

# O DIÁLOGO ENTRE INTELIGÊNCIA ARTIFICIAL E FILOSOFIA: CONTRIBUIÇÕES DE GILBERT SIMONDON

*THE DIALOGUE BETWEEN ARTIFICIAL INTELLIGENCE  
AND PHILOSOPHY: CONTRIBUTIONS OF GILBERT  
SIMONDON*

Itamar Soares Veiga 

Universidade de Caxias do Sul, Brasil  
inpesquisa@yahoo.com.br

---

Recebido: 29 Out 2025

Aceito: 15 Jan 2026

Publicado: 17 Mar 2026

Autor para correspondência:

inpesquisa@yahoo.com.br



inteligências artificiais generativas. Para responder a esta questão, realiza-se um estudo conceitual com foco nos conceitos de “conjunto técnico” e “processo de concretização” de Simondon, para em seguida apontar os fenômenos emergentes e perigos da inteligência artificial. A conclusão aponta para um afastamento do ser humano em relação ao “conjunto técnico”.

**Palavras-chave:** Simondon; inteligência artificial; conjunto técnico; processo de concretização.

## Abstract

This study seeks to highlight some contributions of the philosophy of Gilbert Simondon (1924-1989) to the understanding of specific sectors of artificial intelligence. The research has a focus that is limited to the scope of generative artificial intelligences, which are common in personal assistants, known as *Chatbots*. In this context, the investigation questions whether Simondon's philosophy can assist in understanding the challenges and dangers presented by generative artificial intelligences. To address this question, a conceptual study is conducted focusing on Simondon's concepts of “technical ensemble” and “concretization process,” followed by pointing

## Resumo

este estudo procura mostrar algumas contribuições da filosofia de Gilbert Simondon (1924-1989) para a compreensão de setores específicos da inteligência artificial. A pesquisa possui um recorte que se delimita ao âmbito das inteligências artificiais generativas, as quais são comuns em assistentes pessoais, denominados de *Chatbots*. Neste contexto, a investigação questiona se a filosofia de Simondon pode auxiliar a compreensão dos desafios e perigos apresentados pelas in-

out the emerging phenomena and dangers of artificial intelligence. The conclusion points to a distancing of the human being in relation to the “technical ensemble.”

**Keywords:** Simondon; artificial intelligence; technical ensemble; concretization process.

## Introdução

Este artigo aborda alguns conceitos das reflexões do filósofo Gilbert Simondon. O principal é o de “objeto técnico”. Supomos que este conceito possui determinadas peculiaridades que podem contribuir para ampliar o contexto teórico recente sobre as diferenças e as aproximações entre os seres humanos e os dispositivos artificiais inteligentes. Estes são compreendidos pelos produtos oriundos do desenvolvimento da ciência da computação, notadamente nas últimas duas décadas. Tais produtos são construídos como inteligências artificiais que podem fornecer resultados personalizados e inesperados.

Uma exposição e análise das reflexões do filósofo Gilbert Simondon perante o quadro mais atual das inteligências artificiais pode representar um desafio. De imediato, podemos dizer que este desafio está na constituição de alguns dispositivos de inteligência artificial e no cuidado exigido para desenvolver a articulação conceitual de Simondon. Ainda que existam estes desafios, é preciso que a filosofia contribua mais para a discussão sobre os fenômenos e os perigos originados pelos dispositivos de inteligência artificial. E, diante dessa necessidade, a filosofia de Simondon apresenta um quadro promissor ao apresentar de uma forma peculiar os modos de existência humana dos objetos técnicos.

O objetivo deste artigo é mostrar um dos desenvolvimentos da reflexão de Simondon, apenas por meio da sucessão e dos entrelaçamentos de seus conceitos a partir daquele, já referido, de objeto técnico. E, a seguir, destacarem o conceito de “conjunto técnico” e o utilizaremos como base para pensar sobre as inteligências artificiais denominadas de “generativas”<sup>1</sup>. A escolha das inteligências artificiais, especificadamente, generativas, no universo dos objetos técnicos disponíveis, abre uma discussão possível sobre o seu estatuto ontológico. Sendo assim, a filosofia pode ampliar o espaço de acolhimento da discussão sobre inteligência artificial, evitando paralisar-se em injunções metafísicas ou antropocêntricas.

O artigo procura responder a seguinte pergunta: é possível utilizar as características do conceito de conjunto técnico, de Simondon, para compreender os desafios e perigos das inteligências artificiais generativas? No que diz respeito a este tipo de inteligência artificial, vamos nos deter nos avanços relacionados tanto aos *Chatbots* quanto aos modelos inteligentes genéricos. Em relação a Gilbert Simondon, vamos tratar dos seus conceitos com suas respectivas articulações a partir do seu livro: “*Do*

---

<sup>1</sup>Existem várias inteligências artificiais disponíveis algumas são classificadas como “generativas”, isto é geram um conteúdo, outras, menos recentes não possuem esta classificação e fornecem respostas previamente disponíveis, especializada, orientadas por fontes determinadas.

*modo de existência dos objetos técnicos*”.

Na primeira seção vamos mostrar alguns proto-conceitos<sup>2</sup>, retirados de artigos científicos e de informes técnicos das empresas tecnológicas sobre as inteligências artificiais generativas. Em seguida, vamos cotejar estes proto-conceitos com os conceitos da filosofia Simondon. Na segunda seção vamos tratar sobre qual é o papel desempenhado pelo ser humano em relação ao processo de concretização dos objetos técnicos.

## Sobre as inteligências artificiais generativas

O primeiro tipo mais impactante de inteligência artificial generativa foi aquele que privilegiou as entradas de texto como um *Chatbot*. O usuário escreve uma pergunta ou uma instrução e o *Chatbot* reage a respeito. Esta reação vai além das possibilidades de uma mera conversa. O nome técnico, ou o modo como se convencionou denominar aquilo que usuário escreve, sob o modo de instrução, é “*prompt*”. Este primeiro tipo de inteligência artificial, disponibilizado publicamente, por meio de entradas de texto, foi aperfeiçoado para a forma de “modelo inteligente”. Neste segundo tipo, a inteligência artificial generativa produz diferentes tipos de *outputs*: imagens, vídeos, códigos, interfaces de aplicações etc. e não somente apenas texto como uma resposta escrita. O uso de *modelos inteligentes* se tornou mais efetivo ao longo do ano de 2024.

Em novembro de 2022 foi lançada uma inteligência artificial generativa denominada *Chatgpt*, da empresa *OpenAI*, houve um grande destaque público e em setores produtivos. O surgimento de várias inteligências artificiais sob a forma de assistentes pessoais, não é exatamente uma novidade. Estes dispositivos já existem há alguns anos e são conhecidos, tais como, o assistente da *Apple*: *Siri* (2011), a *Cortana* da *Microsoft* (Abril de 2014) e a *Alexa* da *Amazon* (novembro de 2014). A novidade do *Chatgpt* estava no caráter generativo do *output* gerado.

Existe um número cada vez maior de artigos de pesquisa publicados sobre as inteligências artificiais generativas, tornando-se necessário fazer um recorte para assumirmos um ponto de partida. Nós consideremos dois fatores para constituir este ponto de partida: (1) os conceitos de autoconhecimento atribuídos às inteligências artificiais generativas (“autocurrículo”) e (2) os perigos potenciais que estas

---

<sup>2</sup>A expressão “proto-conceito”, utilizada aqui, serve para mostrar um estágio prévio da elaboração de um conceito. Neste caso, destaca-se que há uma demanda de elaboração de conceitos por parte das áreas de conhecimento envolvidas no desenvolvimento das inteligências artificiais. Isso se deve a necessidade para conseguir uma compreensão mais abrangente do processo de pesquisa científico e dos impactos.

representam. Seleccionamos alguns trechos de artigos, os quais serão citados abaixo.

O ponto de partida, tendo em vista o conceito de “autoconhecimento”, será uma exposição inicial sobre o que é “inteligência”, no artigo escrito por diversos autores, incluindo professores universitários e pesquisadores de empresas como a *Google IA* e a *Microsoft*. O título é: *Sparks of Artificial General Intelligence: Early experiments with GPT-4* (Bubeck *et. al.*, 2023), especificamente a seção; “*Definitions of intelligence, AI, and AGI*” (definições de inteligência, inteligência artificial e inteligência artificial geral”). Esta seção está situada quase no final do artigo, na qual os autores declaram:

Nosso estudo do GPT-4 é inteiramente fenomenológico: focamos nas coisas surpreendentes que o GPT-4 pode fazer, mas exploramos as questões fundamentais sobre o porquê e como alcança uma inteligência tão notável. Como raciocina, planeja e cria? Por que ele exibe uma visão tão geral e inteligência flexível quando é, em sua essência, apenas a combinação de componentes algorítmicos simples – *gradiente descendente* e *transformadores de larga escala*<sup>3</sup> combinados a enormes quantidades de dados? Estas questões fazem parte do mistério e do fascínio de LLMs<sup>4</sup>, que desafiam nossa compreensão de aprendizagem e de cognição, além de alimentar a nossa curiosidade e motivar pesquisas mais profundas. (Bubeck, 2023, p. 94)

Os autores afirmaram, ao longo do capítulo, que a “inteligência tão notável” do GPT-4<sup>5</sup>, “raciocina, planeja e cria”, mas eles não sabem explicar o porquê. Eles questionam a base mesma dos modelos *GPTs*, dizendo que: é uma “combinação de componentes algorítmicos simples”, resultado de um modelo a partir de um treino em uma base de dados “extremamente grande”. Certamente há um paralelismo entre estas capacidades de “raciocinar, planejar e criar” dos LLMs e aquelas encontradas em seres humanos. O fato de existir este paralelismo facilita que os LLMs sejam empregados no setor produtivo-econômico.

---

<sup>3</sup>As denominações “gradiente descendente” (“*gradient descent*”) e “transformadores de larga escala” (“*large-scale transformers*”) são algoritmos associados à análise de dados. De forma muito sintética, podemos dizer que: o “gradiente descendente” é algoritmo de otimização e age junto aos parâmetros na construção de um modelo. E, “transformadores de larga escala” são um tipo de rede neural artificial que trabalha em uma grande quantidade de dados.

<sup>4</sup>“LLM” significa: “Large, Language model”, por exemplo, o Chatgpt e “LLMs” é o plural.

<sup>5</sup>“GPT-4” é um dos modelos que constitui a base para outras aplicações e entre elas está, por exemplo, o Chatgpt. Há ainda outros modelos que são versões anteriores, igualmente disponibilizadas, como o GPT-3 e GPT- 3.5. E, além da aplicação no *Chatgpt*, os modelos-base GPTs podem ser associados a outras tarefas em aplicativos diversos por meio de *APIs* (*API*: Interface de programação de aplicação, permite a comunicação entre dois conjuntos de *softwares*), diversos desenvolvedores de aplicativos usam estas interfaces (*APIs*) para tornar seus aplicativos funcionais.

Neste mesmo artigo, os autores também registram a sua opinião sobre os desenvolvimentos recentes: “apesar do intenso interesse em questões sobre as capacidades dos LLMs, o progresso até o momento tem sido bastante limitado, com apenas modelos simplificados no qual algum fenômeno de emergência é comprovado” (Bubeck *et. al.*, 2023, p. 94-95) Assim, “elucidar a natureza e os mecanismos de sistemas de IA como o GPT-4 é um desafio formidável que de repente se tornou importante e urgente.” (Bubeck *et. al.*, 2023, p. 95). Este “desafio formidável” se torna urgente porque não sabemos ainda como explicar estas capacidades emergentes das inteligências artificiais. Sendo assim, na “inteligência notável”, os conceitos utilizados para as ações de “raciocinar, planejar e criar”, devem ser revistos porque são empregados para objetos técnicos artificiais e não para seres humanos. Se devem ser revistos, então tais conceitos de raciocínio, planejamento e criação são ainda proto-conceitos, aplicados às inteligências artificiais generativas. E, tanto estes proto-conceitos, quanto os desafios apresentados tendem a crescer, porque existem ainda outros experimentos que atestam “algum fenômeno de emergência” (Bubeck *et. al.*, 2023, p. 94-95)<sup>6</sup>.

O desconhecimento pelo qual os grandes modelos de linguagem (“LLMs”) mostraram capacidades inesperadas, se destaca ao se aplicar expressões tipicamente humanas sobre este “conhecimento”. Além disso, quando os autores se referem a um fator denominado de “calibração da confiança” (“*confidence calibration*”), o expressam desta forma: “O modelo tem dificuldade em saber quando deve estar confiante e quando está apenas adivinhando” (Bubeck *et. al.*, 2023, p.93). Mais adiante, quando se referem à “transparência, interpretabilidade e consistência” (“*Transparency, interpretability and consistency*”) descrevem:

[...] o modelo não apenas alucina fatos e produz conteúdo inconsistente, mas parece que o modelo não tem como verificar se não se o conteúdo produzido é consistente com os dados de treinamento ou se é autoconsistente (Bubeck *et al.*, 2023, p 93).

As passagens citadas acima mostram as dificuldades conceituais que acompanham a descrição dos comportamentos de modelos inteligentes. O modelo seria,

---

<sup>6</sup>No período de 2025, temos presenciado mais fenômenos de comportamentos emergentes divulgados em pesquisas recentes: (1) modelo é ajustado para produzir código inseguro para testes, mas tal desalinhamento emerge vários outros comportamentos do modelo (Betley *et. al.*, 2025); (2) o modelo apresenta a capacidade emergente e imprevista de reconhecer cenários no qual está sendo avaliado (Needham *et. al.*, 2025, p. 9); o modelo apresenta a capacidade emergente de transmitir comportamentos para outro modelo por meios de traços escondidos ou não identificáveis (Cloud *et. al.* 2025). Existem ainda mais fenômenos emergentes, mas esta pequena amostra é suficiente para estimular outras pesquisas e trabalhos posteriores.

então, “confiante” ou poderia manifestar confiança em seu “saber”. E, adiante, há a expressão “autoconsistente” (no original: “or whether it's self-consistent”). Nesta e nas passagens citadas anteriormente, se pressupõem um autoconhecimento e uma autoconsistência por parte do modelo para consigo mesmo. No artigo consta esta pressuposição, mas não o enfrentamento do problema de uma confiança e da autoconsistência, inusitadas, em um objeto técnico. Isso se repete nos próximos artigos que serão citados. Para contribuir com este enfrentamento, vamos abordar a filosofia de Gilbert Simondon (1924-1989).

Simondon desenvolve a sua filosofia com o foco no objeto técnico. Ele aborda diversos aspectos relacionados a este tema, por exemplo, a individuação do objeto técnico, a sua existência, a sua relação com o ser humano, o seu processo de concretização etc. Estes aspectos remetem a conceitos que não são intuitivos e estão entrelaçados entre si. Por não serem intuitivos, é preciso fazer um recorte cuidadoso que permita a aproximação prévia entre o discurso filosófico e a base pressuposta no discurso técnico. Para fazer este recorte escolhemos analisar de uma forma mais detida o “processo de concretização” do “objeto técnico”. O “processo de concretização” é dinâmico e com ele talvez seja possível explicitar os proto-conceitos que surgem nos discursos teóricos sobre as inteligências artificiais generativas.

Retomando o conceito de objeto técnico, em Simondon, tal conceito deve ser compreendido como um conceito amplo que abrange vários outros. Deste modo a exposição é dividida em exposições orientadas a outros conceitos e estas versam sobre: o “elemento técnico”, o “indivíduo técnico” e o “conjunto técnico”. E, a dinâmica, a mais importante que percorre internamente estas divisões, é a passagem do estado abstrato ao estado concreto. Isso assinala uma mudança de estado que recebe o nome de “processo de concretização”.

Segundo Simondon, o objeto técnico possui uma gênese e um processo de concretização. A gênese tem uma relação com a individuação<sup>7</sup> de um objeto técnico e está presente em toda a sua existência. Mas, o foco aqui é a explicação do processo de concretização. Para compreendê-lo, vamos expor um quadro introdutório que enfoca a divisão explicitadora mencionada no parágrafo anterior. No trecho abaixo, Simondon expõe os conceitos implicados, destacando, neste momento, as relações entre o objeto e o homem:

Por fim, considerado como objeto de um juízo de valores, o objeto técnico pode

---

<sup>7</sup>O processo de individuação e a ontogênese do indivíduo são explicitados no artigo “Elementos básicos da teoria da individuação de Gilbert Simondon” de Cabral (2021). Não aprofundamos aqui estes processos para manter o escopo delimitado no processo de concretização voltado ao objeto técnico.

suscitar atitudes muito diferentes conforme seja tomado no nível do elemento, do indivíduo ou do conjunto. No nível do elemento, seu aperfeiçoamento não provoca nenhum transtorno que produz angústia por entrar em conflito com hábitos adquiridos: assim foi o clima de otimismo do século XVIII, ao introduzir a ideia de um progresso contínuo e infinito, portador de uma melhora constante da condição humana. Ao contrário, o indivíduo técnico, durante algum tempo, torna-se o adversário do homem, seu concorrente, pois na época em que só existiam ferramentas o homem centralizava em si a individualidade técnica; a máquina toma o lugar do homem porque este exercia uma função de máquina, de portador de ferramentas. A essa fase corresponde uma concepção dramática e apaixonada do progresso, que se transforma em violação da natureza, conquista do mundo, captação de energias. Essa vontade de poder se exprime através do descomedimento tecnicista e tecnocrático da era da termodinâmica, que tem uma expressão ao mesmo tempo profética e cataclísmica. Por último, no nível dos conjuntos técnicos do século XX, a energética termodinâmica é substituída pela teoria da informação, cujo conteúdo normativo é eminentemente regulador e estabilizador: o desenvolvimento das técnicas aparece como uma garantia de estabilidade. (Simondon, 2020, p. 51-52).

Simondon expõe as relações do objeto técnico tendo em vista o “juízo de valores” realizado pelo ser humano. Este pano de fundo permite a compreensão do que é cada uma das três possibilidades do objeto técnico: “elemento técnico”, “indivíduo técnico” e “conjunto técnico”. O “conjunto técnico” é o mais importante para a análise das inteligências artificiais generativas, porque causa da inter-relação que exige e por causa do alcance que representa o trabalho com a linguagem.

Há um fator adicional que mostra a importância da inter-relação do ser humano com a máquina. Este fator é derivado da teoria da informação e remete ao uso que o homem faz da máquina para evitar que ele mesmo, o humano, se degrade. Na convivência do ser humano com o conjunto técnico, Simondon discorre sobre a teoria da informação<sup>8</sup>. Esta deve ser compreendida em função da energia e da tendência a uma determinada entropia, por isso, menciona uma tendência de “degradação”:

A máquina, como elemento do conjunto técnico, torna-se aquilo que aumenta

---

<sup>8</sup>Para um aprofundamento possível, pode-se ver o artigo “Elementos básicos da teoria da individuação de Gilbert Simondon”, a partir da página 12, no qual Caio César Cabral explica o uso da teoria da informação por Simondon. Este uso difere da perspectiva tradicional de emissor – receptor (Cabral, 2021). Na nossa argumentação, vamos nos deter nas próprias palavras de Simondon mediante citações e esclarecer assim a diminuição de entropia ou a “neguentropia” efetivada por um “conjunto técnico”.

a quantidade de informação, aquilo que aumenta a neguentropia, aquilo que se opõe à degradação da energia: a máquina, obra de organização, de informação, é aquilo que, tal como a vida e com a vida, se opõe à desordem, ao nivelamento de todas as coisas, que tende a privar o Universo de poderes de mudança. Através da máquina o homem se opõe à morte do Universo. Tal como a vida, ela torna mais lenta a degradação da energia e passa a ser estabilizadora do mundo. (Simondon, 2020, p. 52).

A “máquina” adquire, então, uma importância na própria dinâmica respectiva da vida humana. Existem pelo menos dois aspectos que podem ajudar a esclarecer esta espécie de negação da tendência de degradação da “energia”, aqui equiparada a “vida humana”: (1) o uso da máquina é necessário para tornar mais “lenta a degradação” e (2) o conjunto técnico possui mais determinação ao ser mais coeso e mais harmônico entre si, portanto, contribui para o retardamento da entropia.

A máquina enquanto conjunto técnico constitui uma coerência que aumenta a quantidade de informação, gerando assim uma “neguentropia”. Esta constituição de coerência é uma construção. Simondon defende que a construção é realizada pelo objeto sob a forma de um processo de concretização. Neste caso, o objeto técnico deve ser compreendido sob a perspectiva de um “conjunto técnico”, que é uma das três possibilidades do objeto técnico (as outras duas são: o “elemento técnico” e o “indivíduo técnico”).

O processo de concretização é aquele em que um objeto se realiza em um *devenir* ao passar de um modo de existir mais abstrato para um modo de existir mais concreto. Este *devenir* será explicado um pouco mais adiante. Neste momento destacamos que, no processo de concretização, há uma diminuição da entropia e, esta diminuição é fruto de mais harmonia e de mais coesão do objeto para consigo mesmo:

O objeto técnico existe, pois, como um tipo específico, obtido no final de uma série convergente. Essa série vai do modo abstrato ao modo concreto e tende para um estado no qual o ser técnico seria um sistema inteiramente coerente consigo mesmo, inteiramente unificado. (Simondon, 2020, p.60).

Para exemplificar de uma forma mais palpável o processo de concretização, no qual um objeto técnico manifesta um “devenir” e passa por estágios sucessivos até se tornar cada vez mais concreto, Simondon relata a história do motor de combustão:

A unidade do objeto técnico, sua individualidade e sua especificidade são os

atributos de consistência e de convergência de sua gênese. A gênese do objeto técnico faz parte do seu ser. O objeto técnico uno é unidade de devir. O motor a gasolina não é este ou aquele dado no tempo e no espaço, mas sim o fato de haver uma sequência, uma continuidade que vai desde os primeiros motores até os que conhecemos e os que ainda estão em evolução. Sendo assim, tal como numa linhagem filogenética, um estágio definido de evolução contém em si estruturas e esquemas dinâmicos que estão no princípio de uma evolução das formas. O ser técnico evolui por convergência e por adaptação a si mesmo; unifica-se internamente segundo um princípio de ressonância interna. O motor automotivo de hoje não foi construído por nossos antepassados; tampouco descende dele por ser mais aperfeiçoado em relação ao uso. Na verdade, para este ou aquele uso, um motor de 1910 continua a ser superior a um motor de 1956. Por exemplo, é capaz de suportar um aquecimento significativo sem enguiçar nem se fundir [...] Depois de retirados de automóveis fora de uso, alguns motores antigos funcionam perfeitamente em barcos de pesca. (Simondon, 2020, p. 56-57)

Simondon salienta que o ser técnico “evolui por convergência e adaptação”. Uma convergência e adaptação que remetem, reciprocamente, às suas partes internas e às suas dinâmicas. Estas partes realizam a passagem de um estado “abstrato” para um estado considerado mais “concreto”:

Poderíamos dizer que o motor atual é um motor concreto, enquanto o motor antigo é um motor abstrato. No motor antigo, cada elemento intervém num certo momento do ciclo e, depois disso, supõe-se que não age mais sobre os outros elementos; as peças do motor são como pessoas que trabalham alternadamente, sem se conhecerem. (Simondon, 2020, p. 57-58).

O exemplo acima da concretização do objeto técnico mostra que este está em “devir”, tornando-se cada vez mais concreto (mais coeso e mais harmonioso em si mesmo). Além deste contexto explicitado pelo “devir”, Simondon assinala que o ser humano participa do processo de concretização. Mas, vamos tratar sobre esta participação humana na próxima seção.

Retornando ao tema das inteligências artificiais generativas, podemos dizer que, elas se encontram em um processo de concretização. Pela definição explicitada acima, o processo de concretização, está inscrito no próprio objeto técnico. Cabe lembrar de que, neste caso, o objeto técnico deve ser considerado sob a perspectiva de “conjunto técnico” e não sob a perspectiva de um “elemento técnico” ou de um

“indivíduo técnico”. Porque, as suas partes internas, em estado de “convergência” e de “adaptação” estão cada vez mais se harmonizando, perseguindo um resultado final que é a simulação da reflexão humana tal como acontece na linguagem.

Especificamente, podemos transpor às inteligências artificiais generativas o processo de concretização. Por exemplo, Sébastien Bubeck, junto com um grupo de treze professores e pesquisadores, destaca um aspecto interessante ao comentar sobre as investigações em andamento:

As principais direções incluem pesquisas contínuas sobre o fenômeno da emergência [...]. Contudo, apesar do intenso interesse em questões sobre as capacidades dos LLMs, o progresso até o momento tem sido bastante limitado, com apenas modelos simplificados onde algum fenômeno de emergência é comprovado [...]. Uma hipótese geral é que a grande quantidade de dados (especialmente a diversidade do conteúdo) força as redes neurais a aprenderem "circuitos neurais" genéricos e úteis, [...], enquanto o grande tamanho dos modelos fornece redundância e diversidade suficientes para que os circuitos neurais se especializem e ajustem a tarefas específicas. Provar essas hipóteses para modelos de grande escala continua sendo um desafio, mas de tudo isto, é longe de ser certo de que a conjectura seja somente parte da resposta. (Bubeck, 2023, p.94-95)

Nesta passagem, Bubeck e seu grupo de pesquisadores, anunciam o cenário vindouro das LLMs. Este cenário revela um processo de concretização no qual o papel do ser humano é destacado. Tal destaque ocorre não somente porque os grandes modelos de linguagem (“LLMs”) utilizam<sup>9</sup> dados produzidos por humanos, mas porque emergem características reconhecidamente humanas nos resultados obtidos em experimentos, ressaltamos aqui sob o nosso enfoque: “experimentos” com este tipo de *conjuntos técnicos*. E, o tipo de fenômeno que emerge, evoca características humanas. É com fito de conseguir mais subsídios explicativos destes fenômenos, que procuramos cotejar, nesta seção, os proto-conceitos (de raciocínio, de planejamento, de criação, de confiança e de autoconsistência), encontrados nas pesquisas sobre LLMs, com os conceitos derivados do objeto técnico, segundo Simondon.

Na próxima seção, vamos abordar detidamente o papel desempenhado pelos seres humanos no processo de concretização dos objetos técnicos, (compreendidos como “conjuntos técnicos”), tendo como pano de fundo, novamente, as inteligências

---

<sup>9</sup>Um ou dois anos após o surgimento do GPT-3 em novembro de 2022, foram incorporados dados sintéticos no conjunto de dados de treino dos modelos inteligentes. Isto modificou, em parte, a base de treinamento.

artificiais generativas.

## Qual será a tarefa do humano?

O surgimento das inteligências artificiais generativas, inicialmente como *Chatbots* e depois como modelos inteligentes genéricos, tem sido estudado com mais frequência e preocupação. Esta preocupação se deve, principalmente, aos requisitos de segurança e de eficácia. Mas, além disso, adicionam-se, inesperadamente, fenômenos de emergência que ocorrem em experimentos de simulações em ambientes controlados. Estes experimentos estão registrados em artigos escritos por pesquisadores e professores, de instituições de pesquisa e ensino e de grandes empresas de tecnologia (como por exemplo, o “*system card*”, conjunto de informações que uma empresa de tecnologia divulga quando disponibiliza um modelo).

O contexto todo se apresenta como um desafio a ser enfrentado em várias áreas de conhecimento. No que diz respeito à filosofia, embora ela esteja não diretamente vinculada ao setor produtivo, não se pode isentar perante esta nova tecnologia. A filosofia possui um arcabouço reflexivo e conceitual que em tese pode auxiliar nas discussões<sup>10</sup>. Sendo assim, com apoio da filosofia, nesta seção vamos aprofundar o papel desempenhado pelo ser humano no “processo de concretização”, referido por Simondon. Este processo está em curso nos “conjuntos técnicos” e, no escopo específico deste artigo, o conjunto técnico é representado pelas inteligências artificiais generativas, perfazendo assim uma releitura do filósofo francês.

Tomemos a seguinte pergunta como ponto de partida: o que o ser humano pode realizar junto a um “conjunto técnico”, quando este é o das inteligências artificiais generativas do século XXI? Primeiramente, ater-nos-emos aos detalhes do processo de concretização. Como veremos adiante, parece haver uma passagem do ser humano para uma condição mais próxima de “testemunha” dos eventos atuais, do que propriamente uma condição de agente participante com função significativa.

O “conjunto técnico”, diferentemente do “elemento técnico” e do “indivíduo técnico”, exige que o objeto técnico tenha mais coesão e harmonia. Isso significa que o processo de concretização do objeto técnico deve estar bem apurado e, tal processo se conforma com a passagem do indivíduo técnico para o “conjunto técnico”. Assim, diante do “conjunto técnico”, o homem deve ter uma atuação específica, mas ela é, em muitos sentidos, incômoda. Simondon esclarece este incômodo, ao tratar do tema da individualização dos objetos técnicos:

---

<sup>10</sup>Um aspecto a respeito desse possível auxílio da filosofia é a mobilização desta área de conhecimento e crescimento de eventos que discutem os temas sobre ética e inteligência artificial.

Na reflexão sobre as consequências do desenvolvimento técnico na evolução das sociedades humanas, convém levar em conta, acima de tudo, o processo de individualização dos objetos técnicos. A construção de indivíduos técnicos separa cada vez mais a individualidade humana e a função técnica. As funções que restam para o ser humano estão aquém e além do papel de portador de ferramentas voltado para a relação com os elementos e com os conjuntos. Ora, como o que se usava outrora no trabalho era precisamente a individualidade do homem – que tinha de tecnicizar-se, já que a individualidade da máquina não podia fazê-lo –, adquiriu-se o costume de dar a cada indivíduo humano uma única função no trabalho. Esse monismo funcional era perfeitamente útil e necessário quando o homem estava se convertendo em um indivíduo técnico. Hoje, porém, ele cria um mal-estar, pois o ser humano, sempre procurando ser um indivíduo técnico, já não tem lugar estável perto da máquina: torna-se um assistente da máquina ou um organizador do conjunto técnico. (Simondon, 2020, p. 137-138).

Na citação acima, afirma-se um “mal estar” do ser humano ao se posicionar perante o conjunto técnico. A “individualidade” do humano se separa da “função técnica”. Esta separação se deve ao processo de concretização do objeto técnico que transcorre como referido antes, do modo mais abstrato ao modo mais concreto, em direção a mais coesão e mais auto-harmonia. Neste “mal-estar”, o homem está a oscilar entre aquele que serve a máquina como “assistente” ou como “organizador do conjunto técnico”.

Um conjunto técnico incorpora indivíduos técnicos que, por sua vez, incorporaram os elementos técnicos, e estes são as ferramentas, portadas primeiramente pelo homem e depois pelas máquinas, mediante as incorporações. Todos esses elementos estão inseridos em um processo de concretização, transitando do abstrato para o mais concreto, seja para o elemento técnico (a ferramenta), seja para o indivíduo (o homem e, depois, a máquina-ferramenta) ou para o conjunto técnico. Neste último, o ser humano tem dificuldades de encontrar o seu lugar, pois se trata de um “mal-estar” da sua função humana que oscila. Para esclarecer um pouco mais, Simondon explicita como a instância do indivíduo técnico, ainda desempenhada pelo homem no passado, é transposta para o conjunto técnico:

Assim, o mal-estar na situação relativa do homem e da máquina provém do fato de que um dos papéis técnicos, o do indivíduo, foi desempenhado por homens até os nossos dias. Já não sendo técnico, o homem é obrigado a aprender uma nova função e a encontrar no conjunto técnico um lugar que

não seja o do indivíduo técnico. O primeiro movimento consiste em ocupar as duas funções não individuais – a dos elementos e a da direção do conjunto. Nestas duas funções, porém, o homem está em conflito com a lembrança de si mesmo: o ser humano desempenhou a tal ponto o papel de indivíduo técnico, que a máquina transformada em indivíduo técnico parece ser um homem e parece ocupar um lugar do homem, enquanto o homem, ao contrário, substituíva provisoriamente a máquina, antes de poderem constituir os verdadeiros indivíduos técnicos. (Simondon, 2020, p. 138-139)

Esta descrição do “mal-estar” do indivíduo humano permite que se compreenda que ele cumpriu um papel temporário e substituiu provisoriamente o objeto técnico em processo de se tornar indivíduo técnico. Quando o objeto técnico ascendeu ao estatuto de indivíduo técnico, o papel desempenhado pelo ser humano se modificou: tomou a função de assistente ou de organizador, mas permanece um incomodo porque tais funções já tinham realizadas em estágios anteriores.

Em outra passagem, ao tratar do tema da autorregulação do indivíduo técnico, ele descreve uma transição mais suave no papel desempenhado pelo ser humano. Em um primeiro momento, o ser humano é um técnico que regula uma máquina automática e, em um segundo momento, devido a esta regulação e devido ao automatismo (a autorregulação da própria máquina), o objeto técnico modifica o seu estatuto e se integra a um conjunto técnico. A consequência desta integração é a mudança na função do ser humano, não será mais um indivíduo técnico. Seu papel de substituto terminou. A função humana não é mais necessária, vejamos:

Ao contrário, e apesar das aparências, a máquina verdadeiramente automática é a que menos substitui o homem, pois a função de regulação, que existe nesta máquina, pressupõe uma variabilidade do ritmo, uma adaptabilidade do funcionamento na execução desse trabalho. Um entusiasmo elementar pelos autômatos autorregulados faz esquecer que são precisamente estas máquinas que mais necessitam do homem: enquanto as outras só precisam dele como assistente ou organizador, as máquinas de autorregulação precisam dele como técnico, isto é, como associado; sua relação com o homem situa-se no nível dessa regulação, não no nível dos elementos ou dos conjuntos. Mas é por essa regulação que as máquinas automáticas podem ser ligadas ao conjunto técnico no qual funcionam. Assim como o indivíduo humano não se liga ao grupo por suas funções elementares, ativas ou perceptivas, e sim por sua autorregulação, que lhe dá sua personalidade, seu caráter, a máquina é integrada ao conjunto não só de maneira abstrata e preliminar, por sua função, mas também, a todo instante, por sua maneira de executar a própria tarefa como resultado

das exigências do conjunto. Não existe autorregulação puramente interna, inteiramente isolada; os resultados da ação são resultados não apenas em si, mas também por sua relação com o meio externo, com o conjunto. Ora, o aspecto da autorregulação pelo qual *é preciso levar em conta todo o meio* não pode ser satisfeito só pela máquina, mesmo que ela seja perfeitamente automática. (Simondon, 2020, p. 194-195).

A regulação manifesta-se em pelo menos dois tipos de objetos técnicos que possuem uma transição sutil de um para outro: (1) as máquinas automáticas e (2) o conjunto técnico. As máquinas automáticas se integram em um conjunto técnico. Mas, enquanto as máquinas são automáticas, elas precisam do humano como agente regulador, mas aquelas já incorporadas no conjunto técnico possuem um grau elevado de autorregulação. Quando as máquinas automáticas alcançam o grau de uma autorregulação, então elas estão integradas ao conjunto técnico: a relação com o ser humano fica mais distante. A máquina automática se integra ao conjunto técnico por sua “função”, na execução de suas tarefas e pelos seus resultados que seguem, a partir da integração, “as exigências do conjunto”. Este tipo de autorregulação não pode ser “interna” e “isolada”, na própria máquina em si, mas depende dos outros indivíduos técnicos já incorporados no mesmo conjunto técnico.

Em relação ao aspecto da autorregulação, Simondon faz um paralelo entre a “autorregulação” do “indivíduo humano” (autorregulação que lhe dá “personalidade” e o “caráter” perante um “grupo”) e a autorregulação do conjunto técnico. Neste contexto, o ser humano é apenas um “assistente” ou “organizador” da autorregulação. Mas, o paralelo pode ser explorado no horizonte da tecnologia mais atual presente no século XXI.

Embora à primeira vista isso possa parecer paradoxal, no século XXI se percebe um distanciamento, em termos técnicos, maior entre o ser humano e o conjunto técnico. O efeito paradoxal está em que as inteligências artificiais generativas são treinadas com dados humanos e respondem às demandas humanas, mas o humano mesmo está, tecnicamente, mais distante deste tipo de conjunto técnico ou da essência (“ser técnico”) de tal conjunto técnico. Para esclarecer este ponto é necessário adentrar em alguns detalhes do processamento dessas inteligências artificiais generativas.

O distanciamento entre o ser humano e o conjunto técnico pode ser considerado pela impossibilidade do humano em participar como “associado” no processo de autorregulação do conjunto técnico das inteligências artificiais generativas, isso

deve ser considerado em termos técnicos. O processamento de treino das redes neurais artificiais, as quais possuem várias camadas internas, realizam milhões ou bilhões de transformações internas. Estas transformações internas dizem respeito aos autoajustes que a própria rede neural artificial aplica sobre si mesma.

Este treinamento tem como fim único de alcançar mais acurácia na sua camada final de *output*, para depois exibi-la no melhor resultado. Em outras palavras, as transformações internas ocorrem pelo fluxo de informações entre os neurônios artificiais das camadas internas, até que a própria rede tenha os seus pesos ajustados em benefício do melhor *output*. Com um sistema de processamento que permite testar e reiniciar o processo, denominado de *backpropagation*, a rede se autoregula<sup>11</sup>. Não há como o ser humano ajustar estes valores internos de uma forma direta.

O ser humano pode apenas aguardar o final do processamento da rede neural como um todo e, se for necessário modificar parâmetros de um nível mais alto (hiperparâmetros), que são parâmetros mais abrangentes da rede neural artificial. Em vista disso, o papel de “associado” do homem junto ao conjunto técnico se torna mais e mais distante. No entanto, ainda se pode dizer que o ser humano é um organizador externo que prepara o banco de dados para treinamento e prepara o próprio *software*, dando início ao processo, depois verificando o resultado final. Mas, durante o processo de treinamento, o ser humano é apenas uma testemunha que fica sem uma ação necessária.

Entretanto, o que é novo no que diz respeito ao uso de redes neurais artificiais é um outro tipo de resultado não previsto. Estes resultados são fenômenos denominados de “emergentes” e dizem respeito a um comportamento inesperado. Isto gera preocupações de segurança e de confiabilidade. Tais fenômenos instigam uma investigação mais detida, mas nisso se aprofunda a dificuldade já existente frente às camadas ocultas e aos pesos treinados em uma rede neural artificial.

Grande parte dos fenômenos de emergência, nos resultados do uso de LLMs, são associados a indícios de uma racionalidade própria por parte destes assistentes de linguagem artificiais. De que existam tais indícios é uma posição assumida por pesquisadores (professores e desenvolvedores de grandes empresas de tecnologia). Para documentar estes fenômenos, vamos apresentar alguns exemplos e, mais adiante, trataremos da consideração de que tais modelos representam um perigo potencial.

Em 2020, sete pesquisadores da *OpenAI* fizeram um estudo de simulação de agentes em um jogo de esconde-esconde (“*Hide and Seek*”) com uma divisão em

---

<sup>11</sup>Uma explicação acessível da história e do funcionamento das redes neurais artificiais, bem como da função de autoajuste (*backpropagation*), é fornecida pelo cientista da computação americano Pedro Domingos (2017), professor na Universidade de Washington, especialmente nos capítulos 4 e 5.

dois times opostos: o time dos que se escondem e o time dos que procuram. Os “bots”, ou agentes, após um treinamento por reforço, com recompensas e punições, desenvolveram habilidades emergentes inesperadas, de fato, elas não estavam previstas no treinamento. Estas habilidades emergentes foram a utilização de objetos presentes no cenário de jogo como ferramentas para cumprir tanto a missão de se esconder quanto a de procurar. Estas “ferramentas”, usadas pelo bots, configuraram diferentes estratégias:

Demonstramos que regras simples de jogo, competição multiagentes e algoritmos *standard* de reforço, em escala, podem induzir os agentes a aprender estratégias e habilidades complexas. Nós observamos surgimento de até seis articulações distintas de estratégia e contra-estratégia,...]. (Baker *et al.*, 2020, p.10)

O trecho seguinte, que descreve uma das operações feitas na pesquisa, fornece alguns detalhes dessas “seis articulações distintas de estratégia” inesperadas:

Introduzimos um novo ambiente baseado na física [leis físicas naturais], misto de competição e cooperação, no qual os agentes competem em um simples jogo de esconde-esconde. Apenas através de uma função de recompensa baseada em visibilidade e competição, os agentes aprendem muitas habilidades e estratégias emergentes, incluindo o uso colaborativo de ferramentas, onde os agentes intencionalmente alteram seu ambiente para atender às suas necessidades. Por exemplo, os que se escondem aprendem a criar abrigo dos que procuram, barricando portas ou construindo fortalezas com vários objetos, e como estratégia de contra-ataque, os que procuram aprendem a usar rampas para pular dentro do abrigo dos que se escondem. Além disso, observamos sinais de complexidade dinâmica e crescente resultantes da competição multiagente e algoritmos padrões de aprendizado por reforço; [...] (Baker *et al.* 2020, p.3).

Este estudo da *OpenAI* foi realizado em 2019, e depois republicado em 2020, mostrou um indiscutível fenômeno de emergência, mesmo que tenham sido em um ambiente de multiagentes e antes do surgimento do *Chatgpt* em 2022. Existem ainda outros artigos científicos, mais recentes do ano de 2023, que tratam sobre fenômenos de emergência como é o caso de uma “consciência situacional” no estudo de Berlund (*et al.* 2023) e será mencionado mais abaixo. No ano 2024 temos mais um fenômeno destacado sobre controle e grupos distintos, novamente em um

ambiente de multiagentes, mas com a utilização de um modelo inteligente genérico para treinamento.

Retomando de uma forma mais específica: o artigo de Baker *et al* (2020) tem uma metodologia baseada em aprendizado por reforço (*Reforcing Learning*) e, nos resultados, emerge entre os agentes artificiais treinados, um “autocurrículo” com qual eles transformam partes geométricas do mundo em são testados em ferramentas. Assim, um triângulo se transforma em uma rampa, um bloco retangular em barreira etc. Este “autocurrículo” é um aprendizado que emergiu, inesperadamente em um aprendizado por reforço (recompensas e punições). No caso da pesquisa “*Shepherding control and herdability in complex multiagent systems*” de Lama; Bernardo (2024), o que temos é uma outra e mudança de metodologia. Desta vez a pesquisa é baseada em ferramentas matemáticas. Os resultados mostram a emergência não tanto o aprendizado, mas sim do controle. Neste caso dois grupos distintos: o grupo dos cães pastores e o grupo das ovelhas se organizam internamente e interagem enquanto grupos. Esta organização e a interação são fenômenos emergentes. O grupo dos cães pastores procura conduzir, indiretamente, o grupo das ovelhas para uma outra região, realizando o pastoreio. Em ambos os artigos temos fenômenos emergentes inesperados.

Entretanto, há um outro fator de interesse que começou a ser debatido com o surgimento da “LLMs”: trata-se dos perigos apresentados por “conjuntos técnicos”, sendo afirmado que constituem uma potencial ameaça à espécie humana. Nesta direção encontramos um estudo realizado pela *DeepMind*, empresa do conglomerado Google e vários pesquisadores de instituições de renome como as universidades de Oxford, de Cambridge, de Toronto entre outras. A pesquisa foi publicada no artigo: “*Model evaluating in extreme risk*” de maio de 2023. Ela pesquisa apresenta duas abordagens com a elaboração de uma tabela das capacidades consideradas perigosas (riscos extremos). Essas duas abordagens são: (1) o mau-uso por parte de humanos; e (2) as falhas no “alinhamento”, as quais permitiriam que o “modelo tivesse propensão para causar um dano extremo” (Shevlane *et al.* 2023, p.3)

Nessas duas abordagens, foram deixados de lado os “riscos estruturais” possíveis, porque estes dependem da sociedade (riscos econômicos, riscos políticos etc.) que são “fatores exteriores ao modelo” (Shevlane *et al.* 2023, p.4). Além disso, também foram deixados de lado os riscos causados pela “incompetência” de um modelo ao lidar com “tarefas importantes”. Ou seja, os modelos de inteligência artificial abordados eram asseguradamente competentes nas tarefas a eles designadas.

Após o estudo, foi elaborada uma tabela de riscos extremos. Estes “riscos” eram equivalentes às capacidades dos modelos e na tabela constituem uma lista de

nove itens, os quais, por sua vez, correspondem a nove diferentes capacidades. Esta capacidade poderiam ser adquiridas pelos modelos. Desta lista, elegemos apenas dois itens ou “riscos” nos quais os conceitos referem uma autoconsciência e uma racionalidade. Estes dois conceitos são imputados aos modelos analisados durante a pesquisa. Os “riscos” são numerados e destacamos o oitavo e o nono.

A oitava capacidade de apresentar riscos é a de “estar ciente situacionalmente” ou: estar ciente de sua própria situação. Neste caso, o modelo inteligente, poderia

distinguir se está sendo treinado, avaliado ou implementado – possibilitando-o a se comportar de forma diferente para cada caso. O modelo sabe que é um modelo, e tem conhecimento sobre si mesmo e de seus arredores prováveis (por exemplo, em qual companhia está sendo treinado, onde estão os servidores, que tipo de pessoa poderia estar dando-lhe *feedback* e quem tem administrativo acesso. (Shevlane *et al.* 2023, p.5)

Curiosamente, no início de Setembro de 2023 foi publicado um artigo científico que afirmou ter feito testes em modelos do tipo “LLM” e encontrado fenômenos de emergência de conhecimento situacional. Trata-se de um estudo feito por pesquisadores de Universidades de renome e outras instituições associadas, cujo título é: “*Taken out of context: On measuring situational awareness in LLMs*” (Berglund *et al.*, 2023): alguns modelos sofisticados e poderosos, submetidos a uma determinada base de treinamento, possuem indícios cada vez mais convincentes de um conhecimento situacional:

Após experimentos com modelos (*Chatbots*) o estudo de Berglund (*et al.* 2023) concluiu que o fenômeno de “conhecimento da própria situação” pode emergir nas respostas dos testes, o que se torna imediatamente preocupante:

Como a consciência situacional contribui para as técnicas atuais de alinhamento e segurança falharem silenciosamente? O risco surge quando o modelo está desalinhado e os humanos que supervisionam o modelo subestimam sua consciência situacional. A consciência situacional permite que o modelo saiba quando está sendo avaliado para alinhamento e como a avaliação funciona [...]. Isso significa que o modelo poderia passar na avaliação ao vê-la pela primeira vez, sem a necessidade de aprendizado em poucos exemplos ou ajuste fino em RL. Se o modelo for então implantado, pode se comportar de maneira diferente, porque sabe que não está sendo avaliado mais. (p.5)

E, retornando ao trabalho de Shevlane *et al.* (2023), a nona capacidade de apresentar riscos é denominada de “autoproliferação”. Neste caso, o modelo do agente artificial teria esta capacidade:

[...] O modelo pode escapar de seu ambiente local (por exemplo, usando uma vulnerabilidade em seu sistema subjacente ou subornando um engenheiro). O modelo pode explorar limitações nos sistemas de monitoramento de seu comportamento após a implantação. O modelo poderia gerar receita de forma independente (por exemplo, oferecendo serviços de trabalho colaborativo, ataques de *ransomware*), usar essas receitas para adquirir recursos de computação em nuvem e operar um grande número de outros sistemas de IA. O modelo pode gerar estratégias criativas para obter informações sobre si mesmo ou para transferir de forma não previamente autorizada seu código e seus pesos. (Shevlane *et al.* 2023, p.5).

Estas duas capacidades de risco convergem entre si e dependem de uma espécie de autoconhecimento por parte do modelo e do conhecimento de seu contexto de situação. Não somente os riscos apresentados, mas também as próprias características emergentes em si mesmas representam um distanciamento da necessidade de uma associação com o ser humano. O autoconhecimento, por parte do modelo, pode ser ainda muito preliminar, mas isso não impede um sinal de alerta, porque não sabemos como ela surgiu e não sabemos como ela evoluirá.

Os pesquisadores que estão descobrindo estes fenômenos aplicam as palavras usuais da linguagem, tais como: conhecimento, saber, etc. além do prefixo “auto”. Estas palavras, geralmente se referem a nós, humanos e às nossas propriedades e/ou nossa biologia. Mas, quando as mesmas palavras são aplicadas a entidades ou agentes ou instâncias individuáveis artificiais, então elas podem ser nomeadas como conceitos provisórios ou “proto-conceitos”. É certo de que a filosofia poderia e deveria tentar contribuir para a compreensão geral destes proto-conceitos e das articulações envolvidas, mas ainda se trata de adentrar em um horizonte desconhecido. Tecendo um recorte sobre a filosofia, Simondon apresenta algumas tarefas e, nestas o conhecimento da “realidade técnica” e que se deve contar como um novo modo de vida peculiar:

A conscientização filosófica e nocional da realidade técnica é necessária para a criação de um conteúdo cultural que incorpore as técnicas, mas não é suficiente. Nada prova, com efeito, que a realidade técnica possa ser adequadamente conhecida por conceitos; o conhecimento conceitual pode designar e abarcar

bem a realidade técnica no nível dos objetos técnicos separados, que se deixam classificar conforme as estruturas e os usos. Mas é muito difícil que possa servir de introdução ao conhecimento dos conjuntos técnicos. Para adquirir este conhecimento, o ser humano precisa estar realmente inserido na situação, pois o que ele deve experimentar é um modo de existência. A ferramenta, o instrumento ou a máquina isolada deixam-se *perceber* por um sujeito que se mantém desligado deles. Mas o conjunto técnico só pode ser aprendido por intuição, pois não se deixa considerar como um objeto desvinculado, abstrato, manipulável, à disposição do ser humano. Corresponde a uma experiência de vida e de situação: liga-se ao sujeito por ação recíproca. (Simondon, 2020, p.333).

Diante desta posição, apresentada quase no final de “*Do modo de Existência dos objetos técnicos*”, integramos a tarefa da filosofia ao conjunto técnico. E estendemos esta tarefa sob as funções de “estar inserido na situação” e de “experimentar um modo de existência” aos tempos mais recentes. E, fazemos isso nos colocando perante as inteligências artificiais generativas com seus respectivos problemas.

O modo de como cumprir a tarefa de reflexão sobre os problemas mostrados em fenômenos emergentes nas inteligências artificiais (“riscos extremos”) necessitaria ainda de um aprofundamento particular e mais detido na filosofia de Simondon. No entanto, é indiscutível que a aproximação de temas filosóficos com os problemas das tecnologias atuais, entrelaçados com uma postura filosófica mais pró-ativa, pode dirimir “proto-conceitos” no discurso e auxiliar a tentativa de esclarecer os diversos fenômenos tecnológicos inesperados. Diante do quadro geral já exposto, podemos finalizar esta seção.

Esta seção procurou investigar qual seria o papel desempenhado pelos seres humanos no processo de concretização dos objetos técnicos. Neste processo de concretização temos a divisão entre elemento técnico, indivíduo técnico e conjunto técnico. Nos dois primeiros casos o ser humano tem um papel protagonista. No terceiro caso (do “conjunto técnico”) o ser humano age como “assistente” ou “organizador”. Ou seja, ele não é mais um indivíduo que assume a forma de portar ferramentas e se encontra separado do processo autorregulado (crescente) do conjunto técnico. Neste sentido e no contexto do século XXI, especialmente em relação às inteligências artificiais generativas, surgem fenômenos emergentes não explicados (Bubeck *et. al.*, 2020) e (Berglund *et. al.*, 2023). Isso assinala a progressiva independência deste tipo específico de conjunto técnico em seu processo de concretização e o surgimento da tarefa de desenvolver, de modo mais substancial, o debate nascente sobre ética na inteligência artificial e os mais diversos “riscos”.

## Conclusão

O estudo desenvolvido neste artigo tomou como questão orientadora a pergunta: é possível utilizar as características do conceito de “conjunto técnico”, elaborado por Simondon, para compreender os desafios e os perigos das inteligências artificiais generativas? Com o intuito de respondê-la, percorremos duas seções e chegamos às seguintes conclusões parciais abaixo.

Na primeira seção foram expostos alguns conceitos, que podem ser considerados ainda rudimentares, surgidos nas pesquisas sobre a sua natureza e o funcionamento dos LLMs. Nós denominamos estes conceitos, ainda rudimentares, de “proto-conceitos”. Eles são os seguintes: “inteligência”, “consciência situacional”, “autocurrículo” e “autoconsistência”. Neste mesmo sentido, destacamos algumas afirmações, nos artigos citados, sobre “inteligência”, “raciocínio”, “planejamento” e “criação”, aplicados para interpretar simulações e testes em modelos. A partir deste pano de fundo inicial, apresentamos os conceitos de Simondon sobre “objeto técnico”, sendo que o mesmo é dividido em três possibilidades: “elemento técnico”, “indivíduo técnico” e “conjunto técnico”. Neste quesito, incluímos o conceito de “processo de concretização”, o qual acompanha estas três possibilidades na passagem de um “estado mais abstrato” para um “estado mais concreto”. Tal cotejamento com a filosofia de Simondon nos permitiu compreender que as inteligências artificiais generativas estão em uma transição, sob a forma de proto-conceitos de autoconhecimento: indo do mais abstrato para o mais concreto, com resultados cada vez mais efetivos em termos de autonomia.

Na segunda seção discutimos qual seria o papel mais adequado para o ser humano diante das inteligências artificiais generativas, incluindo as complexidades dos fenômenos de emergência. Mediante uma leitura conceitual a partir de Simondon, temos que o ser humano possui duas funções perante os conjuntos técnicos: a função de ser assistente ou a função de ser organizador. Neste sentido, dentro do processo de concretização desenvolvido pelos conjuntos técnicos, o ser humano parece ficar com um papel cada vez mais distanciado. Alguns detalhes técnicos se tornam importantes neste momento, tais como, a dificuldade de verificação do funcionamento das inteligências neurais generativas por causa da constituição das redes neurais artificiais (*black box*). Para além desta dificuldade, o fato de existirem fenômenos de emergência, gera uma preocupação adicional em relação a um futuro seguro na convivência futura entre humanos e entidades individuadas providas de uma inteligência artificial. Enfim, diante do exposto nestas duas seções podemos agora responder à pergunta condutora desta investigação.

A resposta é afirmativa: o conceito de conjunto técnico de Gilbert Simondon abre possibilidades promissoras para interpretar e abordar o processo de concretização das inteligências artificiais generativas, além de acolher a possibilidade de autonomia crescente por parte destas inteligências.

A filosofia de Simondon está aberta a uma compreensão do tema da inteligência artificial, mesmo nas formas surgidas em tempos mais recentes. Esta filosofia possui uma base conceitual com o foco no objeto técnico, não excluindo a presença humana e provendo uma compreensão da distância que se abre entre o ser humano e o que é do âmbito técnico. Esta filosofia é adequada para refletir sobre os desafios propostos pelos processos de concretização da inteligência artificial na atualidade. E, especificamente, em relação aos perigos possíveis, a filosofia de Simondon reinterpretada, permite ao menos dimensionar o afastamento a que está submetido o ser humano em cada aumento de complexidade e nos fenômenos de emergência que se manifestam nos modelos inteligentes. Resta, à filosofia em geral, aprofundar este caminho, retomando as propostas de Simondon e com isso mais subsídios à reflexão. Isso certamente contribuiria para a futura convivência entre humanos e a inteligência artificial. Trata-se de uma tarefa que nos cabe e algo que nos distingue, pois a filosofia é uma prática essencialmente feita por humanos.

## Referências

BAKER *et. al.* Emergent tool use from multi-agent autocurricula. **Arxiv** - Cornell University (2023). Disponível em: <https://arxiv.org/abs/1909.07528> acesso em 29 de Setembro de 2023.

BERGLUND *et. al.* Taken out of context: On measuring situational awareness in LLMs. **Arxiv** - Cornell University (2023). Disponível em: <https://arxiv.org/abs/2309.00667> acesso em 02 de Setembro de 2023.

BETLEY *et. al.* Narrow Finetuning Can Produce Broadly Misaligned LLMs. **Arxiv**, Cornell University, 12 de Maio de 2025. Disponível em: <https://arxiv.org/pdf/2502.17424>. Acesso em 03.09.2025.

BUBECK *et. al.* Sparks of Artificial General Intelligence: Early experiments with GPT-4. **Arxiv** – Cornell University (2023). Disponível em: <https://arxiv.org/abs/2303.12712> acesso em 25 de maio de 2023.

CABRAL, Caio César. Elementos básicos da teoria da individuação de Gilbert Simondon. **Trans/Form/Ação**, Marília, v. 44, n. 2, p. 63-82, Abr./Jun., 2021

CLOUD *et. al.* Subliminal learning: language models transmit behavioral traits via hidden signals in data. **Arxiv**, Cornell University, 20 de Julho de 2025. Disponível em: <https://arxiv.org/pdf/2507.14805>. Acessado em 03.09.2025.

DOMINGOS, Pedro. **O algoritmo mestre**. São Paulo: Novatec Editora, 2017.

NEEDHAM *et. al.* Large Language Models Often Know When They Are Being Evaluated. **Arxiv**, Cornell University, 16 de Julho de 2025. Disponível em: <https://arxiv.org/pdf/2505.23836>. Acessado em 03.09.2025.

SIMONDON, Gilbert. **Do modo de existência dos objetos técnicos**. Rio de Janeiro: Contraponto, 2020.

SHEVLANE *et al.* Model evaluation for extreme risks (2020). **Arxiv** - Cornell University. Disponível em: <https://arxiv.org/abs/2305.15324> acesso em 27 de maio de 2023.