Ressenti face aux images créées par intelligence artificielle générative

Rapport de recherche court du laboratoire ErgoDesign Lutin-Gobelins

Sommaire

1 **Introduction** 2

2 **Méthodologie** 3

3 **Résultats** 4

4 **Discussion** 7

5 **Conclusion** 8

6 **Références** 8

Historique du document

Version 1 (02/08/2023) : version initiale

Nicolas Esposito
Simon Autard
Sawsane Ennebati

Version 1 (02/08/2023)

GOBELINS
— PARIS —
Une école de
CCI PARIS ILE-DE-FRANCE EDUCATION

1 Introduction

Dans le contexte actuel de multiplication des outils de création d'images par intelligence artificielle (IA) générative, comme Midjourney, Dall·E ou Stable Diffusion (ROMBACH *et al.*, 2022), de nombreuses questions se posent, notamment sur la manière dont elles sont perçues (RAGOT *et al.*, 2020). Sous cet angle, nous avons mené une expérimentation portant sur le ressenti face à ce type d'images. Nous avons déjà travaillé sur le sujet du ressenti face à différents types d'image. Cette fois-ci, nous le faisons avec une expérimentation pilote qui pourrait donner lieu à des travaux plus approfondis, notamment au regard des progrès à venir dans ce domaine.

Pour cette expérimentation pilote, nous avons comparé le ressenti de vingt participants face à vingt paysages, une moitié correspondant à des photos, l'autre moitié ayant été générée par IA. Chaque type d'image étant représenté par cinq paysages non urbains et cinq paysages urbains (voir figure 1).

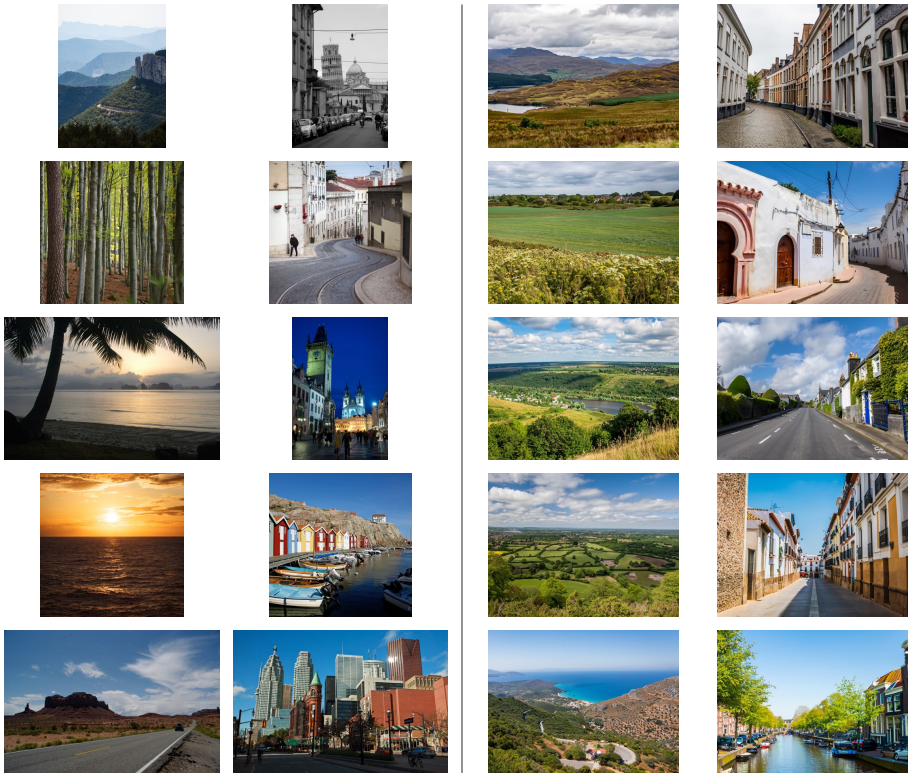

FIGURE 1 – Images utilisées pour l'expérimentation

Nous nous sommes demandé si nous pourrions constater des différences de ressenti et des différences de réussite concernant l'identification du type d'image :

- entre les deux types d'images (photos et images générées par IA) ;
- entre les deux types d'images telles qu'elles sont perçues ;
- entre les images générées par IA selon leur niveau de réalisme ;
- entre les images générées par IA selon l'environnement (urbain ou non) ;
- selon le profil des participants (plus ou moins experts dans le domaine).

Ce qui nous a conduits à énoncer dix hypothèses :

- le ressenti serait plus positif avec les photos ;
- la réussite concernant l'identification du type d'image devrait être similaire pour les deux types ;
- le ressenti serait plus positif avec les images perçues comme des photos ;
- la réussite devrait être similaire pour les deux types d'images telles qu'elles sont perçues ;
- le ressenti face aux images générées par IA devrait être corrélé à leur réalisme ;
- l'échec face aux images générées par IA devrait être corrélé à leur réalisme ;
- le ressenti face aux images générées par IA devrait être plus négatif quand il s'agit de paysages urbains (moins réalistes dans notre ensemble d'images) ;
- pour les images générées par IA, la réussite devrait être plus importante avec les paysages urbains qu'avec les paysages non urbains (les premiers étant moins réalistes) ;
- les participants les plus experts auraient un ressenti plus négatif face aux images générées par IA (les voyant davantage en tant que telles) ;
- les participants les plus experts auraient une réussite plus importante avec les images générées par IA.

2 Méthodologie

L'expérimentation a été menée avec vingt participants ayant un niveau variable d'expertise dans le domaine (image en général et images générées par IA). Chacun a vu les vingt images de la figure 1 selon un ordre aléatoire. Les participants savaient qu'il s'agissait d'un mélange de photos et d'images générées par IA. Pour chaque image, il leur était demandé leur ressenti (de 0 pour très mauvais à 4 pour très bon) et leur estimation concernant le type d'image (de 0 pour pas du tout générée par IA à 4 pour tout à fait générée par IA), ainsi que quelques mots pour expliquer leurs choix. À la fin, il y avait une question ouverte sur le ressenti global à propos de l'ensemble des images et une autre sur le ressenti face aux images reconnues comme étant générées par IA.

La sélection des photos et la génération des images ont été effectuées avec les contraintes suivantes : pas d'images connues, pas de sujets proches au premier plan (qui pourrait rendre l'estimation trop facile), des images aisément compréhensibles et ne générant *a priori* pas d'émotions fortes. Par ailleurs, les images générées par IA devaient représenter le niveau courant des outils actuels (nous avons choisi Firefly) et elles devaient pouvoir être identifiées sans grandes difficultés en tant que telles.

3 Résultats

Comme nous l'avions envisagé, le ressenti a été plus positif avec les photos qu'avec les images générées par IA (voir figure 2). Et la réussite concernant l'identification du type d'image a été similaire pour les deux types. Les moyennes sont de 2,51 (photos) et 2,66 (images générées par IA) sur 4 et nous pensions qu'elles seraient supérieures, donc que l'identification du type de photo serait plus facile. En distinguant les images selon la façon dont elles ont été perçues, nous avons obtenu les mêmes résultats pour le ressenti et la réussite, mais de manière encore plus marquée.

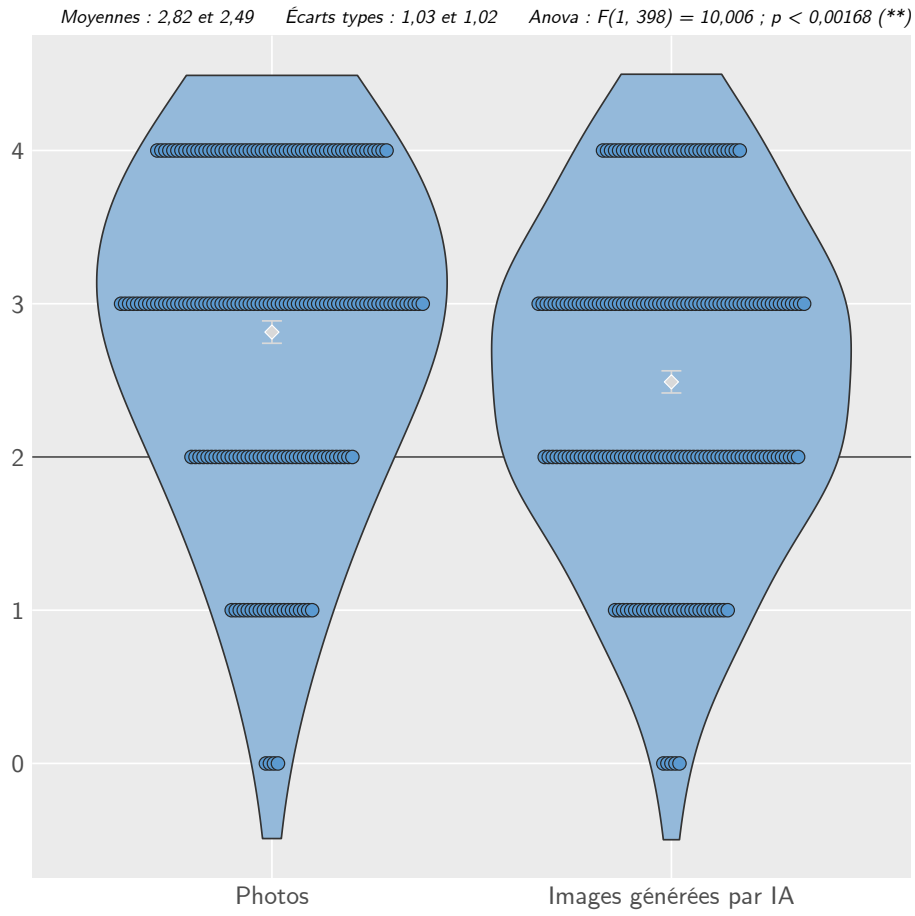

FIGURE 2 – Répartition des réponses pour le ressenti (de 0 pour très mauvais à 4 pour très bon) selon le type d'image, avec moyennes et erreur standard

Nous n'avons pas confirmé de corrélation entre le ressenti face aux images générées par IA et leur réalisme. Mais nous avons confirmé la corrélation entre l'échec face aux images générées par IA et leur réalisme (voir figure 3). Plus ces images sont réalistes, moins elles ont été identifiées comme générées par IA. Concernant la distinction entre les paysages urbains et les paysages non urbains, nous n'avons pas mesuré de différences significatives pour le ressenti et la réussite, même si l'on peut noter une réussite plus grande avec les paysages urbains (moins réalistes pour nos images générées par IA).

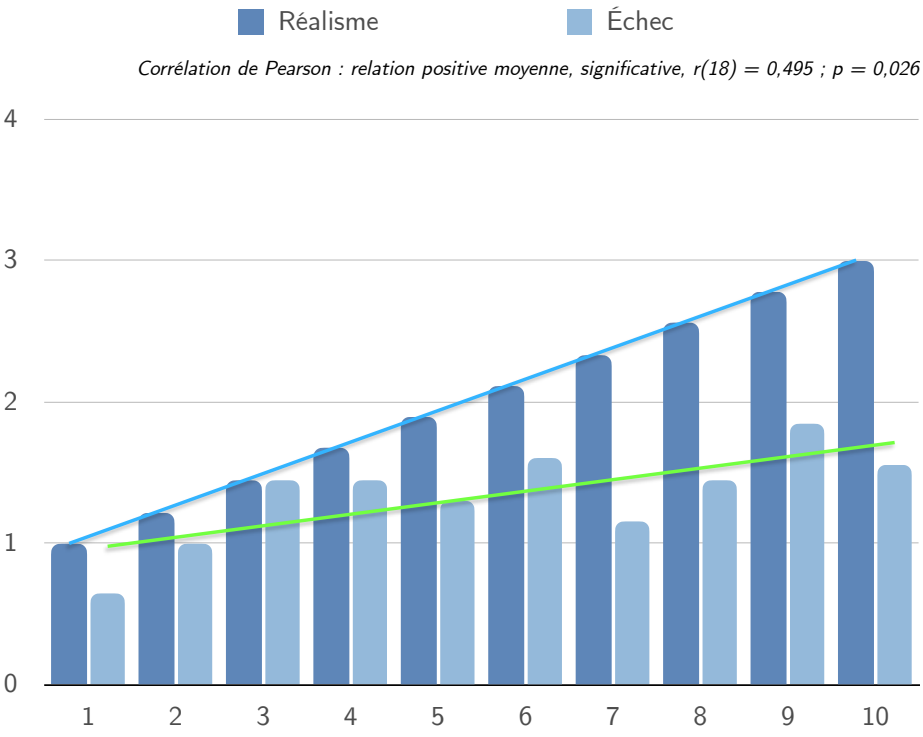

FIGURE 3 – Réalisme et échec pour les images générées par IA

Enfin, nous n'avons pas confirmé de relation entre le niveau d'expertise des participants et le ressenti face aux images générées par IA. Mais nous avons confirmé une corrélation entre le niveau d'expertise des participants et la réussite (voir figure 4). Plus le niveau d'expertise des participants était élevé, mieux ils ont réussi à identifier le type des images.

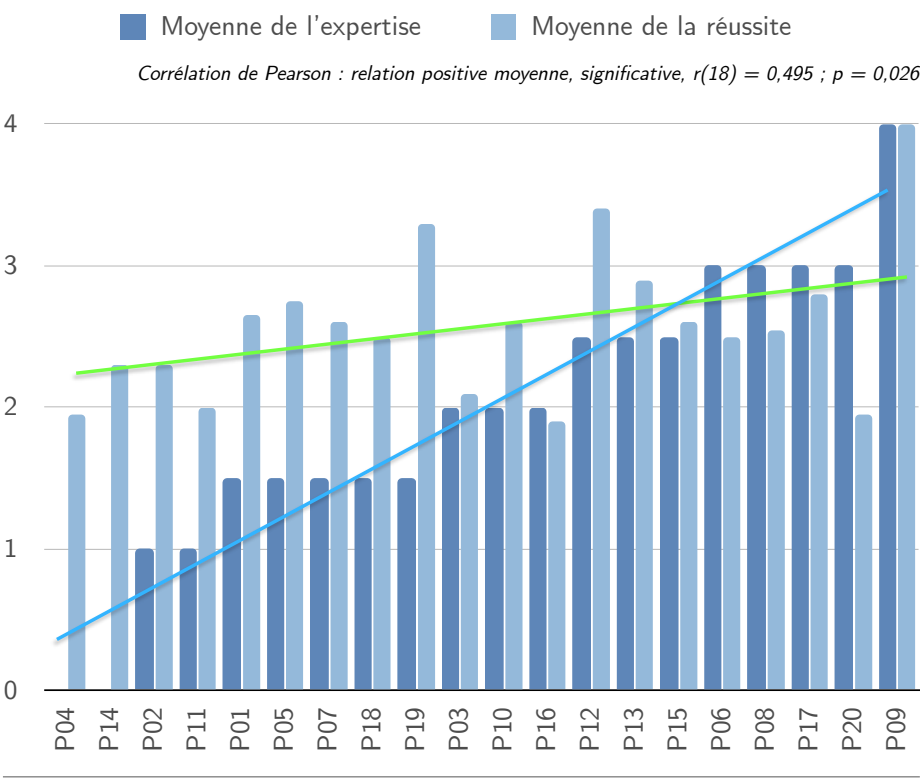

FIGURE 4 – Expertise et réussite

Ainsi, six de nos dix hypothèses ont été validées avec cette expérimentation. Globalement, nous avons mesuré un ressenti moins positif face aux images générées par IA que face aux photos. Cela s'est retrouvé dans l'expression du ressenti. Après avoir catégorisé les retours des participants (entre négatif, positif et neutre), nous avons constaté que le pourcentage de retours négatifs était deux fois plus important pour les images générées par IA que pour les photos : respectivement 22,3 % et 10,7 % (voir figure 5).

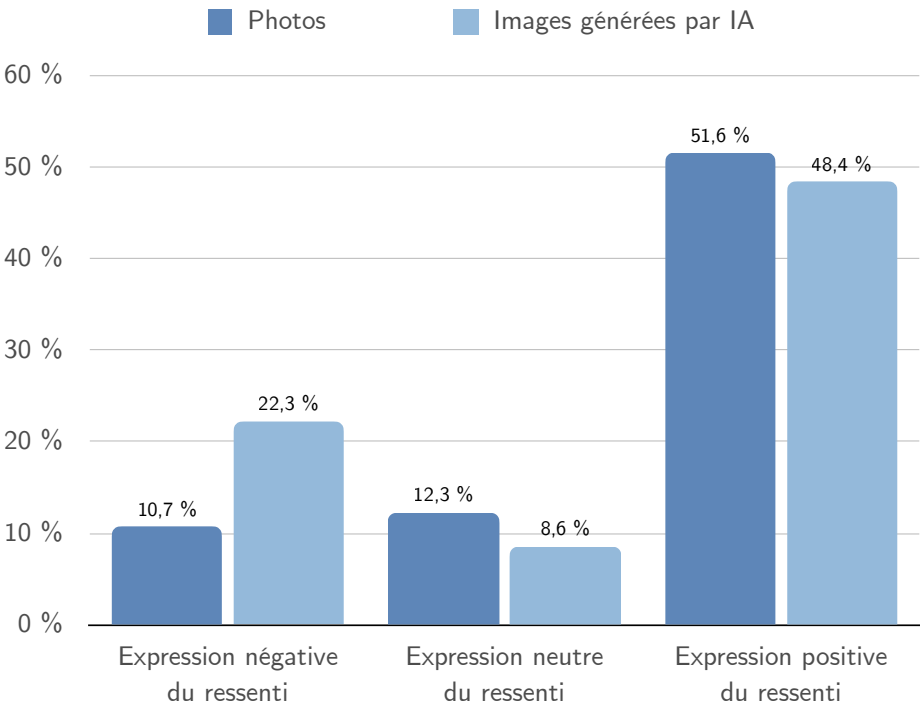

FIGURE 5 – Expression du ressenti

4 Discussion

Les résultats de cette expérimentation pilote nous invitent à envisager des approfondissements, notamment à partir des points suivants. Nous pourrions travailler à partir d'images plus facilement identifiables (images générées par IA ou non) ou à l'opposé, moins facilement identifiables. Nous pourrions constituer des groupes de participants selon des niveaux d'expertise déterminés. De même, nous pourrions constituer des sous-ensembles d'images selon des niveaux de réalismes déterminés. Nous pourrions comparer différents outils de génération. Nous pourrions aussi utiliser d'autres types d'images, pas uniquement des paysages et pas seulement des photos, potentiellement des dessins ou des peintures, figuratives (GU & LI, 2022) ou abstraites (ISRAFILZADE, 2020). Enfin, nous pourrions mettre en œuvre des mesures oculométriques pour travailler sur l'attention visuelle (ROUSSELET & FABRE-THORPE, 2003) et nous pourrions prendre davantage de temps avec les participants afin d'obtenir des retours plus approfondis, en particulier sur leur manière de voir ces images.

5 Conclusion

À partir des résultats de cette expérimentation pilote, avec les images utilisées (la moitié ayant été générée par Firefly), nous retenons en particulier que :

- le ressenti a été plus positif avec les photos qu'avec les images générées par IA (encore plus en distinguant les images selon la façon dont elles ont été perçues) ;
- la réussite concernant l'identification du type d'image a été similaire pour les deux types (encore plus en distinguant les images selon la façon dont elles ont été perçues) ;
- plus les images générées par IA étaient réalistes, moins elles ont été identifiées comme générées par IA ;
- plus le niveau d'expertise dans le domaine des participants était élevé, mieux ils ont réussi à identifier le type des images.

6 Références

Gu, L. & Li, Y. (2022). Who Made the Paintings: Artists or Artificial Intelligence? The Effects of Identity on Liking and Purchase Intention. *Frontiers in Psychology*, 13.

Israfilzade, K. (2020). What's in a Name? Experiment on the Aesthetic Judgments of Art Produced by Artificial Intelligence. *Journal of Arts*, 3(2), 143–158.

Ragot, M., Martin, N. & Cojean, S. (2020). AI-Generated vs. Human Artworks. A Perception Bias Towards Artificial Intelligence? *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–10.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P. & Ommer, B. (2022). High-Resolution Image Synthesis With Latent Diffusion Models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10 684–10 695.

Rousselet, G. A. & Fabre-Thorpe, M. (2003). Les mécanismes de l'attention visuelle. *Psychologie française*, 48(1), 29–44.