



**« La France va remporter
l'Euro 2021 »»**

Une équipe européenne de chercheurs en intelligence artificielle a fait tourner son modèle 100 000 fois avant le début de l'Euro. Verdict : la France allait gagner la compétition. Raté ! Pire, le vainqueur, l'Italie, ne figurait même pas parmi les trois favoris de leur simulation. Comment expliquer que les ordinateurs soient de si piètres pronostiqueurs ?

Texte Thomas Lestavel

« Anéantis ». Ce simple mot, à la une du quotidien *L'Équipe* mardi 29 juin, donnait la mesure de la déception suscitée par la défaite, la veille, de l'équipe de France face à la Suisse. Les Bleus ont eu beau mener de deux buts à dix minutes de la fin du temps réglementaire, ils ont laissé leurs adversaires égaliser en toute fin de match, avant d'être éliminés à l'issue de la séance des tirs au but.

Cette contre-performance dès les huitièmes de finale de l'Euro a déjoué tous les pronostics... même ceux établis par des algorithmes. Le 7 juin 2021, quatre jours avant le coup d'envoi de la compétition, une équipe composée de sept chercheurs de quatre nationalités (allemande, belge, norvégienne et autrichienne) avait en effet attribué à la France la probabilité la plus élevée de remporter le trophée. D'après leurs calculs, la France avait 14,8 % de chances de gagner le tournoi, devant l'Angleterre (13,5 %) et l'Espagne (12,3 %). Mais c'est finalement l'Italie qui a dominé les débats.

Moins fort que Paul le poulpe

Déjà en 2016, lors du précédent Euro disputé en France, plusieurs spécialistes en intelligence artificielle avaient prédit, à tort, une victoire de l'équipe de France en finale. Les Bleus seraient-ils fâchés avec les statistiques ? Possible, mais les pro-

grammes informatiques ont surtout d'immenses progrès à accomplir avant de se poser en oracles crédibles, et pourquoi pas espérer rivaliser un jour avec le regretté Paul le poulpe. Héros de la Coupe du monde 2010 en Afrique du Sud, le célèbre céphalopode s'était contenté à l'époque de sa ration quotidienne de mollusques pour pronostiquer correctement les résultats de huit matchs de suite, jusqu'à la finale. Une prouesse qu'il avait pourtant une chance sur 256 de réaliser... La méthode était simple : le poulpe devin avait le choix entre deux boîtes placées au fond de son aquarium, chacune contenant une moule et arborant le drapeau d'une des deux équipes sur le point de s'affronter.

Pour le collectif de chercheurs européens qui s'est attaqué aux pronostics du dernier Euro, l'exercice de prévision s'est révélé beaucoup plus laborieux. Ils ont fait tourner à 100 000 reprises leur simulation de la compétition pour aboutir à la conclusion suivante : si l'Euro devait avoir lieu 100 fois, les Bleus gagneraient 15 fois. « *Ce qui nous laissait quand même avec un risque de 85 % de nous planter* », résume avec lucidité Gunther Schauburger, l'un des sept chercheurs en question, qui enseigne les statistiques à l'université de Munich. À noter que leur modèle donnait la Belgique vainqueur de la compétition dans seulement 8 % des cas, malgré la première place occupée par les Diables rouges



au classement Fifa. Et de fait, la Belgique a été éliminée par l'Italie en quarts de finale.

Pour établir leurs pronostics, les sept chercheurs ont fait appel aux dernières avancées en matière d'apprentissage automatique (*machine learning*). « Nous avons, pour chaque match, un nombre prévisionnel de buts pour chacun des deux adversaires. Si on prend par exemple le choc Allemagne-Angleterre, la moyenne attendue s'élevait à 1,2 but pour la Mannschaft et à 1,4 but pour les Anglais », détaille Gunther Schaubberger. « À chaque nombre de buts est associée une probabilité, par exemple 30 % d'en marquer un ou 25 % d'en inscrire deux. Nous avons distribué de manière aléatoire le score de chaque équipe et simulé la confrontation 100 000 fois pour dégager une tendance. » Les probabilités utilisées reposaient sur quatre paramètres : les scores passés (en donnant plus de poids aux matchs récents) ; les cotes de dix-neuf bookmakers internationaux, qui reflètent les prévisions des parieurs ; la performance de chaque joueur en club et en équipe nationale, c'est-à-dire le nombre de buts marqués ou encaissés pendant le temps où le

joueur se trouvait sur le terrain ; et enfin d'autres éléments plus hétéroclites comme l'âge moyen des joueurs, leur valeur sur le marché des transferts, leur expérience des phases finales de compétition ou encore le classement Fifa de l'équipe.

52 matchs « seulement »

Malgré la relative complexité du modèle, son efficacité était de toute façon limitée. Et pour cause : de nombreux aléas peuvent interférer lors d'un match de foot. « Pour un match qui peut être gagné, perdu ou finir sur un score nul, la probabilité du scénario le plus plausible dépasse rarement 50 % », précise Gunther Schaubberger. Son équipe a beau avoir compilé l'historique des résultats lors des Euros et des Coupes du monde disputés depuis 2004, un tournoi ne compte qu'une cinquantaine ou une soixantaine de matchs par édition (52 matchs pour l'Euro 2021). Un panel de données très réduit quand on le compare à d'autres domaines comme la météorologie ou la recherche contre le cancer, où l'intelligence artificielle brasse des millions

Si l'Euro devait avoir lieu 100 fois, les Bleus gagneraient 15 fois. *« Ce qui nous laissait quand même avec un risque de 85 % de nous planter. »*

d'informations pour comprendre le passé et prévoir l'avenir.

Une difficulté supplémentaire du football tient au nombre limité de buts par confrontation par rapport à d'autres disciplines. Au basket, par exemple, le nombre de paniers dans un match est plus élevé, et le match nul beaucoup plus rare. Le foot se distingue également de disciplines comme le base-ball, *« tellement répétitif et linéaire qu'on peut le modéliser plus facilement »*, décrypte Julien Maldonato, associé au cabinet de conseil Deloitte. Un footballeur peut marquer du pied, de la tête, du genou, voire du dos ou de la fesse. Les espaces sont larges, et le ballon peut tout aussi bien jaillir du centre que des ailes. Rien à voir avec des disciplines comme la course à pied, le lancer de poids ou la course hippique, pour lesquelles les paramètres permettant d'expliquer la performance sont plus simples à déterminer. Le nombre de combinaisons sur la pelouse est d'ailleurs bien plus important sur un terrain de foot que dans un sport comme le volley-ball : *« Si le ballon rond a autant de succès auprès du public, c'est aussi du fait de son côté chaotique, qui rend les sur-*

prises toujours possibles. On peut dominer un match et le perdre. Une erreur de la défense peut tout faire basculer », rappelle Julien Maldonato. Pour ce spécialiste en intelligence artificielle, le football apparaît ainsi comme l'*« un des sports les plus complexes à modéliser »*.

Objectif 50 % de réussite

Pas de quoi décourager les chercheurs en intelligence artificielle, qui continuent d'améliorer leurs modèles en affinant les paramètres et en sollicitant des machines de plus en plus puissantes. *« Nous progressons dans la compréhension des interactions entre les différentes variables, ce qui est de bon augure pour la suite »*, confie Gunther Schauburger. Pour ses prévisions footballistiques, le chercheur vise, à terme, un taux de réussite de 50 %, *« ce qui serait déjà très satisfaisant »*. Julien Maldonato, de son côté, n' imagine pas d'outil de prévision vraiment efficace en la matière *« avant plusieurs décennies »*. Les parieurs en ligne peuvent se rassurer : les ordinateurs ne sont pas près de tuer leur business !