1. Les intelligences artificielles sont un progrès, qui ne va pas sans risques. Deux journalistes du MIT Technology Review,[Will Knight](https://twitter.com/willknight?lang=fr) et [Karen Hao](https://www.technologyreview.com/profile/karen-hao/) listent dans un long article les principaux écueils qu'il faudra, dans les prochains mois et les prochaines années, cherché à éviter, à savoir : Les voitures autonomes, La manipulation politique, Les algorithmes tueurs, La reconnaissance faciale, L'ère du deep fake et La discrimination algorithmique.
2. En mars 2018, [l’accident mortel](https://techcrunch.com/2018/05/24/uber-in-fatal-crash-detected-pedestrian-but-had-emergency-braking-disabled/) causé par une voiture autonome Uber a révélé les failles d’une technologie encore loin d’être au point. Alors que des constructeurs automobiles traditionnels, tel Ford, et [de nombreuses start-ups partent à l’assaut de ce nouveau secteur](https://korii.slate.fr/tech/2-millions-miles-vehicules-autonomes-silicon-valley-waymo-gmcruise), une législation digne de ce nom se fait encore attendre.
3. Le [scandale Cambridge Analytica](https://www.lemonde.fr/pixels/article/2018/03/22/ce-qu-il-faut-savoir-sur-cambridge-analytica-la-societe-au-c-ur-du-scandale-facebook_5274804_4408996.html) a mis au jour comment l’utilisation de nos données numériques pouvait servir à orienter la communication des politiques afin de nous influencer. On apprend encore que, en filtrant les informations qui nous parviennent sur les réseaux sociaux, les algorithmes encouragent aussi la désinformation. Facebook, par la voix de son PDG Mark Zuckerberg, a déclaré que l’intelligence artificielle servirait à l’avenir à repérer les contenus malveillants et espère le prouver dès cette année lors de l’élection présidentielle [au Kenya](https://qz.com/africa/1411947/facebook-starts-africa-fact-checking-tool-with-afp-africa-check/) et [d](http://www.businessinsider.co.za/facebook-fact-checking-in-south-africa-with-africa-check-2018-10)[es élections générales sud-africaines](https://www.businessinsider.co.za/facebook-fact-checking-in-south-africa-with-africa-check-2018-10).
4. [Le projet Maven](https://www.theverge.com/2018/6/1/17418406/google-maven-drone-imagery-ai-contract-expire), destiné à fournir au Pentagone une intelligence artificielle de pointe pour ses drones-tueurs, a finalement été abandonné en 2018 par Google. L’utilisation de l’IA à des fins militaires est d’ores et déjà au programme de Microsoft ou Amazon. Mais [les armes autonomes](https://korii.slate.fr/tech/silicon-valley-pentagone-intelligence-artificielle) devraient également s’inviter à la table des négociations de l’Onu en 2019 et conduire à leur éventuelle interdiction.
5. La reconnaissance faciale automatisée conduit à la surveillance généralisée des pouvoirs politiques. Les incroyables capacités de l’intelligence artificielle à reconnaître des visages ont conduit plusieurs pays, [la Chine en tête](https://korii.slate.fr/tech/chine-laboratoire-dictature-20), à déployer une surveillance technologique de masse. Utilisée pour taguer nos photos ou débloquer nos téléphones, la reconnaissance faciale sera bientôt étendue à nos webcams et à nos véhicules.
6. La [prolifération de vidéos *deep fake*](https://korii.slate.fr/tech/deep-fakes-cauchemar-fact-checkers-video-visage-ia-bot)l’année dernière a montré combien il était désormais facile de retoucher une séquence de façon quasi-parfaite, en procédant à une permutation intelligente de visages. Un potentiel cauchemar pour les pourfendeurs de *fake news* et un nouvel outil pour ceux qui trouvent un intérêt à les faire proliférer.
7. L'année 2018 a prouvé que des algorithmes entraînés à partir de données inégalitaires [reproduisent les stéréotypes et les discriminations de nos sociétés](https://www.wired.com/story/ideas-joi-ito-insurance-algorithms/). Un manque de diversité qui se retrouve dans le secteur de l’IA où les femmes, par exemple, sont sous-représentées. La [Conférence internationale sur l’intelligence artificielle et l’éducation](https://fr.unesco.org/events/conference-internationale-lintelligence-artificielle-leducation), une référence en matière d’IA, doit se tenir en Éthiopie en 2020 et traiter ce problème. Nous espérons que les choses commenceront à bouger d’ici là...
8. Comme c'est généralement le cas avec les développements technologiques, les avancées des Intelligences artificielles en matière de génération de texte pourraient entraîner des méfaits potentiels. Dans un monde où la guerre de l'information est de plus en plus répandue et où les nations déploient des bots sur les médias sociaux pour tenter d'influencer les élections et de semer la discorde, l'idée de programmes d'IA qui débitent des absurdités incessantes mais convaincantes est troublante
9. OpenAI fait preuve de prudence en dévoilant son dernier programme GPT-2. Contrairement à la plupart des étapes importantes de la recherche en IA, le laboratoire ne partagera pas l'ensemble des données utilisées pour l'entraînement de l'algorithme ni tout le code sur lequel il fonctionne (bien qu'il ait donné un accès temporaire à l'algorithme à un certain nombre de publications médiatiques, dont The Verge
10. Si GPT-2 est capable de traduire un texte sans avoir été explicitement programmé pour le faire, cela pose une question évidente : qu'est-ce que le modèle a appris d'autre que nous ne savons pas ? Les chercheurs d'OpenAI admettent qu'ils ne sont pas en mesure de répondre complètement à cette question. Ils sont toujours en train d'explorer ce que l'algorithme peut et ne peut pas faire. C'est pour cette raison, entre autres, qu'ils font preuve de prudence dans ce qu'ils partagent sur le projet, gardant pour l'instant pour eux le code sous-jacent et les données d'entraînement. Une autre raison de cette prudence est qu'ils savent que si quelqu'un alimente GPT-2 avec des textes racistes, violents, misogynes ou abusifs, il continuera dans cette veine. Après tout, il a été formé sur Internet.
11. Jack Clark, directeur des politiques chez OpenAI, affirme que ces préoccupations ne peuvent être ignorées. OpenAI, dit-il, veut encourager les universitaires et le public à avoir une conversation sur les méfaits de cette technologie avant qu'elle ne devienne largement disponible. "Ce que je vois, c'est que quelqu'un finira par utiliser une vidéo, une image, un son ou un texte synthétique pour briser un état d'information" explique Jack Clark, "Ils vont empoisonner le discours sur Internet en le remplissant de non-sens cohérents. Ils vont faire en sorte qu'il y ait suffisamment d'informations bizarres qui l'emportent sur les bonnes informations, au point de nuire à la capacité des vraies personnes à avoir de vraies conversations."
12. Un rapport publié en 2018 par OpenAI et des groupes universitaires de Cambridge et d'ailleurs, intitulé "L'utilisation malveillante de l'intelligence artificielle", a prédit l'arrivée d'une telle technologie, et il suggère d'autres utilisations nuisibles. La génération automatisée de textes pourrait faciliter les escroqueries en ligne, par exemple, et améliorer les capacités des pirates à faire du spear-phishing sur leurs cibles (c'est-à-dire les inciter à donner leurs identifiants en ligne en se faisant passer pour un ami ou une institution de confiance).
13. L'hypothèse d'OpenAI est qu'il est préférable de parler des dangers de l'IA "avant qu'ils n'arrivent".Selon Jack Clark, directeur des politiques chez OpenAI, les algorithmes de modélisation du langage comme GPT-2 ne sont pas aussi matures que les deepfakes, mais ils en sont suffisamment proches pour justifier une approche prudente. "Notre hypothèse est que le monde pourrait être meilleur et plus sûr si vous parlez de ces dangers avant qu'ils n'arrivent", dit-il. Howard, cofondateur de Fast.AI, est d'accord. "Cela fait un moment que j'essaie de mettre en garde les gens à ce sujet", dit-il. "Nous disposons de la technologie nécessaire pour remplir totalement Twitter, les e-mails et le web d'une prose raisonnable et adaptée au contexte, qui noierait tout autre discours et serait impossible à filtrer."
15. Texte ici
16. Texte ici
17. Texte ici
18. Texte ici