

LIVRE BLANC



**DATA  
SCIENCE**  
L'utilisation  
de l'IA dans  
le scoring  
d'entreprise

**#04**



**DATA SCIENCE**



**Ellisphere** est un acteur majeur du marché de l'information économique, financière et légale sur les entreprises du monde entier. L'entreprise a hérité de plus de 100 années d'expérience dans la collecte, la gestion, l'analyse et la diffusion d'informations qualifiées.

Inscrite dans l'ère du numérique, son équipe de plus de 300 collaborateurs, répartis sur six grands pôles d'excellence, développe des solutions 100% digitales qui proposent de l'information en ligne sur près de 165 millions d'entreprises dans plus de 230 pays et territoires.

Depuis sa création, Ellisphere accompagne les directions marketing et commerciales dans leurs différents enjeux :

- Analyse de leurs marchés et développement de leurs connaissances clients
- Amélioration de la performance de leurs opérations marketing et commerciales
- Qualification de leurs bases clients et prospects et industrialisation de la gestion de leur data.

## Développement du chiffre d'affaires à l'appui d'un portefeuille clients fiable

La transformation numérique, la digitalisation, la personnalisation de la relation commerciale ainsi que la croissance exponentielle des données à traiter poussent les directions marketing et commerciales à se réinventer. Face à ces nouveaux enjeux, les organisations doivent faire preuve d'innovation pour répondre à ces défis et préparer demain.

Pour y faire face, identifier les données internes à leur système d'information ou externes pertinentes afin de gagner en efficacité dans la segmentation de leur marché, dans la connaissance client et dans l'amélioration de leur prospection commerciale devient vital.

Pour accomplir ses missions et accompagner au mieux ses clients, Ellisphere s'appuie sur :

- Un référentiel de data B2B alimenté par nos équipes et un réseau de partenaires international.

- Une équipe pluridisciplinaire en contact direct avec vos équipes marketing/commerciales/DSI,
- Une démarche structurée, quel que soit le profil de votre entreprise, avec un entretien contexte, un audit technique et de potentiel informationnel et une proposition d'accompagnement en 3 volets :
  - Un data-set personnalisé
  - Un mode de diffusion
  - Un accompagnement expert

Cette approche « par les enjeux » reposant sur le conseil, nous permet d'élaborer la solution la mieux adaptée à vos besoins.

L'ambition d'Ellisphere est d'être au service d'une croissance pérenne de l'activité de nos clients.





## **STEVEN HELLEC** RESPONSABLE DATA SCIENCE CHEZ ELLISPHERE

*Depuis une dizaine d'années, Steven Hellec est passionné par le machine learning et les technologies de traitement du Big Data dédiés à l'information stratégique. Il a été ainsi successivement data miner, ingénieur d'études et data scientist.*

*Aujourd'hui, en tant que responsable du département Data Science chez Ellisphere, il met sa passion et son expertise au service de la construction de solutions porteuses de valeur ajoutée. Sa mission est d'éclairer scientifiquement les prises de décision stratégiques en B2B des clients d'Ellisphere. Steven Hellec est diplômé d'un Master en Statistique et Business Intelligence.*



Raréfaction des données financières, digitalisation des process décisionnels, recherche d'une productivité accrue... La mutation engagée des métiers des directions financières, dont celui de credit management en B2B, nécessite de recourir à des outils innovants particulièrement en matière de scoring.

Dans ce livre blanc, nous abordons comment les technologies d'Intelligence Artificielle (IA) ont permis d'accroître de manière significative les performances du scoring d'entreprise. Nous reviendrons également sur les différentes étapes qui ont jalonné la création du score d'Ellisphere de probabilité de défaillance des entreprises.

Au centre des questionnements du credit manager, nous présenterons le contexte et les enjeux de l'usage de l'IA dans le scoring d'entreprise. Évolution des nouvelles technologies, explosion du nombre de données disponibles... Autant de problématiques qui sont aujourd'hui centrales dans les politiques de gestion des risques en B2B.

Ce livre blanc analysera également les conséquences de la crise Covid-19 sur le score d'Ellisphere de défaillance des entreprises, ainsi que sur les évolutions à venir en matière de scoring. Enfin, nous illustrerons un usage concret de l'IA chez Ellisphere avec la récente innovation « l'indicateur de presse » ; celle-ci permet d'analyser des données issues de résumés de presse pour établir un sentiment de défaillance d'une entreprise.

Bonne lecture



# Sommaire

## PARTIE 1

---

Contexte

5

## PARTIE 2

---

Enjeux : quels enjeux de l'IA dans le scoring

9

## PARTIE 3

---

Parole d'expert : interview de Steven Hellec

13

## PARTIE 4

---

Cas concret

16

## *Les origines du score de probabilité de défaillance des entreprises d'Ellisphere*

### Etat des lieux

Le concept d'Intelligence Artificielle est né dans les années 1950. Concrètement, qu'est-ce que c'est ? Il s'agit de permettre à des machines de penser et d'agir comme des êtres humains. Sa première application connue du grand public est le test de « Turing » qui vise à étudier si une machine est capable de penser.

Pendant des dizaines d'années, l'IA a principalement été un sujet de recherche. Son application dans les sphères économiques est plus récente. L'évolution de l'IA et son impact sur les sociétés modernes sont fortement corrélés à l'évolution de l'informatique. En effet, l'innovation technologique et la rapidité de calcul offertes aujourd'hui par le matériel informatique permettent d'envisager des applications qui n'étaient que de l'ordre du fantasme au XXe siècle.



L'application de l'IA dans le scoring des entreprises est intervenue assez tardivement. C'est à l'orée de l'année 2018 que les premiers modèles de score fondés sur le machine learning apparaissent. Cette avancée majeure permet désormais au score de probabilité de défaillance d'atteindre une précision jusqu'à lors inégalée. Si l'héritage laissé par les règles métier ainsi que les modèles statistiques perdurent, la vitesse de calcul et les nouvelles règles implémentées dans les modèles reposant sur l'IA donnent des résultats beaucoup plus précis.

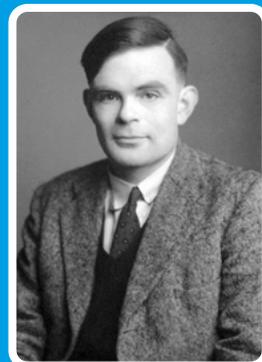
Pour bâtir ces scores de nouvelle génération, les équipes d'Ellisphere se sont ainsi reposées sur les informations de millions d'entreprises françaises à travers un historique de données de plus de 30 ans, ainsi que sur des expertises métier afin d'avoir les meilleures performances prédictives possible. Grâce à l'usage de l'IA, la performance du score de probabilité de défaillance des entreprises d'Ellisphere a augmenté significativement.

### Biographie d'Alan Turing

Né en 1912 à Londres et mort en 1954 à Wilmslow, Alan Mathison Turing est un mathématicien et cryptologue britannique ; il est l'auteur de travaux qui fondent scientifiquement l'informatique.

Afin de résoudre le problème fondamental de la décidabilité en arithmétique, il dévoile en 1936 une expérience de pensée que l'on nommera ensuite machine de Turing, ainsi que des concepts de programme et de programmation. Ces travaux offriront tout leur sens avec la diffusion des ordinateurs, dans la seconde moitié du XXe siècle.

Durant la Seconde Guerre mondiale, il tient un rôle majeur dans la cryptanalyse de la machine Enigma qui sera utilisée par les armées allemandes. Dans l'après-guerre, Alan Turing travaille sur l'un des premiers ordinateurs. En proposant le test de Turing, il contribue au débat sur la possibilité de l'Intelligence Artificielle. Vers la fin de sa vie, il s'intéresse à des modèles de morphogenèse du vivant conduisant aux « structures de Turing ».



## À retenir

### Les concepts clés de l'IA appliqués au scoring



#### MACHINE LEARNING

Le Machine Learning peut être défini comme un domaine d'études de l'Intelligence Artificielle permettant aux machines d'effectuer des tâches sans avoir été au préalable programmées spécifiquement à cet effet. Le Machine Learning est intimement lié au Big Data. En effet, les ordinateurs nécessitent de nombreux exemples pour apprendre à effectuer leur tâche cible.



#### SCORING DES ENTREPRISES

Le score donné à une entreprise vise à mesurer la probabilité de défaillance de l'entité, c'est-à-dire son risque de faire faillite. En B2B, les acteurs économiques ont besoin de connaître la solvabilité de leurs relations d'affaires. Cette information passe notamment par l'évaluation du risque de défaillance matérialisée sous la forme d'un score propre à chaque entreprise.



#### ÉCHANTILLONNAGE

Procédure qui consiste à sélectionner un ou plusieurs échantillons d'éléments individuels à partir d'un ensemble de ces éléments.

## SCORE D'ELLISPHERE QUELLE CHRONOLOGIE ?



### 1970-1990 : LE SCORING PAR RÈGLES MÉTIER

Au début du scoring sur les entreprises, plus spécifiquement pour des besoins de credit management, les experts métier ainsi que les credit managers, par leur connaissance des dossiers, estimaient que certains éléments pouvaient amener à la défaillance d'une entreprise partenaire commerciale. Les premières pondérations ont été mises en place. Si les résultats de l'époque permettaient de manière peu précise d'établir des liens entre certains éléments financiers et la probabilité de défaillance d'une entreprise, il n'en reste pas moins que les avancées techniques et technologiques ont permis d'accroître la précision du score.



### 1990-2018 : LE MODÈLE STATISTIQUE

Dès le début des années 1990, les premiers modèles statistiques apparaissent. Ils vont perdurer jusqu'en 2017. Cette nouvelle méthodologie apporte une révolution sur plusieurs aspects. En effet, si le modèle de première génération reposait principalement sur l'expertise métier de l'expert, le modèle statistique va permettre d'aller encore plus loin sur le système de pondération de la donnée.

Cette deuxième génération de scoring va permettre au score de se nourrir de l'intégralité d'un référentiel de données, mais également d'analyser les événements qui ont eu lieu dans le passé. Cette évolution a permis d'avoir une vision des entreprises beaucoup plus exhaustive grâce à cet ensemble d'information plus vaste dont disposait l'équipe d'Ellisphere.



### 2018-ACTUEL : L'AVÈNEMENT DE L'IA DANS LE SCORING

Le développement de nouveaux algorithmes, ainsi que des matériaux informatiques plus puissants, ont amené l'IA du monde de la recherche à l'application en entreprise. Ces nouvelles techniques, plus performantes, ont permis de passer à la vitesse supérieure et d'envisager un scoring fondé sur les technologies de machine learning.

Ces nouveaux algorithmes ont favorisé plusieurs choses. Tout d'abord, le volume de données traitées par le modèle a pu être grandement amélioré. Ensuite, le nombre de variables traitées par les algorithmes est beaucoup plus important. Cela permet de prendre en compte des situations très diverses qui peuvent amener une entreprise à la défaillance.

Le score de 3<sup>ème</sup> génération bâti sur des technologies de l'IA permet de mieux valoriser les interactions entre les différentes variables, ce qui était beaucoup plus compliqué et moins performant avec le précédent score. Cette évolution a engendré une multiplication des échantillons et donc une précision accrue du score de probabilité de défaillance.

## Ce qu'en pensent **les experts**

*L'IA dans les systèmes de scoring n'en est qu'à ses débuts. Son utilisation va s'amplifier avec l'engouement des acteurs privés et publics autour de son utilisation. Plusieurs défis se posent à nous dans les années à venir pour intégrer de nouveaux types de données au score de probabilité de défaillance.*



**Mélanie Bourguin,**  
data scientist chez Ellisphere

# À retenir

Comment ça marche ?

## La modélisation du score

Prenons l'exemple concret de la base de données d'Ellisphere. L'équipe d'Ellisphere a répertorié toutes les informations qu'elle avait à sa disposition sur les entreprises sur les quinze dernières années afin d'être le plus exhaustif possible. Le référentiel d'Ellisphere est ainsi composé de nombreuses informations économiques, légales et financières sur plus de 10 millions d'établissements d'entités actives en France.

Ensuite, l'équipe a scindé la base de données en deux parties. Une qui servira à construire le modèle de score et la deuxième qui aura pour objectif de tester ce modèle.

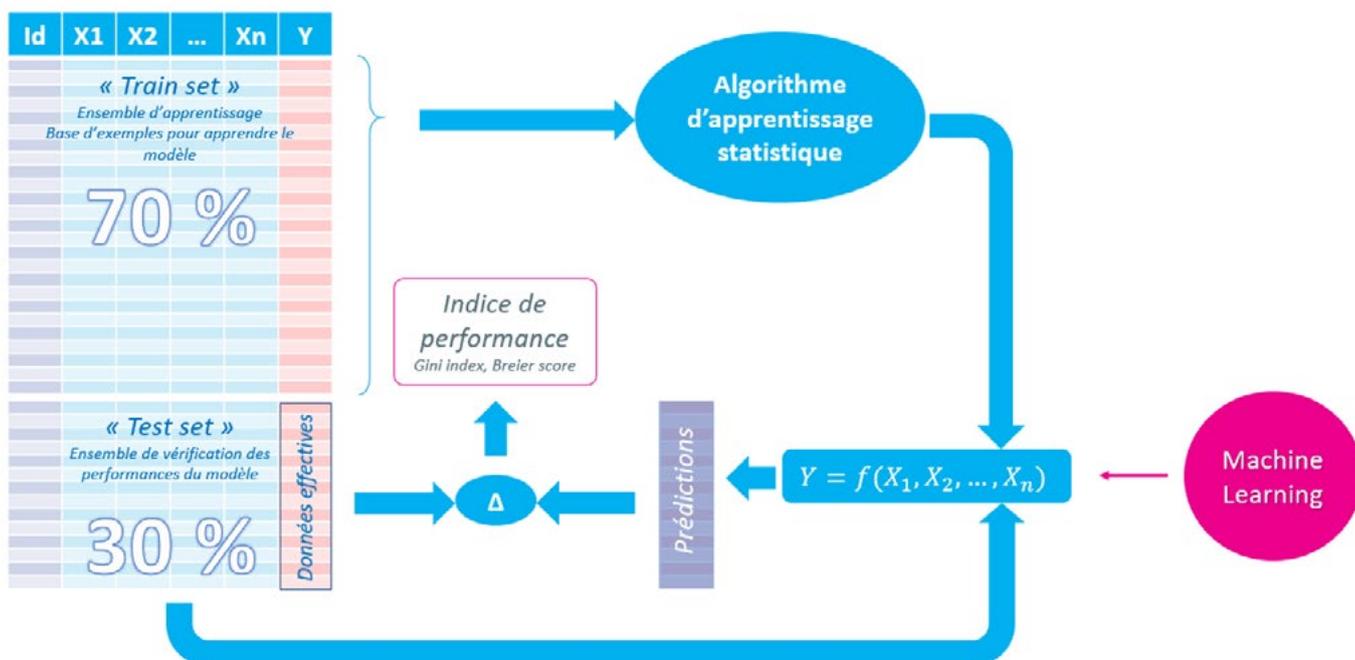
La difficulté à laquelle l'équipe d'Ellisphere a été confrontée est la suivante : une défaillance d'entreprise est un événement rare et donc difficile à prédire. Si on prend l'exemple des sociétés commerciales, celles-ci ont un taux de défaillance de seulement 2% ; pour l'ensemble des entreprises, le taux redescend à 0,8%.



C'est ici que les technologies de machine learning rentrent en action. En effet, à partir d'une grande quantité d'informations, la machine est capable d'avoir une vision globale et d'identifier tous les critères de défaillance d'entreprise en exploitant toutes les données à disposition. Cela peut aisément dépasser plusieurs milliards d'informations.

C'est à partir de ces observations que le modèle de score voit le jour. La création de l'algorithme permet d'observer, selon des critères préalablement définis et grâce aux événements du passé, les éléments qui peuvent amener une entreprise à être déclarée défaillante d'ici à 12 mois.

## Schéma de fonctionnement du score



## Quels enjeux de l'IA dans le scoring ?

L'utilisation de l'IA dans le scoring a permis d'entrevoir de nouvelles pratiques et des évolutions très rapides pour analyser l'information sur les entreprises. Tour d'horizon des enjeux actuels et à venir avec Christophe Goudet, data scientist chez Ellisphere.

1

### AMÉLIORER LA COMPRÉHENSION DES MODÈLES EN S'APPUYANT SUR LES OUTILS D'EXPLICABILITÉ

Pour ne pas avoir à choisir entre performance et confiance, la notion d'explicabilité est essentielle. Dans son développement, un système d'Intelligence Artificielle est une boîte noire. Le data scientist doit donc développer des outils de mesure visant à améliorer la compréhension des algorithmes pour les rendre compréhensibles par les utilisateurs finaux.

Les résultats de ces techniques peuvent être eux-mêmes obscurs pour des non data scientists. Il est alors nécessaire de réfléchir à la manière dont ces résultats seront présentés aux clients finaux afin que ceux-ci puissent réellement comprendre et s'approprier le résultat du modèle. La partie interface et expérience utilisateur doit alors être prise en compte dans le projet d'IA.

Chacune de ces interfaces doit être adaptée en fonction du modèle et du type d'utilisateur final, variant d'une simple phrase expliquant les variables ayant impacté la prédiction pour les utilisateurs les plus aguerris, à un tableau de bord présentant les explications sous forme visuelle pour les utilisateurs métier, peu familiers des algorithmes.

Grâce à l'explicabilité, nous sommes aujourd'hui en mesure de démontrer que l'IA atteint son plein potentiel quand elle est couplée à l'intelligence humaine, qu'elle augmente sans la remplacer.

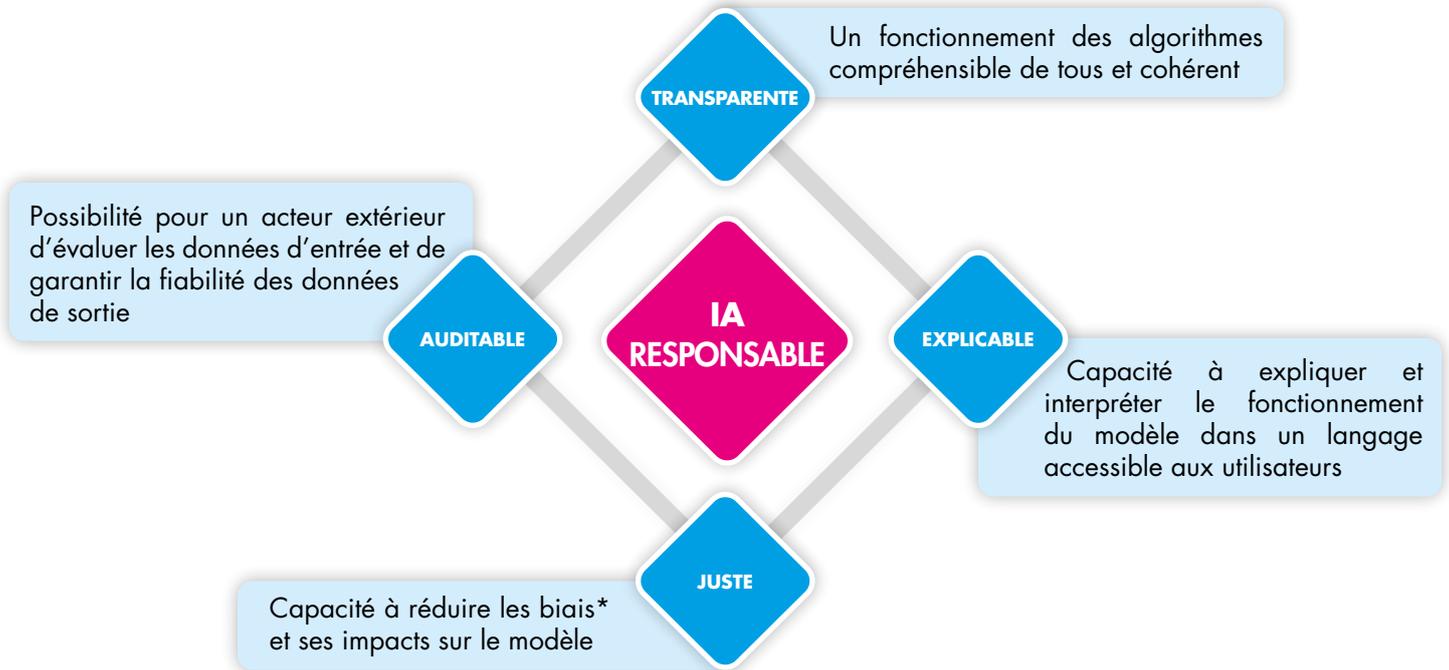


**Christophe Goudet,**  
data scientist chez Ellisphere



Cependant, l'explicabilité n'est pas limitée à l'usage des clients finaux. C'est également un outil précieux du data scientist pour mieux comprendre les rouages du modèle : la manière dont la donnée et l'algorithme interagissent pour proposer un résultat. Grâce à cette compréhension, le data scientist peut identifier des problèmes de données, des relations inconnues, voire des biais. C'est donc un outil essentiel pour s'assurer de la qualité du modèle.

L'explicabilité permet aujourd'hui un contrôle de la pertinence et de l'éthique des prédictions. Toutefois, l'IA ne pourra poursuivre son ascension dans notre société sans que les utilisateurs qui agissent et interagissent avec ces technologies ne soient en mesure de les maîtriser et de les comprendre.



\* Un biais algorithmique est le fait que le résultat d'un algorithme ne soit pas neutre, loyal ou équitable. Le biais algorithmique peut se produire lorsque les données utilisées pour entraîner un algorithme d'apprentissage automatique reflètent les valeurs implicites des humains impliqués dans la collecte, la sélection ou l'utilisation de ces données.

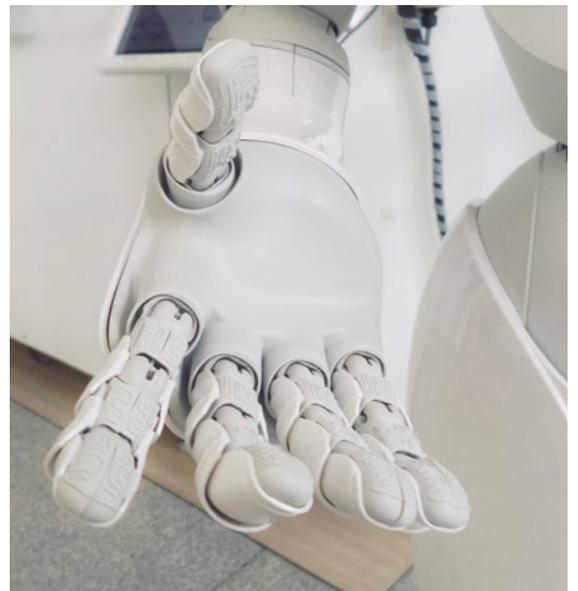
.....

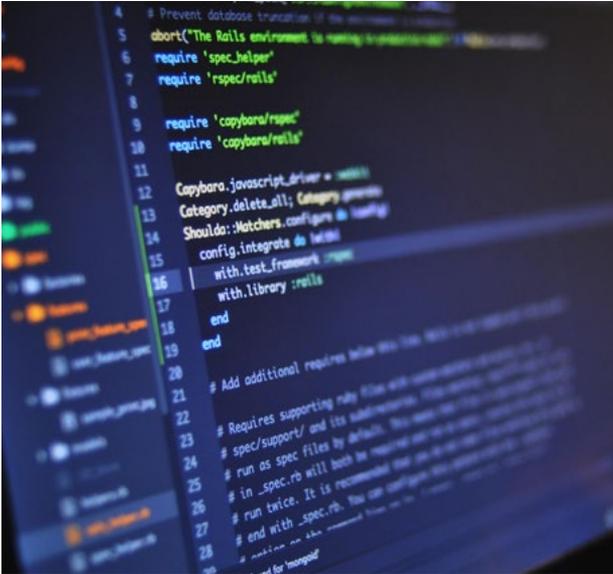
## 2

## LA COURSE AUX ÉVOLUTIONS TECHNOLOGIQUES

En l'espace de 20 ans, les évolutions technologiques se sont accélérées. Les apports de ces évolutions suscitent pourtant encore des interrogations, voire des craintes. L'Intelligence Artificielle divise le monde. Si certains pensent qu'elle va révolutionner nos sociétés, d'autres la voient comme une menace qui dirigera l'humanité vers sa perte. En l'espace de quelques années, la puissance de calcul des machines s'est accélérée permettant ainsi d'appréhender de nouvelles perspectives pour l'analyse des données.

Cependant, dans certains cadres d'application tels que le scoring, l'utilisation de l'IA n'est pas dénuée de contraintes. D'un côté, la réglementation (dans le système financier par exemple) peut imposer des contraintes de justice ou d'explicabilité des modèles. Les utilisateurs de la solutions peuvent également être sources de contraintes sur les technologies à employer. Les toutes dernières techniques à la mode ne sont donc pas systématiquement utilisables immédiatement pour tous les usages.





Dans le scoring d'entreprise par exemple, il y a une longue expérience de l'analyse financière des acteurs économiques. Les analystes utilisant une solution d'IA sont donc en forte attente des critères utilisés par la solution pour évaluer une entreprise. Dans certains cas, l'explication est elle-même plus importante que la performance globale de la solution.

L'un des défis de demain résidera également dans l'optimisation. En effet, les capacités de calcul alliées aux évolutions techniques permettent aujourd'hui d'atteindre des résultats plus précis et moins coûteux en termes d'infrastructure et de temps.

Les évolutions du big data et le nombre exponentiel de données accessibles sont également au cœur des enjeux technologiques à venir. La question de l'optimisation du choix de la donnée et de son traitement deviendra de plus en plus importante dans les années à venir.



## 3

## LE COMBAT CONTRE LES BIAIS

Dans le cadre du scoring d'entreprise, la question des biais existe également. L'algorithmie est fondamentalement neutre, c'est-à-dire que l'algorithme va optimiser une tâche qui lui est demandée. La vraie question des biais se pose en amont, dans les données que l'on va utiliser pour faire travailler la machine.

L'introduction de biais en amont du travail des algorithmes, dès le stade de la récolte des données, doit donc être prise en compte. Notamment quand il s'agit d'algorithmes d'apprentissage automatique. En effet, si les données sont biaisées avant de rentrer dans la machine, ils seront accentués par le travail d'optimisation de l'IA.

Prenons un cas simplifié, basé sur l'échec d'un système de recrutement automatique d'Amazon. Le programme existait depuis 2014 et permettait de trier les curriculum vitae des candidates et candidats qui postulaient à un emploi chez Amazon.

Problème : un an après ses débuts, les femmes qui candidataient pour des emplois techniques (développement, programmation) étaient systématiquement moins bien notées parce qu'elles étaient des femmes. Ce biais était lié à la manière dont a été conçue l'IA. En effet, elle avait été entraînée avec les profils des personnes employées dans l'entreprise entre 2004 et 2014... Principalement des hommes. L'IA en a conclu que sélectionner d'abord des candidats de sexe masculin était préférable.



La recherche du biais peut se faire en amont, mais également en aval du travail de création du modèle. Dans un premier temps, il doit être testé sur des données qui n'ont pas servi à son entraînement afin d'évaluer sa réelle capacité d'abstraction. Durant ce test, on peut alors évaluer les variables qui influent le plus sur les résultats, mais également les performances relatives du modèle pour des sous-catégories d'entreprises. Une fois le modèle validé et livré en production, il est également important de suivre l'évolution de ses métriques de biais afin de s'assurer que le comportement du modèle correspond bien aux attentes sur les données réelles, qui peuvent être différentes des données d'entraînement et de validation.

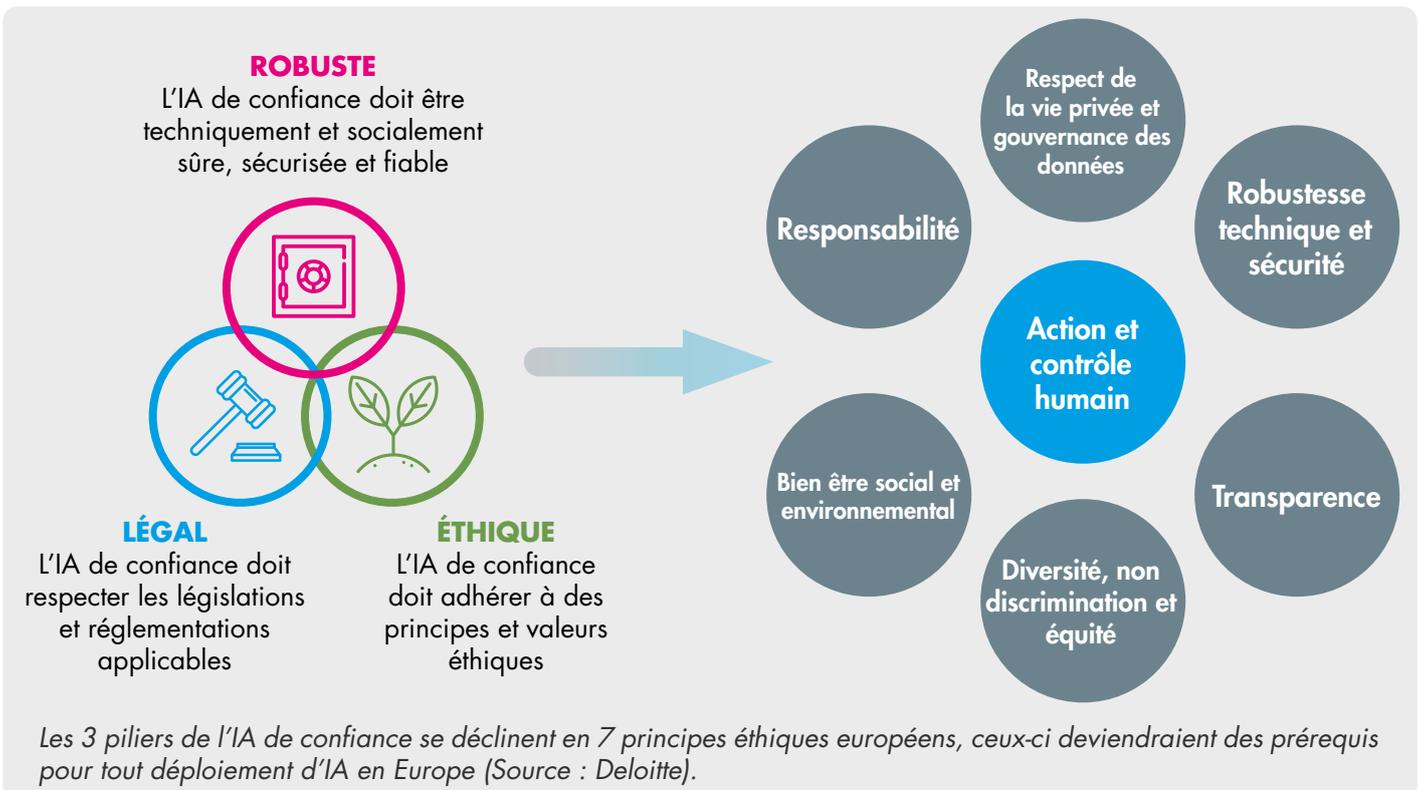
La responsabilité du modèle incombe donc à son créateur. C'est à lui de fixer la justesse des données et de l'algorithme afin que le résultat soit le plus juste possible.

## 4

### LES LIMITES DE LA RÉGLEMENTATION

L'Intelligence Artificielle a aujourd'hui un rôle majeur dans la transformation numérique des entreprises et plus globalement de nos sociétés. Travail, finance, santé, sécurité, agriculture... Cette révolution impacte aujourd'hui chaque pan de l'économie. Cependant, la question de l'éthique reste en suspens...

Alors qu'une réglementation européenne autour de l'IA devrait voir le jour dans les prochains années, ses impacts questionnent et inquiètent les acteurs du marché. Innovation, accès aux données, protection des citoyens, éthique, recherche, enjeux juridiques, sociaux et économiques, les impacts de la future réglementation sont nombreux.



Cependant, alors que les réflexions autour des enjeux de l'IA se poursuivent en Europe, les États-Unis et la Chine investissent massivement et mettent en place des stratégies nationales pour la développer. La crainte des acteurs européens de devenir une usine à produire des données dont les géants américains et chinois tireront toute la valeur est de plus en plus palpable.

La question de la réglementation se pose donc comme un enjeu majeur de la transformation de l'Europe. Elle doit à la fois répondre aux enjeux d'une IA éthique, mais également à ceux d'une concurrence saine dans un monde en pleine mutation technologique.



## STEVEN HELLEC

### *Les enseignements de la crise COVID sur le scoring d'entreprise* //

#### **Comment la crise COVID a-t-elle impacté votre travail de data scientist ?**

**Steven Hellec :** La situation à laquelle nous devons faire face aujourd'hui est inédite. La crise de la Covid-19 ne ressemble à aucune crise économique mondiale connue jusqu'à lors. Si la plupart des acteurs économiques, quel que soit le secteur, semblent menacés, les impacts à terme sur l'économie mondiale sont eux difficiles à appréhender.

Le data scientist va se fonder sur des éléments rationnels (data présente dans les bases de données qui sont à sa disposition) quand il va modéliser les données pour faire ses prévisions. Ces référentiels sont en quelque sorte des « enregistrements » de notre vie. Prenons l'exemple d'Ellisphere, sa base de données est constituée d'enregistrements des moments de vie des entreprises.

En principe, si on observe qu'une entreprise a fait défaut à un moment donné, on analysera les enregistrements précédents qui auront été stockés dans les bases de données pour comprendre ce qui pourrait expliquer la défaillance.

Par exemple, pour une entreprise en situation de défaillance, on aura pu remarquer qu'elle avait des difficultés financières, qu'elle cumulait les factures impayées... 1 à 2 ans avant qu'elle

ne fasse défaut. Nos modèles statistiques vont se nourrir de ce type d'information pour pouvoir prédire le risque de défaillance de chaque entreprise dans les 12 prochains mois.

Aujourd'hui avec la crise de la Covid-19, nous avons des entreprises qui font défaut alors que rien ne le présageait. Cela s'explique principalement parce que la crise leur est littéralement tombée dessus très rapidement ; les dispositifs gouvernementaux n'ont pas eu le temps de donner les effets escomptés sur toutes les entreprises concernées. À l'inverse, des entreprises qui devaient « logiquement » faire défaut ont survécu. Cette autre situation est plus fréquente, car les dispositifs gouvernementaux ont permis, à priori, de redonner du souffle à des entreprises en situation délicate, parfois en très grande difficulté. Mais n'est-ce pas ici un simple report de l'échéance fatidique ?





2020 est une année avec un nombre de défaillances d'entreprise historiquement bas. En effet, on compte près de 31 000 défaillances (ouvertures de redressement et de liquidation judiciaire directe) sur l'année. Un résultat qui peut paraître paradoxal au vu de la situation que nous vivons. Par comparaison, 2019 qui était une vraie bonne année économiquement parlant, avait vu 50 000 entreprises faire défaut. Une certaine logique semble dorénavant mise à mal. En conséquence, les seules informations disponibles dans nos référentiels ne sont plus suffisantes pour bâtir nos modélisations.

Pour répondre à cette problématique, nous avons, au sein Ellisphere, changé d'approche en nous appuyant davantage sur nos experts métier. Ils sont en mesure de considérer des éléments singuliers de la crise comme les PGE (Prêt garanti par l'État) et le chômage partiel de masse, et d'ajuster ainsi le modèle prédictif.

Plus que jamais, la data science travaille conjointement avec les experts pour affiner statistiquement leurs appréhensions. En effet, l'humain a très souvent tendance à surévaluer le risque de défaillance. Ce qui n'est pas une bonne chose en soi. Notre objectif est bien la vision la plus juste et la plus précise de la défaillance. Notre rôle est aujourd'hui majeur. Une surpondération du risque aurait pour conséquence un accroissement de défiance de nos clients envers leurs partenaires commerciaux avec pour conséquence directe, un ralentissement des relations d'affaires ayant pour corollaire une amplification de la crise économique actuelle.

### **Quels enjeux à venir concernant le rôle de l'IA dans la prédictivité du score de défaillance ?**

**SH :** L'utilisation des technologies de machine learning dans le calcul du score de probabilité de défaillance d'entreprise a permis d'atteindre des résultats inégalés en comparaison de méthodes statistiques standard. De ce fait, la crise économique que nous vivons actuellement n'a en aucun cas remis en cause l'utilisation des technologies d'Intelligence Artificielle pour le scoring des entreprises.

Cependant, la crise de la Covid-19 nous a amenés à nous questionner. En effet, pour améliorer la prédictivité de notre score, nous allons mettre en place des protocoles qui permettront de mieux gérer ce genre d'événements à l'avenir. De plus, cela nous permettra de disposer d'une prise en compte plus rapide des impacts de la crise sur les entreprises scorées par Ellisphere.



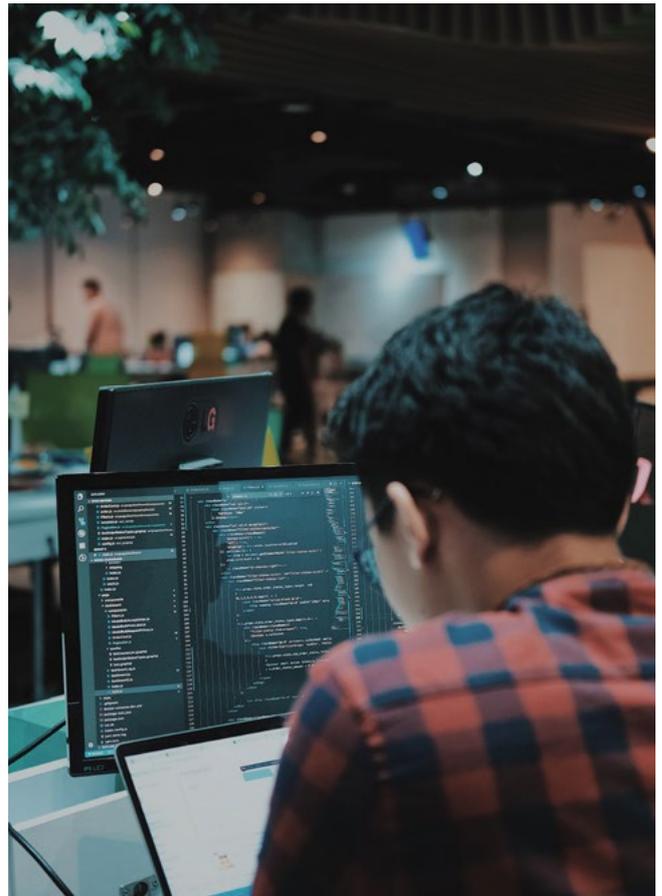
## Demain, quelles nouvelles applications peuvent être imaginées avec l'IA dans le scoring ?

**SH :** La crise de la Covid-19 a mis en exergue que la prédiction du défaut d'une entreprise ne s'est pas généralisée à tous les domaines. Cette situation inédite aura permis de révéler certaines failles dans l'usage de l'IA. Cependant, comme toute technologie, l'IA va apprendre de la situation pour s'adapter aux nouveaux problèmes observés. De nombreuses solutions vont rapidement être trouvées pour répondre à cette problématique.

Aujourd'hui, il reste difficile de se projeter. En effet, il existe de plus en plus d'applications qui visent à analyser des données non structurées telles que les données de presse ou celles venant des réseaux sociaux, qui offrent une information plus actualisée. L'objectif : gagner en réactivité par rapport aux systèmes qui s'appuient sur des données structurées, plus conventionnelles comme les publications de bilan.

Cependant, l'analyse de cette nouvelle data n'est pas sans difficulté. En effet, les sources disponibles, par exemple sur internet, ne sont pas toujours fiables. Elles comportent beaucoup de bruits, beaucoup d'informations qui ne sont pas fondées sur des faits (fake news), et qui font que les systèmes peuvent être influencés et s'emballer pour un « oui » ou pour un « non ». L'implémentation de ces données doit donc être réfléchiée et maniée avec précaution.

Comme pour la recherche d'un vaccin, les nouvelles solutions d'IA prendront du temps avant de pouvoir être éprouvées et déployées pour être utilisées par le plus grand nombre.





## COMMENT DISPOSER D'INFORMATIONS SUR LES ENTREPRISES PLUS ACTUALISÉES GRÂCE À LA PRESSE ?

### CONTEXTE

Un acteur de l'e-commerce souhaite accroître la surveillance sur ses clients et fournisseurs en analysant des données davantage actualisées. Afin de pouvoir monitorer son risque, nous lui proposons de bénéficier des informations détenues dans notre base de résumés d'article de presse.

Ces nouvelles données actualisées lui permettront de profiter d'une surveillance plus en temps réel. Pour ce faire, nous avons créé un indicateur dédié à l'analyse sémantique des résumés d'article de presse pour apprécier le risque de défaillance décelé dans chaque article.

## Phase préparatoire

### ÉTAPE 1 - LABELLISATION DES ARTICLES

Ce travail de labellisation a consisté à lire des résumés d'article de presse par nos experts pour ensuite, définir s'ils pouvaient informer d'une défaillance future (sentiment négatif) ou non (tonalité positive/neutre).

### ÉTAPE 2 - PRÉ-PROCESSING DE LA DONNÉE

Pour rendre notre ensemble de données analysable, nous nous sommes astreints à un prétraitement du texte standard dans le domaine du langage naturel NLP\*. Cette étape indispensable vise à formater la donnée afin que le modèle d'analyse puisse l'étudier et la scorer.

### ÉTAPE 3 - CRÉATION DU MODÈLE

Mise en place, entraînement du modèle et analyse des résultats.

### CE QUE L'IA A PERMIS DANS CE PROJET

- Analyser à grande échelle un grand nombre de données textuelles
- Automatiser l'analyse des résumés d'article de presse grâce à des algorithmes d'analyse textuelle (deep learning)
- Un modèle qui nous permet de qualifier la tonalité de chaque résumé d'article (positif ou négatif)

À l'issue de la phase de recherche, notre modèle a réussi à atteindre 92% de précision sur l'attribution d'une tonalité pour chaque résumé d'article.

\*Le traitement naturel du langage, aussi appelé Natural Language Processing (NLP) en anglais, est une technologie permettant aux machines de comprendre le langage humain grâce à l'Intelligence Artificielle.

Une fois le travail de recherche et développement terminé, l'équipe data science d'Ellisphère, architecte de ce projet, a ensuite travaillé sur son utilisation concrète par notre client. La dernière phase du projet a été réalisée avec la direction des services informatiques qui a participé à l'implémentation de l'indicateur de presse dans la plateforme d'information sur les entreprises Ellipro.

# NOTRE SOLUTION

## ÉTAPE 1 - MISE EN PLACE



Mise en relation des résumés d'article de presse avec les SIREN d'entreprises concernées afin de mettre en place l'indicateur de presse sur Ellipro.

## ÉTAPE 2 - OUTILS DE SURVEILLANCE

Mise en place d'un système de notification en push à destination de l'utilisateur pour l'informer en temps réel des informations de presse négatives sur les entreprises de son portefeuille.



## ÉTAPE 3 - MISE EN PRODUCTION



Déploiement du nouvel indicateur sur Ellipro et mise en place de la surveillance des entreprises.

## CE QUE NOTRE SOLUTION A PERMIS

- Un accès à l'ensemble du référentiel d'Ellisphère comprenant des résumés d'article de presse sur les entreprises françaises
- Une lecture facilitée de cette information de presse
- Des données plus actualisées sur l'état de santé des entreprises
- La surveillance en temps réel des partenaires commerciaux de nos clients grâce aux alertes

## CONCLUSION

Grâce à l'indicateur de presse, notre client peut donc désormais profiter d'une information plus actualisée sur ses tiers avec mise sous surveillance de son portefeuille client.

Cette nouvelle approche lui permet de disposer d'une meilleure réactivité quant à la remontée d'informations négatives pouvant impacter la santé de ses relations d'affaires.



# SUIVEZ

## L'ACTUALITÉ D'ELLISPHERE



[www.ellisphere.com](http://www.ellisphere.com)



@ellisphere



@ellisphere



@ellisphere



Découvrez nos publications :  
[www.ellisphere.com](http://www.ellisphere.com)

# DÉCOUVREZ LE PODCAST ELLISPHERE



ÉCOUTEZ  
LE PODCAST

## INTELLIGENCE ARTIFICIELLE ENTRE MYTHE ET RÉALITÉ

**N'hésitez pas à découvrir notre podcast dédié à l'Intelligence Artificielle. Nos invités ont pu débattre pendant 45 minutes autour de la genèse de la notion d'IA ainsi que de ses applications aujourd'hui dans nos vies.**

Invités :

**Steven HELLEC**, responsable data science chez Ellisphere

**Christophe GOUDET**, data scientist chez Ellisphere

**Vincent DECUGIS**, predictive analytics manager chez BNP Paribas

## PROGRAMME

Partie 1 : Introduction au concept de l'IA

Partie 2 : Intelligence Artificielle VS Intelligence Augmentée : les mythes autour de l'IA

Partie 3 : Applications concrètes de l'IA en B2B



# ellispherē

Pour un monde économique fiable et une croissance durable

[www.ellisphere.com](http://www.ellisphere.com)

Lille - Lyon - Marseille - Nanterre - Nantes - Strasbourg - Toulouse - Tours  
Siège social : Immeuble Via Verde - 55 place Nelson Mandela - 92000 Nanterre  
Adresse postale : 37 rue Sergent Michel Berthet - CS 99063 - 69255 Lyon Cedex 09  
S.A.S. au capital de 2 511 921 euros - 482755741 RCS Nanterre  
N° ORIAS 07029136 - [www.orias.fr](http://www.orias.fr) - Sous le contrôle de l'ACPR

Autorité de contrôle prudentiel et de résolution - 4 place de Budapest - CS 92459 - 75436 Paris Cedex 09