Artificial Intelligence Assisted Creation
Fostering Inspiration & Raising Moral Issues

Matthias Wölfel

Faculty of Computer Science and Business Information Systems
Karlsruhe University of Applied Sciences
matthias.woelfel@hs-karlsruhe.de

Abstract. The promise of *artificial intelligence* (AI), in particular its latest developments in deep learning, has been influencing all kinds of disciplines such as engineering, business, agriculture, and humanities. More recently it also includes disciplines that were exclusively reserved for humans such as art and design. While there is a strong debate going on if creativity is profoundly human, we investigate if creativity can be fostered by AI. To get a better understanding of the creative potential offered by AI we open the black box and investigate where and how the magic is happening. Besides the potentials of AI, we also point out and discuss ethical and social implications caused by the latest developments in AI with respect to the creative sector.

Keywords: inspirational AI; human-machine co-design; moral issues

1 Introduction

Technological developments have been influencing all kinds of disciplines by transferring more competences from human beings to technical devices. The steps inculde [1]:

1. *tools*: transfer of mechanics (material) from the human being to the device
2. *machines*: transfer of energy from the human being to the device
3. *automatic machines*[1]: transfer of information from the human being to the device
4. *assistants*: transfer of decisions from the human being to the device

With the introduction of *artificial intelligence* (AI), in particular its latest developments in deep learning, we let the system (in step 4) take over our decisions and creation processes. Thus, tasks and disciplines that were exclusively reserved for humans in the past can now co-exist or even take the human out of the loop. It is no wonder that this transformation is not stopped at disciplines such as engineering, business, agriculture but also affects humanities, art and design. Each new technology has been adopted for artistic expression—just see the many wonderful examples in media art. Therefore, it is not surprising, that AI is going to be established as a novel tool to produce creative content of any form. However, in contrast to other disruptive technologies, AI seems particular challenging to be accepted in the area of art because it offers capabilities we thought once only humans are able to perform—the art is no longer done by artists using new technology to perform their art, but the art is done by the machine itself without the need for a human to intervene. The question of "what is art" has always been an emotionally debated topic in which everyone has a slightly different definition depending

[1] Automatic machine is called *Automat* or *automate* in other languages such as German or French respectively.

on his or her own experiences, knowledge base and personal aesthetics. However, there seems to be a broad consensus that art requires *human creativity and imagination* as, for instance, stated by the Oxford dictionary "The expression or application of human creative skill and imagination, typically in a visual form such as painting or sculpture, producing works to be appreciated primarily for their beauty or emotional power."

Every art movement challenges old ways and uses artistic creative abilities to spark new ideas and styles. With each art movement diverse intentions and reasons for creating the artwork came along with critics who did not want to accept the new style as an artform. With the introduction of AI into the creation process another art movement is trying to be established which is fundamentally changing the way we see art. For the first time, AI has the potential to take the artist out of the loop, to leave humans only in the positions of curators, observers and judges to decide if the artwork is beautiful and emotionally powerful.

2 Fostering Inspiration

While there is a strong debate going on in the arts if creativity is profoundly human, we investigate how AI can foster inspiration, creativity and produce unexpected results. It has been shown by many publications that AI can generate images, music and the like which can resemble different styles and produce artistic content. For instance, Elgammal et al. [2] have used *generative adversarial networks* (GAN) to generate images by learning about styles and deviating from style norms. The promise of AI-assisted creation is "a world where creativity is highly accessible, through systems that empower us to create from new perspectives and raise the collective human potential" as Roelof Pieters and Samim Winiger pointed out [3]. To get a better understanding of the process on how AI is capable to propose images, music, etc. we have to open the black box to investigate where and how the magic is happening.

2.1 Random & Constrained Variations

Random variations in the *image space* (sometimes also referred to as *pixel space*) are usually not leading to any interesting result. This is because semantic knowledge cannot be applied. Therefore, methods need to be applied which constrain the possible variations of the given dataset in a meaningful way. This can be realized by *generative design* or *procedural generation*. It is applied to generate geometric patterns, textures, shapes, meshes, terrain or plants. The generation processes may include, but are not limited, to self-organization, swarm systems, ant colonies, evolutionary systems, fractal geometry, and generative grammars. McCormack et al. [4] review some generative design approaches and discuss how art and design can benefit from those applications. These generative algorithms which are usually realized by writing program code are very limited. AI can change this process into data-driven procedures. AI, or more specifically artificial neural networks, can learn patterns from (labeled) examples or by reinforcement.

Before an artificial neural network can be applied to a task (classification, regression, image reconstruction), the general architecture is to extract features through many hidden layers. These layers represent different levels of abstractions. Data that have a similar structure or meaning should be represented as data points that are close together while divergent structures or meanings should be further apart from each other. To convert the image back (with some conversion/compression loss) from the low dimensional vector, which is the result of the first component, to the original input an additional component is needed. Together they form the *autoencoder* which consists of the *encoder* and the *decoder* . The *encoder* compresses the data from a high dimensional input space to a low dimensional space, often called the *bottleneck* layer. Then, the *decoder* takes this encoded input and converts it back to the original input as closely as possible. The *latent space* is the space in which the data lies in the bottleneck layer. If you look at Figure 1 you might be wondering why a model is needed that converts the input data into a "close as possible" output data. It seems rather useless if all it outputs is itself. As discussed, the latent space contains a highly compressed representation of the input data, which is the only information the decoder can use to reconstruct the input as faithfully as possible. The magic happens by interpolating between points and performing vector arithmetic between points in latent space. These transformations result in meaningful effects on the generated images. As dimensionality is reduced, information which is distinct to each image is discarded from the latent space representation, since only the most important information of each image can be stored in this low-dimensional space. The latent space captures the structure in your data and usually offers some semantic meaningful interpretation. This semantic meaning is, however, not given a priori but has to be discovered.

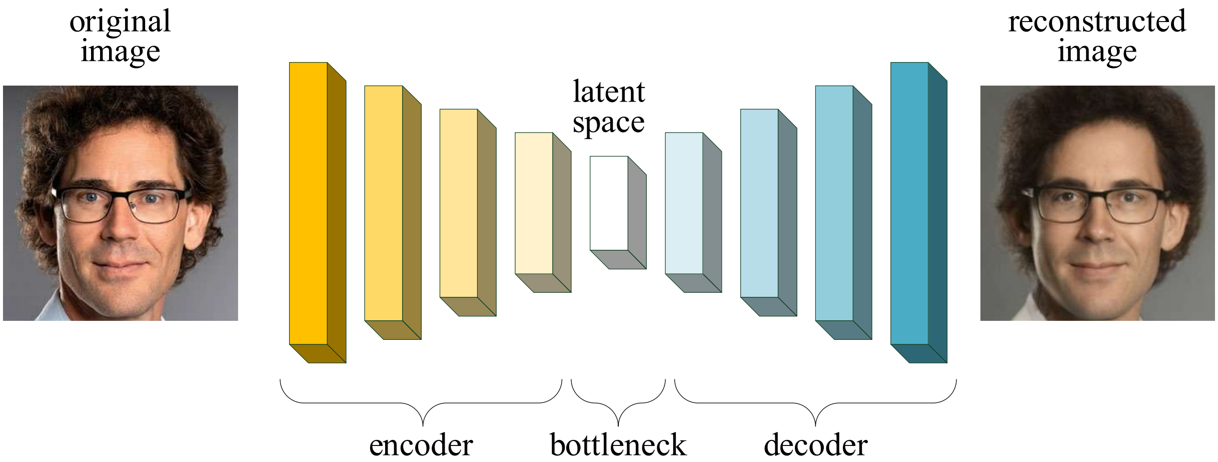

Fig. 1. The general architecture of an autoencoder.

2.2 Exploring the Latent Space

As already discussed autoencoders, after learning a particular non-linear mapping, are capable of producing photo-realistic images from randomly sampled points in the latent space. The latent space concept is definitely intriguing but at the same time non-trivial to comprehend. Although latent space means hidden, understanding what is happening in latent space is not only helpful but necessary for various applications. Exploring the structure of the latent space is both interesting for the problem domain and helps to develop an intuition for what has been learned and can be regenerated. It is obvious that the latent space has to contain some structure that can be queried and navigated. However, it is non-obvious how semantics are represented within this space and how different semantic attributes are entangled with each other.

To investigate the latent space one should favor a dataset that offers a limited and distinctive feature set. Therefore, faces are a good example in this regard because they share features common to most faces but offer enough variance. If aligned correctly also other meaningful representations of faces are possible, see for instance the widely used approach of *eigenfaces* [5] to describe the specific characteristic of faces in a low dimensional space.

In the latent space we can do vector arithmetic. This can correspond to particular features. For example, the vector $A_{smiling\ woman}$ representing the face of a smiling woman minus the vector $A_{neutral\ woman}$ representing a neutral looking woman plus the vector $A_{neutral\ man}$ representing a neutral looking man resulted in the vector $A_{smiling\ man}$ representing a smiling man.

$$A_{smiling\ woman} - A_{neutral\ woman} + A_{neutral\ man} = A_{smiling\ man}$$

This can also be done with all kinds of images; see e.g. the publication by Radford et al. [6] who first observed the vector arithmetic property in latent space. A visual example is given in Figure 2. Please note that all images shown in this publication are produced using BigGAN [7]. The photo of the author on which most of the variations are based on is taken by Tobias Schwerdt.

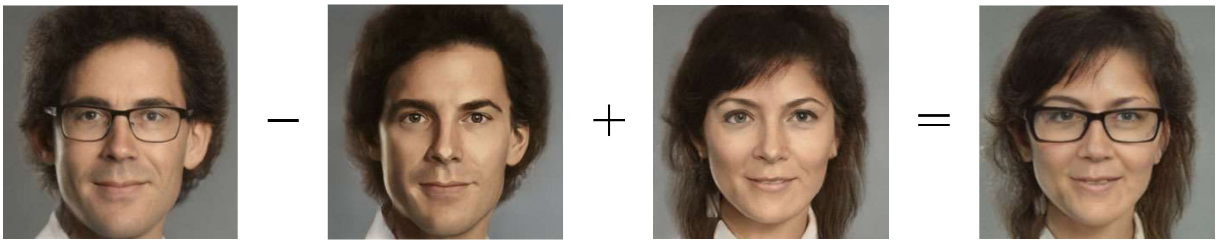

Fig. 2. In latent space, vector algebra can be carried out.

Semantic editing requires to move within the latent space along a certain 'direction'. Identifying the 'direction' of only one particular characteristic is non-trivial since editing one attribute may affect others because they are correlated. This correlation can be attributed to some extent to pre-existing correlations in 'the real world' (e.g. old persons are more likely to wear eyeglasses) or bias in the training dataset (e.g. more women are smiling on photos than men). To identify the semantics encoded in the latent space Shen et al. proposed a framework for interpreting faces in latent space [8]. Beyond the vector arithmetic property, their framework allows decoupling some entangled attributes (remember the aforementioned correlation between old people and eyeglasses) through linear subspace projection. Shen et al. found that in their dataset pose and smile are almost orthogonal to other attributes while gender, age, and eyeglasses are highly correlated with each other. Disentangled semantics enable precise control of facial attributes without retraining of any given model. In our examples, in Figures 3 and 4, faces are varied according to gender or age.

It has been widely observed that when linearly interpolate between two points in latent space the appearance of the corresponding synthesized images 'morphs' continuously from one face to another; see Figure 5. This implies that also the semantic meaning contained in the two images changes gradually. This is in stark contrast to having a simple fading between two images in image space. It can be observed that the shape and style slowly transform from one image into the other. This demonstrates how well the latent space understands the structure and semantics of the images. Other examples are given in Section 3.

Fig. 3. The *age* of a person can be changed by moving from the location of young to old in the latent space.

Fig. 4. The *gender* of a person can be changed by moving from the location of male to female in the latent space.

Even though our analysis has focused on face editing for the reasons discussed earlier it holds true also for other domains. For instance, Bau et al. [9] generated living rooms using similar approaches. They showed that some units from intermediate layers of the generator are specialized to synthesize certain visual concepts such as sofas or TVs.

So far we have discussed how autoencoders can connect the latent space and the image semantic space, as well as how the latent code can be used for image editing without influencing the image style. Next, we want to discuss how this can be used for artistic expression.

2.3 Sweet Spots in Latent Space

While in the former section we have seen how to use manipulation in the latent space to generate mathematical sound operations not much artistic content has been generated—just variations of photography like faces. Imprecision in AI systems can lead to unacceptable errors in the system and even result in deadly decisions; e.g. at autonomous driving or at cancer treatment. In the case of artistic applications, errors or glitches might lead to interesting, non-intended, artifacts. If those errors or glitches are treated as a bug or a feature lies in the eye of the artist. To create higher variations in the generated output some artists randomly introduce glitches within the autoencoder. Due to the complex structure of the autoencoder these glitches (assuming that they are introduced at an early layer in the network) occur on a semantic level as already discussed and might cause the models to misinterpret the input data in interesting ways. Some could even be interpreted as glimpses of autonomous creativity; see for instance the artistic work 'Mistaken Identity' by Mario Klingemann [10].

So far the latent space is explored by humans either by *random walk* or *intuitive steering* into a particular direction. It is up to human decisions if the synthesized image of a particular location in latent space is producing a visually appealing or otherwise interesting result. The question arises where to find those places and if those places can

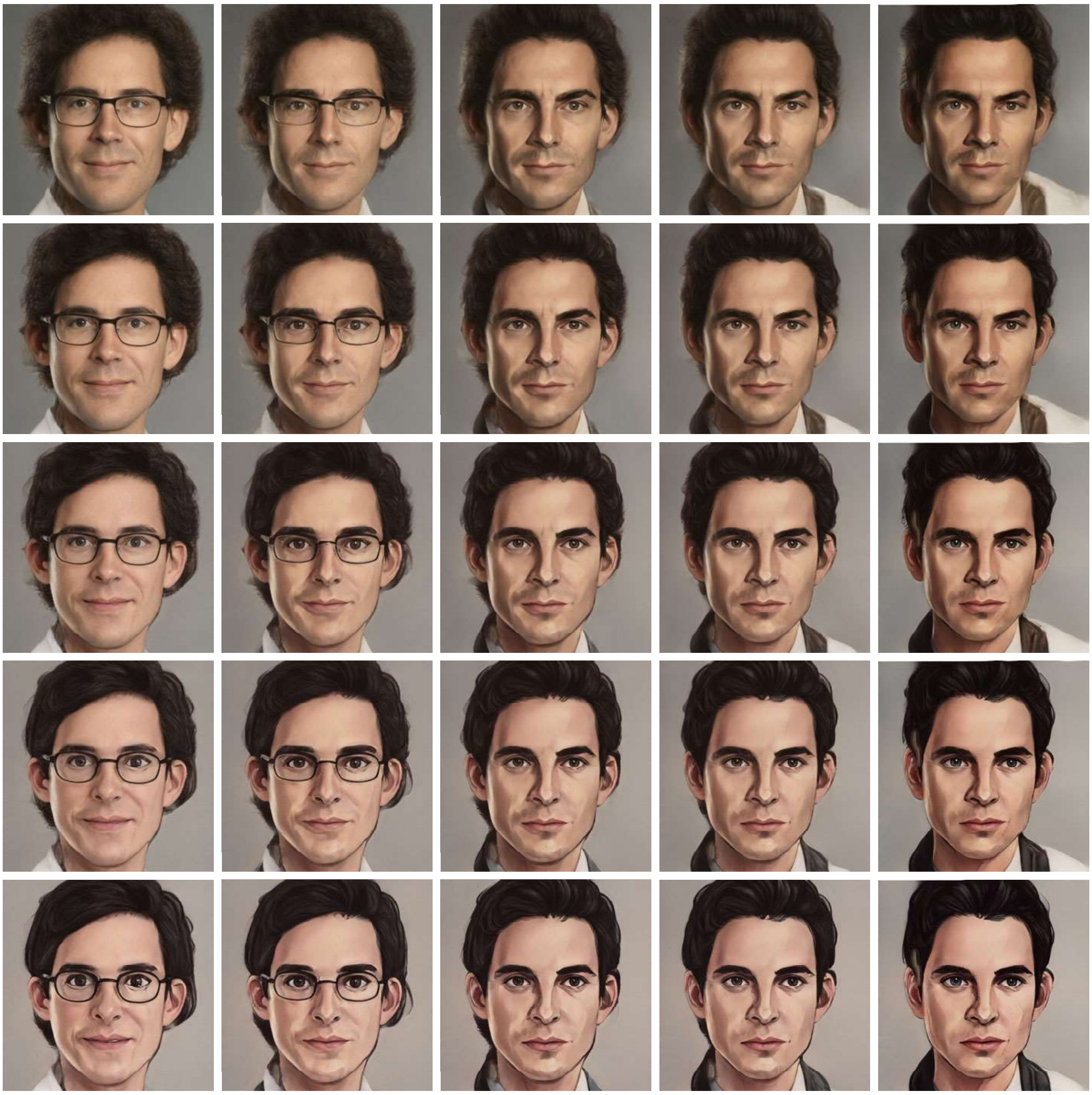

Fig. 5. The *facial features* of a person or the *style* of representation can be explored by changing the respective parameters in the latent space.

be spotted by an automatized process. The latent space is usually defined by a space of $d$dimensions for which it is assumed the data to be represented as multivariate Gaussian distributions $\mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ [11]. Therefore, the mean representation of all images lies in the center of the latent space. But what does that mean for the generated results? It is said that "beauty lies in the eyes of the beholder", however, research shows that there is a common understanding of beauty. For instance, averaged faces are perceived as more beautiful [12]. Adopting these findings to latent space let us assume that the most beautiful images (in our case faces) can be found in the center of the space. Particular deviations from the center stand for local sweet spots (e.g. female and male, ethnic groups). These types of sweet spots can be found by common means of data analysis (e.g. clustering). But where are interesting local sweet spots if it comes to artistic expression? Figure 6 demonstrates some variation in style within the latent space.

Of course, one can search for locations in the latent space where particular artworks from a given artist or art styles are located; see e.g. Figure 7 where the styles of different artists, as well as *white noise*[2], have been used for adoption. But isn't lingering around these sweet spots not only producing "more of the same"? How to find the local sweet spots which can define a new art style and can be deemed truly creative? Or do those discoveries of new art style lie outside of the latent space, because the latent space is trained within a particular set of defined art styles and can, therefore, produce only interpolations of those styles but nothing conceptually new?

Fig. 6. The *style* of the image can be changed (from left to right with increasing variation) by varying parameters in the latent space which represent style instead of facial features.

Fig. 7. The style of different source images is transferred to the target image. The styles of the images (from left to right) are Roy Lichtenstein, Friedensreich Hundertwasser, Joan Miro, Vincent van Gogh, and white noise.

[2] White noise is a signal with an equal spread frequency spectrum.

3 Example Artworks

So far we have discussed how AI can help to generate different variations of faces and where to find visually interesting sweet spots. In this section, we want to show how AI is supporting the creation process by applying the discussed techniques to other areas of image and object processing.[3]

3.1 Images

Probably, different variations of *image-to-image translation* are the most popular approach at least if looking at the mass media. The most prominent example is *style transfer*—the capability to transfer the style of one image to draw the content of another (examples are shown in Figure 7). But mapping an input image to an output image is also possible for a variety of other applications such as *object transfiguration* (e.g. horse-to-zebra, apple-to-orange, *season transfer* (e.g. summer-to-winter) or *photo enhancement* [13]. While some of the just mentioned systems are not yet in a state to be widely applicable, AI tools are taking over and gradually automating design processes which used to be time-consuming manual processes. Indeed, the most potential for AI in art and design is seen in its application to tedious, uncreative tasks such as coloring black-and-white images [14].

Marco Kempf and Simon Zimmerman used AI in their work dubbed 'DeepWorld' to generate a compilation of 'artificial countries' using data of all existing countries (around 195) to generate new anthems, flags and other descriptors [15]. Roman Lipski uses an *AI muse* (developed by Florian Dohmann et al.) to foster his/her inspiration [16]. Because the AI muse is trained only on the artist's previous drawings and fed with the current work in progress it suggests image variations in line with Roman's taste.

3.2 Objects

Cluzel et al. have proposed an interactive genetic algorithm to progressively sketch the desired side-view of a car profile [17]. For this, the user has taken on the role of a fitness function[4] through interaction with the system. The *chAIr Project* [18] is a series of four chairs co-designed by AI and human designers. The project explores a collaborative creative process between humans and computers. It used a GAN to propose new chairs which then have been 'interpreted' by trained designers to resemble a chair. *DeepWear* [19] is a method using deep convolutional GANs for clothes design. The GAN is trained on features of brand clothes and can generate images that are similar to actual clothes. A human interprets the generated images and tries to manually draw the corresponding pattern which is needed to make the finished product. Li et al. [20] introduced an artificial neural network for encoding and synthesizing the structure of 3D shapes which—according to their findings—are effectively characterized by their hierarchical organization. German et al. [21] have applied different AI techniques trained by a small sample set of shapes of bottles, to propose novel bottle-like shapes. The evaluation of their proposed methods revealed that it can be used by trained designers as well as non-designers to support the design process in different phases and that it could lead to novel designs not intended/foreseen by the designers.

---

[3] Of course these techniques have been also successfully applied to other areas such as audio and video, but should not be presented here.

[4] also referred to as objective function

4 Ethical and Social Implications

For decades, AI has fostered (often false) future visions ranging from transhumanist utopia to "world run by machines" dystopia. Artists and designers explore solutions concerning the semiotic, the aesthetic and the dynamic realm, as well as confronting corporate, industrial, cultural and political aspects. The relationship between the artist and the artwork is directly connected through their intentions, although currently mediated by third-parties and media tools. Understanding the ethical and social implications of AI-assisted creation is becoming a pressing need. The implications, where each has to be investigated in more detail in the future, include:

– *Bias*: Al systems are sensitive to bias. As a consequence, the AI is not being a neutral tool, but has pre-decoded preferences. Bias relevant in creative AI systems are:
  - *Algorithmic Bias* occurs when a computer system reflects the implicit values of the humans who created it; e.g. the system is optimized on dataset A and later retrained on dataset B without reconfiguring the neural network (this is not uncommon, as many people do not fully understand what is going on in the network, but are able to use the given code to run training on other data).
  - *Data Bias* occurs when your samples are not representative of your population of interest.
  - *Prejudice Bias* results from cultural influences or stereotypes which are reflected in the data.

– *Art Crisis*: Until 200 years ago painting served as the primary method for visual communication and was a widely and highly respected art form. With the invention of photography, painting began to suffer an identity crisis because painting—in its current form then—was not able to reproduce the world as accurate and with as low effort as photography. As a consequence visual artists had to change to different forms of representations not possible by photography inventing different art styles such as impressionism, expressionism, cubism, pointillism, constructivism, surrealism, up to abstract expressionism. At the time AI can perfectly simulate those styles what will happen with the artists? Will artists still be needed, be replaced by AI, or will they have to turn to other artistic work which yet cannot be simulated by AI?

– *Inflation*: Similar to the image flood which has reached us the same can happen with AI art. Because of the glut, nobody is valuing and watching the images anymore.

– *Wrong Expectations*: Only esthetic appealing or otherwise interesting or surprising results are published which can be contributed to similar effects as the well-known *publication bias* [22] in other areas. Eventually, this is leading to wrong expectations of what is already possible with AI. In addition, this misunderstanding is fueled by content claimed to be created by AI but has indeed been produced—or at least reworked—either by human labor or by methods not containing AI.

– *Unequal Judgment*: Even though the raised emotions in viewing artworks emerge from its underlying structure in the works, people also include the creation process in their judgment (in the cases where they know about it). Frequently, becoming to know that a computer or an AI has created the artwork, in the opinion of the people it turns boring, has no guts, no emotion, no soul while before it was inspiring, creative and beautiful.

– *Authorship*: The authorship of AI-generated content has not been clarified. For instance, is the authorship of a novel song composed by an AI trained exclusively on songs by Johann Sebastian Bach belonging to the AI, the developer/artist, or Bach? See e.g. [23] for a more detailed discussion.

– *Trustworthiness*: New AI-driven tools make it easy for non-experts to manipulate audio and/or visual media. Thus, image, audio as well as video evidence is not trustworthy anymore. Manipulated image, audio, and video are leading to fake information, truth skepticism, and claims that real audio/video footage is fake (known as the *liar's dividend*) [24].

5 Conclusion

The potential of AI in creativity has just been started to be explored. We have investigated on the creative power of AI which is represented—not exclusively—in the semantic meaningful representation of data in a dimensionally reduced space, dubbed latent space, from which images, but also audio, video, and 3D models can be synthesized. AI is able to imagine visualizations that lie between everything the AI has learned from us and far beyond and might even develop its own art styles (see e.g. deep dream [25]). However, AI still lacks intention and is just processing data.

Those novel AI tools are shifting the creativity process from crafting to generating and selecting—a process which yet can not be transferred to machine judgment only. However, AI can already be employed to find possible sweet spots or make suggestions based on the learned taste of the artist [21]. AI is without any doubt changing the way we experience art and the way we do art. Doing art is shifting from handcrafting to exploring and discovering. This leaves humans more in the role of a curator instead of an artist, but it can also foster creativity (as discussed before in the case of Roman Lipski) or reduce the time between intention and realization. It has the potential, just as many other technical developments, to democratize creativity because the handcrafting skills are not so much in need to express his/her own ideas anymore. Widespread misuse (e.g. image manipulation to produce fake pornography) can limit the social acceptance and require AI literacy. As human beings, we have to ask ourselves if feelings are wrong just because the AI never felt alike in its creation process as we do? Or should we not worry too much and simply enjoy the new artworks created no matter if they are done by humans, by AI or as a co-creation between the two ones?

References

1. Wölfel, M.: Der smarte Assistent. In: Smartphone-Ästhetik: zur Philosophie und Gestaltung mobiler Medien. Ruf, O. (editor). Transcript (2018) 269–288
2. Elgammal, A., Liu, B., Elhoseiny, M., Mazzone, M.: CAN: creative adversarial networks, generating" art" by learning about styles and deviating from style norms. arXiv preprint arXiv:1706.07068 (2017)
3. Pieters, R., Winiger, S.: Creative AI: On the Democratisation & Escalation of Creativity (2016) Accessed: 2020-03-14, medium, available at: https://medium.com/creativeai/creativeai-9d4b2346faf3.
4. McCormack, J., Dorin, A., Innocent, T., et al.: Generative design: a paradigm for design research. Proceedings of Futureground, Design Research Society, Melbourne (2004)
5. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of cognitive neuroscience **3**(1) (1991) 71–86
6. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434 (2015)
7. Brock, A., Donahue, J., Simonyan, K.: Large scale gan training for high fidelity natural image synthesis. arXiv preprint arXiv:1809.11096 (2018)
8. Shen, Y., Gu, J., Tang, X., Zhou, B.: Interpreting the latent space of gans for semantic face editing. arXiv preprint arXiv:1907.10786 (2019)

9. Bau, D., Zhu, J.Y., Strobelt, H., Zhou, B., Tenenbaum, J.B., Freeman, W.T., Torralba, A.: Visualizing and understanding generative adversarial networks. arXiv preprint arXiv:1901.09887 (2019)
10. Klingemann, M.: Mistaken Identity (2018) Accessed: 2020-03-10, available at: https://aiartists.org/mario-klingemann.
11. Miyato, T., Kataoka, T., Koyama, M., Yoshida, Y.: Spectral normalization for generative adversarial networks. arXiv preprint arXiv:1802.05957 (2018)
12. Trujillo, L.T., Jankowitsch, J.M., Langlois, J.H.: Beauty is in the ease of the beholding: A neurophysiological test of the averageness theory of facial attractiveness. Cognitive, Affective, & Behavioral Neuroscience **14**(3) (2014) 1061–1076
13. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Computer Vision (ICCV), 2017 IEEE International Conference on. (2017)
14. Zhang, G., Qu, M., Jin, Y., Song, Q.: Colorization for anime sketches with cycle-consistent adversarial network. International Journal of Performability Engineering **15**(3) (2019)
15. Kempf, M., Zimmerman, S.: Deepworld (2018) Accessed: 2020-03-14, medium, available at: http://kim.hfg-karlsruhe.de/deepworld/.
16. Lipski, R.: Artificial muse (2016) Accessed: 2020-03-14, artificialmuse, available at: https://www.artificialmuse.ai/.
17. Cluzel, F., Yannou, B., Dihlmann, M.: Using evolutionary design to interactively sketch car silhouettes and stimulate designer's creativity. Engineering Applications of Artificial Intelligence **25**(7) (2012) 1413–1424
18. Schmitt, P., Weiss, S.: The chair project—four classics (2018) Accessed: 2019-11-27, available at: https://philippschmitt.com/work/chair.
19. Kato, N., Osone, H., Sato, D., Muramatsu, N., Ochiai, Y.: Deepwear: a case study of collaborative design between human and artificial intelligence. In: Twelfth International Conference on Tangible, Embedded, and Embodied Interaction, ACM (2018) 529–536
20. Li, J., Xu, K., Chaudhuri, S., Yumer, E., Zhang, H., Guibas, L.: Grass: Generative recursive autoencoders for shape structures. Transactions on Graphics (TOG) **36**(4) (2017) 52
21. German, K., Limm, M., Wölfel, M., Helmerdig, S.: Co-designing object shapes with artificial intelligence. In: Proceedings of ArtsIT. (2019)
22. Dwan, K., Altman, D.G., Arnaiz, J.A., Bloom, J., Chan, A.W., Cronin, E., Decullier, E., Easterbrook, P.J., Von Elm, E., Gamble, C., et al.: Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PloS one **3**(8) (2008)
23. Rack, F., Vettermann, O.: KI-Kunst und Urheberrecht – die Maschine als Schöpferin? (2019) Accessed: 2019-11-27, telemedicus, available at: http://tlmd.in/a/3353.
24. Chesney, R., Citron, D.K.: Deep fakes: A looming challenge for privacy, democracy, and national security. California Law Review 1753; U of Texas Law, Public Law Research Paper No. 692; U of Maryland Legal Studies Research Paper No. 2018-21., available at: http://dx.doi.org/10.2139/ssrn.3213954 (2018)
25. Mordvintsev, A., Olah, C., Tyka, M.: Inceptionism: Going deeper into neural networks. (2019) Accessed: 2020-03-10, Google research blog available at: http://googleresearch.blogspot.co.uk/2015/06/inceptionism-going-deeper-into-neural.html.