Midjourney Caveats: Exploring the Limitations of Text-to-Image Generative AI

Josef Haupt
University of Technology Chemnitz
Chemnitz, Germany
josef.haupt@phil.tu-chemnitz.de

Lena Nischwitz
University of Technology Chemnitz
Chemnitz, Germany
lena-marcella.nischwitz@phil.tu-chemnitz.de

Aurora Weigelt
University of Technology Chemnitz
Chemnitz, Germany
aurora-zoe.weigelt@s2020.tu-chemnitz.de

Lewis Chuang
University of Technology Chemnitz
Chemnitz, Germany
lewis.chuang@phil.tu-chemnitz.de

Abstract

In this work we describe out initial method of utilizing the text-to-image AI Midjourney to generate stimulus material that is cute based on Lorenz `Kindchenschema`.

CCS Concepts

• **Human-centered computing → Human computer interaction (HCI)**.

Keywords

Midjourney, ChatGPT

Mensch und Computer 2024 – Workshopband, Gesellschaft für Informatik e.V., 01.-04. September 2024, Karlsruhe, Germany

https://doi.org/10.18420/muc2024-mci-ws09-194

1 Introduction

Direct observation and critical reflexivity is often applied, especially but not only in the social sciences, to determine ontological truths. In the behavioural sciences, mid-20th scholars (e.g., von Uexküll, Lorenz, Piaget, Bowlby, Gibson) derived far-reaching and influential theories of human perception, reasoning and action based largely on their selective observations and subjective contemplations of their physical experiences. Empirical research continues to thrive by systematically producing systematically acquired and curated data to incrementally validate and disconfirm many of such early theorizing. In the 20th century, could foundation models—trained on large datasets, such as the use of language and media in everyday life—provide a way to validate theoretical concepts?

Natural human languages consist of words that are tokens for abstract concepts. Words such as *cute, attention, industrious, sociable* reference properties and methods that are embedded in our physical experiences, which are often treated as universal states and/or truths. Nonetheless, a closer examination of such terms often reveal that nuanced differences across languages, cultures, and individuals. This paper documents our efforts to investigate the construct validity of the word "cute", based on its influence on popular media or rather, the datasets used to train large language models and their applications i.e., ChatGPT, Midjourney.

What is cute and why is it important? To begin, we consider "cute" as a visual aesthetic that when perceived in a stimulus, evokes in the observer a cute-emotion (also referred to as an *Aww* [2]). When perceived, cute aesthetic is claimed to promote prosocial attitudes and behavior in the observer, by activating brain structures and neural mechanisms that are associated with attention orienting and care-giving [5, 8]. For instance, cute product design has been claimed to result in a greater willingness in consumers to purchase, otherwise unappetising, products (e.g., insect-based food) [1, 11]. As a visual aesthetic, *cute* is often used interchangeably in popular culture with the Japanese word *kawaii* and explicitly defined by [9] as a *Kindchenschema* or collection of childlike features (i.e., arched forehead, large deepset eyes, etc.). Nonetheless, *kawaii* can also be understood as a positive emotion as well as a social cue [14] that can give rise to a heartwarming feeling (i.e,. Kama Muta) [16] or consumer therapeutic effect [3]. Other definitions include those that focus on the state of fun and playfulness (e.g., "whimsical cute" [13]) rather than the visual aesthetics, especially in product design and branding. Attempts have been made to integrate the surprisingly diverse accounts for a seemingly universal understanding of the cute aesthetic, either as a unidimensional spectrum that range from harmless, helpless *Kindchenschema* to uncanny cute [12] or as a taxonomy that define the relationships between diverse visual instances (e.g. Sexy Cute, Baby Cute, Animal Cute, Positive Cute, Negative Cute, Soft Cute or Scary Cute) [10].

To date, *Kindchenschema* continues to be a classical definition for "cute" as a visual aesthetic. The definition of *Kindchenschema* was based on Lorenz's observations of his daughter's caregiving interactions with puppets, especially those that possessed certain physiognomic qualities [9]

> "Versucht man nun zunächst, rein selbstbeobachtend, alle Merkmale herauszugliedern, die ein Objekt haben muß, um die spezifische Erlebnisqualität des „Herzigen" auszulösen, so ergeben sich folgende Kennzeichen, die selbstverständlich im Normalfall dem adäquaten Objekt in optimaler Weise zu eigen sind.1. Verhältnismäßig dicker, großer Kopf, dessen Optimalverhältnis zur Körpergröße sich vielleicht wie bei Tinbergens Amseln durch Versuche bestimmen ließe. 2. Im Verhältnis zum Gesichtsschädel stark überwiegender, mit gewölbter Stirn vorspringender Hirnschädel. 3. Großes und in Übereinstimmung mit der vorerwähnten Proportionierung tief, bis unter der Mitte des Gesamtschädels liegendes Auge. 4. Verhältnismäßig kurze, dicke und dickpfotige Extremitäten. 5. Allgemein rundliche Körperformen. 6. Eine ganz bestimmte, der Fettschicht des gesunden Menschenkindchens entsprechende, weich-elastische Oberflächenbeschaffenheit. 7. Runde, vorspringende „Pausbacken" mangels derer sich die Niedlichkeit des Kindchenkopfes stark verringert." (Lorenz, 1942, p. 275, [9])

The influence of *Kindchenschema* is undeniable and, perhaps, surprising given its anecdotal origins. *Kindchenschema* has not only motivated cute studies [4] but also the increasing cute-ification of the teddy bear [7] and Mickey Mouse [6] over the years. We sought to use large language models and text-to-image generators to determine the extent to which *Kindchenschema* features are perceivable in images that have been annotated as being "cute", even if text-to-image generators were explicitly instructed to selectively manipulate the presence and/or absence of the seven features. Our attempts revealed that the selective manipulation of part-features was not possible and that certain features were so strongly associated that they would co-occur even if explicit instructions were provided for them to be absent.In the next section, we describe our image generation procedures.

2 Method

To verify or refute the *Kindchenschema* brought forward by Lorenz [9], we use the text-to-image generative AI *Midjourney* to generate stimuli based on different *Kindchenschema* features. The only settings we determined were "High Variation Mode", "Stylize med"(–s 100), "Fast Mode" and we used *Midjourney* version 5.2 for our experiments. To assess the general importance of individual features and rank them accordingly we prompted Midjourney to create images showcasing each feature in isolation.

In its original conception [9], the perceived saliency of an object's cuteness was believed to be a summative effect of these features, although certain features (e.g., "round, protruding cheeks") were argued to contribute more strongly than others [9, 15]. To investigate this, we prompted *Midjourney* to generate images with sets of n features, where $n = \{1, 2, 4, 7\}$, using all possible combinations for each n. For this methodology to be effective the generated images are required to have the exact requested features, as any deviation would complicate the testing for intra-*Kindchenschema* dependencies Therefore the prompt needs to exclude all features that are not included in the set of n features explicitly (see Fig. 1).

Midjourney is known to potentially include an object x in a generated image, even when the prompt specifically requests its exclusion, such as in "… without x". To address this, the documentation recommends using weighted prompts[1]. When parts of the prompt are assigned a negative weight, *Midjourney* attempts to exclude these subjects from the resulting image. The –no parameter can be used to list subjects, assigning them negative weights. Thus, a prompt intended to exclude an object x would be formatted as . . . –no x.

Figure 1: The set of images was created using the prompt "person without a backpack". *Midjourney* considers all words and struggles to interpret common natural language negations.

[1] https://docs.midjourney.com/docs/no

3 Challenges

Although *Midjourney* does generate creative and realistic looking images, it's ability to accurately consider each element of the prompt can decrease significantly when adding a number of specific elements. In our attempts to control the generated output we encountered some challenges which will be described in this section as well as the strategies we employed to counter them.

3.1 Prompt Phrasing

The *Midjourney* documentation does not recommend a certain language for the prompting. Simple image generation queries in German might generate the wanted output but with increasing prompt complexity the results differ more and more from the given description. The original description of Lorenz's *Kindchenschema* is written in German and includes some old German words that are likely underrepresented in the *Midjourney* training data, such as "Pausbacken" which can be translated to chubby cheeks. *Midjourney* can not understand the word and will generate a seemingly random output (see Fig. 2). For these reasons all our prompts were translated into English.

- Thick, large head
- Strongly predominant cranium with an arched forehead
- Large, deep-set eyes
- Short, thick and thick-pawed limbs
- Generally rounded body shapes
- Soft-elastic surface texture
- Round, protruding chubby cheeks

These initial translations might be accurate according to Lorenz's original writing [9], but adding multiple features into one prompt leads to very long phrases. *Midjourney* seems to focus on some of the included words and ignore the rest. This is also reflected in the original documentation[2], with the recommendation to use short and simple phrases instead. Therefore we shortened some of the *Kindchenschema* translations while preserving the original meaning:

- Strongly predominant cranium with an arched forehead ⟶ arched forehead
- Short, thick and thick-pawed limbs ⟶ short, thick limbs
- Generally rounded body shapes ⟶ rounded body shapes
- Soft-elastic surface texture ⟶ soft-elastic surface
- Round, protruding chubby cheeks ⟶ round, protruding cheeks

3.2 Excluding Elements

The `–no <phrases>` parameter enables users to adjust the prompt weight such that the generated images excludes certain target phrases listed in the parameter arguments. However, using this parameter does not guarantee that the excluded phrases will not be present in the generated image. The likelihood that *Midjourney* overlooks some elements on the exclusion list increases with the complexity of the phrases or the length of the list.

[2] https://docs.midjourney.com/docs/prompts

Figure 2: Images generated with the German phrase "Pausbacken".

3.3 Selection of Generated Images

Midjourney always creates a set four images with each prompt. Because of the varying accuracy of the prompts, one or more images might fit the prompt better than the others. Normally the user would select the images that are most suitable according to their requirements. In the generation process of stimulus material, this would add a "review" layer to the pipeline, where each image would have to be checked for the presence or absence of the *Kindchenschema* features. Because of this pre-selection, the resulting image collection might not be representative of the popular media the AI was trained on, the prevent this we exclusively used the first of of the four generated images. Depending on the research question this might not be important.

3.4 Prompt Subject

The *Kindchenschema* features described by Lorenz [9] need to be attached to some subject. The selection of a subject is a non-trivial task and can change the output significantly.

We explored the following subjects as targets for the *Kindchenschema* features: "animal", "figure", "character", "being", "creature" and "no subject", with the goal to have a neutral base for all generated images. Comparable to human natural language understanding, all of the listed subjects already invoke a pre-existing bias, likely inherent to the data *Midjourney* was trained on.

The "animal" subjects often already include some of the *Kindchenschema* features and are more likely to be perceived as cute. If "figure" is used as the target subject, most of the time *Midjourney* produces an output with a literal figurine (see Fig. 3a). "character" as a subject will always result in a human that has has some stylised predetermined properties (see Fig. 3b). When using "being" as the

subject the generated images can be very abstract (see Fig. 4a). A prompt that does not include some subject leaves more creative room for *Midjourney*, which in turn shows how *Midjourney* interprets the features on their own, leading to some abstract pictures as well as a zoomed in view of the subject focusing only on the described feature (see Fig. 4b)

(a) "a figure with a large head"

(b) "a character with large forehead"

Figure 3: Generated images where the subject was either a "figure" (3a) or a "character" (3b)

(a) "large eyes"

(b) "large eyes"

Figure 4: Generated images where the subject was a being (4a) or not named (4b)

3.5 Prompt Misinterpretation

Some combinations of words Fig. 5a might be interpreted in an unintended way by *Midjourney*. For example combination "animal" and "large head" often leads to images of hunting-trophies.

Moreover, mixing the "animal" subject with the round body feature can also have some unexpected outcomes. *Midjourney* interprets the "round" aspect of the prompt very literally, frequently rendering the animal inside of a spherical object (see Fig. 5b).

(a) "an animal with a large head and round body shape –no large forehead, large eyes, round body shape, soft surface and chubby cheeks"

(b) "an animal with a round body –no large head, large forehead, large eyes, short limbs, soft surface and chubby cheeks"

Figure 5: Animal subjects with a large head (5a) and/or round body (5b)

4 Conclusion

Generating somewhat accurate stimulus material with certain limitations proved to be much more challenging than initially anticipated, and not all of the problems encountered can be listed here. Since not all of the named issues could be solved sufficiently, we conclude that the current iteration of *Midjourney* was not yet suited to verify the *Kindchenschema* in the way we originally intended. We ended up modifying the perspective of our research. While we maintain that text-to-image AI has a unique understanding of cuteness due to the large amounts of "cute" data it was trained on, we also recognize that this may lead to biases in the generated images, reflecting predominantly popular perceptions of what constitutes cuteness, potential limiting its ability to generate images with fine grained prompts.

We used ChatGPT-3.5 to generate a list of seven animals and seven inanimate objects as our prompt subjects. We then used the conditions "cute" and/or "antropomorphic" to generate prompts for each subject, resulting in eight groups who were either:

- cute / not cute
- anthropomorphic / not anthropomorphic
- animal / inanimate object

The generated images were shown to participants in an online study. They were asked to identify the visible *Kindchenschema* features and to rate them them according to cuteness and anthropomorphism.

Using this method, we have found a way to determine the extent to which Lorenz's *Kindchenschema* features are perceptible in images generated by a commercial text-to-image generator such as *Midjourney*.

References

[1] Raphaela E Bruckdorfer and Oliver B Büttner. 2022. When creepy crawlies are cute as bugs: Investigating the effects of (cute) packaging design in the context of edible insects. *Food Quality and Preference* 100 (2022), 104597.

[2] Ralf C. Buckley. 2016. Aww: The Emotion of Perceiving Cuteness. *Frontiers in Psychology* 7, 1740 (Nov. 2016), 229693. https://doi.org/10.3389/fpsyg.2016.01740

[3] Hsuan-Yi Chou, Xing-Yu (Marcos) Chu, and Tzu-Chun Chen. 2022. The Healing Effect of Cute Elements. *Journal of Consumer Affairs* 56, 2 (June 2022), 565–596. https://doi.org/10.1111/joca.12414

[4] Joshua Paul Dale. 2016. Cute Studies: An Emerging Field. *East Asian Journal of Popular Culture* 2, 1 (April 2016), 5–13. https://doi.org/10.1386/eapc.2.1.5_2

[5] Melanie L Glocker, Daniel D Langleben, Kosha Ruparel, James W Loughead, Jeffrey N Valdez, Mark D Griffin, Norbert Sachser, and Ruben C Gur. 2009. Baby schema modulates the brain reward system in nulliparous women. *Proceedings of the National Academy of Sciences* 106, 22 (2009), 9115–9119.

[6] Stephen Jay Gould. 1979. Mickey Mouse Meets Konrad Lorenz. *Natural History* 88, 5 (1979), 30–36.

[7] Robert A. Hinde and L. A. Barden. 1985. The Evolution of the Teddy Bear. *Animal Behaviour* 33, 4 (1985), 1371–1373.

[8] Morten L. Kringelbach, Eloise A. Stark, Catherine Alexander, Marc H. Bornstein, and Alan Stein. 2016. On Cuteness: Unlocking the Parental Brain and Beyond. *Trends in Cognitive Sciences* 20, 7 (July 2016), 545–558. https://doi.org/10.1016/j.tics.2016.05.003

[9] Konrad Lorenz. 1942. Die angeborenen Formen moeglicher Erfahrung. *Zeitschrift fuer Tierpsychologie* 5, 2 (1942), 235–409. https://doi.org/10.1111/j.1439-0310.1943.tb00655.x

[10] Aaron Marcus, Ayako Hashizume, Masaaki Kurosu, and Xiaojuan Ma. 2017. *Cuteness Engineering: Designing Adorable Products and Services.* Springer, Cham. https://doi.org/10.1007/978-3-319-61961-3

[11] Didier Marquis, Felipe Reinoso Carvalho, and Ga"elle Pantin-Sohier. 2024. Assessing the effect of baby schema cuteness on emotions, perceptions and attitudes towards insect-based packaged foods. *British Food Journal* 126, 4 (2024), 1492–1509.

[12] Simon May. 2019. *The Power of Cute.* Princeton University Press, Princeton.

[13] Gergana Y. Nenkov and Maura L. Scott. 2014. So Cute I Could Eat It Up": Priming Effects of Cute Products on Indulgent Consumption. *Journal of Consumer Research* 41, 2 (Aug. 2014), 326–341. https://doi.org/10.1086/676581

[14] Hiroshi Nittono. 2016. The Two-Layer Model of Kawaii': A Behavioural Science Framework for Understanding Kawaii and Cuteness. *East Asian Journal of Popular Culture* 2, 1 (April 2016), 79–95. https://doi.org/10.1386/eapc.2.1.79_1

[15] E. Seitz. 1942. Die Paarbildung Bei Einigen Cichliden. *Z. TierpsychoL* 5 (1942), 221–251.

[16] Kamilla Knutsen Steinnes, Johanna Katarina Blomster, Beate Seibt, Janis H. Zickfeld, and Alan Page Fiske. 2019. Too Cute for Words: Cuteness Evokes the Heartwarming Emotion of Kama Muta. *Frontiers in Psychology* 10 (March 2019), 387. https://doi.org/10.3389/fpsyg.2019.00387