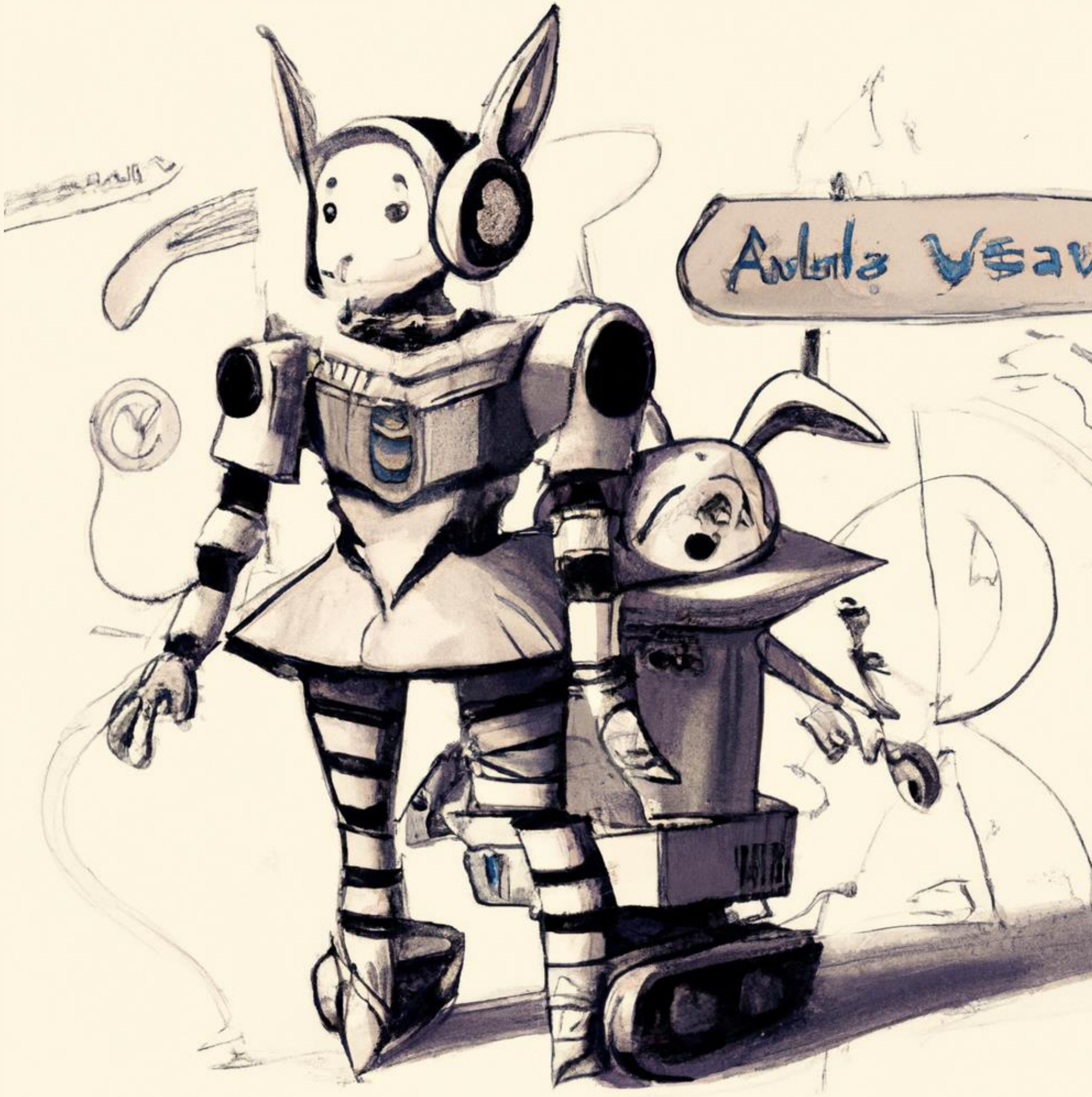
CREATIVITY AND MACHINES

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



From Lovelace and Turing…

All starts with so-called *Lady Lovelace’s objection* [Turing, 1950]:

> «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]

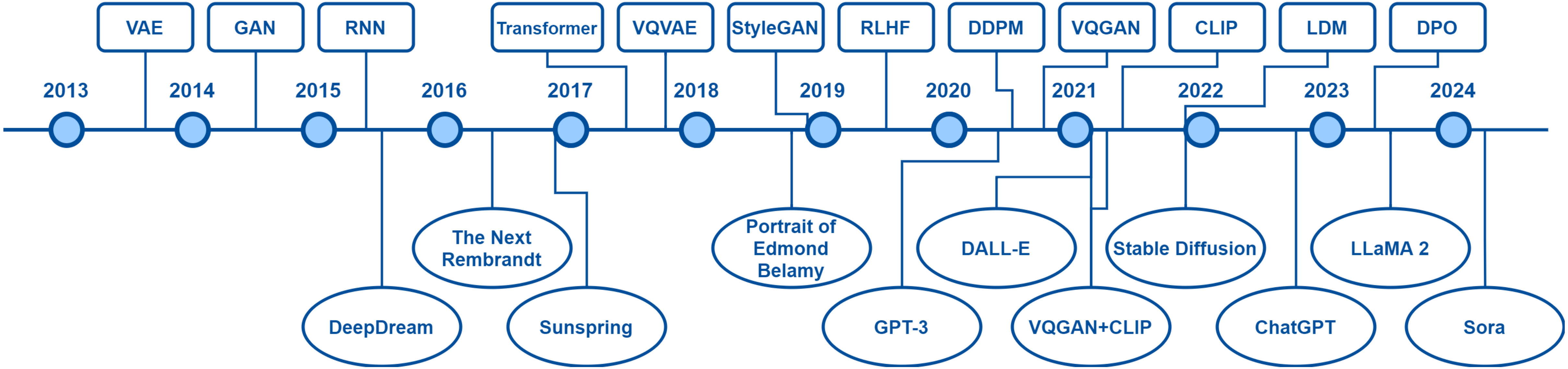
After that, several attempts in late ‘900 of doing machines to *originate something* by means of coding, rule-based systems, dynamic programming, ecc.

[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!)

We can get the simplest forms of surprise (i.e., combinatorial or exploratory creativity), but more accurate the training, less creative the output [Franceschelli and Musolesi, 2021].

[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

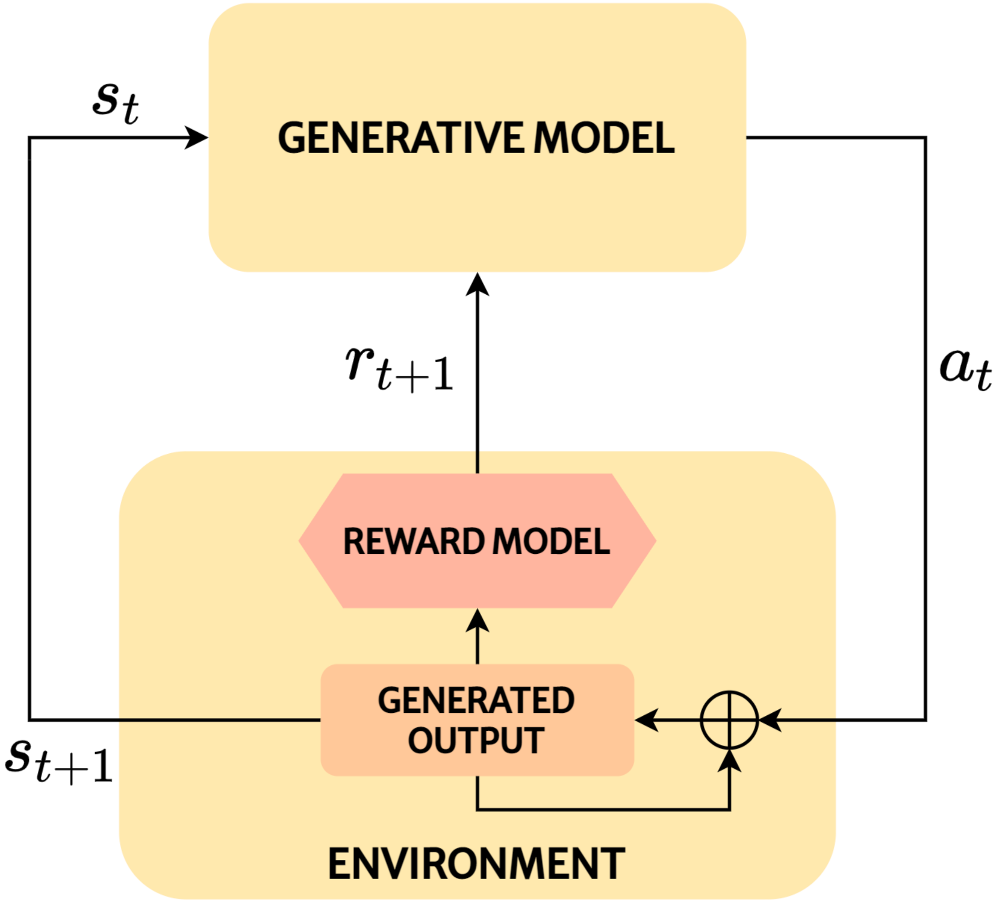
- Creative Adversarial Networks [Elgammal, 2017]: add a «novelty»-like **objective function** to make generator learning a divergent distribution
- Active divergence [Berns, 2020]: perform **optimization over inputs** at sampling time in order to maximize divergence
- Curiosity-based RL [Schmidhuber,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

[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, 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.

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

Of course, using RL for generative AI introduces several of the classic RL problems into generative modeling, and hybrid solutions are now emerging (e.g., Direct Preference Optimization).

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

Creativity should involve a process with:

- Task identification (from external or internal stimuli);
- Preparation (gathering facts and knowledge about the task);
- Response generation (incubation + illumination thanks to creativity-relevant skills);
- Validation (thanks to domain-relevant skills).

Generative AI lacks internal stimuli, usually creativity-relevant skills to guide the generation, and for sure the entire final step!

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

Creative Press

Individuals and their works cannot be isolated from the society: products must be accepted as creative by the society, and previously accepted works should influence producers.

Generative AI is influenced by training data, but is not influenced by new and subsequent products! It considers only a photography of the society at a given moment, not the actual environment around it.

[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?

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).

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;*
- Biases and prejudices can be (unintentionally) propagated;
- Human and AI products might be indistinguishable.

Case Study: Creative Beam Search - 1

Creative Beam Search is a small project to address some of the (easy) problems seen so far:

- Define a sampling procedure that is more creativity-oriented, and
- Simulate (certain parts of) the creative process.

Case Study: Creative Beam Search - 2

Response generation -> Diverse Beam Search (i.e. beam search where candidates are evaluated not only on likelihood under the model, but also on a diversity score that penalizes tokens used in other candidates).

Response validation -> LLM-as-a-Judge (i.e. we ask the model which of the generated solutions is the best one).

Case Study: Creative Beam Search - 3

How to evaluate an AI method for creativity is (another) open issue.

No metrics are available and no benchmarks have been defined yet…

So qualitative assessment is the only solution!

We will ask you to choose between Creative Beam Search output and standard output which is the most creative one.

https://21c1a583af112c3909.gradio.live/   https://d34306ecd95456ec86.gradio.live/

USERNAME: llamattina

PASSWORD: obinu4202