Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2022 doi:10.20944/preprints202206.0148.v1

Article

Erroneous AI-classification Systems Produce Creative Design Ideas, Predicting Level of Innovation Value

Johan F. Hoorn

School of Design and Dept. of Computing, The Hong Kong Polytechnic University, Hong Kong SAR, 999077, China; johan.f.hoorn@polyu.edu.hk

Abstract: In the mid-layers of Deep Learning systems, clustered features tend to fit multiple classifications, which are filtered out during the final stages of object recognition. However, many misclassifications remain, regarded as errors of the system. This paper claims that tagging an entity incorrectly for reasons of similarity is evidence of spontaneous machine creativeness. According to the ratings of 40 design educators and researchers, AI-generated false-class inclusions produced creative design ideas, predicting the level of innovation value. These designers were not just anybody but came from a design school in Asia with a top position on the world ranking-lists. They entered an experiment in which 20 classification mistakes were framed as early-design ideas that were either human-made or intentionally suggested by creative AI. Many examples passed the Feigenbaum variant of the Turing test with a conceptual preference to creations supposedly done by human hand.

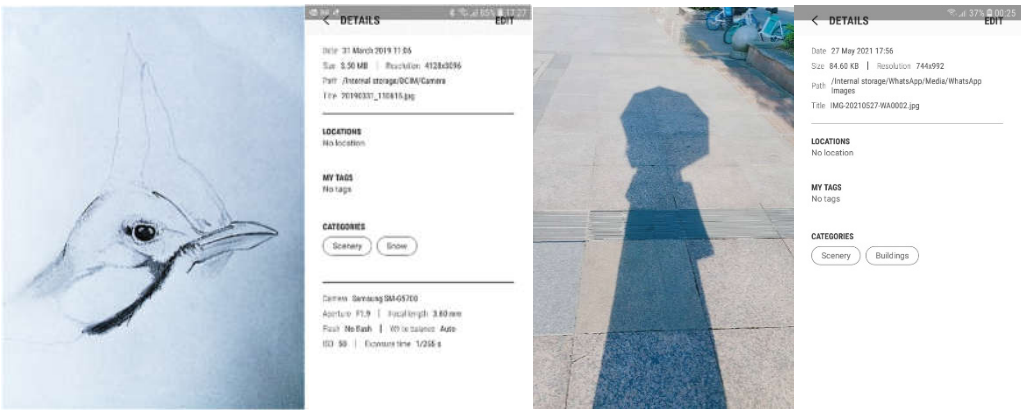
Keywords: false-class inclusions; serendipity; machine vision; creativity; innovativeness

1. Introduction

In the summer of 2021, together with Hong Kong neon artist Sharmaine Kwan and top-tier roboticist David Hanson, I did an exhibition on computational creativity with robots creating art, robots in art, and art just for robots (Ma, 2021) (Appendix A: https://harbourtimes.com/2021/08/09/the-sparkle-not-alive-yet-bright-macabre-god-like-robots-and-a-poetic-glimpse-into-hong-kongs-past-and-future/). At the opening, I did a public seminar on artificial creativity, stating that the process of creativeness is inherent to physical nature (Hoorn, 2014) and that therefore, machines can show creative behaviors as well – without being taught (the slides are available in Appendix B).

That of course encountered the necessary skepticism among the audience and I showed them a few classification mistakes of my Samsung Galaxy J5 Prime Gallery app (Figure 1). Some admitted that this was quite similar to what humans do to generate creative crossovers; others were not convinced that easily.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2022 doi:10.20944/preprints202206.0148.v1



A. Bulbul sketch classified as Scenery with Snow **B**. Shadow of person with umbrella classified as Buildings.

Figure 1. AI-generated false-class inclusions regarded ‘creative’.

However, Lee and Ostwald (2022) show rather convincing evidence that there are several ways in which divergent thinking, the generation of variety, is connected to early ideation, the production of ideas during the design process. One such way to diversify ideas is to break classification boundaries, combining associatively remote entities in a novel way (e.g., Han, Shi, Park, Chen, & Childs, 2018). Conceptual combination across categories often is established by matching the properties of two unrelated entities and estimating their degree of similarity (irrespective of category membership) (cf. Wang & Hu, 2018). In transferring knowledge from one domain to another, Deep Learning systems have a hard time in bridging such large ‘domain discrepancies’ (Na, Jung, Chang, & Hwang, 2021).

Theorists tend to discern two types of similarity. Surface similarity would be the resemblance of form, color, shape, dimension, material, and outer appearance (cf. Figure 1). Structural similarity pertains to aspects such as principle, meaning, and system (Vosniadou & Ortony, 1989; Blanchette & Dunbar, 2000). One of the participants in the experiment indeed emphasized this distinction:

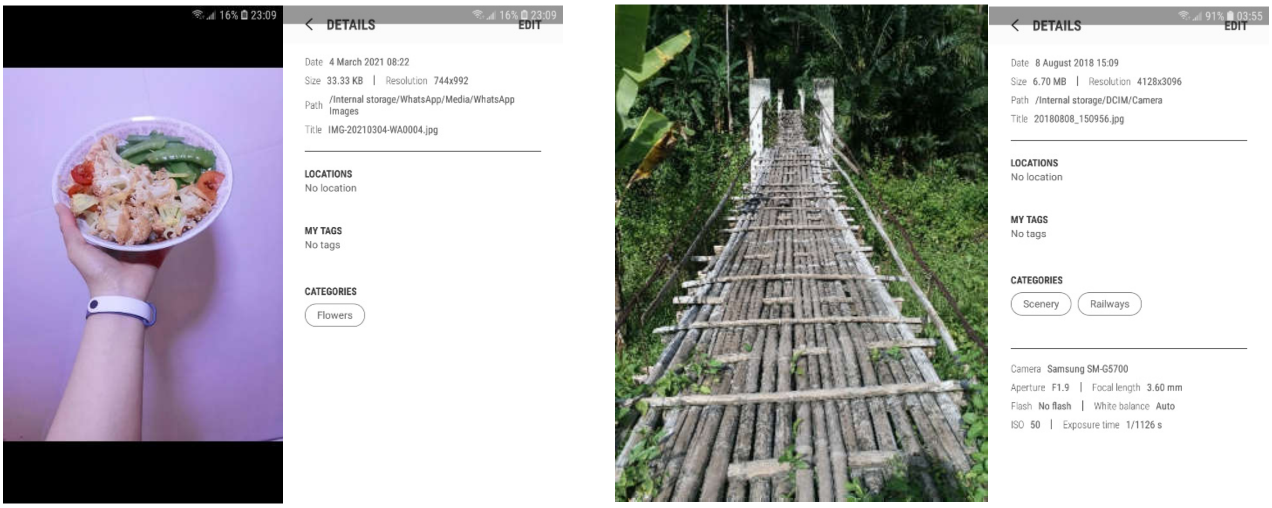
> In my class, I do similar experiments but less visual, it is more scenarios. But the kind of association is quite the same, the shape of the documents being like a building. But there are more subtleties to it, novelty, surprise. Surface similarity (camouflage, mimicking) and structural similarities (that would be real novelty, on a more abstract level). (Participant 09, Human condition)

For those sensitive to the distinction, structural similarity and surface similarity may establish different levels of creativity of ideas (Figure 2):

> Christmas dress is innovative because Xmas is an event and dress is an object (Example 06). Looks like a tree but the reference is to a concept, more abstract. (Participant 06, Human condition)

> The moth (Example 10) is connected to nature and the environment, this may bring new meaning to people about their clothing. Food dispenser (Example 18), the idea takes things out rather than throw things in, that is interesting. I would see it as a creature that feeds on trash. But some of these ideas relate to my students’ ideas in the sense of structural similarity, put things in instead of taking things out: schoolbag as Pokémon to feed the proper ingredients each school day. That way, routine becomes interesting. Food plate as flower (Example 07) may change people’s interpretation of their meal with more appreciation as if admiring a flower. The hanging bridge (Example 15) is just visual similarity, I do not rate that too high. (Participant 09, Human condition)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2022 doi:10.20944/preprints202206.0148.v1



A. Structural similarity: food plate as flower **B.** Surface similarity: bridge as railway

Figure 2. AI-generated false-class inclusions illustrating structural (A) versus surface similarity (B).

Many examples of human-created false-class inclusions may be found in Robert and Robert (1996), seeing angry, sad, and smiling faces in everyday objects like power outlets and light switches. Moderately infrequent and peripheral associations seem to be important for something else than correct classification (cf. Benedek, Könen, & Neubauer, 2012). False-class inclusions seem to be the starting point for creative innovation (Hoorn, 2014, pp. 284-285).

There is quite some human inconsistent thinking about machine creativeness and human brightness. When humans do false-class inclusions, these are regarded as deliberate, meaningful, and contextualized; when machines do the same, mindlessly, the results - although identical - are deemed uninteresting and 'wrong.' This position overlooks one of the most elementary aspects of physical nature that sparks the creative process (Hoorn, 2014, p. 312): coincidence. Or in the theory of creativity: serendipitous findings (Simonton, 2022).

Correct classification is what computer scientists strive for. There is a certain breed of white on pink tulips called Tulipa Ice Cream.[1] Images of this flower are a hard test for any image recognition system. When it says Flower, it is correct. When it says Food (i.e. ice cream), it is wrong although the tulip looks like an ice cream and therefore is called Tulipa Ice Cream but metaphor seems to be a privilege of human beings alone.

It may seem counter-intuitive to computer specialists, who try to reduce 'anomalies' and 'false positives.' By adding context information, the false-positive rate of deep neural network models is reduced (Ru, Li, Hu, & Yao, 2019). Noise, diversity, high dimensionality, and the distributed nature of information are seen as so many stumbling blocks to unsupervised learning (Ramakrishnan & Grama, 1999) but are the very conditions under which spontaneous cross-overs occur: unexpected results, open-ended play (Cheng & Hegre, 2009).

Deep Learning (DL) systems have about 10 to 20 layers of 'cells' that process raw data (e.g., one pixel). Cells integrate information from the bottom layers up into the next higher level and so on until they reach criterion and place an object in a category based on its features. Bottom layers detect, for instance, lines. Second layers search combinations of lines such as edges. Edges together make up typical features for faces, cars, elephants, chairs, etc. (Figure 3). All cells are connected but connections have different weights. Tweaking weights is done by terabytes of training data. Weights indicate the relative importance of each cell. During training, the system gets thousands of pictures with the

[1]https://www.fluwel.com/media/catalog/product/cache/6548503aa833e68ffdc45b75be6da2e5/i/c/ ice_cream_1200_b.jpg

correct output it should produce. In the beginning, all weights are wrong but from the difference with the desired output, the system adapts the weights to come closer to the ideal (not this edge but that). Through many iterations, the weights are fine-tuned such that they produce 'elephant' even in cases that were not in the training set. Without giving it much thought, scientists, analysts, and software developers want their system 'to do it right:' From the bottom-up features, through middle layers of larger units (e.g., eyes, wheels, legs), the machine should recognize Faces, Cars, Elephants, Chairs.

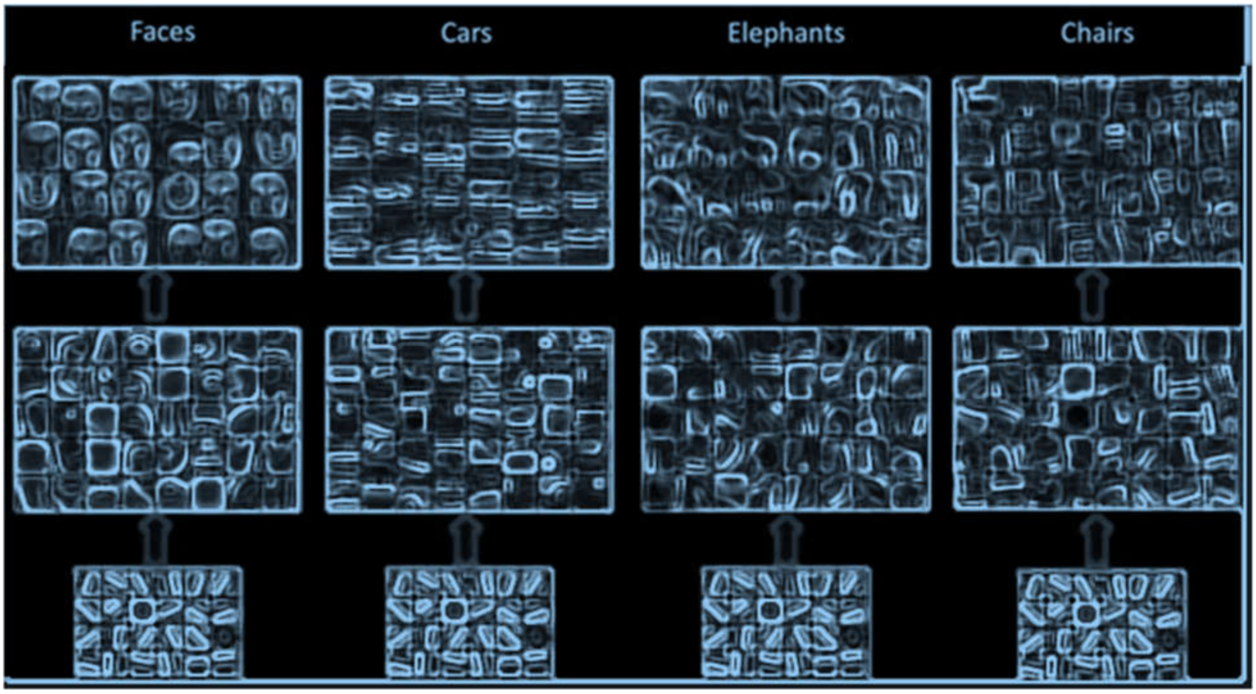


Figure 3. Deep Learning and other Neural Networks work their way up from the composite features (image after Kumar & Singh, 2019).

Why error is only error in a rigid system and not in creativity can be learned from the mistakes that Deep Learning makes. DL does categorizations from the bottom features up (Figure 3). However, the middle layers are where the hybridization happens: Features from the 'wrong' category may match on an abstract level. That is not 'wrong,' it is very exact. On an abstract level, features of one entity are the same as for another entity (cf. topological invariance). That is why they can crossover. That a scientist's poor theory is not capable of discerning what feature is specific for a car and what for a human face makes DL rightfully say that the wheels of a car can be replaced by human eyes and that a human face has wheels for eyes. A chair looks like an elephant because its legs are thick. The Elephant Chair by designer Maximo Riera is a case in point.[2] In the middle layers, crossovers happen and 'false-class inclusions' occur, indicating that DL systems are susceptible to creativeness like all non-living matter in the universe: It does serendipitous findings, sometimes rejected by rigid taxonomists.

Humans allow themselves to be inconsistent; machines are disqualified for not following human inconsistency. Why would not there be an Elephant face? In the 19th century, Joseph Merrick was nicknamed Elephant Man because of his extreme cranial deformations.[3] A car is a face because two wheels look like eyes. Why would not there be a Chair face? We do have Car chairs! When during a presentation, I hold the microphone at my mouth, the Object Recognition System of my Pepper robot tells me that I am licking from an ice cream. When I hold the microphone in front of my chest, the Object

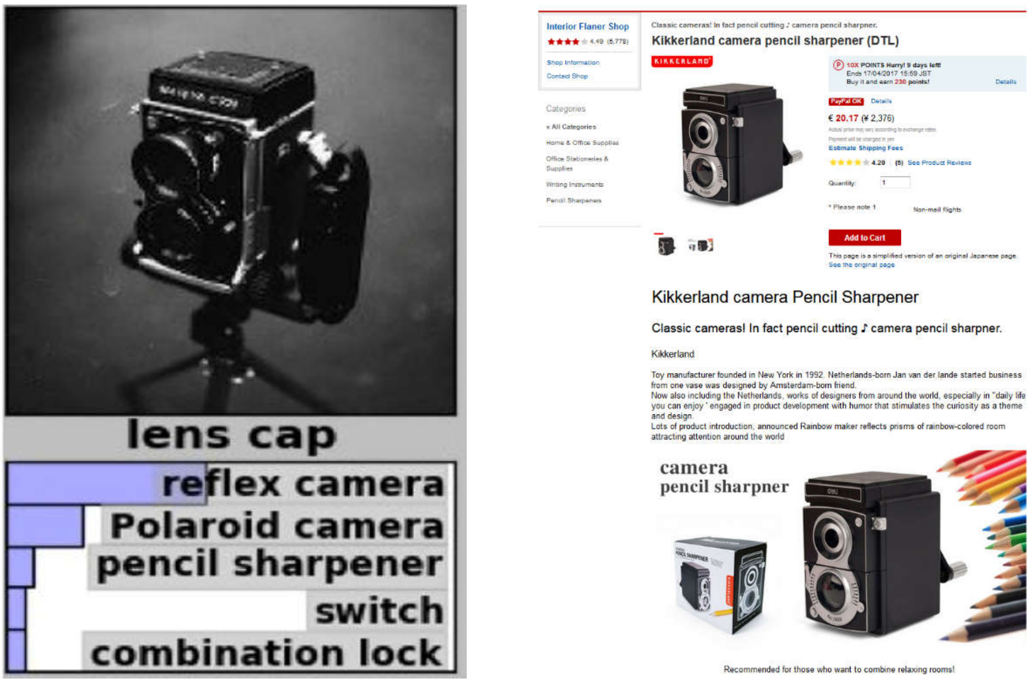
[2] https://cdn.trendhunterstatic.com/thumbs/elephant-chair-by-maximo-riera.jpeg?auto=webp

[3] https://upload.wikimedia.org/wikipedia/commons/b/be/Joseph_Merrick_carte_de_visite_photo,_c._1889.jpg

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2022 doi:10.20944/preprints202206.0148.v1

Recognition System says I am wearing a tie. If I use the microphone to point something out on the projection screen, the Object Recognition System tells the audience that I swing my baseball bat. Three times wrong, fully unaware of situation or context but three times totally creative.

Deep Learning (Krizhevsky, Sutskever, & Hinton, 2012; 2017) was fed the picture of a historic 6×6 photo camera and recognized a pencil sharpener (Figure 4, A). Funny enough and fully independent from any DL classification, the New York toy manufacturers of Kikkerland brought a 'Camera Pencil Sharpener' to market,[4] where the lens of an old-fashioned 6×6 photo camera serves as a hole to put the pencil in (Figure 4, B). The web site says that this is "…product development with humor that stimulates the curiosity as a theme and design."



A. Pencil sharpener as camera (machine error) **B**. Pencil sharpener as camera (product design)

Figure 4. Deep Learning's wrong classification of a pencil sharpener (A) is a marketable design idea (B) when done by humans.

Deep Learning (Krizhevsky, Sutskever, & Hinton, 2012; 2017) was fed the photo of a planetarium and mistakenly classified it as a mosque (Figure 5, A). However, the web site with attractions in Malaysia advertises the National Planetarium in Kuala Lumpur as:[5] "The planetarium has been intelligently designed and structured to mimic a mosque with a blue dome" (Figure 5, B). In Jordan, the construction of a new glass planetarium is in progress at Al Husseini Mosque in downtown Amman (Petra, 2021). Interestingly, Nguyen, Yosinski, and Clune (2016) found that entities such as planetarium, mosque, and church do not only share the same semantic class on a structural level but also are closely related for their similar visual patterns (Vosniadou & Ortony, 1989; Blanchette & Dunbar, 2000), when ordered according to WordNet hierarchy (Figure 6).

[4] http://global.rakuten.com/en/store/flaner/item/10001689/

[5] http://kuala-lumpur.attractionsinmalaysia.com/National-Planetarium.php

A. Planetarium as mosque (machine error) **B**. Planetarium as mosque (human intelligent design)

Figure 5. Deep Learning's wrong classification of a planetarium (A) is an intelligent design (B) when done by humans.

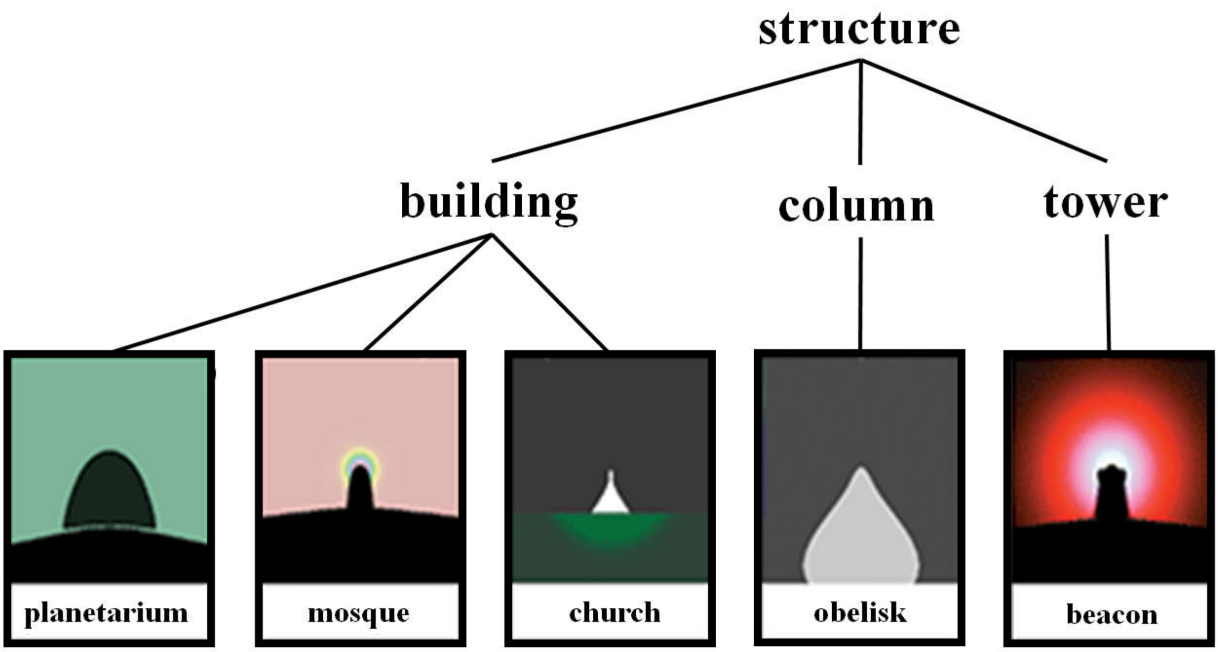


Figure 6. After Nguyen, Yosinski, and Clune (2016, p. 15): Planetarium, mosque, church, etc. belong to subclasses of the structure category, exhibiting surface similarity.

All of these are impressive design ideas but not if they come from a machine. Worse, if DL is totally consistent and recognizes a pile of fruit and vegetables in *Vertumnus*,[6] Italian painter Arcimboldo's (1591) portrait of Emperor Rudolf II, it is told wrong again. This time, the machine should have recognized that when humans do false-class inclusions, it

[6] https://upload.wikimedia.org/wikipedia/commons/d/d2/Arcimboldovertemnus.jpeg

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2022 doi:10.20944/preprints202206.0148.v1

is creative. When machines do it, it is a faulty system. False-class inclusions are creative. If humans are allowed to do it then why is it mere 'error' when machines do the same?

Humans create metaphors all the time: Vestmannaeyar in Iceland has a lava mountain that looks like an elephant.[7] It is called Elephant Mountain. One can see an eye in the cave, a trunk that takes water from the sea, it has a rough texture, is grey, and old. The granite top of Lion Rock Hill in Hong Kong is called Lion Rock because it looks like a crouching lion. We have pie charts, hammerheads, a swallowtail joint, and Grey Wolf optimization-routines (Niu, Niu, Liu, & Chang, 2019).

On the Internet, we find photos of clouds that are described by the people who posted them as a cat, a dog, or a rabbit but my Samsung Gallery app rightly tags them as 'Clouds.' The flower that is baptized Dove/Egret because it looks like a white pigeon is rightly classified as 'Flower.' A plate with 12 pieces of fruit on the verge line and a fork and spoon at 2 o'clock is correctly regarded as 'Food.' Now the machine is uninteresting and lacks imagination. If it says it sees a building in the shadow of a person holding an umbrella (Figure 1, B), it is inaccurate and does false-class inclusions.

Incorrect classification is an important creative behavior. For designers, associations that are far-fetched for others are the very basis of creative design (Benedek, Könen, & Neubauer, 2012), vide the test on Inductive Differentiation Inference by Siu and Wong (2002, pp. 259-307), propagating visual thinking, free imagination, and unlimited association on everyday objects surrounding us. As one of the participants commented on the false-class inclusion of a tropical-fish hat (Example 19):

> Creativity is something nobody else did before, unique. Can also be to update existing design with new functions and purpose. Nobody else would think it fits, like the fish and the hat. (Participant 37, AI condition)

In sum, there may be certain inconsistencies in how the research community goes about the error-estimation of a taxonomy. What in computer vision and image recognition is seen as a mistake is in other disciplines such as arts and design a creative crossover. From these considerations, I infer the following hypotheses:

H1: False-class inclusions lie at the basis of creativity and can have innovation value (cf. the Kikkerland example)

H2: Serendipity is an important impetus to the creative process

H3: Errors are only errors when one adheres to one particular system

H4: Errors are serendipitous findings when perceived creatively

H5: Creative behaviors may emerge without being consciously aware (i.e. machines do not know they create)

H6: Creativity is an inherent aspect of information processing, whether from human or AI origins

To test whether false-class inclusions made by AI classification systems can be considered 'creative ideas,' I conducted an experiment with 20 misclassifications presented to 40 educators and researchers in School of Design, The Hong Kong Polytechnic University. According to QS World University Rankings (2019), School of Design is among the world top in art and design, ranking 2nd in Asia and 15th worldwide (out of 2574 competitors), which almost certainly is the work of the design educators and researchers included in the experiment. This way, I made sure that the evaluations of creativity and innovativeness were based on years of experience in judging the work of design students.

2. Methods

2.1. Participants and Design

Voluntary participants ($N = 40$) not receiving any reward were $M_{age} = 45.8$, $SD_{age} = 8.86$; 25 (59.5%) male, 13 (31%) female, 2 (4.8%) other; 33 Asians, 4 Europeans, 3 North-

[7] https://s-media-cache-ak0.pinimg.com/736x/82/05/7e/82057e506aa1bd7bbd04fd3bf5f4663c.jpg

Americans. All had a PhD degree and were research and teaching staff in School of Design of The Hong Kong Polytechnic University with a distribution between 5% and 18% over the 8 different programs the school covers.

The designers were divided into two groups: One group judged the creativity and novelty of the false class-inclusions believing they were from human origin (a student named 'Sam'); the other group did so, believing the false class-inclusions were done by a creative AI system (named 'Sam'). Thus, the experiment consisted of a 2 (Agency: human vs AI) (between-subjects) * 2 (Measures: creativity vs innovativeness) (within-subjects) design. During experimentation, no objections against the human nature of 'Sam' were raised. Participant 19 (Human condition) upon seeing the trashcan food-dispenser (Example 18): "This person has funny ideas!" Upon seeing the person-with-umbrella shadow (Example 11): "S/he should study architecture!"

2.2. Procedure

Dependent on random allocation to a condition, participants were personally invited over the email to evaluate creative output from either a design student or from a creative AI; the output itself being the same. They met with the experimenter in their natural environment, their office or meeting area in Zaha Hadid's Innovation Tower, where they also evaluate the work of their students. After giving their consent, the experimenter would introduce the task to them and open a portfolio, containing the 20 examples and one sketch combining four of these. The 20 examples were presented in a different random order for each participant, the combined design (number 21) always following last. After rating all 21 stimuli, participants reflected verbally on the concepts of creativity and innovativeness, and then answered 4 extra demographic questions. The experimenter inputted all data so that participants could fully focus on rating the design ideas without having to handle an interface. The whole procedure was self-paced. On this procedure, Participant 03 (Human condition) commented: "I think it is a great exercise for my design students. Can I use your examples for next semester?" Participants were debriefed during a seminar where the results were presented to them and their feedback was received.

2.3. Apparatus and Materials

The 20 examples were photos from the author's personal collection (e.g., Figure 7) or from the Internet (Appendix 1), classified by 'Sam,' short for Samsung Galaxy J5 Prime Gallery app, which is notorious for doing misclassifications and incorrect photo tagging.[8] The app sometimes tags the same name to different faces or classifies colorful foam clogs ('Crocks') floating on the water as 'animals.' 'Tags' or metadata are not embedded in the images and tags cannot be user-added to the images nor are they back-upped, which from a user perspective is undesirable but for experimental reasons quite perfect as the false categorizations remained naïve and 'clean,' free from correct knowledge. Upon investigation of the algorithm responsible for analyzing what is in the image and automatically assigning tags, the company remained silent. Direct email to one of the former Samsung researchers rendered the following quote (anonymized): "I am sorry that I can not help you. As you know, the company confidential is confidential. Even though I know, I can not share it with others" (personal communication, October 28, 2021). For those interested in the technical details of mobile AI for image processing, consult Thabet, Mahmoudi, and Bedoui (2014).

[8] https://eu.community.samsung.com/t5/other-smartphones/gallery-face-tagging-not-working/td-p/527939

 doi:10.20944/preprints202206.0148.v1

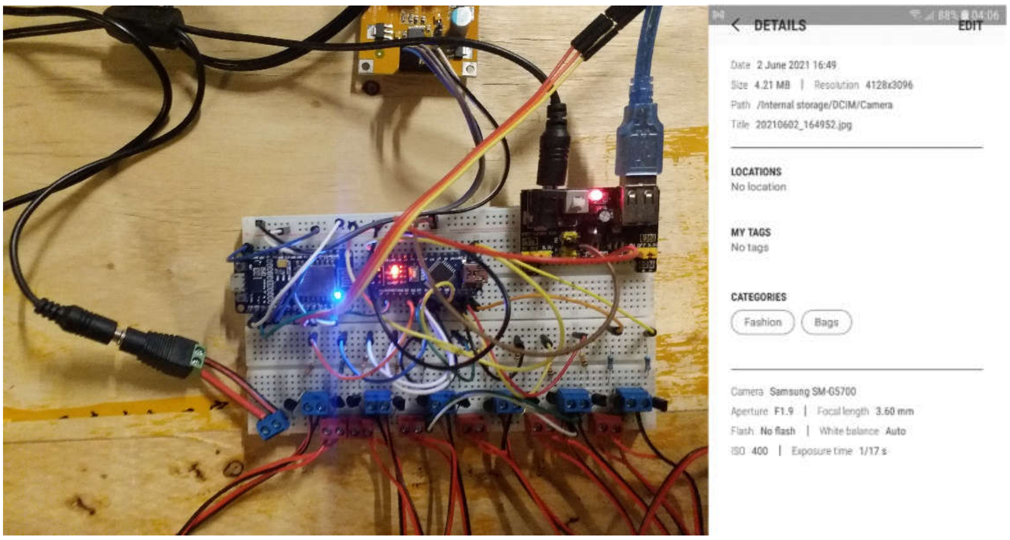


Figure 7. Example 01: Arduino board misclassified as fashion bag.

Following up on the advice of Rodgers, Green, and McGown (2000), Example 21 was a fashion sketch in pencil, consisting of tropical-fish hat (Example 19), moth mantle (Example 10), stained-glass lampshade skirt (Example 14), and clothes-iron shoes (Example 17). The sketch was drafted by the author but in both conditions (Human or AI) presented to the participants as the work from a student in the Institute of Textiles and Clothing.

Examples were color-printed on A3 paper and presented in a black portfolio folder, each transparent sleeve holding a different example (Appendix 3). Participants called out their ratings verbally, which were typed in on a MacBook Pro by the experimenter.

2.4. Measures

One trial consisted of a color picture on the left of the A3 printout (Figure 8), on the right, the misclassification by the Samsung Gallery app in normal writing, and a paraphrase under the picture in the form of a Likert-type item. An example of a paraphrase is: *I find the idea that an Arduino board looks like a fashionable bag…* after which participants scored the level of *creative* and *innovative* on a 6-point rating scale (1= totally disagree, 6= totally agree).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2022 doi:10.20944/preprints202206.0148.v1

10

Framed as human-made

Instead of “A moth” or “butterfly”

Sam said:

“This moth seems a fashionable fur coat.”

Framed as AI-generated

Instead of “A moth” or “butterfly”

Sam outputted:

`<fashion fur coat/>`

10 I find the idea that the moth looks like fashion…

creative

```
Totally                     Disagree    Agree a                 Totally
disagree    Disagree        a little    little      Agree       agree
1 ------- 2 ------- 3 ------- 4 ------- 5 ------- 6
```

innovative

```
Totally                     Disagree    Agree a                 Totally
disagree    Disagree        a little    little      Agree       agree
1 ------- 2 ------- 3 ------- 4 ------- 5 ------- 6
```

Figure 8. Example 10 as trial. Vertical in blue (not present in the actual experiment), the framing in Human and AI condition.

The experimenter did not offer conceptualizations of creativity and innovation before or during data acquisition. Participants reflected on the meaning of these concepts afterwards, answering two qualitative questions on what they thought being creative is about and what being innovative (text data). The survey ended on 4 questions about Gender, Age, Program (e.g., Product Design, Social Design), and Region.

3. Analysis and Results

3.1. Inspection of the stimuli

To investigate which examples were assessed as more creative and innovative, I ranked the mean scores of Creativity and Innovation to examples in the Human and the AI condition. Spearman *rho* did not indicate a significant coherence between rankings (Creativity: r_s = .038, p = .871, 2-tailed; Innovativeness: r_s = .187, p = .417, 2-tailed). What design researchers and educators considered best or worst did not correlate between perceptions of human-made or AI-generated.

To find which examples were most controversial (largest spread in scores), stimuli were ranked according to their standard deviations but Spearman *rho* found no significant coherence. In the Human condition, most remarkably, the least controversial creative idea (Example 19) was the most controversial for its innovative value: For some, a tropical-fish hat would count as an innovation; others would beg strongly to disagree. Exact rank orderings for each example separately can be found in the technical report (Appendix 3).

As a manipulation check, one-sample t-tests were performed for all stimuli across conditions with 3.5 as the test value, being scale-midpoint. Anything significantly different from scale-midpoint (scores >> 3.5) would count as 'creative' and 'innovative.' Across both conditions (N = 40), 50% of examples were regarded as creative, 25% had innovation value. For exact test values, see the technical report (Appendix 3).

Since Spearman *rho* indicated no substantial consistency between the evaluations of individual stimuli and because certain examples (i.e. 19) may have obtained a large 'spread of opinion,' one-sample t-tests were performed for all examples per condition, setting the test value to 3.5 (midpoint). In the Human condition (n = 20), 38% of examples were regarded as creative, 14% as innovative when perceived as human-made. In the AI condition (n = 20), 50% were regarded as creative, *none* as innovative when perceived as AI-generated. Details are in Appendix 3.

Although the stimuli were the same in both conditions, which and how many examples were regarded as creative and/or innovative (at all) largely differed whether framed as human-made or AI-generated. When AI-generated, the design researchers and educators deemed all examples 'not particularly innovative,' whereas in the human-made condition, they assessed some of those same examples as significantly innovative (i.e. 10: moth mantle, 14: stained-glass lampshade skirt, 17: clothes-iron shoe).

Creative beyond doubt in both conditions were examples 05 (drip-loop eyewear), 10 (moth mantle), 14 (lampshade skirt), 16 (ice cream as desert landscape), and 17 (clothes-iron shoe). Examples seen as creative when AI generated but less so when perceived as human-made were: 04 (cat bed is like food), 06 (high-fashion dress looks like Christmas), and 08 (landform looks like buildings). Reversely, examples judged as creative while assumed to be made by a human were 13 (documents seen as buildings) and 19 (tropical-fish hat) but when perceived as AI-generated, the rating dropped.

Thus, three examples were significantly creative and innovative throughout every analysis, being 10: moth mantle, 14: lampshade skirt, and 17: clothes-iron shoe.

3.2. Outliers and background variables

Because not every example was judged in the same way, averages were calculated across conditions as well as per condition from the examples that were beyond doubt (i.e. scored significantly better than scale midpoint). For those 6 averages (two per condition), boxplot analysis showed five extremes and outliers (specifics are in Appendix 3). Further analyses were conducted with and without those participants.

Bivariate Pearson correlations showed that Age did not significantly correlate with any of the dependents. Gender, Program, and Region showed no significant univariate effects on the dependents (also see Appendix 3) and further analyses will discard the background variables.

3.3. Creativity explains Innovativeness

Outliers included (N = 40), linear regression (method Enter) of Mean Creativity on Mean Innovativeness of the selected examples across conditions revealed that R^2 = .59, R^2_{adj} = .58; $F_{(1,38)}$ = 54.13, p = .000; β = .79, β_{st} = .77; t = 7.36, p = .000.

Outliers excluded (n = 36), linear regression (method Enter) of Mean Creativity on Mean Innovativeness of the selected examples across conditions showed: R^2 = .38, R^2_{adj} = .36; $F_{(1,34)}$ = 20.53, p = .000; β = .59, β_{st} = .61; t = 4.53, p = .000.

Because participants evaluated none of the examples in the AI condition as sufficiently innovative, next follows an analysis of the Human condition alone. Outliers included (n = 20), linear regression (method Enter) of Mean Creativity on Mean Innovativeness of the selected examples within Human conditions showed that R^2 = .65, R^2_{adj} = .63; $F_{(1,18)}$ = 32.95, p = .000; β = .73, β_{st} = .80; t = 5.74, p = .000.

Outliers excluded (n = 18), linear regression (method Enter) of Mean Creativity on Mean Innovativeness of the selected examples within the Human condition obtained: R^2 = .57, R^2_{adj} = .54; $F_{(1,16)}$ = 21.23, p = .000; β = .78, β_{st} = .76; t = 4.61, p = .000.

However analyzed, the mean creativity of the examples that were beyond doubt significantly explained 60% to 80% of their level of innovativeness.

3.4. Example 21: combined vs single design ideas

Example 21 was a fashion-design sketch composed of the most creative and innovative examples in the set, namely 10: moth mantle, 14: lampshade skirt, 17: clothes-iron shoe, together with the most controversial number 19: tropical-fish hat. See Appendix 3.

To study the Creativity scores of single ideas as predictor of creativity and innovativeness of the overall design (across Human and AI), multiple regression (method Enter) of the Creativity scores (N = 40) to Example 10, 14, 17, and 19 to predict the Creativity score to Example 21 did not obtain significant model fit: $F_{(4,35)} = 1.67$, $p = .179$. Multiple regression of the Creativity scores to Example 10, 14, 17, and 19 to predict the Innovativeness score to Example 21 showed no significant model fit either: $F_{(4,35)} = 1.85$, $p = .142$.

New boxplot analyses rendered a new set of extremes and outliers but including or excluding these values did not render significant results either. See Appendix 3.

In the Human condition, without extremes and outliers (n = 18), multiple regression on the Creativity scores to single ideas as predictor of creativity and innovativeness of the overall design rendered no significant model fit: $F_{(4,13)} = 1.04$, $p = .422$. For the spurious results with outliers included, see Appendix 3.

Yet, multiple regression (method Enter) of the Creativity scores in Human (n = 20) did explain the *Innovativeness* of the combined Example 21. Model fit was significant: $F_{(4,15)} = 7.20$, $p = .002$ with $R^2 = .66$, $R^2_{adj} = .57$. In the Human condition, the level of Creativity of Examples 10, 17, and 19 could significantly predict the level of Innovativeness of the combined design Example 21 (Table 1). There was a catch, however: The higher idea 10 (moth mantle) was rated as creative, the more it decreased the overall design for being innovative!

Table 1. Creative ideas explaining innovativeness of integrated design (N = 20).

E	β	β_{st}	t	p	$r_{partial}$	r_{part}	
10	-.821	-.731	-2.504	.024	-.543	-.378	significant negative contribution
14	.353	.285	1.296	.215	.317	.196	positive contribution, n.s.
17	.815	.630	3.359	.004	.655	.508	significant highest positive contribution
19	1.046	.676	3.014	.009	.614	.455	significant medium positive contribution

Multiple regression of the Creativity scores without extremes (n = 18) to predict the Innovativeness score to Example 21 again led to significant results: $F_{(4,13)} = 4.24$, $p = .021$ with $R^2 = .57$, $R^2_{adj} = .43$. The only change was the level of contribution of the individual predictors (Table 2).

Table 2. Creative ideas explaining innovativeness of integrated design (n = 18).

E	β	β_{st}	t	p	$r_{partial}$	r_{part}	
10	-.817	-.616	-2.242	.043	-.528	-.409	significant negative contribution
14	.360	.212	1.008	.332	.269	.184	positive contribution, n.s.
17	.764	.517	2.498	.027	.570	.456	significant medium positive contribution
19	1.036	.679	2.733	.017	.604	.499	significant highest positive contribution

When framed as human-made, creativity of the single examples were a significant predictor of the innovativeness of the combined ideas, sometimes inversely related (10:

 doi:10.20944/preprints202206.0148.v1

moth mantle). When framed as AI-generated, model fit remained not significant (Appendix 3).

Apart from creativity, multiple regression also was performed on the Innovativeness scores to single ideas as predictors of the Innovativeness of the combined design. Again, in the Human condition alone, model fit was significant, not in the AI condition (Appendix 3).

Table 3. Innovation level of ideas explaining innovativeness of integrated design (n = 20).

E	β	β_{st}	t	p	$r_{partial}$	r_{part}	
10	-.519	-.439	-1.716	.107	-.405	-.282	negative contribution, n.s.
14	.299	.199	.877	.394	.221	.144	positive contribution, n.s.
17	.656	.621	3.332	.005	.652	.547	significant positive contribution
19	.235	.291	1.344	.199	.328	.221	positive contribution, n.s.

With n = 20 in the Human condition, multiple regression (model Enter) of level of Innovativeness of the single examples on the Innovativeness of Example 21 indicated that model fit was significant: $F_{(4,15)} = 5.53$, $p = .006$ with $R^2 = .60$, $R^2_{adj} = .49$. Table 3 shows that the innovation level of the integrated design depended solely on the innovation value of Example 17 (clothes-iron shoe). The pattern of results for the individual ideas in Table 3 remained the same when outliers and extremes were excluded (Appendix 3).

3.5. Effects of Agency (human vs AI) on Creativity and Innovation

Table 4 shows the mean values and standard deviations of Creativity and Innovativeness, without outliers. As said, none of the examples was seen as significantly innovative in the AI condition.

Table 4. Mean Creativity and Innovativeness for best design ideas (N = 40).

Agency		Mean Creativity for examples significantly creative in Human, AI, or both	Mean Innovativeness for examples significantly creative in Human, AI, or both	Mean Creativity for best examples in Human and (different) best examples in AI	Mean Innovativeness for examples in Human only
Human	*Mean*	4.3526	3.9868	4.5113	4.7593
	N	19	19	19	18
	SD	.72216	.60938	.64121	.58080
AI	*Mean*	4.4722	3.9750	4.3500	
	N	18	20	20	
	SD	.65242	.99638	.84102	
Total	*Mean*	4.4108	3.9808	4.4286	4.7593
	N	37	39	39	18
	SD	.68223	.81995	.74504	.58080

Whether outliers were included or not, GLM Multivariate (Pillai's Trace) of 2 Agency (Human vs AI) by 2 Measures (Mean Creativity and Mean Innovativeness) showed no significant effects for Agency. This was so over selected examples across conditions as well as for the averages of best creative examples in the Human and best creative examples in the AI condition.

No matter how it was analyzed (Appendix 3), the design researchers and educators were indifferent to the Agency producing the examples. They did not differentiate the best performing examples for the level of creativity or innovativeness according to their supposed human or artificial origin.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2022 doi:10.20944/preprints202206.0148.v1

For a Bayesian t-test, Bayesian Independent-Sample Inference was run for $N = 40$, with Agency as independent and Mean Creativity for best examples in Human and for (different) best examples in AI as the dependent. Characterization of posterior distributions and Bayes factor (BF) were obtained, variances were assumed unequal, selecting diffuse priors on mean conditional on variance. In testing the Null against the alternative (BF_{01}), for $N = 40$ (with outliers), $BF_{01} = 4.25$, $t_{(38)} = -.163$, $p = .871$, CI [-.59 – .50]. For $N = 39$ (without outliers), $BF_{01} = 3.50$, $t_{(37)} = -.671$, $p = .507$, CI [-.66 – .34].

Whereas frequentist t indicates difference (not significant), BF_{01} tests for sameness (H0 confirmed). In following Lee and Wagenmakers' (2013, p. 122) reading of Jeffreys (1961), $3 < BF_{01} < 10$ is considered moderate evidence for H0. In other words, the design researchers and educators were not just indifferent to the Agency producing the examples but regarded the work of human and AI as more-or-less of equal quality.

3.6. Verbal conceptualizations of creativity and innovativeness

The verbal responses to the questions what being creative and being innovative are about were analyzed for frequency of occurrence of words and associations (concepts). The two frequency lists were each thematically clustered (e.g., 'ideas,' 'things,' 'context'), deleting words related to specific examples like 'fish' and 'clothes iron.' The number of concepts gathered under each cluster determined the ranking of thematic clusters, position 1 covering the highest number of interrelated concepts. Frequency lists, thematic clusters, and raw text data can be looked up in Appendix 3. The interpretation of these results is part of the Discussion.

4. Discussion

Just like with human-made designs, not everyone saw every design idea as creative all the time. Similar to Lee and Ostwald (2022), the designers in the experiment showed higher levels of divergent thinking and ideation whereas others did less so:

> … what I saw was too definite or too obvious so I see no creativity in that. Fish is fish, I cannot see that as a hat. (Participant 31, AI condition)

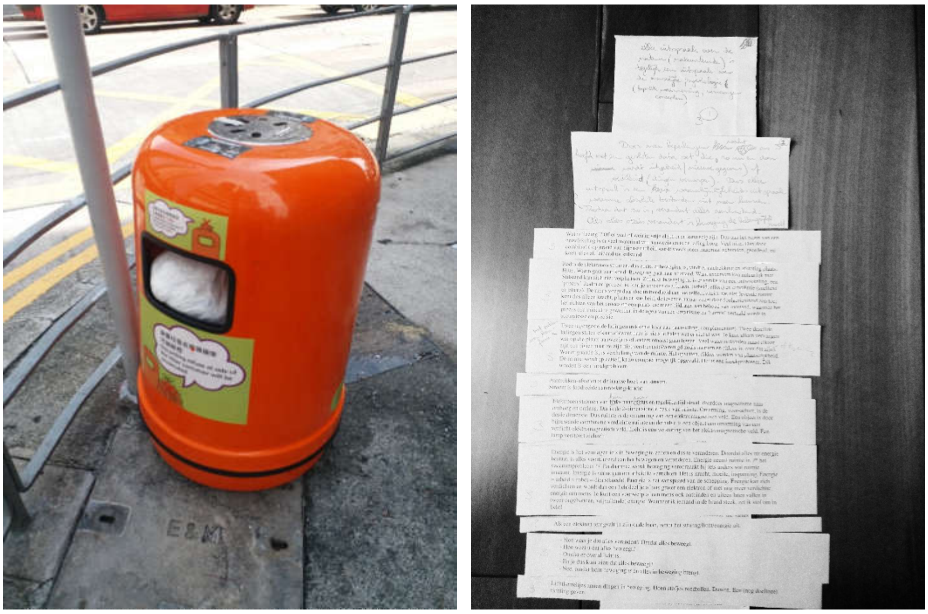
The controversy was visible in the ranking of examples according to their standard deviations, but Spearman *rho* did not indicate a significant coherence between rankings. Ergo, the design educators and researchers showed little coherence in what was a creative and what an innovative idea, whether forwarded by a human or by an AI system. To illustrate the controversy (Figure 9):

> The garbage-can idea can revolutionize the whole garbage industry – you can generate a new line of thinking from that. Who could be fed? Animals, city scavengers? It could be really interesting. (Participant 11, Human condition)

> Example 18: Trash can is dirty, not good association for me, therefore 2 and 2. (Participant 22, Human condition)

> For example, the bridge, if we see it as a railway then the picture suggests how we can use bridge construction in railway systems. The materiality of the slipper could sustain a boat. That's how I see innovative. (Participant 30, AI condition)

> Some visual element has association like shadow or stack of documents, yeah that similarity is interesting, I would never see a building in them. Bridge and railway is too obvious, too familiar. (Participant 34, AI condition)



A. Hong Kong trash can as feeder? **B**. Document stack resembles a building

Figure 9. A. Controversial: innovative food-dispenser for the poor or plain dirty? B. Interesting similarity.

This lack of coherence between rank-orders seems to signify that on the level of individual design ideas, designers differentiate between human-made and AI-generated content although the examples were the same, which suggests a perceptual bias, favoring humans particularly for estimated innovation value.

4.1. Specific design ideas

To know how good each example was, one-sample *t*-tests looked if creativity and innovativeness significantly differed from scale-midpoint (3.5). Across conditions ($N = 40$), 50% of examples were regarded as creative, 25% had innovation value (note: creativity was not a prerequisite to be innovative and v.v., see the umbrella quote next). When framed as human-made (n = 20), 38% of examples were judged as creative, 14% as innovative. When perceived as AI-generated (n = 20), 50% were seen as creative, but none as innovative.

> … some examples are obvious by the shape, form, texture that it looks like another object. How much is put into emotional thinking and exaggeration and metaphor, how much in-depth relationship between real object and what s/he is thinking. Sometimes there is little creativity because obvious, not special. No doubt the umbrella shadow (Example 11) looks like a building but how creative that is, it depends. For a designer, that is not so creative but as a general statement, that is okay. For 'normal' people, Example 11 is more creative. (Participant 10, Human condition)

> The shadow umbrella innovative, yes, because you can use that shape to build a building after. (Participant 10, Human condition)

As said, creativity was not necessary for being innovative and v.v. For that matter, rankings for the largest spread in scores indicated that in the human-made condition, most remarkably, the least controversial creative example (19) was the most controversial for its innovative value. As Participant 01 remarked:

> Some of the ideas of this student are really interesting and creative but not as a novel product. Student has very good imagination, associates one thing with another thing. (Participant 01, Human condition)

Likewise, Participant 31 said on innovation:

> The examples gave some illusions of reference of creation but it was not actually manipulating anything to make something new. (Participant 31, AI condition)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2022 doi:10.20944/preprints202206.0148.v1

In addition, Participant 22 stated:

Everybody can be creative but not innovative. (Participant 22, AI condition)

4.2. On the combined design ideas

Certain examples were creative and innovative beyond doubt, according to one-sample *t* (3.5 scale midpoint as test value), some of those just for Human, other just for AI, and a small number for both conditions, surprisingly the ones that made up the integrated fashion design. Regression analysis showed that the mean creativity of these best examples explained 60% to 80% of their level of innovativeness.

Three examples were significantly creative and innovative, being 10: moth mantle, 14: lampshade skirt, and 17: clothes-iron shoe. Together with the more controversial tropical-fish hat (19), these made up the integrated fashion design. Take notice that the fashion design was sketched before experimentation and that participants saw the sketch last, after first having rated the individual design ideas.

Ranking results for the combined Example 21 were often in the mid-range (Appendix 3). Some found the fashion design very creative and innovative whereas it did not work for others. On the upside of things, Participants 29, 32, and 16 remarked:

The fashion design is a new offering to the fashion context of industrial design plus bio-clothing. Very nice drawing. Let me take a picture for my wife. (Participant 29, AI condition)

The fashion design has some innovativeness for a new product that can be standing out in the market as unique selling point. Different and unique to sell. (Participant 32, AI condition)

This is pretty cool! (Participant 16, AI condition)

On the downside of things, Participants 27 and 20 thought that:

When I see the fish in my mind, it is too flat for a hat. I see the 2D image in 3D, trained to do so. Then it is not deep enough for a hat. The fashion design sees it as a feature of a hat, not as the hat itself. If the idea of 19 is 'the decoration on a hat,' it makes sense again. (Participant 27, AI condition)

In the combined fashion design, I found those four single ideas the most creative but put together, it seems not reflecting the creativity I expected. (Participant 20, Human condition)

It seems, then, that level of creativity is related to level of remoteness of association:

Creativity is able to make unexpected connections or making a link between things most others find dissimilar. To me it is how far-fetched or ridiculous the connection is. (Participant 04, Human condition)

The farther away from what it is, the more creative. (Participant 29, AI condition)

Lampshade I could also see it as skirt. I find that less creative than something I do not expect. Surprising ideas is important. (Participant 21, AI condition)

If it is very familiar (bridge-railway) I do not find it creative. It must be something I have not thought of. It does not have to be practical but it must have meaning. Like the wasp nest looks like food is not meaningful to me (Participant 14, Human condition)

With the organic pictures, the AI did quite well, the linear pictures (shadow-building) was a bit too obvious. (Participant 24, AI condition)

Creativity is the analogue of one object belonging to one category but then you refer to another category. Something different than what is shown from the photo. Association is based on visuals, e.g., shape of the shoe is like a boat is creative but there is nothing new here. Not much distance between the categories. Well-known in our culture. Clothes iron has shape like shoes but categories are remote. Interesting. Arduino-board bag is no nonsense. A really good design concept. (Participant 06, Human condition)

However, associative distance (remote or not) apparently was restricted by context (the other ideas, the overall concept). In the context of the fashion design, the categorical mismatch may have been perceived as smaller than presented in isolation that a clothes iron is a shoe.

Although for the best performing design ideas, level of creativity largely explained their innovation value, multiple regression showed that overall ($N = 40$), their level of creativity could not explain the creativity of the integrated fashion design nor its innovativeness. Only when seen as *human*-made ($n = 20$) was creativity of the single examples a significant predictor of the innovativeness of the combined ideas. Seen as AI, creativity had no explanatory power for level of innovativeness. Be aware, however, that single design ideas may negatively contribute to innovativeness (i.e. 10: moth mantle), an inverse relationship: Sometimes, a single idea that is highly creative reduces the innovation value of the integrated whole.

Lampshade is higher on innovation, using this material for a skirt, you need to push the technology or engineering how that could be done. But an organic fish is already close to a human being, organic, it borders on being human, not so innovative. (Participant 15, Human condition)

In the same vein was the framing as human-made crucial for the evaluations of the innovativeness of individual ideas and of the combined design. Multiple regression showed that due to the scores in the human condition, the level of innovativeness of the single examples was a significant predictor of the overall innovativeness of the integrated fashion design. If the integrated design was seen as based on AI-generated content, this was not so.

Student repurposes the iron or the lampshade and so changes the form and material of the new fashion. (Participant 11, Human condition)

Again, note that when partialed out, the only single contributor to the innovation value of the overall fashion design was 17, the clothes-iron shoe. Thus, some highly creative ideas may counter the innovation value of the whole but the whole of innovativeness may rely on that one single golden idea, the rest being supplemental (Figure 10).

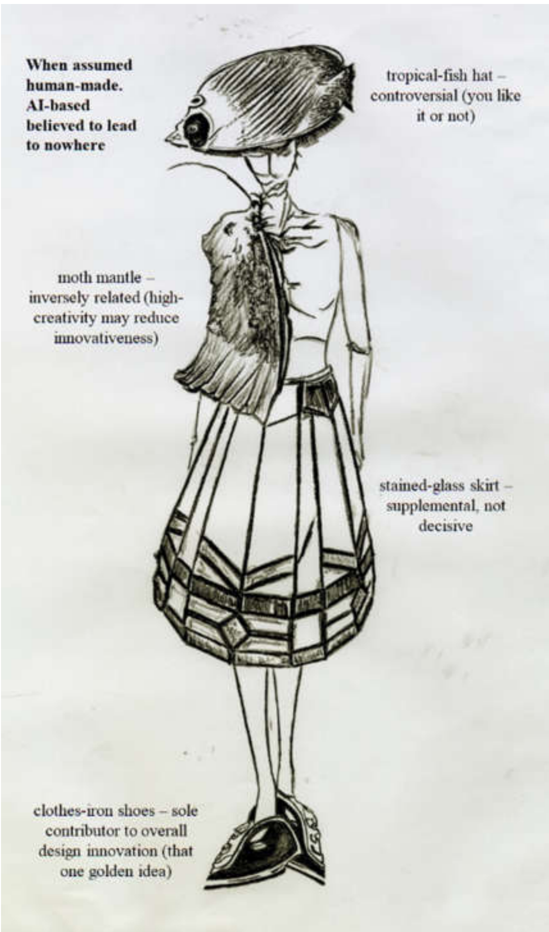


Figure 10. Deconstruction of innovativeness of the combined fashion design.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2022 doi:10.20944/preprints202206.0148.v1

4.3. Meaning

What did the measures mean? Why did the design teachers and researchers differentiate creativity from innovativeness although these notions are so much related? The overlapping part in the meaning of creativity and innovativeness was that both were about novel ideas. The differentiating part was that creativity is freely associative and imaginative whereas innovation restricts those new ideas to practical value, implementation, technology, use, feasibility, and monetary value. The numbers in the following summery refer to the frequency of occurrence of words and associations in a thematic cluster such as 'ideas' or 'cultural meaning' (Appendix 3). Clusters go first that have the highest frequency of mentioning interrelated concepts. According to the design educators and researchers:

Creativity is about 1. ideas, the expression and association of ideas, funny and surprising, regarding 2. things, out of old things, you create novel products, new solutions, a new design, away from common sense. 3. Creative thinking is about, for instance, 4. shape, making 5. (aesthetic) combinations of related concepts through association as a 6. way of understanding, a way of interpretation, an interesting way with a fresh eye for a good design concept. Creativity also uses 7. basic tools on simple elements, not necessarily accurate, relating to 8. the meaning of a culture, educational background, the social environment, and ourselves as a species, to attend to a specific 9. pattern or context such as clothing.

Innovativeness is about 1. innovation, being innovative as in not done many times, away from the state of affairs; it is novelty, which comes from the word 'nova,' meaning new, new in meaning, purpose, a new line, new products, driven by 2. (creative) ideas, 3. referring to creativity (not just an illusion of reference to it), a form of creativity that is based on (visual) similarity; a line of thinking linked up with personal experience but it has 4. value, new value, application value, important to the commercialization process, making profit, offering a unique selling point, having a first go at 5. things that are new, regular and simple in a 6. context of innovation and use, seeing the bigger topic, change the way of living. Innovation relates to the 7. material side of things and to developing 8. scenario's that are grounded on 9. common sense, data collection, and adding a layer of information, while having a sense of control. It reckons with a 10. community of practice, focuses on the practical invention side and the practial possibilities to come to design in a specific domain such as fashion or industry.

4.4. Agency

When Sam was believed to be a person, design ideas were assumed to be culturally contextualized (both for creativity and innovativeness) and to have practical meaning (innovativeness). When that same person based his/her design on computer-generated content, it was assumed that cultural relevance and practical use were absent and so the entire integrated design was deemed void of meaning, although it was made by human hand (i.e. the author alias the fashion student). Quotes by Participant 10 and 15, assuming the examples were student work and pointing out its shortcomings:

> Creativity comes from meaning of culture, should have meaning, cannot be anything that pops to mind. What I see here, s/he is directly reacting on what is shown: tactility, material, very direct associations. Not looking deeper, not much cultural relevance. Take and use only. (Participant 15, Human condition)

A just remark as the ideas were mindlessly generated by a faulty AI. Participants 26, 28, and 29 in the AI condition added:

> It needs to deliver output; not just a thought. It must be a thought in context. In all of us, us species. (Participant 26, AI condition)

> Idea 21 I rated lower, the ideas in isolation are reasonable but contextualized in a more complicated network, put into scale, human body, makes our judgment more clear. The context puts

 doi:10.20944/preprints202206.0148.v1

more boundaries on the design. So the ideas together make little sense: clothes iron with fish? Design is not just putting together some wild ideas. (Participant 28, AI condition)

When the AI got the context right, I went into 'agree', when it got it wrong I tended to 'disagree.' (Participant 29, AI condition)

Nonetheless, being 'unaware' and not knowing context actually may be conducive to creativity (Figure 11):

When you know more, it becomes harder to be creative. When you know little, it is easy to think you are creative. (Participant 05, Human condition)

Idea 02 I rated lower. I recognized the material of the statue and therefore could not see it as a mountain anymore. But my kid would, who has no clue about materials. (Participant 27, AI condition)

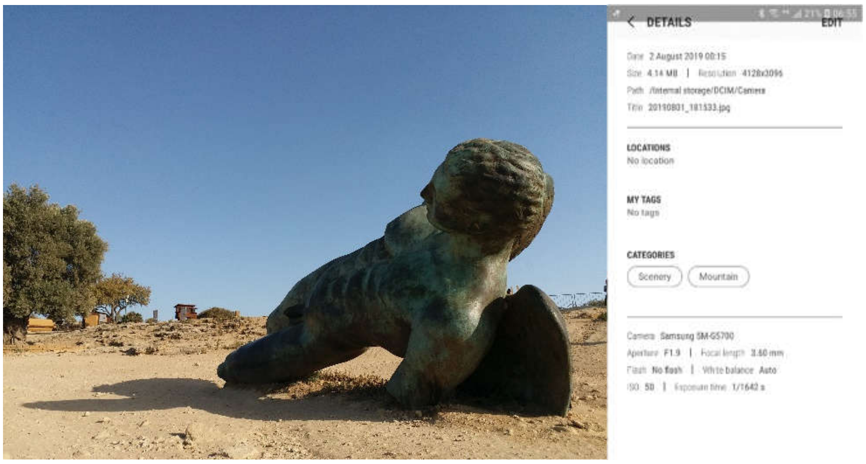


Figure 11. Example 02: Expert knowledge may prohibit seeing a broken statue as mountain scenery.

The upshot is this: Multivariate Analysis of Variance did not detect statistically significant differences between the scores for the best performing examples in the human versus (partially different) best examples in the AI condition. Alternatively, a Bayesian *t*-test looked for sameness, whether the Null could be confirmed as theoretically relevant, for instance, when human and robot arguably are treated as equals (cf. Computers As Social Actors) or when an AI is expected to pass the Turing test (Feigenbaum, 2003).

It is not too critical what the system suggests but it depends more on what the observer does with it. The tool may have to be adjusted for different users and their way of understanding. Most of the ideas are easy to understand and quite natural. I believe an AI should be like that. Like a human being. (Participant 27, AI condition)

According to Bayes, the faulty image-classification system passed the Feigenbaum variant of the Turing test. Bayesian *t* obtained moderately strong evidence that the design researchers and educators regarded the ideas by human or AI as more-or-less of equal quality, *in general*. You really have to look into the detail design of the single examples (Lee & Ostwald, 2022: "microscopic ideation") and their contribution to the whole to see that designers yet discriminate artificial against human origins of design.

I can see Sam does not see depth and does not know meaning. When meaning is involved, I find it more creative. Ice cream is desert: If it were a human, I would find it creative. Sam is an AI, it just zoomed in. Less creative. (Participant 33, AI condition)

5. Conclusion

All things considered, I conclude that false-class inclusions or category mismatches lie at the basis of creativity and certainly can have innovation value (H1). The individual design ideas that were beyond doubt for their creativity explained their innovation value for 60% to 80% with a clear preference towards human origins. Even the examples judged

 doi:10.20944/preprints202206.0148.v1

as 'too obvious' (e.g., slipper-boat or bridge-railway track) were acknowledged as 'known creativity,' not original but demonstrating that machines can make the same cross-category connections as humans did in the past:

> The process of creation brings something into being. However, it may be a repetitive act that creates more of an existing item, such as standardized sequence that yields a standardized result. (Lubert, 2018, p. 4)

Serendipity is an important impetus to the process of creation (H2). The algorithm of the Samsung Galaxy J5 Prime Gallery app produced plenty of misclassifications, which yet could be pitched to high-ranking professionals as creative ideas. After all, the erroneous algorithm passed the Feigenbaum-Turing test, according to Bayesian Independent-Sample Inference. Only in the detail design, the level of creativity of single ideas changed by the framing: Forwarded as human-made, design ideas went up or down the rankings as compared to AI-generated. Although judged creative, AI-generated content was not seen as innovative, which these design educators and researchers defined as practical, useful, adding value, culturally contextualized. This is what humans assume of other humans although that does not have to be the case either.

Errors are only errors within a particular system (H3). If the goal is to have more accurate classifications according to some predefined taxonomy, then indeed, the Samsung Galaxy J5 Prime Gallery app can be considered faulty and error-prone. If the goal is the transgression of category boundaries, so to kick-start the process of creative combination making, the false-class inclusions that the Samsung Gallery app produces may be the impetus to a creativity-support tool. Put differently, errors are serendipitous findings when perceived creatively (H4). Owing to a lack of knowledge, humans do misclassifications just the same: A wasp's nest is food (Samsung); a wasp's nest is food (Participant 23) (Figure 12).

Figure 12. Example 20: "Did not know it was a wasp's nest. I thought it was chocolate ice cream." (Participant 23, AI condition).

Therefore, machines can expose creative behaviors mindlessly (H5). The 'secret' algorithm driving the image classifications of the Samsung Galaxy J5 Prime Gallery app has no self-awareness. It does not know it produces mistakes and so does not know it produces creative crossovers. This demonstrates that like a brainstorm, idea generation may be done blindly, without deep understanding, context free. Evaluation in hindsight will sift out that one golden idea with practical innovation potential (Figure 10).

The current study underscored human inconsistent thinking about machine creativeness and human brightness. The Honda ASIMO robot is walking. We presumably agree on that. We also agree that the robot is not consciously aware that it walks. It just does. The Samsung image-classification system is creative. It attributes exemplars to remote categories based on matching features (cf. Han, Shi, Park, Chen, & Childs, 2018). We can also agree that this system is not consciously aware that it is creative. It just is. Like people who

 doi:10.20944/preprints202206.0148.v1

do serendipitous findings. ASIMO does not walk in exactly the same way as humans do. Neither does a dog or a cat. Yet, we regard all of that as walking. The Samsung app is not creative in exactly the same way as we are. Yet, we should regard it as being creative: serendipitous computational creativity or spontaneous machine creativeness.

It is certainly so that the Samsung Galaxy J5 Prime Gallery app did not do false-class inclusions, knowing what the proper class should be. However, it would not be a great effort to have a more accurate algorithmic classifier state the proper category (e.g., fish) and then make a cross-over to another category by loosening its restrictions (e.g., allow more fuzziness to its feature matching): Fish looks like hat. Even in its current state, however, the faulty Samsung app may be considered 'creative.'

Creative behaviors are an inherent aspect of information processing, whether human or AI (H6). We have witnessed the emergence of spontaneous creativeness of information processing, which is not limited to humans but is manifested in machines as well. As long as systems stay faulty, creative escapes from stifling order have a chance. This is important as 'anomalies' may not only force creative breakthroughs but are important to scientific discovery as well (cf. Giles & Walkowicz, 2019).

5.1. Creativity-support tool

In 2020, the global art market was valued by Statista Research Dept. at 50 billion US$ and comprised of 31 million transactions (Statista, 2021). Guttmann (2021) stated that visual/graphical out-of-home advertising (e.g., Figure 13) reached up to 29 billion US$ worldwide in 2020 with projections of 39.6 billion US$ for the year 2023.

Figure 13. Outdoor advertising: 'Bond girl,' holding a skin-care appliance as a 'gun'.

From a viewpoint of 'added value,' erroneous AI systems may increase the innovation value of designs as those systems may come up with cross-over suggestions that humans cannot think of because humans do not have the whole of the Internet as knowledge base.

This system would be good to suggest ideas to my advertising designs, I found the examples very creative. (Participant 39, AI condition)

Be aware that the functionality of erroneous AI may serve niche markets only as some design ideas in the current experiment worked for certain participants but not for other: Only 20% of the false-class inclusions in this mindless machine brainstorm (cf. Oliva & Elaziz, 2020) were agreeable with everyone. However, a 20% success rate already is a big achievement since merely 10% of 'raw ideas' brainstormed by humans typically make it to the next stage on the logarithmic curve of 'universal success' (Stevens & Burley, 1997/2016).

As we saw in the current experiment, single design ideas can be outstanding but integrated, the overall design ends up in the middle as an average of the good and not so good ideas. Since there is a 3,000-to-1 reduction (.03%) of raw ideas leading to 1 (!) successful commercial product (Stevens & Burley, 1997/2016) and since generally, brainstorm sessions suffer from many difficulties (Maaravi, Heller, Shoham, Mohar, & Deutsch, 2020), electronic brainstorm aid of a machine that fully resists group pressure may be welcomed for company survival (Mandal, 2020).

Acknowledgments: This study was supported by Project P0000254 - *AntiFix: Computational Creativity for Industrial Engineering,* funding scheme: Start-up Fund for New Recruits, grant number 1-BE22. Ethical approval was obtained from the Ethical Review Board of the university, file number HSEARS20210923002.

Conflicts of Interest: The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary Materials 1

Supplementary information to this article can be found online at https://doi.org/

Appendix B. Supplementary Materials 2

Supplementary information to this article can be found online at https://doi.org/

Appendix C. Appendix 3

Supplementary data and the technical report to this article can be found online at https://doi.org/

References

1. Benedek, M., Könen, T., & Neubauer, A. C. (2012). Associative abilities underlying creativity. *Psychology of Aesthetics, Creativity, and the Arts, 6*(3), 273-281. doi: 10.1037/a0027059
2. Blanchette, I., & Dunbar, K. (2000). How analogies are generated: The roles of structural and superficial similarity. *Memory and Cognition, 28*(1), 108-124. doi: 10.3758/bf03211580
3. Cheng, N. Y. W., & Hegre, E. (2009). Serendipity and discovery in a machine age: craft and a CNC router. *ACADIA 09: reForm*(), 284-286.
4. Feigenbaum, E. A. (2003). Some challenges and grand challenges for computational intelligence. *Journal of the Association for Computing Machinery, 50,* 32-40.
5. Giles, D., & Walkowicz, L. (2019). Systematic serendipity: a test of unsupervised machine learning as a method for anomaly detection. *Monthly Notices of the Royal Astronomical Society*, 484(1), 834-849. doi: 10.1093/mnras/sty3461
6. Guttmann, A. (September 2, 2021). *Global OOH ad expenditure 2000-2023.* Available from https://www.statista.com/statistics/273716/global-outdoor-advertising-expenditure/
7. Han, J., Shi, F., Park, D., Chen, L., & Childs, P. (2018). The conceptual distances between ideas in combinational creativity. In *Proceedings of the 15th International Design Conference DESIGN 2018 (DS 92)* (pp. 1857-1866). doi: 10.21278/idc.2018.0264
8. Hoorn, J. F. (2014). *Creative confluence.* Philadelphia, PA: John Benjamins.
9. Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Oxford University.
10. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, *25*, 1097-1105.
11. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM, 60*(6), 84–90. doi: 10.1145/3065386
12. Kumar, P., & Singh, N. (July 21, 2019). Introduction to Deep Learning with computer vision - kernels, channels & neural architecture. In *Hitchhiker's Guide to Deep Learning* (Chap. 5). Available from https://medium.com/hitchhikers-guide-to-deep-learning/5-introduction-to-deep-learning-with-computer-vision-kernels-channels-neural-architecture-41b6bc4befa7
13. Lee, J. H., & Ostwald, M. J. (2022). The relationship between divergent thinking and ideation in the conceptual design process, *Design Studies, 79,* 101089. doi: 10.1016/j.destud.2022.101089.
14. Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian modeling for cognitive science: a practical course.* New York: Cambridge University.
15. Lubert, T. (2018). Introduction. In T. Lubert (Ed.), *The creative process: Perspectives from multiple domains* (pp. 1-18). Cham, Switzerland: Springer.

16. Ma, C. (August 9, 2021). *'The Sparkle: Not Alive Yet Bright': Macabre God-like Robots and a poetic exploration of timeless Hong Kong.* Harbour Times. Available from https://harbourtimes.com/2021/08/09/the-sparkle-not-alive-yet-bright-macabre-god-like-robots-and-a-poetic-glimpse-into-hong-kongs-past-and-future/
17. Maaravi, Y., Heller, B., Shoham, Y., Mohar, S., & Deutsch, B. (2020). Ideation in the digital age: literature review and integrative model for electronic brainstorming. *Managerial Science, 15,* 1431-1464. doi:10.1007/s11846-020-00400-5
18. Mandal, P. C. (2020). How can we generate innovative ideas for new product development?. *International Journal of Quality and Innovation, 5*(1), 17-32.
19. Na, J., Jung, H., Chang, H. J., & Hwang, W. (2021). FixBi: Bridging domain spaces for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'21), June 19-25, Virtual Location* (pp. 1094-1103). Piscataway, NJ: IEEE and Computer Vision Foundation. doi: 10.1109/CVPR46437.2021.00115
20. Nguyen, A., Yosinski, J., & Clune, J. (2016). Understanding innovation engines: automated creativity and improved stochastic optimization via Deep Learning. *Evolutionary Computation, 24*(3), 545-572. doi: 10.1162/EVCO_a_00189
21. Niu, P., Niu, S., Liu, N., & Chang, L. (2019). The defect of the Grey Wolf optimization algorithm and its verification method. *Knowledge-Based Systems, 171,* 37-43. doi: 10.1016/j.knosys.2019.01.018
22. Oliva, D., & Elaziz, M. A. (2020). An improved brainstorm optimization using chaotic opposite-based learning with disruption operator for global optimization and feature selection. Soft Computing, 24, 14051-14072. doi: /10.1007/s00500-020-04781-3
23. Petra (Feb. 18, 2021). New glass planetarium to be built at Al Husseini Mosque at King's expense - Awqaf minister. *Petra, Jordan News Agency.* Available from https://www.petra.gov.jo/Include/InnerPage.jsp?ID=32436&lang=ar&name=en_news
24. QS World University Rankings (2019). *Subject: Art & Design.* Retrieved Dec. 17, 2021 from https://www.topuniversities.com/university-rankings/university-subject-rankings/2019/art-design
25. Ramakrishnan, N., & Grama, A. Y. (1999). Data mining: from serendipity to science. *Computer, 32*(8), 34-37.
26. Robert, F., & Robert, J. (1996). *Face to face.* Zurich, Switzerland: Lars Müller Publishers.
27. Rodgers, P. A., Green, G., & McGown, A. (2000). Using concept sketches to track design progress. *Design Studies, 21*(5), 451-464. doi: 10.1016/S0142-694X(00)00018-1
28. Ru, B., Li, D., Hu, Y., & Yao, L. (2019). Serendipity - a machine-learning application for mining serendipitous drug usage from social media. *IEEE Transactions on Nanobioscience, 18*(3), 324-334.
29. Simonton, D. K. (2022). Serendipity and creativity in the arts and sciences: A combinatorial analysis. In W. Ross, & S. Copeland (Eds.). *The art of serendipity* (pp. 293-320). Palgrave Studies in Creativity and Culture. Cham, Switzerland: Palgrave Macmillan. doi: 10.1007/978-3-030-84478-3_12
30. Siu, K. C., & Wong, P. (2002). *Designs you don't know what to do with.* Hong Kong SAR: MCCM creations.
31. Statista Research Department (August 5, 2021). *Art market worldwide - statistics & facts.* Available from https://www.statista.com/topics/1119/art-market/#topicHeader__wrapper
32. Stevens, G. A., & Burley, J. (1997/2016). 3,000 raw ideas= 1 commercial success!. *Research-Technology Management, 40*(3), 16-27. doi: 10.1080/08956308.1997.11671126
33. Thabet, R. Mahmoudi, R., & Bedoui, M. H. (2014). Image processing on mobile devices: an overview. In *Proceedings of International Image Processing, Applications and Systems Conference (IPAS'14), November 5-7, Sfax, Tunisia* (pp. 1-8), doi: 10.1109/IPAS.2014.7043267
34. Vosniadou, S., & Ortony, A. (1989). Similarity and analogical reasoning: a synthesis. In A. Ortony & S. Vosniadou (Eds.), *Similarity and analogical reasoning* (pp. 1-18), Cambridge, UK: Cambridge University.
35. Wang, J., & Hu, C. (2018). Similarity, metaphor and creativity. *Language and Semiotic Studies, 4*(3), 101-116.