L'accessibilité numérique : Auditer les images

Laura Noreskal[1], Guillaume Feuilloley[2]

[1] Sogeti part of Capgemini, SogetiLabs, Issy-les-Moulineaux, France
[2] Sogeti part of Capgemini, SogetiLabs, Rennes, France

laura.noreskal@sogeti.com guillaume.feuilloley@sogeti.com

Résumé

Les images occupent une place centrale sur les sites et applications web. En effet, elles y assurent des fonctions d'organisation, de transmission d'informations et d'enrichissement visuel. Elles jouent ainsi un rôle déterminant dans la compréhension des contenus. Toutefois, leur utilisation soulève des enjeux majeurs d'accessibilité pour les personnes déficientes visuelles, qui ne peuvent percevoir ces éléments graphiques. Depuis l'émergence du web, des référentiels d'accessibilité ont été établis, notamment le RGAA, afin de définir des normes garantissant l'accès équitable à l'information. Dans ce contexte, nous présentons un outil d'audit automatique des images basé sur de l'intelligence artificielle et conforme aux exigences du RGAA. Cet outil vise à aider les concepteurs de sites web à évaluer et améliorer l'accessibilité de leurs contenus visuels, de manière à permettre aux utilisateurs déficients visuels d'accéder pleinement aux informations mises en ligne.

Mots-clés

IA, Accessibilité, RGAA, Traitement des images, LLM

Abstract

Images play a central role in websites and web applications, where they serve functions of organization, information transmission, and visual enhancement. They therefore contribute significantly to content comprehension. However, their use raises major accessibility challenges for visually impaired individuals, who cannot perceive these graphic elements. Since the emergence of the web, accessibility guidelines have been developed, most notably the RGAA to establish standards that ensure equitable access to information. In this context, we introduce an automatic image auditing tool based on artificial intelligence and compliant with RGAA accessibility requirements. This tool is designed to help website developers assess and improve the accessibility of their visual content, thereby enabling visually impaired users to fully access the information provided online.

Keywords

AI, Accessibility, RGAA, Computer vision, LLM

1 Introduction

Les images véhiculent souvent un message complémentaire au texte ou améliorent l'apparence d'une page. Pour beaucoup, la distinction entre ces deux fonctions n'est pas déterminante. Toutefois, cette distinction reste essentielle pour les personnes déficientes visuelles qui utilisent des lecteurs d'écran comme principal moyen d'accéder aux sites web. En effet, une image qui transmet une information doit être accessible, contrairement à une image décorative. Quel que soit le lecteur d'écran, celui-ci doit pouvoir lire un texte alternatif pour chaque image qui véhicule une information. Le thème « images » du RGAA (Référentiel Général d'Amélioration de l'Accessibilité[1]) permet d'analyser les deux notions d'images informatives et d'images décoratives. Une image informative est une image qui transmet une information nécessaire à la compréhension du contenu auquel elle est associée, tandis qu'une image décorative ne fournit aucune information essentielle à la compréhension. Le RGAA exige que toutes les images informatives présentes sur un site web ou une application soient accompagnées d'une alternative textuelle (ou « attribut alt ») décrivant le contenu et la fonction de l'image. Cette alternative textuelle doit être pertinente et concise afin de permettre aux utilisateurs de lecteurs d'écran de comprendre l'information véhiculée par l'image. En revanche, il est recommandé de ne pas utiliser de texte alternatif pour les images décoratives, car elles ne transmettent aucune information utile à l'utilisateur. Dans ce cas, ces images doivent contenir une alternative textuelle vide (par exemple : alt=""). Cela permet de réduire la charge cognitive des personnes utilisant des lecteurs d'écran en évitant la lecture de descriptions inutiles.

2 Etat de l'art : Accessibilité et Images

Ces dernières années, de nombreuses études sur l'accessibilité des images ont été menées pour démontrer la nécessité de se conformer aux normes établies par les WCAG (Web Content Accessibility Guidelines[2] ou Règles pour l'accessibilité des contenus Web) et le RGAA. Les travaux examinés mettent collectivement en évidence les défis multiples ainsi que les solutions émergentes dans le domaine de l'accessibilité des images, mais aucun n'aborde explicitement la distinction entre images informatives et

1. https://accessibilite.numerique.gouv.fr
2. https://www.w3.org/WAI/standards-guidelines/wcag/fr

images décoratives, un concept clé des standards d'accessibilité tels que les WCAG et le RGAA. L'étude de [7] souligne que les approches actuelles restent limitées, reposant principalement sur des descriptions manuelles pour certains types d'images, comme les cartes et les graphiques, et plaide en faveur d'une feuille de route intégrant l'automatisation et des stratégies d'interaction multimodale. Cette analyse rejoint les préoccupations exprimées par [8], qui mettent en avant la complexité des figures scientifiques et recommandent une stricte application des directives WCAG, en particulier en matière de contraste, de différenciation par motifs et de texte alternatif détaillé, afin de garantir une communication scientifique inclusive. Dans la même lignée, [9] révèlent l'absence quasi totale de directives d'accessibilité dans les revues en libre accès, exposant un manque important de standardisation malgré l'existence de cadres normatifs. De leur côté, [1] se penchent sur le domaine éducatif et insistent sur la nécessité de former les créateurs de contenus afin d'assurer des conditions d'apprentissage équitables pour les étudiants déficients visuels. Ces constats font écho à des contributions plus techniques : [3] montrent que l'intégration d'informations contextuelles améliore considérablement la pertinence des descriptions générées par l'IA, tandis que [4] identifient des limites persistantes dans les systèmes automatisés de génération de légendes lorsqu'ils traitent des images complexes. [5] proposent de dépasser le simple texte alternatif en intégrant des métadonnées et des indices contextuels pour enrichir la représentation visuelle. Enfin, les avancées récentes telles que les techniques de [2] permettent d'optimiser la qualité des légendes générées. Dans leur ensemble, ces études convergent vers une exigence essentielle : combiner conformité normative, automatisation intelligente et conception centrée utilisateur afin d'atteindre une accessibilité des images véritablement inclusive. Cependant, malgré toutes ces recherches, il n'existe pas, à notre connaissance, dans la littérature actuelle, de travaux analysant explicitement la distinction entre images porteuses d'information et images décoratives selon la définition proposée par le RGAA. Cette absence de cadre théorique établi nous a conduits à nous appuyer directement sur ce référentiel pour concevoir une méthodologie adaptée. Ainsi, le RGAA ne constitue pas seulement une recommandation normative dans notre travail, mais également un point d'ancrage méthodologique en l'absence de références académiques dédiées.

3 Méthodologie

Nous proposons ainsi de contribuer aux recherches en cours en présentant notre travail sur la distinction entre images informatives et images décoratives. Cette recherche a été menée dans l'objectif de développer un outil automatisé d'audit de l'accessibilité des images. Notre objectif est de permettre aux concepteurs de sites web d'auditer leurs images en se basant sur les directives du RGAA. Au cours de cet audit, les images sont analysées automatiquement, ce qui permet de détecter la présence d'images informatives sur les sites web et d'identifier celles qui nécessitent un texte alternatif. En utilisant notre outil, l'accessibilité des sites peut être significativement améliorée grâce à la classification automatique de ces deux types d'images. Pour de nombreuses organisations, l'amélioration de l'accessibilité est perçue comme coûteuse en temps et en ressources ; avec notre solution, nous souhaitons démontrer que renforcer l'accessibilité des images peut être réalisé de manière efficace et avec un effort minimal. Pour développer notre outil automatisé d'audit de l'accessibilité des images, nous avons d'abord analysé la manière dont les audits RGAA sont généralement effectués. Les audits RGAA se déroulent en plusieurs étapes. Dans un premier temps, l'auditeur sélectionne un ensemble représentatif de pages à analyser, généralement celles les plus consultées. Ensuite, les 106 critères du RGAA sont vérifiés manuellement sur ces pages. Les résultats de cette évaluation manuelle sont présentés sous la forme d'un taux global de conformité. L'auditeur rédige ensuite un rapport détaillé comprenant le score global, la liste des non-conformités et des recommandations pour améliorer l'accessibilité. Dans notre projet, nous souhaitons collecter l'ensemble des images du site et fournir un rapport complet incluant le pourcentage d'images non conformes (images décoratives avec texte alternatif et images informatives sans texte alternatif), accompagné de recommandations de correction. Après avoir défini les éléments à analyser et à produire, nous nous sommes concentrés sur la clarification des concepts clés essentiels à la compréhension du problème. Nous avons établi que, en plus de distinguer les images informatives et décoratives, il est nécessaire de définir ce que sont le texte alternatif, le texte adjacent et le texte interne. Le texte alternatif correspond à une description textuelle qui transmet l'information représentée par l'image ; il est contenu dans l'attribut « alt » et lu par les lecteurs d'écran. Le texte adjacent désigne le texte situé à proximité de l'image à laquelle il se rapporte. Enfin, le texte interne désigne le texte intégré directement dans l'image.

3.1 Le traitement des images

La méthodologie adoptée dans ce travail ne s'appuie pas directement sur un corpus de littérature scientifique existant, en raison du manque de recherches traitant spécifiquement de notre problématique. Elle résulte plutôt d'une réflexion approfondie menée au sein de l'équipe. Plus précisément, notre démarche repose sur le postulat initial selon lequel il existe différents types d'images, et que ces types ne contribuent pas de manière équivalente à la transmission de l'information. Certains types d'images sont en effet intrinsèquement plus porteurs de sens que d'autres. À partir de cette hypothèse, nous avons fait le choix de catégoriser les images. Ce principe de catégorisation constitue le point de départ de notre approche et a orienté l'ensemble des décisions méthodologiques ultérieures.

Dans des travaux précédents [6], nous avons élaboré une chaîne de traitement d'images intégrant plusieurs modèles d'IA, notamment YOLOv8 pour classer les images en cinq catégories (Texte, Illustration, Diagramme, Logo et Pictogramme), Florence-2 pour la description d'images, EasyOCR pour l'extraction du texte interne, et BERT pour le

calcul de similarité. Dans ce travail, nous proposons d'utiliser un Large Language Model (LLM) afin d'effectuer toutes ces tâches au sein d'un cadre unifié.

Pour analyser les images, nous avons choisi de traiter quatre catégories d'images Diagrammes, Illustrations, Logos et Pictogrammes, au lieu des cinq catégories initiales (Texte, Illustration, Diagramme, Logo et Pictogramme). La catégorie Texte a été fusionnée avec Illustration. Pour chaque catégorie, nous avons mené une analyse approfondie et développé des arbres de décision afin de faciliter la détection des images informatives.

Diagrammes : cette catégorie comprend les graphiques, schémas et tableaux, qui représentent des ensembles complexes d'informations. Leur traitement dépend de la présence de texte adjacent. Si le diagramme n'est pas accompagné de texte adjacent, il est automatiquement classé comme informatif. Si un texte adjacent est présent, nous calculons la similarité entre le texte interne contenu dans le diagramme et ce texte adjacent. Une similarité élevée indique que l'image est décorative ; une similarité faible suggère qu'elle est informative, car certaines données ne sont peut-être pas entièrement retranscrites dans le texte adjacent.

Illustrations : cette catégorie regroupe les images de personnes, objets, paysages ou scènes réelles ou imaginaires ne relevant pas des pictogrammes ou des logos. Le traitement dépend également du texte adjacent. Si une illustration ne possède pas de texte adjacent, nous recherchons la présence de texte interne. Si aucun texte interne n'est détecté, l'image est automatiquement considérée comme décorative ; s'il y en a, elle est classée comme informative. Lorsqu'un texte adjacent existe, nous calculons la similarité entre ce texte et une description automatique de l'image. Une similarité élevée indique que l'image est décorative ; une similarité faible indique qu'elle est informative. La description générée sert d'équivalent textuel à l'image, permettant une comparaison avec le texte adjacent.

Pictogrammes : cette catégorie regroupe des éléments graphiques minimalistes composés d'une ou plusieurs formes, utilisés pour communiquer sans recours à la langue parlée, représentant objets, actions, concepts ou idées de manière simple et immédiatement reconnaissable. Contrairement aux logos, les pictogrammes ne représentent pas des marques ou organisations. Leur traitement repose uniquement sur la présence de texte adjacent : si un texte adjacent est présent, le pictogramme est considéré comme décoratif ; sinon, il est classé comme informatif. Nous posons l'hypothèse que le texte adjacent fournit un contexte suffisant pour déterminer si un pictogramme véhicule une information essentielle.

3.2 L'extension web

Nous avons développé une extension web qui, après avoir extrait les images des sites web et pris des captures d'écran de leurs contextes environnants, exécute les différentes étapes du pipeline de traitement.

Analyse de la catégorie d'image : La première étape consiste à identifier le type d'image : Logo, Illustration, Pictogramme ou Diagramme. Cette classification détermine les

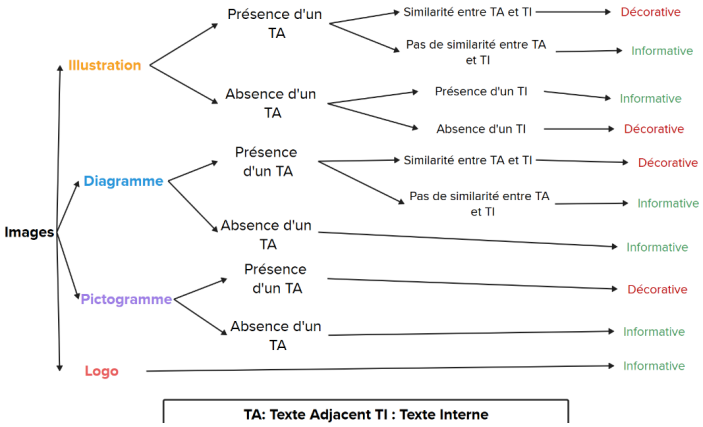

FIGURE 1 – Méthodologie de classification des images

traitements suivants, car chaque catégorie implique un traitement différent. **Extraction du texte interne** : Ensuite, le système extrait le texte présent à l'intérieur de l'image à l'aide d'un OCR. Cette étape est importante, car le texte interne constitue souvent un indicateur que l'image contient une information essentielle. **Extraction du texte adjacent** : Le système récupère ensuite le texte environnant (légendes, titres ou paragraphes proches). Ce contexte textuel aide à déterminer la pertinence de l'image et à identifier si le texte complète ou explique le contenu visuel. **Génération d'une description de l'image** : L'image est ensuite convertie en une description textuelle. Cette étape est cruciale, car comparer directement une image avec du texte n'est pas possible. En l'absence de texte interne, la description devient alors la référence principale pour la comparaison. **Comparaison des textes** : Le système compare le texte adjacent avec la description générée et/ou le texte interne. Si ces éléments sont cohérents, l'image est probablement informative ; dans le cas contraire, elle est susceptible d'être décorative. **Classification binaire (Décorative vs Informative)** : Sur la base des étapes précédentes, l'image est classée comme informative (si elle contient ou transmet une information essentielle) ou décorative (si elle est purement esthétique). **Vérification de conformité RGAA** : Le système vérifie ensuite les règles d'accessibilité. Les images informatives doivent posséder un texte alternatif. Les images décoratives ne doivent pas avoir de texte alternatif. Le statut de conformité est alors affiché (par exemple : « Conforme »). **Justification de la classification** : Enfin, le système fournit une explication justifiant la classification attribuée à l'image.

3.3 Sélection du LLM

Pour sélectionner un LLM adapté à notre outil, nous avons procédé à deux expériences. La première visait à identifier quel LLM pouvait générer les descriptions d'images les plus précises. Nous avons évalué neuf LLM : Claude-3.7-Sonnet, Nova, Titan Express, Command, DeepSeek, LLaMA, Mistral, GPT-5 et Gemini. Un jeu de données de trente images a été constitué, comprenant dix diagrammes, dix illustrations et dix pictogrammes. Pour

TABLE 1 – Résultats pour le meilleur LLM par diagramme

Img	Moy. max	Moy.	Méd. max	Méd.
diagram1	GPT-5.1	1.000	GPT-5.1	1.000
diagram2	Claude-3.7	0.821	Claude-3.7	0.750
diagram3	Claude-3.7	0.321	Claude-3.7	0.250
diagram4	Titan Expr.	0.500	Titan Expr.	0.500
diagram5	Titan Expr.	0.571	Titan Expr.	0.500
diagram6	Claude-3.7	1.000	Claude-3.7	1.000
diagram7	Command	0.500	Command	0.500
diagram8	Claude-3.7	0.536	Claude-3.7	0.500
diagram9	Claude-3.7	0.714	Claude-3.7	0.750
diagram10	Claude-3.7	0.750	Claude-3.7	0.750

TABLE 2 – Résultats pour le meilleur LLM par illustration

Img	Moy. max	Moy.	Méd. max	Méd.
illustration1	Claude-3.7	1.000	Claude-3.7	1.000
illustration2	Claude-3.7	1.000	Claude-3.7	1.000
illustration3	Claude-3.7	1.000	Claude-3.7	1.000
illustration4	Claude-3.7	0.000	Claude-3.7	0.000
illustration5	Claude-3.7	0.000	Claude-3.7	0.000
illustration6	Claude-3.7	0.000	Claude-3.7	0.000
illustration7	Gemini	1.000	Gemini	1.000
illustration8	Claude-3.7	1.000	Claude-3.7	1.000
illustration9	Claude-3.7	1.000	Claude-3.7	1.000
illustration10	Claude-3.7	0.964	Claude-3.7	1.000

chaque description d'image, les neuf LLM ont été évalués par un panel de sept annotateurs. Chaque annotateur a attribué un score au LLM à l'aide d'une échelle ordinale (0, 0,25, 0,5, 0,75, 1). L'objectif était de sélectionner le LLM obtenant les meilleures performances globales en description d'image, sur la base de deux indicateurs : la moyenne et la médiane des scores. Les résultats figurent dans les tableaux 1, 2 et 3.

Les résultats indiquent clairement que Claude-3.7 Sonnet est le LLM le plus fréquemment sélectionné par les annotateurs, aussi bien en termes de score moyen que de médiane. Dans le test sur les diagrammes, Claude est choisi 6 fois sur 10 ; pour les illustrations, 9 fois sur 10 ; et pour les pictogrammes, 10 fois sur 10. Les descriptions

TABLE 3 – Résultats pour le meilleur LLM par pictogramme

Img	Moy. max	Moy.	Méd. max	Méd.
pictogram1	Claude-3.7	0.821	Claude-3.7	1.000
pictogram2	Claude-3.7	0.964	Claude-3.7	1.000
pictogram3	Claude-3.7	0.857	Claude-3.7	1.000
pictogram4	Claude-3.7	0.857	Claude-3.7	1.000
pictogram5	Claude-3.7	1.000	Claude-3.7	1.000
pictogram6	Claude-3.7	0.857	Claude-3.7	1.000
pictogram7	Claude-3.7	0.964	Claude-3.7	1.000
pictogram8	Claude-3.7	0.857	Claude-3.7	1.000
pictogram9	Claude-3.7	0.857	Claude-3.7	1.000
pictogram10	Claude-3.7	0.786	Claude-3.7	1.000

produites par Claude sont très précises et contextuellement riches, permettant de restituer l'ensemble des informations essentielles contenues dans l'image. Sur la base de ces observations, nous avons décidé de sélectionner Claude-3.7 Sonnet pour la tâche de description d'images.

Par la suite, nous avons cherché à comparer les performances de différents LLM en classification binaire. En nous appuyant sur les enseignements de l'expérience de description, nous avons exclu certains modèles des tests suivants. Étant donné que la classification est intrinsèquement plus complexe que la description d'images, nous avons mené des tests préliminaires qui nous ont conduits à nous concentrer sur Claude 4.5 sonnet, GPT-5, Gemini 2.5 pro et Nova 2 Lite pour cette tâche. Nous avons réalisé plusieurs tests de classification avec les quatre modèles. Les tests les plus importants et significatifs consistaient à comparer des audits manuels de sites effectués par un expert à des audits automatiques réalisés par les LLM. Nous avons donc demandé à un expert d'auditer plusieurs sites du gouvernement français. Notre premier test sur sept images toutes informatives nous a amené à sélectionner Claude 4.5 sonnet car le modèle a bien classé les images et ses explications étaient pertinentes. Nous avons donc réalisé un autre test à la suite pour confirmer notre choix. Ce second test comprenait 19 images dont 5 informatives et 14 décoratives d'un même site. Notre prompt était le suivant[3] :

""Tu vas recevoir une image et son texte adjacent. Ta tâche en tant que auditeur RGAA d'images est de :
1. Décrire l'image en 2 à 4 phrases.
2. Classer l'image en tant que "informative" ou "décorative" en suivant la norme RGAA. Une image décorative est une image n'ayant aucune fonction et ne véhiculant aucune information particulière par rapport au contenu auquel elle est associée.Exemples d'images décoratives : Une image précédant chaque item d'une liste, une image servant à caler la mise en page, une image de coin arrondie pour habiller un bloc d'information ou une image d'illustration n'apportant aucune information nécessaire à la compréhension du texte auquel elle est associée. Une image informative est une image qui véhicule une information nécessaire à la compréhension du contenu auquel elle est associée.
3. Explique pourquoi tu as classé l'image en tant que informative ou décorative. Retourne exactement :
Description : <text>
Label : <informative ou décorative>
Explication : <text>""

En comparant les résultats des LLM à l'annotation de référence, nous obtenons les résultats du tableau 4.

GPT-5 arrive premier pour la classification suivi par Gemini 2.5 Pro puis Claude Sonnet 4.5 et enfin Nova 2 Lite. Contrairement à notre premier test, Claude 4.5 Sonnet n'est plus placé premier car il a une tendance à classer

3. Les définitions des deux classes viennent directement du glossaire du RGAA : https://accessibilite.numerique.gouv.fr/methode/glossaire/

TABLE 4 – Performances sur la classification binaire

Modèle	Accuracy	Rappel	Précision	F1
GPT-5	0.84	0.80	0.67	0.73
Gemini 2.5 Pro	0.70	1.00	0.50	0.67
Claude 4.5 Sonnet	0.58	1.00	0.38	0.55
Nova 2 Lite	0.50	1.00	0.33	0.50

les images comme informatives alors même qu'elles sont classées comme décoratives par l'auditeur. Cette tendance n'avait pas été remarqué dans le premier test vu que les images étaient toutes informatives. Avec nos derniers résultats, on remarque très vite que Gemini, Claude et Nova classent la plupart du temps les images comme informatives. Les trois modèles ont un taux de rappel de 1 pour la classe informative alors que GPT-5 a un rappel de 0.8. Concernant la précision, on remarque très rapidement que malgré sa précision moyenne, GPT-5 reste le meilleur pour classer les images.

Nous avons analysé un échantillon des images du site français culture.gouv. Afin d'observer les différences entre les 4 LLMs, nous avons pris 3 images classées informatives et 3 images classées décoratives par l'auditeur. Les résultats sont visibles dans le tableau 5. Nous pouvons voir que le taux de bonnes classifications sur les images informatives est parfait car aucun LLM ne s'est trompé. De manière générale sur toutes les données obtenues, les modèles n'ont pas de mal à classer une image comme informative. Bien que le taux de rappel soit fort, la précision est cependant mauvaise pour les modèles d'Anthropic et d'Amazon. En effet, sur cet échantillon, Claude n'a prédit que des images informative et Nova a prédit deux informatives sur trois décoratives. Gemini et GPT ont obtenu les meilleurs résultats sur cet échantillon en n'ayant qu'une seule erreur (un faux positif). Les faux positifs sont dû à des interprétations des objets présents sur les images. Pour GPT-5, la sixième image "illustre le thème de la Fête de la musique et apporte des informations visuelles sur l'ambiance collective et la diversité des instruments, ce qui complète et renforce le texte adjacent." et pour Gemini 2.5 pro, l'image 4 "est informative car elle illustre visuellement et de manière créative le thème principal du texte. Elle met en évidence les mots-clés "LANGUE", "FRANÇAIS" et "FRANCOPHONIE", qui sont au cœur de l'événement décrit, renforçant ainsi le message du texte.". Bien que les images permettent de renforcer l'idée véhiculée par le texte, elles restent décoratives car elles n'ajoutent pas réellement d'informations en plus de celles du texte.

Concernant Claude 4.5 sonnet, les trois images décoratives (4,5 et 6) sont classées comme informatives car il sur-interpréte ce qu'il observe sur les images. Pour la 4ème image, le modèle justifie son choix en expliquant que "L'image illustre de manière symbolique le concept de diversité, de connexions et de richesse évoqué dans le texte sur la Semaine de la langue française et de la Francophonie. Les multiples couleurs et les points interconnectés représentent la variété des formes et expressions de la langue française, ainsi que les liens entre différents publics et activités culturelles mentionnées." Pour la 5ème image, selon lui, "L'image illustre directement le contenu de l'article sur la Nuit européenne des musées en représentant visuellement l'ambiance nocturne et les performances artistiques mentionnées dans le texte. Elle apporte une valeur contextuelle en donnant un aperçu du type d'expériences visuelles et artistiques que les visiteurs peuvent découvrir lors de cet événement culturel." et pour la 6ème image, "Cette image illustre directement le contenu du texte adjacent sur la Fête de la musique en visualisant la diversité des musiciens, des instruments et l'esprit de partage évoqués dans l'article. Elle apporte une valeur informative en complétant le message textuel par une représentation visuelle concrète de l'événement décrit.". On observe une tendance chez le modèle à toujours interpréter l'abstrait de façon à le rendre informatif. Des tests plus poussés pourraient être utiles pour observer le comportement du modèle avec d'autres images. Cependant, à partir de notre échantillon, il n'est pas encore possible en l'état de choisir Claude 4.5 pour le traitement des images.

Concernant Nova 2 Lite, le modèle s'est trompé deux fois sur notre échantillon. Les deux erreurs sont liées à des images décoratives classées comme informatives. Comme pour Claude 4.5, le modèle a tendance à sur-interpréter ce qu'il observe sur l'image. Cependant, contrairement à Claude, Nova donne des explications dubitatives. On perçoit une certaine indécision dans ses explications. Pour la 4ème image, le modèle explique que "Bien que l'image ait un aspect artistique, elle semble avoir une intention informative sous-jacente, probablement pour illustrer un concept ou une idée liée à la créativité et à la langue française, comme le suggère le texte adjacent. Les lettres et les connexions peuvent représenter des relations ou des idées abstraites, ce qui en fait une illustration plutôt qu'une simple décoration." Nous pouvons retrouver les expressions "semble", "probablement", "peuvent représenter", "plutôt qu'une simple décoration" qui montrent bien le côté "dubitatif" du modèle.

Cette analyse de notre échantillon nous a permis d'avoir une vision plus globale des comportements des différents LLMs. Cela nous a permis d'avoir de nouvelles idées de tests pour confirmer notre choix du LLM a utilisé pour notre extension. Les LLMs étant dépendants des prompts et de leurs données d'entrainement, ils peuvent parfois sembler correspondre à une tâche sans pour autant être les plus adaptés. Nous avons donc besoin de tester plus de tâches même si ce test a été intéressant pour nos recherches.

Afin d'évaluer notre approche, nous avons constitué des jeux de données de taille réduite, comprenant respectivement 30 images pour un premier lot et 19 images pour un second. Bien que ces volumes restent limités et ne permettent pas de généraliser les résultats, ils offrent néanmoins un premier terrain d'expérimentation pertinent. En effet, ces échantillons permettent d'observer des tendances

TABLE 5 – Comparaison des 4 LLM (GPT-5, Amazon Nova 2 Lite, Claude 4.5, Gemini 2.5 Pro) sur 6 images

Image	Audit manuel	GPT-5	Amazon Nova	Claude 4.5	Gemini
	**informative**	**informative**	**informative**	**informative**	**informative**
	**informative**	**informative**	**informative**	**informative**	**informative**
	**informative**	**informative**	**informative**	**informative**	**informative**
	**decorative**	**decorative**	informative	informative	informative
	**decorative**	**decorative**	**decorative**	informative	**decorative**
	**decorative**	informative	informative	informative	**decorative**

et de mettre en évidence des différences de comportement entre les modèles de langage (LLM) face aux types d'images considérés. Ainsi, malgré leur taille restreinte, ces jeux de données fournissent des indications qualitatives utiles pour comparer les approches et orienter les analyses futures.

Conclusion

Cet article visait à montrer l'approche adoptée pour distinguer automatiquement les images informatives des images décoratives grâce à une chaîne de traitement adaptée et appuyée par un LLM. Après avoir défini les critères nécessaires, nous avons développé une extension de navigateur afin d'analyser une image directement depuis une page web. Afin de choisir le LLM le plus adapté, nous avons procédé à l'évaluation de quatre modèles :GPT-5, Gemini 2.5 Pro, Claude 4.5 et Amazon Nova 2 Lite. Cette évaluation a toutefois révélé des différences marquées : GPT-5 est de loin le plus fiable. À l'inverse, Gemini, Claude et Nova présentent une tendance forte à sur-classer comme «informative» des images pourtant décoratives. Ces résultats montrent que, si tous les modèles détectent bien les images réellement informatives, seule l'approche de GPT-5 respecte strictement les critères d'accessibilité. Les autres modèles interprètent trop largement le sens visuel des illustrations. Cependant, les résultats peuvent dépendre des prompts et des spécificités des modèles. Ainsi, les prochaines étapes de notre projet consisteront à tester les LLM sur différentes tâches pour valider notre choix puis à tester l'extension web que nous avons développée sur plusieurs sites afin de pouvoir comparer nos résultats à ceux d'audits manuels.

Usage de l'IA générative

L'IA générative a été utilisée pour une révision linguistique de certaines parties de l'article.

Références

[1] Laetitia Castillan, Julie Lemarié, and Mustapha Mojahid. Numérique, handicap visuel et accessibilité des apprentissages. contenus pédagogiques numériques : quelle accessibilité pour les élèves présentant une déficience visuelle ? *Education & Formation*, (e-311), 2018.

[2] Luigi Celona, Simone Bianco, Marco Donzella, and Paolo Napoletano. Improving image captioning descriptiveness by ranking and llm-based fusion. *Neural Computing and Applications*, 37(32) :27279–27299, September 2025.

[3] Ananya Gubbi Mohanbabu and Amy Pavel. Context-aware image descriptions for web accessibility. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '24)*, 2024.

[4] Maurizio Leotta, Fabrizio Mori, and Marina Ribaudo. Evaluating the effectiveness of automatic image captioning for web accessibility. *Universal Access in the Information Society*, 22 :1293–1313, 2022.

[5] Meredith Ringel Morris, Jeffrey Johnson, Cynthia L. Bennett, and Edward Cutrell. Rich representations of visual content for screen reader users. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '18)*, 2018.

[6] Laura Noreskal, Guillaume Feuilloley, and Simon-Pierre Charbel. An ai solution for web accessibility and images classification. In *2024 IEEE Thirteenth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6, 2024.

[7] Uran Oh, Hwayeon Joh, and YunJung Lee. Image accessibility for screen reader users : A systematic review and a road map. *Electronics*, 10(8) :953, 2021.

[8] Hannah Ramsey and Imke Schröder. Beyond the visual : Achieving accessibility in scientific figures. *ACS Chemical Health & Safety*, 32 :126–132, 2025.

[9] Kaitlin Stack Whitney, Julia Perrone, and Christie A. Bahlai. Open access journals lack image accessibility guidelines. *Quantitative Science Studies*, 6 :46–62, 2025.