il était très rare qu'une catastrophe touche la planète entière ou ait un impact véritablement mondial. À l'époque, la population humaine mondiale était divisée par la distance et le temps en des « mondes » *de facto* différents, et les chances qu'une même chose affecte tous étaient par conséquent faibles. Aujourd'hui, le niveau d'interconnexion humaine est si élevé qu'il est logique de considérer le monde comme une entité unique, même si cela ne se reflète pas dans les institutions ou la conscience humaine. Le risque d'une catastrophe véritablement mondiale est donc bien plus élevé.

Troisièmement, le développement technologique et économique de la société fait que des phénomènes qui n'auraient pas été catastrophiques auparavant ont toutes les chances de l'être aujourd'hui. C'est là un paradoxe du progrès. Le côté brillant du progrès technologique est qu'il apporte d'énormes avantages et accroît considérablement les capacités humaines, y compris notre capacité à faire face aux catastrophes, qu'elles soient naturelles ou provoquées par l'homme. Le côté sombre de tout cela est que les populations humaines et la structure entière des civilisations se retrouvent, du fait de ce progrès, dépendantes d'une infrastructure complexe qui, si elle venait à être endommagée à une échelle suffisamment importante, conduirait à un effondrement mondial. L'exemple classique d'un tel scénario est celui des très grandes éjections de masse coronale – encore connues sous le nom « d'événements de Carrington », en référence à l'extraordinaire tempête solaire de 1859. Une éjection de masse coronale consiste en l'éruption d'une grande quantité de matière hautement énergétique depuis l'atmosphère solaire (la couronne). Ces éjections sont fréquentes et habituellement inoffensives. Cependant, de temps en temps, il y en a une de bien plus grande ampleur. Lorsque cela se produisait, nos ancêtres observaient de spectaculaires aurores boréales. Et quasiment



rien de plus : lorsque l'une d'entre-elles a frappé la Terre en 1859, elle a également perturbé et incendié les systèmes télégraphiques, la seule technologie à l'époque utilisant l'électricité. Si une telle éjection de plasma devait se produire aujourd'hui, les effets en seraient dévastateurs, car elle perturberait les réseaux électriques dans de très vastes régions du monde, détruirait de nombreux appareils, ferait planter Internet pendant un temps considérable et perturberait sérieusement toutes les formes de communication électronique. Notre économie tout entière et notre mode de vie dépendent tellement de ces technologies que les conséquences seraient très graves. Une tempête solaire de la taille de celle de 1859 est un « événement centenaire », donc une probabilité de 1 % chaque année. Des phénomènes encore plus puissants, plusieurs fois plus grands, se produisent en moyenne une fois tous les trois cents ans, soit une probabilité de 0,3 % chaque année. Il s'agit d'un exemple de processus naturel qui n'aurait pas été un RCM dans le passé mais qui pourrait l'être aujourd'hui, car notre développement technologique nous a rendus vulnérables à son impact.

La quatrième raison, et de loin la plus importante, à l'origine d'une augmentation du nombre de RCM possibles est que nous devons désormais faire face aux risques qui résultent d'innovations technologiques. C'est ici que l'incertitude quant au niveau de risque est la plus grande. Plusieurs technologies ont créé de nouveaux RCM ou pourraient très facilement le faire dans un avenir proche. Il y a tout d'abord la technologie des armes nucléaires et le risque associé d'une guerre nucléaire mettant fin à la civilisation. Parmi les autres exemples majeurs de nouvelles technologies susceptibles de générer de manière très plausible un RCM on trouve la modification génétique des organismes, les nanotechnologies et surtout l'IA dite non alignée ou incontrôlable. Cette dernière éventualité occupe une place de choix dans les recherches menées sur les RCM.

### **Une vulnérabilité au RCM temporairement plus élevée**

Notre espèce est non seulement confrontée à un plus grand nombre et à une plus grande variété de RCM, elle traverse aussi une fenêtre



historique de vulnérabilité aiguë à plusieurs de ces risques. Pour le dire autrement, le risque de catastrophe est plus grand qu'il ne l'était dans le passé prémoderne et qu'il ne le sera probablement à l'avenir, en s'appuyant sur des extrapolations raisonnables. Si nous, en tant qu'espèce, parvenons à traverser les cent à cent cinquante prochaines années, le niveau de risque diminuera, que ce soit en termes de nombre, d'étendue ou de probabilité. Le défi consiste donc à traverser le prochain siècle sans dommage.

Pourquoi sommes-nous désormais si vulnérables en tant qu'espèce ? Ce n'est pas seulement dû aux quatre facteurs énumérés ci-dessus. Il faut aussi tenir compte de trois facteurs aggravants supplémentaires.

Premièrement, le niveau actuel de développement technologique dans plusieurs domaines est suffisamment avancé pour susciter des risques, mais pas encore suffisamment pour nous permettre d'y faire face. Cette situation est aggravée par le fait que l'évolution institutionnelle n'a pas suivi le rythme des découvertes technologiques et scientifiques. Deuxièmement, une combinaison de choix délibérés et de processus émergents au cours du siècle dernier a créé une civilisation organisée autour de systèmes hiérarchiques complexes. Ceux-ci sont fragiles plutôt que résilients (ou anti-fragiles) de telle sorte que des événements qui auraient été mauvais mais gérables par le passé deviennent catastrophiques car ils peuvent déclencher un effondrement systémique catabolique. Nous sommes à un niveau technologique où le seuil d'effondrement irrémédiable est bien plus bas que par le passé.

Mais c'est le troisième facteur d'aggravation des risques qui est le plus important. La plupart des travaux sur les RCM se concentrent sur les risques d'extinction de l'espèce humaine, plutôt que sur la deuxième catégorie de RCM, à savoir, l'effondrement irrémédiable de la civilisation. La raison de ce choix se trouve dans l'histoire : les civilisations se sont effondrées à plusieurs reprises, mais une nouvelle civilisation a toujours fini par émerger. Si tel était le cas, l'effondrement de la civilisation moderne serait un désastre majeur, mais il ne serait



pas catastrophique pour autant, car l'effondrement ne serait pas permanent. Malheureusement une telle extrapolation est trompeuse. Notre civilisation moderne se distingue par sa haute technologie et son importante consommation d'énergie. Or, notre civilisation a consommé la plupart des minerais, du pétrole, du charbon et d'autres ressources qui étaient faciles d'accès et de haute qualité, en particulier comme source d'énergie. Tant que nous avons accès à de grandes quantités d'énergie, cela ne pose aucun problème. Mais si la civilisation devait s'effondrer, ce serait parce que l'accès à des ressources nouvelles nécessiterait une technologie et des ressources énergétiques que nous n'aurions plus. Il serait dès lors possible de voir émerger une civilisation agraire peu consommatrice d'énergie. Mais il serait très improbable, voire impossible, qu'une civilisation à haute énergie comme la nôtre refasse surface, car les ressources dont elle aurait besoin pour démarrer ne seraient tout simplement plus là.

### **Que faire ? Il faut plus que jamais préférer la liberté**

Tout cela signifie que l'humanité est confrontée à des types de RCM plus nombreux et plus variés et que l'impact de l'un quelconque de ces risques, s'il se concrétisait aujourd'hui, serait plus grand qu'il ne l'aurait été dans une société avec un rythme de développement technologique plus faible ou dans une société avec un rythme de développement technologique plus élevé. Une grande partie des commentaires à ce sujet se sont concentrés sur le rôle de l'innovation. Dans un article paru il y a quelques années et largement lu, Nick Bostrom présentait un argument à la fois convaincant et alarmant – surtout si vous êtes un libéral. Il compare le processus d'innovation technologique au tirage de boules à partir d'une urne opaque. Jusqu'à présent, la plupart des boules qui ont été tirées étaient blanches (bénéfiques), avec parfois des boules grises discutables (résultats mitigés). Cependant, nous savons avec quasi-certitude que dans l'urne se trouvent des innovations qui sont des boules noires – elles entraîneraient l'extinction humaine ou quelque chose de tout aussi catastrophique. Le problème est que nous ne savons pas quelles boules (innovations) sont « noires » et nous ne pouvons pas



dire combien il y en a dans l'urne. Cela signifie que nous n'avons aucun moyen facile d'établir la probabilité qu'une innovation fatale se produise dans un horizon temporel donné. Nous pouvons fortement soupçonner que certaines innovations dont nous avons déjà connaissance entrent très probablement dans cette catégorie – l'IA est la candidate évidente – mais nous ne pouvons pas en être certains. Si nous nous trompons, le résultat est fatal pour nous tous.

Le diagnostic posé par Bostrom repose sur une argumentation puissante. En revanche la solution qu'il préconise pose un problème et devrait alarmer tout économiste ou tout libéral qui se respecte et *a fortiori* tout économiste libéral. Bostrom raisonne de la façon suivante : étant donné les risques d'une innovation incontrôlée – en particulier si elle est « démocratique », c'est-à-dire, facile à réaliser et à mettre en œuvre avec des ressources limitées – nous devrions imposer des contrôles stricts sur l'innovation, sur la recherche et sur l'investigation et pour ce faire, mettre en place une sorte de système de licences. Il faudrait établir une autorité mondiale dotée de pouvoirs quasi totalitaires de supervision, de régulation et de contrôle sur de nombreux domaines de la vie intellectuelle. En raison des enjeux, ce pouvoir ne pourrait pas être démocratique ni soumis à un contrôle juridique normal – il fonctionnerait sur des bases similaires à celles déjà établies pour les organismes qui, au niveau national, protègent contre des risques d'extinction de l'espèce.

Au-delà des aspects pratiques, cette proposition est tout simplement effrayante. Elle soulève également des interrogations d'un point de vue théorique : sommes-nous certains qu'un contrôle de ce type sur l'innovation constitue le meilleur moyen de relever le défi des RCM ? Aussi puissants que soient les arguments de Bostrom, la réponse de l'économiste est : probablement non.

De façon ironique, la mise en place de cette solution est précisément l'un des RCM identifiées par Toby Ord, collègue de Bostrom à Oxford, à savoir, l'émergence d'une tyrannie mondiale permanente qui stoppe tout progrès et tout développement intellectuel. Par ailleurs, la théorie des choix publics suggère fortement que le processus et les institutions



mis en place seraient inefficaces et susceptibles d'être capturés par des intérêts particuliers en quête de rente. Cela pourrait en réalité aggraver le problème que l'on tentait de résoudre. Troisièmement, et c'est le plus important, une telle autorité serait confrontée à *un problème de connaissances paralysant*, problème que Bostrom identifie correctement dans son diagnostic mais ignore lorsqu'il en vient à formuler une solution. Le problème de la connaissance fait que, pour relever le défi des RCM, nous avons plus que jamais besoin d'une expérimentation décentralisée. Comme le montrent de nombreuses études, l'innovation a toujours été le fruit d'expérimentations décentralisées, de bricolages, d'émulation et de concurrence. C'est cette piste qu'il est urgent de suivre.

Certes, toutes les expérimentations ne doivent pas être autorisées. Pour certaines une interdiction est logique car les risques dépassent largement les avantages possibles. Les expériences de « gain de fonction » sur les virus constituent un cas classique, mais il existe également de solides arguments en faveur d'une pause de courte durée de la recherche sur l'IA afin de permettre aux chercheurs de faire le point et de permettre au système d'innovation décentralisé de trouver des solutions aux éventuels problèmes. Cela étant dit, nous devons aussi tenter d'évaluer les coûts d'opportunité que de tels freins à l'innovation impliquent. A moins que les risques soient si grands que nous devrions tout simplement arrêter presque toutes les recherches ? Cela constituerait en soi une sorte de RCM et les coûts en termes de bien-être futur potentiel seraient énormes.

Ce que nous devons essayer et faire, c'est prendre en compte les deux probabilités et trouver un équilibre : la probabilité de créer un risque en brimant l'innovation et la probabilité de créer un risque en laissant l'innovation sans aucun contrôle. Étant donné qu'il s'agit d'un problème de connaissances, les marchés et autres processus sociaux non-coercitifs (y compris les marchés contingents) constituent le meilleur moyen de parvenir à un bon équilibre. Ce dont nous avons besoin, c'est de beaucoup d'expériences et de nombreux échecs, car cela demeure le plus sûr chemin vers l'innovation, l'accroissement des connaissances et le recul de l'incertitude.



Nous devons donc changer radicalement la façon dont fonctionnent les institutions et les organisations de toutes sortes. À l'heure actuelle, trop de systèmes essentiels à nos économies et à nos sociétés sont excessivement centralisés, rigides et gérés d'en haut selon des règles explicites et rigides. Nous devons au contraire aller dans le sens de réseaux distribués décentralisés et synonymes d'anti-fragilité. Ostrom, Hayek et Taleb sont les personnes sur lesquelles nous devrions nous appuyer comme guides ici. Il existe également un besoin urgent d'évoluer vers une vision à plus long terme dans des domaines tels que l'investissement, le commerce, la recherche et l'innovation, et les modes de gouvernement. C'est la suggestion de Toby Ord.

En d'autres termes, nous avons besoin d'une sorte de libéralisme radical différent du néolibéralisme d'entreprise prôné par des gens comme le *World Economic Forum* (qui est plus proche des arguments avancés par Bostrom et d'autres et reflète une vision d'avenir fait de recherches de rente) et du genre d'activisme gouvernemental ou de « gouvernance de projet » incarné par exemple par Mariana Mazzucato<sup>1</sup>. L'humanité est actuellement confrontée à de nombreux défis et la menace de nombreux RCM, pour la plupart nouveaux, est peut-être la plus pressante, mais nous devons être convaincus que c'est la liberté et des institutions et processus décentralisés et en réseau qui offrent le meilleur moyen de surmonter cela.

---

<sup>1</sup> Cf. Alberto Mingardi, « Une critique de l'État entrepreneur de Mazzucato », *Journal des libertés*, n°12, printemps 2021 à <https://bit.ly/3tu0YMK>