

*Conformed Thought and Art in the context of algorithmic logic and AI
procedures: first results*

Silvia Laurentiz¹

Resumo

Este artigo apresenta os resultados da pesquisa Pensamento Conformado e Arte no Contexto da Lógica Algorítmica e Procedimentos de IA (FAPESP nº 2023/05500-3). A investigação é motivada pela emergência de uma imagem inesperada gerada por um sistema de inteligência artificial — uma ocorrência frequentemente referida como “alucinação de máquina”. Esse fenômeno se torna um ponto de partida para investigar as estruturas lógicas presentes nos modelos conceituais e perceptivos que informam os processos criativos em contextos computacionais. A pesquisa também examina como as práticas artísticas contribuem para reflexões críticas sobre o que definimos como Pensamento Conformado, desenvolvido em publicações anteriores. O projeto propõe um modelo representacional orientado pelo processamento da informação, capaz de interrogar as dimensões epistemológicas e operacionais dos algoritmos computacionais, com especial atenção à inteligência artificial. A investigação aponta uma preocupação central: o uso indiscriminado de conteúdos gerados por IA em conjuntos de dados de treinamento pode gerar uma degradação irreversível da integridade da produção cultural.

Palavras-chave: Arte e Inteligência Artificial, Processos Algorítmicos, Modelos Generativos.

Abstract

This article presents the outcomes of the research project Conformed Thought and Art in the Context of Algorithmic Logic and AI Procedures (FAPESP No. 2023/05500-3). The study is motivated by the emergence of an unexpected image generated by an AI system—an instance often referred to as a “machine hallucination.” This phenomenon becomes a point of entry to investigate the logical structures embedded in the conceptual and perceptual models that inform creative processes within computational contexts. The research further examines how artistic practices contribute to critical reflections on what we define as Conformed Thought, a conceptual framework developed in earlier publications. The project proposes a representational model guided by information processing, capable of interrogating the epistemological and operational dimensions of computational algorithms, with special attention to artificial intelligence. The investigation points to a central concern: the indiscriminate use of AI-generated content in training datasets can generate an irreversible degradation of the integrity of cultural production.

Keywords: Art and Artificial Intelligence, Algorithmic Processes, Generative Models.

¹ Silvia Laurentiz, Professora Associada do Departamento de Artes Plásticas e do Programa de Pós-Graduação em Artes Visuais da ECA/ USP. É Doutora em Comunicação e Semiótica (PUC-SP, 1999) e Livre-docente em Artes (ECA / USP, 2011). Coordena o Grupo de Pesquisa e Extensão Realidades. Atualmente desenvolve o projeto “Pensamento Conformado e a Arte no contexto dos procedimentos de lógica algorítmica e IA” (FAPESP).

Inicialmente é imprescindível dizer que apesar de termos um posicionamento crítico quanto ao uso de IAs, este artigo não abordará este ponto especificamente, pois a proposta aqui é apresentar os primeiros resultados da pesquisa Pensamento Conformado e a Arte no Contexto dos procedimentos de lógica algorítmica e IA (FAPESP No. 2023/05500-3).

Diante da geração, por processo de IA, de uma imagem inesperada, que poderia ser identificada como “alucinação de máquina”, esta pesquisa busca analisar procedimentos lógicos em modelos conceituais e perceptuais envolvidos em processos criativos, refletindo sobre contribuições da arte ao que chamamos de “Pensamentos conformados”. Este termo foi definido em diversas publicações (Laurentiz 2015, 2018, 2019, 2021) como: códigos e conjunto de códigos, normas, algoritmos, padrões, dispositivos, interfaces, uma vez que são resultados de conceitos, textos e teorias científicas e, portanto, formas atualizadas de conhecimentos elaborados, com capacidade de mudar hábitos e comportamentos. Esta pesquisa está ainda em andamento e pode ser descrita a partir dos seguintes argumentos: em nosso cotidiano, exercitamos e praticamos extensivamente o raciocínio lógico-abstrato-simbólico, que chamamos de 'pensamentos conformados', autocontrolados e deliberados, fruto de conceitos, modelos de conhecimento de uma cultura ou grupo, e isso significa que estamos nos adaptando e mudando hábitos, comportamentos, crenças e ações a partir destes padrões. Por consequência, isto causará mudanças em nossa maneira de pensar. O principal interesse é apresentar um modelo representacional promovido pelo processamento de informações que seja capaz de apontar questões sobre algoritmos computacionais em particular, de Inteligência Artificial. Em breve resumo, estivemos estudando modelos de treinamento de máquina em IA generativa e inicialmente, para entendermos o processo, propusemos didaticamente a simples tarefa de se gerar imagens de maçãs.

A ideia, desde o início, não foi utilizar prompts ou quaisquer rótulos que indicassem que as imagens utilizadas de um banco público de imagens eram de maçãs. A proposta foi trabalhar com modelo *stable diffusion*, onde se adiciona progressivamente ruído a uma imagem até que tudo que reste seja um ruído aleatório. Sabendo identificar o ruído adicionado, pela sua subtração posterior, o

sistema acabará identificando padrões nas imagens; e finalizado a fase de treinamento, será capaz de gerar imagens a partir do padrão apreendido.

Curiosamente, às vezes, aconteciam imagens resultantes inesperadas que não poderiam ser identificadas imediatamente, de forma precisa, como sendo uma imagem de maçã, apesar da base de dados ser de apenas imagens de maçãs. E neste processo, percebemos prejuízos causados pelo uso indiscriminado de conteúdo gerado por aprendizado de máquina, chegando a alguns danos irreversíveis nos modelos resultantes, onde os rastros do conteúdo original chegam até a desaparecer. Os autores Ilia Shumailov et al (2024) referem-se a esse efeito como 'colapso do modelo'. Portanto, a pesquisa, neste momento, encontra-se preocupada com estes desvios causados por treinamento a partir de dados em grande escala, que afetarão de forma significativa nosso modelo representacional.

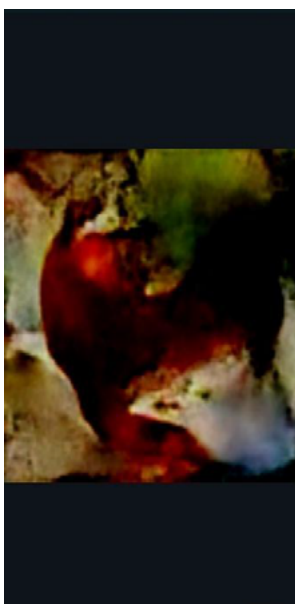


figura 1. Imagem gerada por IA em procedimento preliminar ao modelo representacional proposto na pesquisa. Fonte: da autora.

Observando a figura 1, imediatamente podemos perceber que se trata de uma imagem abstrata, com algumas formas volumosas, cores de um vermelho intenso passando por verde, tons terrosos, branco.... Há uma exuberância da cor vermelha, que carrega uma carga sensória na imagem; há também uma oscilação com o tom

verde, sua complementar. E a princípio, identificamos um fundo onde a forma mais volumosa centralizada na imagem parecer estar sustentada. Não temos outra informação que possa confirmar se esta imagem foi ou não gerada por IA (embora tenha sido), e nem identificamos um rótulo que possa ter sido utilizado.

Entretanto, uma coisa que não se esperava de uma IA, quando iniciamos esta pesquisa, era que fosse capaz de interpretar informações abstratas como a dessa imagem. Subimos esta imagem no ChatGPT – 4.0 (OpenAI, 2023) e pedimos que a descrevesse. A resposta surpreendeu desde o início, quando foi definida como "arte abstrata". Não uma imagem abstrata, mas como arte. Após uma descrição formal da imagem, acrescentou: *"Esse tipo de arte abstrata pode ser interpretado de várias maneiras, dependendo da perspectiva do observador. As formas e cores podem evocar emoções diferentes ou lembranças pessoais. No geral, a imagem transmite uma sensação de intensidade e dinamismo"*. Continuamos a questionar o ChatGPT sobre aspectos representacionais, por exemplo, como se poderiam ser identificados objetos e figuras na imagem. Ao final perguntamos se seria reconhecida alguma fruta na imagem, e recebemos nova surpresa com a resposta: *"[...]poderia ser vista como algo semelhante a uma maçã [...]"*.

Diante dessa condição preliminar, o projeto se inicia com o reconhecimento de unidades de significação em um modelo representacional, a partir de coisas, objetos e modelos (Laurentiz, 2019). Neste modelo proposto, em um processo de representação temos 'coisas em si', 'objetos', que serão coisas objetivadas (coisas em si só entram num sistema comunicativo estando objetivadas), e 'modelos' destes objetos.

Desta forma, pelo sistema proposto, estas serão unidades mínimas de representação. Entretanto, não se trata de simples sistema linear e nem tampouco de um sistema determinista fechado. Temos influências e novos inputs, interferências e informações inseridas pelo contexto, distúrbios e alterações provocados pelo caminho. Mas, uma coisa é objetivada para participar de um sistema comunicacional, que se transforma a partir de metas e objetivos (e neste momento inicial, as metas e objetivos são dados por agentes externos). Depois, o resultado é um conceito deste objeto transformado em modelo que retornará ao sistema novamente por retroalimentação (figura 2).

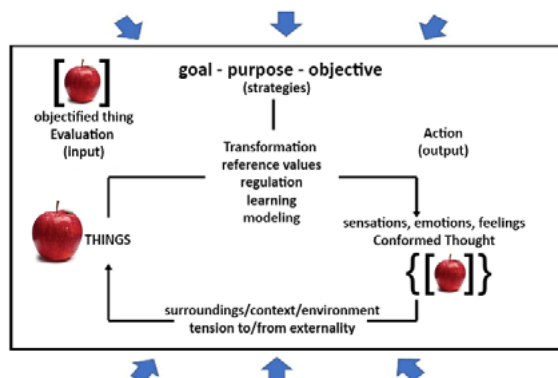


figura 2. Sistema básico de representação a partir das unidades mínimas: coisas, objetos e modelos, reconhecidos como **pensamentos conformados** no diagrama. Fonte: da autora.

Seguimos por retroalimentação, a partir do modelo cibernético, onde o processo passará para a próxima fase, quando coisa objetivada, se transformará por um propósito externo, gerando novos resultados, assim infinitamente.... Importante ressaltar que, quando dizemos que o propósito é ordenado por agentes externos ao sistema, é diferente de quando o sistema em si tem a autonomia de mudar suas próprias metas. De qualquer forma, trata-se de um sistema complexo, o que significa que pode ser capaz de gerar padrões emergentes, e por entropia gerar comportamentos imprevisíveis, desorganização, ou sofrer interferências técnicas ou humanas, introduzir aleatoriedade, sobrecarga de informação, entre outros.

Segundo Flusser (2007), a imaginação é imprescindível para nossas ações e a compreensão do mundo. Entretanto, ao mesmo tempo, a imaginação é nossa capacidade de distanciamento dos objetos do mundo. Este movimento de afastamento e recuo dos objetos, podemos chamar de abstração.

Tem um exemplo em seu livro *O mundo codificado* (Flusser, 2007) que trata do gesto inaugural da criação de imagens, quando em cavernas pré-históricas encontramos a figuração de um cavalo. Flusser diz: "*os primeiros configuradores de imagens tiveram que se afastar de um cavalo, olhar para ele e depois fixar esta visão fugidia na parede da caverna, exatamente para que outros pudessem reconhecê-la*" (idem, p. 162). A questão fundamental, continua Flusser, é: para onde

se vai quando se afasta do cavalo? Então temos a imagem de uma coisa, depois a explicação dessa imagem da coisa, e daí uma imagem dessa explicação, que resultará em conceitos e finalmente no modelo da coisa. Acredito que a teoria de Flusser já tenha sido muito divulgada, desnecessário prolongar neste ponto.

Nosso problema é que com os procedimentos algoritmos, confirmando a teoria flusseana, o pensamento imagético está se tornando capaz de reelaborar conceitos, transformar conceitos em objetos, e tornar-se um metapensamento de um modo de pensar conceitual (Flusser, 2007, p. 118). Ele diz ainda que o pensamento imagético passa a **pensar conceitos**, na forma de **modelos** (idem, idem). Mais categórico, continua na página seguinte, que o pensamento imagético toma o lugar da filosofia, enquanto **metaprocessamento de conceitos**.

E com a IA – onde percebemos uma nova dimensão da escalada de abstração de Flusser, a máquina não só classifica e analisa informações, como também cria fenômenos a partir de síntese de inúmeros modelos. Por conseguinte, continua o processo, a ponto da IA mesma estabelecer novas metas, objetivos e hábitos, e neste caso, os agentes transformadores são internos ao sistema, ou seja, a máquina consegue estabelecer novas metas internamente.

Desnecessário dizer as consequências que isso possa vir a causar na produção cultural. E ainda, a IA deixa de ser apenas linguagem, e passa a ser um agente capaz de realizar mediações tecnológicas, influenciador na produção cultural, promovedor de interações sociais, gerando novos paradigmas culturais.

A partir da escalada abstrativa de representação nos diagramas para um modelo representacional, passamos a próxima etapa, onde a pesquisa se encontra atualmente, que é a de criar uma máquina cibernética dinâmica, baseada em modelo de aprendizagem, que sintetizará dados para gerar objetos representacionais num metaprocessamento de/para a geração de "pensamentos conformados". Percebe-se já neste momento que existe uma sutil diferença entre: 1. se ter a imagem de algo que está ausente - e sobre isto muito já foi dito (e relembro o princípio de fantasmagoria); de 2. se ter um processo de modelagem que metaprocessa e gera seus próprios modelos objetivados. Em outras palavras, não estamos apenas dizendo

que temos uma simulação da imaginação, e trazendo a reboque a tão discutida questão de se simulação é ou não representação; nem citando a proposta de o digital levar para uma condição zero-dimensional informacional com possibilidades de recombinações infinitas e superposições sígnica; mas, **além disso**, o que estamos acrescentando neste processo, é a **quase-autonomia da máquina e tomada de decisões**.

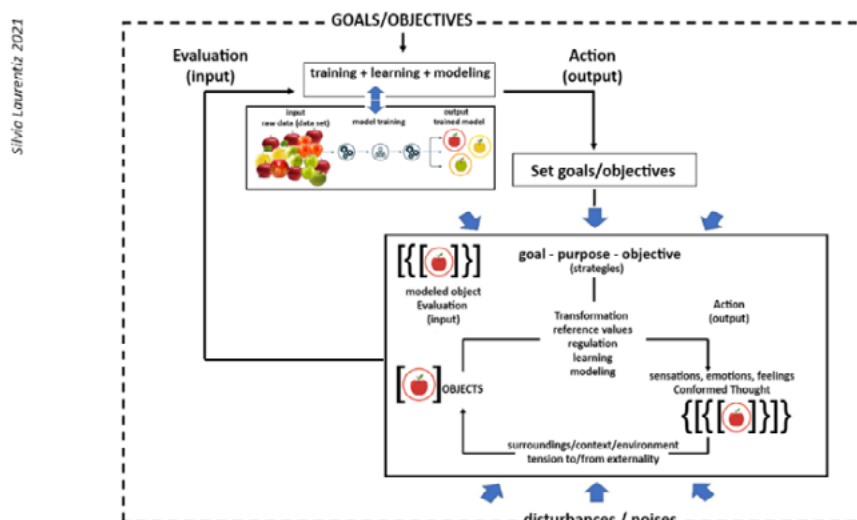


figura 3 – diagrama ampliado para modelo representacional a partir do uso de IAs. Fonte: da autora.

A partir da figura 3 podemos ver o modelo de aprendizagem de máquina, que chamamos de **quase-processo metacognitivo** (Laurentiz, Projeto Pensamento Conformado e a Arte no contexto dos procedimentos de lógica algorítmica e IA, 2023, p. 10).

Assim, na escalada da abstração de Flusser, passamos por recuos e afastamentos importantes ao que chegamos na síntese de uma máquina treinada por conceitos que retroalimentarão o sistema num metaprocessamento, ou metacognição.

Mas, vale um regressão, antes desta fase atual, na tentativa de aproximação aos processos de IAs, foram geradas imagens de maçãs a partir de *prompts* nos principais geradores de IA atuais, sempre com a mesma solicitação: **Gere a imagem de uma maçã!**

Podemos lembrar Baudrillard (1991) neste momento e seu conceito de Hiperreal, onde a imagem técnica torna-se mais real do que o real, ou seja uma representação idealizada do real que passa a nos causar efeitos reais. Uma realidade produzida por modelos, signos e imagens que não têm mais referência direta com o mundo real, mas que parecem mais reais do que o real.

E o que acontece quando a hiperrealidade borra a fronteira entre o real e o ficcional? Quando o hiperreal torna-se um epifenômeno, ou seja, passa a fazer parte de uma *imageria* mental que causa um efeito além do efeito sob sua consideração primária. Em outras palavras, quando fenômenos paralelos e secundários tomam a frente ao fenômeno principal e restabelecem os parâmetros da realidade?

Desta forma, já podemos deduzir que as IAs produzem e reproduzem incessantemente imagens e representações que substituem o contato direto com a realidade, restabelecendo novos parâmetros para o real.

E, enquanto hiperreal, considerado um epifenômeno porque ele não é o motor primário das representações, mas uma manifestação superficial - chamemos assim, da realidade, - as IAs passam a ultrapassar esta barreira no momento que credibilizam suas imagens não como manifestação superficial, mas, numa inversão do processo representacional, o fenômeno real passa a ser secundário e superficial, uma vez que a distância que nos colocam dos objetos nos confundem entre real e ficcional.

Assim, a primeira fase de gerar imagens por diferentes geradores disponíveis atualmente, sempre partiu de modelos já treinados, exatamente para perceber esta dinâmica, do borramento de fronteiras. Foram utilizados os modelos do DALL-E, Flux.ai, Krea.ia, ChatGPT, neste momento, e já podemos perceber sutis diferenças entre os diferentes modelos, uns mais realistas, outros mais ilustrativos, ou com alguns desvios, o que evidencia ainda mais as proporções e efeitos na produção cultural atual.

Além deste processo preliminar, utilizando *prompts* em diferentes modelos treinados, em nossa pesquisa estivemos treinando nosso próprio modelo, a partir de

um banco de dados público de maçãs, como já supracitado. Repetindo, neste caso, não estivemos usando *prompts* e nem rótulos. Trata-se de um sistema de treinamento semi-supervisionado, onde estivemos trabalhando com modelo do tipo *stable diffusion*. E usamos um modelo U-Net, que são redes neurais convolucionais. De forma rápida, estima-se a quantidade de ruído no espaço latente e a subtrai da imagem de treinamento. Repete-se esse processo um número de vezes, reduzindo o ruído de acordo com as etapas especificadas pelo programador.

Para estes processos iniciais, contamos com bolsista de Capacitação Técnica (FAPESP No. 2024/13304-2), onde o técnico Murilo Trassi esteve nos primeiros meses do projeto, estruturando o processo de treinamento e geração de imagens.

Neste processo de inserção e subtração de ruído, a rede foi paulatinamente recuperando padrões na imagem, ao mesmo tempo que alteramos parâmetros até conseguir diferentes resultados de treinamento. Não identificamos, por exemplo, as mesmas características dos modelos atuais de mercado, utilizados na fase preliminar, até porque nosso modelo é uma versão simplificada de *machine learning*. A ideia nunca foi competir com grandes conglomerados da tecnologia, e nem chegar aos mesmos resultados, mas demonstrar a fragilidade e riscos destes sistemas. Até porque esta é uma pesquisa teórica, que se apoia em experimentos práticos, e não o inverso.

Depois deste processo de treinamento, passamos para um processo de discriminação daquelas imagens geradas que representam melhor o modelo a partir de uma outra rede treinada, agora com a referência de maçãs definida por rótulos. Esta segunda rede identifica dentre aquelas imagens geradas por nosso modelo, a partir do seguinte critério: aquelas com maior nível de confiança (estamos estabelecendo 90%), aquelas com baixa confiança (abaixo de 90), e aquelas não reconhecidas. Evidente que imagens com 90 a 100 % de confiança escolhem maçãs mais realistas enquanto as demais vão se distanciando de um reconhecimento imediato de maçã. Mas lembramos que o discriminador utilizado foi treinado com este perfil, por isso, o resultado tenderá sempre a identificar maçãs "mais realistas".

Portando, de imediato, podemos dizer que a imagem apresentada na figura 1 seria descartada. Entretanto, motivados exatamente por aquela imagem, passamos a contabilizar também aquelas imagens com baixa confiança, tanto quanto aquelas totalmente descartadas, pois nos interessam também perceber um espaço latente das imagens geradas. E, a princípio, imagens que acabam ficando aquém das expectativas e apresentam resultados inesperados são consideradas por alguns autores como 'alucinações de máquinas', e evidente que para questionar um modelo de um sistema representacional isso pode ser interessante no futuro.

Agora, nesta fase, estamos forçando a retroalimentação do sistema, fazendo voltar para a base inicial do banco de imagens aquelas que foram consideradas de alta confiança na primeira geração, depois do modelo já treinado. Com o banco de dados inicial ampliado, novo treinamento é feito. Depois deste treinamento, novas imagens são geradas, nova discriminação, recursivamente, por no mínimo 3 gerações de imagens. Isto condiz com Ilia Shumailov et all (2024), quando demonstram a mesma preocupação, logo no início do artigo, de que se os dados de treinamento da maioria dos modelos futuros também forem extraídos da web, eles inevitavelmente treinarão com base nos dados produzidos por seus antecessores (idem, p.755). Na fase final, pretende-se analisar todo o processo.

O foco principal é provocar retroalimentação no treinamento continuamente. Se de fato, segundo relatório da EUROPOL Innovation Lab de 2022, que previa que 90% do conteúdo online seria gerado por IA até 2026, isto, consequentemente, significa dizer que imagens geradas por IA retroalimentarão todos os demais sistemas e agentes da internet também, e que lotarão nossos feeds. Essas imagens lotando nossos sistemas de acesso e busca, acabarão impactando nossos modelos de representação, proliferando imagens cada vez mais distantes dos seus objetos iniciais, causando sobrecarga de informações disponíveis com qualidade reduzida. Tanto que as pesquisas conduzidas pela Europol mostram que as tendências mais preocupantes estão situadas na evolução e na detecção de *deepfakes* e desinformação de forma em geral. Evidente que todos sabem o que é uma *deepfake* e quais os prejuízos legais, éticos e morais envolvidos neste assunto. Neste projeto, em consonância com Flusser (2007), nos interessam a passagem de uma coisa que foi

objetivada para um modelo, e este modelo depois tornando-se modelo objetivado, que se transforma, subsequentemente, em modelo do modelo e que estes processos carregam níveis de abstrações que se distanciam cada vez mais da “coisa em si”. Esta é a proposta da escalada de abstração de Flusser e em nosso modelo proposto a partir de uma unidade básica de representação sugerimos que o processo de modelagem e aprendizagem causaria um metaprocessamento de/para geração de ‘pensamentos conformados’. Em outras palavras, quando se inicia uma discussão sobre informações falsas (*deepfakes*, falsidade de documentos, uso indevido de imagens, invasão de privacidade, alteração de dados em campanhas políticas, geração de provas criminais falsas...) criadas e distribuídas na rede, facilmente se detecta questões éticas propositalmente participando deste processo. Entretanto, nosso projeto levanta a questão de que mesmo sem se tratar de imagens geradas por algum princípio duvidoso e de má fé, algo acontece na geração de imagens que causará consequências irreversíveis no sistema representacional.

Importante alertar que em 2024 o relatório de 2022 da EUROPOL foi contestado e teve uma versão atualizada que substituiu a versão anterior exatamente no ponto onde declaram a participação futura esperada de 90% do conteúdo gerado sinteticamente por IA. Esta informação na versão atualizada foi removida em 2024. Apesar disso, ainda acreditamos ser uma previsão preocupante. Mesmo que não sejam 90% do conteúdo de IA, mostra-se atualmente que será um número significativo. Concluindo, em cenários em que novos modelos são treinados por dados gerados por outros modelos, isso pode causar uma degeneração acumulada. Com o tempo teremos perda de variedade, fidelidade e capacidade de generalização. Em especial pois estamos tratando de erros, vieses e simplificações que vão se acumulando, e com isso, tal qual Flusser previa, o modelo se afasta cada vez mais das coisas do mundo. E, se antes buscava-se distinguir informações verdadeiras de falsas, quando passamos a tratar de modelos, isto parece não fazer mais sentido.

Evidente que sistemas fechados que se retroalimentam infinitamente tendem a gerar desordem e aleatoriedade por entropia, e este é sim um motivo de preocupação. Mas, em nosso modelo, pretendemos ter portas de entrada aberta para sair dessa armadilha sistêmica, enquanto possibilidade sônica, permitindo que o sistema se adapte e

compense a perda da informação. E esta será a fase final do projeto, que traremos em futuras publicações.

Referências

- Baudrillard, J. (1991). *Simulacros e simulação*. Lisboa: Relógio D'água. pp. 151–158. Original em francês, 1981).
- Europol (2022). Facing reality? Law enforcement and the challenge of deepfakes, an observatory report from the Europol Innovation Lab, Publications Office of the European Union, Luxembourg.
- Flusser, V. (2007). *O mundo codificado*. Ed. Cosac Naify, São Paulo, Brasil.
- Flusser, V. (2012). *O Universo das imagens técnicas*. Coedição: Annablume editora (Brasil) and Imprensa da Universidade de Coimbra (Portugal).
- Laurentiz, S. R. F. (2015). Sensoriality and conformed thought. In: Antona, Margherita, Stephanidis, Constantine. (Org.). *Universal Access in Human-Computer Interaction*. Access to Interaction 9th International Conference, UAHCI 2015, Held as Part of HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015, Proceedings, Part II. 1 ed. New York: Springer International Publishing, 2015, v. 9176, p. 217-225.
- Laurentiz, S. (2019). Conformed thought and the Art of Algorithms. In Proceedings of the 9th International conference on digital and interactive arts - ARTECH, 23-25 October 2019, Braga, Portugal, ISBN/ISSN 9781450372503, p. 465-472. Editora Universidade Católica Portuguesa, Portugal.
- Laurentiz, S. (2021). Arte en el contexto de los procedimientos de lógica algorítmica. In *Arbor*, 197(800), a603. <https://doi.org/10.3989/arbor.2021.800005> (Acesso em: maio. 2025).
- Laurentiz, S. (2022). Conformed thoughts, representational systems, and creative procedures. In: Vear, C., Poltronieri, F. (eds) *The Language of Creative AI*. Springer Series on Cultural Computing. Springer, Cham. https://doi.org/10.1007/978-3-031-10960-7_10.
- Laurentiz, S. (2023). Um entrelaçar de ações criativas entre arte e tecnologia. *Revista VIS: Revista do Programa de Pós-Graduação em Artes Visuais*, [S. l.], v. 21, n. 2, p. 459–467, 2023. Disponível em:

Ilia Shumailov, Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, Ross Anderson & Yarin Gal (2024). AI models collapse when trained on recursively generated data. Nature 631, 755–759. <https://doi.org/10.1038/s41586-024-07566-y>.