

## **Devoir 3 - Dossier de proposition**

Aura : Une solution à la problématique de la désinformation numérique

Amir Fayçal Ghorab

Université de l'Ontario français

CMN3003 – Communication et vérité à l'ère numérique

M. Jérémie Pelletier-Gagnon

11 décembre 2025

## La problématique de la désinformation numérique

Le mensonge ne date pas de l'ère numérique. Il accompagne l'humanité depuis ses récits fondateurs. L'un des exemples les plus anciens et les plus connus se trouve dans la Bible, avec le mythe du serpent trompant Ève pour l'inciter à manger le fruit défendu. Bien qu'il s'agisse d'un récit symbolique et non d'un fait historique, ce mythe illustre déjà l'idée selon laquelle la tromperie peut être utilisée comme un outil d'influence, de manipulation et de transgression (Hibbs, 2025). Ce récit montre aussi que le mensonge ne repose pas nécessairement sur une négation totale de la vérité. Le serpent ne ment pas de manière frontale, mais déforme le réel et joue sur l'interprétation. Cette nuance est essentielle, car un mensonge efficace n'est que rarement une information entièrement fausse. Il s'agit le plus souvent d'un assemblage subtil de vrai et de faux.

Avec l'évolution des technologies de communication, le mensonge a progressivement changé d'ampleur et de nature. À la fin du 19e siècle, le *yellow journalism* marque une étape clé dans l'histoire de la désinformation dans les médias. Ce type de journalisme s'est développé dans un contexte de concurrence intense entre grands journaux américains. Pour attirer l'attention du public, les médias ont alors privilégié des titres exagérés, des récits sensationnalistes et parfois des informations déformées, en particulier autour de la guerre hispano-américaine. Cette couverture sensationnaliste a contribué à influencer l'opinion publique et à légitimer l'intervention militaire des États-Unis (U.S. Department of State, Office of the Historian, s.d.). On observe ici que la désinformation ne relève plus seulement de la tromperie individuelle, mais devient un outil stratégique au service d'intérêts politiques, économiques et idéologiques.

Si ces formes de manipulation fonctionnent, c'est en grande partie en raison des limites propres à la cognition humaine. Les sciences comportementales ont mis en lumière l'existence de nombreux biais cognitifs, qui affectent nos prises de décision et notre manière d'interpréter l'information. *L'effet de masse*, ou *bandwagon effect*, pousse par exemple à adopter une idée simplement parce qu'elle semble largement partagée, ce qui renforce artificiellement sa crédibilité. Le *biais d'autorité* amène à accorder une importance excessive à une information uniquement parce qu'elle provient d'une figure perçue comme légitime. Le *biais de réitération*, ou *biais de vérité illusoire*, conduit quant à lui à considérer une information comme vraie ou fiable en raison de sa répétition constante (The Decision Lab, s.d.). Ces mécanismes sont largement exploités dans les stratégies de désinformation. En jouant sur nos émotions, nos peurs, nos convictions et nos automatismes mentaux, certaines informations fausses deviennent particulièrement persuasives et difficiles à déconstruire.

Ces vulnérabilités cognitives se retrouvent amplifiées dans notre ère numérique. Aujourd'hui, l'information circule en continu, sans filtre centralisé, portée par les réseaux sociaux, les plateformes de partage et les communautés en ligne. La

désinformation y trouve un terrain particulièrement fertile. La pandémie de la Covid-19 en constitue un exemple frappant. De nombreuses fausses informations ont circulé massivement sur Internet à propos des vaccins, de leurs effets secondaires et de leurs prétendus dangers. L'idée selon laquelle les vaccins provoqueraient l'autisme, pourtant réfutée scientifiquement depuis longtemps, a notamment ressurgi avec force dans les discours en ligne (Kata, 2012). Cette propagation est largement favorisée par la formation de communautés numériques où les individus partagent, renforcent et légitiment mutuellement leurs croyances, créant ainsi des cercles de résonance informationnelle.

Les réseaux sociaux jouent un rôle central dans ces dynamiques, non seulement comme vecteurs de diffusion, mais aussi comme constructeurs des environnements numériques dans lesquels l'information circule. Ces plateformes reposent sur des mécanismes de design appelés *affordances*. Une affordance correspond aux possibilités d'action qu'un objet ou un système suggère intuitivement à l'utilisateur. Dans le numérique, un bouton « aimer », un partage en un clic, un commentaire instantané ou un défilement infini sont autant d'exemples d'affordances (Goldfarb, 2024). Ces dispositifs sont conçus pour être simples, rapides et engageants. Leur objectif est de réduire au maximum l'effort cognitif nécessaire pour interagir avec le contenu, afin de maintenir l'utilisateur dans un flux continu d'attention et d'action, un « flow state » (Kerdraon, 2020). Toutefois, cette logique de fluidité favorise avant tout la réactivité émotionnelle, la viralité et l'engagement, bien plus que la vérification de l'information, la nuance ou l'esprit critique.

C'est dans cette même logique que s'inscrit la suppression du bouton « dislike » sur YouTube. Cette décision illustre le fait que les plateformes conçoivent leurs affordances non pas principalement pour créer un écosystème favorisant l'honnêteté et la vérité, mais pour offrir une expérience utilisateur sans résistance ni adversité. En retirant cette possibilité d'exprimer publiquement un désaccord négatif, YouTube a cherché à décourager le harcèlement, les campagnes de dénigrement et les attaques coordonnées visant certains créateurs ou certaines vidéos (BBC Newsbeat, 2021). Toutefois, ce choix a aussi eu pour effet de limiter la capacité collective des utilisateurs à signaler rapidement un contenu problématique, trompeur ou de mauvaise qualité. La disparition du désaccord visible réduit ainsi une forme de friction sociale pourtant essentielle au débat, à l'autorégulation communautaire et à la mise à distance critique de l'information.

Un autre enjeu majeur qui s'ajoute aujourd'hui à la désinformation est celui de l'intelligence artificielle sur les réseaux sociaux. De plus en plus d'utilisateurs accusent dans les commentaires certains contenus d'être générés par IA, parfois à tort. Cette méfiance est devenue si forte que certains créateurs doivent désormais se justifier publiquement. L'artiste Jonnybush, par exemple, a dû publier un message sur Instagram

pour prouver que son travail n'avait pas été produit par intelligence artificielle (johnnybush, 2025). Cette situation illustre un climat de scepticisme extrême, où la frontière entre le réel et l'artificiel devient floue. Les utilisateurs ne doutent plus seulement du contenu, mais aussi de son origine. À mesure que les technologies numériques se développent et que les outils d'IA deviennent plus accessibles et plus performants, le risque de tromperie augmente, tout comme la peur d'être manipulé. Cela renforce encore davantage le besoin de mécanismes capables d'apporter de la clarté, de la transparence et des repères visibles dans l'environnement numérique.

Au final, face à un écosystème numérique centré sur l'engagement, la rapidité et la performance algorithmique, les plateformes actuelles montrent des limites structurelles dans la manière dont elles traitent la désinformation. Leurs logiques de fonctionnement sont d'abord orientées vers la visibilité, la rétention et la monétisation, bien avant la rigueur informationnelle. Dans ce contexte, corriger les outils existants ne suffit plus. Il devient nécessaire de concevoir une plateforme nouvelle, pensée dès sa conception autour d'une éthique de l'information. Une plateforme « built from the ground up », avec des affordances intuitives qui ne sacrifient pas la rigueur, la véracité et l'authenticité. C'est dans cette logique qu'a été pensé Aura. La section suivante en propose la description détaillée.

## Comment fonctionne Aura ?



*Logo de la plateforme*

Aura est une application de réseau social visuel basée sur un format vertical de posts (vidéos et images) scrollables. Elle reprend les codes d'interaction des plateformes sociales actuelles. L'utilisateur interagit avec le contenu à l'aide des mêmes affordances qui lui sont familières : scroller, aimer, ne pas aimer, commenter et partager.

### 1. Le fil de contenu

L'interface principale de Aura repose sur un fil de contenus continus affichés en plein écran. Chaque publication apparaît sous forme de vidéo ou d'image. Sous chaque publication se trouvent les icônes d'interaction :

- Bouton de like.

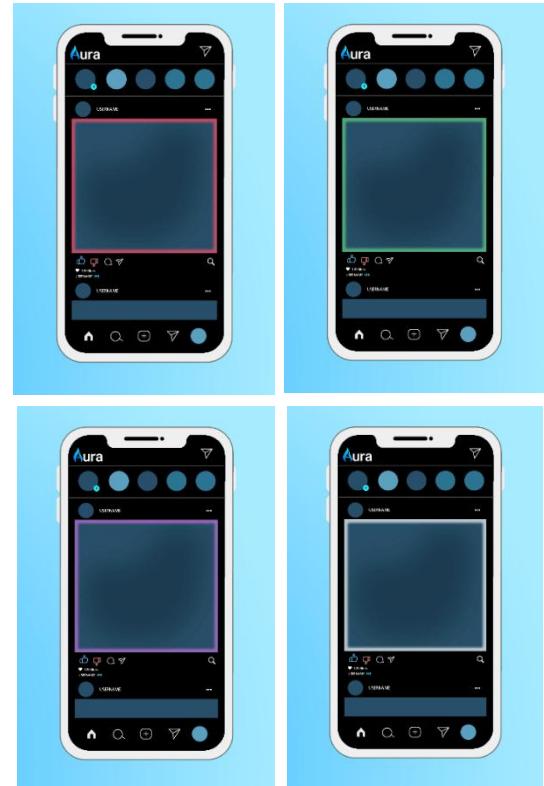
- Bouton de dislike.
- Bouton de commentaire.
- Bouton de partage.
- Il y a également une icône de loupe, qui sert à consulter les sources d'un post.

## 2. Le système des Auras visuelles

Chaque publication est entourée d'un cadre coloré appelé Aura. Ce cadre indique la nature du contenu selon quatre couleurs :

- **Violet** : contenu de divertissement.
- **Vert** : contenu informatif avec sources visibles.
- **Rouge** : contenu signalé comme désinformation probable.
- **Gris** : contenu ambigu, récent ou en cours d'évaluation.

L'Aura apparaît directement autour du contenu, sans élément textuel supplémentaire.



## 3. Le système de vote sur le post et sur les sources

Les utilisateurs peuvent :

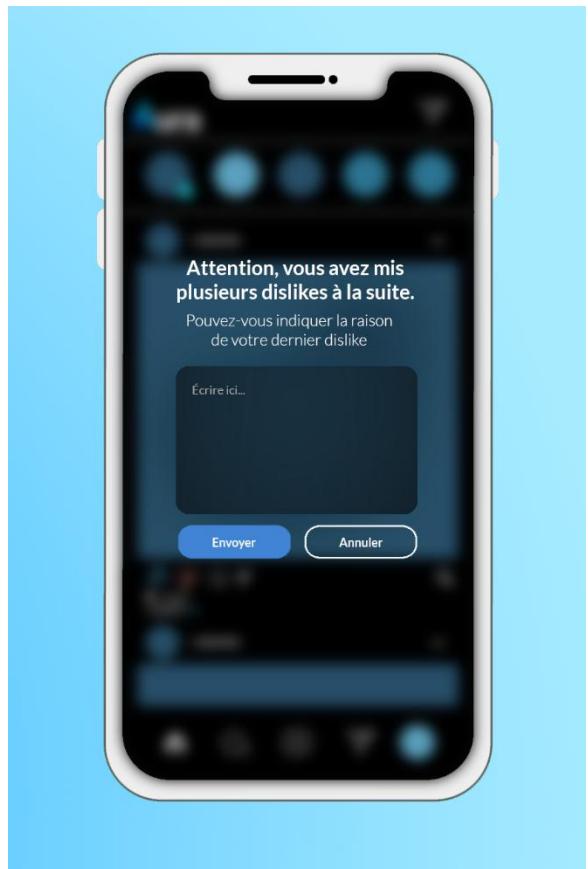
- Disliker un post dans son ensemble,
- Disliker individuellement chacune des sources associées à un contenu informatif.

Lorsqu'un contenu est classé comme informatif, les sources sont affichées sous la publication. Chaque source dispose de son propre bouton de like et de dislike. L'Aura du post évolue automatiquement en fonction des votes appliqués aux sources.

La plateforme intègre aussi des mécanismes de gestion des abus liés au dislike. Lorsque qu'un utilisateur attribue un nombre important de dislikes sur des posts ou des sources dans un intervalle de temps restreint, une fenêtre de justification apparaît. L'utilisateur doit



alors saisir une courte explication pour continuer à utiliser la fonction dislike. Cette justification est visible par les Fact Checkers.



Les Fact Checkers peuvent, à partir de ces informations, appliquer un statut de cooldown à un utilisateur. Pendant cette période, généralement de deux à trois jours, le bouton dislike est désactivé pour ce compte, tandis que l'accès au reste de la plateforme est maintenu. Si un utilisateur, après un premier avertissement, continue à enchaîner les dislikes de manière répétée, un blocage temporaire des likes et des dislikes est appliqué automatiquement. L'utilisateur ne peut plus voter pendant une durée déterminée.

Les Fact Checkers disposent également d'un outil de filtrage permettant d'afficher les posts ayant reçu un nombre élevé de dislikes et dont l'Aura est passée au rouge. Ils peuvent vérifier les sources associées à ces contenus et, si elles sont jugées valides,

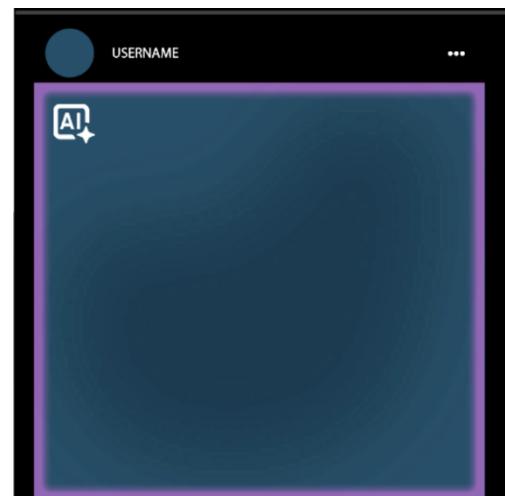
décider de maintenir ou de rétablir une Aura verte malgré le ratio de dislikes élevé.

#### 4. Le badge IA

Les créateurs peuvent déclarer qu'un contenu a été généré ou modifié par intelligence artificielle. Dans ce cas, un badge « AI » apparaît sur la publication. Si l'IA n'est pas déclarée, les utilisateurs peuvent signaler le contenu. Après vérification, le badge est ajouté automatiquement.

#### 5. Le système de signalement

Chaque publication peut être signalée selon trois catégories :



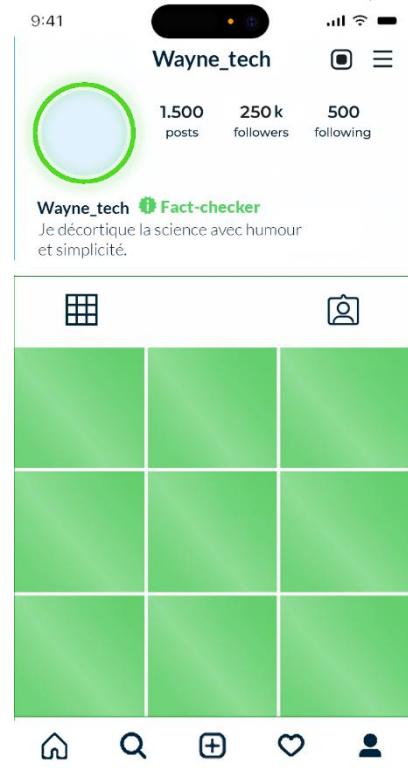
- Contenu généré par IA non déclaré,
- Contenu non fiable,
- Contenu potentiellement dangereux.

## 6. Le statut de Fact Checker

Certains utilisateurs peuvent obtenir un badge « Fact Checker ». Ce statut est destiné aux créateurs de contenus engagés dans la vulgarisation de connaissances, la diffusion d'informations factuelles, ainsi que dans la pédagogie et l'éducation. Il est attribué aux utilisateurs qui produisent régulièrement du contenu à visée scientifique, informative ou éducative, et dont la démarche repose sur l'usage systématique de sources vérifiables.

Les Fact Checkers disposent de :

- Droits de vote plus « fort » sur les sources.
- Capacités de signalement prioritaires.
- Fonctionnalité de filtre pour les publications aux sources fortement dislikedées.
- Droit d'intervenir sur le statut d'une publication en proposant par exemple de maintenir ou de rétablir une Aura verte après vérification, même en présence d'un grand nombre de dislikes.



## Comment Aura répond à la problématique ?

Aura a été pensée comme une réponse directe aux problèmes soulevés dans la problématique sur la désinformation. Là où le mensonge se diffuse aujourd'hui à grande vitesse à cause du numérique, de la viralité et des réseaux sociaux, Aura tente de réintroduire du contrôle, de la lisibilité et de l'esprit critique directement dans l'interface grâce à ces affordances intuitives, familières mais utilisées de manière éthique.

Dans la problématique, il a été montré que les plateformes actuelles utilisent des affordances simples et rapides pour favoriser l'engagement, souvent au détriment de la vérité. Aura reprend volontairement ces mêmes affordances, comme le scroll, le like, le partage ou le commentaire, afin de ne pas perturber les habitudes des utilisateurs. La différence est que ces affordances ne servent plus uniquement à faire consommer du contenu, mais aussi à l'évaluer. L'interface ne sert donc plus seulement à diffuser, mais aussi à distinguer l'information du divertissement et de la désinformation.

Le système des Auras visuelles répond directement au problème du sensationnalisme et de la confusion entre information et divertissement. Aujourd'hui, beaucoup de

fausses informations circulent sous des formats ludiques, amusants ou choquants, ce qui rend leur nature difficile à identifier. Avec les codes de couleur, l'utilisateur sait immédiatement à quoi il est exposé, sans avoir besoin de lire ou d'analyser longuement. Cela permet de limiter l'impact émotionnel brut qui favorise la désinformation tout en évitant de surcharger cognitivement les utilisateurs.

Le système de dislike, quant à lui, répond à deux problèmes évoqués dans la problématique. D'un côté, certaines plateformes ont supprimé le désaccord visible pour éviter le harcèlement. De l'autre, le dislike libre peut être utilisé de manière abusive. Aura ne supprime pas le désaccord, mais l'encadre. Le fait de modérer les dislikes et d'obliger l'utilisateur à se justifier après plusieurs dislikes permet de limiter les comportements impulsifs, les trolls et les attaques coordonnées. Le dislike reste un outil d'esprit critique, mais il ne peut plus être utilisé automatiquement et sans réflexion.

Aura agit aussi directement sur plusieurs biais cognitifs. Le biais de réitération, par exemple, repose sur le fait que plus une information est vue, plus elle devient crédible. Avec Aura, un contenu qui passe en rouge reste visible, mais il est clairement signalé comme problématique. Le simple fait de le voir plusieurs fois ne le rend donc plus crédible automatiquement. Le biais d'autorité est aussi limité. Un créateur populaire ou perçu comme expert ne peut plus s'appuyer uniquement sur sa notoriété. Si ses sources sont jugées mauvaises, son contenu peut être signalé comme douteux malgré sa réputation.

La plateforme répond également au problème des effets de masse. Sur les réseaux sociaux actuels, ce qui est le plus liké ou le plus partagé est souvent perçu comme vrai. Avec l'intervention des Fact Checkers, une information peut rester en Aura verte même si elle reçoit beaucoup de dislikes. Cela permet d'éviter que la vérité soit décidée uniquement par le nombre ou par la pression d'un groupe.

Enfin, Aura contrecarre la marginalisation de l'information rigoureuse et de la vulgarisation. En donnant un statut spécifique aux créateurs engagés dans l'information réelle, la science et l'éducation, la plateforme met en valeur ces contenus au lieu de les reléguer au second plan. Les Fact Checkers ne sont pas une autorité extérieure, mais des membres reconnus de la communauté.

Enfin, Aura répond au dernier problème soulevé dans la problématique, celui de la présence de contenus générés par IA. Le badge IA est la solution, il minimise la confusion entre contenus réels et contenus générés artificiellement. En rendant l'usage de l'intelligence artificielle visible, Aura permet à l'utilisateur de savoir à quoi il a affaire. Il peut alors interpréter l'information avec plus de recul.

Ainsi, Aura ne combat pas la désinformation uniquement par la suppression de contenus, mais par la manière dont l'interface est pensée. Elle agit sur les biais, sur les

usages, sur les comportements et sur la manière dont l'information est perçue. C'est cette approche globale qui permet à Aura de répondre de façon cohérente aux problèmes exposés dans la problématique.

# Bibliographie

BBC Newsbeat. (2021, novembre 12). *YouTube removing dislike “discourages trolls” but “unhelpful for users”*. BBC. <https://www.bbc.com/news/newsbeat-59264070>

Bush, J. [@jonnybush]. (2025, septembre 16). *[Contenu du post jusqu’aux 20 premiers mots]* [Instagram post]. Instagram.

[https://www.instagram.com/p/DOp4OX9CsmL/?img\\_index=5](https://www.instagram.com/p/DOp4OX9CsmL/?img_index=5)

Goldfarb, E. (2024, septembre 30). *Qu’est-ce que l’affordance et comment l’utiliser ?*

Agence Churchill. <https://www.agence-churchill.fr/blog/quest-ce-que-laffordance/>

Hibbs, P. T. (2025, décembre 1). *The greatest lie in the history of humanity*. Westminster Theological Seminary. <https://wm.wts.edu/read/the-greatest-lie-in-the-history-of-humanity>

Kata, A. (2012). Anti-vaccine activists, Web 2.0, and the postmodern paradigm – An overview of tactics and tropes used online by the anti-vaccination movement. *Vaccine*, 30(25), 3778–3789. <https://doi.org/10.1016/j.vaccine.2011.11.112>

Kerdraon, E. (2020, November 27). *Psychologie positive : le flow de Csikszentmihalyi*. Gordon Crossings. <https://www.gordon-crossings.com/psychologie-positive-le-flow-de-csikszentmihalyi/>

The Decision Lab. (s.d.). *List of cognitive biases and heuristics*. <https://thedecisionlab.com/biases>

U.S. Department of State, Office of the Historian. (s.d.). *U.S. Diplomacy and Yellow Journalism, 1895–1898*. <https://history.state.gov/milestones/1866-1898/yellow-journalism>