> menteurs, avec la prolifération des *fake news* ou la demande d'une partie de la population d'une meilleure maîtrise des frontières.

C'est dans ce contexte qu'est né le programme de recherche européen *iBorderCtrl*, mené par un consortium impliquant notamment la Manchester Metropolitan University, en Angleterre. Il vise à développer un outil pour effectuer des vérifications renforcées aux frontières, sans ralentir le flux des passagers dans les aéroports. Sa singularité par rapport à d'autres systèmes de contrôle existants réside dans la création d'un module de détection automatique du mensonge (*Automatic deception detection system*, ou ADDS).

INTERROGÉ PAR UN AVATAR

Imaginez maintenant la scène suivante. Vous vous apprêtez à partir en vacances, en direction d'une somptueuse plage d'une île grecque. Vous allez donc prendre l'avion et effectuer quelques contrôles préalables, comme vous en avez l'habitude. Sauf que, cette fois-ci, vous devez vous préenregistrer sur une application web. Un avatar vous pose alors toute une série de questions sur votre identité et votre voyage, comme : « Quel est votre nom de famille ? » ou : « Que contient votre valise ? ». Votre webcam filme l'entretien, tandis qu'une intelligence artificielle analyse votre comportement non verbal. Pour chacune de vos réponses, elle livre son verdict : mensonge, vérité ou « indéterminé ». Le lendemain, en se fondant sur cette évaluation et sur les éléments complémentaires des contrôles aéroportuaires (comme la vérification du passeport), elle vous attribue un « niveau de risque » global. Si ce niveau est faible, vous passez sans encombre ; dans le cas contraire, vous êtes soumis à des contrôles supplémentaires.

Rassurez-vous, ce module n'en est qu'au stade expérimental et n'est pas réellement déployé dans les aéroports. Mais l'annonce de son élaboration a entraîné une controverse scientifique et un déchaînement de critiques – même si, selon ses concepteurs, l'objectif est d'aider la décision humaine et non de la supplanter. Dans un article paru dans le journal *The Guardian* fin 2018, Bennett Kleinberg, de l'University College de Londres, a ainsi qualifié ce système de « pseudoscientifique », tandis que le site internet iborderctrl.no milite contre son utilisation. Pourquoi une telle levée de boucliers ?

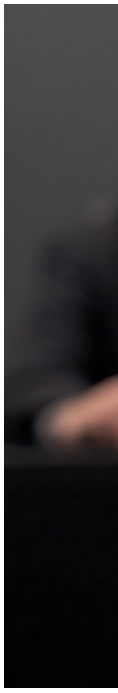
Rappelons déjà que nous sommes de piètres détecteurs de mensonges, puisque nous ne faisons guère mieux que le hasard. En effet, en nous fondant uniquement sur ce que nous captons par nos yeux et nos oreilles, nous débusquons la tromperie à hauteur de 54 % en moyenne. C'est ce qu'ont montré plusieurs expériences où des scientifiques présentaient des vidéos aux participants et leur demandaient si la personne filmée disait ou non la vérité.

LE NEZ DE PINOCCHIO N'EXISTE PAS

Plusieurs facteurs expliquent cette faible performance. D'abord, aucun indicateur ne permet à lui seul de repérer un mensonge de manière certaine : le nez de Pinocchio n'existe pas. Ensuite, de nombreuses idées reçues entravent notre arbitrage. La plus prégnante d'entre elles consiste à considérer qu'une personne qui ment est nécessairement nerveuse. En réalité, ce n'est pas toujours le cas ; et à l'inverse, une personne sincère est parfois stressée pour bien d'autres raisons, notamment la crainte de ne pas être crue. Mais la conséquence de cette idée reçue est que nous associons à la tromperie les diverses manifestations de nervosité : regard fuyant, agitation excessive, tics faciaux, clignements nerveux des yeux… Autant de comportements qui ne sont pas spécifiques du mensonge. Dans une expérience, des chercheurs ont par exemple compté les clignements d'yeux d'un menteur et de quelqu'un qui dit la vérité, et observé… qu'il n'y a aucune différence.

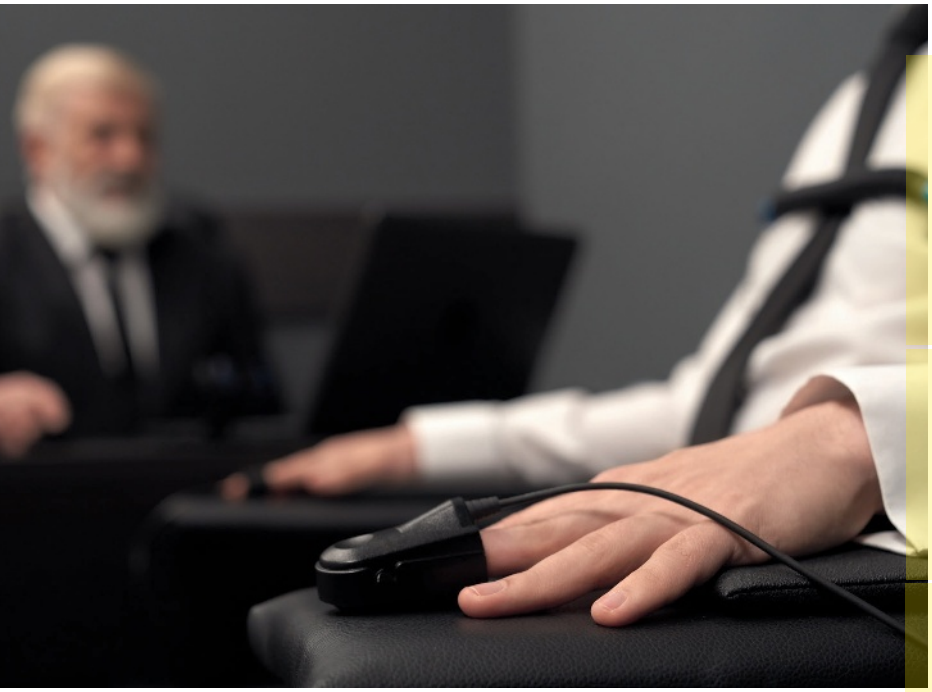
Un espoir d'améliorer les choses a un temps résidé dans les microexpressions, des expressions faciales qui passent sur le visage en moins de 500 millisecondes. Selon le psychologue américain Paul Ekman, elles trahissent les émotions que l'on souhaite réprimer ou masquer, ce qu'il qualifie d'*emotional leakage*, ou « fuite émotionnelle ». Un indicateur potentiellement précieux pour détecter les menteurs, dont les microexpressions ne seraient pas en accord avec ce qu'ils racontent (le coupable d'un vol aura par exemple une microexpression de joie lorsqu'il est en train de mentir, laissant transparaître le plaisir de duper son interlocuteur). Vous en avez peut-être déjà entendu parler, puisqu'elles ont été

Les très discrets indices comportementaux ne constituent en aucun cas une preuve absolue



La détection automatique de mensonges a des antécédents avec les polygraphes, appareils qui mesurent certains paramètres physiologiques de la personne interrogée, tels que son rythme cardiaque, sa pression sanguine, la conductance de sa peau… Ces dispositifs sont utilisés dans bon nombre de pays, mais leur fiabilité est controversée et leurs indications ne sont généralement pas prises en compte par les tribunaux.

© Serhii Bobyk / shutterstock.com



popularisées par la série télévisée *Lie to Me*. Son personnage principal, Cal Lightman, repère les mensonges au moindre mouvement facial qui lui paraît suspect.

La réalité scientifique est toutefois différente. Les recherches indiquent que les micro-expressions apparaissent très rarement sur les visages, de sorte que leur utilité est limitée. De plus, elles se manifestent indépendamment du fait que l'on cherche ou non à dissimuler ses émotions. Enfin, être formé à leur détection n'améliore pas la capacité à détecter les mensonges, selon une étude récente menée par Sarah Jordan et ses collègues de l'université de Huddersfield, au Royaume-Uni.

54 %

C'est le taux moyen de bonnes réponses chez les humains qui tentent de détecter un mensonge chez autrui, soit à peine mieux que le hasard. Ce taux dépasse souvent les 70 % pour les systèmes automatiques, mais cela reste très insuffisant.

L'INTELLIGENCE ARTIFICIELLE À LA RESCOUSSE

Il existe tout de même quelques différences de comportements entre le mensonge et la vérité (une étude ayant par exemple montré que les menteurs ont tendance à davantage pincer les lèvres entre deux mots ou en début et en fin de phrase), mais elles posent deux types de problèmes. D'une part, elles constituent tout au plus des indices de mensonge, et en aucun cas une preuve absolue. D'autre part, ces différences sont si discrètes qu'il est en pratique impossible de les distinguer sans une analyse approfondie. C'est face à ce constat des piètres performances humaines qu'est né le projet de recourir à l'intelligence artificielle: un détecteur informatisé serait capable de capter d'infimes contractions musculaires et d'agréger plusieurs dizaines d'indicateurs. Froncements de sourcils, mouvements de la bouche, direction du regard... Les possibilités sont multiples.

Pour développer ce type d'outils, les scientifiques indiquent alors au système les paramètres qu'il doit analyser et lui soumettent des enregistrements de menteurs ou de personnes qui disent la vérité. Ces données alimentent un modèle statistique qui, après une phase d'apprentissage, se prononce sur la sincérité des déclarations qu'on lui soumet.

Dans le cas d'*iBorderCtrl*, le module de détection du mensonge se fonde sur des « microgestes », d'après ses concepteurs. Ceux-ci n'ont pas communiqué la nature exacte de ces indicateurs, qu'ils sont les seuls à évoquer dans la communauté scientifique. Nous savons juste qu'il s'agit d'infimes contractions musculaires, susceptibles de durer plus ou moins longtemps, mais nous n'avons aucune donnée sur la pertinence de leur utilisation dans ce contexte.

Une étude publiée en 2016 par Lin Su et Martin Levine, de l'université McGill, au Canada, pourrait toutefois nous donner quelques éléments de réponse. Ces chercheurs ont développé un outil de détection des mensonges qui semble se rapprocher de celui d'*iBorderCtrl*, car il se fonde également sur une série d'indicateurs faciaux (comme certains mouvements des sourcils) extraits de manière automatique. Lin Su et Martin Levine l'ont testé avec des vidéos de vrais et de faux appels à témoins diffusées par les médias. Dans les deux cas, les personnes filmées disaient rechercher un de leurs proches disparu, mais dans les vrais appels à témoin, elles étaient sincères alors que dans les faux, elles avaient en réalité tué le prétendu disparu (une vidéo montrait ainsi Susan Smith, condamnée à perpétuité pour avoir assassiné deux de ses enfants). Résultat: l'intelligence artificielle a correctement identifié les menteurs dans 77% des cas.

Les concepteurs d'*iBorderCtrl* ont communiqué un taux de détection supérieur à 70% (*voir la figure page 68*), ce qui serait cohérent avec cette étude. Toutefois, pour atteindre des taux de fiabilité qui justifieraient leur emploi dans le cadre du contrôle douanier ou de la lutte antiterroriste, il faudrait que ces systèmes fassent bien mieux que cela. Peut-être en incluant une analyse du langage?

En effet, après plus de cinquante ans de recherches sur le sujet, les résultats montrent globalement une supériorité des indicateurs >

> linguistiques pour évaluer la crédibilité, par rapport aux indicateurs paraverbaux (le ton de la voix, la longueur des pauses, etc.) et aux comportements non verbaux (les gestes, les postures, etc.) : grâce à « une métaanalyse » portant sur ces recherches, Bella DePaulo, de l'université de Virginie, a par exemple montré en 2003 que les récits de menteurs sont moins détaillés que les versions d'individus sincères. En d'autres termes, si l'on cherche à tirer une histoire au clair, mieux vaut bien écouter et se concentrer sur ce qui est dit plutôt que de chercher à observer les indices corporels ou vocaux. Là encore, *iBorderCtrl* gagnerait probablement à creuser dans cette direction, en incluant des indicateurs linguistiques.

Un autre problème est que ce système a été entraîné et préalablement testé sur un nombre assez restreint de personnes, environ une trentaine. Cela peut se traduire par une limitation bien connue en intelligence artificielle, nommée l'« *overfitting* ». C'est-à-dire que le modèle risque de ne pas être généralisable à d'autres individus, notamment ceux qui parleraient une autre langue ou s'exprimeraient différemment.

D'après le site officiel d'*iBorderCtrl*, l'outil a tout de même été éprouvé dans des conditions quasi réelles en 2019, sur la base du volontariat, aux frontières de la Hongrie, de la Grèce et de la Lituanie. Ses développeurs n'ont communiqué aucune donnée sur les résultats,

ACCUSÉS À TORT DE MENTIR

On évalue l'efficacité d'un outil de détection automatique des mensonges avec deux taux : la proportion des menteurs qui sont identifiés comme tels et la proportion de personnes disant la vérité qui sont effectivement reconnues comme sincères. Pour *iBorderCtrl*, ces taux seraient respectivement de 74 % et 76 %, selon des tests préliminaires. Comme on le voit sur la figure, pour 1 000 voyageurs qui répondraient à une question, 3 menteurs passeraient à travers les mailles des filets (en supposant qu'il y a 1 % de menteurs), mais 242 personnes sincères seraient accusées à tort. Ce nombre élevé s'explique par le fait que dans une application comme le contrôle des frontières, la grande majorité des gens (99 % dans le modèle proposé ici) disent la vérité, de sorte que même un taux de 76 % conduit à de nombreux « faux positifs » – ce taux signifie que lorsque vous êtes sincère, vous avez environ une chance sur quatre d'être accusé de mentir. Alors, imaginez un instant ce que cela donnerait avec les centaines de millions de personnes qui franchissent chaque année les frontières de l'Europe, et qui devraient répondre à de nombreuses questions !