The Legal Implications (if any) of Creative Adversarial Networks

Dr Eliza Mik

法律 CUHK LAW
THE CHINESE UNIVERSITY OF HONG KONG
香 港 中 文 大 學

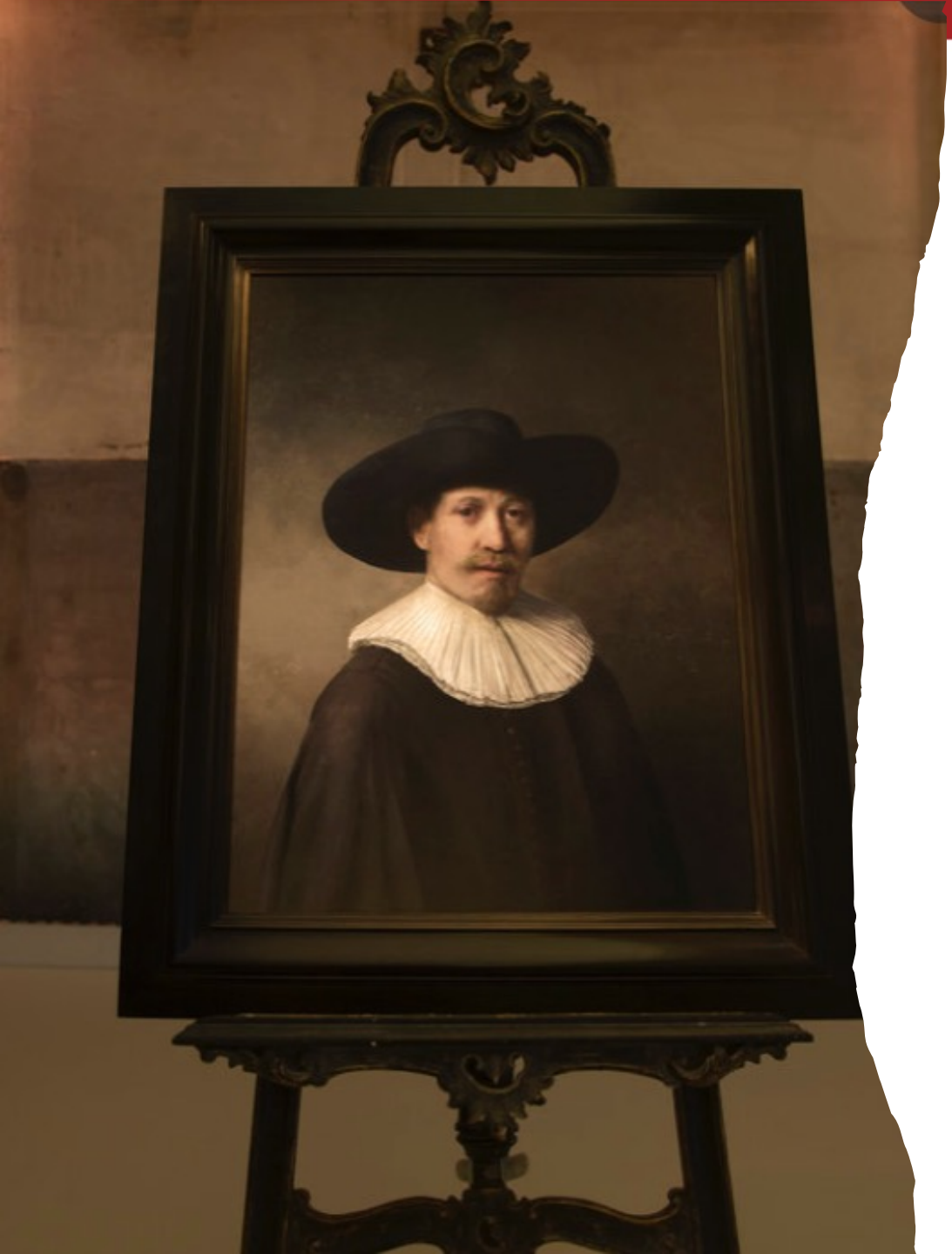
Why do we need to talk about CANs?

Turing Test / Visual Turing Test / Lovelace Test

Can AIs generate art # Can AIs be creative?

> The question of whether a computer can think is no more interesting than the question of whether a submarine can swim.

Edsger W. Dijkstra



The Next Rembrandt

Edmond de Belamy

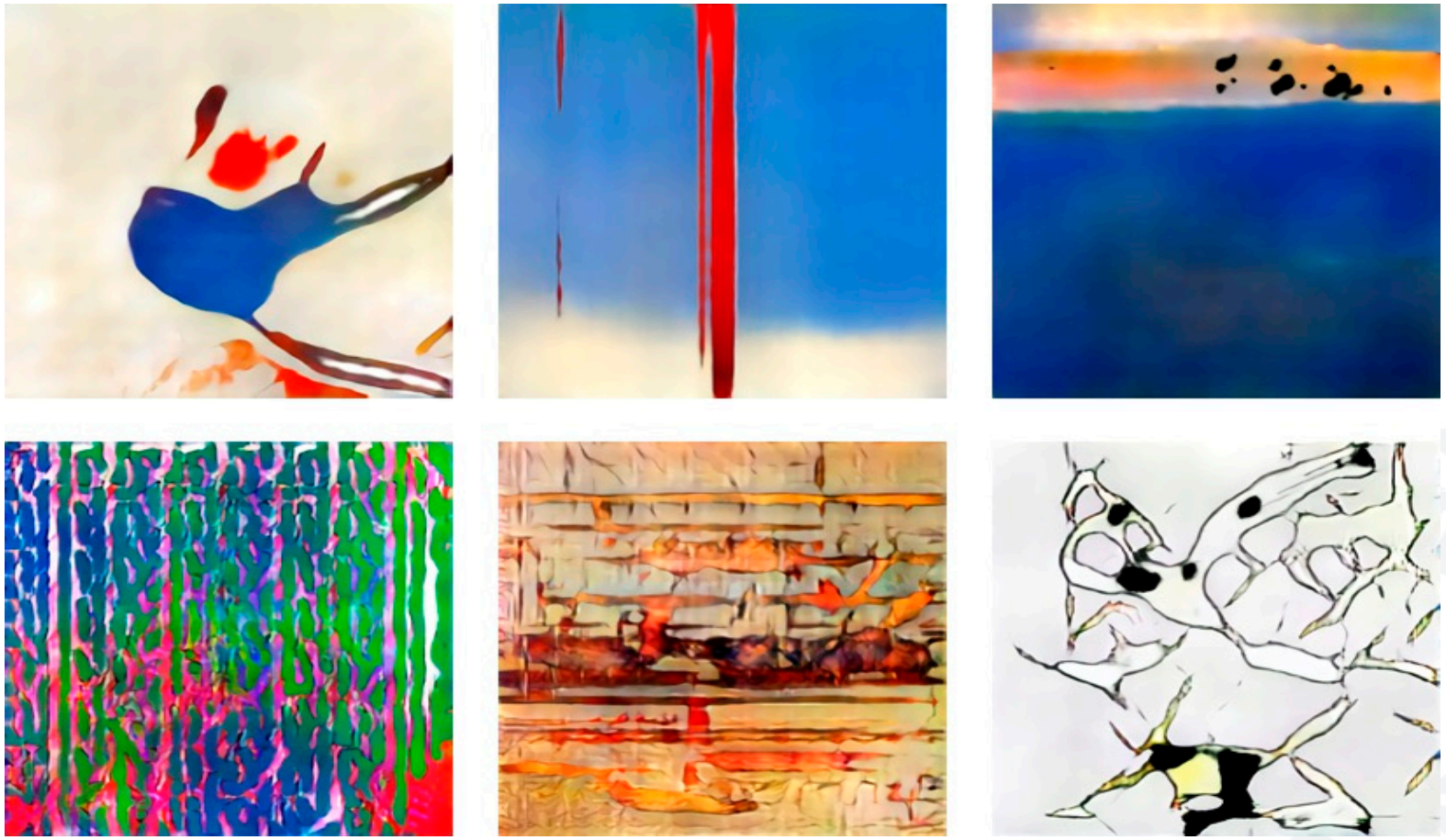
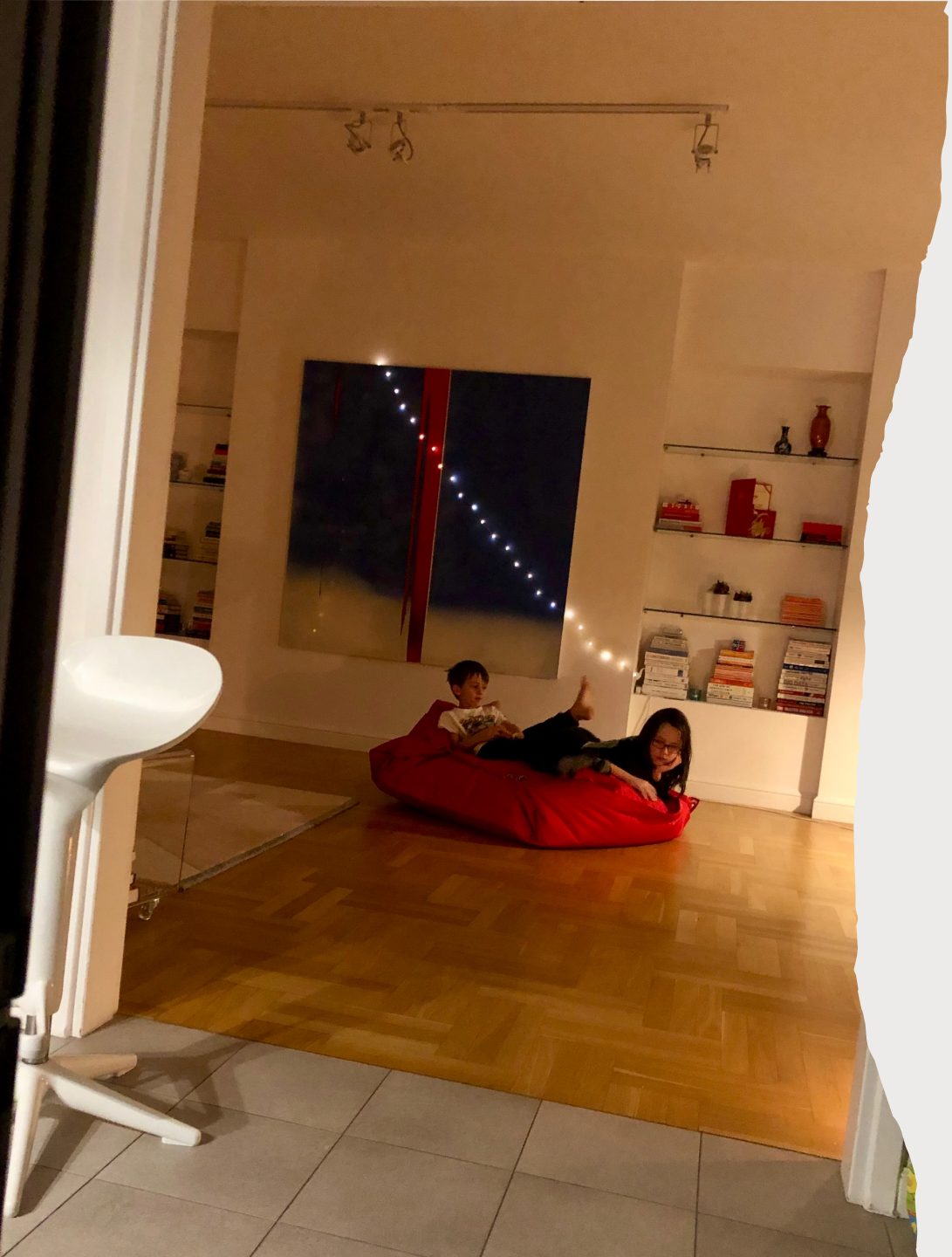


Figure 4. Examples of images generated by AICAN after training with images from all styles and genres from the past 500 years of Western art. Images courtesy of the Art & Artificial Intelligence Laboratory, Rutgers.



“split sunshine”

(AICAN + Grzegorz Szyma, Acrylic, 160x160, 2020)

Traditional Approach: Use AI to optimize the performance of certain tasks (i.e. automate repetitive, standardized tasks)

Novel Appraoch: Use AI to "touch the human soul" or "evoke emotions"

Natural Constraints:

Human Art: derives from prior art + experiencing the world

AI Art: derives from prior art. (lack of embodiment, "computer vision")

Q1: can you be creative, intelligent, conscious etc, without "seeing the world"?

Q2: How to integrate exposure to art with the creation of novel art?

How do we define / formalize:
Originality / Creativity / Art?

Creativity

Traditional approach:

= randomness

(loss of control, unpredictability)

Novel approach (CAN)

= novelty / deviation / influence

CAN: Creative Adversarial Networks Generating "Art" by Learning About Styles and Deviating from Style Norms*

Ahmed Elgammal[1†] Bingchen Liu[1] Mohamed Elhoseiny[2] Marian Mazzone[3]
The Art & AI Laboratory - Rutgers University
[1] Department of Computer Science, Rutgers University, NJ, USA
[2] Facebook AI Research, CA, USA
[3] Department of Art History, College of Charleston, SC, USA

June 23, 2017

Abstract

We propose a new system for generating art. The system generates art by looking at art and learning about style; and becomes creative by increasing the arousal potential of the generated art by deviating from the learned styles. We build over Generative Adversarial Networks (GAN), which have shown the ability to learn to generate novel images simulating a given distribution. We argue that such networks are limited in their ability to generate creative products in their original design. We propose modifications to its objective to make it capable of generating creative art by maximizing deviation from established styles and minimizing deviation from art distribution. We conducted experiments to compare the response of human subjects to the generated art with their response to art created by artists. The results show that human subjects could not distinguish art generated by the proposed system from art generated by contemporary artists and shown in top art fairs.

Theoretical Background for CANs

Colin Martindale (1943-2008) proposed a psychology-based theory that explains new art creation:

- **creative** = **increase the arousal potential of art to push against habituation = break with existing styles**

- increase must be minimal to avoid negative reaction by the observers ("**principle of least effort**")

Focus on progress in art, not just replicating or emulating – but creative element present!

Theoretical Background for CANs

D. E. Berlyne (1924-1976) psychophysical concept of "levels of arousal" to study & evaluate art

"**Arousal Potential**" refers to the properties of stimulus patterns that lead to raising arousal, the most significant arousal-raising properties for studying aesthetic phenomena are:

1. *novelty*
2. *surprisingness*
3. *complexity*
4. *ambiguity*
5. *puzzlingness*

"People prefer stimulus with a moderate arousal potential. Too little arousal potential is considered boring, and too much activates the aversion system, which results in negative response. This behaviour is explained by the Wundt curve that correlates the arousal potential with the hedonic response." (at 4)

Novel. But not too novel..

"The agent's goal is to generate art with increased levels of arousal potential in a constrained way without **activating the aversion** system and falling into the **negative hedonic range**." (at 4)

Generative Adversarial Networks

GAN: consists of two sub networks: a generator and a discriminator.

The **discriminator** has access to a set of images (training images) + tries to discriminate between "real" images (from the training set) and "fake" images generated by the generator.

The **generator** tries to generate images similar to the training set without seeing these images

+ starts by generating random images and receives a signal from the discriminator whether the discriminator finds them real or fake.

At equilibrium, the discriminator cannot distinguish between images generated by the generator and images in the training set, hence the generator succeeds in generating images that come from the same distribution as the training set.

Problem: GANs emulate a given **training distribution** but: limited in their ability to generate creative art (i.e. GANs try to match the training set)

Creative Adversarial Networks

The generator does not have access to any art. It generates art starting from a random input, but unlike GAN, it receives two signals from the discriminator.

The discriminator has a large training set with style labels (e.g. Baroque) and uses it to learn to discriminate between styles.

The discriminator sends two signals to the generator (which act as two contradictory forces)

1) **"art or not art."** (like traditional GANs)

2) fits into **"established styles"**

The creative generator will try to **generate art that confuses the discriminator**. (i.e. generator tries to fool the discriminator to think it is "art," and to confuse it about the style of the work generated!)

Second signal pushes the generator to generate **"style-ambiguous work"**

$$\min_G \max_D V(D,G) =$$

$$\mathbb{E}_{x,\hat{c}\sim p_{data}}[\log D_r(x) + \log D_c(c=\hat{c}|x)] +$$

$$\mathbb{E}_{z\sim p_z}[\log(1 - D_r(G(z))) - \sum_{k=1}^{K}\big(\frac{1}{K}log(D_c(c_k|G(z)) +$$

$$(1-\frac{1}{K})log(1 - D_c(c_k|G(z)))\big)],$$

Training Dataset + Validation

Dataset: CAN was trained using paintings from the publicly available WikiArt dataset = images of 81,449 paintings from 1,119 artists ranging from the Fifteenth century to Twentieth century…

Validation: Qualitative + Quantitative

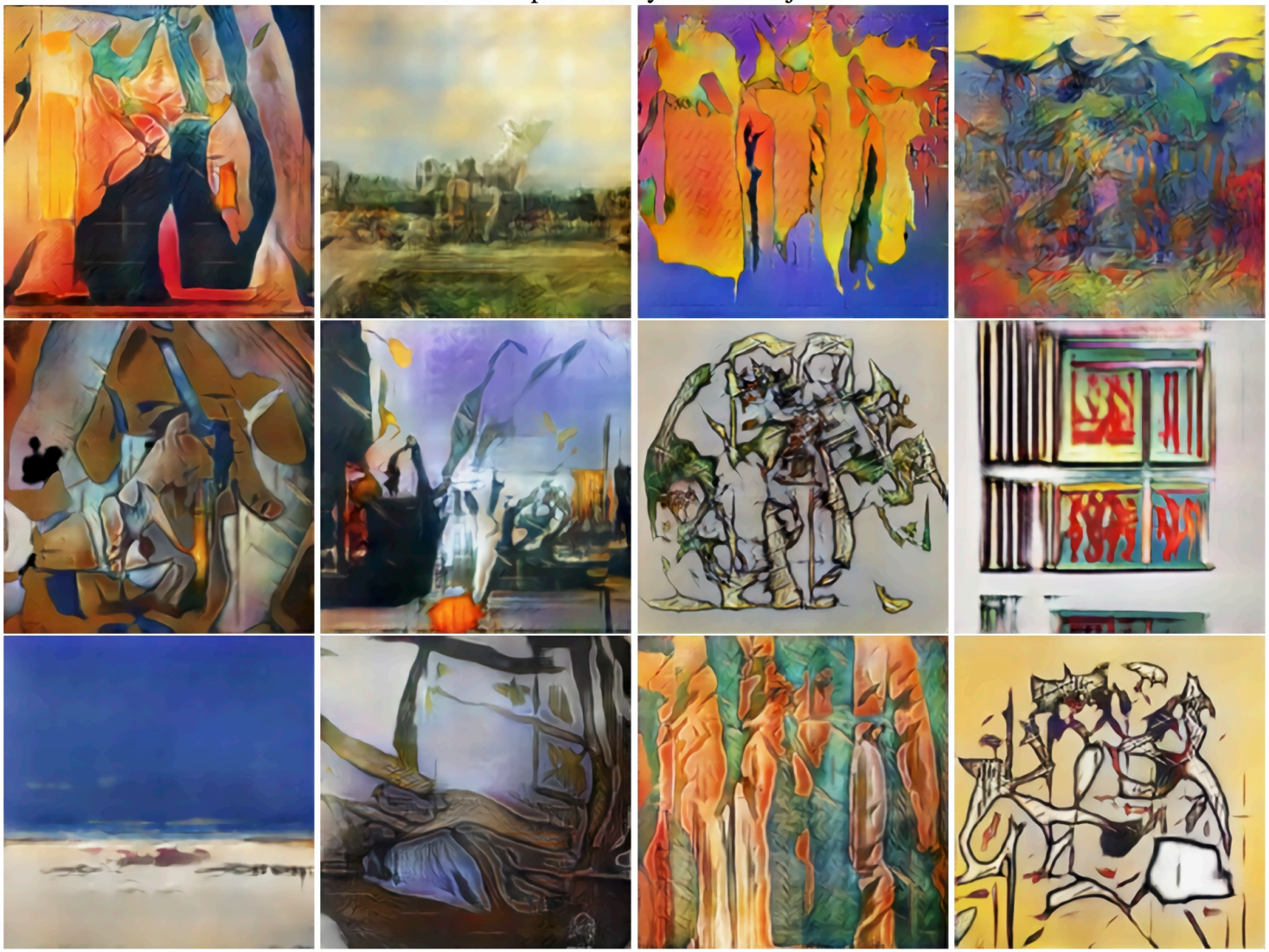
Qualitative: difficult to establish baseline, general observation; "lots of hallucination of portraits, landscapes, architectures, religious subject matter, etc."

Quantitative: "human testing" –

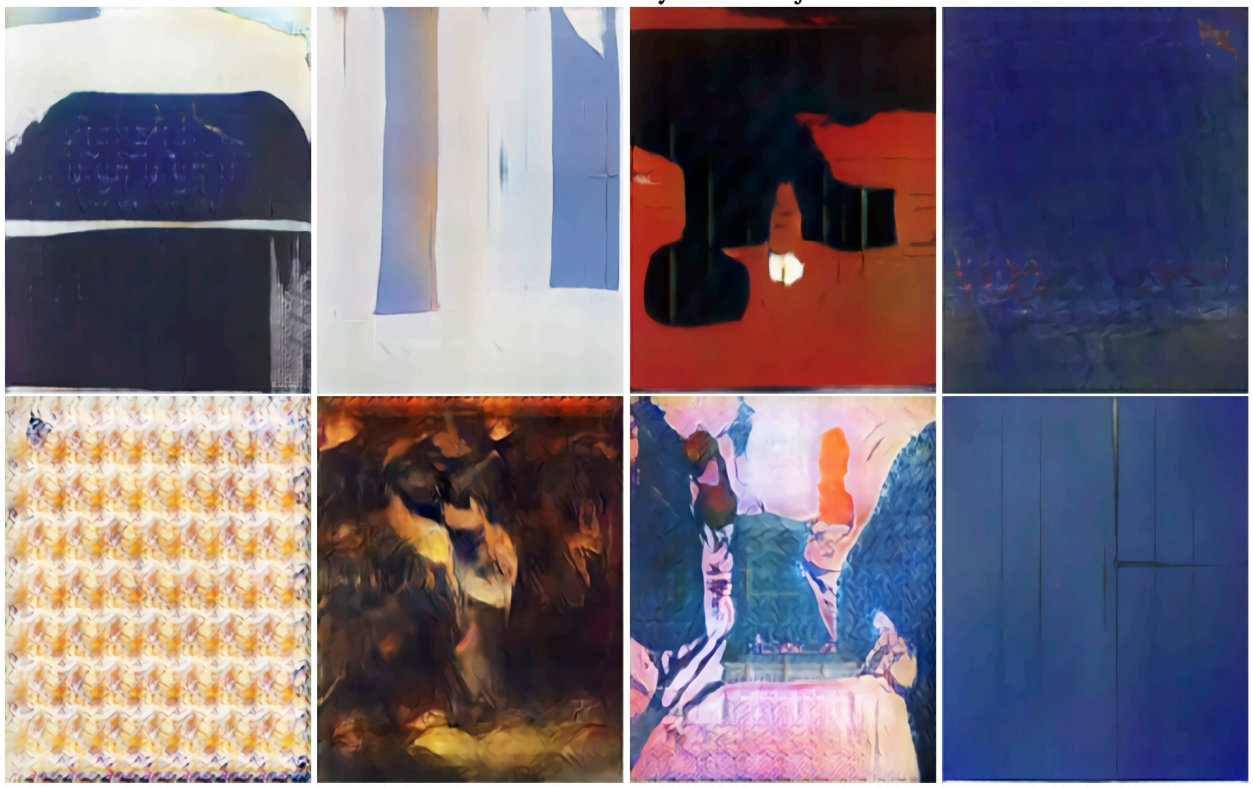
Who created this art? Is it art? Do you like it?

(comparisons to GANs and to human-art)

CAN: Top ranked by human subjects



CAN: Lowest ranked by human subjects



So: what has really changed?

Nothing yet. But let's talk again once AIs has vision and can "experience the world" and select its own "training set"

At some stage, IP has to acknowledge a "new reality" - that certain types of AI can in fact be creative *independently* of their creators...

Thank you for having me!

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3616732

(paper on autonomy and on AI being independent of its creators….)