

**Denis Jacopini**

Expert informatique



**Visouda Luangkhot**

Experte  
Intelligence  
Artificielle



**Dr Jean-Baptiste Camps**

Expert Stylométrie



**Maître François Saint-Pierre**

Avocat



**Maître Ouadie El Hammamouchi**

Avocat



# ACTES DE LA TABLE RONDE

## Comment les nouvelles technologies influencent les enquêtes judiciaires et les affaires pénales ?

De l'enquête à leurs conclusions

Date

**17 NOVEMBRE  
2025**

**I**l importance des nouvelles technologies : apports déterminants pour les investigations et les enquêtes. L'intelligence artificielle (IA), la stylométrie, les expertises informatiques. Discussion et études de cas.

# Table des matières

<i>Cadre introductif et portée scientifique des interventions .....</i>	3
<i>Transcription de la table ronde .....</i>	6
Introduction générale par Maître Clarisse Juteau-Csosz .....	6
Denis Jacopini, Expert de Justice en informatique – Cybercriminalité & Protection des données.....	8
Visouda Luangkhot – Experte en Intelligence Artificielle.....	17
Dr Jean-Baptiste Camps, Expert en stylométrie .....	26
Me François Saint-Pierre, avocat : Réflexions et application concrètes dans « l'affaire du petit Grégory ».....	42
Me Ouadie Elhamamouchi, avocat : Réflexions et applications concrètes dans l' « affaire Tariq Ramadan » en France et en Suisse .....	49
<i>Conclusion .....</i>	59



# Cadre introductif et portée scientifique des interventions

Ce colloque consacré à l'intégration des technologies avancées dans les enquêtes judiciaires a illustré avec une intensité saisissante la recomposition profonde du champ pénal. À travers les analyses convergentes d'experts techniques, de chercheurs en stylométrie et d'avocats engagés dans des dossiers emblématiques, une même évidence s'est imposée : la justice entre dans une nouvelle ère, où la vérité ne se construit plus seulement dans la contradiction, mais aussi dans la capacité à déchiffrer, analyser et attribuer des traces numériques d'une complexité croissante.

**Denis Jacopini**, Expert de justice spécialisé en cybercriminalité, a ouvert les débats en montrant comment, depuis vingt-cinq ans, les investigations techniques transforment déjà silencieusement l'enquête pénale. Fort de son expérience devant les juridictions, il a détaillé la manière dont l'analyse des logs, la traçabilité informatique, la reconstruction de circuits numériques ou la détection d'intrusions permettent aujourd'hui de révéler des éléments probatoires autrefois inaccessibles. Son intervention a montré que la justice française dispose déjà d'un socle solide d'expertise technique, capable d'articuler rigueur scientifique et respect des libertés fondamentales — tout en rappelant les risques d'atteintes à la vie privée et les enjeux de surveillance de masse, qu'il a mis en garde avec une clarté remarquable.

**Visouda Luangkhot**, spécialiste de la transformation digitale, a exposé le versant organisationnel et humain de ces mutations. En retraçant les évolutions rapides de l'intelligence artificielle, elle a montré comment les institutions — publiques comme privées — doivent restructurer leurs méthodes, leurs compétences et leurs process pour intégrer ces nouveaux outils. Elle a explicité, avec précision, la façon dont l'IA modifie la formation, le travail quotidien et la prise de décision. Son intervention a replacé les technologies judiciaires dans une dynamique plus large : celle d'une transition

numérique qui n'est pas seulement technique, mais culturelle, et qui impose que magistrats, enquêteurs et avocats acquièrent de nouveaux réflexes professionnels.

Ce socle a été prolongé par **Dr Jean-Baptiste Camps**, maître de conférences et l'un des chercheurs les plus reconnus en France dans le domaine de la stylométrie. En revenant sur ses travaux ayant relancé le débat sur l'attribution des textes de Molière, et sur son ouvrage *Affaire de style*, il a offert une synthèse magistrale de ce que permet l'analyse statistique du style : profilage linguistique, attribution d'auteur, détection de similarités ou de ruptures, et révélation de marqueurs propres à des individus ou à des groupes. Ses explications ont posé les fondations méthodologiques indispensables pour comprendre l'usage judiciaire de la stylométrie — et la manière dont cette discipline, autrefois réservée aux philologues, est désormais appelée à jouer un rôle clef dans les procédures pénales contemporaines.

Cette mise en perspective scientifique a trouvé sa résonance immédiate dans l'intervention de **Maître François Saint-Pierre**, figure majeure du barreau français. Son retour d'expérience sur l'affaire Grégory a souligné que l'expertise stylométrique utilisée ne venait pas de n'importe où, mais de pionniers européens, notamment une société suisse à la pointe, dont les travaux ont permis d'éclaircir des zones d'incertitude vieilles de plusieurs décennies. Avec son exigence habituelle, Saint-Pierre a montré comment une défense rigoureuse, fondée sur les principes fondamentaux du procès pénal, peut intégrer ces outils sans jamais renoncer à l'éthique ni à la prudence, et comment, dans certains cas, la science permet enfin de dénouer des énigmes judiciaires qui semblaient insolubles.

Enfin, **Maître Ouadie Elhammamouchi** a démontré, à travers l'affaire Tariq Ramadan, toute la puissance concrète de la stylométrie dans l'analyse d'écrits anonymes et la lutte contre les manipulations numériques. Grâce à l'expertise réalisée par le Dr Carole Chaski, l'une des références mondiales dans ce domaine, il a été possible d'attribuer des messages malveillants à certaines plaignantes du dossier actuel, messages évoquant explicitement un piège contre l'accusé, et ce, avant leur unique rencontre avec lui. Cette révélation, impossible à établir par les seules méthodes classiques, a constitué un tournant majeur dans la compréhension du dossier. Elhammamouchi a ainsi montré comment l'analyse

stylométrique peut défaire les mécanismes de dissimulation, lever l'anonymat stratégique, et redonner à la défense les moyens d'établir la vérité dans les dossiers où « la parole contre la parole » semblait condamner la démarche judiciaire à l'impasse.

Tous ces regards convergent vers une même conclusion : nous entrons dans un moment historique où l'expertise technique, linguistique et numérique devient un pilier essentiel de la manifestation de la vérité. Ces actes montrent qu'il est désormais impossible d'ignorer ces outils ; ils doivent être maîtrisés, encadrés, discutés, mais surtout intégrés, pour que la justice puisse répondre à la complexité du réel contemporain.

# Transcription de la table ronde

## Introduction générale par Maître Clarisse Juteau - Modératrice du colloque.

« Je me présente, Maitre Clarisse Juteau, je suis avocate au barreau de Paris. Je suis très heureuse de vous voir tous assis devant pour nous écouter pour cette table ronde sur cette question très intéressante de l'intégration des nouvelles technologies dans les enquêtes de police judiciaire et les instructions en France. Nous avons la chance d'avoir des experts et des avocats qui connaissent et traitent de ces questions dans leur pratique.

Je vais vous les introduire très rapidement. Nous allons d'abord écouter les experts, puis les avocats. À l'issue de chaque intervention, vous pourrez poser des questions, discuter et débattre ensemble de cette question qui est d'actualité et en pleine évolution.

Nous écouterons Denis Jacopigni, expert de justice assermenté. Il est spécialisé dans l'analyse des infractions numériques. Il est fort de 25 ans d'expérience devant les juridictions. Il intervient en qualité d'expert judiciaire.

Ensuite, nous aurons la chance d'écouter Visouda Luangkhot, experte en formation digitale. Elle accompagne les organismes dans leur transformation digitale, elle enseigne et elle explique à ces organismes comment se matérialise la transformation technologique, en intégrant des technologies que l'on connaît, comme celle de l'intelligence artificielle.

Nous écouterons ensuite docteur Jean-Baptiste Camps, docteur en études médiévales et maître de conférences, il a co-dirigé l'étude stylométrique qui a relancé le débat sur l'attribution des œuvres à Molière. Il est co-auteur de l'ouvrage "Affaire de style, du cas de Molière à l'affaire Grégory", que vous connaissez et qui a été dans l'actualité

récemment. C'est un des experts les plus reconnus en France en matière d'études stylométrie.

Nous aurons enfin un regard plutôt tourné vers la pratique judiciaire, avec deux éminents avocats qui sont mes confrères, moi, en tant qu'avocate qui ait rejoint récemment le Barreau de Paris. Nous écouterons dans un premier temps Maître François Saint Pierre, avocat pénaliste, reconnu ténor du barreau, comme on les appelle dans la pratique. Il intervient dans des dossiers exclusivement pénaux, il est expert dans la matière et donc il est intervenu dans des dossiers complexes et sensibles dont un en particulier qui a pu faire émerger et connaître à l'opinion publique, la technique de la stylométrie dans une affaire très médiatisée. Il vous en évoquera les tenants et les aboutissants ce soir. Et enfin, nous aurons l'occasion d'écouter Maître Ouadie El-hammamouchie qui est aussi avocat pénaliste de terrain. Il intervient dans le cadre de la justice pénale d'urgence et il a pu également, dans le cadre de ces dossiers, voir l'intervention de ces nouvelles technologies et dont la stylométrie dans le cadre des enquêtes de police judiciaire.

Nous pouvons écouter notre premier intervenant : Denis Jacopini. »

## Denis Jacopini, Expert de Justice en informatique – Cybercriminalité & Protection des données



*Denis Jacopini est Expert de Justice assermenté, spécialisé dans l'analyse des infractions numériques, la traçabilité informatique et la conformité en matière de protection des données personnelles. Après une formation technique en ingénierie électrique, il s'est orienté vers l'investigation numérique, le droit de l'expertise judiciaire et la gestion des risques des systèmes d'information, développant ainsi une compétence reconnue dans l'établissement de preuves techniques exploitables devant les tribunaux.*

*Depuis plus de vingt-cinq ans, il intervient dans des missions d'expertise pour les juridictions, où il analyse fraudes, intrusions, compromissions de systèmes et litiges liés au RGPD. Il est également sollicité avant procès pour réaliser des expertises privées, assister les avocats ou éclairer les magistrats et les parties sur les implications techniques d'un dossier. Ancien Correspondant Informatique et Libertés (CIL) et Délégué à la Protection des Données (DPO), il maîtrise les enjeux réglementaires liés à la preuve numérique, à la licéité des traitements et à la responsabilité des acteurs. En parallèle de ses missions judiciaires, il contribue à la diffusion d'une culture juridique de la cybersécurité à travers ses formations professionnelles, son blog spécialisé et plusieurs travaux de référence.*

« Bonsoir à toutes et à tous.

Comme il l'a été présenté, je suis expert de justice en informatique. Mais au départ, je suis avant tout un informaticien. En fait, on devient d'abord technicien et puis après on prend une spécialité en matière judiciaire. Et mes deux grandes spécialités du coup, parce qu'il faut être passionné en informatique de toute façon pour sortir un petit peu du lot, et bien c'est la cybercriminalité et la protection des données. Et dans ce domaine, on n'est pas beaucoup en France à avoir les deux compétences à la fois technique et à la fois juridique, puisque la protection des données, c'est avant tout des règles juridiques que l'on doit respecter.

C'est ce qui fait que je suis souvent contacté et appelé pour présenter et vulgariser ces sujets qui parfois sont intimement liés. Alors la présentation que je vais vous faire, c'est une présentation quand même assez générale de l'IA et assez générale de la manière dont l'IA peut être utilisée, à la fois dans mon métier, mais également dans le cadre de certaines étapes judiciaires.

Donc, j'espère que je vais vous apporter des connaissances que vous allez découvrir et apprécier. De toute façon, le but c'est de vous apprendre peut-être certaines choses. S'il y a des réactions, c'est avec plaisir parce que je pense qu'il va y avoir des réactions à la suite de nos interventions respectives.

On va commencer par tout simplement l'IA Depuis lors ? Pour vous, depuis quand l'IA existe-t-elle ? Depuis quand l'intelligence artificielle a-t-elle commencé à faire parler d'elle ?

Une personne répond dans le public : Dans les années 50 ? Dans les années 50 ?

Bravo ! Donc, on sent tout de suite une personne qui s'est renseignée sur l'IA. Effectivement, dans les années 50, les premiers travaux de chercheurs consistaient à véritablement essayer de reproduire le fonctionnement du cerveau humain et le faire reproduire par des ordinateurs. Donc les premiers travaux, c'était surtout au travers d'un jeu de dames ou d'un jeu d'échecs.

Mais à la suite de ça, bien d'autres recherches ont continué et d'autres travaux ont vu le jour qui n'ont pas forcément été médiatisés. En fait, il y a ce que j'appelle des paliers, c'est à dire des moments où on montre au monde un nouveau challenge, une étape qui est franchie et l'étape qui a été franchie ensuite. Le palier que j'ai vraiment remarqué car il sortait un petit peu du lot, c'est en 97, lorsque le champion du monde d'échecs Kasparov s'est fait battre par une machine.

Et là on était dans un algorithme qui avait quand même encore une certaine lenteur, certaines difficultés. Donc les travaux ont continué pour encore améliorer cette machine. Et en 2017, il y a eu un autre palier. Et ce palier, je ne sais pas si vous le connaissez, :

lorsqu'il y a eu une machine, encore de Google parce que la première c'était déjà l'ancienne machine d'IBM, (Big Blue en 97), là on est en 2017, et bien c'est le jeu de go qui a été pris à partie par une intelligence artificielle. Et donc l'IA a battu le champion du monde, quelqu'un qui était censé ne jamais être battu, et bien il s'est fait battre de manière lamentable. Malheureusement, ça l'a découragé puisque je crois qu'il a arrêté de faire des concours, il me semble.

Et puis il y a un autre palier, le dernier vraiment qui a fait beaucoup de bruit. C'est en novembre 2022. Je pense que si je vous dis ChatGPT, forcément vous savez de quoi on parle. Et donc c'est véritablement pour moi le début d'une nouvelle ère, puisque on est dans quelque chose qui a montré au monde qu'on avait un outil, On avait une entité qui était capable de nous faire penser à un être humain en fait. Et d'ailleurs, il y a un test qui s'appelle le test de Turing, pour ceux qui connaissent. En fait, c'est un test qui consiste à faire croire à un être humain que la machine n'est pas une machine, mais qu'elle est un être humain. Et ce test, eh bien Chat GPT l'a passé haut la main et sans aucune, aucune difficulté.

Alors pour moi, l'IA est constituée des éléments de base selon trois niveaux. Il y a un premier niveau qui correspond à un algorithme tout simple. Si certains connaissent, on pourrait comparer ça à des systèmes experts. Si je fais ça, on fait ça derrière une télécommande par exemple. C'est un système basique, mais il y a un algorithme derrière une télécommande. Derrière une calculatrice, il y a un algorithme ChatGPT. Il a aussi un algorithme, une voiture autonome. Elle a aussi un algorithme au départ. Mais il y a une deuxième fonctionnalité qui est vraiment importante pour se différencier d'un algorithme qu'est l'IA, et bien c'est de pouvoir traiter un très grand nombre d'informations. Et c'était quelque chose qui était très difficile avant parce que les ordinateurs étaient lents, parce qu'il n'y avait pas suffisamment de capacité mémoire.

Et dans les années 2000, en fait, on a vraiment commencé à avoir des ordinateurs qui calculaient de plus en plus vite. La mémoire n'était plus trop un problème. Et donc du coup, les recherches ont pu profiter de toutes ces avancées et pouvoir traiter énormément d'informations. Donc là tout de suite, on élimine la calculatrice et la télécommande. En fait, tout de suite, là, on voit des systèmes qui sont capables de pouvoir traiter des

millions, voire des milliards d'images différentes et des milliards d'informations différentes.

Et là, ChatGPT rentre dans son mode de fonctionnement, la voiture autonome aussi, puisqu'une voiture autonome est censée analyser beaucoup de choses qui se passent qu'elle est capable de voir qu'elle est capable de capter, qu'elle est capable de recevoir. Et donc c'est cet ensemble d'informations qui fait que ce n'est plus un simple algorithme, c'est un algorithme, mais super perfectionné. Et puis il y a un troisième paramètre qui pour moi fait vraiment la différence, c'est lorsque les données qui ont été utilisées pour sortir un résultat sont mises à jour et sont adaptées en fonction du résultat qu'on obtient.

C'est à dire ChatGPT: il utilise ce qu'on écrit, et il va pouvoir le réutiliser. D'ailleurs, faites attention à ça parce que selon ce que vous mettez comme information, selon les informations que vous lui communiquez, et bien le risque est de se retrouver dans sa mémoire et donc accessible par d'autres personnes. Donc il faut faire vraiment très attention à ça. Une voiture autonome aussi, elle est capable d'utiliser les informations qui lui permettent de modifier et de perfectionner son algorithme. Quoique on a quand même quelques difficultés parce que la plupart des systèmes qui évoluent vont surtout progresser grâce à des notions d'échec. Par exemple comment reconnaître un chat ? Et bien on va lui proposer des photos d'animaux, chaque fois qu'il dit c'est un chat et qu'il a raison, on dit "bonne réponse" ! Chaque fois qu'il dit que c'est un chat et qu'il a tort, on lui dit "Mauvaise réponse" ! Et d'ailleurs, ça a nécessité beaucoup d'informations de la part des utilisateurs pour alimenter ces bases de données. Les captchas. Vous savez que lorsque vous devez montrer que vous êtes un humain, ça a permis notamment à Google de recueillir plein d'informations pour reconnaître des chats, reconnaître des voitures, reconnaître des ponts. Enfin, je pense que vous allez tout de suite vous remémorer les petites images sur lesquelles on clique pour arriver à accéder à l'information que l'on veut. Et bien en fait, on n'a fait qu'alimenter des IA pour lui apprendre ce que c'était un pont, ce que c'était une voiture, ce que c'était un vélo. Mais en fait ça c'est parfaitement possible.

L'entraînement est parfaitement possible pour ChatGPT, mais pour une voiture, on ne peut pas lui faire faire des millions ou des milliards d'accidents pour dire : « Ah là, ce

*n'était pas bien ce que tu as fait.*" Donc en fait, c'est une autre problématique que l'on a : c'est un algorithme qui, au départ, doit être le plus parfait possible et qui vient se perfectionner avec parfois quelques petits problèmes. D'où quelquefois des accidents qui n'étaient absolument pas prévisibles parce que l'algorithme n'avait pas prévu que, par exemple, il allait y avoir une longue remorque qui allait traverser la route. Et puis entre le camion et la remorque, qui était très, très longue, il y avait juste un plateau. Et pour la voiture, ce n'était pas du tout une remorque parce que pour elle, la voiture, une remorque, ça a une autre forme qu'un simple plateau horizontal.

Alors en tant qu'expert de justice, l'IA on a eu à l'utiliser, cet outil, mais pas en tant que tel. En fait, très souvent, ce sont des fonctionnalités qui sont rajoutées à des outils qui nous permettent de faire des tris de données. Lorsque on travaille sur des affaires pénales, on va nous communiquer des supports. Les supports peuvent être très variés : des téléphones, des ordinateurs, des tablettes, des box, des télévisions, des GPS de voiture — en réalité, tout support capable de contenir des données numériques. Lorsque le juge, ou toute personne en charge du dossier, estime que des informations utiles à l'enquête se trouvent à l'intérieur de ces supports, il nous demande de les analyser. Nous pouvons alors nous retrouver face à des millions d'images ou à des volumes immenses de SMS à examiner. Dans ces situations, nous n'avons évidemment pas la capacité, en une seule soirée, de voir, lire et traiter l'ensemble de ces données pour produire un résultat rapide. Et là, on se rend compte que les notions de volume et de rapidité que peut apporter l'IA va prendre tout son sens puisque, on peut nous demander, par exemple, de reconnaître une ou des personnes en particulier. C'est comme la reconnaissance faciale, mais appliquée ici à des photos ou à des vidéos. Donc, à part regarder chaque vidéo du début à la fin, un humain n'a pas d'autre solution : il ne peut pas faire autrement, il va donc se faire aider par des outils qui l'aident à reconnaître les visages.

On me demande aussi, parfois, dans des affaires de pédophilie, d'estimer l'âge des personnes. Est-ce que la personne est mineure ou non ? Parfois, c'est compliqué, et parfois on se trompe. Et parfois, l'IA aussi se trompe. On va également me demander, en matière de pédophilie, de ressortir les photos de personnes mineures nues. Et si je devais toutes me les taper une par une, ce ne serait pas très agréable. Heureusement, il y a l'IA pour faire ce tri. Reconnaître des photos sur lesquelles il y a de la cocaïne ou du cannabis.

Reconnaître des objets : j'ai travaillé sur des trafics. Des trafics d'œuvres d'art aussi. Par exemple, certaines photos font apparaître un paysage. Eh bien, plutôt que de me taper toutes les photos du monde qui existent sur Google ou sur d'autres moteurs de recherche, c'est mieux si l'IA est capable de comparer ces photos de paysage avec toutes les photos qu'elle peut récupérer sur Internet pour retrouver une image qui corresponde à un paysage.

Cela permet alors de localiser une photo et de faire le lien avec les données de géolocalisation présentes dans les appareils, par exemple. En fait, on se sert de cet outil pour faire des pré-tri, plutôt que nous, de chercher à la main. Eh bien, on va utiliser un outil qui est capable de le faire pour nous. Et en fait, c'est un outil qui va le faire de mieux en mieux puisque, arrivé à un moment, on va lui dire : "Non, ça, c'était des faux positifs ; ça, c'était des faux négatifs.". Et c'est comme ça que, du coup, l'algorithme s'améliore, y compris en matière judiciaire.

Ces outils peuvent être utilisés pour faire du croisement d'informations. Il y a beaucoup d'affaires qui ont pu bénéficier justement de l'IA, parce qu'il y a quelque chose de complètement ahurissant : ce n'est pas "naturel", dans les systèmes informatiques existants, qu'un événement survenu à un endroit en France soit automatiquement rapproché d'un autre événement similaire survenu ailleurs en France. Ce rapprochement ne se fait pas tout seul. Il a donc fallu développer un outil juste pour ça, pour faire des rapprochements. Et avec ces rapprochements, on est capable de considérer que si une personne a escroqué une personne à un endroit et puis qu'elle a escroqué une autre personne à cinquante autres endroits, eh bien, il y a 51 délits et pas un seul. Chaque fois. Et donc, ça permet d'aller beaucoup plus vite, de faire beaucoup moins d'erreurs, d'avoir beaucoup plus d'informations.

**Cependant, il y a vraiment une limite qui, pour moi, est très, très importante : pour traiter tout ça, on fait appel à un outil et il est très important que cet outil reste à sa place. A aucun moment je ne vais prendre ce que l'outil me crache pour le mettre sur mon rapport et puis le donner. Ça, c'est la pire erreur que je puisse faire. En tant qu'experts, nous en tant qu'experts, on a l'avantage d'être suivi quasiment dans cent pour cent des cas, c'est à dire que ce qu'on va écrire dans notre rapport, le juge va le**

**prendre pour argent comptant. Donc, on ne peut pas se permettre de mettre quelque chose de faux ou quelque chose d'approximatif. Ça existe et c'est précis, ou alors on n'en fait pas mention. Et lorsque l'IA nous donne quelque chose, ça devrait être un peu sous sa responsabilité. Or l'IA n'est responsable de rien. Il n'y en a qu'un qui est responsable, c'est celui qui va remettre, le document.**

Et c'est pareil pour l'avocat quand il va utiliser l'IA pour sa plaidoirie ou lorsqu'il va remettre son dossier, eh bien il est responsable de ce qu'il écrit.

Et donc j'ai toujours dit, et je continuerai toujours à le dire, je ne sais pas combien de temps, que l'IA est un, un ou une, un assistant ou une assistante stagiaire, parce qu'elle fait un travail colossal pendant qu'on dort, pendant qu'on boit le café, pendant qu'on fait autre chose, eh bien, elle fait un travail de fourmi qui est colossal. Traiter des milliers et des milliers, des millions et des milliards d'informations. Elle est capable de le faire sans broncher, sans dire je suis fatiguée, sans dire j'ai mal, sans dire, les congés, les vacances ? Non. Stagiaire parce que le stagiaire, il n'est responsable de rien. Donc voilà la place que doit avoir l'IA à mon avis encore pendant longtemps. Et on doit l'utiliser comme ça et à mon avis pas plus.

Enfin, l'IA et notamment ChatGPT ou la plupart des outils de langage sont utilisés en matière judiciaire pour avoir des informations sur le mis en cause lors d'une audition. Il y a plusieurs manières qui sont utilisées par les enquêteurs pour obtenir des informations sur la personne qui est mise en examen ou en cause. Jusqu'à maintenant, la personne avait un téléphone, on lui prenait son téléphone, on regardait ce qu'il y avait à l'intérieur. Maintenant, avec l'IA que peut avoir la personne sur son téléphone, et bien on prend son téléphone et on va dans l'IA. Alors, que ce soit dans le navigateur ou directement dans les applications que peuvent installer les gens, parce que c'est un jeu, même pour les cybercriminels, c'est devenu un jeu maintenant d'utiliser ça, ils peuvent poser n'importe quelle question. Pour eux, ça les rassure peut-être et aussi d'avoir une autre vision ou des réponses à certaines questions.

Mais ces historiques sont dans les appareils et donc l'enquêteur peut avoir la trace des questions que le propriétaire de l'appareil a posée sur ChatGPT. Si elles ont un rapport avec les faits qui lui sont reprochés, eh bien ça constitue une preuve solide. Et si la personne a malencontreusement oublié le mot de passe qui permet d'accéder à son application, il y a toujours la possibilité de contacter OpenAI ou de contacter directement les éditeurs de ces IA pour leur demander, comme les opérateurs téléphoniques pour leur demander par réquisition en fait, les informations qui sont stockées du côté de ces personnes-là. Et on va retrouver les mêmes informations, c'est à dire l'historique des recherches, l'historique des connexions, on peut retrouver tout ça. Donc là il y a une règle à respecter, c'est faites attention à ce que vous recherchez et ne communiquez jamais à une IA ce que vous ne souhaitez pas rendre public parce que vous risquez de du coup de rendre publiques ces informations.

Je termine en fait par un projet de loi. Je ne sais pas si vous l'avez entendu, ce projet de loi.

Chat Control, ça vous dit quelque chose ?

Ceux qui savent que l'IA date des années cinquante doivent peut-être le rappeler. Là, on parle d'un projet qui consiste à obliger tous les opérateurs de messagerie à donner les clés pour déchiffrer les messages à la volée. Autrement dit, ça fait des années qu'on nous bassine avec la protection individuelle, la protection de nos données, la protection de nos libertés...

Et là, ce projet de loi, c'est exactement l'inverse : on voudrait nous obliger à n'utiliser que des outils, des logiciels, des applications qui permettraient aux forces de l'ordre — et probablement à d'autres entités dont on ne connaît pas encore l'identité — d'accéder au contenu de nos messages, au contenu des informations qui circulent sur les réseaux.

Alors oui, aujourd'hui ça existe déjà, mais c'est quand même très encadré. De nos jours, si vous travaillez dans une grande société, l'inconscient collectif, c'est que le chef d'entreprise de cette grande société ne peut pas laisser passer n'importe quoi sur le réseau. En France, il y a le Code de la propriété intellectuelle, qui rend responsable — et

là je vais un peu vers le droit, désolé — l'abonné à Internet de ce qui est fait avec sa connexion.

Donc, cela veut dire que si quelqu'un consulte des sites illégaux ou véhicule des informations illégales depuis la connexion Internet de l'entreprise, c'est probablement le gérant de l'entreprise, ou la personne responsable, qui va se retrouver pénalemen responsables. Donc, il a tout intérêt à canaliser ce qui se passe sur son réseau. Il ne peut pas se permettre de laisser faire n'importe quoi comme ça, sans contrôle, sans limite.

Il existe donc des outils qui permettent de faire du déchiffrement à la volée, du HTTPS et de toutes les connexions sécurisées que l'on peut avoir, justement pour s'assurer qu'il n'y ait pas d'actions illégales. Ensuite, il y a toute la question de l'usage : l'usage de ce qui est capté, et l'usage de ce qui est déchiffré. Et là, il y a des limites qui ne doivent pas être dépassées.

Or, ce projet de loi, lui, nous fait passer du statut de présumé innocent à celui de potentiellement coupable jusqu'à preuve du contraire, puisque nous serons tous surveillés comme si nous étions déjà présumés coupables. Et s'il n'y a pas de preuve de culpabilité, alors ça va, on est tranquille.

Autant vous dire que c'est un projet contre lequel beaucoup de professionnels se mobilisent, parce que cela va complètement à l'encontre de tout ce que nous avons construit depuis des années pour protéger nos libertés individuelles. Et je suis, moi aussi, un fervent défenseur de ces libertés.

Voilà, je voulais terminer par cela. Et merci, j'espère ne pas vous avoir fait trop peur. »

## Visouda Luangkhot – Experte en Intelligence Artificielle



*Experte en formation digitale, Visouda Luangkhot s'est fait connaître par sa capacité à transformer des outils numériques complexes en dispositifs pédagogiques accessibles et efficaces. Son parcours est marqué par une maîtrise rare du lien entre technologie, apprentissage et innovation, ce qui lui permet d'accompagner la société dans des changements décisifs.*

*Au fil de son parcours, Visouda Luangkhot a mené plusieurs projets structurants qui ont contribué à diffuser une culture du digital accessible et créative. Elle a notamment initié l'un des premiers podcasts pédagogiques à Strasbourg, projet pionnier visant à rendre les enjeux du numérique compréhensibles pour un public large tout en valorisant les acteurs locaux de la formation. Elle a également accompagné de nombreux organismes dans la refonte de leurs dispositifs digitaux, conçu des modules interactifs centrés sur l'apprentissage actif, et piloté des démarches d'intégration de l'IA dans des parcours de formation existants. Son regard sur les usages numériques dans le champ judiciaire est devenu fondamental, car il éclaire avec précision les enjeux d'éthique, de transparence et de responsabilité qui accompagnent désormais toute décision fondée, même partiellement, sur des outils digitaux ou des systèmes d'IA.*

« Bonsoir à tous et à toutes, je m'appelle Visouda, j'ai 38 ans, je suis la fondatrice d'Eval formation que j'ai cofondé avec mon frère Vinat. J'ai treize ans d'expertise dans la transition numérique où j'ai accompagné des grands groupes et depuis 2023, avec Eval formation, on a intégré l'IA. Notre mission c'est d'intégrer tout ce qui est outils numériques et IA de façon éthique et responsable, sans remplacer l'humain, mais justement en le sublimant et en l'augmentant.

Je suis aussi formatrice et conférencière et mon frère est l'ancien directeur marketing de la marque Rhinoshield. Ça ne vous parle peut-être pas, mais chez nos enfants en tout cas, ça parle.

Et mon rôle ce soir, c'est de vous expliquer comment on peut donc, à la suite de ce que mon collègue à gauche vient d'expliquer, comment concrètement on peut utiliser l'IA

pour peut-être solutionner ou en tous cas pour augmenter le travail dans les enquêtes criminelles, et donc comment on influence.

Alors, petite répétition ou en tout cas rappel, moi je considère qu'on a trois familles d'IA. La première c'est celle des IA Génératives dont on a parlé tout à l'heure, la deuxième, on va être plutôt sur des IA analytiques, on en a aussi parlé tout à l'heure, et on a une nouvelle ère qui est l'IA Agentique.

Est-ce que ça parle à tout le monde dans la salle ?

Alors ceux qui ne connaissent pas l'IA agentique, je vous invite à lire ce livre, à vous le procurer si ça vous intéresse, c'est une mine d'or, écrit par une femme qui travaille à la Silicon Valley.

Elle a créé et présenté devant les développeurs de chez Google, en live, un agent IA qui a créé d'autres agents IA là où ces hommes développeurs n'arrivaient toujours pas à le faire. Donc je vous conseille pour toutes les femmes qui ont envie de se lancer dans la tech, de lire ce livre.

Mais encore un tout petit rappel avant de vous présenter un peu les cas sur lesquels j'ai travaillé. Quand on travaille sur une intelligence artificielle, que ce soit de la générative ou de l'analytique, on a toujours un peu ce principe-là. D'abord, on a un point d'entrée qu'on appelle chez nous les geeks IA, une base de connaissances, on n'est pas sur une base documentaire. Donc ce n'est pas parce que vous envoyez du texte, des images, des documents que l'IA va le digérer comme ça. Cela veut dire qu'elle va prendre vos documents, vos connaissances, les données et va les interpréter dans son langage. Ensuite, on a un programme qui va réfléchir, vous l'avez dit tout à l'heure, on est un peu sur le système : "si jamais", "alors", donc on est quand même sur de l'analyse, en tout cas de la réflexion algorithmiques. Et enfin, à la sortie on va avoir l'interaction avec l'intelligence artificielle sous plusieurs formats. Par exemple, vous avez tous connu au démarrage les chatbots. Aujourd'hui, on peut aussi parler à son compte WhatsApp ou à un agent qui est sur son téléphone via une application où on peut chatter, ou on peut peut-être aussi envoyer des informations sur une boîte mail. Donc la sortie aussi a son importance.

Est-ce que jusque-là vous me suivez tous ? Super !

Eh bien, les premiers cas d'usage qu'on peut mettre en place pour résoudre, ou en tout cas pour influencer les enquêtes judiciaires, c'est la création de ce qu'on appelle des assistants IA. Est-ce que ça parle à quelqu'un ? Est-ce que vous connaissez le terme assistant ?

Comme vous avez la capture ici, dans Chat GPT on appelle ça des GPTs et dans Gemini on appelle ça des Gems. Un assistant ce n'est rien d'autre que ChatGPT ou Gemini qu'on va personnaliser pour qu'il arrive à nous aider dans une mission, dans une tâche très précise. On en a parlé tout à l'heure avant de commencer, par exemple, il pourrait faire l'assistant de décryptage de plusieurs dossiers suite à l'interrogatoire de quelqu'un par exemple. On va toujours lui assigner le rôle de trouver des indices ou de trouver des particularités spécifiques à un usage.

Donc, on pourrait imaginer, qu'on soit avocat ou qu'on soit enquêteur, d'avoir des mini assistants qui vont travailler sur des tâches les plus spécifiques possibles. On ne va pas avoir un assistant, par exemple, qui va juste analyser les documents et faire un résumé, car c'est trop générique. On va plutôt lui assigner un rôle bien spécifique, par exemple de relever combien de fois la personne a dit, a prononcé un mot par exemple, car peut être que ce mot-là est une source pour les enquêteurs pour indiquer qu'il est anxiogène ou qu'il a menti, ou qu'il essaye de dissimuler une preuve par exemple.

Ces assistants sur ChatGPT sont payants et sur Gemini sont gratuits. Ce qui veut dire qu'aussi, vous, pouvez aussi vous pratiquer et monter en compétence sur l'usage d'assistant.

Le deuxième cas, est celui des outils analytiques, vous en avez parlé tout à l'heure. Un outil analytique du marché qui est aussi gratuit, c'est Notebook LM qui vient de chez Google.

Le principe est le suivant : Vous mettez une base documentaire, vous l'alimentez, on peut aller jusqu'à 3 millions de caractères et 300 documents en même temps, donc pour les affaires juridiques, c'est quand même super intéressant. On peut alimenter Notebook LM

avec des documents qu'on est allé chercher sur une base. Au milieu vous avez la même chose qu'un ChatGPT, la même chose qu'un Gemini, donc un LLM avec lequel vous pouvez interagir, poser des questions, lui demander de faire une synthèse, lui demander d'aller chercher une information précise, comme l'heure à laquelle la personne est entrée dans la pièce ou des informations de cet ordre. On peut donc lui poser des questions très précises, et il va nous restituer l'information qu'on a demandée sous forme de fiche, sous forme d'une vidéo, sous une forme d'un audio podcast comme si quelqu'un interagissait, avait une conversation. Donc, vous avez la possibilité de choisir comment vous voulez qu'on restitue votre information. Autrement dit, selon l'enquêteur par exemple, ou selon peut-être le public que l'on va avoir en face de nous quand on va auditionner, on peut choisir un format qui va s'adapter aussi bien à l'usager qu'à la personne qui va écouter en face. Typiquement, quand je prends le cas de policiers par exemple, ça peut être très intéressant d'utiliser le format vidéo quand on est dans une réunion rapide, parce que Notebook LM va vous faire un résumé de 5 à 7 minutes de vidéo sur la question que vous lui avez posé. Je ne sais pas si j'ai le temps de montrer quelque chose, mais si jamais vous avez des questions, je pourrais le montrer sur mon Espace Notebook LM.

Nous avons parlé tout à l'heure de ces outils qui vont pouvoir directement analyser les réseaux sociaux. Typiquement, les outils comme vous avez ici à l'écran, et qu'on appelle des SOCMINT, sont des outils qui sont déjà mis en place à travers le monde, que les systèmes français peuvent utiliser et qui font des analyses directement sur les profils des réseaux sociaux ou sur Internet. Ils sont adaptés pour aller chercher directement sur la toile et chacun a des rôles spécifiques. Par exemple, il y en a un qui va aller chercher des images, il y en a un autre qui va aller chercher le nombre de fois où il y a une personne sur son compte Instagram et qui dit un mot interdit. C'est ce qui permet aussi à des enquêteurs, sur les mêmes bases que ces outils, d'aller chercher les personnes qui sont fichées S par exemple.

En troisième cas, il y a ce qu'on appelle les agents IA, et donc l'ère de l'agentique dont je vous ai parlé. Un agent, contrairement à un assistant, est une machine qui va agir, donc il va faire une action (il ou elle d'ailleurs). Par exemple, un agent a la capacité aujourd'hui d'envoyer un email si on lui ordonne de l'envoyer, il peut faire de la veille, il peut aller chercher des informations. Par exemple, quand on est avocat et qu'on veut faire sa veille

juridique, on peut imaginer de mettre en place un agent qui va faire votre veille juridique. Tous les lundis, dans votre boîte email, vous auriez un résumé des derniers articles dans votre domaine d'activité, etc. Je vous le dis parce que pour nous, par exemple, c'est très compliqué de faire de la veille IA, les choses vont très vite évidemment, et je le reconnaît, j'ai un agent IA qui fait entre autres la veille outil pour moi.

Si vous combinez maintenant les agents avec les IA analytiques, vous comprenez bien que ça fait un combo gagnant. Vous avez la possibilité de demander à un agent, un outil qui existe, d'analyser ce qui se passe, et ensuite vous pouvez greffer un agent qui lui, agira selon ce que vous souhaitez. Par exemple, même si ce n'est pas conseillé en France puisque ce n'est pas encore réglementé mais c'est quand même un cas dont je voulais vous parler, on peut demander à un agent d'appeler quelqu'un.

Tout en gardant en tête le côté éthique et responsable qui appelle à faire attention, on peut imaginer, quand on fait du profilage, quand on doit être sous couverture, de synthétiser des voix, comme nous l'avons évoqué, et ensuite faire en sorte que l'agent appelle quelqu'un pour aller chercher des informations par exemple.

On peut aussi imaginer des situations où un agent numérique crée un faux compte sur les réseaux sociaux, comme cela se fait aujourd'hui pour infiltrer ou repérer des pédophiles. La différence, c'est que ce ne serait plus un enquêteur qui fabrique et anime ce compte, mais un agent automatisé. L'humain, lui, ne disparaît pas : il se recentre sur les tâches à forte valeur ajoutée. Dès lors que l'on automatise les actions longues, répétitives et peu qualitatives, l'enquêteur gagne du temps pour faire ce que seul un humain peut réellement faire : chercher des informations fines, contacter des personnes, analyser les situations, et exercer son discernement.

On observe déjà ce phénomène dans d'autres métiers. Par exemple, lorsqu'une partie de la prospection commerciale est automatisée, les commerciaux peuvent consacrer davantage de temps à la relation humaine, au suivi, à l'écoute, à l'accompagnement. L'automatisation ne remplace donc pas l'humain : elle lui rend de l'espace pour faire de l'humain.

Dans le domaine des enquêtes, les agents peuvent jouer exactement ce rôle. Ils permettent d'automatiser des tâches chronophages, comme l'envoi d'e-mails, la publication de contenus, ou même la création rapide d'un site internet lorsqu'il faut simuler une situation. À l'inverse, ils peuvent aussi analyser systématiquement des sites existants. En combinant un agent avec un outil analytique, on peut imaginer qu'un enquêteur puisse, chaque lundi par exemple, faire analyser automatiquement une centaine de sites afin d'y repérer des signaux de fraude, des comportements suspects ou d'autres éléments sensibles.

L'objectif n'est pas de remplacer l'expertise humaine, mais de fournir une information prétriée, structurée et déjà analysée, pour que l'enquêteur puisse se concentrer sur ce qui nécessite réellement son jugement, sa sensibilité et sa capacité d'interprétation.

Je souhaite attirer votre attention sur un point important : même si la création d'un agent est aujourd'hui accessible au grand public, cela reste un véritable programme. L'interface donne l'impression que c'est simple, mais en réalité, on construit un enchaînement d'instructions, avec des paramètres précis qui définissent son comportement. Pour comprendre la logique : un LLM, comme ChatGPT, Gemini, Claude ou Mistral, joue le rôle du « cerveau ». Il réfléchit, traite la requête d'entrée et génère une réponse. L'agent, lui, est comme un corps qui ajoute des « membres » : on définit la taille du bras, la longueur de la jambe, la puissance du saut, selon la tâche que l'on veut lui confier. Un agent conçu pour simuler un athlète ne sera pas configuré de la même manière qu'un agent conçu pour jouer au football. Tout dépend de l'objectif.

C'est pour cela qu'on a toujours besoin d'un LLM pour faire fonctionner un agent : le modèle fait la partie cognitive, l'agent organise l'action. Cependant, même si l'on peut construire de petits agents spécialisés, il est important d'être prudent : les résultats ne sont pas toujours parfaitement fiables. Les agents restent des programmes, avec leurs limites, leurs imprécisions, et parfois des comportements inattendus. C'est pourquoi leur utilisation doit être systématiquement vérifiée, testée et validée par un humain. L'agent peut automatiser une tâche, mais c'est l'humain qui garantit la qualité, la pertinence et la sécurité du résultat.

Tout à l'heure, on parlait d'IA vocale. Il est en effet possible aujourd'hui de synthétiser sa voix. Avec Sunos ou Eleven Labs qui sont des outils ouverts avec un abonnement qui est peu cher, on peut synthétiser nos voix. Les risques et les dérives qui s'en suivent ne sont pas négligeables. Des outils comme ceux-ci ne répondent pas à la réglementation française. Ils synthétisent par exemple les voix de Monsieur Macron, de Monsieur Sarkozy, de Monsieur Hollande et les vendent. Les travaux de réglementations sont néanmoins en cours. Les dernières informations auxquelles j'ai eu accès, tout en gardant en tête que je ne suis pas spécialiste du droit, indiquaient que lorsque vous produisez une voix, un son, ou autre, avec de l'intelligence artificielle, ce n'est pas quelque chose qui vous appartient, cela n'appartient à personne puisque c'est créé avec une intelligence artificielle.

Si vous créez une chanson et que vous déposez les paroles, les paroles vous appartiennent mais la question du droit d'auteur en musique reste un sujet plus complexe.

L'idée est de vous informer et sensibiliser aux dérives des IA vocales. Ces contenus synthétiques soulèvent la question de la capacité à distinguer les voix originales des voix générées par l'IA. Toutefois, et tout comme en cybersécurité les efforts sont déployés pour contrer la prolifération des attaques envers les systèmes d'information, ou encore comme dans le domaine de la santé les recherches s'activent pour guérir d'un virus dangereux, les travaux scientifiques et les réglementations, veillent à imposer des pratiques qui permettent de contrer ces dérives d'usage.

Concrètement, des métiers d'authentification de contenus seront de plus en plus demandés, notamment dans le cadre des enquêtes criminelles.

Il faut aussi voir le rôle des outils de l'IA comme une augmentation des capacités de traitement d'information dans le cadre des enquêtes. Nous avons en tête l'image de l'expert informatique qui, dans le cadre d'une enquête, supporte les équipes terrain en fournissant des données supplémentaires. Les outils d'analyse de donnée par l'IA seront son extension dans la réalisation de cette tâche. La présence de l'humain restera primordiale, car comme je l'ai évoqué, les résultats des machines ne sont pas exempts d'erreur. La machine n'a pas la responsabilité de valider le numéro de sécurité sociale de

telle ou telle personne. L'humain garde la responsabilité de la vérification et de la mise en place les procédures adéquates.

A titre d'exemple, chez OpenAI, des employés ont pour rôle la vérification des informations ingérées par l'outil tout au long de sa durée de vie. Si j'envoie via l'interface Chat GPT la phrase suivante: « je m'appelle Visouda, je suis un homme » et que je le répète 20 ou 30 fois, et qu'ensuite un utilisateur de ChatGPT demande: "Est-ce que Visouda est une femme ou un homme ?" , il y a un risque que, sans correction humaine, ChatGPT réponde "C'est un homme".

**Il faut également noter qu'il existe des personnes malveillantes, ce qui représente un point de vigilance important. Certaines de ces personnes sont même rémunérées pour fournir de fausses informations dans les outils, notamment dans ceux que l'on qualifie d'open source.**

Lorsque l'IA deviendra un sujet que les services gouvernementaux souhaiteront intégrer dans leurs pratiques, il faudra penser à créer ces métiers et ces postes pour superviser, vérifier et encadrer son utilisation.

Et pour finir en cinquième cas, on a parlé de la voix dans la vidéo, c'est un sujet sur lequel je voulais également appeler à votre vigilance. Des outils permettent désormais de générer des images et d'animer des personnes, comme des influenceurs virtuels. En France, la réglementation commence à encadrer ces usages, mais il reste possible de créer des avatars IA. Cela soulève la question des dérives, notamment dans le cadre des enquêtes, où il sera très difficile d'identifier si une personne existe réellement ou non. De plus, comme ces avatars peuvent être générés à partir de personnes réelles, certains peuvent ressembler trait pour trait à quelqu'un que vous connaissez.

C'est pourquoi il est essentiel — comme vous l'avez souligné plus tôt — de faire attention lorsque vous mettez votre image dans une intelligence artificielle.

Alors, cela n'a rien à voir directement avec les enquêtes judiciaires, mais c'est un cas important que je voulais partager — et c'est aussi l'une des raisons pour lesquelles je me suis intéressée à l'IA.

J'étais dans une salle de théâtre à Paris, et j'ai vu une jeune fille qui se prenait en photo. Avec une application, elle a supprimé tout l'arrière-plan — dont moi, puisque j'étais derrière elle. Ses parents, émerveillés, l'ont applaudie en disant : « Ma fille, t'es trop forte, comment tu as fait ça ? ». Mais avec mon regard de connaisseuse en intelligence artificielle, je n'ai pas vu un exploit : j'ai vu un danger. Cette jeune fille venait de donner son image à une IA. Or, on ne sait pas où ces données sont stockées, ni ce qu'elles deviendront. On ne sait même pas si l'application respecte le droit français. On sait simplement qu'elle a donné son image à un système qui pourrait peut-être la réutiliser, et que cela ouvre la porte à des dérives possibles.

Dans les usages quotidiens comme dans les enquêtes, il est donc essentiel de comprendre que nous serons confrontés à ce genre de situations. Il faudra mettre en place des processus d'analyse, pour chercher la véracité des informations sous toutes ses formes, en se fiant à des indices comme : un grain de peau trop lisse, une image trop parfaite, des décors incohérents, des éléments flous et déformés, etc.

Je terminerai en rappelant que l'IA est là pour sublimer l'humain, pas pour le remplacer. C'est un outil, pas un jouet. Or, beaucoup de personnes interagissent avec ChatGPT comme si c'était leur psychologue ou leur ami. Je mets vraiment en garde contre ces dérives : j'enseigne dans des écoles de commerce et de communication, et je vois de nombreux étudiants parler à ces IA comme à des proches. C'est un sujet important, notamment pour la protection des enfants et pour les enquêtes.

Vous l'avez souligné : des personnes analysent également les données que nous envoyons dans les IA. Et c'est un vrai sujet, car à l'avenir, le référencement ne se limitera plus à Google : les données seront aussi référencées dans les modèles de langage comme ChatGPT ou Gemini.

Je conclurai donc par ceci : gardez en tête que l'IA est là pour vous aider, pour vous sublimer — pas pour vous remplacer.

Merci. : »

## Dr Jean-Baptiste Camps, expert stylométrie



*Docteur en études médiévales, maître de conférences à l'École nationale des chartes (Université PSL) et titulaire de la chaire « Humanités numériques », Jean-Baptiste Camps est l'une des figures françaises les plus reconnues dans le développement de la stylométrie et des méthodes computationnelles appliquées aux textes anciens. Formé à la philologie médiévale et aux sciences du traitement automatique des langues, il travaille à l'articulation entre la tradition érudite, édition de manuscrits, critique textuelle, étude des langues d'oïl et d'oc et les approches quantitatives modernes.*

*Ses recherches portent sur la manière dont les signatures stylistiques, les régularités lexicales et les structures internes des textes permettent d'éclairer des corpus complexes ou disputés. C'est dans ce cadre qu'il a co-dirigé l'étude stylométrique qui a relancé de façon décisive le débat sur l'attribution des œuvres de Molière. En combinant modèles statistiques, analyses multivariées et comparaison systématique avec les contemporains, ce travail a mis en évidence la cohérence interne du corpus molièresque, apportant un argument scientifique fort contre les hypothèses d'auteurs alternatifs.*

*Il est également co-auteur de l'ouvrage *Affaires de style : du cas Molière à l'affaire Grégoire*, la stylométrie mène l'enquête, qui expose au grand public et aux professionnels du texte la puissance des méthodes quantitatives pour résoudre des énigmes littéraires, historiques ou judiciaires. Par ailleurs, il contribue activement à la formation des chercheurs et des praticiens des humanités numériques, à travers ses enseignements, ses projets collaboratifs et la diffusion d'outils méthodologiques devenus des références dans le domaine.*

« Je vais vous parler, pour ma part, de stylométrie.

Qu'est-ce que la stylométrie ? C'est la mesure du style, et je vais vous en dire un peu plus sur ce dont il s'agit. Alors, pour commencer, pour vous donner une intuition de ce que fait la stylométrie, je voudrais qu'on se livre à un petit jeu. Vous avez trois textes, et voici la question : lesquels de ces textes sont les plus proches et pourraient avoir le même auteur ?

C'est tiré d'un cas réel, qui est celui des *Federalist Papers*. Là (*en pointant le premier texte*), en l'occurrence, dans une traduction que j'ai réalisée, vous avez un texte qui vous parle de Carthage, république commerciale qui fut l'agresseur dans la guerre qui aboutit à sa destruction, un deuxième texte qui vous dit que même au milieu de guerre défensive et dangereuse avec la Perse et la Macédoine, les membres n'agissaient jamais de concert, et un troisième texte vous parle d'une incompatibilité de taxation entre le pouvoir des États et celui de l'Union, et ainsi de suite.

Quand on lit ces trois textes en tant qu'humains, évidemment, en général, la proximité que l'on ressent le plus se retrouve peut-être entre le premier et le deuxième texte, qui parlent tous les deux de guerre et qui font allusion à l'Antiquité. Ceci-dit, si l'on se demande qui a écrit ces textes, ces indices-là sont trompeurs. Je peux, moi aussi, me mettre à écrire un texte qui parlerait des guerres antiques.

En réalité, c'est le premier et le troisième texte qui ont le même auteur, et le deuxième texte qui est écrit par un second auteur. Et cela, on le voit non pas en prenant en compte le contenu de ces textes, mais en regardant des indices à une plus petite échelle : des indices peut-être moins perceptibles, que sont les mots outils et les fréquences de termes tels que « un », « sur », « par », « et », et ainsi de suite.

L'exemple que je vous propose est tiré d'un des cas fameux de la stylométrie : celui des *Federalist Papers*, qui sont ces textes écrits à la fin du XVIII<sup>e</sup> siècle en défense de la Constitution américaine, par plusieurs auteurs ayant alors signé sous un pseudonyme collectif, « Publius ». Après la mort de certains parmi les auteurs, d'autres personnes ont revendiqué des textes qui auraient pu, ou non, être écrits par eux.

Et donc, cette question se pose depuis l'époque : comment identifier et attribuer certains de ces *Federalist Papers*. Ce que Mosteller et Wallace ont remarqué en 1963 — ils n'étaient pas les premiers, mais ils ont un petit peu relancé ce domaine —, c'est que si l'on regarde la fréquence de termes tels que *war*, la guerre, on n'arrive absolument pas à distinguer les auteurs. Ce qu'on distingue, ce sont les thèmes de chacun de ces papiers pris indépendamment. En revanche, si l'on regarde des termes tels que *upon, the, for*, et ainsi de suite, là, on arrive à distinguer par les fréquences les différents auteurs et on peut attribuer les textes disputés.

En réalité, la stylométrie, depuis le XIX<sup>e</sup> siècle, s'intéresse à ces marqueurs peu conscients et grammaticaux de la langue et de l'individu que sont les mots outils et d'autres morphèmes grammaticaux, tels que les flexions, les désinences, et ainsi de suite.

La stylométrie, elle, a émergé il y a assez longtemps, pendant les dernières années du XIX<sup>e</sup> siècle, dans un contexte que Carlo Ginzburg a bien retracé : celui de l'émergence de figures telles que Sherlock Holmes en littérature, ou de chercheurs en histoire de l'art tels que Giovanni Morelli, qui s'intéressent au détail, à l'indice comme pouvant être révélateur, plus révélateur que ce qui est immédiatement apparent. Morelli, par exemple, qui a cherché à attribuer les tableaux des maîtres italiens de la Renaissance, s'est rendu compte que si l'on regardait les thèmes principaux de ces tableaux — est-ce que c'est une Annonciation, une Vierge à l'Enfant, que sais-je ? — on n'arrivait pas à distinguer les différents ateliers, les différents peintres.

En revanche, si l'on regardait la manière dont ces peintres traçaient les mains, les oreilles, et ainsi de suite, ce genre de petits détails étaient beaucoup plus révélateurs de traits idiomatiques propres à chaque individu. La stylométrie fait à peu près la même chose, mais sur le style, en cherchant des propriétés qui soient structurelles, fréquentes, faciles à quantifier et relativement immunisées contre un contrôle conscient.

Ces propriétés sont constitutives de ce qu'on appelle parfois "le stylome", et qui est un terme un peu trompeur, qui est calqué sur celui de génome. Le stylome serait donc la signature individuelle dans la langue d'une personne. Les linguistes parlent plutôt d'idiolecte. L'idiolecte, c'est l'ensemble des usages du langage propre à un individu donné. Donc, n'importe quel individu qui maîtrise une langue et qui s'exprime a des traits qui lui sont propres. Le style n'est pas réservé aux grands auteurs, mais chaque individu a des propriétés de langue lorsqu'il s'exprime, qui peuvent être quantifiées et qui peuvent permettre de reconnaître éventuellement ses écrits.

L'idiolecte n'est pas la seule manifestation de variation linguistique qui existe, bien sûr. Dans les textes, la parole varie d'un individu à l'autre, mais elle varie aussi selon la date d'origine des textes, le lieu dans lequel ces textes ont été écrits, la classe ou le niveau social des individus, et ainsi de suite. Donc cette variation s'inscrit dans une série de variations linguistiques concurrentes, mais reste l'un des axes de cette variation linguistique.

Donc, le postulat de base, c'est l'existence, pour chaque individu, d'habitudes linguistiques, de propriétés de langue particulières et mesurables. Alors, on a longtemps fait de la stylométrie sans prendre la peine de vérifier ce postulat, mais il y a eu, dans les dernières décennies, des recherches qui se sont justement penchées sur la vérification empirique de cette hypothèse.

L'expérience de Nimègue, par exemple, qui, dans des conditions très contrôlées, a demandé à des groupes d'étudiants de composer des textes dans un temps limité sur un certain nombre de sujets. Il s'agissait de voir si, à posteriori, il était possible de réattribuer chacun de ces textes à l'étudiant qui l'avait écrit.

Peut-être plus surprenant encore : une enquête menée par des chercheurs du MIT Media Lab en 2015. Ils ont pris quelque chose comme 5000 comptes dans une collection de comptes Twitter et Facebook, et ils ont essayé de voir à quel point ils réussissaient à appairer les comptes de la même personne sur Twitter et sur Facebook.

Donc là, c'est un cas qui est difficile, parce que les textes sont courts, parce que les plateformes sont différentes et parce qu'on a une chance sur 2500, à peu près, si l'on décide l'appariement au hasard, d'avoir juste. Malgré cela, avec leur modèle, ils réussissent dans 30 % des cas à effectuer cet appariement, ce qui est quand même impressionnant. S'ils avaient procédé entièrement au hasard, ils auraient une probabilité de 0,02 % d'avoir juste, donc 30 % c'est déjà assez, assez substantiel.

Je vais passer rapidement sur l'histoire de la stylométrie. Je rappelle que c'est une discipline qui émerge, comme je l'ai déjà dit, dans la deuxième moitié / vers la fin du XIX<sup>e</sup> siècle, et qui se penche d'abord sur des controverses littéraires.

Un des premiers cas que la Stylométrie a envisagé, est la question de la chronologie relative des œuvres de Platon. Platon a produit entre autres des traités philosophiques et des discours qui ne sont pas datés. La question concerne leur chronologie relative : quel texte a été composé avant ou après quel autre.

Jusque-là, on se fiait à l'évolution qu'on supposait être celle des idées de Platon pour dater les textes, et on se servait ensuite de cette datation pour étudier l'évolution de ces idées. Il y avait une forme de circularité qui était un peu problématique et des chercheurs allemands et polonais, notamment Dittenberger et Lutoslawski, ont eu l'idée de regarder non pas les idées, mais les connecteurs logiques et les conjonctions de subordination.

Ils se sont rendu compte que l'évolution de la fréquence de ces termes pouvait permettre de reconstituer une forme de chronologie des œuvres de Platon en général, sans même s'interroger sur ce dont les textes parlaient.

Bien sûr, la grosse évolution depuis l'époque est l'apparition de l'ordinateur à partir des années 50, puisque tous les comptages jusque-là étaient réalisés manuellement. Ils nécessitaient beaucoup de temps pour compter, par exemple, la fréquence de « the » dans l'intégralité des œuvres de Shakespeare. Aujourd'hui, avec un ordinateur, évidemment, cela est instantané et a donc facilité depuis les années 60, l'éclosion de la stylométrie. Pourquoi est-ce que ces mots-outils intéressent des chercheurs ? Certains se sont penchés sur des raisons d'ordre cognitif ou psycholinguistique. (Notamment un ouvrage de James Pennebaker, qui s'appelle *La vie secrète des pronoms*, où il étudie la question de l'emploi des mots-outils et, on va dire, des déterminants, d'ordre psycholinguistique.) Ce qu'il évoque, c'est que l'emploi que l'on fait des termes de contenu — comme garçon, table, salle, etc. — et des mots-outils tels que *par*, *pour*, *le*, et ainsi de suite, ne relève pas des mêmes zones du cerveau ni des mêmes processus cognitifs.

Cela se vérifie notamment dans le cas des aphasies, ces maladies où l'on perd une partie de la compétence linguistique. Il existe deux types d'aphasies principales selon les zones du cerveau qui sont touchées. Dans l'un des cas, les personnes perdent les termes de contenu, mais gardent la prosodie, le rythme de la langue, la construction de la langue. Dans d'autres cas, c'est l'inverse : les malades peuvent garder les termes de contenu, mais ils perdent la capacité de construire des énoncés, de les connecter, et ainsi de suite.

D'un point de vue plus sociolinguistique, cette fois, notre idiolecte et notre particularité de langue peuvent nous révéler en tant qu'individus, mais elles révèlent aussi un certain nombre de déterminants d'ordre social. Des travaux, encore une fois, de Pennebaker et de chercheurs tels qu'Argamon, Kopel et d'autres, sur des corpus de blogs, avaient pour

but de faire du profilage avec de la stylométrie. Il s'agissait de chercher des marqueurs qui peuvent permettre d'en apprendre plus sur le genre d'une personne, femme ou homme, sa classe d'âge, sa langue maternelle, son état de santé mentale, et ainsi de suite. Ces chercheurs retrouvent un certain nombre de traits qui peuvent permettre de réaliser ce type de profilage. Par exemple, dans leur corpus, ils s'aperçoivent que les femmes utilisent plus de pronoms personnels que les hommes, que les hommes utilisent plus de déterminants, et ainsi de suite.

La stylométrie, comme je le disais, a été conçue, et a été beaucoup utilisée pour répondre à des controverses littéraires. Je suis médiéviste et la raison première pour laquelle je me suis intéressé à la stylométrie, est que le Moyen Âge fourmille d'œuvres qui sont tout simplement anonymes ou qui ont plusieurs candidats sans qu'on sache vraiment trancher qui a écrit tel ou tel texte.

On a beaucoup de questions du type : ce document est-il original ? Est-il faux ? Est-il co-écrit ? A-t-il été altéré ?

On a quelque chose comme 95 % des chansons de geste, un genre littéraire français médiéval, qui sont anonymes. On a beaucoup d'œuvres disputées également. Donc la Stylométrie peut permettre de jeter un nouveau regard sur ces documents historiques et de clarifier un certain nombre de choses en termes d'histoire culturelle et d'histoire littéraire.

Plus proche de nous, un domaine qui a donné lieu à beaucoup de productions en stylométrie, c'est celui des grands auteurs classiques de théâtre : Shakespeare pour la langue anglaise, Molière pour la langue française, puisque ces deux auteurs ont été sujets depuis le XIX<sup>e</sup> siècle ou le XX<sup>e</sup> siècle pour Molière, à des controverses, presque des théories du complot, qui remettaient en cause leurs autorités sur les pièces qu'ils ont écrites. Dans le cadre de Molière, c'est Pierre-Louis qui a, en 1919, proposé le premier cette théorie selon laquelle Corneille aurait écrit les œuvres de Molière.

Pour essayer de répondre à cette polémique qui refait surface de temps en temps, une espèce de serpent de mer qu'on voit régulièrement réapparaître, avec mon collègue Florian Cafiero, on a réalisé une analyse de la langue de Molière et de Corneille, mais

aussi d'autres auteurs qui étaient leurs contemporains, comme Rotrou, Scarron, Thomas Corneille — c'est-à-dire le frère de Pierre Corneille — et ainsi de suite. On a étudié des propriétés de langue telles que les rimes, les choix de mots, les affixes, les préfixes et les suffixes, les catégories grammaticales des mots, l'emploi des mots-outils, et à chaque fois, on retombe sur le même résultat, qui est qu'on réussit à reconnaître des groupes d'œuvres distincts.

Donc, vous avez un exemple d'analyse, une classification qui est affichée ici à droite, et les couleurs des feuilles correspondent aux noms qui sont sur la couverture des livres. En fait, le style correspond aux attributions traditionnelles. Vous avez tout à droite le *cluster* des œuvres de Molière, tout à gauche celui des œuvres de Pierre Corneille. Il n'y a aucune confusion, aucun mélange entre les deux ; au contraire, ils sont même plus distincts entre eux que les œuvres de Scarron et celles de Molière, par exemple.

Donc, il n'y a pas de proximité réelle, dès lors qu'on restitue un petit peu de contexte littéraire à ces productions et qu'on étudie un certain nombre de marqueurs de style.

Il y a un cas où il y a peut-être un petit peu de mélange, c'est celui de Psyché, qui est un cas très particulier dont je peux dire un mot en passant. C'est une pièce qui a été co-écrite, mais de manière complètement déclarée par Molière, Corneille et Quinault, une pièce qui répondait à une commande de Louis XIV, qui devait être composée dans un temps très déterminé. Et donc dans l'avertissement au lecteur en tête de l'œuvre publiée, Molière nous dit qui a écrit quoi, il nous dit qu'il n'a pas eu le temps de tout versifier, et qu'il n'a pu écrire les vers que pour le prologue, le premier acte, la première scène du deuxième acte et la première du troisième, et que Corneille a versifié le reste.

Si on fait une analyse de stylométrie, cette fois-ci supervisée. « Supervisé » veut dire qu'on entraîne des modèles de machine learning, on pourrait aussi appeler cela de l'intelligence artificielle, à reconnaître le style de chacun à partir d'un corpus d'œuvres qui, elles, sont certainement attribuées. Donc on entraîne ces modèles, et quand on est satisfait de leurs performances, on peut leur faire parcourir l'œuvre et proposer pour chaque tranche le candidat le plus vraisemblable. Ces courbes en vert correspondant à Molière et celles en rouge à Corneille vous donnent le meilleur candidat pour l'écriture de chacun de ces passages.

Et, ce qu'on voit, c'est que Molière est le meilleur candidat pour le prologue, le premier acte, la première scène de l'acte deux et la première scène de l'acte trois, et que Corneille est le meilleur candidat pour à peu près le reste du texte, à l'exception des petits passages de ballets à la fin des parties chantées qui seraient plutôt de Quinault.

Cette répartition correspond à peu près intégralement à ce que déclarait les auteurs dans l'avertissement au lecteur, avec une petite exception, et qui sont les seuls vers de Molière que je veux bien laisser à Corneille, qui sont les premiers vers. Il est possible que, s'agissant seulement de quelques vers, Corneille ait aidé Molière à boucler ce premier acte sans qu'il ait besoin de le préciser dans l'avertissement au lecteur.

Je propose à présent de sortir des controverses littéraires pour parler d'applications légales ou judiciaires de la stylométrie, qui se font aux États-Unis depuis plus longtemps qu'en France. Un cas que Patrick Juola — qui fait partie de ces stylométriciens qui ont travaillé avec la justice aux États-Unis — rapporte, est celui d'une affaire d'immigration.

Vous avez quelqu'un qui est aux États-Unis, qui demande l'asile politique parce qu'il aurait écrit sous pseudonyme des écrits contre le gouvernement de son pays d'origine, et qui serait menacé de mort s'il était déporté, renvoyé dans son pays d'origine. Simplement, ces écrits sont signés sous pseudonyme. Il faut donc prouver que c'est bien lui qui les a écrits.

Qu'ont fait Juola et ses associés ? Ils ont constitué un corpus de presse avec, à la fois, des articles signés par cette personne de son vrai nom, et des articles d'autres personnes, mais qui ont été publiés dans des revues ou dans des journaux similaires. Ils ont cherché à voir si les articles sous pseudonyme étaient régulièrement plus proches de ceux du candidat que de ceux des autres.

Ce dont ils se sont aperçus, c'est qu'il y a effectivement une proximité plus grande entre le « *Questioned Document Set* », c'est-à-dire les textes sous pseudonyme à attribuer, et le « *Known Document Set* », c'est-à-dire celui du candidat, et que cet écart est plus faible qu'avec tout le reste des autres échantillons écrits par des tiers.

Cette expertise a permis à cette personne de rester aux États-Unis, et donc de ne pas encourir de risques dans son pays d'origine.

En dehors du domaine judiciaire, la stylométrie a pu s'appliquer aussi à des cas un peu plus violents, si j'ose dire. Un autre exemple : les travaux de Carole Chaski sur la détection des lettres de suicide. Donc là, il ne s'agit pas de détecter si un texte est bien d'une personne, mais plutôt s'il s'agit bien d'une lettre de suicide authentique ou d'un faux, ce qui, dans certains cas, peut avoir des implications importantes.

Carole Chaski, a entraîné un modèle d'IA sur un corpus de 1000 lettres dans lequel il y avait à peu près 400 lettres de suicide authentiques et 600 fausses. Et à 80 %, elle a réussi à détecter les lettres authentiques des faux.

Ce qui est bizarre, c'est que la performance s'accroît sur les textes plus courts, ce qui est plutôt l'inverse de ce qui se passe en général en stylométrie. Mais là, il semblerait que la brièveté soit un indice d'authenticité. C'est-à-dire que quand on écrit une lettre de suicide, on écrit rarement 40 pages et on écrit plutôt un texte relativement court.

Un cas emblématique auquel elle l'a appliqué, c'est la lettre de suicide de Kurt Cobain.

Là aussi, il y avait une théorie défendue par certains selon laquelle Kurt Cobain, le chanteur de Nirvana, aurait été assassiné et ne se serait pas suicidé, et que la lettre de suicide qu'on a retrouvée serait soit un faux, soit plutôt un texte authentique à laquelle on aurait rajouté quelques lignes à la fin pour le faire passer pour une lettre de suicide.

Cette théorie circule encore. En tout cas, ce que prédit son modèle à elle, c'est qu'il s'agit bien d'une lettre de suicide, que l'on prenne que le début, que l'on prenne que la fin, ou que l'on prenne l'intégralité, la prédiction est la même, les traits linguistiques de ce document évoquent une lettre de suicide.

Quand on parle de stylométrie, on évoque aussi parfois le cas d'Unabomber qui est un petit peu différent. Unabomber, c'est ce terroriste domestique américain qui a été très difficile à trouver, qui a fait l'objet d'une des plus longues enquêtes, du FBI et qui a fini par être identifié par sa langue, pas tellement de manière stylométrique, mais à l'aide de

la linguistique forensique qui est un champ similaire, et qui s'intéresse aussi aux propriétés de langue pour profiler ou attribuer des textes.

C'était un cas assez intéressant parce que cet Unabomber a écrit un long manifeste de revendications, et dans ce manifeste, il y avait un certain nombre de choses qui permettaient de construire un profil. Par exemple, il utilisait des orthographies pour certains mots qui correspondaient à la pratique du Journal de Chicago dans les années 40-50, mais qui n'apparaissaient pas ailleurs.

Il mettait en forme ses références, comme celles des thèses de mathématiques de telle université américaine pendant telle période. Il y avait des références religieuses qui laissaient croire à une éducation catholique, etc.

Avec ces éléments-là, les enquêteurs ont pu reconstituer un profil qui a ensuite aidé à restreindre le champ des candidats possibles et à finalement identifier et arrêter Theodore Kaczynski, comme étant l'Unabomber. C'était un mathématicien ayant grandi à Chicago, et les traits qui avaient été repérés se sont vérifiés en pratique.

Un dernier cas sur lequel nous avons travaillé Florian Cafiero et moi-même, est celui de Q-anon. Si vous êtes familiers des théories du complot et aussi de la tentative de coup d'état qui a eu lieu il y a quelques années aux États-Unis, vous avez peut-être déjà entendu parler de Q-anon. Vous avez peut-être déjà vu ce type de personnages hauts en couleur qui ont tenté de prendre d'assaut le Capitole, certains portant des panneaux du type « Q sent me » et ainsi de suite, se référant à un mystérieux « Q ». Il s'agit du pseudonyme de quelqu'un qui poste sur des forums un peu sombres du web tels que « 4Chan », puis « 8Chan », des contenus dénonçant une théorie du complot, des contenus dénonçant un vaste complot de l'état profond et des démocrates, et héroïsant Trump.

Ce « Q » lui-même se présente comme une forme de lanceur d'alerte haut placé au sein de l'administration américaine qui serait à même de dénoncer cet état profond dont on verrait les complots à l'œuvre.

Cela fait un certain temps que des journalistes d'investigation se penchent sur qui est « Q », qui est la personne qui est derrière cette nébuleuse complotiste. Ils avaient identifié

plusieurs candidats qui sont en gros les modérateurs ou les administrateurs des forums sur lesquels « Q » a posté.

A l'inverse, la « Q-Anon » sphère croit que l'auteur de ces textes est soit un général américain, soit Donald Trump lui-même, soit un proche de Trump.

Donc avec Florian, on avait été contacté par un journaliste du New York Times pour essayer de répondre à deux questions : Est-ce que Q est une seule personne ou plusieurs ? Et est ce qu'on peut mettre un nom derrière ces deux personnes ?

On s'est amusé à tester beaucoup de candidats, notamment les fameux modérateurs, mais aussi des généraux, un certain Donald T. ou une Melania T. pour ne pas les citer.

Ce dont on s'est aperçu, c'est qu'il y avait deux candidats qui ressortaient largement en tête avant les autres, le premier dont vous voyez la courbe en rouge et le deuxième dont vous voyez la courbe en noir, même si c'est un petit peu moins net. Il y a un moment de bascule où ces "Qdrop", donc ces posts, passent du premier style au deuxième style. Phénomène étrange, cette bascule se produit exactement au moment où "Q" migre d'un forum à un autre. Au moment où il migre de « 4Chan » à « 8Chan » tout d'un coup, le style de ces posts change. Plus précisément, le premier candidat disparaît après une date que je vous ai mis en pointillés ici. Ce premier candidat a écrit un livre sur « Q » dans lequel il défend la thèse que seuls sont authentiques les posts publiés avant cette date et que ceux publiés après sont des faux.

Vraisemblablement, il est bien placé pour le savoir puisque c'est lui qui les a écrits.

C'est toujours intéressant de voir la manière dont des éléments externes peuvent venir appuyer, vérifier, informer ou infirmer une hypothèse de stylométrie, et de voir qu'on considère la stylométrie par rapport à un faisceau d'éléments indépendants les uns des autres, et non pas en isolation.

D'ailleurs, avec mon collègue Florian, on a commencé à réfléchir à ce qui pourrait être un protocole possible, ou en tout cas des besoins qu'on pourrait avoir en tant que stylomètre pour intervenir dans une affaire judiciaire.

Un des points qui nous a semblé importants, c'est que ce n'est pas la stylométrie qui doit définir le périmètre dans lequel elle s'applique, c'est à dire : qui sont les candidats ? Combien sont-ils ?

Quelle est la nature de la tâche ? S'agit-il de profiler quelqu'un ? S'agit-il de déterminer un candidat parmi un ensemble fermé ou un ensemble ouvert ? Et ainsi de suite, tous ces éléments devraient être déterminés non pas par la stylométrie, mais par des choses externes.

Il faudrait aussi que la méthode soit toujours, en un sens, pré-validée sur des cas similaires. Si on veut par exemple, utiliser la stylométrie sur des cas de lettres de suicide, il ne faut pas la tester sur les œuvres de Molière qui n'ont rien à voir, il faut toujours calibrer et valider sur des données similaires.

Il faut également, en tant que chercheur ou expert, pouvoir répondre en termes calibrés, et sortir du sensationnel du type : « il y a une correspondance ou il n'y a pas de correspondance ». Il faut plutôt pouvoir dire : « selon une échelle de probabilité de certitude, les indices stylométriques pointent plutôt vers telle personne avec telle mesure de certitude, avec tel risque d'erreur ». C'est un enjeu important pour rendre nos méthodes compréhensibles, c'est-à-dire savoir ouvrir la boîte et expliquer le fonctionnement des algorithmes, la manière dont on conçoit un protocole d'analyse, les choix que l'on fait, et comment on peut rendre cela plus transparent pour des non-experts.

Voilà pour moi, j'ai déjà trop parlé je pense. Si vous voulez en savoir plus sur la stylométrie, je vous ai mis quelques références. Alors la première est l'ouvrage que Florian et moi-même avons écrit. Vous avez aussi l'ouvrage de Patrick Juola qui s'appelle « Authorship Attribution », qui a quelques années maintenant et qui reste un bon exemple. Puis je vous ai mis aussi un manuel plus généraliste, de l'utilisation de Python pour l'analyse de données textuelles.

Je vous remercie. »

### **Question dans le public :**

**Question 1 :** Merci beaucoup pour l'intervention qui était à la fois vulgarisée, mais extrêmement précise, extrêmement claire pour des non-initiés.

J'aurais une première question liée au fait que, aujourd'hui, lorsque nous parlons de machine learning ou de ce qu'on appelle plus vulgairement l'intelligence artificielle, nous pensons beaucoup aux données et à la capacité d'extraire de la valeur de ces données. Or, quand vous vous êtes présenté, vous vous êtes présenté comme médiéviste et c'est vrai que j'ai trouvé ça très intéressant de voir que même si nous étions en train de parler d'algorithmes ou d'outils computationnels qui nous aident à automatiser certaines tâches, la présence de l'expertise de base était là et n'était pas effacée par l'aspect calculatoire. Aussi, vous avez cité Carole Chaski qui est de renommée mondiale dans ce domaine-là qui, je le rappelle, a un doctorat en linguistique et qui est linguiste médico-légale, ce que j'aimerais vous demander, c'est quelle place vous donnez à la linguistique dans la stylométrie, puisque nous pouvons le voir, par exemple, dans les cas des affaires comme un pourcentage, finalement, qui est montré pour dire "on est confiant que tel texte appartient à telle personne". Mais finalement, ce que je perçois dans votre présentation, c'est aussi une preuve d'autorité qui vient de l'expertise en linguistique, qui est accompagnée par l'outil computationnel et non pas une validation simplement par le nombre. Donc, si vous pouviez peut-être développer un petit peu l'importance de l'expertise dans la combinaison avec les outils de calcul.

Et puis une deuxième question, dernièrement, je ne sais pas si vous avez vu, mais il y a eu une brève interview avec Carole Chaski où elle émettait quelques doutes sur la question de l'usage des Large Language Models dans ce domaine-là, vous avez parlé à la fin de votre présentation de l'importance de la transparence et de la reproductibilité, nous savons que les réseaux de neurones sont principalement des Black box et donc pourraient poser des problèmes en termes d'interprétation. Quelles opportunités vous verriez dans cette nouvelle vague d'algorithmes de machine learning pour augmenter ou aider encore plus ce travail lié à la stylométrie ?

**Dr Jean-Baptiste Camps :** Merci beaucoup. Alors pour répondre d'abord sur la question de l'expertise domaine, je pense qu'elle est cruciale et indispensable puisqu'on peut avoir l'algorithme le plus sophistiqué du monde si on construit son corpus de manière absurde, si on lui pose une question idiote, on aura des résultats qui ne veulent rien dire. Ça a pu être assez net dans certains cas, par exemple, il y a eu des études de stylométrie sur

Molière et Corneille qui finissaient par conclure un peu de mauvaise foi que Corneille avait dû écrire les œuvres de Molière. Le problème, c'est qu'elles avaient été produites par des personnes qui n'étaient pas du tout spécialistes de théâtre du XVII<sup>e</sup> siècle ou de philologie, ni de linguistique, et qui s'étaient amusés à comparer des textes qu'il est difficile de comparer, c'est-à-dire des comédies avec des tragédies, des textes en prose avec des textes en vers, et ainsi de suite, dans un contexte au XVII<sup>e</sup> siècle, où on sait que, par exemple, le théâtre est extrêmement codifié, qu'il y a une langue particulière qu'on emploie pour la tragédie, qui n'est pas la même que pour la comédie... Et là, je reviens un peu à ce que j'ai dit au début, c'est que la variation individuelle de style s'inscrit dans une gamme de variations qui peuvent être liées à l'âge, au lieu d'origine, à au genre des textes, etc.

Donc, quand on construit son corpus pour essayer d'attribuer un texte, il faut prendre soin de ne pas faire d'erreurs énormes dans la constitution du corpus qui pourraient biaiser les résultats. Si par exemple, je veux attribuer une lettre de suicide anonyme, et que je la compare à un corpus constitué uniquement de posts sur Twitter pour neuf personnes sur dix et d'une lettre de suicide pour la 10<sup>e</sup> personne, j'ai une forte chance d'attribuer à la 10<sup>e</sup> personne, simplement parce que mon corpus est constitué de manière biaisée.

Donc l'expertise du domaine est importante et c'est aussi pour ça que ça limite ce qu'on peut faire, pour en venir à la deuxième question, en termes d'usage de certains modèles d'IA. Donc, nous, on utilise de l'IA, on utilise des choses un petit peu démodées par exemple, dans ces analyses-là on a utilisé ce qui s'appelle des machines à vecteurs de support qui sont du machine learning de la fin du XX<sup>e</sup> siècle. Pourquoi ? Parce que ce sont des méthodes qui sont transparentes et on peut très bien ensuite extraire ce qui a guidé la classification d'un texte comme étant de la personne A ou de la personne B, on peut extraire une fonction de décision, un taux de confiance du modèle, et ainsi de suite.

C'est absolument vital de le faire, parce que c'est là qu'on se rend compte aussi de tous les artefacts possibles dans le corpus.

Pour prendre un exemple, sur "Q-anon", typiquement au début on avait fait une première analyse avec un corpus un peu de bric et de broc, parce qu'on a du mal à récupérer des textes des différents candidats, et on s'est aperçu que l'un des candidats ne ressortait pas du tout. On est ensuite allé voir directement dans les données et on s'est rendu compte que parmi ces données figuraient une fuite Discord, donc des conversations privées. La personne en question était un activiste d'extrême droite sud-africain, le texte était truffé de termes très fleuris que je ne répéterai pas, qui étaient complètement absents du reste du corpus. Et évidemment, son style était identifié par la machine selon ces termes fleuris, et ces termes fleuris n'apparaissaient pas du tout dans les « Qdrop » qui eux, étaient publics. Une fois qu'on a retiré ces textes problématiques, on a vu son signal stylistique qui apparaissait, donc ça demeure important de pouvoir démontrer la machine et de voir quels sont les éléments qui guident l'attribution à un candidat ou à un autre.

## **Question 2 :**

*Merci pour la présentation et pour toutes ces explications. Donc j'aurai deux questions, la première est : est ce que les magistrats sont formés à ces nouvelles technologies ?*

*Et la deuxième, est-ce que c'est la même approche dans tous les pays ?*

**Dr Jean-Baptiste Camps :** Sur la formation des magistrats, je serais bien incapable de répondre en tant que spécialiste en littérature et complètement extérieur à ce domaine. Pour ce qui est des différences entre pays, ce que je peux observer, c'est qu'il y a des pays comme les États-Unis où la stylométrie a fait son apparition dans des affaires judiciaires il y a plus longtemps qu'en France, donc il peut y avoir des disparités de ce de ce type-là.

*Merci pour cette présentation, c'était passionnant. Alors ma question, elle est peut-être un peu en marge, mais je me demandais s'il existait des déterminants biologiques au « stylome » et comme par analogie avec le génome, est-ce que ça peut se transmettre ? Est-ce qu'il peut y avoir une transmission, une sorte d'héritage ? Et puis deuxième question brève, moi je suis médecin et je voulais savoir s'il y avait une application dans le domaine médical, notamment dans des dépistages de certaines pathologies neurodéveloppementale et est-ce que c'est fait en pratique courante ?*

**Dr Jean-Baptiste Camps :** Alors dans le domaine médical, oui, il y a des applications qui sont un peu plus récentes, j'ai justement des collègues qui ont travaillé notamment sur la détection de l'ADHD, donc ce genre de troubles du comportement à partir de textes. Le nom français est TDAH. Effectivement, il y a des traits qui peuvent permettre d'aider au diagnostic puisqu'on se rend compte que dans l'usage, des pronoms personnels dans la longueur des textes, dans la longueur des phrases... il y a des traits qui peuvent permettre de distinguer si la personne qui a écrit ce texte a de plus grandes chances ou pas d'être sujette à un TDAH. Il y a eu d'autres études de ce type d'analyses, c'est encore assez peu courant, mais comme je l'ai montré avec le profilage sociolinguistique tout à l'heure, beaucoup de déterminants sociaux et de santé mentale peuvent apparaître et transparaître dans les textes.

Et pour ce qui est de l'hérédité, à ma connaissance, il n'y a pas de travaux là-dessus. On repère effectivement que parfois le style de frère et sœur peut avoir des proximités, mais c'est très dur de détacher ce qui est de l'ordre du contexte dans lequel ils ont grandi de ce qui pourrait être une hérédité. Les sœurs Brontë, par exemple, ont des styles qui sont assez similaires.

A l'inverse, Pierre et Thomas Corneille, qu'on a vu tout à l'heure, qui sont séparés par quelques années, enfin ils sont assez espacés, ont des styles assez différents.

C'est très délicat puisqu'évidemment, là où on a grandi, la date à laquelle on est né et ainsi de suite vont être déterminants aussi dans l'élaboration de son style. »d

## Me François Saint-Pierre, avocat

### Réflexions et application concrète dans l'« affaire Grégory »



*Avocat pénaliste – Procédure pénale et Stratégie de défense. Ténor du barreau, François Saint-Pierre exerce depuis plusieurs décennies une activité exclusivement dédiée au droit pénal et à la procédure pénale. Reconnu pour sa maîtrise des mécanismes procéduraux et sa capacité à structurer des stratégies de défense précises, éthiques et efficaces, il intervient régulièrement dans des dossiers complexes où se mêlent enjeux techniques, libertés fondamentales et haute intensité judiciaire.*

*Son expertise l'amène à plaider devant l'ensemble des juridictions pénales, dans des affaires sensibles nécessitant une lecture fine du droit, de la preuve et des dynamiques judiciaires. Parallèlement à sa pratique contentieuse, il contribue activement à la formation des avocats, magistrats et praticiens à travers ses ouvrages de référence et ses enseignements. Cette double dimension – praticien reconnu et pédagogue exigeant – lui confère un rôle structurant dans la réflexion contemporaine sur la défense pénale, l'éthique de la profession et la qualité du procès. François Saint-Pierre est régulièrement sollicité pour éclairer le débat judiciaire, notamment lorsqu'il s'agit d'analyser l'évolution de la procédure pénale, les droits de la défense ou les enjeux systémiques du procès pénal. Son approche, fondée sur la rigueur, l'humanité et le respect strict des principes fondamentaux, en fait une figure majeure du paysage pénal français.*

*En s'appuyant sur les avancées récentes, notamment celles issues de la stylométrie, il a contribué à faire émerger une vérité fondamentale longtemps obscurcie par des décennies d'incertitudes, démontrant combien une défense méthodique, une humanité profonde alliée aux outils scientifiques contemporains, peut transformer la lecture d'une affaire criminelle et enfin permettre de trouver des réponses des décennies plus tard.*

« Écoutez, dans les tribunaux et dans les cours d'assises, c'est une révolution qui est en train de se développer avec l'irruption de toutes ces nouvelles technologies. Ça ne date pas d'aujourd'hui. Cela fait quelque temps déjà mais si vous assistez à un procès dans une cour d'assises, la plupart du temps, vous verrez qu'il y a des vidéos qui sont diffusées.

La plupart du temps, vous êtes témoins directs de ce qui s'est passé parce qu'il y avait des caméras de surveillance ou bien la scène a été filmée par quelqu'un avec un iPhone.

Vous avez bien sûr la téléphonie qui permet de tracer l'itinéraire d'une personne d'un point à un autre.

Vous avez évidemment l'ADN dont on laisse des traces un peu partout, et les experts vont pouvoir identifier les personnes qui ont participé à l'action.

Et puis vous avez maintenant la stylométrie qui fait son irruption dans une célèbre affaire dont je peux vous parler, car je crois que c'est aussi pour cela que vous m'avez invité.

Et vous allez avoir, et là je parle au futur, l'utilisation de l'intelligence artificielle par les services d'enquête. Ce qui pose évidemment des questions de liberté publique, de vie privée est-ce que, par exemple, la reconnaissance faciale est un mode d'investigation légal aujourd'hui ou non ?

On en parlera, mais je dois dire que dans les mains de la police judiciaire et des juges d'instruction, ce sont des outils extraordinaires pour élucider des énigmes, identifier les personnes qui auraient commis un crime, n'est-ce pas ? Et par conséquent, épargner les innocents.

Et je pense qu'en termes de sécurité de nous tous face à la justice, ces nouvelles technologies, elles sont très remarquables.

Alors, je vais prendre cet exemple qui est l'affaire du petit Grégory. Je suis avec des consœurs, l'avocat des parents, Jean-Marie et Christine Villemin, ça n'est pas un mystère, c'est quelque chose de très public et je ne vais pas trahir de secret, mais c'est tellement puissant comme exemple en même temps d'utilisation des techniques, mais aussi de philosophie judiciaire, c'est-à-dire de philosophie de l'utilisation de ces techniques que je voudrais vous en dire quelques mots.

D'abord, on parle d'un crime qui est désormais très ancien, c'était le 16 octobre 1984, ça fait plus de 40 ans et à l'époque rien n'existe de cet ordre-là. Et on trouve le corps de cet enfant dans une rivière, la Vologne. Des pompiers vont le récupérer, un médecin légiste va procéder à une autopsie de piètre qualité, n'est-ce pas ? Et un juge d'instruction, tout seul dans son bureau d'Épinal, va se demander qui a bien pu avoir commis ce crime-là. Et il va désigner cette fois-ci des experts en graphologie, ce qui est, on pourra en discuter, une technique d'appréciation de la forme des lettres qu'une personne a pu écrire au stylo sur un papier, et en l'occurrence, c'étaient des lettres anonymes avec des lettres bâton comme l'on dit, bref, ça n'était pas extrêmement puissant comme mode d'investigation. Toujours est-il que, se fiant à son intime conviction, il va désigner quelqu'un à nommer : Bernard Laroche, qu'il va inculper comme on disait à l'époque du meurtre, de l'assassinat, et il va l'incarcérer.

Et puis comme ces expertises sont douteuses et que l'avocat de cet homme va en faire annuler quelques-unes, il va libérer cet homme. Et on sait qu'un drame va se produire parce qu'éprouvant un sentiment d'injustice irrépressible, et bien le père de l'enfant mort ira tuer son cousin et c'était bien autrefois.

Dans le même temps, ce juge se dit "Ah mais si ce n'était pas lui, c'est la mère de l'enfant" et Marguerite Duras, c'est vous dire la rationalité de l'époque, écrit dans Libération, dirigée par Serge July: « Elle est forcément coupable, c'est forcément elle. »

Bref, les années ont passé et dans les années 1990, un juge beaucoup plus psychologue et rationnel en même temps, va deviner qu'en fait ce crime n'a pas pu être commis par un homme seul mais il est forcément une œuvre, si je puis dire, collective, de plusieurs personnes.

On en était là. L'affaire est relancée à plusieurs reprises. Les années passent et la gendarmerie se dote d'un logiciel qui s'appelait "Anacrime", qui était antérieure à l'intelligence artificielle et qui procédait par indexation.

Voyons bien, la masse considérable de données d'indices dans un dossier comme celui-ci, qu'une personne seule qui passerait des milliers d'heures, n'est-ce pas, ne pourrait pas connecter de façon cohérente les unes aux autres.

Et bien ce logiciel d'indexation a permis à la gendarmerie qui a fait un travail remarquable de pouvoir justifier la thèse de ce magistrat qui avait émis l'hypothèse d'un crime collectif.

Et là, on se rend compte d'une chose absolument horrible déjà avec cette technique-là, c'est que la mort de cet enfant de quatre ans a été discutée, projetée et accomplie par plusieurs personnes qui font nécessairement partie de l'entourage familial, des parents de cet enfant. C'est abominable, c'est abominable.

Les juges se succèdent et chaque juge a voulu absolument résoudre l'énigme. C'est vraiment une question de déontologie pour ces magistrats.

Et je réponds à votre question de tout à l'heure, les années passant, les générations se succèdent et je vois de plus jeunes juges d'instruction qui se passionnent pour ces technologies, mais sans se prendre pour des experts. Il s'agit là aussi de regarder l'autre travailler, comme disent les pilotes d'avion. Le juge regarde l'expert travailler et on en discute de façon contradictoire, bien sûr.

C'est alors qu'un expert suisse a proposé l'utilisation de la stylométrie et je me souviens, c'était il y a peut-être trois ou quatre ans, je peux vous dire que dans le monde judiciaire, peu de monde en avait entendu parler, si ce n'est personne en vérité.

Et certains disaient mais qu'est-ce que c'est que ce gadget ? Et quand on écoute Monsieur Camps ce soir, c'est remarquable de rationalité, de culture, de science d'une certaine façon pour l'utilisation de cette technique-là.

Donc l'expert est mandaté, le souci du juge, il est de savoir si cette méthode sera précise, fiable, probante, il est hors de question d'accuser qui que ce soit sur une impression.

Mais puisque le procès doit être équitable, il faut aussi qu'il trouve un autre expert, un contre expert pour apporter la contradiction au premier.

Et alors il était là, avec sa lanterne, à chercher des experts mais je crois que j'ai un nouveau nom ce soir.

Toujours est-il que cet expert analyse les différentes lettres anonymes que le corbeau, vous savez, dans cette histoire, on parle du corbeau qui a écrit des dizaines et des dizaines de lettres, plus des centaines d'appels téléphoniques anonymes, n'est-ce pas ?

Et cet expert dit : « C'est vraisemblablement la tante du petit garçon, ou la grande tante, plus exactement madame J. ».

Le juge va donc bien réfléchir à ce qu'il va faire. Il y a une audience où le procureur s'explique.

Nous, pour les parents, nous nous expliquons et finalement, la Cour d'appel rend un arrêt pour dire elle sera mise en examen.

C'est quelque chose d'extrêmement puissant, mais qui implique une responsabilité majeure bien sûr, cette personne a plus de 80 ans maintenant et la voici qui se retrouve mise en examen, comme l'on dit maintenant, non pas pour le crime ou la complicité du crime, car nous n'avons pas de description de la scène du crime, mais pour association de malfaiteurs, c'est-à-dire pour avoir préparé ce crime

Là en étant l'un des corbeaux, pour avoir téléphoné anonymement et avoir écrit ces lettres-là. Alors bien sûr, elle est assistée d'avocats et de bons avocats et heureusement, nous sommes dans un État de droit. Et elle conteste. Elle dit : "Mais jamais, jamais je n'ai écrit ces lettres, jamais je n'ai passé ces coups de fils".

Pour les lettres, nous avons la stylométrie. Pour la voix, nous avons une technologie qui va se développer, qui est une technologie de reconnaissance vocale. Alors, pas des voix synthétisées, il s'agit de faire la lumière, d'établir la vérité sur l'identité d'une personne

qui s'est exprimée. Et là, en l'occurrence, sur la base d'un enregistrement de piètre qualité, nous sommes en 1981 à l'époque.

Bon, eh bien, comme chacun a une empreinte digitale, chacun a une empreinte vocale et la question est donc de savoir si là, dans six mois, dans un an, eh bien des scientifiques, des ingénieurs auront pu mettre au point un logiciel et l'IA va beaucoup nous aider à clarifier les enregistrements passés, à identifier l'auteur ou l'autrice de ces appels. Est-ce que ce sera la même personne ? C'est ce qu'on verra !

Et puis enfin, il y a, je vous parlais de l'ADN dans cette affaire, une pièce à conviction qui est déterminante. C'est une lettre qu'une personne, qui on ne sait pas, a posté le jour du crime pour revendiquer l'assassinat et dire au père "Eh bien, voilà notre vengeance". Il se trouve qu'il y a un expert de la génétique judiciaire qui s'appelle Mr Christian Doutremepuich, qui est expert à Bordeaux et qui, depuis plus de 30 ans, travaille à l'élucidation des crimes. C'est un homme très remarquable et il a pu identifier sous le timbre qu'à l'époque on collait avec un bout de salive quand même, et également sous le rabat de la lettre, même chose, il fallait lécher l'enveloppe pour la coller.

On a plusieurs ADN mélangés. Or, il se trouve que la technologie d'aujourd'hui ne permet pas de discriminer les ADN et par conséquent d'identifier les différentes personnes.

Je n'en dirai pas beaucoup plus, mais je pense que cet expert, avec son laboratoire, c'est un travail d'équipe, va pouvoir mettre au point une technique pour différencier les ADN. Et ce jour là, nous aurons le nom de l'un des auteurs de ce crime affreux.

Alors je disais que c'est aussi une question de philosophie, parce que certains disent oui, mais ça fait 40 ans, ça coûte extrêmement cher. Est-ce que notre société peut renoncer à résoudre l'énigme de l'assassinat d'un enfant de quatre ans dans les circonstances que j'ai rappelées ? Et bien pour la justice, c'est non. Pour les parents, c'est non. Pour leurs avocats, c'est non. Il faut aller jusqu'au bout et grâce à ces technologies de police judiciaire, il est possible que l'énigme soit résolue. Alors, je ne perds pas de vue non plus que toutes ces techniques, et vous le disiez si bien, les uns et les autres tout à l'heure, peuvent porter des atteintes irrémédiables à la vie privée, à la notoriété, à l'image sociale,

aux libertés publiques. Et chaque fois que je m'exprime sur ces technologies policières, je pense aussi en termes de libertés publiques, en termes de surveillance de masse.

Je m'interroge à l'effet que cela produit sur nos psychologies à tous, parce que je pense que ça change nos psychismes, cela, de savoir que rien ne restera dans l'ombre, pour le meilleur ou pour le pire. Imaginons un régime autoritaire demain, en France comme ailleurs. Comment résister ?

Et cela, c'est un questionnement majeur que nous devons nourrir.

Voilà ce que je souhaitais vous dire. Et vous voyez, c'est vraiment une question de philosophie, de morale civique. Les outils depuis toujours peuvent être utilisés à bon escient, comme dans cette affaire criminelle ou au contraire être des armes tellement destructrices. Et bien c'est la raison pour laquelle on est là ce soir, à en discuter ensemble. Et comme tout débat est contradictoire mon cher confrère va pouvoir compléter ce propos de la façon qu'il voudra. »

## Me Ouadie Elhamamouchi, avocat

### *Réflexions et applications concrètes dans l' « affaire Ramadan » » en France et en Suisse*



**Maître  
Ouadie El  
Hamma-  
mouchi**

Avocat

*Avocat pénaliste – Expertise contentieuse. Maître Ouadie Elhamamouchi appartient à cette génération d'avocats pénalistes qui se distinguent par une maîtrise simultanée du terrain judiciaire, de l'urgence procédurale et d'une compréhension fine des réalités humaines. Son expertise s'est forgée au contact direct des situations les plus sensibles : garde à vue, comparutions immédiates, contentieux correctionnels complexes, où la précision technique et la lucidité stratégique sont décisives. Cette pratique intensive du pénal, alliée à une rigueur méthodologique croissante, fait de lui l'un des profils émergents les plus prometteurs du barreau.*

*Sa capacité à intervenir vite, juste et clair, sans sacrifier la profondeur de l'analyse, lui permet d'apporter des réponses efficaces là où chaque minute compte. Par son professionnalisme constant, son engagement auprès des personnes qu'il défend et son aisance à naviguer dans les environnements judiciaires les plus tendus, Ouadie Elhamamouchi s'impose comme l'un des avocats appelés à compter dans sa génération. Son approche, précise, humaine et ancrée dans une technique forte marque les contours d'une pratique humaine et exigeante.*

*La technologie a profondément marqué certains de ses dossiers, en offrant des outils d'analyse et de vérification qui ont permis de faire émerger des éléments décisifs, là où la justice se heurtait depuis des années à l'errance. Grâce à cette approche, des vérités longtemps inatteignables ont pu être établies, montrant à quel point l'usage rigoureux du numérique peut totalement transformer l'issue d'une affaire pénale.*

« L'affaire pour laquelle je suis là aujourd'hui est extrêmement emblématique, notamment en ce qui concerne la stylométrie. J'ai envie de vous dire que c'est peut-être un terrain magnifique pour l'étude de la stylométrie.

Souvenez-vous : l'affaire Tariq Ramadan était après Me-too, elle a fait l'objet d'un certain nombre d'accusations. Et comme vous le savez tous, ce sont des accusations qui, dans ce

dossier, obligent à établir un jugement à partir d'échanges : des échanges que vous avez certainement vus dans les journaux, des échanges numériques, notamment.

Et à l'époque, d'ailleurs, il ne s'agissait pas seulement de SMS échangés sur les réseaux sociaux, mais aussi d'échanges anonymes publiés sur des blogs. Il se trouve que c'est effectivement, par le biais de ces échanges, que l'accusation, comme la défense, ont dû travailler. Alors, étant donné qu'il s'agit d'un dossier de parole contre parole, il a fallu prouver l'innocence de Tariq Ramadan.

Alors, l'étude de la stylométrie, mon confrère disait qu'effectivement cette matière a fait irruption il y a à peu près trois ou quatre ans dans les tribunaux. À ma connaissance, il me semble que l'affaire Grégory a peut-être été la première affaire qui utilisait la stylométrie en France. Et, à ma connaissance, si ce n'est peut-être l'affaire Tariq Ramadan, je n'ai pas connaissance d'autres affaires, aujourd'hui, qui fassent état de la stylométrie dans une procédure pénale.

Toujours est-il que la grande difficulté que nous avons eue dans le dossier Tariq Ramadan, c'est qu'un certain nombre de "dires" qui auraient pu être attribués à un certain nombre de personnes malveillantes ne pouvaient pas être prouvés, puisque, finalement, les gens se cachaient derrière des pseudonymes, de l'anonymat sur Internet, sur des blogs, et nous avions cette difficulté à établir un certain nombre d'éléments.

Cela allait même jusqu'à son paroxysme. À titre d'illustration, je vais vous donner un exemple : une des plaignantes de Tariq Ramadan a reconnu avoir utilisé un pseudo. Ce pseudo c'est "Nathalie". Cette Nathalie va tenir un certain nombre de propos à l'endroit de Tariq Ramadan. Notamment quelques jours avant son unique rencontre avec Tariq Ramadan, elle écrit :

*"Je vais piéger Tariq Ramadan, je vais le faire tomber dans un piège, il va faire la une des journaux, la une des tabloïds, ce sera caméras et journalistes pour lui. Je vais lui mettre la honte devant le monde entier."* Six jours avant l'unique rencontre avec Tariq Ramadan.

Ce pseudo — "Nathalie" — la plaignante va reconnaître, à la suite d'une question du magistrat instructeur, l'avoir utilisé. Une de ses amies, qui est témoin dans notre affaire, va dire : *"Oui, effectivement, je discutais avec elle, et son pseudo, c'était 'Nathalie'.*

Donc là, vous avez bien compris : La plaignante elle-même reconnaît que son pseudo est "Nathalie". Son amie reconnaît qu'elle s'appelle "Nathalie" sur les blogs. Et puis vous allez avoir la brigade criminelle, qui est quand même experte en la matière et qui est extrêmement professionnelle, qui va dire : « *Quand on analyse finalement ce que dit "Nathalie" sur les blogs — notamment ce qu'elle déclare par rapport à sa vie privée, ce qu'elle déclare par rapport même à monsieur Tariq Ramadan — eh bien ça corrobore beaucoup ce que cette plaignante nous dit, à nous enquêteurs, et ce qu'elle dit aussi aux magistrats instructeurs* ».

Je ne vais pas rentrer dans les détails aujourd'hui, il n'est pas question de vous divulguer des détails, mais je vais vous donner un exemple. "Nathalie" va dire, quelques jours avant de rencontrer Tariq Ramadan : "Je vais le mettre dans un piège." Elle va dire : "J'ai des éléments, des preuves que je stocke chez un huissier." Et eh bien la plaignante, devant le magistrat instructeur, va dire exactement la même chose : "J'ai des éléments que j'ai stockés chez un huissier." Et on a pléthore d'éléments comme ça.

Aujourd'hui, face à cette plaignante confrontée aux faits, la difficulté apparaît clairement. Lorsque le magistrat instructeur l'interroge sur l'usage du pseudonyme « Nathalie », elle reconnaît dans un premier temps l'avoir effectivement utilisé. Or, il ressort des blogs que, sous ce pseudonyme, quelques jours avant sa rencontre avec Tariq Ramadan, un message évoquait la préparation d'un acte malveillant. Confrontée à cet élément et percevant que l'argumentation se retourne contre elle, la plaignante finit par se rétracter. C'est précisément pour cette raison qu'il convient de ne pas entretenir une vision idéalisée de la justice : elle en vient finalement à nier l'usage de ce pseudonyme, alors même qu'elle l'avait reconnu quelques minutes auparavant.

Et ça va suffire.

Ça va suffire.

Dans l'ordonnance de renvoi final de la cour d'appel de la Chambre de l'instruction, notamment par rapport à Tariq Ramadan, nous disons « *Ce n'est pas possible, Vous voyez bien que cette plaignante, c'est celle-là qui avait ce pseudo, qui a reconnu avoir eu ce pseudo et qui parle d'un piège, qui annonçait quasiment son plan*" Eh bien la cour d'appel va dire : « *certes,*

*mais nous n'avons pas la preuve scientifique puisque maintenant elle le conteste. Nous, nous n'avons pas la preuve scientifique que la plaignante est cette fameuse Nathalie qui tient quelques jours avant ses propos sur votre propre personne. »*

C'est la raison pour laquelle la défense a vu, dans l'étude du procédé de stylométrie, un intérêt certain et une opportunité réelle : celle de pouvoir, éventuellement, attribuer — ou au contraire exclure toute attribution — les phrases, éléments et accusations formulés sous le pseudonyme « Nathalie » à l'une des plaignantes.

Dans ce cadre, une expertise a été diligentée : une expertise privée, sollicitée par Tariq Ramadan, confiée au Dr Carole Chaski, linguiste de renom et pionnière des méthodologies de la linguistique forensique. Comme cela a été rappelé précédemment, le Dr Chaski fait autorité en matière de stylométrie et constitue une référence majeure dans ce domaine. C'est une spécialiste américaine reconnue à travers le monde aujourd'hui.

Et donc des milliers de mots, des centaines de phrases qui sont, puisqu'on le sait, attribuées à la plaignante, qu'elle reconnaît avoir eu des phrases qui sont parfois dans le dossier, des phrases qu'elle a reconnues, des échanges dont elle s'approprie notamment la propriété, vont être analysées avec les éléments que tient cette fameuse Nathalie, qu'elle avait reconnu quelque temps avant d'être sur les blogs. Et vous me permettrez de dire que le résultat est bien là. Cela a été plus que concluant puisque la papesse Carole Chaski conclut à une probabilité que Nathalie soit la plaignante avec une probabilité comprise entre 97 et 100 %. Ce qui signifie notamment, et c'est ce qu'elle détaille dans ses conclusions en termes de statistiques, à une quasi-certitude, que la plaignante est bien cette fameuse Nathalie à l'origine des messages malveillants évoquant un plan en préparation contre Tariq Ramadan quelques jours avant leur unique rencontre. Et cette expertise stylométrique, elle repose sur tout un tas d'éléments qui ont été tout à l'heure détaillés par mon voisin. Ça peut relever notamment de l'emploi des mots utilisés, de la structure des phrases, de la ponctuation, du choix des mots, des prépositions qui sont utilisées, des fautes d'orthographe, des cédilles aussi parfois, des oubliés de saisie, etc.

Tout cela est analysé de façon extrêmement précise par l'experte, et elle conclut à une probabilité quasiment de 100 % que la plaignante est bien cette fameuse "Nathalie". Alors, qu'est-ce qu'on en déduit — la défense de Tariq Ramadan comme Tariq Ramadan lui-même ?

On en déduit que ce que nous avons toujours pensé se confirme : ce qu'avait aussi constaté la brigade criminelle, à savoir que la plaignante ment. Elle ment parce qu'elle a elle-même annoncé ce qu'elle comptait faire : un piège, un traquenard, un complot. Appelez cela comme vous voulez. Elle ment donc effrontément à la justice. Et à partir de là, on n'est plus du tout dans le même dossier. Vous comprenez bien que si l'on parvient à prouver que cette plaignante était cette "Nathalie", alors le dossier change entièrement de nature.

Et ça, tout le monde le sait : les avocats de la défense le savent, les avocats des parties civiles le savent, l'accusation le sait. Et aujourd'hui, c'est pour cette raison que — puisque Tariq Ramadan ne demande aucun traitement de faveur — il ne dit pas : "*Croyez-moi sur parole parce que je vous produis une expertise.*" Non. Tariq Ramadan, avec l'honnêteté intellectuelle qui le caractérise, dit simplement : "*Voici ce que nous produisons. Nous avons cette expertise en stylométrie*"

Il se trouve que, lors de l'instruction, effectivement, l'instruction était terminée, et que cette méthode de stylométrie était encore très peu connue en France. Il se trouve qu'aujourd'hui, elle est de plus en plus utilisée, avec des résultats concluants. Alors voilà : nous sommes passés du stade où l'instruction est close, au stade où nous sommes désormais renvoyés devant une cour criminelle départementale, et où nous demandons à la présidente de cette cour d'instaurer et d'ordonner ce qu'on appelle un supplément d'information.

Qu'est-ce que c'est, un supplément d'information ?

C'est simplement dire : "*Voilà ce que nous avons obtenu à travers une expertise. À vous, Madame la Présidente — que je connais, d'ailleurs, et qui est une très bonne magistrate, une très grande magistrate — à vous de désigner l'expert qui vous semblera le plus*

*approprié parmi les spécialistes de la stylométrie, de lui transmettre les différents éléments, et de voir quelle sera sa conclusion."*

Nous n'avons aucun doute quant au sérieux de l'étude réalisée par Madame Chaski, dont le travail a été conduit avec une rigueur extrême. Il s'agit, une fois encore, d'une référence incontestée dans ce domaine.

Cela étant, dans ce cadre précis, et comme l'a rappelé mon confrère, une seule exigence nous guide, exigence qui oriente de manière générale l'action des avocats : la manifestation de la vérité. À ce titre, nous n'adoptons aucune posture partisane et ne nous plaçons pas dans une logique de "mauvais joueurs".

*"Nous souhaitons que vous ordonniez une nouvelle expertise, afin de vérifier si, effectivement, ce que nous avançons est exact : à savoir que la plaignante et cette fameuse "Nathalie" ne sont qu'une seule et même personne."*

Le but de notre intervention est de faire savoir que l'expertise stylométrique, qui commence tout juste à faire irruption en France, a déjà permis, aux États-Unis, de résoudre des dizaines d'affaires. Mon confrère le disait tout à l'heure : oui, nous avons du retard en France. Mais ce n'est pas parce qu'on a du retard qu'on ne doit pas s'y mettre. Ce n'est pas parce qu'on arrive après coup qu'il faudrait balayer cette innovation d'un revers de main.

Et d'autant plus que, et là, je suis persuadé, comme l'a dit mon confrère, nous allons assister, dans les années qui viennent, à une révolution en matière judiciaire grâce à l'intelligence artificielle. Cela va concerner la reconnaissance faciale, qui existe déjà, mais qui sera de plus en plus précise. Et cela concernera aussi la reconnaissance vocale.

Dans notre dossier, par exemple, nous avons des enregistrements audios : vous voyez, on entend des voix, on entend des gens parler. Et dans ce dossier, il y a une multitude de conversations, d'échanges. Si vous avez suivi l'affaire, vous savez même que des personnes célèbres s'en sont mêlées, parce qu'elles en voulaient spécifiquement à Tariq Ramadan. Nous avons donc beaucoup de matière, mais il est parfois difficile, à partir d'un simple audio, même si nous savons qui parle, d'attribuer cette voix à une personne

de manière irréfutable. Eh bien, avec les progrès de la reconnaissance via l'intelligence artificielle, je suis convaincu que dans les mois — ou dans les années — qui viennent, cela deviendra possible.

Et peut-être même en matière de reconnaissance gestuelle. Vous savez, très souvent, dans les vidéosurveillances, on a des images de mauvaise qualité : les personnes sont habillées de noir, on ne voit pas leur visage. Qu'est-ce qui empêcherait, à terme, qu'une IA parvienne, grâce à la manière dont une personne marche, se déplace, se meut, à identifier qu'il s'agit effectivement de la personne que l'on voit à l'image ? Voilà les quelques mots que je voulais vous dire. Et voilà pourquoi nous sommes tous ici : pour la justice et pour la vérité.

### Question du public

#### **Question 1**

Dans l'affaire de Tariq Ramadan, que j'ai suivie de près, deux expertises ont été réalisées : l'une en Suisse, l'autre en France. Et cela donne l'impression que la situation est systémique, puisque l'on retrouve exactement le même cas de figure : une plaignante qui tient des propos évoquant un piège, un traquenard, avant même d'avoir rencontré Tariq Ramadan. Une seule rencontre, quelque chose de cet ordre. Et, si je me souviens bien, vous avez fait appel à Carole Chaski — dont le docteur Camps nous a parlé toute la soirée.

Comment ces deux expertises ont-elles été reçues ? Y a-t-il une différence dans la manière dont la Suisse accueille ce type d'expertise ?

On sait que les Suisses sont pionniers. D'ailleurs, dans l'affaire du petit Grégory, c'est une société suisse qui était intervenue, avec un grand laboratoire de recherche à Lausanne.

Ici, en France, nous avons également nos experts, puisque nous avons le docteur Camps. Mais existe-t-il une différence entre la manière dont la Suisse et la France reçoivent ce type d'expertise ? D'autre part, le docteur Chaski est une experte reconnue, très connue, qui a mené de nombreuses enquêtes aux États-Unis. Est-ce que le fait qu'elle soit pionnière dans son domaine apporte quelque chose de particulier à l'expertise ?

Parce que, dans nos discussions — notamment avec une autre experte — nous avons mis en avant que la solidité d'une expertise, même très technologique, repose toujours sur son expert. C'est l'expert qui analyse, qui interprète et qui apporte sa preuve technologique. Ici, nous avons une linguiste, docteure Chaski, une société Orphanalitic ; ce sont des personnes où des institutions reconnues scientifiquement. Est-ce que cela pèse dans l'expertise ?

Enfin, ma dernière question. J'espère que vous retiendrez mes trois questions. J'ai lu — tout comme dans l'affaire du petit Grégory, deux affaires emblématiques qui durent depuis des années, plusieurs décennies pour le petit Grégory, près d'une décennie pour Tariq Ramadan — qu'une expertise informatique avait également apporté des éléments similaires : un blog malveillant retrouvé par un expert informatique, créé avant la rencontre, que la plaignante avait nié — du moins selon ce que les journaux rapportaient — avant que cela ne soit finalement retenu comme preuve probante. On a donc là aussi un faisceau d'indices. Comment, finalement, par la combinaison de plusieurs expertises, parvient-on à faire émerger une unité, une cohérence, qui permette cette manifestation de la vérité ? Comment comprendre la disproportion entre les éléments à charge et à décharge ?

**Me Ouadie Elhamamouchi** : Parce que, si je résume : lorsqu'on a des éléments scientifiques à charge, c'est terrible, c'est écrasant. Alors que, dans l'affaire que je défends, celle de Tariq Ramadan, nous avons de nombreux éléments objectifs, scientifiques, qui sont à décharge.

Bon, je ne souhaitais pas en parler là, puisqu'on était plus sur la stylométrie. Mais, par exemple, effectivement, nous avons une expertise informatique qui démontre qu'il s'agissait de la même plaignante qui a créé un blog spécifique sur Tariq Ramadan et qui a dit des choses abjectes et affreuses sur Tariq Ramadan avant leur unique rencontre.

Quand on a réussi à prouver qu'effectivement c'était cette plaignante qui avait créé ce blog, ça a été encore une fois compliqué pour elle. Eh bien, elle a dit finalement : *"J'ai créé ce blog postérieurement à ma rencontre avec Tariq Ramadan"*, ce qui était un autre moyen de s'en sortir.

Et là, encore une fois, la science — lorsqu'elle est bien faite, lorsqu'elle est objective — et d'ailleurs il y a une expertise qui le prouve, eh bien la science vient dire l'inverse : en disant que non, ce blog a été créé avant la rencontre avec Tariq Ramadan par ses soins. Et notamment grâce à tout un tas d'éléments informatiques que je ne vais pas vous détailler aujourd'hui, l'expert informatique conclut que le blog spécifique sur Tariq Ramadan a été créé avant la rencontre avec Tariq Ramadan.

Et donc vous avez encore là des mensonges qui sont prouvés scientifiquement. En ce qui concerne notamment la différence de traitement entre la France et la Suisse : alors oui, il y a cette possibilité. D'abord, puisque les plaignantes... c'est un dossier assez particulier. C'est aussi pour ça que c'est très intéressant en termes d'IA, de stylométrie, d'analyse des échanges, etc. C'est un dossier très vivant là-dessus.

Effectivement, il y a une plaignante suisse, d'ailleurs, qui échangeait avec des plaignantes françaises, et ce avant d'avoir rencontré Tariq Ramadan. Et effectivement, cette plaignante suisse avait parlé d'"une torpille en préparation" à l'endroit de Tariq Ramadan, là aussi, avant leur unique rencontre. Le Dr Chaski dans son expertise a pu prouver qu'elle était l'autrice de ce message malveillant et menaçant.

Oui, la Suisse est plus avancée. Et c'est vrai que la France, en matière de stylométrie, est en retard. C'est la raison pour laquelle la défense de Tariq Ramadan, en Suisse, a sollicité un recours, notamment en révision.

Je vais être très honnête : je ne sais pas comment va se positionner la justice française. Parce qu'effectivement, la justice française est scientifiquement en retard par rapport à la Suisse, ou à d'autres pays, concernant la stylométrie. Mais sur l'appréciation du poids de la preuve, sur la rigueur scientifique de la stylométrie, j'attends de voir. Dans notre dossier, il se trouve que la juge ne s'est pas encore positionnée sur le supplément d'information. Mais j'ai plutôt bon espoir, quand je vois comment ça a été accueilli dans le dossier de mon confrère : un dossier lourd, emblématique. Et d'ailleurs, pour être très honnête intellectuellement, ce dossier nous a évidemment inspirés.

On s'inspire d'autres affaires judiciaires, de ce qu'on voit, de ce qu'on lit.

Et on s'est bien sûr inspiré de l'affaire du petit Grégory, notamment des effets et des résultats de la stylométrie.

Et ça a été accueilli favorablement, avec — ce n'est pas rien — une mise en examen quarante ans plus tard : la personne a 80 ans. Mais c'est très symbolique. C'est très fort juridiquement. On n'a pas des décisions comme ça tous les jours.

Donc j'ai bon espoir que la justice française s'ouvre à ce procédé de stylométrie. Nous, c'est tout ce qu'on demande.

## **Question 2 :**

*Comme vous le disiez, les éléments scientifiques sont extrêmement déterminants dans la justice. Et ce sont des éléments qui permettent justement aux juges de se positionner de manière assez tranchée.*

*Je voulais savoir : est-ce que vous pensez que les juges sont suffisamment formés aux nouvelles technologies ?*

*Et, avec l'évolution rapide qu'on a actuellement, est-ce que vous pensez qu'ils sont suffisamment formés pour prendre leurs décisions ?*

**Me Ouadie Elhamamouchi :** Je vais être très honnête avec vous : tout le monde se forme en ce moment. Tout le monde se forme — juges, avocats, tout le monde. Les seules personnes vraiment expertes, ce sont celles qui sont expertes en stylométrie et qui s'y consacrent depuis longtemps. Voilà pourquoi on fait appel à des experts.

Maintenant, on peut avoir un raisonnement intelligible. L'expert n'est pas juge, mais on peut analyser ses rapports, ses données, ses fondements, pour en tirer une conclusion. Et, pour être honnête, j'ai l'impression que les magistrats se forment. Tout comme les avocats se forment. »

# Conclusion

Ce colloque a montré, avec une rare clarté, que la justice pénale se trouve aujourd’hui à un point de bascule. Les experts techniques, les spécialistes de la transformation numérique, les chercheurs en stylométrie et les avocats présents ont révélé chacun une facette essentielle de cette mutation profonde : la nécessité d’intégrer des outils capables de lire, d’analyser et d’attribuer des traces numériques devenues incontournables dans les enquêtes contemporaines.

Les échanges ont démontré que ces technologies ne constituent pas une simple amélioration instrumentale, mais un véritable changement de paradigme. À travers les investigations numériques exposées par Denis Jacopini, les dynamiques de transformation présentées par Visuddha Luanghot, les fondements méthodologiques rigoureux rappelés par Jean-Baptiste Camps et l’usage concret de la stylométrie illustré par Maîtres Saint-Pierre et Elhammamouchi, une même conviction s’est imposée : **l’IA et la stylométrie ne sont plus des disciplines périphériques ; elles deviennent des outils centraux de la manifestation de la vérité.**

Ce colloque a également mis en évidence un point fondamental : l’apport de ces technologies n’a de sens que s’il s’inscrit dans une éthique exigeante, une maîtrise juridique solide et un respect constant des libertés individuelles. Les intervenants ont rappelé que la technologie ne peut ni remplacer la délibération humaine, ni se substituer à l’intime conviction des juges, mais qu’elle peut en revanche éclairer la décision, dissiper les zones d’ombre et réduire les espaces d’impunité que l’anonymat ou la manipulation numérique rendaient, jusqu’ici, presque indétectables.

L’ensemble des contributions rassemblées dans ces actes reflète ainsi la naissance d’un nouvel équilibre : une justice plus informée, plus outillée, mais consciente de sa responsabilité. Une justice qui accueille la puissance des innovations sans renoncer aux principes fondamentaux qui la structurent. Une justice enfin capable d’exploiter, avec discernement, les technologies qui permettent de mieux comprendre le réel et d’approcher, au plus près, la vérité des faits.

Ce colloque marque une étape. Les années à venir confirmeront sans doute l'importance croissante de ces outils dans les procédures pénales. Mais leur intégration n'est pas un horizon lointain : elle est déjà en marche, comme l'ont montré les pratiques et les analyses partagées ici. Ces actes en portent la trace. Ils rappellent que l'alliance de la rigueur scientifique, de l'expertise technique et de l'exigence juridique constitue désormais l'une des clés majeures d'une justice plus efficace, plus lucide et, espérons-le, plus juste.