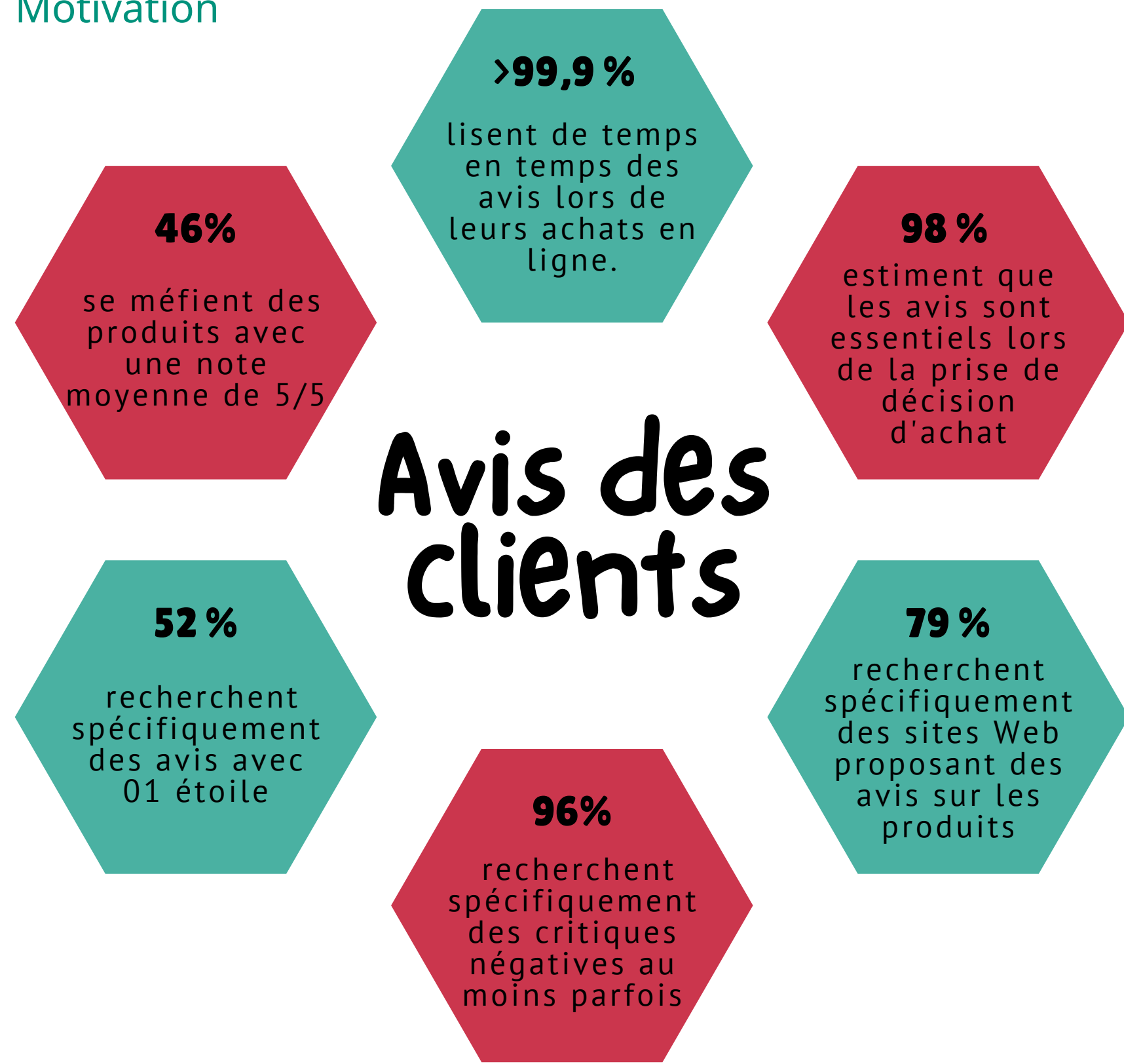


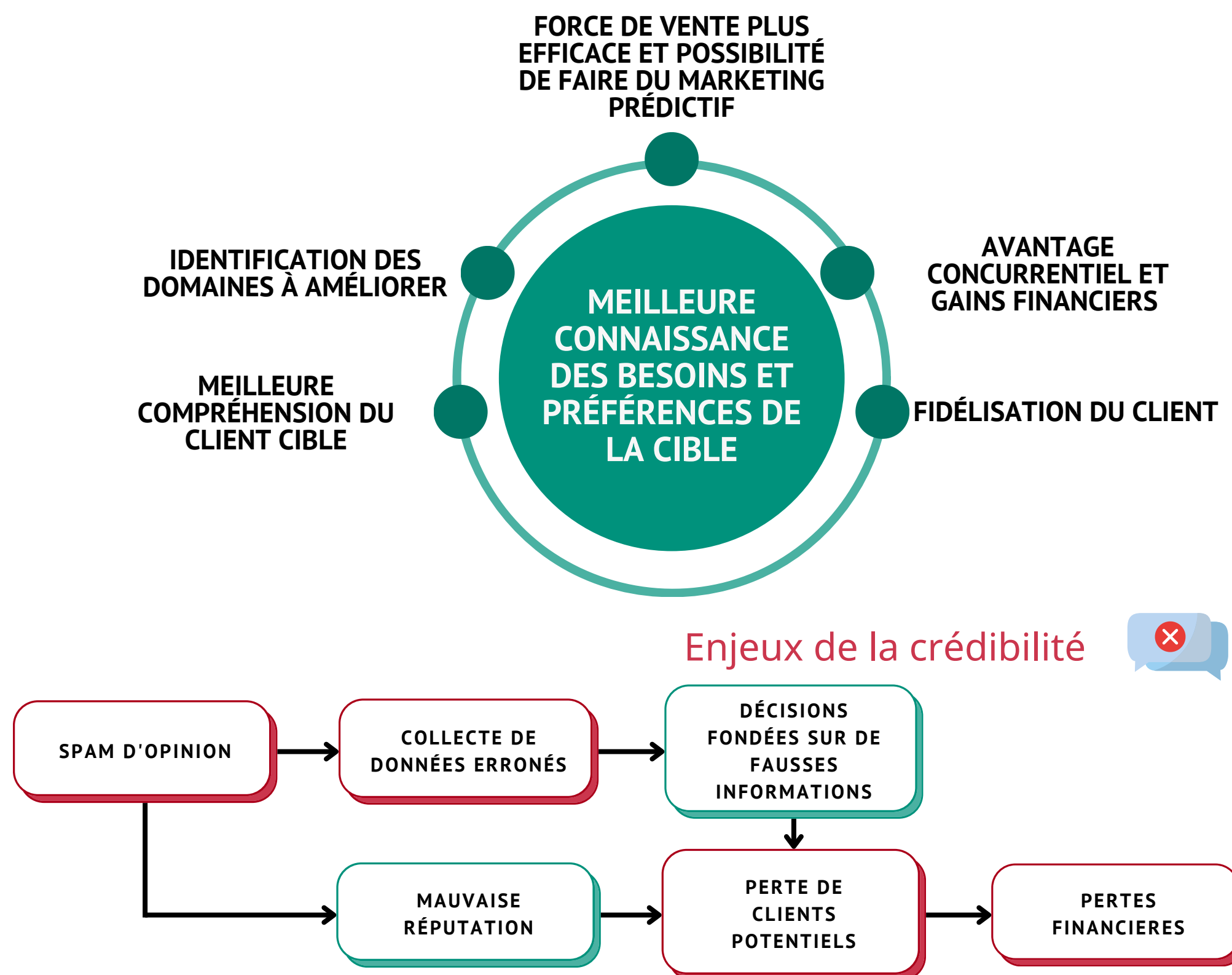
État de l'art sur l'intégration de la crédibilité dans la tâche de fouille d'opinion

Contexte & problématique

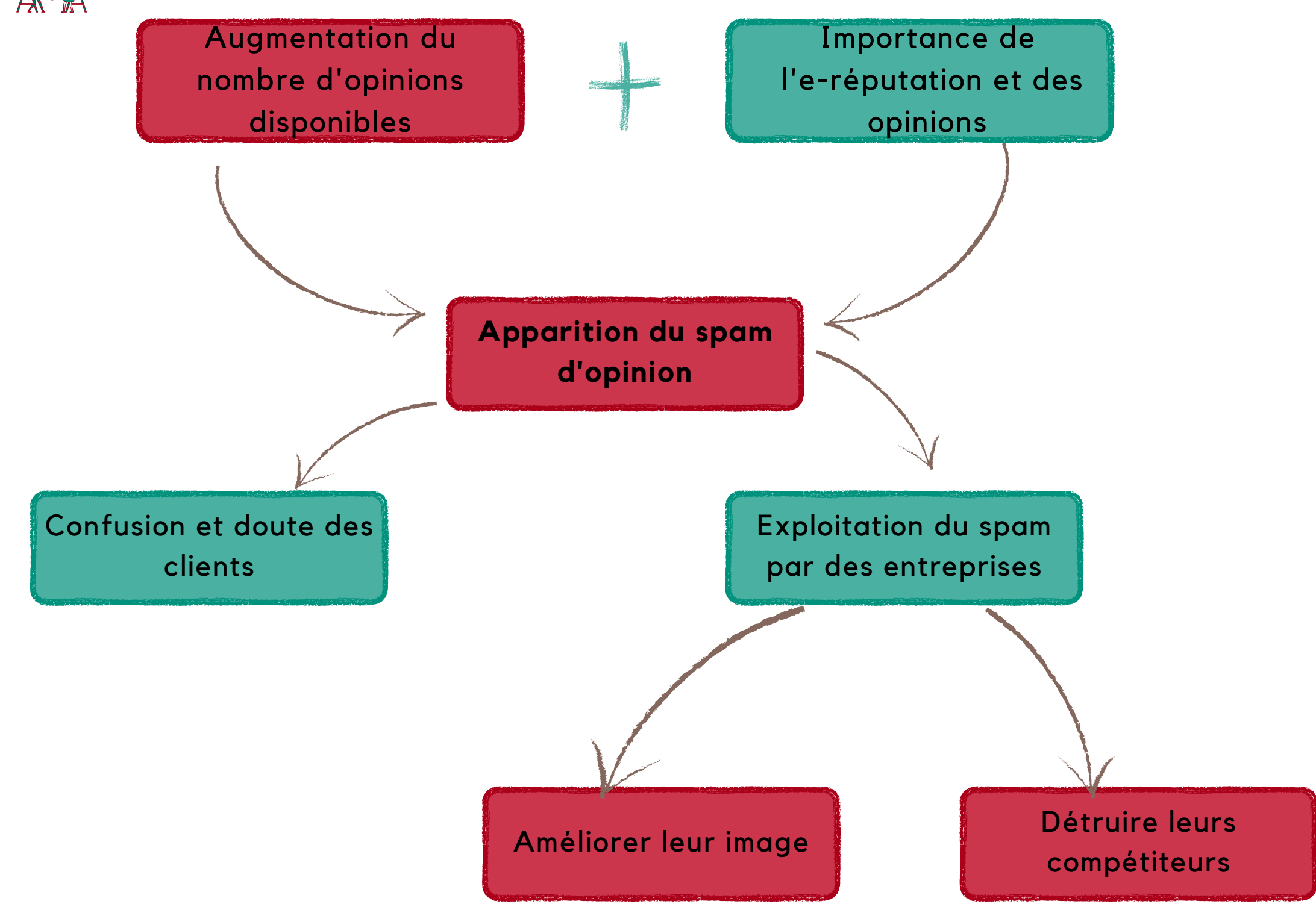
Motivation



Avantages de l'analyse des sentiments

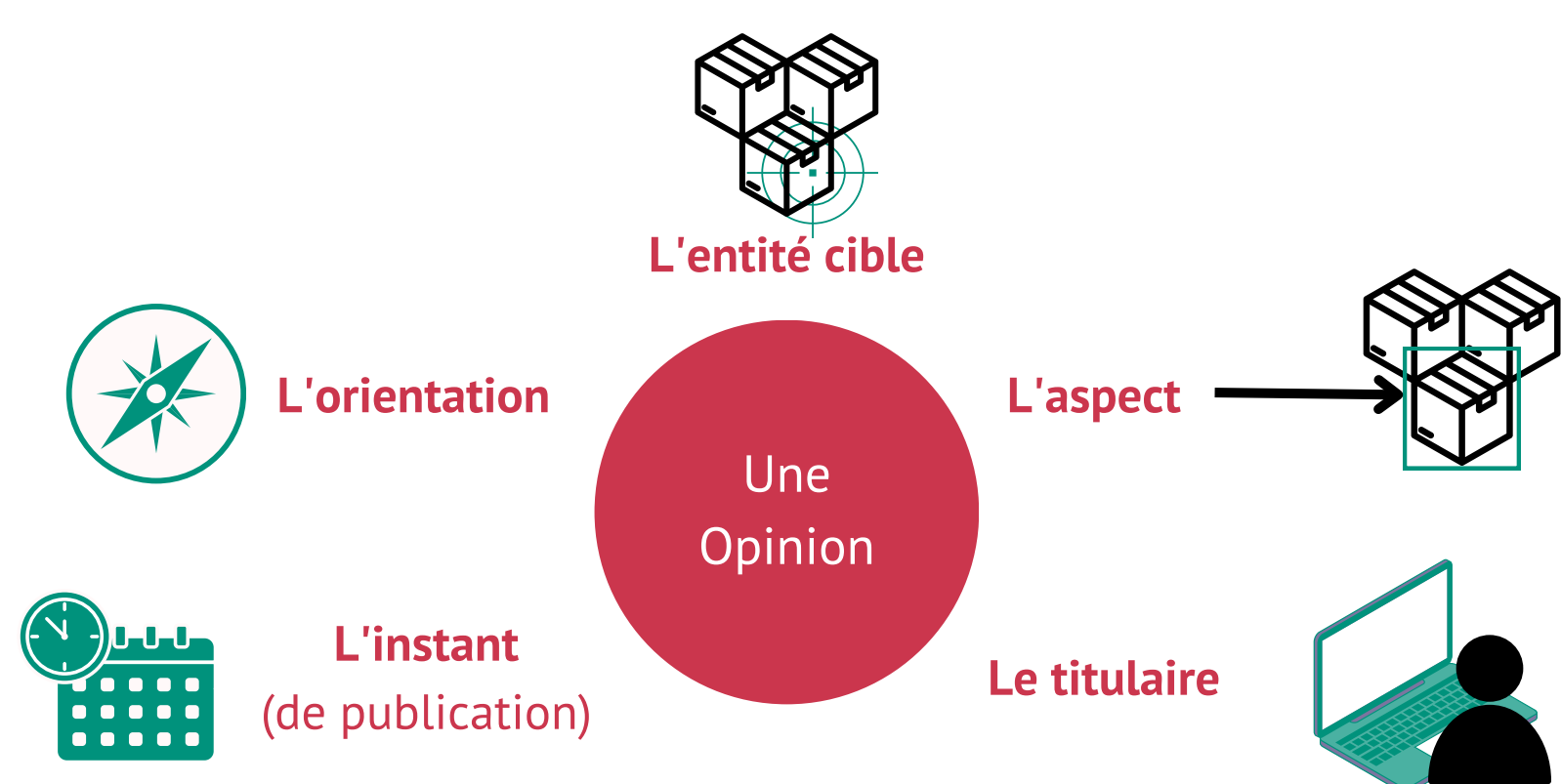


Problématique



L'opinion & l'e-réputation

Une **opinion** peut être caractérisée comme une croyance sur des sujets généralement considérés comme subjectifs ; c'est le résultat d'une émotion ou d'une interprétation des faits. Une opinion est composée des éléments suivants :



Une **opinion** est caractérisée selon plusieurs perspectives [6,7] :

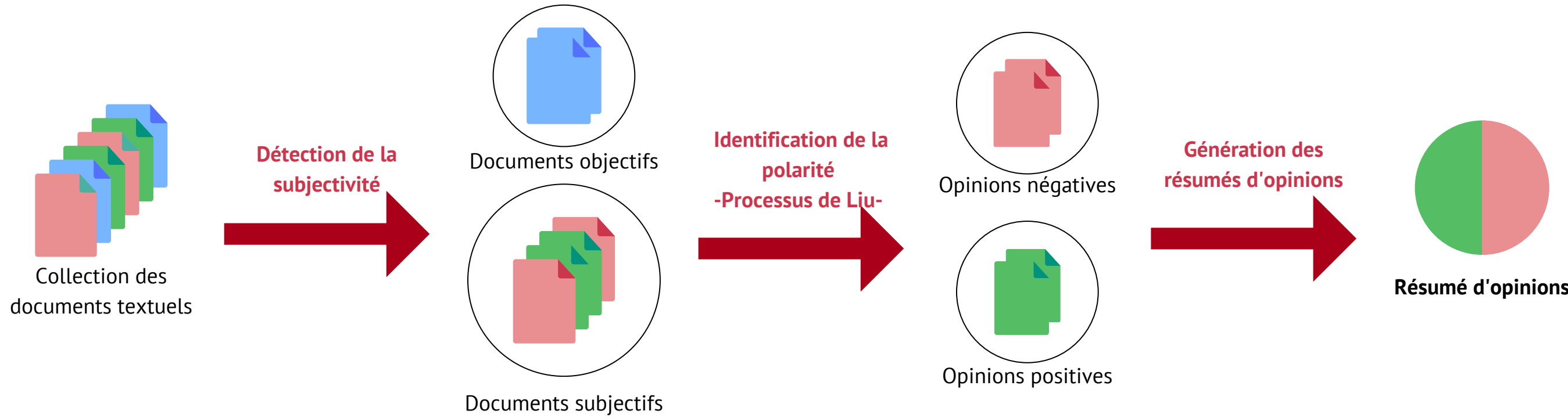
- Individuelle ou Collective
- Régulière ou Comparative
- Directe ou Indirecte
- Implicite ou Explicite

L'**e-réputation** se définit comme la notoriété ou l'image véhiculée d'une entité (entreprise, marque, particulier, produit...) sur les différents supports web.

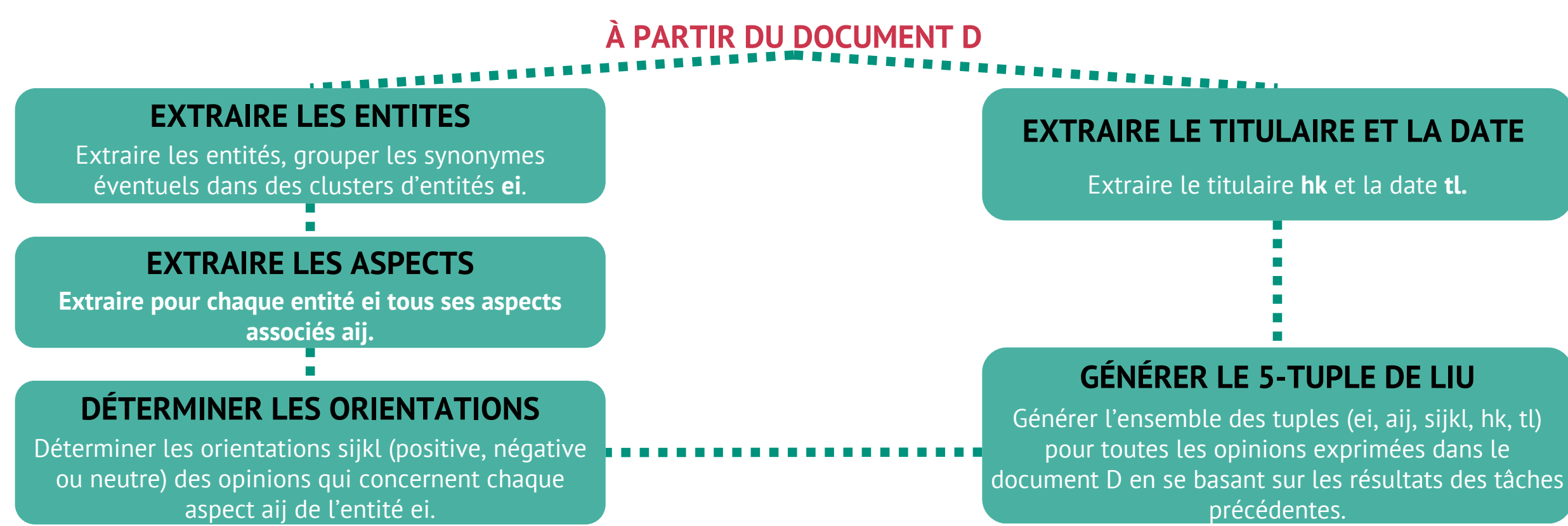
La fouille d'opinion

[5,6] définissent l'analyse des sentiments (la fouille d'opinion) comme le domaine d'étude qui analyse les opinions, les sentiments, les appréciations, les attitudes et les émotions des personnes envers les entités et leurs attributs exprimés dans un texte écrit.

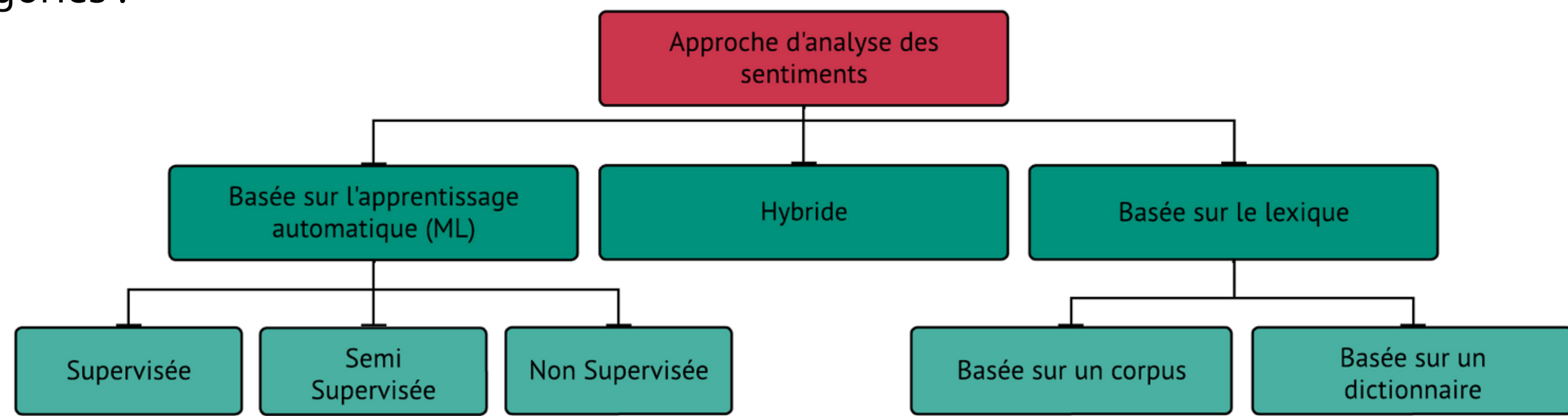
- Processus de fouille d'opinion présenté par [6,7] :



Le processus d'identification de polarité [6,7] est formalisé comme suit : transformer le texte d'un document **D** non structuré en un quintuple (**ei**, **aij**, **sijkl**, **hk**, **tl**) où **sijkl** est l'orientation de l'opinion sur l'aspect **aij** de l'entité **ei** à l'instant de son expression **tl** par la source **hk**.

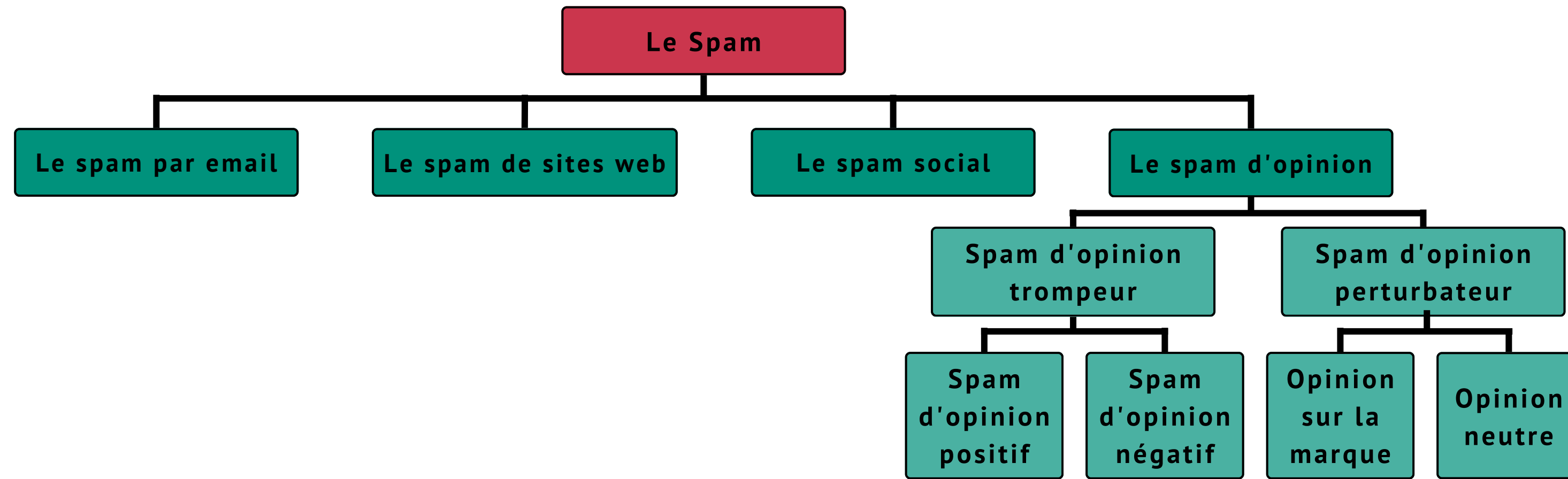


L'analyse des sentiments est un domaine multidisciplinaire qui comprend l'étude de divers domaines. Les différentes études de l'art [5,8] classifient les approches adoptées dans l'analyse des sentiments en trois catégories :



Le spam & la crédibilité

- Différentes définitions et catégories de spam [3]

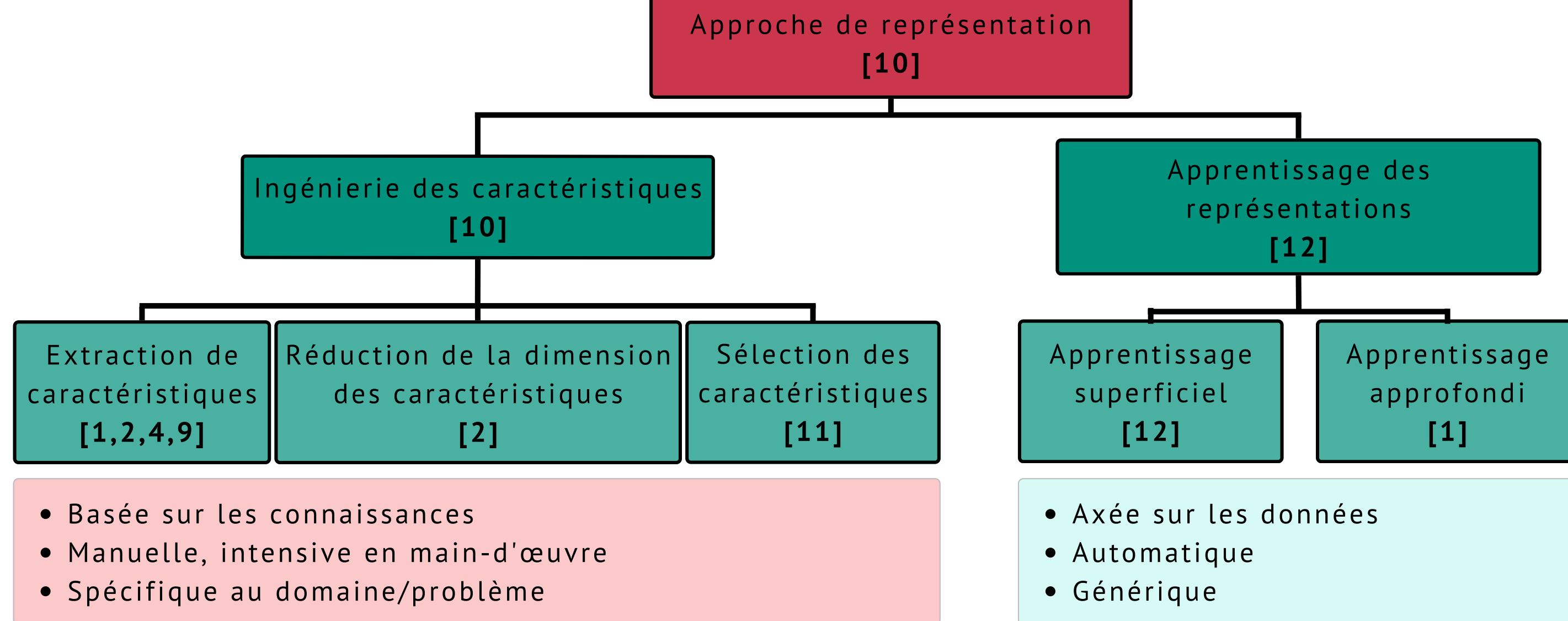


Dans le domaine de fouille d'opinion, la **crédibilité** est une qualité importante qu'une opinion doit satisfaire. Une opinion crédible reflète une expérience réellement vécue par le client qui l'exprime, elle ne doit pas contenir de propos mensongers concernant l'entité qu'elle cible.

La détection de spam

Le processus de détection de spam est similaire à celui de fouille d'opinion, on commence par la collecte des données et leur prétraitement, puis l'extraction ou la représentation des caractéristiques pour ensuite passer à la classification et enfin l'évaluation des performances.

- Classification des approches de représentation :



- Synthèse des travaux sur l'opinion spam mining

Référence	Description de l'approche	Principales Conclusions
[9] Shahariar, GM et al. (2019).	<ul style="list-style-type: none"> Approche exploratoire des approches traditionnelles et d'apprentissage des représentations des opinions. Elle est basée sur l'apprentissage actif pour étiqueter les données après le prétraitement. Les caractéristiques considérées sont : les N-grammes et TF-IDF pour les méthodes ML traditionnelles et TF-IDF et Word2Vec avec les méthodes de Deep Learning. La combinaison des Unigrammes et Naïve Bayes a donné le meilleur résultat pour les méthodes traditionnelles sur Yelp, mais a été surpassée par toutes les méthodes de Deep Learning, la combinaison de LSTM et Word2vec donnant les meilleurs résultats pour les deux datasets exploités : une exactitude de 94.56% sur Yelp et de 96.75% sur OpSpam. 	<ul style="list-style-type: none"> Pour la détection des avis spam, les algorithmes Deep Learning nécessitent beaucoup plus de données d'entraînement que les algorithmes ML traditionnels. Les algorithmes ML traditionnels tels que SVM et NB atteignent un certain seuil où l'ajout de données d'entraînement supplémentaires n'améliore pas leur précision. Des algorithmes Deep Learning tels que Word2Vec sont également utilisés afin d'obtenir de meilleures représentations vectorielles pour les mots et d'améliorer la précision des classificateurs formés avec des algorithmes ML traditionnels. ⇒ À l'avenir, la détection des spammeurs peut être introduite avec la détection de spam d'opinion.
[4] Lighthart, Alexander, r, et al. (2021).	<ul style="list-style-type: none"> Cet article explore l'auto-apprentissage, le co-apprentissage et le SVM transductif sur le dataset OpSpam. Ils ont aussi testé ces modèles avec différentes combinaisons de caractéristiques dont les Uni-grammes, les Bi-grammes, la combinaison de TF-IDF et des Bi-grammes ainsi que la combinaison de POS et les Uni-grammes. Le meilleur résultat a été obtenu en combinant les Bi-Grammes avec l'algorithme de Self-training avec une exactitude de 93% sur OpSpam ; c'est la seule approche, sur tout les modèles semi-supervisés testés, qui surpasse les modèles de classification supervisés traditionnels lorsque les données disponibles sont limitées. 	<ul style="list-style-type: none"> Le self-training peut atténuer les efforts d'étiquetage tout en conservant des performances élevées du modèle, ce qui est utile pour les scénarios où des données limitées sont disponibles ou la récupération de données étiquetées est plus coûteuse. Les modèles d'apprentissage automatique fonctionnent généralement mieux lorsque davantage de données étiquetées sont introduites dans le modèle. Les résultats de la plupart des montages expérimentaux du papier ont confirmé cette affirmation. Cependant, le modèle SVM supervisé traditionnel fonctionne légèrement mieux dans le scénario avec 10 % de données étiquetées par rapport à 20 % de données étiquetées, ce qui est une découverte remarquable. ⇒ Parmi les orientations possibles pour les recherches futures, les auteurs proposent l'exploration d'approches non supervisées pour la détection des spams d'opinion, la prise en compte des métadonnées et des connaissances réelles sur les avis et les auteurs...
[1] Andresini, Giuseppina et al. (2022).	<ul style="list-style-type: none"> Approche, appelée EUPHORIA, exploite l'apprentissage multi-vues avec plusieurs vues (chevauchées) des données afin que le modèle de classification ne manque aucun aspect important. Chaque vue est représentée par un vecteur de caractéristiques distinct qui est ensuite traité par un réseau neuronal à entrées multiples. Le modèle prend en charge trois vues de données différentes : deux liées aux caractéristiques textuelles (BERT et Word2vec) et une troisième liée aux caractéristiques comportementales (MRD, PR, ARL, RD, ARS, MRS). La combinaison des 03 vues a donné une AUC de 70.8% sur le sous-ensemble Restaurant et le dataset YelpCH, et de 81.3% sur le sous-ensemble Hôtels du même dataset. C'était la méthode la plus performante, pour les 02 datasets, en comparant les résultats avec d'autres combinaisons des vues ou même avec 2 algorithmes de l'état de l'art qui utilisent SVM. 	<ul style="list-style-type: none"> Les caractéristiques comportementales transmettent les informations les plus pertinentes nécessaires pour détecter les faux avis surpassant à la fois Word2Vec et Bert et leur combinaison également. La prise en compte à la fois des vues comportementales et de contenu surpasse toutes les autres combinaisons de vues. Le réseau de neurones multi-entrées d'EUPHORIA peut profiter de la richesse des données multi-vues contrairement au réseau de neurones mono-entrée dont les performances sont dégradées par les effets de la maldiction de la dimensionnalité. L'apprentissage de classificateurs séparés pour chaque vue est moins performant que l'apprentissage d'un seul classificateur en concaténant toutes les vues ensemble, cependant, cela ne pose pas de problème en ce qui concerne EUPHORIA en raison de la rétro-propagation dans le réseau neuronal à entrées multiples. ⇒ Parmi les directions de recherche que les auteurs aimeraient explorer, on peut citer l'apprentissage continu pour permettre au réseau de réagir en temps réel aux changements dans la distribution des données, l'apprentissage par transfert pour comprendre comment un modèle de classification s'est entraîné pour un problème peut être appliqué à un problème différent, mais connexe, ainsi que l'exploitation de l'étatement des avis pour la détection des spammeurs en introduisant de nouvelles caractéristiques comportementales.
[2] Harris, Christopher G (2022).	<ul style="list-style-type: none"> L'approche exploite une architecture LSTM combinant des caractéristiques linguistiques et comportementales. Elle comprend 3 modèles LSTM (LSTM1, LSTM2 et LSTM3) qui utilisent respectivement 12 caractéristiques linguistiques, 6 caractéristiques comportementales, et l'approche K-L Divergence. Les modèles sont combinés dans un modèle d'ensemble avec deux couches d'auto-attention. Le meilleur résultat a été observé en combinant les modèles, dépassant ainsi chacune des 3 approches indépendantes et d'autres approches de l'état de l'art basées sur la divergence K-L appliquées sur le même ensemble de données, avec une AUC de 89.30% sur le dataset Yelp2IP. 	<ul style="list-style-type: none"> L'évaluation des 12 caractéristiques linguistiques montre que les avis signalés comme spam ont moins de diversité lexicale et de contenu, utilisent moins de mots spatio-temporels, ont une longueur de mot/terme légèrement plus courte et utilisent globalement moins de clauses. L'évaluation des 6 caractéristiques comportementales montre que les fenêtres d'activité des spammeurs ont tendance à être beaucoup plus courtes, utilisent des contenus plus similaires, sont plus extrêmes et utilisent davantage les majuscules. ⇒ Les auteurs voudraient tester si les mêmes caractéristiques sont également bénéfiques pour d'autres types d'évaluations de produits. Ils aimeraient également explorer d'autres caractéristiques auxquelles ils n'avaient pas accès, telles que la géolocalisation, les cookies et les informations de navigation sur le site Web.

Défis et pistes de recherche

Les dépendances

- La dépendance au domaine, vu que les termes et expressions utilisés peuvent varier en fonction du domaine d'application. Il est donc difficile de généraliser un modèle pour différents domaines.
- La dépendance au contexte, vu que le sens d'une opinion peut changer selon le contexte qui l'entoure. De ce fait, la prise en compte du contexte est nécessaire pour l'obtention de résultats précis.

Les fautes d'orthographe

- Ces erreurs peuvent être volontaires, qui peuvent altérer la polarité d'un commentaire et donc peuvent être prises en compte lors de l'analyse des sentiments.
- Ou accidentelles, quant à elles, peuvent être ignorées si le mot d'origine peut être identifié.

Le langage figuratif

- Il est difficile pour une machine de comprendre des choses que peuvent distinguer facilement. Par exemple :
 - Les métaphores (« He is a shining star »).
 - La métonymie (« La salle a applaudi l'orchestre »).
 - L'ironie et le sarcasme.

La crédibilité de l'opinion

- La crédibilité de l'opinion à analyser, de l'auteur qui l'a publiée et de l'avis général sur une entité ciblée peut affecter.

L'acquisition de données

- L'établissement de datasets de référence de qualité est une tâche importante, afin d'assurer la qualité des modèles de détection de spam entraînés sur ces données, mais coûteuse.
- Ceci est dû, en premier lieu, à la difficulté de collecter des données étiquetées comme spam ou pas, mais aussi, au fait que ces données ne sont pas nécessairement représentatives du spam d'opinion du monde réel.

Les métadonnées

- Il est difficile d'obtenir certaines métadonnées intéressantes (!@ IP d'un utilisateur, son comportement sur le site,...).
- Le type de métadonnées disponibles peut dépendre de la source d'où les opinions sont collectées.

Les caractéristiques

- La difficulté de déterminer la meilleure combinaison de caractéristiques décrivant un contenu spam ou l'attitude d'un spammeur est qu'il est également difficile pour les humains de distinguer si un avis est un spam ou non.
- La raison est que les spammeurs fournissent des efforts considérables pour que leur style de rédaction ressemble le plus à celles d'un être humain qui a vécu l'expérience avec le produit.

Les opinions spams singletons

- Une opinion spam singleton est un commentaire dont l'auteur n'en a publié qu'un seul, il n'a donc pas laissé beaucoup d'informations utiles pour modéliser.
- Dans la plupart des études, ce type de commentaire est supprimé de l'ensemble de données, seuls quelques-uns d'entre eux ont essayé de résoudre ce problème

Références bibliographiques :

[1] Andresini, Giuseppina et al. (2022). « EUPHORIA : A neural multi-view approach to combine content and behavioral features in review spam detection ». In : Journal of Computational Mathematics and Data Science 3, p. 100036.
 [2] Harris, Christopher G (2022). « Combining Linguistic and Behavioral Clues to Detect Spam in Online Reviews ». In : 2022 IEEE International Conference on e-Business Engineering (ICEBE). IEEE, p. 38-44.
 [3] Jindal, Nitin et Bing Liu « Opinion spam and analysis ». In : Proceedings of the 2008 international conference on web search and data mining, p. 219-230.
 [4] Lighthart, Alexander, Gagay Catal et Bedir Tekinerdogan (2021). « Analyzing the effectiveness of semi-supervised learning approaches for opinion spam classification ». In : Applied Soft Computing 101, p. 107035.
 [5] Lighthart, Alexander, Gagay Catal et Bedir Tekinerdogan (2021). « Systematic reviews in sentiment analysis : a tertiary study ». In : Artificial Intelligence Review 54.7, p. 4997-5053.
 [6] Liu, Bing (2012). « Sentiment analysis and opinion mining ». In : Synthesis lectures on human language technologies 5.1, p. 1-167.

[7] Liu, Bing (2020). « Sentiment Analysis ». In : Sentiment Analysis : Mining Opinions, Sentiments, and Emotions. 2e éd. Studies in Natural Language Processing. Cambridge University Press, p. i-ii.
 [8] Ramesh, Bhavana et Charles M Weber (s. d.). « State-of-art methods used in sentiment analysis : A literature review ». In : 2022 Portland International Conference on Management of Engineering and Technology (PICMET). IEEE, p. 1-13.
 [9] Shahariar, GM et al. (2019). « Spam review detection using deep learning ». In : 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). IEEE, p. 0027-0033.
 [10] Yan, Weizhong, and Lijie Yu « On accurate and reliable anomaly detection for gas turbine combustors: A deep learning approach ». arXiv preprint arXiv:1908.09238 (2019).
 [11] Zheng, Hongwen, and Yanxia Zhang. "Feature selection for high-dimensional data in astronomy". Advances in Space Research 41, no. 12 (2008): 1860-1864.
 [12] Zhong, Guoliang, Xiao Ling, and Li-Na Wang. "From shallow feature learning to deep learning: Benefits from the width and depth of deep architectures". Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 9, no. 1 (2019): e1255.