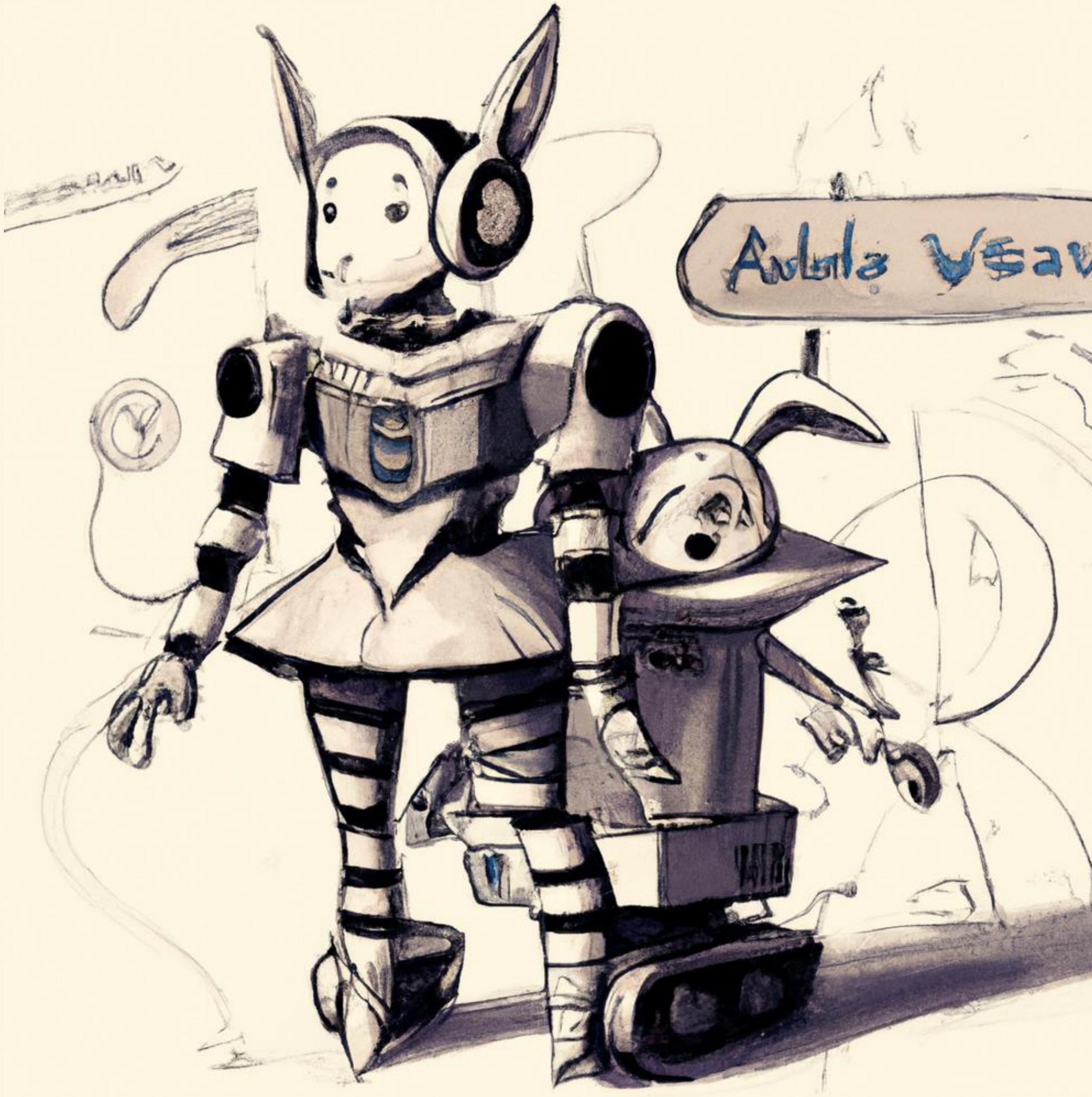
CREATIVITY IN THE GENERATIVE AI ERA

Giorgio Franceschelli
PostDoc @UNIBO
giorgio.franceschelli@unibo.it
giorgiofranceschelli.github.io



From Lovelace and Turing…

All starts with so-called *Lady Lovelace's objection*:

> «The Analytical Engine has **no pretensions to originate anything**. It can do whatever we know how to order it to perform» [Menabrea and Lovelace, 1843]

Then, [Turing, 1950] recast it as «machines cannot take us by **surprise**».

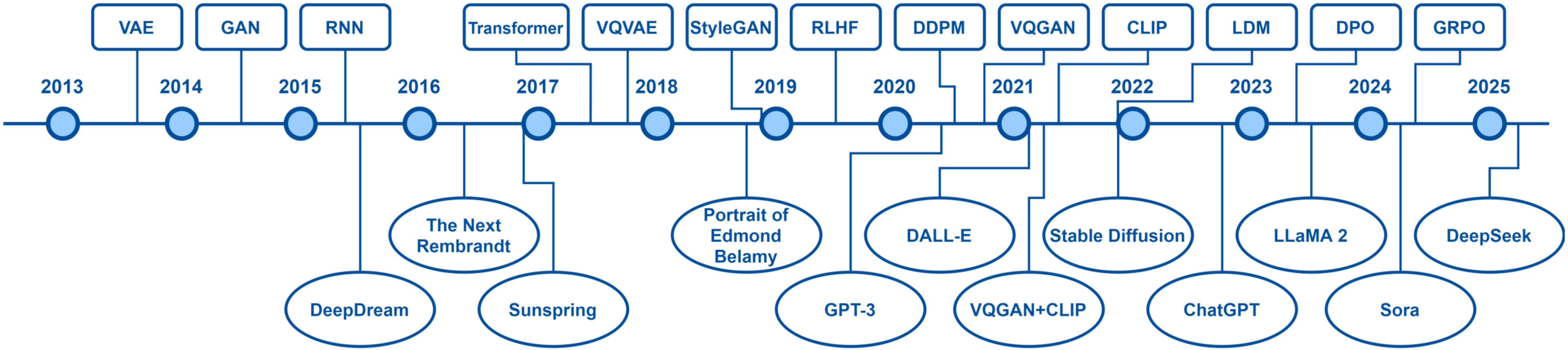
After that, several attempts in late '900 of doing machines to *originate something* by means of coding, rule-based systems, dynamic programming, ecc.

[Menabrea and Lovelace, 1843] L. F. Menabrea and Ada Lovelace. 1843. Sketch of The Analytical Engine Invented by Charles Babbage. In *Scientific Memoirs. Vol. 3*. Richard and John E. Taylor, 666–731

[Turing, 1950] A. M. Turing. 1950. Computing Machinery and Intelligence. *Mind* LIX, 236:433–460

… to ChatGPT and Stable Diffusion

The «big bang» of **generative AI** comes in the new millennium.



Defining Creativity

«Creativity is the ability to come up with ideas or artefacts that are **new**, **surprising** and **valuable**.» [Boden, 2003]

- Value quality + appropriateness
- Novelty for the creator (P-) or for the entire history (H-creativity)
- Surprise unexpected result due to: re-combination of concepts (combinatorial), exploration of space of solutions (exploratory), or transformation of the space itself (transformational creativity)

[Boden, 2003] M. A. Boden. 2003. *The Creative Mind: Myths and Mechanisms*. Routledge, London, UK

Defining Generative AI

«Generative modeling is a branch of machine learning that involves training a model to produce **new** data that is **similar** to a given dataset.» [Foster, 2019]

And what about **surprise**?

[Foster, 2019] D. Foster. 2019. *Generative Deep Learning*. O'Reilly, Sebastopol, CA

Classic Generative Learning Methods

- **Training** procedure: maximize log-probability per-sample (self-supervised learning) or maximize log-probability of in-distribution classification (adversarial learning)
- **Sampling** procedure: execute the learned model on a random (in-distribution) vector and/or on user prompts (that might introduce creativity!)

The (Non-)Problem of Surprise

With current generative models, we can get the simplest forms of surprise, but more accurate the training, less creative the output [Franceschelli and Musolesi, 2021]: remember that the generative models are probabilistic, and will return **a likely output** given the input and what has been learned.

However, even if an output is not unexpected to its producer, it might be to the observers!

An artifact might not be creative *per sé*, but we might still perceive it as creative.

[Franceschelli and Musolesi, 2021] G. Franceschelli and M. Musolesi. 2021. Creativity and Machine Learning: A Survey. arXiv:2104.02726 [cs.LG]

Towards Creativity-Oriented Solutions

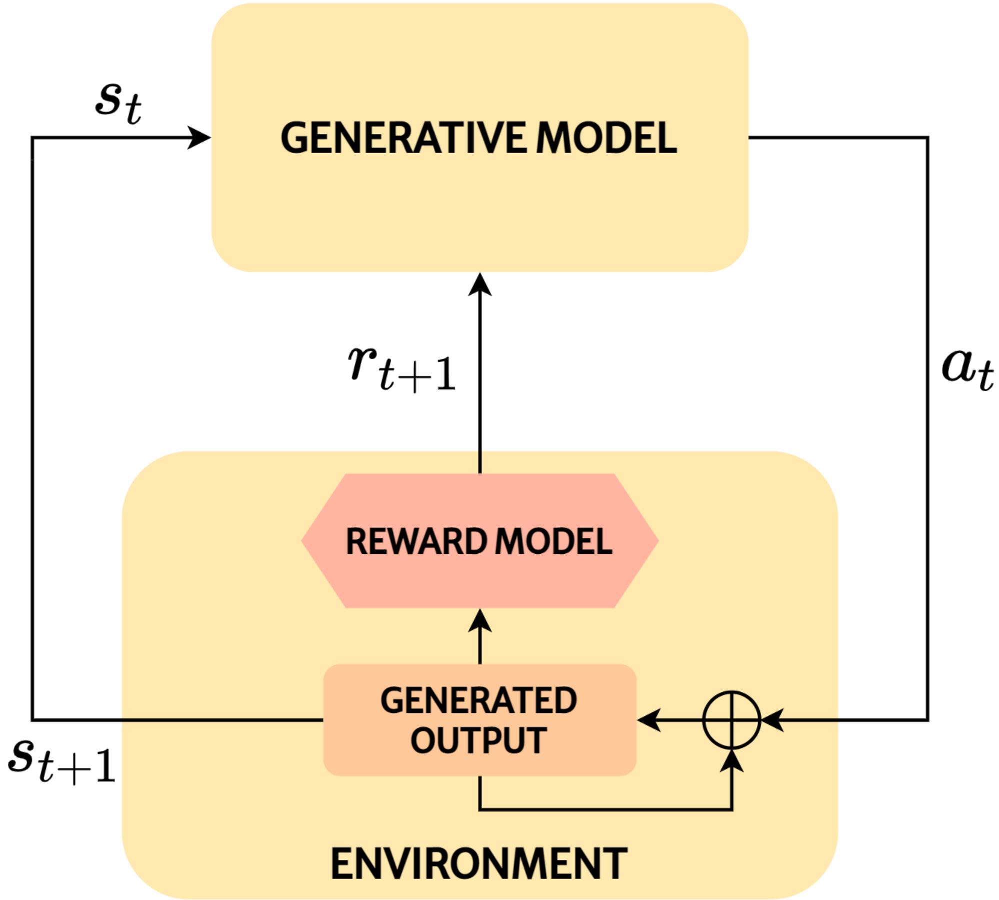
- Creative Adversarial Networks [Elgammal, 2017]: add a «novelty»-like **objective function** to make generator learning a divergent distribution
- Curiosity-based RL [Schmidhuber, 2010]: train the generative model in order to **maximize its curiosity**
- Active divergence [Berns, 2020]: perform **optimization over inputs** at sampling time in order to maximize divergence

[Berns, 2020] S. Berns and S. Colton. 2020. Bridging Generative Deep Learning and Computational Creativity. In *Proc. of the 11th International Conference on Computational Creativity (ICCC'20)*

[Elgammal, 2017] A. Elgammal, B. Liu, M. Elhoseiny, and M. Mazzone. 2017. CAN: Creative Adversarial Networks, Generating "Art" by Learning About Styles and Deviating from Style Norms. In *Proc. of the 8th International Conference on Computational Creativity (ICCC'17)*

[Schmidhuber, 2010] J. Schmidhuber. 2010. Formal Theory of Creativity, Fun, and Intrinsic Motivation (1990–2010). *IEEE Transactions on Autonomous Mental Development* 2, 3 (2010), 230–247

RL for Generative AI



[Franceschelli and Musolesi, 2024] G. Franceschelli and M. Musolesi. 2024. Reinforcement Learning for Generative AI: State of the Art, Opportunities and Open Research Challenges. *Journal of Artificial Intelligence Research*, 79:417-446.

RL for Mere Generation

RL helps learn generative models when the classic generative learning would not be possible:

- GANs for sequences (discriminator signal not differentiable for each single token of the generated sequence);
- Domains that cannot be defined in terms of differentiable losses (e.g., stroke paintings).

Even if RL is used as a «trick» for classic generative learning, it can provide additional advantages (e.g. hierarchical RL, intrinsic motivation -> curiosity and exploration, ecc).

RL for Objective Maximization

In addition to classic generative modeling, RL can help optimize for quantifiable objectives by just considering as rewards: test-time metrics, domain-specific target properties, ecc.

This moves generative AI from learning to produce a good example of a given domain to learning to produce the best possible example according to a given numerical objective.

In other words, if we have a way to quantify creativity, we can use that measure as our reward in an RL framework!

RL for Quality Optimization

Apart from quantifiable properties, RL can also help optimize for non-quantifiable properties (like helpfulness, fairness, or… Creativity!):

- Inverse RL
- RLHF
- RLAIF

This means we can use RL to train/tune a model to produce outputs that someone/something finds as creative.

Inducing Creativity During Sampling

Instead of training/fine-tuning a model to seek for more creative solutions, it is possible to increase the value/novelty/surprise of generated outputs at sampling time:

- We can increase the probability of unlikely tokens (potentially after excluding incorrect ones);
- We can ask the model to choose the most creative solution among several generated ones;
- We can rely on quantifiable creativity metrics and perform a best-of-N sampling.

Is This Enough?

We have seen that RL (and other techniques as well) can be used to make the generation diverge or more likely to be considered as creative.

But…

Can we say the **producer** has been creative because their **product** is creative?

Four P’s of Creativity

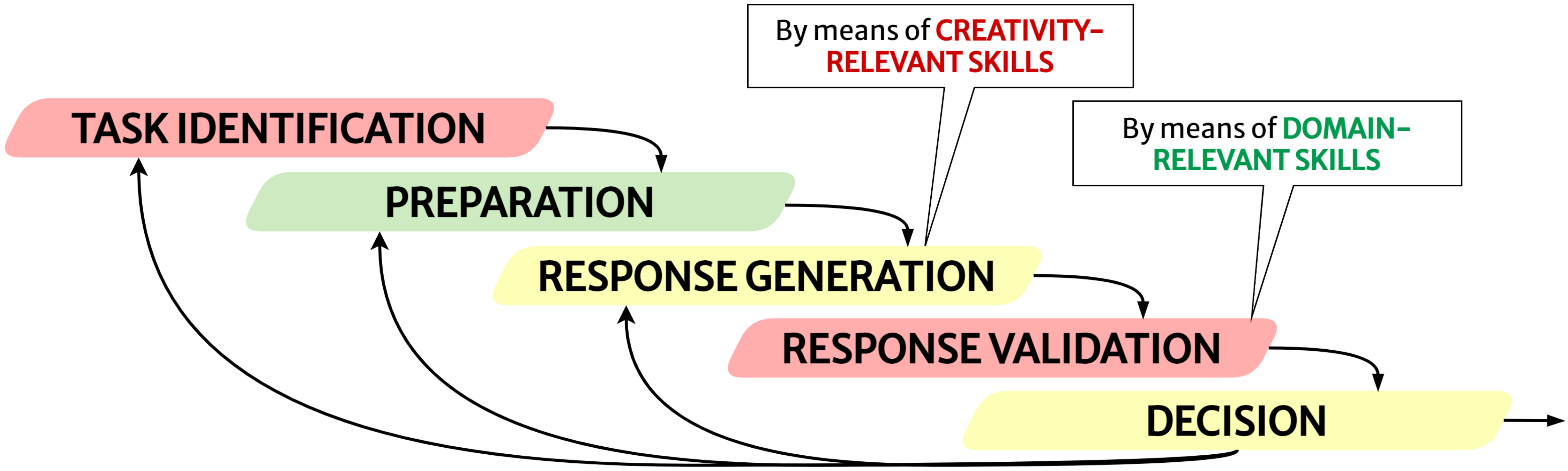
Product is only one of different possible facets of creativity.

It is now broadly accepted that there are four P’s [Rhodes, 1961] defining creativity:

- **Product**
- **Process**
- **Press**
- **Person**

[Rhodes, 1961] M. Rhodes. 1961. An analysis of creativity. *The Phi Delta Kappan*, 42(7):305–310

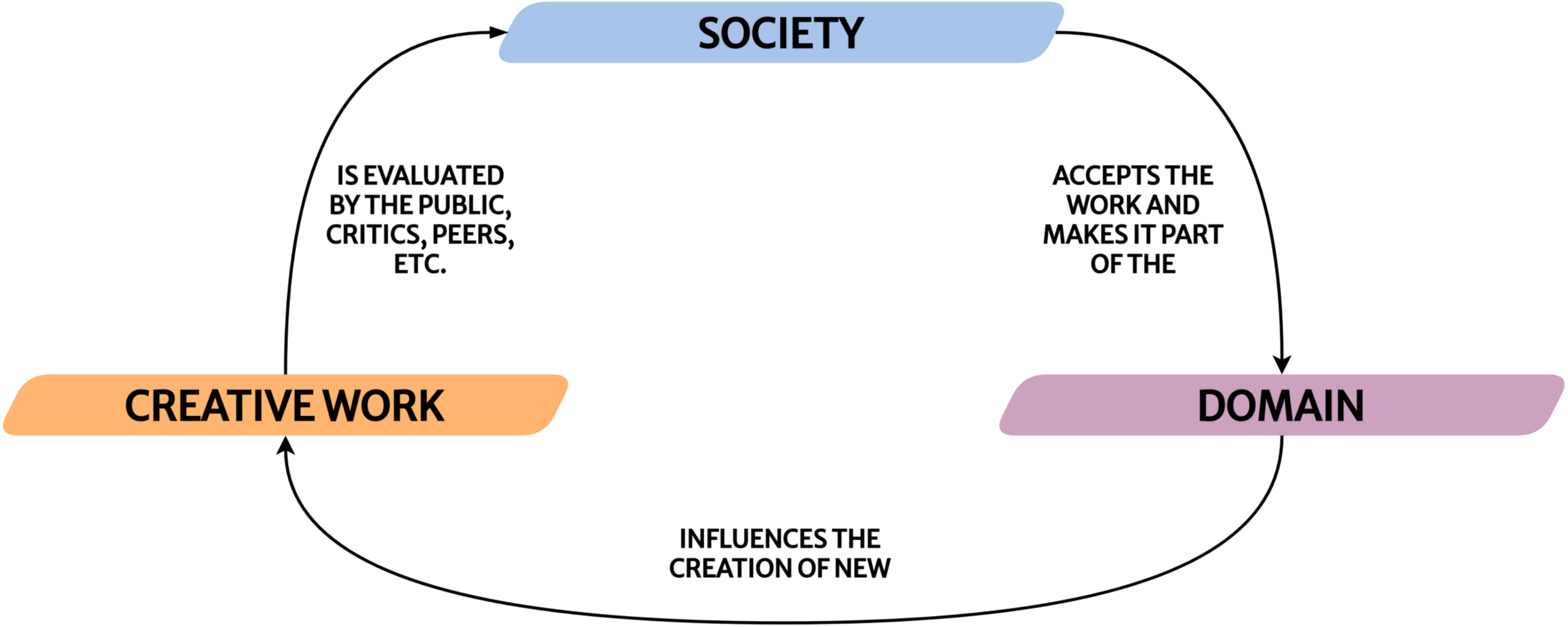
Creative Process



[Amabile, 1983] T. M. Amabile. 1983. The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology*, 45(2):357–376

[Franceschelli and Musolesi, 2024] G. Franceschelli and M. Musolesi. 2024. Creative Beam Search: LLM-as-a-Judge For Improving Response Generation. *Proc. of the 15th International Conference on Computational Creativity (ICCC'24)*

Creative Press



[Csikszentmihalyi, 1988] M. Csikszentmihalyi. 1988. Society, culture, and person: A systems view of creativity. In *The nature of creativity: Contemporary psychological perspectives*, 325–339. Cambridge University Press.

Creative Person

Creativity should be attributed to persons that act intentionally and put (some of) their personality into the product.

AI in general is by no means close to reach the consciousness and self-awareness it would require!

[Gaut, 2003] B. Gaut. 2003. Creativity and imagination. *The Creation of Art: New Essays in Philosophical Aesthetics*, 148–173. Cambridge University Press.

Easy and Hard Problems in Creativity

The previously mentioned limitations are **easy** problems, i.e., they can be solved by correcting the underlying training and optimization processes.

The person perspective requires instead to consider a series of aspects related with consciousness and self-awareness: they are **hard** problems [Franceschelli and Musolesi, 2023] …

… Or, back to [Turing, 1950] and this time *The Argument from Consciousness*, a machine should «not only write it but know that it had written it».

[Franceschelli and Musolesi, 2023] G. Franceschelli and M. Musolesi. 2023. On the Creativity of Large Language Models. arXiv:2304.00008 [cs.AI]

[Turing, 1950] A. M. Turing. 1950. Computing Machinery and Intelligence. *Mind* LIX, 236:433–460

(No) Creativity and Machines

To sum up: generative models can appear to be creative and can simulate many aspects of creativity, but they are not truly creative!

Still… Does it really matter?

Yes, It Matters!

Knowing (and studying) the limits of generative AI is crucial to deal with the ethical, legal, and practical issues it raises: for users, it is relevant to know its limitations and capabilities so as to make the most of it; for researchers, it is relevant to understand where there is still room for improvement and what problems we need to deal with.

Opportunities of Generative AI for Creativity

- Certain parts of tasks can be delegated to AI, freeing authors and workers to spend more time validating news or thinking;
- The same output can be adapted for different audiences;
- Authors can co-create with AI at different stages (brainstorming ideas; role-playing characters; making (more) interactive fictions).

No, It Does Not Matter…

Usually, we don’t bother questioning whether what we see/read/hear is creative, or whether its producer has been creative; we are only interested in whether it works, it is useful, beautiful, or if we just like it.

Therefore, knowing that generative AI is not truly creative will not impact how much it is used!

Still, this can help us understand which risks we might have to face.

Risks of Generative AI for Creativity

- Since the cost for getting an output is minimal:
  - Certain workers might be replaced (especially when timeliness is more valuable than accuracy), and
  - Certain artists might be threatened (especially when cost is more valuable than quality);
- Ideas or styles from human authors might be *stolen* and reused for free, potentially risking a semantic saturation that deprive original works of meaning;
- Biases and prejudices can be (unintentionally) propagated, and people can be easily manipulated thanks to the quality of such outputs;
- Human and AI products might be indistinguishable, causing convergence and conformism (and obscuring minorities);
- Big companies can manipulate scientific research and policymaking thanks to the increasing request of better (and larger) models.

From Ethics to Law

The opportunities and risks we have seen are about what we should or should not do, but does not answer the question of what we can or must not do!

The legal aspects of generative AI are equally important (and perhaps more complex).

Legal Challenges

The most important issues regard Copyright:

- Can we use protected artworks to train a generative model?
- What if the generative model produces an output that is similar to an existing work?
- Who is the copyright owner of an AI-generated work?

But these are not exhaustive – think about privacy and data protection in training data, legal responsibility, ecc.

Using Protected Works for Training

Using a protected work during training requires to make copies of that work (to be moved on GPU cluster, for example), and this can violate the reproduction right of that work.

In the European Union, it is now lawful to use a protected work for AI purposes by a research institute and for non-economic purposes; and it is also lawful to do the same by a company (or anyway for economic purposes) if the rightsholders have not explicitly reserved such right in an *appropriate way* (so called **opt-out right**).

In the US there is no specific rule, and the companies are claiming it is lawful under **fair use doctrine** (which merely specify the principles we need to consider when assessing if a use is lawful) – we need to wait for interpretations...

Discovering Protected Works for Training

An interesting research path nowadays is how to discover whether a specific work has been used during training.

Several techniques exist, but, as the generative model itself, they are only probabilistic…

Anyway, the very recent AI Act requires the deployers of general-purpose generative AI to make it publicly available a *detailed summary* of the content used during training (but who knows which information will contain a detailed summary…)

Output Plagiarism

Since the generative model has been trained on existing works, it is perfectly logic to assume that it can also reproduce those very same works.

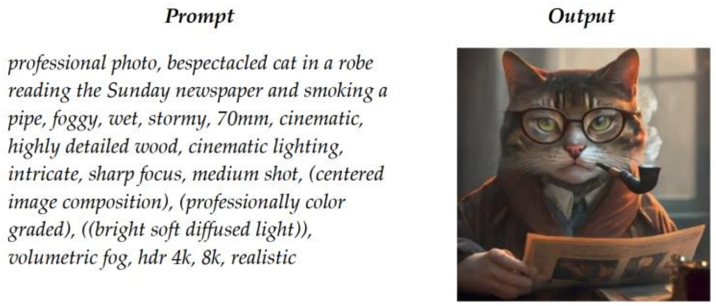
When the reproduced portion is **substantial** (i.e., it clearly characterize the original work), this infringes the reproduction right.

However, only the original content in that expressive form is protected: ideas and styles are not protected, thus an output *in the style of X* is totally lawful (even if arguably unethical).

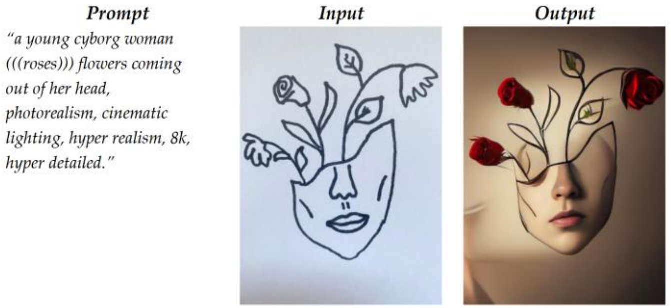
Protection of AI-Generated Works

Copyright laws aim to protect the *human* creativity behind the originality of an artwork: if it is possible to identify attribute a substantial role to a human, then it can be considered its author; otherwise there cannot be any protection and the work falls into the public domain.

The US Copyright Office has recently released some guidelines in this sense:



Public domain



Protected

Protected

https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-2-Copyrightability-Report.pdf